

Characterization of Binding Pocket Flexibility of Aldose Reductase

DISSERTATION

ZUR

ERLANGUNG DES DOKTORGRADES

DER NATURWISSENSCHAFTEN

(DR. RER. NAT.)

dem

Fachbereich Pharmazie der

PHILIPPS-UNIVERSITÄT MARBURG

vorgelegt von

Matthias Zentgraf

aus Fulda

Marburg an der Lahn, 2006

Vom Fachbereich Pharmazie der PHILIPPS-UNIVERSITÄT MARBURG als Dissertation

angenommen am: 15.11.2006

Erstgutachter: Prof. Dr. G. Klebe

Zweitgutachter: Dr. C. Sotriffer

Tag der mündlichen Prüfung: 16.11.2006

Table of Contents

Aldose Reductase and the sorbitol pathway.....	1
Introduction.....	1
Structure of AR and related enzymes.....	1
Mechanism and kinetics of the AR catalyzed reaction.....	2
The sorbitol pathway and its influences on metabolism.....	4
Aldose Reductase inhibitors.....	5
Why consider protein flexibility in structure-based drug design?.....	9
Addressing protein flexibility and ligand selectivity by "in-situ cross-docking".....	21
Introduction.....	21
Results and Discussion.....	22
Summary and Conclusion.....	28
Supporting Information A: Computational Methods.....	28
Protein setup and grid calculations.	28
Ligand setup.....	29
Docking	30
Supporting Information B: Docking to trypsin and thrombin.....	31
Supporting Information C: Docking to multiple conformers by grid-based averaging.....	32
Extending charted space: comparative MD simulations of Aldose Reductase.....	35
Introduction.....	35
Materials and Methods.....	38
Crystal structure analysis.....	38
MD simulations.....	41
Generation of 2D rms plots.....	42
Docking with AutoDock.....	42
Results and Discussion.....	43
Comparative crystal structure analysis.....	43
Comparative MD simulations analysis.....	63
Simulations of AR complexes.....	64
MD of AR in complex with sorbinil.....	64
MD of AR in complex with tolrestat.....	67
MD of AR in complex with idd 594.....	70
MD of AR in complex with the Pfizer compound.....	73
MD of AR in complex with the zopolrestat compound.....	76
MD of AR in complex with the 47d compound.....	78
MD of AR in complex with the JFD compound.....	81
Simulations of the holo enzyme.....	83
Summarizing the results.....	89
Implications for the binding pocket.....	104
Addressing the new binding pocket conformations.....	108
Summary and Conclusions.....	120

Evaluating MM-PBSA in case of a flexible binding pocket: the AR test case.....	122
Introduction.....	122
Methodology.....	123
Setup and workflow.....	124
Literature overview.....	129
The test system: Aldose Reductase.....	135
Materials and Methods.....	138
Data set.....	138
ITC measurements.....	138
Molecular Dynamics Simulations	140
MM-PBSA calculations.....	141
Calculation of Predictive Indices (PI).....	143
Results and Discussion.....	144
MD trajectories.....	144
Sampling the reference state.....	146
MM-PBSA results.....	151
Summary and Conclusions.....	159
Expect the Unexpected while working with AR.....	163
Introduction.....	163
Results and Discussion.....	165
Conclusions.....	176
Summary and Outlook.....	178
Zusammenfassung.....	185
Appendix A.....	193
The Molecular Dynamics Database (MDDDB).....	193
Introduction.....	193
Database and program architecture.....	194
Data content.....	194
Using MDDDB.....	195
Appendix B.....	197
Additional crystal structures for MD simulations.....	197
Structure Determination.....	197
The sorbinil complex structure.....	198
The tolrestat complex structure.....	199
Usage of crystal structures from collaboration partners.....	202
References.....	203
Publications arising from this work.....	223
Articles.....	223
Oral Communications.....	224
Posters.....	224
Acknowledgments.....	226
Curriculum vitae.....	228
Erklärung.....	229

Aldose Reductase and the sorbitol pathway

Introduction

Aldose Reductase (AR) is a cytosolic NADPH-dependent enzyme that catalyzes the reduction of various aldoses and aldehydes to the corresponding alcohols. It was first described in 1956 by Hers et al.¹ as a glucose-reducing activity. The authors demonstrated that the conversion of blood glucose to fructose was used as an energy source for sperm cells.

Today it is known that AR is the first and rate-limiting enzyme of the sorbitol pathway. Via this pathway glucose is first reduced to sorbitol by AR. In a second step sorbitol is then re-oxidized to fructose by an enzyme called sorbitol dehydrogenase (SDH) using NAD⁺ as cofactor. Galactose is an even better substrate for AR, but the corresponding product galactitol is not further metabolized by SDH^{2,3}. The net reactions of the sorbitol pathway for glucose are conversions of glucose to fructose as well as of NADPH to NADH.

Both enzymes are expressed in almost all human tissues^{4,5}. However, the ratio of AR to SDH differs in different tissues⁶. The kidney, for example, is one of the tissues showing the highest concentrations of AR⁷.

Under physiological conditions glucose is usually phosphorylated by the enzyme hexokinase to glucose-6-phosphat which is then further metabolized using glycolysis or the pentose phosphate metabolism. Under raised blood glucose levels, as found in people suffering from diabetes mellitus, up to one third of the available glucose is processed via the sorbitol pathway^{8,9}. The accumulation of sorbitol on the one hand and the consumption of NADPH on the other have been postulated to be responsible for some of the late-onset complications of diabetes mellitus¹⁰⁻¹⁶.

Structure of AR and related enzymes

AR belongs to the superfamily of the aldo-keto reductases (AKR). It comprises 315 amino acids and folds to an (β/α)₈-TIM barrel structure with the catalytic site deeply buried at the center of the barrel¹⁷. The protein core is composed of eight parallel β -strands. Adjacent

strands are connected by eight parallel α -helices running anti-parallel to the β -sheets. The active site is located at the C-terminal end of the barrel. The bottom of the barrel is closed by two short anti-parallel β -strands near the N-terminus. Three large loops partially cover the top of the barrel. Two additional α -helices are found outside the barrel structure preceding two of three additional loop regions in sequence. The cofactor NADPH binds at the bottom of the active site, with its nicotinamide moiety pointing towards the active site^{18,19}. The shape and properties of the binding pocket of AR as well as the binding modes of different inhibitors in complex with AR will be discussed in great detail in the chapter on Comparative Crystal Structure Analysis (see page 43).

Structurally related enzymes of the AKR family which also catalyze the pyridine nucleotide-dependent reduction of carbonyl functions are abundant in most organisms. However, the accepted substrates differ between the single enzymes. AR and the very closely related enzyme aldehyde reductase reduce aldo and keto sugars as well as aromatic and aliphatic aldehydes²⁰⁻²². Hydrosteroid dehydrogenase accepts steroids as substrates²³, and the AR like prostaglandin synthase is responsible for the production of prostaglandin f2 alpha²⁴. However, no strict rules can be applied for AR since substrate promiscuity is very pronounced for this enzyme. A broad variety of different substrates is accepted. Among these substrates are glutathiolated^{25,26}, phospholipid²⁷, saturated, and unsaturated²⁸ aldehydes as well as 4-hydroxyalkenals²⁸ and steroids^{29,30}. The best known physiological substrate for AR is the 2-oxoaldehyde methylglyoxal³¹⁻³³. In general, many AR substrates are more hydrophobic than expected for an enzyme involved in sugar metabolism.

Mechanism and kinetics of the AR catalyzed reaction

The reaction mechanism of AR follows a sequential binding of the involved molecules. The cofactor, NADPH, binds first and forms a binary complex with AR. Upon binding large movements in a protein region called the 'cofactor safety belt' are induced³⁴. Thus, the cofactor becomes deeply buried in the protein. Subsequently, the ternary complex of substrate, cofactor and protein, is formed. The substrate is kept properly aligned with respect to the cofactor by a network of hydrogen bonds with residues around the catalytic pocket. Then the hydride transfer takes place reducing the aldehyde to the corresponding alcho-

late. To finalize the reaction a proton needs to be transferred from the protein to the product. Three residues around the active site are at suitable distances to act as proton donor: Cys 298, His 110, and Tyr 48. Kinetic³⁵, computational³⁶, as well as thermodynamic³⁷ studies favor Tyr 48 as the source of the proton. After the product has been released, exchange of the cofactor is necessary to enable the next catalytic cycle. Therefore, the conformational change which occurred during binding of the cofactor needs to be reversed. Several kinetic studies³⁸⁻⁴⁰ suggest that removing the cofactor from its binding site after the reaction is indeed the rate limiting step of the whole process.

Compared to other substrates, glucose is a rather poor substrate of AR having K_m values in the range of 50-200 mM which is much higher than physiological concentrations^{41;42}. However, less than 0.1% of glucose exists in the acyclic carbonyl form which is mandatory for binding to AR. It is, thus, conceivable that the enzyme processes significant amounts of glucose only if the glucose concentrations are pathologically high. Surprisingly, it has been shown that AR catalyzes the reduction of saturated and unsaturated medium- to long-chain (C-6 to C-18) aldehydes, which are generated during lipid peroxidation, with 10^3 to 10^4 -fold higher efficiency than glucose^{28;43;44}.

Several studies present results on post-translational modifications of AR to regulate the activity of the enzyme⁴⁵⁻⁴⁹. These modifications are in part held responsible for the reduced efficiency of hydantoin-based AR inhibitors such as sorbinil^{12;47}. Furthermore, it was observed that depending on the presence or absence of reducing agents the kinetic parameters of AR differed significantly^{50;51}. The same effects were encountered upon *in vitro* thiol modifications using different agents^{43;52-55}. Oxidative modifications⁵⁵ and glutathiolation^{42;53} of Cys 298, which is located directly at the binding site are deemed responsible for these effects. Thus, Cys 298 is considered to be a modulator for AR activity. Depending on the conditions of the reactions, AR becomes either S-thiolated (inactivated) or S-nitrosated (activated). Therefore, nitric oxide (NO) is thought to be the physiological regulator of AR activity^{56;57}.

If NADPH is bound, the Cys 298 side chain is less prone to oxidative modifications. Due to the high binding affinity of NADPH to AR most of the enzyme in the cell will be in the complexed state. Exchange of the cofactor will only take place if substrate concentra-

tions have reached a level where the catalytic reaction can occur. Thus, the switch for regulating the AR activity by modifying the Cys 298 side-chain will only be available at situations of high substrate concentration.

The sorbitol pathway and its influences on metabolism

Several hypotheses have been discussed with respect to the question how the activation of the sorbitol pathway may induce diabetic complications. Oxidative stress is one of the major issues in this context. The sorbitol pathway contributes to oxidative stress in different ways:

- The increased activity of AR depletes its cofactor NADPH. A sufficient concentration of NADPH is required in cells due to its role as cofactor of glutathione reductase (GR). GR is responsible for several critical reductive metabolic steps, such as the detoxification of reactive oxygen species (ROS) and hydroperoxides. Thus, by consuming NADPH, AR weakens the ability of cells to protect themselves from oxidative stress.
- In the second step of the pathway sorbitol is oxidized to fructose by SDH. Thereby NAD^+ is reduced to NADH. However, NADH is the substrate of NADH oxidase, which in turn produces additional ROS⁵⁸.
- The conversion of glucose to fructose is also problematic in itself, since fructose and its metabolites are more potent non-enzymatic glycation agents than glucose. The formed glycation products of these reactions are known to cause oxidative stress within cells.

The accumulation of sorbitol within cells and the resulting increase of osmotic stress is the second hypothesis discussed as putative explanation for the occurrence of diabetic complications. Since sorbitol can be oxidized by SDH to fructose, which then re-enters the glycolytic pathway, accumulation of sorbitol occurs primarily in cells and tissues where only low levels of SDH are present. For diabetic cataract it could be demonstrated that osmotic stress is the main reason for this complication^{59;60}. However, it remains unclear whether this mechanism can be transferred to other diabetic pathological phenomena.

Another factor, contributing in particular to diabetic vasculopathy, is the activation of protein kinase C (PKC) which is associated with the sorbitol pathway. PKC is activated by diacylglycerol (DAG)⁶¹. The concentration of the latter is elevated by raised blood glucose levels in many tissues. Several studies have shown that the elevated activation of PKC under diabetic conditions could be prevented by AR inhibitors⁶²⁻⁶⁴. This might be due to the fact that fructose metabolites are precursors of the DAG synthesis⁶⁵. Furthermore, there are hints that PKC is activated by oxidative stress⁶⁶.

Despite the large amount of knowledge about putative patho-physiological effects, the exact mechanism by which AR contributes to the development of diabetic complications remains unclear. Concerning the physiological function of AR, two main hypotheses are discussed in literature:

- osmoregulation in the kidney: sorbitol is one of the organic osmolytes that balance the osmotic pressure of extracellular hyperosmotic fluids during antidiuresis⁶⁷⁻⁷⁰. Considering the reported increased expression of AR under hyperosmotic stress⁷¹⁻⁷³, it can be assumed that the enzyme plays an osmoregulatory role in the renal homeostasis.
- detoxification and housekeeping: the broad distribution of AR over different tissues combined with the already discussed wide spectrum of substrates accepted by the enzyme suggest a role as a general detoxifying enzyme. In addition, AR might be involved in metabolizing environmental toxins⁷⁴ as well as some drugs⁷⁵.

Again, despite a number of intriguing studies, the beneficial role of AR under physiological conditions has not yet been entirely clarified.

Aldose Reductase inhibitors

Aldose Reductase inhibitors (ARI) have received considerable attention as potential therapeutic agents preventing for diabetic complications. A large variety of structurally diverse compounds have been identified as potent *in vitro* ARI. Currently known ARI can be divided into different classes according to their structure^{76;77}:

- acetic acid derivatives;

- cyclic imides (hydantoins);
- phenolic derivatives ;
- pyridazinones;
- phenylsulfonylnitromethane derivatives;
- structurally diverse ARI from natural sources

Generally speaking, compounds derived from acetic acid show the highest *in vitro* affinity towards AR. However, in many cases these compounds lack suitable *in vivo* potency. In contrast, the hydantoin-based molecules are superior concerning pharmacological efficacy. This behavior is usually attributed to the differing acidity of the two head groups. Carboxylic acids are fully deprotonated under physiological pH conditions. Thus, their ability to penetrate biological membranes is limited. Hydantoins, however, have pK_A values around 9. Thus, at physiological pH the protonated as well as the deprotonated forms are present in considerable amounts. Therefore, hydantions can more easily cross biological membranes in their uncharged protonated form, whereas the deprotonated polar form is needed to bind to the enzyme.

More than ten drugs have reached phase II or III clinical trials (see Figure 1). All of them belong either to the carboxylic acid or the hydantoin groups. Only one ARI (epalrestat) is currently on the market. Most of the ARI were withdrawn before reaching the market due to inadequate therapeutic potential or serious side-effects. The results for many of the clinical trials have recently been reviewed by Hamada et al.⁷⁸ and will be summarized shortly in the following section.

The first ARI taken to clinical trials in the 1970s was alrestatin. Despite promising initial results, the development of alrestatin was stopped due to side-effects^{79;80}. Numerous studies were carried out for sorbinil⁷⁸. The compound failed to prove its efficacy in larger, statistically more significant studies. Again, side-effects prevented the compound from reaching the market⁸¹. Tolrestat made it to the market in Italy and other countries for the treatment of diabetic neuropathy⁸². However, in later studies tolrestat failed to confirm its efficacy⁸³ and was suspected to cause hepatic damage⁸⁴. Therefore, the compound was withdrawn. The clinical studies for ponolrestat were disappointing. This was probably due

to the poor penetration of the drug into human tissues⁸⁵. Zopolrestat reached phase III of clinical trials⁸⁶. However, the results of this study were never reported and further trials have been stopped. For zenarestat clinical trials were carried out in the US, Japan, and the UK. Again the compound looked promising, reached phase III, and was withdrawn without publishing the results of the final study⁸⁷. For fidarestat a phase III trial was finished with encouraging results. However, the approval is still pending in Japan because of incomplete data⁸⁸. Three additional ARI are in phase II or III clinical trials: NZ-314⁸⁹, AS-3201⁹⁰, and minalrestat⁹¹. Epalrestat is the only ARI currently on the market and has been available for more than ten years by now. The results of several studies have supported the usefulness of epalrestat⁹²⁻⁹⁴. However, the duration of the studies may not have been long enough to assess alterations in nerve functions. Only one longer study has been carried out, where it was shown that epalrestat minimized progression of early sensory deficits⁹⁵.

In summary the data from the different studies do not provide a conclusive picture about the clinical potential of ARI. While some compounds, such as tolrestat and epalrestat, showed promising results, others failed to show beneficial effects. Clearly, there is still the need for the development of further ARI with better pharmacological properties⁷⁸.

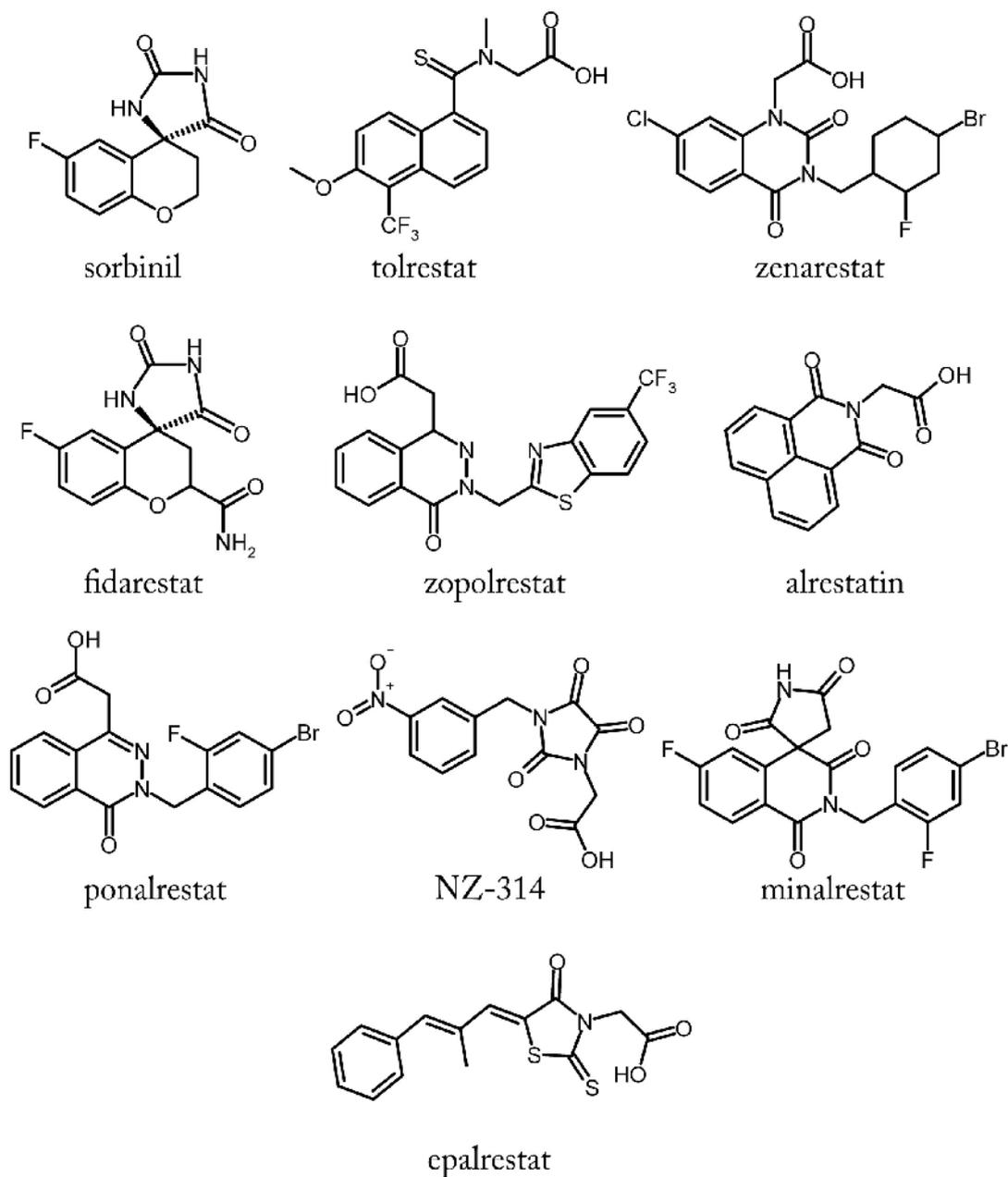


Figure 1: AR inhibitors which reached phase II or III of clinical trials for diabetic neuropathy. Only epalrestat is currently on the market in Japan.

Why consider protein flexibility in structure-based drug design?

Proteins are dynamic systems. Their plasticity is a key characteristic of their function. To underline this hypothesis Eisenmesser et al.⁹⁶ were able to show by NMR experiments that the characteristic enzyme motions necessary for catalysis are already present in the free enzyme and, thus, are an intrinsic property of the enzyme. Of course, inherent flexibility has consequences for the rational design of small molecules as inhibitors of protein function⁹⁷⁻⁹⁹. Therefore, it is crucial to study and characterize the flexibility exhibited by a target enzyme of a structure-based drug design (SBDD) project.

In recent years various attempts have been made to include protein flexibility into one of the most important steps of a SBDD project: the docking of small molecules into the binding site of a target enzyme¹⁰⁰⁻¹¹². The fundamental problem when considering protein flexibility is the enormous amount of degrees of freedom available to a protein. Therefore, all suggested methods rely on assumptions and simplifications to keep the calculation times within reasonable boundaries for routine use in an industrial or academic environment. As a consequence, these methods only include a limited range of flexibility of the target protein. Rotamer libraries, ensembles of pre-generated structures or different techniques to average multiple conformations are used to cope with side-chain flexibility. Movements of the main chain or larger scale domain movements are usually not within the scope of these methods.

Knowing that current methods to consider protein flexibility are of reduced value for every day use, the question arises, whether one should deal with this problem at all. If a high-quality crystal structure of a certain protein conformation is available, then why not take this structure as a reference and start designing small molecules which address this pocket conformation?

Aldose Reductase (AR) is a good example for an enzyme which shows pronounced flexibility of its binding pocket and where many high-resolution structures are available. It exhibits an 'induced-fit' binding mechanism where three major binding-pocket conformations have been described¹¹³. A previous virtual screening approach with subsequent docking experiments which focused on the ultra-high resolution complex crystal structure with

idd 594¹¹⁴ resulted in hits which were active in the low micromolar range¹¹⁵. The binding mode determined later by means of X-ray crystallography is in good overall agreement with the proposed binding mode¹¹⁶ from the docking studies using FlexX¹¹⁷. Thus, this project can be considered a success, since new inhibitors were discovered for a flexible enzyme using only one of the pocket conformations as template. However, the idea that one pocket conformation is sufficient to correctly predict the binding mode of AR inhibitors is misleading. As with other flexible binding pockets, unexpected changes of the overall pocket conformation can occur, leading to a binding mode totally different than expected.

To emphasize the need for including multiple discrete pocket conformations into the docking process, one example for such an 'undesigned' binding mode shall be discussed briefly.

Da Settimo et al.¹¹⁸ published a series of AR inhibitors which were designed as potential analogues of the long known AR inhibitor tolrestat¹¹⁹, whose complex structure was solved already in 1997¹²⁰. This inhibitor induces a conformation of the binding pocket which is still unique until today. The two most potent inhibitors **1** and **2** of the series inhibit AR with IC₅₀ values of 140 and 550 nM, respectively (see Figure 2). The only difference between these two compounds is a different side chain in the 4-position of the ring scaffold.

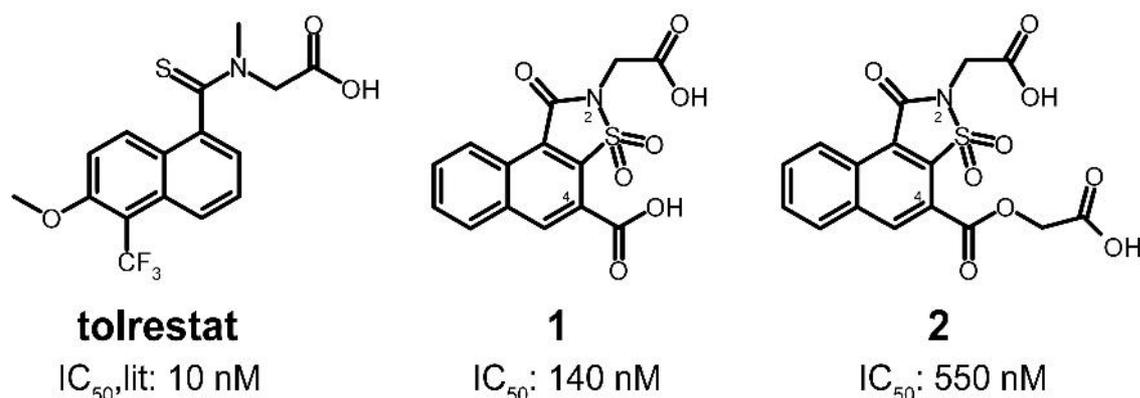


Figure 2: Chemical formulae of the AR inhibitor tolrestat and the two most most potent compounds from the series of Da Settimo et al.¹¹⁸. IC₅₀ values for 1 and 2 are taken from the publication of De Settimo¹¹⁸, whereas the tolrestat value was published by Urzhumtsev et. al¹²⁰.

The hypothesis that **1** and **2** induce a binding-pocket conformation similar to tolrestat is based on two observations: First, the high structural similarity to tolrestat analogs published by Wrobel et al. in 1990¹²¹. In these compounds, the thioamide nitrogen of tolrestat is inserted into rigid cyclic structures. However, no complex crystal structure of AR with any of these analogs is available. The second observation is the good superposition of compounds **1** and **2** with the protein-bound conformation of tolrestat (see Figure 3 A).

To support the hypothesis of a binding mode of **1** and **2** similar to tolrestat, docking experiments were carried out using AutoDock 3.0.5¹²². The complex crystal structure of porcine AR and tolrestat (pdb code: 1ah3¹²⁰) was used as template and both inhibitors **1** and **2** were docked into this structure. Subsequently, the best docking solutions for each compound and the side chains of the corresponding binding-pocket residues were subjected to an energy minimization to further optimize the binding geometries. The resulting binding modes are shown in Figure 3 B (**2**) and Figure 3 C (**1**). In both cases the carboxylic function added in 2 position of the ring system addresses the catalytic subpocket. The overall binding mode of these docking solutions is indeed similar to the one of tolrestat. Thus, the authors conclude that this tolrestat-like binding mode is the most likely for their new AR inhibitors which is reasonable looking at the experiments carried out.

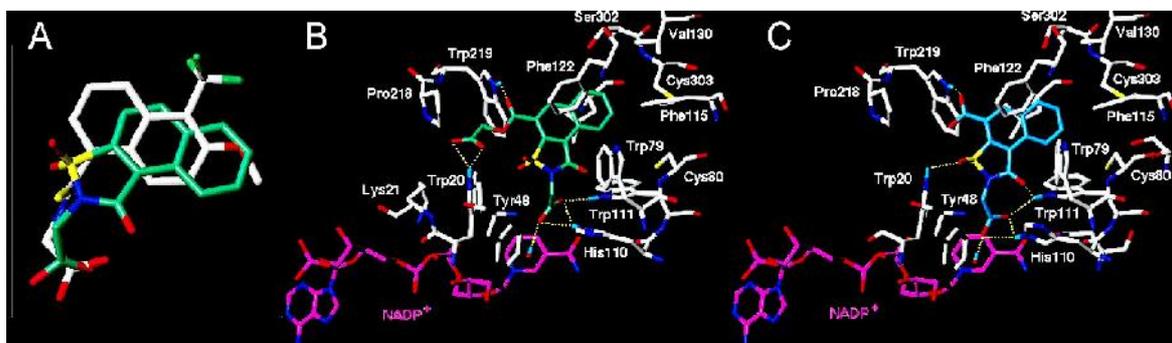


Figure 3: *A* shows the superposition of **1** (green) onto the AR-bound conformation of the inhibitor tolrestat (white). *B* and *C* show the docking solutions for **2** (*B*) and **1** (*C*) with the PDB deposited structure of tolrestat (1ah3). All pictures were taken from Da Settimo et al.¹¹⁸.

However, the binding pocket of AR can adopt at least three different conformations (see page 43) and AutoDock by itself can not handle protein flexibility. Thus, the program tries to fit the inhibitor into the predefined binding-pocket conformation and neglects all other

possible binding modes. It is, therefore, instructive to dock both inhibitors not only to the pocket conformation of tolrestat but also to the two other main conformations. Besides tolrestat, the sorbinil and the idd 594 complexes were chosen to represent these two additional AR binding-pocket conformations. Sorbinil binds to the closed form of the specificity pocket and idd 594 addresses the specificity pocket, but in a different conformation than tolrestat.

To allow best comparison with the results from Da Settimo et al., AutoDock¹²² 3.0.5 was used to dock **1** and **2** into all three pocket conformations. The search- and scoring grid for AutoDock was centered on the AR binding pocket. The grid size was set to 46 X 50 X 54 points using a spacing of 0.5 Å. 100 runs of the Lamarckian genetic search algorithm were performed using a population size of 50 and an upper limit for the number of energy evaluations of 1.5×10^6 . The parameters for mutation and cross-over were kept at their default settings of 0.02 and 0.80, respectively. The local energy minimization algorithm was limited to 300 steps.

The results of the docking experiments for compound **1** are presented in Figure 4. Shown is the top-scoring docking solution for each pocket (sorbinil A (green), tolrestat B (silver), idd 594 C (yellow)), the corresponding AutoDock energy score, and a superposition of the docking solution with the bound crystallographic conformation of the corresponding inhibitor. Reasonable docking solution of **1** are found for each of the binding-pocket conformations. Surprisingly, the docking solution in the tolrestat pocket (B) is scored worst, obtaining a score almost 1kcal/mol less favorable compared to the solution in the other two pockets. The solutions addressing the sorbinil and idd 594 pocket are scored equally. Furthermore, for the sorbinil pocket the carboxylic acid in 4-position of **1** addresses the catalytic pocket. This is in contrast to the original design hypothesis where the carboxylic acid in 2-position is used for this purpose. Looking at the superpositions in the lower part of Figure 4, it can be seen that for the sorbinil pocket the docking solution and the bound conformation of the inhibitor show a very reasonable overlap. The carboxylic moiety addressing the catalytic pocket occupies the same point in space where the nitrogen and an oxygen of the hydantoin moiety of sorbinil are located, thus enabling the same interactions with surrounding residues. For the other two cases the fit of the polar head groups is not as convincing. Although in both cases the carboxy function in 2-position is

placed into the anion pocket, the conformations are different and the interaction geometries with the catalytic residues are not ideal. The rigid aromatic scaffold of **1** superimposes well with sorbinil and tolrestat. The fit with idd 594 is not convincing, since the corresponding ring systems are rotated by almost 45° against each other.

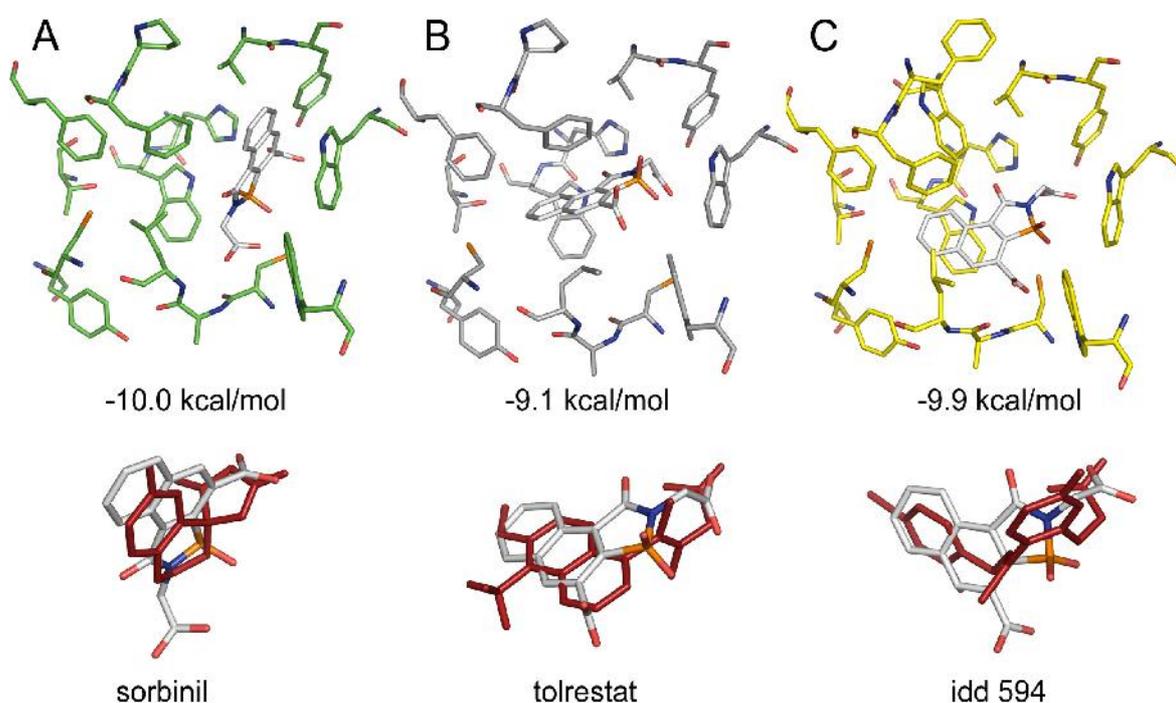


Figure 4: Results for docking compound 1 into different binding pocket conformations formed with: sorbinil (A), tolrestat (B), and idd 594 (C). The best scored docking solutions for each pocket are shown in the upper part of the figure. In addition, the corresponding scores are given as computed by the scoring function implemented into AutoDock. In the lower part of the figure superpositions of compound 1 with the crystallographically determined binding mode of the bound inhibitor in the corresponding binding pocket are shown.

As briefly mentioned, the only difference between compounds **1** and **2** is the length of the side chain in 4-position of the ring system. The corresponding docking results for compound **2** are shown in Figure 5.

As with compound **1** the docking solution in the sorbinil pocket is scored best, followed by the one in the idd 594 pocket. Again, docking into the tolrestat pocket results by far in the worst energy score. Compared to **1**, the best docking solution in the sorbinil pocket has

been flipped, and it is now the carboxylic function in 2-position which addresses the anion pocket. For the tolrestat pocket, the best-scored solution has none of its carboxylic acids placed into the anion pocket, which is rather unlikely, since addressing this pocket is mandatory for every AR inhibitor. While the docking solution in the idd 594 pocket addresses the anion pocket, it places the second carboxylic-acid moiety into the hydrophobic specificity pocket, a fact which is very unlikely, in particular, since this part of the binding pocket is certainly not very suitable to accommodate a carboxylic function. In the sorbinil pocket the docking solution addresses the anion pocket in a favorable geometry with respect to possible interaction partners, and the second carboxy function points towards the solvent. The superpositions in the lower part of Figure 5 reflect these findings very well. For sorbinil a good superposition is obtained, especially with respect to the polar head groups. For tolrestat and idd 594, however, the superpositions are less convincing.

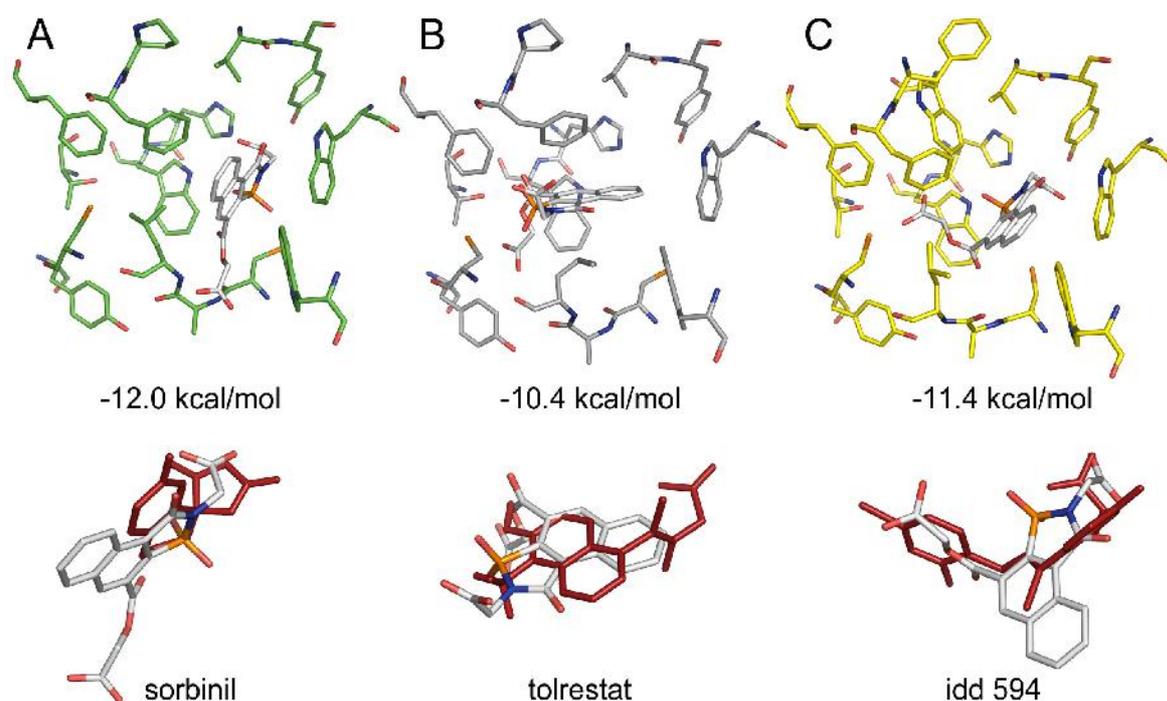


Figure 5: Results for docking compound 2 into different binding pocket conformations formed with: sorbinil (A), tolrestat (B), and idd 594 (C). The best scored docking solutions for each pocket are shown in the upper part of the figure. In addition, the corresponding scores are given as computed by the scoring function implemented into AutoDock. In the lower part superpositions of compound 2 with the crystallographically determined binding mode of the bound inhibitor in the corresponding binding pocket are shown.

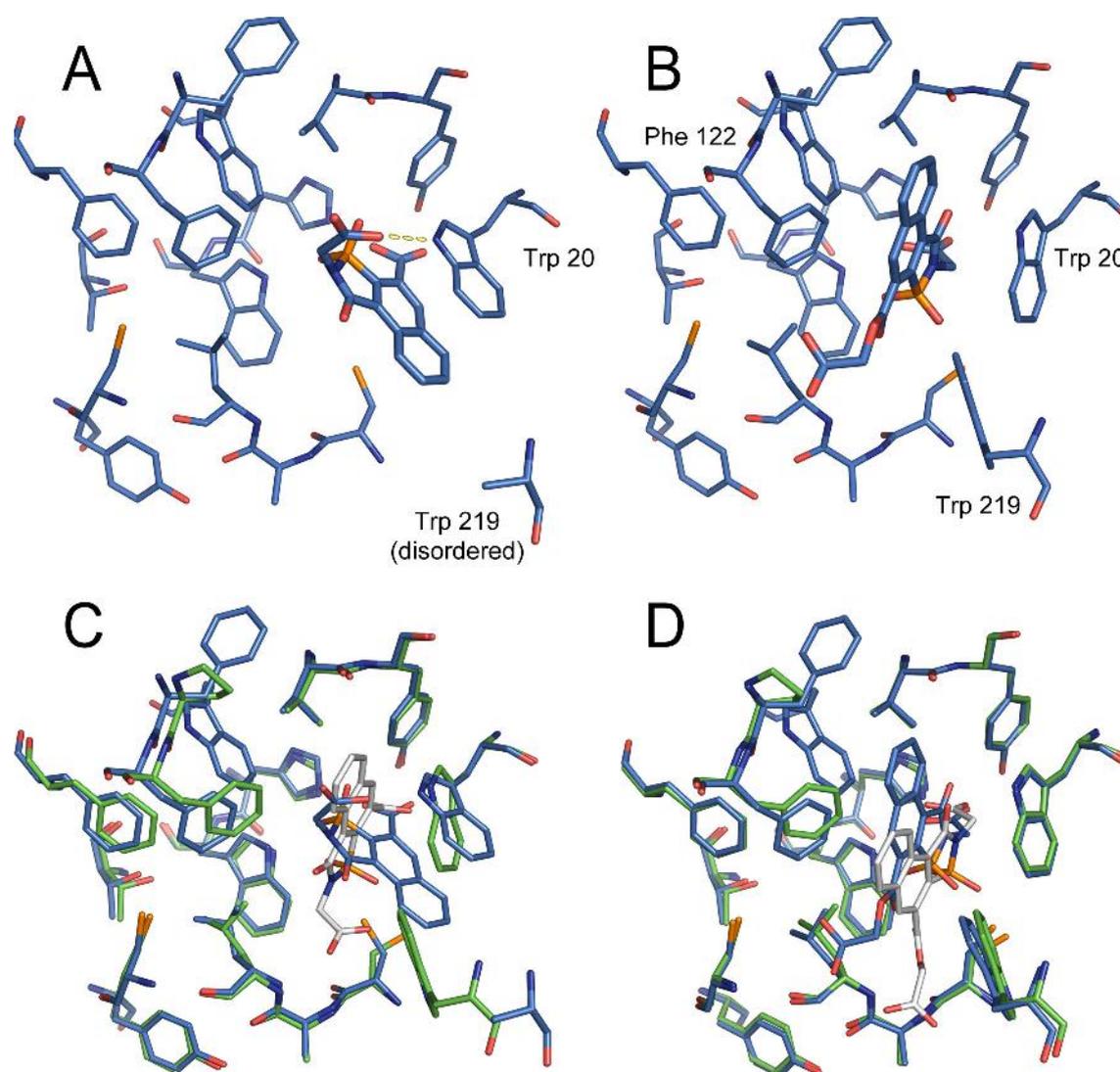


Figure 6: In the upper part the crystal structures of compound 1 (A) and 2 (B) are shown in marine blue. In the lower part of the figure superpositions of the best docking solutions and the crystal structures are shown (compound 1: C, compound 2: D) The protein conformation used for docking is shown in green, the crystal structures including the ligand are shown in marine blue. The corresponding best scored docking solution is shown in silver.

In summary, based on these docking experiments to multiple binding-pocket conformations, neither for compound 1 nor 2 a binding mode similar to tolrestat seems to be the most likely solution. Instead, a binding mode similar to sorbinil where the specificity pocket is in closed conformation, is favored by AutoDock and supported by visual inspection.

Fortunately, crystal structures for both compounds in complex with AR could be solved in-house¹²³ (Figure 6 A for **1**, Figure 6 B for **2**). Since the structures themselves are not part of this work, they will be discussed only as reference to validate the results obtained by the docking experiments.

The most important finding is that in both cases the specificity pocket is in a closed conformation, resulting necessarily in a different binding mode than adopted by tolrestat. Instead, the binding pocket is more similar to the sorbinil-bound conformation as indicated by the docking experiments.

For compound **1** it is indeed the carboxylic function in 4-position which addresses the anion pocket (Figure 6 A). This is also in good agreement with the docking results in the sorbinil pocket. The carboxylic moiety of the 2-position points towards the solvent and forms an H-bond with the side-chain nitrogen of the nearby Trp 20, which rotates around its χ_2 angle by $\sim 30^\circ$ compared to all other known AR structures. Thus, **1** induces a new conformation of the anion-binding pocket. Alternative conformations for the anion pocket have not been seen so far for AR, except for the structure of the apo enzyme (pdb code: 1xgd)³⁴. Instead of pointing upwards as in the docking solution, the aromatic ring system of **1** points downwards (Figure 6 C). Besides the rotation of Trp 20 it displaces Trp 219, which becomes disordered. This binding mode of **1** is incompatible with all previously known AR conformations since it would clash with Trp 20 and Trp 219 in all other structure. Interestingly, among the 100 docking solutions from AutoDock there are 15 solutions where the aromatic ring system of **1** points downwards similar to the crystal structure and the 2-carboxylic function forms a hydrogen bond with Trp 20. However, the first of these solutions appears on rank 75 with an energy score of -9.6 kcal/mol, and the geometries do not match perfectly (see Figure 7 A).

Compound **2** binds as intended with the 2-carboxylic function into the anion pocket (see Figure 6 B). The overall pocket conformation is very similar to the one of the sorbinil complex. Phe 122 is the only binding-site residue which has to shift in order to accommodate the ligand. Again, this pocket conformation was correctly predicted by AutoDock to be the most favorable for accommodating ligand **2**. The other carboxy function points towards the solvent and forms two H-bonds with the backbone N and the side-chain OG of Ser 302.

Comparing the docking solution and the crystal structure of **2** shows that the position of the carboxylic function addressing the anion pocket is very well predicted. This is also true for the sulfon moiety of **2**. The entire ring system is shifted upwards in the crystal structure and the side chain in 4-position points in a slightly different direction. Three positions where ligand oxygens are placed in the docking solution are occupied by water molecules in the crystal structure (see Figure 7 B). This implies that these are good positions for hydrogen bonds and, thus, the placements of the oxygens of the docking solutions are reasonable from this point of view.

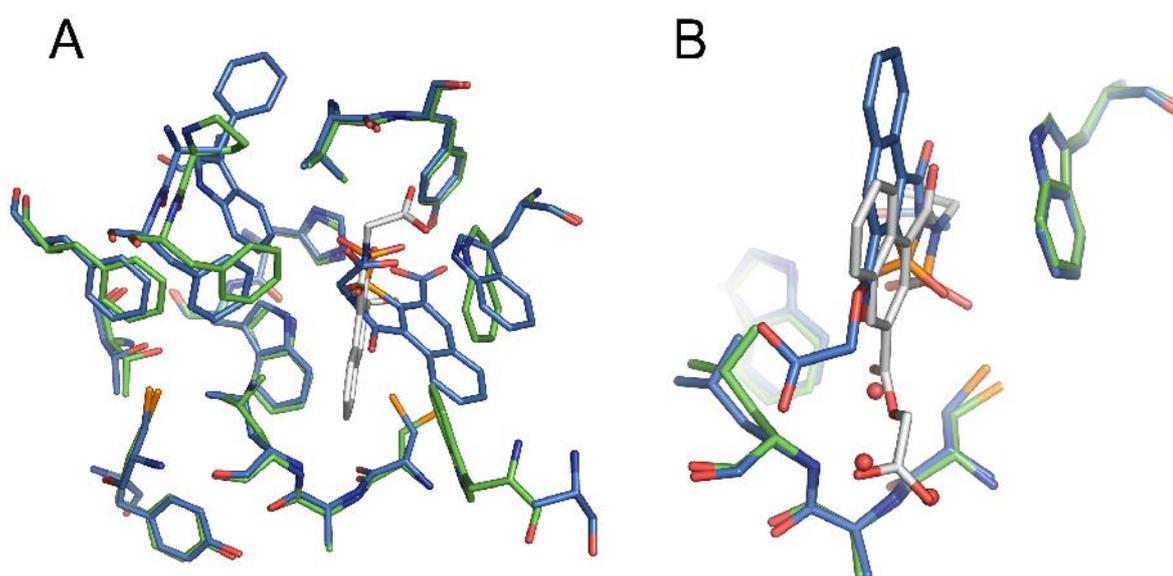


Figure 7: A shows the docking solution on rank 75 of compound 1 (silver) in the sorbinil-pocket conformation (green) compared to the crystal structure (marine blue). The ring moiety of the docking solution of 1 points downwards instead of upwards as in the more favorably scored docking result. This binding mode is in better agreement with the crystal structure. B shows a superposition of the best-scored docking solution from compound 2 (silver) and the crystal structure (marine blue). The spatial positions occupied by the 4-position side chain oxygens in the docking solution show waters nearby in the crystal structure.

In summary, this example clearly demonstrates why it is important to include knowledge about protein flexibility in a structure-based drug design project. In contrast to the failure to predict the binding modes of the two compounds using only one crystal structure

as template, the favored binding pocket conformation was predicted correctly by using multiple protein conformations as templates for docking. The overall binding mode for compound **2** has been predicted reasonably well. Deviations were seen between docking solution and crystal structure that result in consequence to a slight movement of a phenyl alanine side and the attempted placement of the polar ligand substituent onto positions in the binding pocket previously occupied by water molecules. In addition, the exchange of the carboxylic function addressing the anion pocket for compound **1** was also predicted in the docking experiments. The strong conformational changes of the anion pocket and the consequences for the binding mode of compound **1** would have been hardly possible to predict by any method available today.

The question remains, whether using the corresponding native binding-pocket conformations of **1** and **2** as templates for docking leads to improved binding-mode predictions. This would indicate that a suitable binding-pocket conformation is indeed a mandatory prerequisite to obtain reasonable binding modes from a docking experiment. In Figure 8 A the results for docking compound **1** into the corresponding crystal structure are shown. Due to an unusual conformation in the complex crystal structure of the N2 atom of compound **1**, which adopts a tetrahedral geometry in the crystal structure rather than a planar one, experiments for two different ligand setups were carried out. For the first docking experiment the ligand geometry extracted from the pdb file (silver, tetrahedral N2 geometry) was used unaltered, whereas for the second experiment the ligand (yellow, planar N2 geometry) was subject to an energy minimization prior to carrying out the actual docking experiment. However, only small deviations in the placement of the side chain attached to the N2 atom can be observed in the results for the two different setups, whereas virtually no differences are observed for the rest of the molecules. Using 1 Å rms deviation as cluster criterion, all 100 generated docking solutions fall into the same cluster in both cases and none of the solutions has an rmsd value higher than 0.82 Å with respect to the crystal structure. Therefore, it can be concluded that predicting the binding mode of **1** correctly is feasible given the correct binding-pocket conformation.

The binding mode obtained for docking compound **2** into its crystallographically determined binding-pocket conformation (Figure 8 B) is very similar to the one resulting from docking to the sorbinil complex pocket conformation (Figure 6 D). This finding is reason-

able, since there are only small deviations between the two complex binding pocket conformations of **2** and sorbinil. However, if the three water molecules described in Figure 7 B are included in the docking process as part of the binding pocket, the docking solutions improve significantly. The 88 top-ranked solutions show an rmsd below 1 Å with respect to the crystal-structure binding mode, thus also the binding mode for **2** can be correctly predicted if the actual state of the binding pocket is known.

On the one hand, this highlights the need for better methods to cope with protein flexibility and the treatment of water; on the other hand, one should always be aware of unexpected changes in parts of the protein which had previously been regarded as rigid.

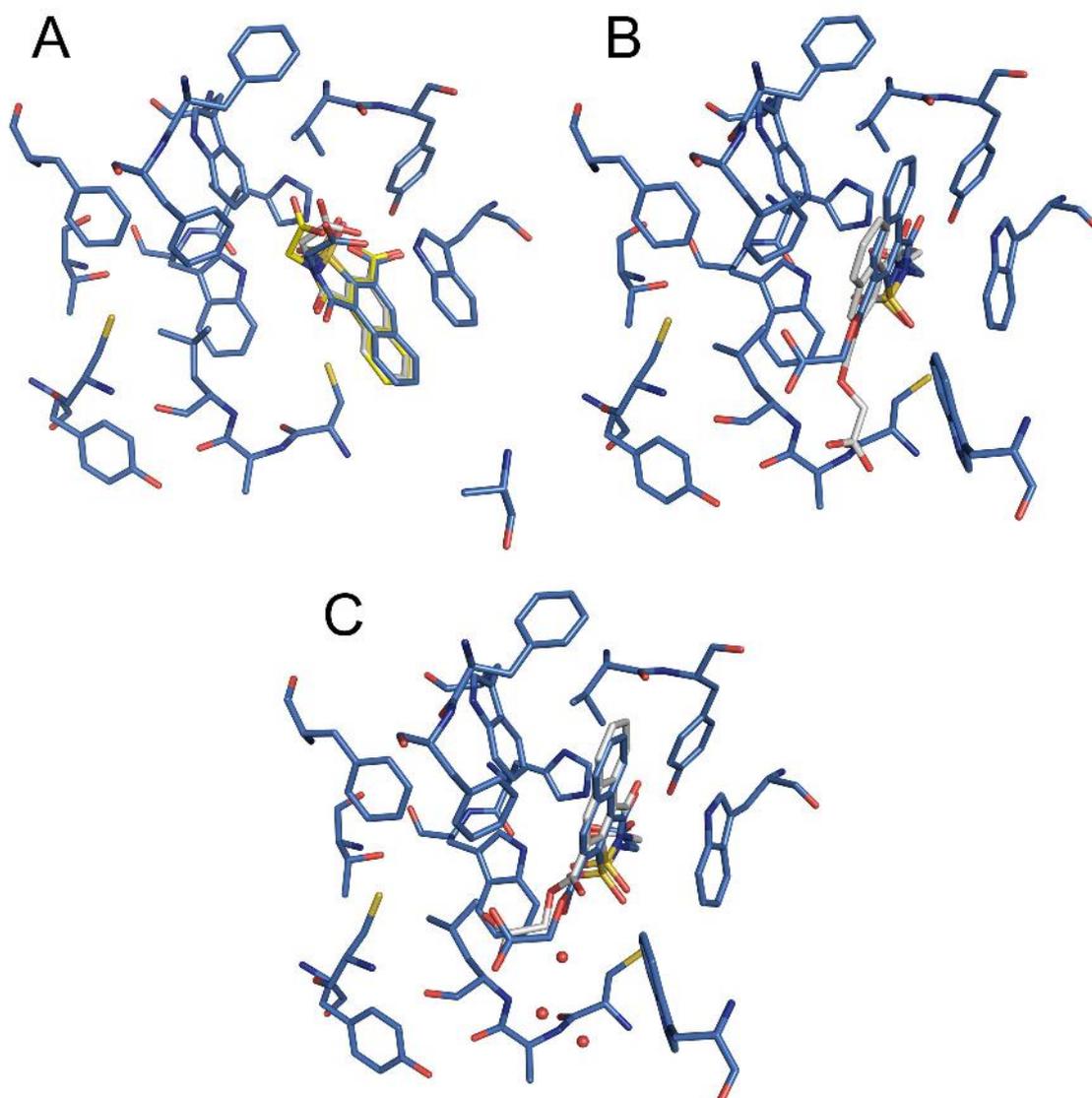


Figure 8: A shows the best scored docking solutions obtained from docking compound 1 into the corresponding binding-pocket conformation obtained from crystallography (marine blue). Results for two different ligand setups are presented. The conformation shown in silver was obtained for using the ligand as extracted from the pdb file as input for docking. The ligand shown in yellow was subject to an energy minimization prior to carrying out the docking experiment. B depicts the top-scored docking results for docking 2 to its crystallographically determined binding-pocket conformation, respectively. C shows the best-scored docking solution for 2 which is obtained, if the three water molecules occupying the similar spatial positions in the complex crystal structure of 2 where the side chain in 4-position is placed in B, are treated as part of the protein during docking.

Addressing protein flexibility and ligand selectivity by "in-situ cross-docking"

Introduction

To overcome the "single-structure-paradigm" in current methods for computational protein-ligand docking, we have recently introduced the "in-situ cross-docking" (ISCD) approach to simultaneously address multiple targets,¹²⁴ using the grid-based AutoDock program as search engine.¹²² While the feasibility of ISCD for dealing with well-differing binding sites and non-cross-reactive, tight-binding ligands had been demonstrated, it remained to be shown whether it could also be applied to different conformations of the same target (to address protein flexibility) or to closely related targets binding the same ligand with varying affinity (to address selectivity). Here, we investigate the first issue using aldose reductase as a test case, and the second using a recently introduced series of thrombin and trypsin inhibitors.

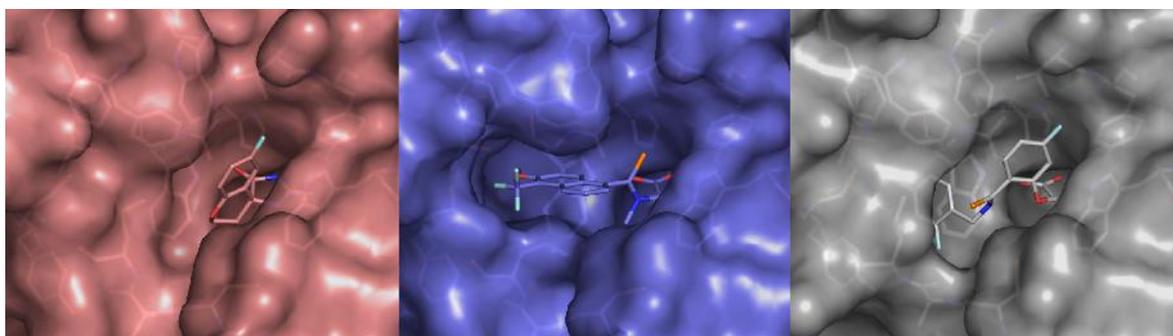


Figure 9: Comparison of the three different AR binding-site conformers determined by crystal structure analysis, as exemplified by the complexes with sorbinil (red), tolrestat (blue), and IDD549 (grey), seen from an identical perspective.

Aldose reductase (AR), a target against late-onset diabetic complications, catalyzes the reduction of aldoses and other aldehydes to the corresponding alcohols.¹⁴ Its substrate promiscuity is in part due to an "induced-fit"-like mechanism of ligand binding, whereby a specificity pocket can be closed or opened in different conformations, depending on the ligand being bound.¹¹³ Crystal structures of AR-inhibitor complexes have revealed three

major binding-pocket conformations, best represented by the complexes with sorbinil (PDB 1AH0), tolrestat (1AH3), and IDD594 (1US0) as shown in Figure 9.

For docking and structure-based ligand design AR poses the obvious problem that a single conformation of the protein is not sufficiently representative as target structure; instead, at least the three major conformations need to be addressed. In standard docking, this would be done sequentially, using each protein conformer for separate docking simulations, thus requiring for each ligand as many separate simulations as there are protein conformers to investigate. With ISCD, instead, the conformers can be combined to a single search space such that only one simulation must be run per ligand.

Results and Discussion

Using 1AH0, 1AH3, and 1US0 as structures for the three AR binding-site conformers, separate AutoDock grids were first calculated for each of them (further details about the methods are provided as Supporting Information A) and AutoDock runs were carried out on the separate single grids, proving that the experimental binding mode can indeed be reproduced by standard flexible docking (sorbinil to 1AH0: docking result on rank 1 shows a root-mean-square deviation, RMSD, of 0.26 Å with respect to the crystal structure; tolrestat to 1AH3: 0.94 Å; IDD594 to 1US0: 1.39 Å for rank1, 0.84 Å for rank 2).

Ligand	RMSD [Å]	E_d [kcal/mol]	Cluster size	$N \leq 2 \text{ Å}$ RMSD
Sorbinil	0.26	-9.38	96	96
Tolrestat	1.00	-11.77	35	51
IDD594	0.84	-12.16	5	6
Fidarestat	0.51	-9.58	52	99
Zenarestat	1.37	-13.15	8	15
Pyridazinone	0.49	-10.98	30	34

Table 1: Results of 100 ISCD runs for 6 AR ligands, using the joined 1AH0-1AH3-1US0 grid. For each ligand, the top-ranked docking result is reported, with the RMSD to the experimental binding mode, the docking score ("docked energy" E_d), the size of the top-ranked result cluster, and the total number of results showing an RMSD less than 2 Å.

To setup ISCD, the single grids were combined to a joined grid with repulsive layers between them.¹²⁴ The joined grid representing all three binding-site conformers was used to test whether ISCD is able to identify the native binding pocket of a given ligand (sorbinil, tolrestat, and IDD594) in a single docking calculation. For this purpose the same standard docking parameters were used as applied before in the single-binding-site docking.

Table 1 illustrates that with ISCD the native binding pocket and binding mode is indeed found in all three cases: for each ligand, the top-ranked cluster corresponds to the correct binding mode in the correct protein conformer, and the cluster size (corresponding to the occurrence frequency of the result) is excellent for the most rigid ligand (sorbinil) and fully acceptable for the most flexible compound (IDD594).

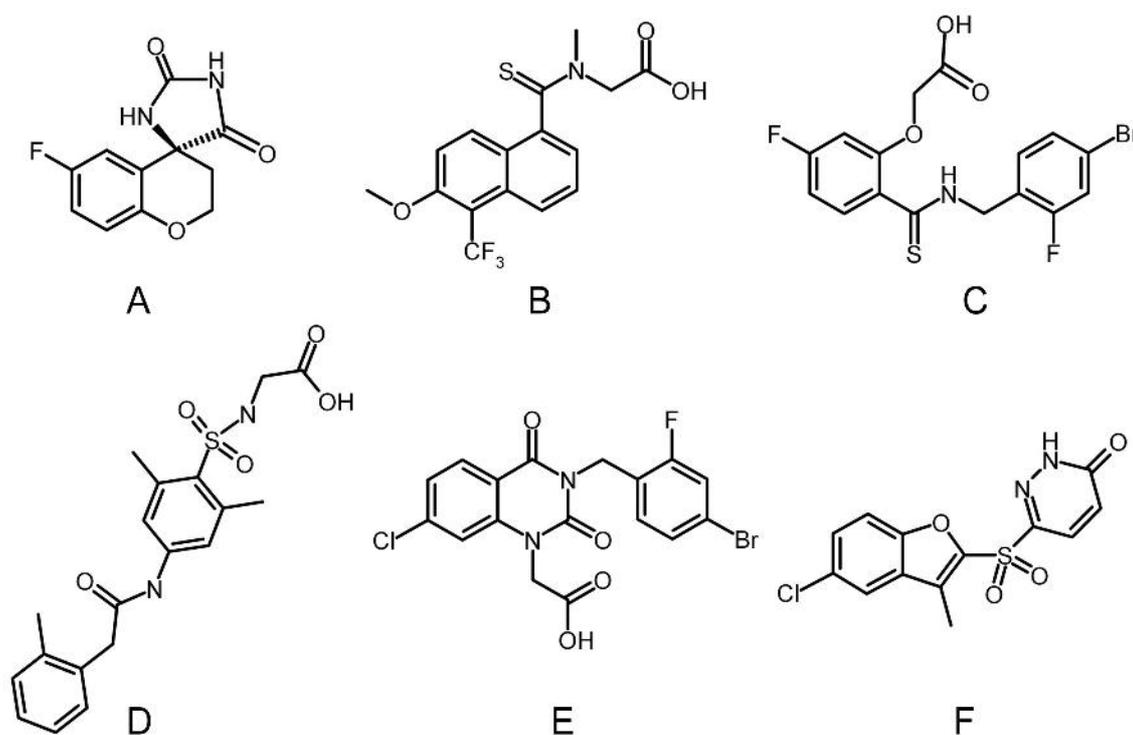


Figure 10: Chemical formulae of the AR inhibitors investigated by ISCD: sorbinil (A), tolrestat (B), IDD594 (C), fidarestat (D), zenarestat (E), and the pyridazinone inhibitor (F).

Since the inhibitors tested so far were exactly the ligands of the protein conformers used to construct the joined grid, a further test for running ISCD with the joined 1AH0-1AH3-1US0 grid was carried out using the compounds fidarestat and zenarestat (cf. Figure 10). In

the crystal structure, fidarestat (PDB 1PWM) shows a sorbinil-like binding mode, whereas zenarestat (PDB 1IEI) binds to a similar AR conformer as IDD594 (to date, no other ligand than tolrestat is known to induce a 1AH3-like conformation).

For zenarestat, ISCD correctly suggests preferential binding to the 1US0 conformation (cf. Table 1): the top-ranked result shows a clear score difference of 1.14 kcal/mol with respect to binding modes in other pockets, and all top 15 results yield an RMSD <1.4 Å with respect to the crystal structure (after best-fit alignment of the 1IEI protein structure with 1US0). For fidarestat, the correct binding mode is obtained in almost all of the 100 runs, though in two different pockets: 47 times in the 1AH0 pocket (with an RMSD of 0.69 Å after best-fit alignment of the 1PWM protein structure with 1AH0) and 52 times in the 1US0 pocket (RMSD 0.51 Å after best-fit alignment of 1PWM with 1US0). Both results show a virtually identical score (-9.46 kcal/mol versus -9.58 kcal/mol, respectively), indicative of the fact that although 1US0 corresponds to the open conformation and fidarestat is expected to bind to the closed conformation (1AH0), the 1US0 conformation is also compatible with the native fidarestat binding mode. This is in contrast to sorbinil, where a clash with Cys298 in 1US0 precludes docking in the native mode. Fidarestat shows a larger distance to Cys298, thus allowing a near-native binding mode to be observed also in 1US0.

In a further test, predictive docking was carried out for a new and very potent sulfonyl-pyridazinone inhibitor for which no crystal structure was available at that time (compound 24 described by Mylari et al.¹²⁵, cf. Figure 10). ISCD suggested a binding mode similar to IDD594, with the clearly best score achieved in the 1US0 pocket and 34 out of 100 runs ending up with this solution. The correctness of the prediction was later confirmed by the crystal structure of the pyridazinone-AR complex¹²⁶, which yields an RMSD of 0.49 Å for the docking result after superposition of the protein in the new crystal structure with 1US0.

Taken together, the results indicate that with ISCD alternative protein conformations can successfully be addressed simultaneously. The major advantage of the method, however, is that differences among the target structures need not to be confined to small side-chain reorientations. Accordingly, issues of selectivity with respect to different targets can also be addressed. While in the original proof-of-concept very different proteins (and

ligands with no expected cross-reactivity) were used¹²⁴, here the selectivity with respect to two closely related enzymes is investigated.

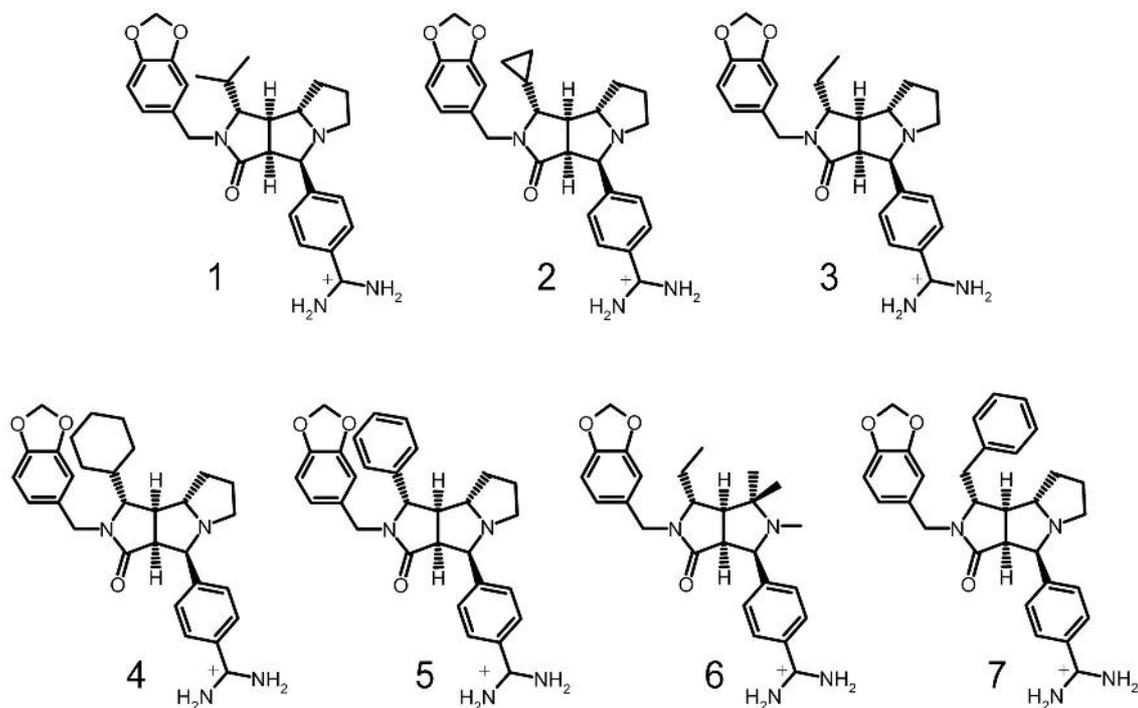


Figure 11: Chemical formulae of the seven trypsin and thrombin ligands investigated by ISCD.

The serine proteases thrombin and trypsin and a set of seven new inhibitors with varying selectivity for thrombin were used for this purpose (derivatives 1-7 in Fokkens et al.¹²⁷). The inhibitors consist of a common rigid non-peptidic core, substituted in position 1 with different aliphatic and aromatic groups, as shown in Figure 11 (ligands 3 and 6 differ only with respect to the core ring system). The selectivity ratio of the inhibition constants $K_i(\text{trypsin})/K_i(\text{thrombin})$ ranges between 2.6 for ligand 4 and 760 for ligand 1 (measured for the racemic mixture).¹²⁷ For all ligands except ligand 4, crystal structures in complex with trypsin could be obtained^{127;128}, for ligands 1, 2, 3, and 6 crystal structures in complex with thrombin are available as well¹²⁷⁻¹³⁰. The structures show that all ligands share a common binding mode and that exclusively the (+)-enantiomer is bound to the enzyme; this is also the enantiomer displayed in Figure 11 and used in all docking calculations.

Ligand	Trypsin					Thrombin				
	Cluster rank	E_d [kcal/mol]	RMSD [Å]	Cluster size	$N \leq 2 \text{ Å}$ RMSD	Cluster rank	E_d [kcal/mol]	RMSD [Å]	Cluster size	$N \leq 2 \text{ Å}$ RMSD
Lig1	7	-11.43	0.54	42	42	1	-13.72	0.90	2	8
Lig2	5	-11.66	0.52	40	40	1	-13.11	0.89	3	9
Lig3	3	-11.51	0.85	42	42	1	-13.65	0.85	3	6
Lig4	2	-12.21	0.86	7	43	1	-13.28	1.06	1	1
Lig5	5	-10.84	1.10	28	29	1	-11.70	8.60	5	2
Lig6	4	-11.20	0.97	1	39	1	-13.59	1.03	5	9
Lig7	2	-12.59	0.55	45	45	1	-12.76	8.65	1	0

Table 2: Results of 50 ISCD runs for the joined trypsin and thrombin grid. The columns have the same meaning as in Table 1. The RMSD was calculated for the common ligand substructure consisting of the core ring system and the benzamidine anchor, which is the part of the ligands that is well defined in all available crystal structures.

Using the crystal structures of trypsin and thrombin in complex with ligand 1, grids were calculated for the two binding sites and combined to one joined grid. As shown in Table 2, performing ISCD with 50 independent runs per ligand revealed energetically preferred binding to thrombin: for all seven ligands, the top-ranked result was found in the thrombin grid, whereas the best-ranked docking result for trypsin occurred between rank 2 and 7. All best-ranked trypsin results are structurally correct, as indicated by the low RMSD values. For thrombin, this is the case for five of the seven ligands, since the top-ranked result for ligand 5 and 7 does not show the presumably correct binding mode (for both ligands no experimental complex structure with thrombin is available).

Thus, looking at the best-ranked result in each grid, ISCD reveals the correct binding mode of all ligands for trypsin and of five ligands for thrombin. In addition, the binding preference is correctly predicted: for all ligands, the better score is observed for thrombin, with a difference of 0.17 - 2.39 kcal/mol in favor of thrombin, roughly correlating with the experimental affinity difference. Despite this energetic preference, however, the large majority of favorable binding modes is found in the trypsin grid (cf. Table 2 and Figure 12). Apparently, the trypsin binding site represents a well-defined and easily accessible local minimum, whereas the global optimum given by the thrombin binding site is more difficult to reach with a finite stochastic search in this large search space. Even with different or enhanced search protocols the situation does not change significantly (results not shown).

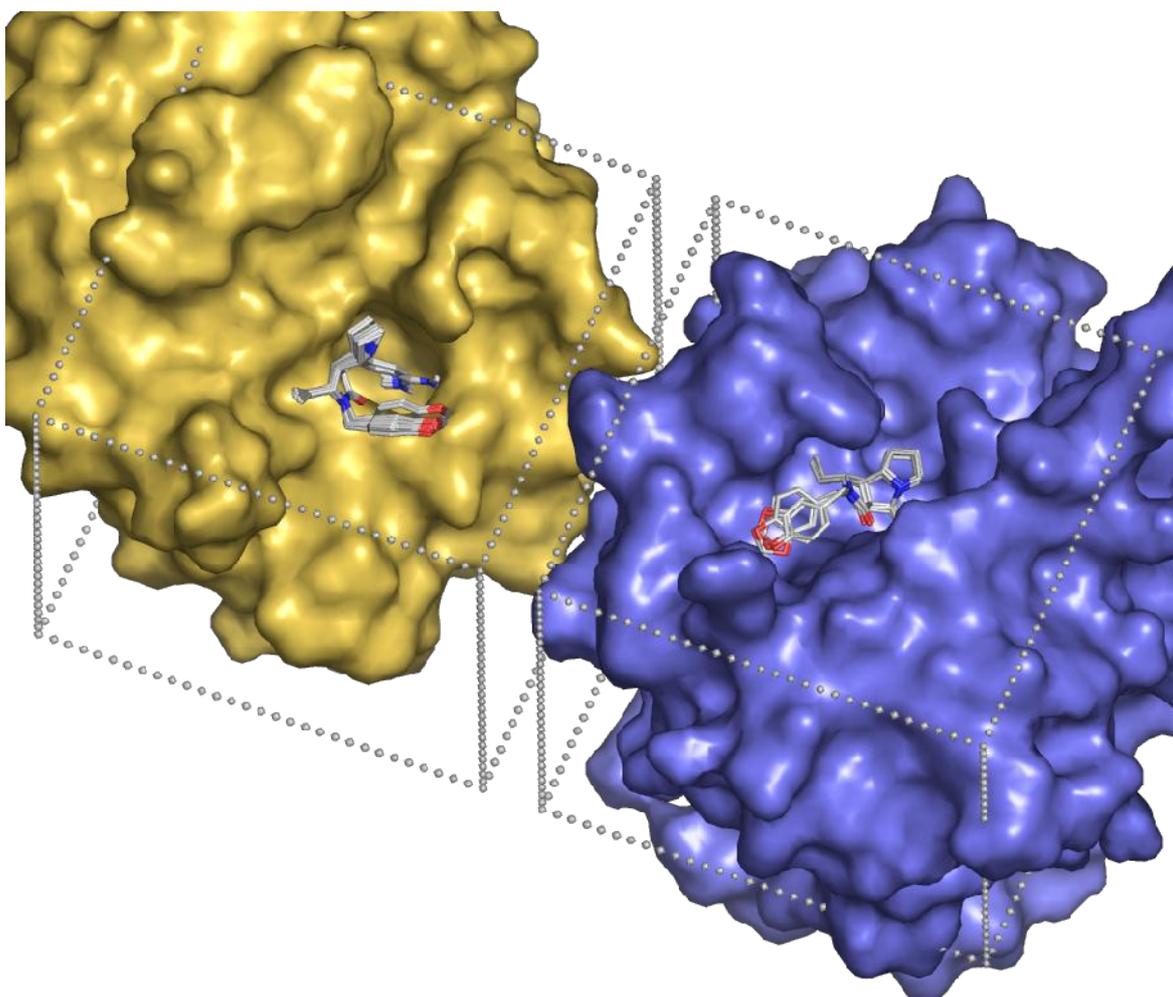


Figure 12: ISCD docking results for ligand 3, illustrating the members of the three best-ranked clusters, which together cover 96% of the results (48 of 50 docking runs). Clusters 1 and 2 are found in thrombin (blue) and contain three results each; cluster 3 is found in trypsin (yellow) and contains 42 closely overlapping results. The borders of the two grids joined to a single ISCD grid are shown; the gap between them corresponds to the repulsive layer required for avoiding artifacts.

While it may be argued that converging to the global minimum with a probability of $\leq 16\%$ is not sufficient, it is also necessary to see that full convergence to the global optimum would not be desired in ISCD, because otherwise binding to the protein showing lower affinity would not be found at all. Instead, ISCD should indeed reveal all experimentally observable binding modes, and the best-ranked result in each grid should provide the correct binding mode for the corresponding protein, as observed here for most of the lig-

ands. (Further details about the results, as well as results of separate docking runs to trypsin and thrombin are provided as Supporting Information B.)

Summary and Conclusion

In summary, this study has shown that addressing protein flexibility and ligand selectivity with ISCD is feasible. With a single docking calculation, the appropriate binding-site conformation can be selected from three different alternatives in case of the flexible AR enzyme. Likewise, a single docking calculation is sufficient to reveal the preferred binding modes of a given ligand for two closely related proteins, trypsin and thrombin.

Alternative approaches for dealing with protein flexibility in ligand docking are available, such as FlexE¹⁰⁰ (which had also been tested on AR¹⁰⁰) and grid-based averaging^{101;107} (which we have applied to AR for comparative purposes; cf. Supporting Information C), but these are all restricted to relatively small structural differences among the investigated systems. ISCD does not have this limitation and can be applied to protein structures of any desired difference.

The challenge, however, will be to develop efficient protocols for simultaneously considering more than two or three structures in order to extend the scope of ISCD.

Supporting Information A: Computational Methods

Protein setup and grid calculations.

For aldose reductase (AR), the following three PDB structures corresponding to different binding-site conformers were used: 1AH0 (complex with sorbinil), 1AH3 (complex with tolrestat), and 1US0 (complex with IDD594).

For trypsin, the complex structure with ligand **1** was obtained in house by X-ray crystallography and deposited in the PDB under code 1Y3U¹²⁷. The structure of thrombin in complex with ligand **1** was kindly provided by Hoffmann La Roche (Basel, Switzerland)^{129;130}.

Water molecules and ligands were removed from the complexes, and polar hydrogens were added to the proteins using the PROTONATE utility distributed with AMBER¹³¹.

Partial atomic charges from the AMBER united atom force field were assigned to the protein atoms, and solvation parameters were added with the ADDSOL utility of AutoDock 3.0¹²².

In the actual docking process, AutoDock uses a grid-based representation of the protein. Accordingly, for each point in a regularly spaced grid and for any probe of interest (i.e., any ligand atom types and a probe charge), the interaction energy of that probe with the entire protein is calculated and saved in files that serve as look-up tables for faster energy evaluation upon docking. Therefore, grids were generated with the help of AutoGrid, using a grid spacing of 0.5 Å in case of the AR structure and of 1 Å in case of trypsin and thrombin.

First, a separate standard binding-site grid was calculated for each protein. Each grid was centered on the corresponding active site and had a dimension of 23 Å × 25 Å × 27 Å in case of AR and 30 Å × 30 Å × 30 Å in case of trypsin and thrombin; this is sufficiently large to include the entire binding site and significant parts of the surrounding surface.

To perform in-situ cross-docking (ISCD), the grids were joined to a single large grid readable by AutoDock. In case of AR, the joined grid consisted of the grids calculated for 1AH0, 1AH3, and 1US0; in case of trypsin and thrombin, the joined grid consisted of the grids calculated for each of the two separate proteins. Since only already available grid files had to be manipulated, AutoGrid was not required at this stage. The manipulations corresponded to a linear alignment of the two grids along the x-axis. To avoid docking results across the border of the two grids (which, in structural terms, would be pure artifacts), a "spacer" was inserted between the grids (1 Å in case of the AR grids, 3 Å in case of trypsin and thrombin). Each grid point in the spacer region was assigned an energy value of +100 kcal/mol, leading to a "repulsive layer" in which no docking run would terminate.

Ligand setup

Coordinates of the AR ligands were taken from the corresponding PDB files, with the exception of the pyridazinone compound (**F**), which was generated with Sybyl¹³² since no crystal structure was available at that time. Coordinates of ligand **1** binding to thrombin and trypsin were taken from the complex structure with thrombin. Using Sybyl, ligands **2-7** were built based on the coordinates of ligand **1**.

Hydrogens were added to the initial ligand structures, the structures were subjected to short energy minimization using the Tripos force field in Sybyl, and atomic charges were assigned according to the Gasteiger-Marsili formalism¹³³, which is the type of charges used in calibrating the AutoDock free energy function.

The minimized structures were superimposed with the experimental binding mode of ligand **1** in trypsin and thrombin, respectively; the binding modes obtained this way served as reference for the RMSD measurements of the docking results.

Finally, the compounds were setup for docking with the help of AutoTors, the main purpose of which is to define the torsional degrees of freedom to be considered during the docking process. Accordingly, the following numbers of rotatable bonds were defined: 0 in sorbinil, 5 in tolrestat, 6 in IDD594, 1 in fidarestat, 4 in zenarestat, 2 in the pyridazinone compound; 4 in ligand **1** to ligand **6**, and 5 in ligand **7** (the bond between the aromatic ring and the amidine group in **1-7** was kept fixed at a torsion angle of 10°).

Docking

Docking was carried out with AutoDock 3.0, using the empirical free energy function and the Lamarckian Genetic algorithm (LGA)¹²².

The standard protocol used an initial population of 50 randomly placed individuals, a maximum number of 1.5×10^6 (AR) or 3.0×10^6 (trypsin, thrombin) energy evaluations, a mutation rate of 0.02, a crossover rate of 0.80, and an elitism value of 1. For the local search, the pseudo Solis and Wets algorithm was applied, using a maximum of 300 iterations per local search. The probability of performing local search was 0.06, and the maximum number of consecutive successes or failures before doubling or halving the local search step size was 4.

With this protocol, 100 independent runs were carried out for each ligand in case of AR, 50 in case of trypsin and thrombin. Results differing by less than 1 Å RMSD were clustered together and represented by the result with the best docking score in the cluster.

Supporting Information B: Docking to trypsin and thrombin

For comparative purposes, standard docking runs with the individual (separate) grids were carried out for trypsin and thrombin. The results obtained with the same settings as used in trypsin-thrombin ISCD are summarized in Table 3.

Ligand	Trypsin					Thrombin				
	Cluster rank	E_d [kcal/mol]	RMSD [Å]	Cluster size	$N_{\leq 2 \text{ Å}}$ RMSD	Cluster rank	E_d [kcal/mol]	RMSD [Å]	Cluster size	$N_{\leq 2 \text{ Å}}$ RMSD
Lig1	1	-11.43	0.49	50	50	1	-13.85	0.92	16	39
Lig2	1	-11.67	0.48	50	50	1	-13.08	0.80	8	42
Lig3	1	-11.48	0.54	50	50	1	-13.81	0.90	22	41
Lig4	1	-12.29	0.67	11	50	1	-13.27	1.05	17	22
Lig5	1	-10.85	1.10	40	43	1	-11.70	8.95	9	6
Lig6	1	-11.35	1.02	1 [#]	50	1	-13.58	1.00	19	41
Lig7	1	-12.57	0.51	50	50	1	-12.91	2.36	4	5

[#]The result on rank 2 shows a cluster size of 49 ($E_d = -11.17$ kcal/mol, RMSD = 0.60 Å).

Table 3: Docking results with separate grids for trypsin (left) and thrombin (right). In each case, 50 independent docking runs were carried out. For each ligand, the top-ranked docking result is reported, with the root-mean-square deviation to the experimental binding mode (RMSD), the docking score ("docked energy" E_d), the size of the top-ranked result cluster, and the total number of results showing an RMSD less than 2Å. The RMSD was calculated for the common ligand substructure consisting of the core ring system and the benzamidine anchor, which is the part of the ligands that is well defined in all available crystal structures.

Except for the different occurrence frequencies, the results are very similar to those obtained with ISCD. The docking runs for trypsin reproduce the native binding mode in virtually all cases. Also the thrombin results show good occurrence frequencies of the correct binding mode (with the exception of ligand **5** and **7**, as further discussed below). The scores are in all cases almost identical to those observed by ISCD (with a maximum difference of 0.16 kcal/mol). The score differences between thrombin and trypsin (i.e., $\Delta E_d = E_d(\text{thrombin}) - E_d(\text{trypsin})$) are comparable to the experimental $\Delta\Delta G$ values as estimated from the corresponding K_i ratios (cf. Fokkens et al.¹²⁷), as they allow to separate the highly selective thrombin ligands (**1**, **2**, **3**, **6**) from those with low selectivity for thrombin (**4**, **5**,

7). A more quantitative agreement is not to be expected, given that the accuracy of the AutoDock scoring function in reproducing experimental binding free energies is approximately 2 kcal/mol¹²² and racemate data are considered as experimental reference.

The separate-grid docking results also indicate that the problems for ligands **5** and **7** in ISCD are not primarily due to the ISCD method. In fact, even when docking to thrombin alone, a near-native binding mode is not obtained on the first ranks. The reason, however, is not necessarily a short-coming of the docking procedure, since for both ligands a crystal structure with thrombin could not be obtained and the actual binding mode can only be assumed.

Furthermore, in the crystal structures with trypsin, the phenyl and benzyl substituent of ligand **5** and **7**, respectively, is not visible in the electron density, thus pointing to disorder or flexibility of these side chains. It is likely that in the geometrically more restrictive environment of the thrombin binding site these substituents can not be well-placed without small conformational changes either in the neighboring protein residues or in the central ring system.

Supporting Information C: Docking to multiple conformers by grid-based averaging

As an alternative approach to incorporate protein flexibility in grid-based docking, a weighted averaging of grids calculated for different binding-site conformers has been suggested^{101;107}. In this method, a "normal" grid is used (covering geometrically only one binding site), but every grid point holds a weighted average energy calculated from the alternative protein conformations. This "clamped-grid" approach¹⁰¹ was tested here with the AR conformers. The separate grids calculated for the three AR conformers 1AH0, 1AH3, and 1US0 (cf. above; grid spacing 0.5 Å, grid dimension 23 Å × 25 Å × 27 Å), were averaged according to the procedure described by Österberg et al.¹⁰¹: the weight is given a value of 1 if the interaction energy of the grid point is less than 0 kcal/mol; if the energy is greater than 0 kcal/mol, a small weight of 0.0001 is applied. Upon averaging, the weights are nor-

malized to give a sum of 1. The resulting "clamped-grid" was then used for docking, with all parameters and settings kept identical to the other AR docking runs.

The results are summarized in Table 4. While the correct binding mode was obtained with high fidelity for sorbinil and fidarestat, and with sufficient quality and occurrence frequency for tolrestat, the method failed to produce good results for IDD594, zenarestat, and the pyridazinone ligand. For IDD594 one result with 2.00 Å RMSD was obtained on a very low rank, all other results showed much larger deviations. For the pyridazinone compound, reasonable binding modes started occurring at rank 3, but the two top ranks did not yield the correct binding geometry.

Ligand	RMSD [Å]	E_d [kcal/mol]	Cluster size	$N \leq 2 \text{ Å RMSD}$
Sorbinil	0.32	-8.95	100	100
Tolrestat	1.72	-10.90	3	31
IDD594	2.39	-11.11	1	1
Fidarestat	0.73	-9.27	99	100
Zenarestat	7.19	-11.04	35	0
Pyridazinone	7.13	-8.99	3	22

Table 4: Results of 100 "clamped-grid" docking runs for the six AR ligands. For each ligand, the top-ranked docking result is reported, with the root-mean-square deviation to the experimental binding mode (RMSD), the docking score ("docked energy" E_d), the size of the top-ranked result cluster, and the total number of results showing an RMSD less than 2Å.

The advantage of the grid-averaging approach over ISCD is that a grid of normal size (as for a single binding site) can be used. However, the averaging process smears out certain differences among the conformations, and the ligands being docked seek for optimal features which arise from different conformations, but which may not be realizable in a single protein conformation. In some cases this combinatorial power of the averaging method might lead to better results since new possible binding modes might be found which would have been missed otherwise.

For AR, this advantage seems not to be given, since a clear differentiation between the alternative conformations is possible. Furthermore, rescoring (and structural interpretation) of docking results may sometimes be difficult, since there is no structural representation of the clamped grid in terms of a corresponding protein structure. Such difficulties do not

arise with the simpler ISCD method. It does not allow for the recombination (generation) of new conformations, but it is also not limited to a certain amount of conformational differences.

Extending charted space: comparative MD simulations of Aldose Reductase

Introduction

High quality structural data are the basis for any structure-based drug design project. Nowadays computational methods are abundantly used at various stages of the drug development process¹³⁴⁻¹³⁶. The success of many of these methods relies on accurate structural information about the binding mode of a given inhibitor, its interactions with the protein, and the conformation of the binding pocket itself. X-ray crystallography is the method of choice to deliver this information¹³⁷⁻¹⁴¹. Improvements in the technology, fine-tuned experimental procedures as well as the increased availability of synchrotron radiation sources have taken great impact on the quantity and quality of the available crystal structures.

Despite the enormous wealth of protein structures, there is an intrinsic danger in a coordinate file downloaded from the PDB. These files contain the exact positions for each atom of the system. The coordinates are given with a precision of three positions after the decimal point, suggesting that a protein is a rigid system of atoms which basically do not move. This is a misleading assumption, since proteins are of course dynamic systems. Different types of motions which occur on different time scales are ubiquitous in proteins: thermal atomic fluctuations, motions around torsional angles, displacements and librations of single amino acids, secondary structure elements or even entire protein domains. These motions are not a design flaw of nature, but are essential to evolve the biological function of the protein. However, while moving from physiological conditions to the environment of a protein crystal many of these motions are either inevitably restricted or become less pronounced by the experimental conditions, especially due to the low temperature at which crystal structures are measured.

The dynamical behavior of molecular systems is of utmost importance to understand protein-ligand interactions and has to be considered in structure-based drug design projects. One of the many ways in which flexibility can influence drug design is the occurrence of an "induced-fit" binding mechanism. In this case, a protein can accommodate different ligands by adapting its binding pocket to the shape of the ligand. Aldose Reductase

(AR) is such a protein. Its biological function is to reduce various aldoses and alcohols to the corresponding alcohols using NADPH as cofactor. Not much is known about its function under physiological conditions. However, a role as general detoxifying enzyme is discussed^{11;12}. Therefore, it is mandatory for this enzyme to operate on various substrates, since otherwise, for every aldehyde which has to be reduced, a special enzyme would be needed. Thus, a flexible binding pocket is a prerequisite for an enzyme having a broad range of substrates. Very similar properties have been attributed to the recently intensively studied P450 enzymes which play a crucial role in the metabolism of xenobiotics¹⁴².

AR is build-up of 315 amino acids and folds to an α/β -TIM-Barrel structure (Figure 13). Eight parallel β -sheets are surrounded by eight anti-parallel α -helices. The eight β -strands enclose a tightly packed hydrophobic core formed by their side-chains. In the next layer of the barrel hydrophobic side-chains of the α -helices are packed against hydrophobic side-chains of the β -sheets. Thus, the barrel is equipped with a shell of hydrophobic residues. As with most enzymes of this fold¹⁴³ the active site is formed by the loop regions that connect the β -strands with the adjacent α -helices. The TIM-barrel fold provides the stable and rigid framework for the catalytic reaction, whereas the binding site itself is built-up from loop regions.

For AR, many high-quality crystal structures are available which were determined in complex with a large variety of ligands. Thus, from a structural point of view, AR is a very well characterized enzyme. A previous comparative analysis¹¹³ of the crystal structures available at that time gave insight into the binding pocket of AR. The pocket can be divided into two sub-pockets: one where the catalytic reaction takes place, i.e. the anion or catalytic pocket; and another, which is named specificity pocket. 'Induced-fit' phenomena are executed only by the specificity pocket, which can adopt at least two different opened and a closed conformation leading to three major pocket conformations. Not all residues within the specificity pocket show the same degree of flexibility. A small loop region near the C-terminal end of the protein is used to facilitate most of the adaptations, a behavior which will be analyzed in more detail in this study. In recent years many high-resolution structures of AR in complex with different ligands have been determined. They provide a wealth of information which has not been considered in the last comparative crystal structure analysis¹¹³. The adaptations resulting from binding of a variety of different ligands can

now be analyzed in greater detail. In addition to the analysis of available crystal structures, ten molecular dynamics (MD) simulations were carried out to further investigate the flexibility of this enzyme. In total seven different complex simulations and three simulations of the holo enzyme were carried out. Using a multitude of different complex structures as starting geometries for MD simulations improves the mapping of the conformational space of the enzyme, thus providing a better data basis for examining the conformational space of the protein.

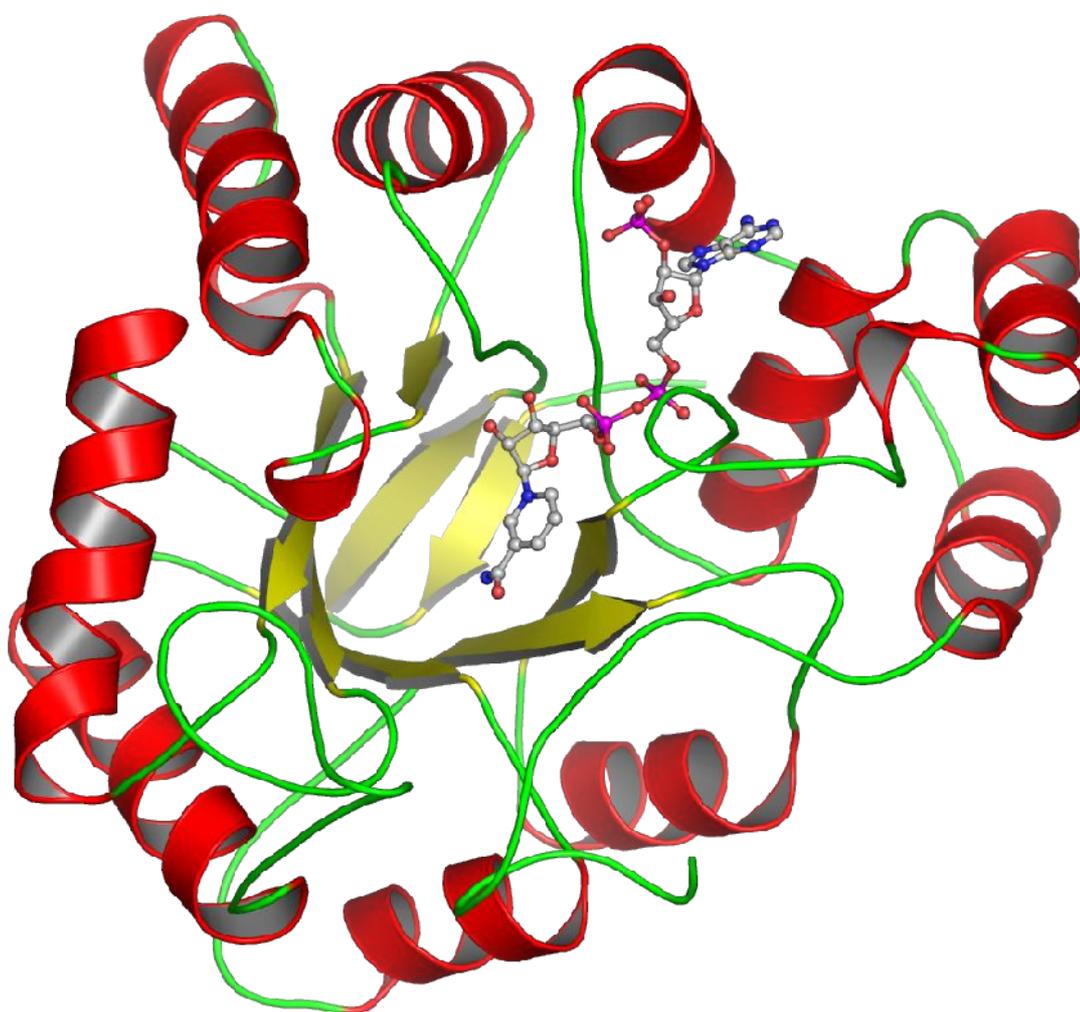


Figure 13: Overall fold of AR; α -helices are shown in red, β -sheets are depicted in yellow, and loop regions are shown in green. The cofactor is shown in a ball-and-stick representation.

Besides its interesting structural features, AR is also a relevant drug target. It is the first enzyme of the sorbitol pathway. As with AR itself, the physiological role of the whole pathway is still not fully understood. However, if the human body is faced with raised blood glucose levels, some of the excessive glucose is processed via the sorbitol pathway¹⁴⁴. Elevated blood glucose levels are frequently found with patients suffering from diabetes mellitus. Through the conversion of glucose to sorbitol cellular NADPH is consumed and not available for other processes. Thus, oxidative stress is induced in the cells. Furthermore, sorbitol accumulates within the cells and increases the osmotic stress of these cells. The two combined factors are made responsible for many of the late-onset diabetic complications such as retinopathy and angiopathy¹⁴⁵⁻¹⁴⁷. In numerous studies inhibition of AR has shown to alleviate these complications^{62;88;148;149}. Consequently, AR has become a drug target to counteract these effects.

Materials and Methods

Crystal structure analysis

Comparison of available crystal structures was carried out using Relibase⁺ Version 2.1¹⁵⁰⁻¹⁵². The inhouse structures of human AR in complex with the JFD compound¹¹⁶, 47d, zopolrestat¹⁵³, sorbinil (see Appendix B), and tolrestat¹⁵³ were added to Relibase⁺ using the Relibase⁺ data processing module. Querying the resulting database for aldose reductase structures with a resolution of ≤ 2.5 Å results in 38 hits. For three entries (1abn, 1dla, and 1mar) only the coordinates for the Ca atoms are deposited in the PDB¹⁵⁴. Based on the ultra-high resolution structure of the inhibitor idd594 in complex with AR (1us0¹¹⁴) a 'similar binding sites' search was performed using a minimum sequence identity of 95%. This discards structures of non-human species and results in a dataset of 29 protein chains of which four structures of AR mutants were removed (2agt, 2acu, 1az1, 1az2). For the inhibitor fidarestat and its stereoisomers four different structures are available: two for the native ligand (2-S,4-S configuration) (1pwm, 1ef3) and one for the isomers 2-R,4-S (1x97) and 2-S,4-R (1x98). 1pwm was chosen as representative for the native ligand due to its higher resolution, whereas both structures for the stereoisomers were included. For the inhibitor idd 552 also two different structures are available. However, in these two cases the protein was crystallized under different pH conditions: pH 5 (1t41) and pH 8 (1t40), respectively.

Since it is known that changing the pH of the crystallization conditions can influence the resulting crystal structure¹⁵⁵, both structures were included in the dataset. Also for the Pfizer compound two structures are available¹²⁶: one at 0.95 Å resolution (1z8a) and one which diffracted to 1.43 Å (1z89). For the high-resolution structure, the occupancy of the inhibitor in the binding pocket is only ~50%, whereas for 1z89 the inhibitor possesses full occupancy. Therefore, the latter included in the dataset.

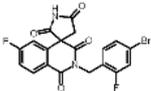
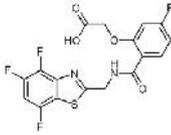
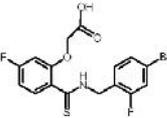
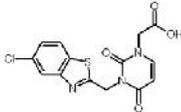
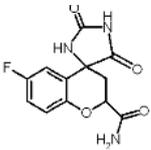
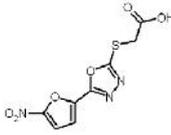
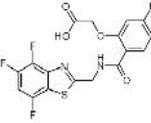
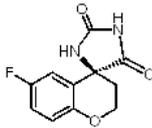
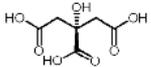
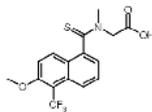
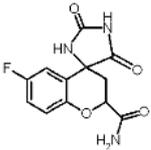
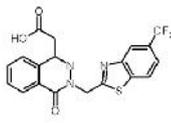
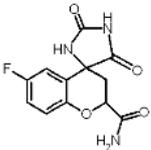
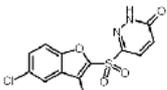
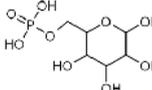
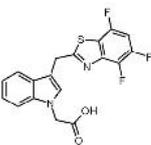
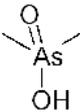
Finally, the structure of AR in complex with DEG (2f2k)¹⁵⁶ was removed, since the inhibitor shows an average B-factor of 77 Å² at 100 % occupancy in comparison to 21 Å² of the surrounding protein. It is therefore questionable, whether meaningful structural conclusions can be drawn from such a structure. The resulting dataset comprises the 20 remaining structures. The pdb codes and chemical formulas of the bound ligands are shown in Table 5.

Removal of the non-human complex structures as well as the mutated forms provides an unbiased data sample only affected by adaptations of one single protein species induced by the different bound ligands. All possible superimposed consequences arising from mutations or sequence differences of the protein remain unconsidered. However, possible influences of deviating crystallization conditions remain and it is very difficult to trace and quantify such effects.

The only mutant structure which was added to the dataset was the apo form of the enzyme (1xgd)³⁴. This structure was obtained from an Arg268Ala mutant. According to the authors the mutation decreases the affinity of the cofactor to the enzyme and was crucial to obtain an NADPH-free structure. The mutation is located more than 12 Å away from the binding pocket, thus the impact on the binding-pocket conformation is presumably low.

All structures in the dataset were superimposed based on the residues within 6 Å of the ligands. All subsequent data analyses were carried out using built-in Relibase⁺ features. All values for assessing differences between the structures were kept at their default values. Therefore, C α movements of single residues were considered significant if greater than 0.5 Å, the threshold for movements of side-chain centers was set to 1 Å, and torsional angles were considered distinct if more than 10° apart. Visual inspection of the superimposed

structures as well as measurements of main-chain torsion angles were performed using Re-liview, the visualization tool of Relibase⁺.

structure	res.[Å]	ligand name	ligand structure	structure	res.[Å]	ligand name	ligand structure
1pw1	1.10	minalrestat		1t40	1.80	idd552	
1us0	0.66	idd594		947d*	1.42	47d	
1pwm	0.92	fidarestat		9jfd*	1.55	jfd	
1t41	1.05	idd552		9sbi*	1.47	sorbinil	
1x96	1.40	2 citrates		9tol*	1.08	tolrestat	
1x97	1.40	fidarestat (2R4S)		9zop*	1.6	zopolrestat	
1x98	1.30	citrate, fidarestat (2S4R)		1ads	1.65	none	
1z89	1.43	Pfizer compound		2acq	1.76	glucose-6- phosphat	
1z3n	1.04	lidorestat		2acr	1.76	cacodylate	

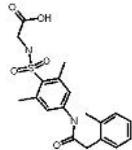
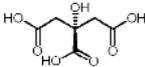
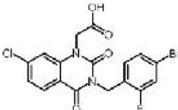
structure	res.[Å]	ligand name	ligand structure	structure	res.[Å]	ligand name	ligand structure
1el3	1.70	idd384		2acs	1.76	citrate	
1iei	2.50	zenarestat		1xgd	2.10	none	

Table 5: AR crystal structures used for analysis. The first column gives the pdb code, the second the crystallographic resolution, the third and fourth contain the name of the ligand and its 2D chemical formula.

* *in-house crystal structure which was added to Relibase⁺ using the data processing tool*

MD simulations

Starting geometries for each protein ligand complexes were taken from the corresponding crystal structures. Starting geometries for the simulations of the holo enzymes were derived by extracting the ligands sorbinil, tolrestat and idd 594 from their corresponding crystal structures. All preparing steps and simulations were carried out with the AMBER 8.0¹⁵⁷ suite of programs, using the Cornell et al.¹⁵⁸ force field in the variant of the parm99 parameter set¹⁵⁹. Parameters for the ligand were taken from the GAFF force field¹⁶⁰ and assigned using ANTECHAMBER. Parameters for the cofactor NADP⁺ were taken from a previous study by Sotriffer et al¹¹³. Atomic charges for both, the ligand and the cofactor, were calculated by fitting to the HF/6-31G* electrostatic potential using the conformation observed in the crystal structure. The corresponding ab-initio calculations were performed with GAUSSIAN98¹⁶¹, the restrained electrostatic potential fit with the ANTECHAMBER program using the RESP method^{162;163}. Hydrogen atoms were added using AMBER templates. The simulated protein system thus consisted of 316 amino acids or 5071 atoms, plus the cofactor with 73 atoms and the ligand atoms. The crystal structure was subjected to a short initial energy minimization in vacuo using 20 steps of steepest descent followed by 180 steps conjugate gradient minimization. After addition of two sodium counterions to ensure neutrality (placed at minima of the electrostatic potential), the systems were solvat-

ed in a box of TIP3P¹⁶⁴ water molecules, which resulted in box sizes of $\sim 80\text{\AA} \times 65\text{\AA} \times 75\text{\AA}$ and a total of ~ 9000 water molecules. The MD simulations were then started by heating the solvent to 300K over a period of 20 ps and cooling to 100 K over a period of 5 ps, keeping the solute fixed. After this procedure, the entire system was gradually brought to 300 K over a period of 25 ps. The simulation was then carried on for 6050 ps under constant temperature and pressure (NPT), applying periodic boundary conditions. The temperature was kept constant by coupling to a heat bath via the Berendsen algorithm¹⁶⁵. Pressure was adjusted by isotropic position scaling using a Berendsen-like algorithm. Covalent bonds to hydrogen atoms were constrained by the SHAKE¹⁶⁶ algorithm and a time step of 2 fs was used. A cutoff of 8 Å was applied to the van der Waals interactions, while the electrostatics were treated by the Particle Mesh Ewald method¹⁶⁷. The simulations were carried out with the SANDER MD module of AMBER 8.0. For analysis, energy data were saved every 10 time steps, solvent and solute coordinates every 0.5 ps. All results presented refer to the 6.0 ns trajectories which exclude the first 100 ps required for temperature adjustment and equilibration.

Generation of 2D rms plots

To obtain this plot every 10 ps snapshots were extracted from the simulation and fitted to the starting structure using the C α atoms only. This leads to an ensemble of 600 snapshots representing the time course of the simulation. Using the *2drms* command of the PTRAJ module of AMBER 8¹⁵⁷ a pairwise rmsd matrix based on the binding site atoms between the individual snapshots was calculated. This matrix was then plotted using the program PLOTMTV. The pairwise rmsd values between the individual snapshots are color-coded on a relative scale using a color ramp raising from green over yellow to red.

Docking with AutoDock

All docking experiments were carried out using AutoDock 3.0.5¹²² which is a grid-based docking tool using a Lamarckian genetic algorithm as search engine and a force-field based scoring function as fitness function. If not noted otherwise, 100 runs of the genetic algorithm were performed for each docking experiment. A population size of 50 and a maximum number of energy evaluation steps of 1.5×10^6 were applied. The grid consisted of $46 \times 50 \times 54$ points with a spacing of 0.5 Å between the single grid points. Mutation and crossover rate were set to 0.02 and 0.80, respectively. The local energy minimization was

set to a maximum of 300 steps. For the docking experiments using the 'in-situ' cross-docking approach, the grids and settings were used as previously described^{124,168}.

Results and Discussion

Comparative crystal structure analysis

Since a previous preliminary analysis of available crystal structures¹¹³ many novel high-quality structures of potent inhibitors in complex with AR have been deposited in the protein data bank (PDB¹⁵⁴) or solved in our laboratory, thus expanding the experimental basis to analyse binding-site flexibility of AR. Considering these new examples allows removing limited resolution structures and structures from non-human or mutated enzymes.

The following comparative crystal structure analysis is carried out roughly following the strategy published by Sotriffer et al.¹¹³. As already briefly mentioned in the introduction, this analysis demonstrated that the AR binding pocket can be divided into two sub-pockets: an 'anion binding pocket' where the cofactor as well as the residues involved in catalysis are located and a hydrophobic 'specificity pocket'. The flexibility of the binding pocket is almost entirely confined to the region of the specificity pocket, whereas the residues of the anion pocket show considerably larger spatial conservation.

As mentioned in Materials and Methods, 20 high-quality complex structures of human AR plus an apo and a holo structure are available. In contrast to the previous analysis no mutant or porcine AR are included in the dataset. Thus, the PDB structures of sorbinil and tolrestat in complex with porcine AR were replaced in the present study by their human complex counterparts. Additionally, in the last analysis only three structures of ligands addressing the specificity pocket were included: tolrestat, idd594, and zenarestat. In the meantime, many structures of inhibitors addressing this specificity pocket have been published. Also for structures solved in our laboratory the emphasis was put on ligands penetrating into the specificity pocket. Hence, to extend the scope and relevance of our previous study considering a larger dataset, we present the results analyzing structural differences of the binding pocket based on multiple ligands addressing the specificity pocket.

As a starting point for the analysis the proposed holo conformation (1ads) of the enzyme was chosen. The results for this analysis are presented in Table 6. With the exception of the

apo form of the enzyme (1xgd), all pockets superimpose very well, showing rmsd values below 0.9 Å.

protein chain	rms binding site (Å)	$\Delta C\alpha$	Δ side chain
2acr	0.10	0	0
2acq	0.12	0	0
1el3	0.16	0	0
1x96	0.19	0	0
1x98	0.20	1	0
9sbi	0.22	0	0
1x97	0.22	1	0
2acs	0.23	0	0
1pwm	0.26	1	0
9jfd	0.44	3	2
947d	0.59	6	5
9tol	0.62	10	7
1z89	0.71	5	2
1t40	0.72	5	5
1z3n	0.73	5	5
9zop	0.74	4	5
1pwl	0.76	7	6
1t41	0.76	8	5
1us0	0.77	8	5
1iei	0.81	10	6
1xgd	2.86	21	11

Table 6: Binding-site analysis based on the holo reference structure 1ads. The second column reports the the rmsd considering the $C\alpha$ atoms of the binding-site residues. The third column gives information on significant $C\alpha$ movements (if any) with respect to the reference structure using a threshold of 0.5 Å. The fourth column contains information on the number of significant side-chain movements using a threshold of 1.0 Å. The movement is measured between the centroids calculated for all the heavy atoms within the side chain, for both, reference and superimposed chains. The values are sorted according to their ascending rmsd with respect to the binding-site residues.

Based on the rms deviation, the dataset can be divided into three clusters. The first nine structures with low rmsd values ≤ 0.26 Å form the first cluster. The rmsd values of the next cluster range from 0.44 Å for the JFD complex to 0.81 Å for the zenarestat-bound structure

1iei. These 11 structures form the second cluster. The structure of the apo form (1xgd) exhibits much larger rms deviations, thus forming its own cluster; the reasons will be discussed below.

Besides the overall fit of the binding-site residues, the number of C α atoms moving by more than 0.5 Å is reported in Table 6. Again the first nine structures show very little deviations compared to the holo structure. Six of them have no C α atom displacement exceeding 0.5 Å. Only the structures with fidarestat or one of its isomers (1pwm, 1x97, 1x98) have the C α carbon of Trp219 displaced by 0.63, 0.57, and 0.50 Å, respectively.

	Trp 79	His 110	Trp 111	Thr 113	Phe 115	Phe 121	Phe 122	Trp 219	Cys 298	Ala 299	Leu 300	Leu 301	Ser 302	Cys 303	Tyr 309	noc §
1iei	0.5	0.5		0.8						0.5	2.1	1.3	1.6		0.8	8 [#]
1pwl							0.7	0.8		0.8	2.3	1.9	1.1	0.5		7
1t40									0.6	1.1	2.2	1.9	1.2			5
1t41							0.6	0.9	0.5	1.3	2.4	1.9	1.1	0.5		8
1us0				0.5				0.8	0.5	0.8	2.1	1.8	1.1		0.5	8
1z3n								0.5	1.2	1.2	2.4	1.9	1.2			5
1z89							0.8			0.6	2.2	2.0	0.9			5
947d							0.5			0.7	2.1	1.6	1.0	0.6		6
9jfd							0.6				1.5	1.2				3
9tol			0.5		0.6	0.7	0.7				1.1	1.9	1.4	1.0	0.8	9 ^{##}
9zop										1.3	2.3	1.8	1.3			4
noc [§]	1	1	1	2	1	1	3	6	4	10	11	11	10	4	2	

Table 7: Size of the C α displacements for crystal structures of the second cluster from Table 6 with respect to the reference structure 1ads, using a threshold of 0.5 Å. Additionally, in the last column and row the total number of observed displacements is presented.

§ number of changes; # the missing 2 residues are Pro 310 and Asn 160 which are shifted by 0.7 and 1.0 Å, respectively; ## the missing residue is Tyr 209 which is shifted by 0.6 Å

None of the structures in cluster 1 shows a deviation of the side-chain centers by more than 1 Å. However, for the χ_2 torsional angle of Leu300 two distinct states can be seen. The reference angle from the holo structure is \sim -135° and in many of the other complex structures of cluster 1 it falls into the range of 70-75°. The flipped hydrophobic side chain

of Leu 300 exposes very similar spatial requirements and structural features towards the binding pocket. The question arises whether this observation can be regarded as a distinct difference between the individual structures or whether the corresponding electron densities allows for an ambiguous assignment. Given the little deviations with respect to structural features between these nine crystal structures, one can conclude that all are very similar to the holo structure. As a consequence, the protein can obviously bind any of the ligands within this cluster without the requirement to perform major 'induced-fit' adaptations.

Within the second cluster the variability between the individual structures increases significantly. Between three and ten C α atoms are displaced by more than 0.5 Å with respect to the holo structure. Thus, all of these examples show significant differences between holo and complexed form of the enzyme. The JFD (9jfd) structure shows the least amount of displacements, whereas tolrestat (9tol) and zenarestat (1iei) show the most pronounced changes. Details about the residues involved in adapting to the different ligands are given in Table 7.

The C-terminal loop region ranging from Cys 298 to Ser 302 is mainly responsible for the changes related to the "induced-fit" adaptation. Except for the JFD and tolrestat complexes, the Leu 300 C α atom moves at least 2 Å with respect to the holo structure in order to accommodate this set of ligands. Interestingly, this also affects the neighboring residues: Leu 300 and 301 are displaced in all cases, whereas Ala 299 and Ser 302 are shifted in 10 out of 11 cases, with the only exception of the JFD complex. In the latter case, a comparably small moiety penetrates into the corresponding region of the binding pocket. Hence, local movements in terms of C α shifts restricted to Leu 300 and 301 are apparently sufficient to adapt the specificity pocket to this ligand.

Also Trp 219 shows a considerable amount of flexibility. Its C α position is shifted in six out of the eleven cases. This residue is attached to a larger loop region, called the 'cofactor safety belt'^{19,34}. The electron density for this loop region of the protein is often insufficiently defined, supposedly indicating elevated levels of intrinsic flexibilities. From a biochemical point of view, this loop plays an important role in binding and releasing the cofactor. Hence, flexibility of this loop region is mandatory for the biological function of the enzyme, which relates to the exchange of the cofactor as the rate-limiting step.

The C α atoms of Cys 298 and 303 are shifted four times. Cys 298 is at the N-terminal end of the small C-terminal loop region which shows the highest flexibility. Thus, movements affecting this loop region affect also the backbone of Cys 298. The position of Cys 303 is partly responsible for the size of the specificity pocket at the far end of the anion pocket. Hence, it adopts its position in a way to give sufficient room depending on the size of the ligand moiety penetrating into the specificity pocket.

Phe 122 is shifted in three cases between 0.6 and 0.7 Å. It is located in a loop region directly above Leu 300 and borders the upper part of the specificity pocket.

An interesting case is Thr 113 whose C α atom is shifted in two cases. The corresponding ligands are idd 594 and zenarestat, which both place a bulky bromine atom into the distant part of the specificity pocket, thus displacing Thr 113. Also Tyr 309, which is located

Protein Chain	Phe 115	Phe 121	Phe 122	Ala 299	Leu 300	Leu 301	Ser 302	Cys 303	Tyr 309	noc [§]
1iei	2.0			1.1	3.7	1.2	2.3		2.1	6
1pwl			1.0		4.5	4.2	1.2	1.2	1.1	6
1t40				1.1	4.6	3.9	1.2	1.2		5
1t41				1.4	4.4	4.2	1.2	1.3		5
1us0					4.2	4.1	1.1	1.0	1.2	5
1z3n				1.2	4.5	4.4	1.2	1.2		5
1z89					4.2	4.4				2
947d				1.1	4.8	3.8	1.1	1.4		5
9jfd					4.0	3.1				2
9tol		1.0	2.5	1.2	3.4	2.1	1.3	1.2		7
9zop				1.2	4.7	3.8	1.3	1.1		5
noc [§]	1	1	2	7	11	11	9	8	3	

Table 8: Movements of side-chain centers of all structures of cluster 2 from Table 6 with respect to the holo conformation of the enzyme Iads. As for Table 6 the movement is determined between the centroids calculated for all the heavy atoms within the side-chains of the reference and the superimposed chains. Displacements are reported, if they are larger than a threshold of 1 Å.

§ number of changes

below Thr 113, is shifted in the complexes with these two ligands. The other residues mentioned in Table 7 are shifted only once, as observed in the zenarestat (1iei) and tolrestat (9tol) complexes, two structures which will be discussed in more detail below.

Movements of the side-chain centers are analyzed in Table 8. As expected, similar trends are observed as reported in Table 7. The majority of the "induced-fit" movements resulting from ligand binding occur in the C-terminal loop. Leu 300 and 301 are involved in all 11 cases, whereas Ser 302 and Ala 299 are found to move in only nine and seven cases, respectively. For the Pfizer compound and the JFD compound no movements of Ser 302 are observed. This can be attributed to overall smaller movements necessary to accommodate these comparatively small ligands. Consequently, Leu 300 and 301 are the only residues which significantly shift their side-chain positions upon binding one of these two ligands.

A residue which shows a surprisingly high rate of side-chain shifts compared to its $C\alpha$ movements is Cys 303. In eight of the cases, the side chain moves away from its holo position once a ligand addresses the specificity pocket. The side chain of Tyr 309 is moved in three cases. The corresponding ligands are minalrestat, idd 594, and zenarestat. As shown above, the latter two are also responsible for the $C\alpha$ shifts of this residue, most likely induced by a bromine atom in the vicinity of Tyr 309. Interestingly, also minalrestat has a bromine at the corresponding position. Thus, a movement of Tyr 309 seems to occur only if a bulky atom or group is placed in its vicinity. Finally, the side-chain centers of Phe 115, 121, 122 are occasionally involved in the 'induced-fit' adaptations.

Not all relevant motions induced upon ligand-binding can be captured solely considering $C\alpha$ and side-chain center movements. Changes of side-chain torsional angles virtually conserving the positions of side-chain centers or $C\alpha$ atoms can also be of relevance for the exposure of recognition determinant features. All torsional angles were measured on a scale ranging from -180° to 180° . The results from this analysis are presented in Table 9. In general, the amino acids exhibiting the highest degrees of flexibility with respect to $C\alpha$ atom and side-chain center movements also show the largest torsional variances. Hence, the Cys 298 to Leu 301 loop region displays the highest diversity with respect to the corresponding side-chain torsion angles. The Cys 298 side chain adopts three distinct conforma-

tions. In the holo structure, the χ_1 angle is 167° and the sulfur points away from the binding pocket. However, the side chain can also rotate to values about 70° (1t41), orienting the sulfur atom towards the binding pocket. Interestingly, in the ultra-high resolution structure with idd594 (1us0), Cys 298 shows a split conformation with values of 158 and 74° . Additionally, values around 120 - 130° are also possible, as can be seen in the zopolrestat structure. This value represents a conformation where the sulfur lies in bisecting position between the other two orientations.

Leu 300 shows dramatic variations in both its χ_1 (holo: 110°) and χ_2 (holo: -136°) angles. If the specificity pocket is addressed by a ligand, χ_1 fluctuates between 60 and 90° . However, values between 140 and 160° can also be seen. The highest deviation can be observed for the tolrestat complex, where χ_1 adopts a value of -174° . The χ_2 angle is confined to two distinct states in the complexes, corresponding to torsions of $\sim 60^\circ$ and close to 180° . The next residue in the sequence, Leu 301, also shows a high degree of variability. However, since its side chain points away from the binding pocket into the solvent, changes are most likely not relevant for the recognition of small molecules in the binding pocket.

To accommodate tolrestat, Phe 122 χ_1 is shifted by 14° (169 to -177°). Changes of further residues are primarily observed in the zenarestat complex (1iei). Thr 113 shows a shift of the χ_1 angle from -72 to 132° . Tyr 309 also adopts a unique position: instead of forming an H-bond with the backbone carbonyl of Ala 299, Tyr 309 is rotated around its χ_1 angle and points now towards the ligand. A water molecule is found in the former position of the Tyr 309 OH atom.

The conformations observed within the zenarestat structure are in many ways unique. Positional changes are more frequent than in the other structures. However, it remains unclear whether these changes can solely be attributed to restraints imposed by the ligand, considering that this structure exhibits by far the lowest resolution (2.5 \AA) in the dataset. Hence, it cannot be ruled out that some of the observed 'abnormalities' are related to the higher uncertainty of the atomic positions.

Residue	torsion	1ads	liei	1pwl	1t40	1t41	1us0	1z3n	1z89	947d	9jfd	9tol	9zop
TYR 48	χ^1	178	166										
GLN 49	χ^1	161		172		176	173						174
GLN 49	χ^2	176		-170		-171			-168	-171	-168		
LYS 77	χ^1	-162	-143										
LYS 77	χ^2	-173	-126										
LYS 77	χ^3	-173	150							86			
LYS 77	χ^4	74									92	163	
TRP 79	χ^1	62	77										
CYS 80	χ^1	180	153										
TRP 111	χ^2	93					82			83			
PRO 112	χ^1	-103	-138										
THR 113	χ^1	-72	132										
THR 113	χ^1	173	13										
PHE 115	χ^1	-177	-67										173
PHE 115	χ^2	-15	-135										
PHE 122	χ^1	169											-177
PHE 122	χ^2	83	95						-94	-97	-92	-85	-93
ASN 160	χ^1	177	-31										
TYR 209	χ^2	68	86							57	56		54
TRP 219	χ^2	84	96										
CYS 298	χ^1	167	139		86	71		127	155				119
LEU 300	χ^1	110	72	69	91	166	62	152	79	65	140	-174	71
LEU 300	χ^2	-136	158	63	46	170	60	-175	64	59	50	175	71
LEU 301	χ^1	137	163	64	76	69	62	57	80	57	98	170	85
LEU 301	χ^2	0	-65	61	54	58	66	55	-80	62	-176	-177	33
SER 302	χ^1	-60	-81								-44	-40	
CYS 303	χ^1	174											76
TYR 309	χ^1	62	106										
TYR 309	χ^2	72	54										
PRO 310	χ^1	-93	-107										
PHE 311	χ^1	176	-169				-171						

Table 9: Torsional analysis for structures addressing the specificity pocket. Values are listed if deviating by more than 10° from the corresponding value in the holo structure 1ads. All angles are given in $^\circ$, the reference angles of 1ads are shown in bold face.

In addition to the positions of the amino-acid side chains, the conformation of the backbone plays a major role in the recognition of small ligands. Unfortunately, no built-in functions to automatically analyze changes in the backbone beyond $C\alpha$ -movements exist in Relibase⁺. Hence, these changes were assessed by visually inspecting the superimposed binding pockets and measuring the corresponding torsion angles using Reliview.

Among the residues of the binding pocket only the C-terminal loop stretch of Cys 298 to Leu 300 shows noticeable variations of the backbone which exceeds beyond simple $C\alpha$

atom shifts. Detailed analyses of the backbone torsional angles of this loop region are presented in Table 10.

	Cys 298		Ala 299		Leu 300	
	Ψ	Φ	Ψ	Φ	Ψ	Φ
1ads	92	-108	4	47	51	
1el3	97	-105	-3	46	44	
1iei	-34	44	-37	17	121	
1pwl	100	-152	157	-150	-150	
1pwm	97	-114	7	43	52	
1t40	95	-160	164	-141	123	
1t41	93	-166	162	-132	130	
1us0	107	-154	165	-156	113	
1x96	101	-114	2	48	54	
1x97	98	-110	0	47	58	
1x98	100	-108	0	46	53	
1xgd	85	-106	1	41	61	
1z3n	105	-175	167	-142	133	
1z89	97	-144	153	153	114	
2acq	97	-108	1	43	59	
2acr	98	-113	4	46	58	
2acs	91	-112	2	48	54	
947d	100	-167	75	-70	126	
9jfd	98	-143	76	-71	112	
9sbi	100	-115	-4	51	53	
9tol	75	-87	-40	50	60	
9zop	102	-155	-155	-165	112	

Table 10: Backbone torsion flexibility in the C-terminal loop region Cys 298 - Leu 300. Values of the torsion angles are given in $^{\circ}$.

Starting at the N-terminal end of the segment, the ψ angle of Cys 298 coincides in most cases between 90 and 105 $^{\circ}$. Two exceptions can be apparent: the structures of zenarestat (1iei) and tolrestat (9tol). For zenarestat the torsion changes by $\sim 130^{\circ}$ and adopts a value of -34° . This results in a flip of the carbonyl function of Cys 298, which usually points towards the binding pocket. Here, instead, this carbonyl function points upwards towards the solvent. However, this flip is partly compensated by the next two torsions, which adopt un-

usual values compared to the other structures. For the complex with tolrestat, the Cys 298 and Ala 299 main-chain torsions also adopt unique values. These torsion angles enable the unique position of the Leu 300 side chain observed in this complex. Except for 1iei and 9tol, the ϕ angles of Ala 299 cluster around two different values. If either no ligand is bound or the bound ligand does not address the specificity pocket, the values for ϕ Ala 299 fall between -115° and -105° . If the ligand, however, addresses the specificity pocket, the ϕ angle adopts values around -143° and -175° . Thus, this torsion can be considered to act as a hinge which facilitates the opening and closing of the specificity pocket.

The ψ angle of Ala 299 shows a similar but not identical behavior. If the specificity pocket is closed, the torsion angle adopts values between -3° and 7° . However, if the specificity pocket is addressed, two different states can be seen: the angle can either fluctuate between -155° and 167° or have values around 75° . The latter corresponds to a conformation where the backbone nitrogen of Leu 300 forms an H-bond with the ligand. Hence, with the open specificity pocket this conformation only occurs if the corresponding ligand presents a suitable H-bond acceptor to the Leu 300 backbone, as seen in the 47d and JFD structures. A clear dependence on the bound ligand can also be observed for the ϕ (Leu 300) angle. A closed specificity pocket corresponds to values between 43° and 51° in ϕ (Leu 300) and $44^\circ - 59^\circ$ in ψ , respectively. For the open pocket two clusters are seen for ϕ Leu 300. One shows a relatively large variability within the conformations, with values between -132° and 153° . This cluster corresponds to the conformation where no H-bond is formed with the inhibitor. Within the other cluster the ϕ angle adopts values of $\sim -70^\circ$ and the N-H of the Leu 300 is spatially aligned to facilitate H-bond formation with a bound inhibitor. This dependence of the backbone torsional angles on the presence of an H-bond to the inhibitor is lost for ψ Leu 300, where the only noticeable difference is apparent between the opened ($>110^\circ$) and closed ($\sim 50^\circ$) specificity pocket.

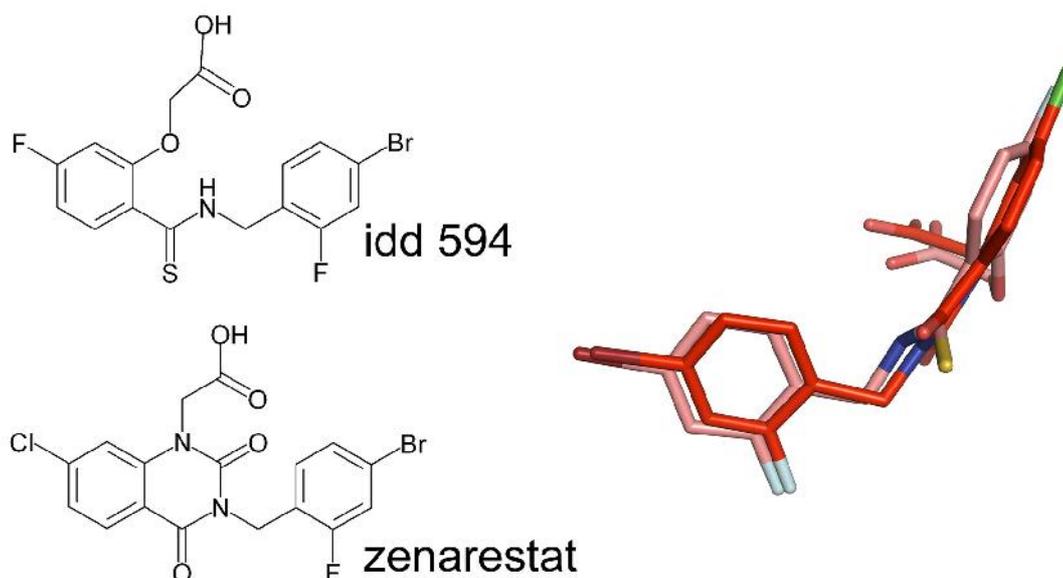


Figure 14: Chemical formulas and superposition of zenarestat (red) and idd594 (salmon) based on a fit of the binding site residues. Both structures superimpose very well and use the same moiety to address the specificity pocket.

Again one has to note that the backbone conformations for the zenarestat and tolrestat complex are unique, which can be clearly deduced from the binding mode in case of tolrestat. For zenarestat the driving forces which promote these unique backbone conformations are not easily identified. Figure 14 shows a superposition of the bound conformations of zenarestat and idd594 based on the binding-site residues fit as calculated by Relibase⁺. Both structures superimpose quite well and show the same structural features in those parts which address the specificity pocket and the C-terminal loop region. However, different conformations of many binding-site residues are observed in the two complexes, for reasons which remain ultimately unclear apart from the mentioned lower resolution of 1iei.

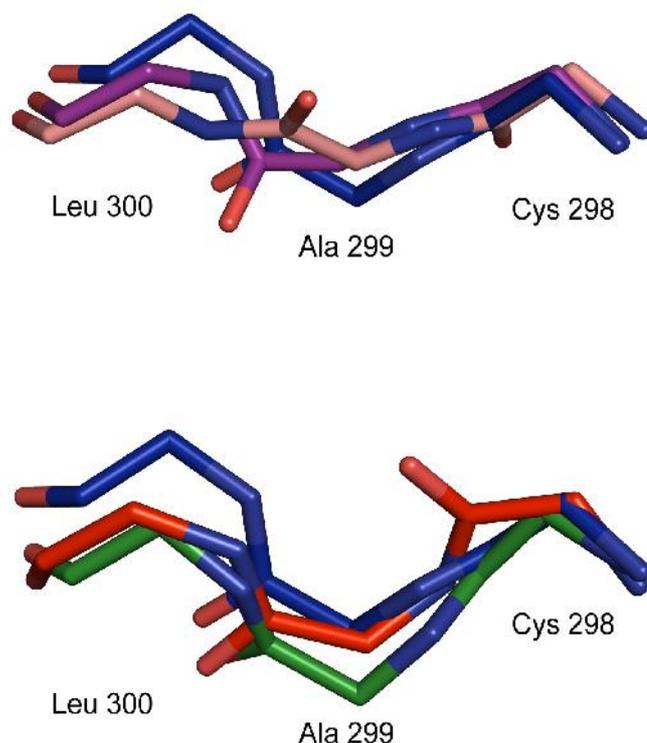


Figure 15: Backbone conformations of the Cys 298 - Leu 300 stretch. In the upper picture the three major conformational clusters are given. The holo conformation (1ads), which represents the closed specificity pocket, is shown in blue, the conformation where the specificity pocket is open and no hydrogen bond is formed with the ligand is shown in salmon (1us0) and the corresponding conformation with an H-bond is given in purple. In the lower picture the holo conformation is shown in blue, whereas two unique conformations are given in green (tolrestat) and red (zenarestat).

By analyzing the backbone flexibility of the C-terminal loop region it becomes clear that this part of the protein is of utmost importance for understanding the mechanism which enables the enzyme to accommodate a variety of different ligands. As shown by the analysis of the torsional angles, within this region the leverages are located which enable the opening and closing of the specificity pocket. Figure 15 shows representative examples for the different conformations found for the backbone of this small loop stretch. In the upper picture the three major conformational clusters are displayed. The holo conformation (1ads) representing the closed specificity pocket is shown in blue, the conformation of the open specificity pocket with no hydrogen bond formed towards a suitable ligand moiety is

shown in salmon (1us0), and the corresponding conformation with an H-bond present is given in purple. In the lower picture the two remarkably different conformations of tolrestat (green) and zenarestat (red) are shown in comparison to the holo conformation.

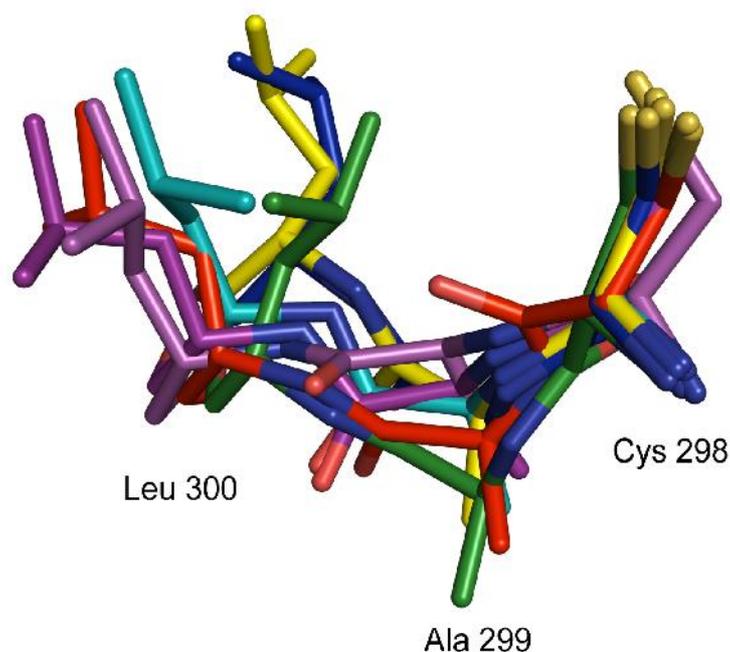


Figure 16: Representative structures which illustrate the conformational flexibility of the C-terminal loop region including the side chains. The backbone flexibility, which is illustrated in Figure 15, is amplified by the Leu 300 side chain. Thus, this residue shows enormous positional variability.

Considering movements of C α atoms and torsion angles in an isolated way is clearly not sufficient to describe the full conformational space of the AR binding pocket. Accordingly, Figure 16 shows a composite picture of the residues Cys 298 to Leu 300 for selected structures which represent the major conformations seen in the dataset.

Departing from Cys 298, the side chain of this residue adopts three different conformations. As indicated by the side-chain analysis, the sulfur of Cys 298 can either point into or off the binding pocket, and it can adapt an intermediate orientation.

Moving on to Ala 299, a remarkable flexibility in the backbone torsions is revealed. On the one hand, this flexibility enables a flip of the main-chain conformation as indicated by in the different orientations of the carbonyl function. On the other hand the opening and closing of the specificity pocket is facilitated by this part of the backbone.

For Leu 300 three main conformations can be distinguished: the first corresponds to the closed form of the specificity pocket, the second to the opened form, and the third to the unique conformation of the tolrestat-bound structure. Within the cluster of the open specificity pocket conformations one can further dissect the backbone conformations with respect to the already discussed formation of an H-bond to the ligand. Independent of the opened or closed state of the specificity pocket the side chain of Leu 300 itself can adopt different conformations in which the CG atom points either towards or off the binding pocket.

In summary, what picture can be drawn about the binding pocket of AR and its 'induced-fit' effects?

Based on the analysis of the crystal structures one can divide the binding pocket of AR into two different sub-pockets. The first sub-pocket is the cavity where the catalytic reaction takes place. As the name suggests, the residues involved in catalysis^{35;169;170} as well as the cofactor²² border this cleft. The corresponding residues are Trp 20, Val 47, Tyr 48, Trp 79, and His 110. The positions of these residues are very well conserved among the different structures. Despite very low thresholds, hardly any variations could be noticed. This behavior has been described earlier¹¹³. Such a behavior is to be expected for a sub-site within a binding pocket which contains the catalytic residues¹⁷¹. In order to catalyze the reduction of aldehydes, a precise alignment of the substrate, the cofactor as well as the residues involved in stabilizing reaction intermediates is mandatory. Therefore, structural integrity is a key feature for this sub-pocket.

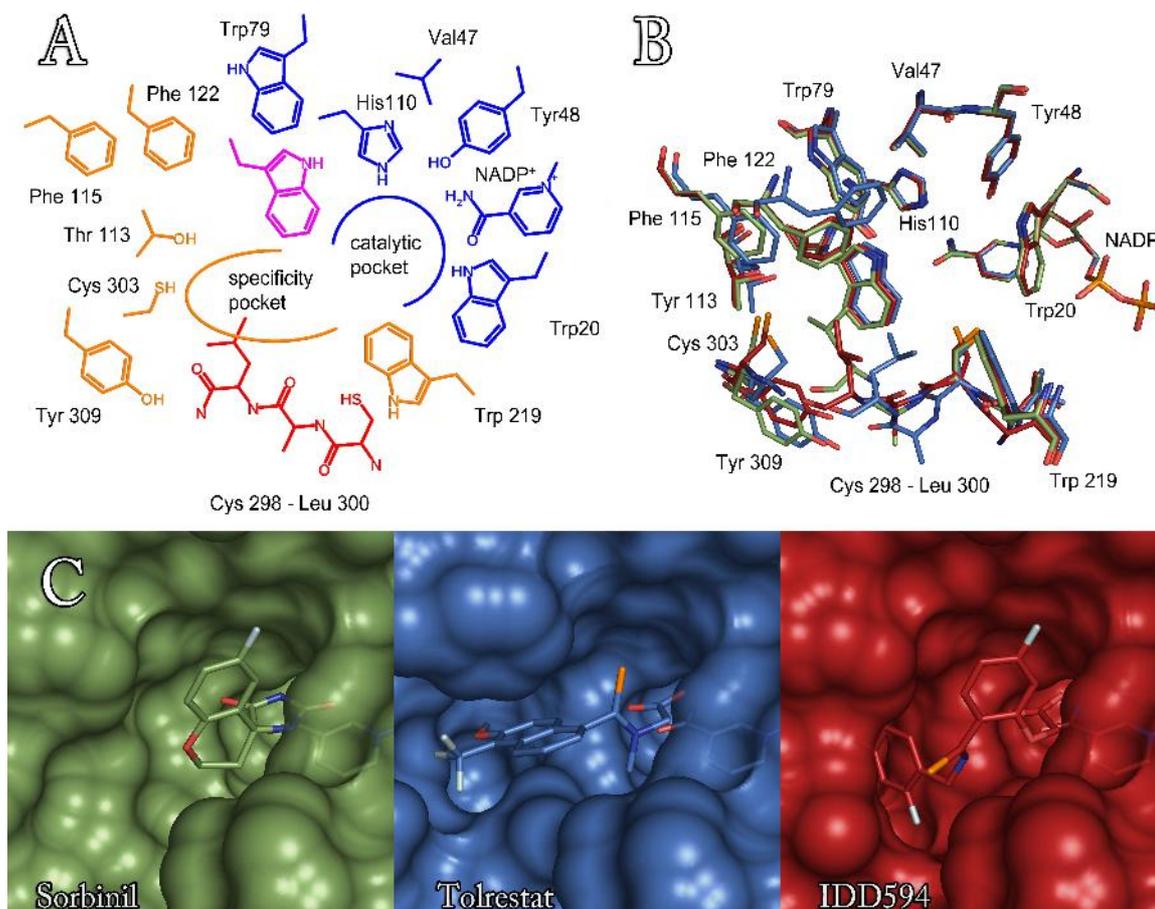


Figure 17: Different representations of the AR binding pocket. **A** shows a 2D representation of the AR binding pocket with its two sub-pockets. The amino acids are color coded with respect to the flexibility observed in the comparative crystal structure analysis. Blue indicates the least and red the highest amount of flexibility. **B** depicts a superposition of the binding-site residues of the three major binding pocket conformations. The closed specificity pocket (sorbinil complex) conformation is shown in green, the two major opened conformations are shown in blue (tolrestat) and red (idd 594). **C** shows a surface representation of the different binding pocket conformations with the corresponding ligand shown as sticks. The chosen complexes and the color coding are the same as for **B**.

The second important sub-pocket for the functionality of the enzyme is the specificity pocket. It is a hydrophobic pocket which is composed by the residues Thr 113, Phe 115, Phe 122, Leu 300, Cys 303, and Tyr 309. The accessibility and shape of this pocket varies

depending on the bound inhibitor. Despite this variability, the available crystal structures suggest that the intrinsic flexibility is limited to only a handful of residues.

Opening and closing of the specificity pocket is mainly determined by Leu 300 as gate-keeper. The spatial position of the Leu 300 side chain with respect to Trp 111 determines whether the specificity pocket is in an open or closed state. Trp 111 plays an important role within the binding pocket of AR: it connects the catalytic anion pocket and the specificity pocket. The nitrogen of its aromatic side chain helps to anchor bound inhibitors to the active site of AR by forming an H-bond with the respective polar head group of the inhibitor. The aromatic ring system of the side chain is involved in hydrophobic interactions with the aromatic or hydrophobic moiety of the ligand penetrating into the specificity pocket. If no ligand is bound, the side chain performs hydrophobic contacts to the Leu 300 side chain, thereby closing the specificity pocket.

While Leu 300 can be considered as amplifier for the opening mechanism of the specificity pocket, the analysis of the available crystal structure showed that the whole loop stretch from Cys 298 to Leu 300 is involved in regulating the opening and closing as well as the shape of the specificity pocket. Within the backbone of this loop region intrinsic flexibility is found, which is amplified by the Leu 300 side chain. In Figure 17 A a schematic overview of the AR binding pocket is presented and the amino acids are colored according to their observed flexibility, ranging from 'rigid' (blue) to highly flexible parts (red).

Superimposed structures of the binding-pocket residues representing the corresponding conformations are shown in Figure 17 B, whereas Figure 17 C shows the different shapes of the pockets. As expected, the catalytic pocket superimposes very well whereas spatial diversity is observed for the specificity pocket.

It has been shown in previous studies, that the binding modes of ligands addressing the specificity pocket in an open conformation are incompatible with the other binding pocket conformations¹¹³. However, this issue deserves revision: How different are the binding modes of inhibitors addressing the same of the three overall pocket conformation? How do these differences compare with differences seen for inhibitors addressing one of the parent pocket conformations? To analyze this issue, four complex structures representing the dif-

ferent binding-pocket conformations were selected: sorbinil for the closed conformation; idd594 for the open pocket with no H-bond formed between Leu 300 and the ligand; 47d for the similar open pocket but with this H-bond present, finally, tolrestat representing its own unique binding conformation.

To analyze the compatibility of the binding modes of the four reference ligands with respect to the other structures in the dataset, all structures were subsequently superimposed to the four representative binding pockets, respectively. The results are shown in Table 11, where the number of clashes between the ligands bound to the four above-mentioned corresponding reference structures and all other superimposed structures of the dataset is given. A clash is defined as a distance between two atoms smaller than the sum of their vdW radii as implemented in Relibase⁺.

The binding mode of sorbinil (9sbi) is largely compatible with most of the other complex structures. This can be expected since it is a rather small ligand which binds to the closed form of the specificity pocket. Between zero and three clashes are observed. Sorbinil would experience most clashes if placed into the tolrestat binding pocket (5 clashes). No clashes are observed for the binding pocket of two ligands which bind to the closed form of the specificity pocket (1ads, 1el3). Interestingly, also the binding pocket conformation of the Pfizer compound complex (1z89) could well accommodate sorbinil in its binding mode, although the Pfizer ligand binds to the open state of the specificity pocket. Most of the clashes between the bound conformation of sorbinil and the other binding pocket conformations occur with Tyr 48, which is surprising, since Tyr 48 superimposes very well across all structures. However, the comparatively large hydantoin head group of sorbinil might be responsible for this unexpected finding. Furthermore, Cys 298, Phe 122, His 110, and Trp 111 show small clashes with sorbinil in some of the structures. No major differences in terms of the number of clashes can be detected between structures which address the specificity pocket to those who do not.

Idd 594 (1us0) shows a different behavior. Structures addressing the specificity pocket show between three and six clashes with the bound conformation of idd 594. The involved residues are Trp 111, Phe 122, Trp 219, and Ala 299. Interestingly, the most compatible binding pocket is formed in the 47d complex, where, in contrast to the idd 594 complex, an

H-bond between the ligand and Leu 300 is formed. For complexes which do not open the specificity pocket, the number of clashes with idd 594 increases to eight to 11 bad contacts. Not surprisingly, Leu 300 and Trp 111 are responsible for most of the clashes, since these residues are responsible for opening the specificity pocket.

	reference complex structure			
	idd 594 (1us0)	sorbinil (9sbi)	47d (947d)	tolrestat (9tol)
1ads	8	None	10	14
1eB	8	None	10	14
1iei	3	3	6	11
1pwl	4	1	3	12
1pwm	9	1	10	15
1t40	3	3	3	11
1t41	6	2	5	10
1x96	9	2	11	14
1x97	8	3	11	14
1x98	9	2	11	14
1xgd	9	1	10	12
1z3n	5	1	2	9
1z89	3	None	1	11
2acq	11	2	12	15
2acr	9	2	12	15
2acs	10	3	13	15
947d	2	2	-	11
9jfd	8	2	9	12
9sbi	9	-	10	13
9tol	14	5	13	-
9zop	3	3	4	13
1us0	-	2	2	11

Table 11: Analysis of the mutual compatibility of different binding modes. Given is the number of clashes of the ligand bound to the respective reference structure with the corresponding structure in column 1. A clash is defined as a contact shorter than the sum of the vdW radii as implemented in Relibase+. All structures corresponding to the same cluster of Table 6 are colored equally: closed specificity pocket (green); opened no H-bond (red); opened with H-bond (yellow); tolrestat conformation (blue). Within the columns 2 to 5, structures belonging to the same cluster as defined in Table 6 are colored in light gray.

The situation for the structures exhibiting a hydrogen bond between the ligand and Leu 300, as represented by the 47d complex, is similar to the 1dd 594 case. Between ten and 13 clashes with the bound conformation of 47d are observed for conformations where the specificity pocket is in a closed state. As expected, complexes with ligands addressing the open conformation show significantly less (1-5) clashes.

The most incompatible binding mode with respect to all complexes in the dataset is the one of tolrestat. At least 9 clashes (mostly 11-15) are present for any of the comparisons. Leu 300 and Phe 122 clash in every structure, but there is virtually no binding-site residue which is never involved in any of the clashes.

In summary, none of the larger reference ligands opening the specificity pocket would be fully compatible with any of the other structures even if they address the same overall binding-pocket conformation. This illustrates that the protein undergoes slight adjustments and adaptations that are specific for the bound molecule.

One structure published by Bohren et. al.³⁴ in 2005 has been neglected so far but deserves discussion in greater detail: the crystal structure of the apo form of AR at 2.1 Å resolution. To obtain this structure it was necessary to introduce the R268A mutation. In the holo form of AR this Arg forms a salt bridge to one of the phosphate groups of NADPH. By removing this interaction site, the affinity of AR towards its cofactor is sufficiently reduced to obtain a structure with an unoccupied cofactor binding pocket.

In Figure 18 A, a superposition of the binding site residues of the sorbinil complex (9sbi) and the apo structure (1xgd) based on a C α -atom fit is shown. Overall, the binding-site residues match very well, except for some differences in the side-chain positions of Trp 20. This residue is adjacent to the cofactor and it is reasonable that its position is shifted upon NADPH binding.

However, not all parts of the structure superimpose with similar low rmsd values. In Figure 18 B the area surrounding the cofactor-binding cleft is shown. The most striking difference between the two structures is the position of the 'cofactor safety belt'. This loop region hosting residues 214 to 226 is characterized by high temperature factors. Trp 219, which is usually located adjacent to the binding pocket, is moved by as much as 17 Å go-

ing from apo to holo conformation. Instead of forming a sulfur-aromatic interaction with Cys 298, Trp 219 now experiences a cation- π interaction with the side chain of Arg 293.

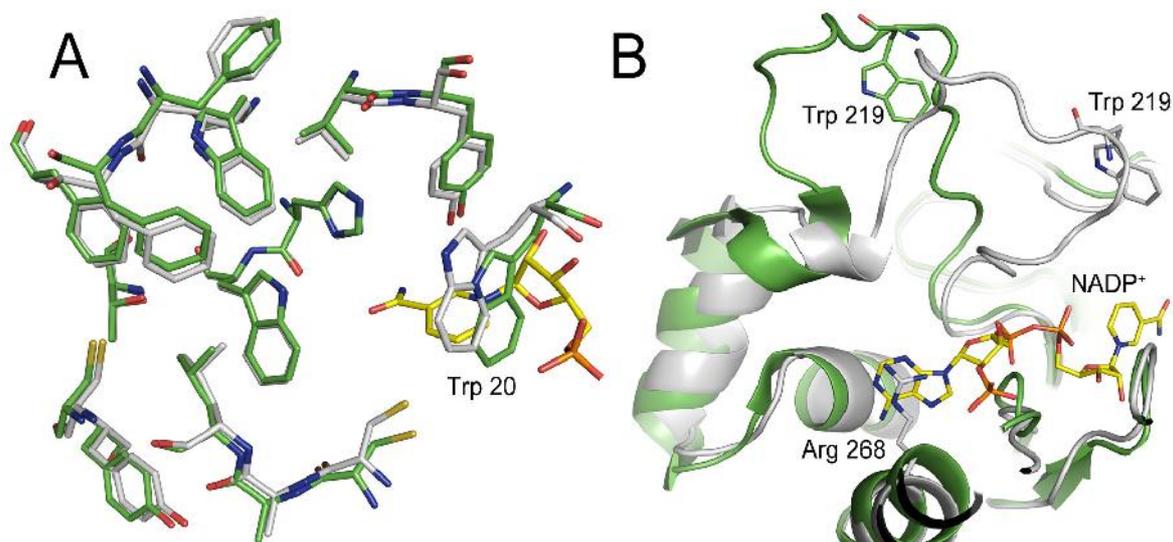


Figure 18: Superposition of the apo structure (1xgd, green) and the sorbinil complex structure (9sbi, silver) based on a Ca fit for all protein residues. The cofactor (yellow) of the sorbinil structure is included for orientation purposes. A shows the superposition of the binding pocket residues. Except for Trp 219 all residues superimpose very well. B shows a view towards the cofactor binding pocket. The dramatic shift in the orientation of the cofactor safety belt can be observed. Trp 219 is moved 17 Å between the two structures. Arg 268 from the sorbinil complex is also included: in the complex structure it forms a Coulomb interaction with the C2-phosphate group of NADP⁺. The Arg268Ala mutation eliminates this interaction.

Considering the superposition of the amino acids adjacent to the cofactor safety belt, it is interesting to notice that the residues preceding the loop region overlap very well, whereas this is not the case for the residues following the C-terminal end of the loop region. Due to the movement of the whole loop, the cleft where the cofactor binds is uncovered and the mandatory exchange of the cofactor preceding each reaction cycle is enabled. This large movement is most likely facilitated by the backbone torsional changes of the three residues Gly 213 ($\Delta\Psi=162^\circ$), Ser 214 ($\Delta\Psi=85^\circ$, $\Delta\Phi=71^\circ$), and Ser 226 ($\Delta\Psi=59^\circ$, $\Delta\Phi=70^\circ$). As for the 'induced-fit' adaptations in the small C-terminal loop region from Cys

298 to Leu 300, the enzyme uses the backbone conformation of a few residues as hinges about which the loop is moving in a relatively rigid fashion.

Comparative MD simulations analysis

The comparative analysis of crystal structures describes the experimentally observed conformational space of the binding pocket, thus it can be seen which parts of a binding pocket tend to adopt different conformations or are restrained to one conformation.

However, X-ray structures represent both, time and spatial averages and are measured at rather low temperatures (100 K). The analysis of crystal structures might suggest that proteins are rigid frozen-in systems with some flexible regions that cannot be located in the experimental electron density. The ultra-high resolution 0.66 Å structure of AR¹¹⁴ indicates that the rigidity seen in crystal structures exaggerates reality due to limited resolution. Almost one third of all amino acids in this structure are present in multiple conformations, indicating higher degrees of flexibility than normally expected. The same experience could be collected with other systems resolved to very high resolutions.

To overcome these limitations of crystal structure determinations, complementary techniques have to be consulted. Nuclear magnetic resonance spectroscopy (NMR) can be used to elucidate the three-dimensional structure of macromolecules and to detect binding of small molecules to proteins¹⁷². Therefore, it is frequently used in structure-based drug design¹⁷³⁻¹⁷⁵. Instead of one definite model, NMR experiments result in an ensemble of conformations which are in agreement with the spatial restraints set by the experiment. Thus, NMR probably indicates in a more appropriate way where proteins exhibit flexibility^{176,177}. However, not all proteins are suitable for NMR experiments, particularly since the size of a macromolecule can become the limiting factor.

From a computational point of view molecular dynamics (MD) simulations are the method of choice to assess the dynamic properties of a protein¹⁷⁸. Using MD simulations, a trajectory of snapshots is generated representing the states of the system at a given point in time. Due to the inherent assumptions and simplifications of the model used for the simulations, formation and breaking of bonds cannot be simulated. Conformational changes and flexibility of amino acids, however, are certainly within the scope of the method. There-

fore, MD was chosen as a tool to obtain deeper insights into the flexibility of the AR binding pocket.

Simulations of AR complexes

To examine the intrinsic flexibility of AR, MD simulations of seven AR complexes were carried out. In addition, three simulations of the holo enzyme were conducted. All MD simulations were started from experimentally derived crystal structure coordinates (see Materials and Methods) In the following section each simulation will be discussed briefly. After that the most interesting parts of the binding pocket, which show the largest amount of flexibility will be investigated further.

MD of AR in complex with sorbinil

The rigid inhibitor sorbinil binds to the enzyme with closed specificity pocket and shows a free energy of binding of $\Delta G = -9.1$ kcal/mol.

It utilizes a hydantoin moiety to anchor to the catalytic sub-pocket of AR. The analysis of crystal structures (see above) has shown that for the sorbinil complex the binding pocket adopts a conformation which is likely to be very similar to the holo form of the enzyme. Figure 19 depicts some of the key data to analyze the MD simulation of this complex. The average $C\alpha$ rmsd across the entire simulation is $1.16 (\pm 0.10)$ Å with respect to the starting structure. Thus, the overall structure and fold of the protein are very well conserved during the simulation. The average rmsd value for the binding site residues is $0.93 (\pm 0.13)$ Å. For the first picoseconds an increase of the $C\alpha$ rmsd can be observed. This increase is not reflected for the binding site residues where no drift can be seen for the first part of the simulation.

Analysis of the 2d-rms plot for the binding-site residues fluctuations indicate a clear separation into three clusters. Two larger clusters emerge at the beginning and the end of the simulation. As indicated by the 'cross-correlation' between these two clusters they are similar to each other, whereas the smaller cluster in the central part is different compared to the other two. However, for a short period within the the small cluster the system returns to a conformation similar to the other clusters as indicated by the small 'cross-peak' next to frame 365. This clustering of the binding pocket conformations is also well reflected by the

binding site rms deviations shown in the center right plot of Figure 19. The deviating smaller cluster is accompanied by increased rms values in the same time frame. Figure 19 also depicts the structural changes within the binding pocket that cause the clustering observed in the 2d-rms plot. The first cluster is represented by the conformation shown in green, the corresponding structure from the second cluster is shown in silver. The most striking difference between these two snapshots is the conformational change of Phe 122. Its side-chain rotates and now points away from Trp 79. In a representative geometry of the third cluster, shown in yellow, the Phe 122 side-chain has flipped back and adopts an orientation similar to the first cluster.

Within the last cluster two states deviating from the rest of the simulation are observed. One is located around frames 445-450 and the other around frames 540-560. The structural change causing this deviation consists mainly of a movement at the position of Trp 219.

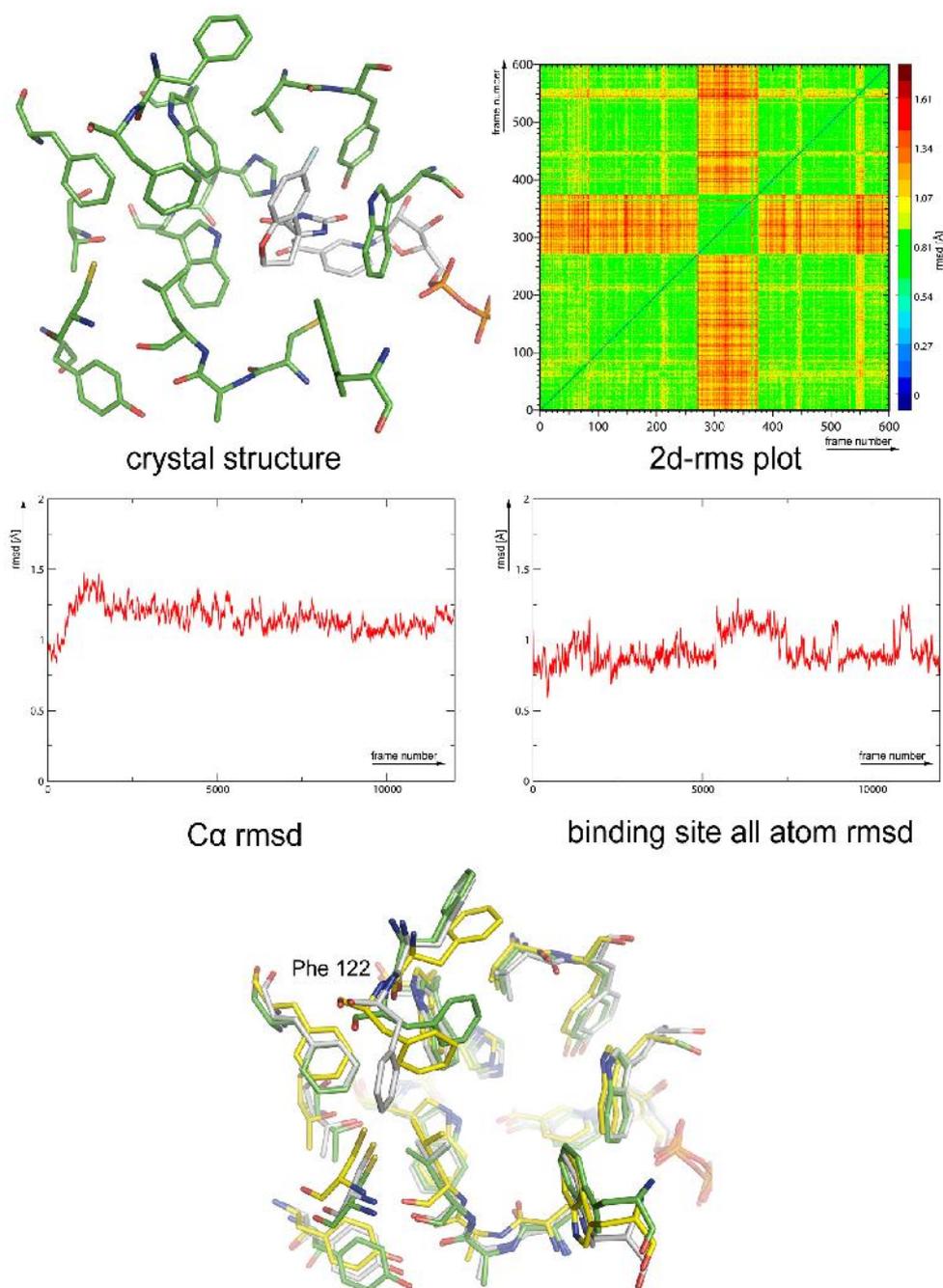


Figure 19: Overview over the MD simulation of sorbinil in complex with AR. In the upper left corner the starting conformation of the MD simulation is shown as obtained from the X-ray experiment. On the upper right corner a 2d-rms plot (see Materials and Methods) for all atoms being part of binding side residues is shown. In the center the rms values are presented as running averages with a window size of 20 snapshots (left: Ca atoms of the enzyme, right: all atoms of binding site residues). In the lower part the following representative snapshots for the clustering are given: 150 (green), 320 (silver), 500 (yellow)

MD of AR in complex with tolrestat

As observed in the crystal structure analysis, tolrestat forces the specificity pocket of AR into a unique conformation, thus it is interesting to study whether this conformation affects the flexibility of the pocket. Tolrestat uses a carboxylate moiety as an anchor group and binds to AR with a free energy of binding of $\Delta G = -9.2$ kcal/mol.

An overview over the MD tolrestat complex simulation is given in Figure 20. The average $C\alpha$ rmsd for the whole simulation is $1.24 (\pm 0.16)$ Å. The corresponding rmsd for the binding site residues is $1.18 (\pm 0.17)$ Å. Those for the $C\alpha$ -atoms are slightly higher than in the case of the sorbinil simulation. Considering the evolution over time of the rmsd values it becomes obvious that in the central part of the simulation the system drifts towards higher rmsd values. However, over the course of the simulation the system returns back to lower rmsd values compared with the starting geometry. These deviations mainly result from movements within the cofactor safety belt region. Especially the region next to Pro 222 shows substantial movements as indicated in Figure 21. The rms deviations of the $C\alpha$ atoms parallel the rmsd values of the binding site residues. However, the mobility of the latter portions is not as pronounced as for the $C\alpha$ atoms. These movements have only minor impact on the binding site since Trp 219, the only residue close to this region, shows no unusual movements during this time span.

Concerning the 2d-rms plot a clear clustering is observed. Three major clusters can be extracted, two larger clusters in the beginning and at the end of the simulation and a smaller one in the center. As indicated by the cross-correlations the similarity between the first and last cluster is higher than for each of them and the central part. Within the first cluster a short period can be detected where the system fluctuates between two distinct states. One is similar to the first and last cluster, the other shows remarkable similarity with the smaller cluster in the center.

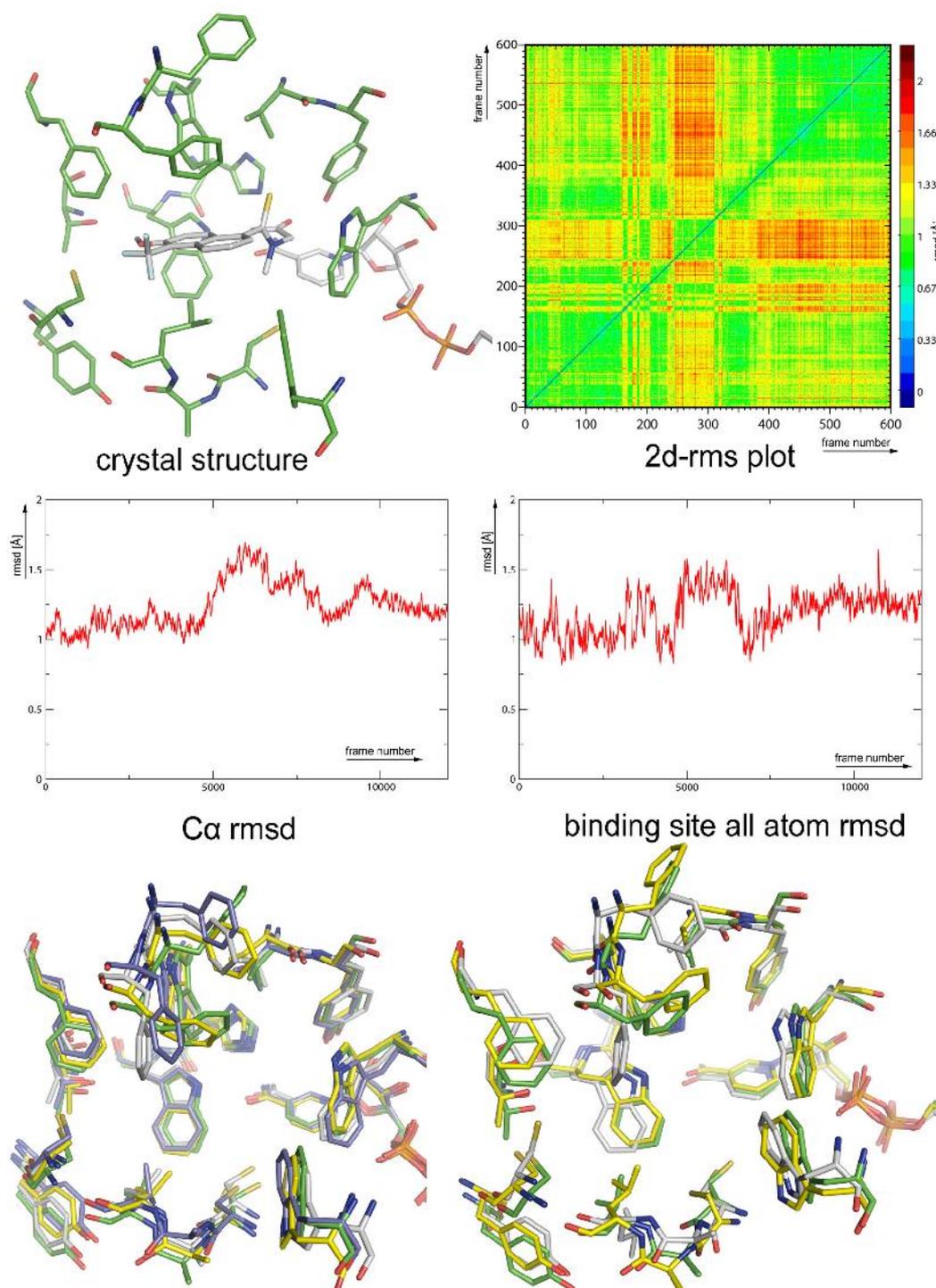


Figure 20: Overview over the MD simulation of the tolrestat complex. Data are presented similarly to Figure 19. Lower left column: snapshots representing the movements in the first cluster from snapshots 1-250: 100 (green), 163 (silver), 173 (yellow), 183 (blue). Lower right column: representatives for the three main clusters: 100 (green), 280 (silver), 450 (yellow)

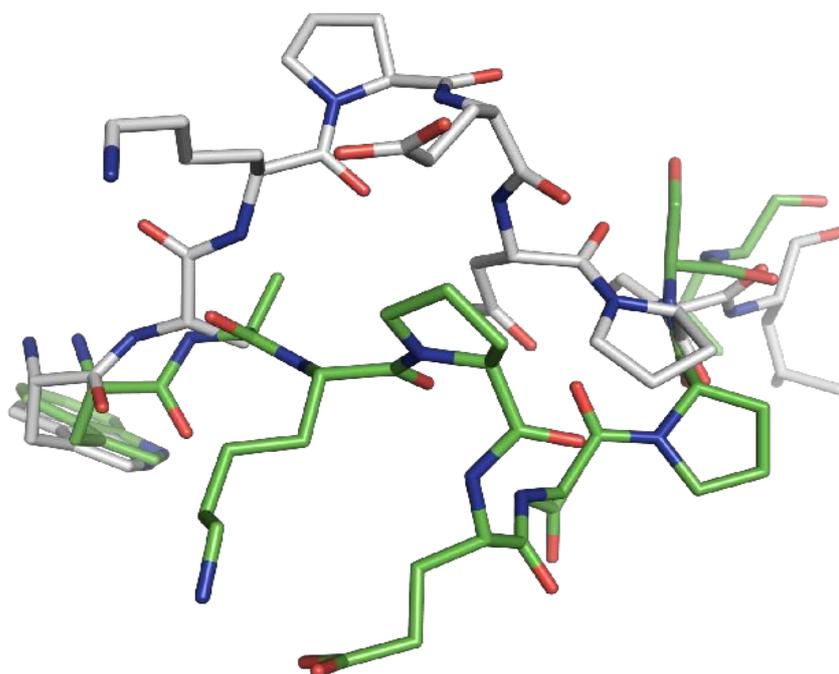


Figure 21: Movements of the cofactor safety belt region in the tolrestat complex simulation. Two snapshots depicting Trp 219 – Leu 227 from the tolrestat complex simulation are shown. The green conformation represents the first main cluster of the simulation and the silver structures shows the movements in this region which cause the formation of the smaller cluster in the middle of the simulation.

At the bottom of Figure 20 the structural changes which cause the clustering of the snapshots are depicted. In the lower left column, the conformational transitions occurring in the first large cluster are shown. Starting from the conformation shown in green the first transition at frame 160 is characterized by a concerted motion of the side-chains of Phe 121 and Phe 122. In the conformation shown in silver both side-chains adopt a different conformation as previously. The next transition is characterized by a movement of Phe 122 back to its original conformation while Phe 121 remains at the new position (yellow conformation). At frame 180 Phe 122 flips back into the intermediately populated conformation. Parallel to these side-chain movements the loop region around Ala 299 shows backbone movements. The lower right part of Figure 20 shows the changes associated with the transitions between the three main clusters. The variability of the side-chain conformations

of the Phe 121 and 122 is the main event observed in the binding pocket. In addition, side-chain motions of Cys 298 and Cys 303 are observed towards the end of the simulation.

In summary, the cofactor safety belt region next to Pro 222 is the most flexible part in this simulation, whereas the side chains of Phe 121 and 122 turn out to be the most flexible binding site residues. Compared to the sorbinil simulation, the backbone next to Ala 299 shows enhanced flexibility and additional side-chain flips of Cys 298 and 303 can be observed.

MD of AR in complex with idd 594

The complex of the inhibitor idd 594 and AR has been solved to an unprecedented resolution of 0.66 Å¹¹⁴.

As in the case of tolrestat a carboxylate moiety is used as anchor group in the anion binding pocket. Idd 594 addresses the specificity pocket *via* an aromatic moiety. A special feature of this ligand in the crystal structure is the placement of a bulky bromine atom into the distant part of this pocket forming a short contact to the side-chain oxygen of Thr 113. Idd594 binds with a free energy of binding of $\Delta G = -9.8$ kcal/mol.

Figure 22 shows the summary for the simulation of this complex. The mean C α rmsd is 1.08 (± 0.08) Å. Both the absolute value as well as the standard deviation are lower than in the previous two simulations. The overall system remains remarkably stable over the course of the simulation. The corresponding values for the binding site residues are 1.40 (± 0.20) Å. Surprisingly, these values are higher than in case of sorbinil and tolrestat.

For this simulation, no distinct clustering in conformational families is indicated by the 2d-rms plot as in the other cases. The similarity between most of the snapshots throughout the simulation is very high. At the beginning, a small cluster can be seen expanding from frame 20 – 40. In addition, a cluster which clearly separates from the rest of the simulation can be found between frames 520 and 530. The associated changes are depicted in Figure 22.

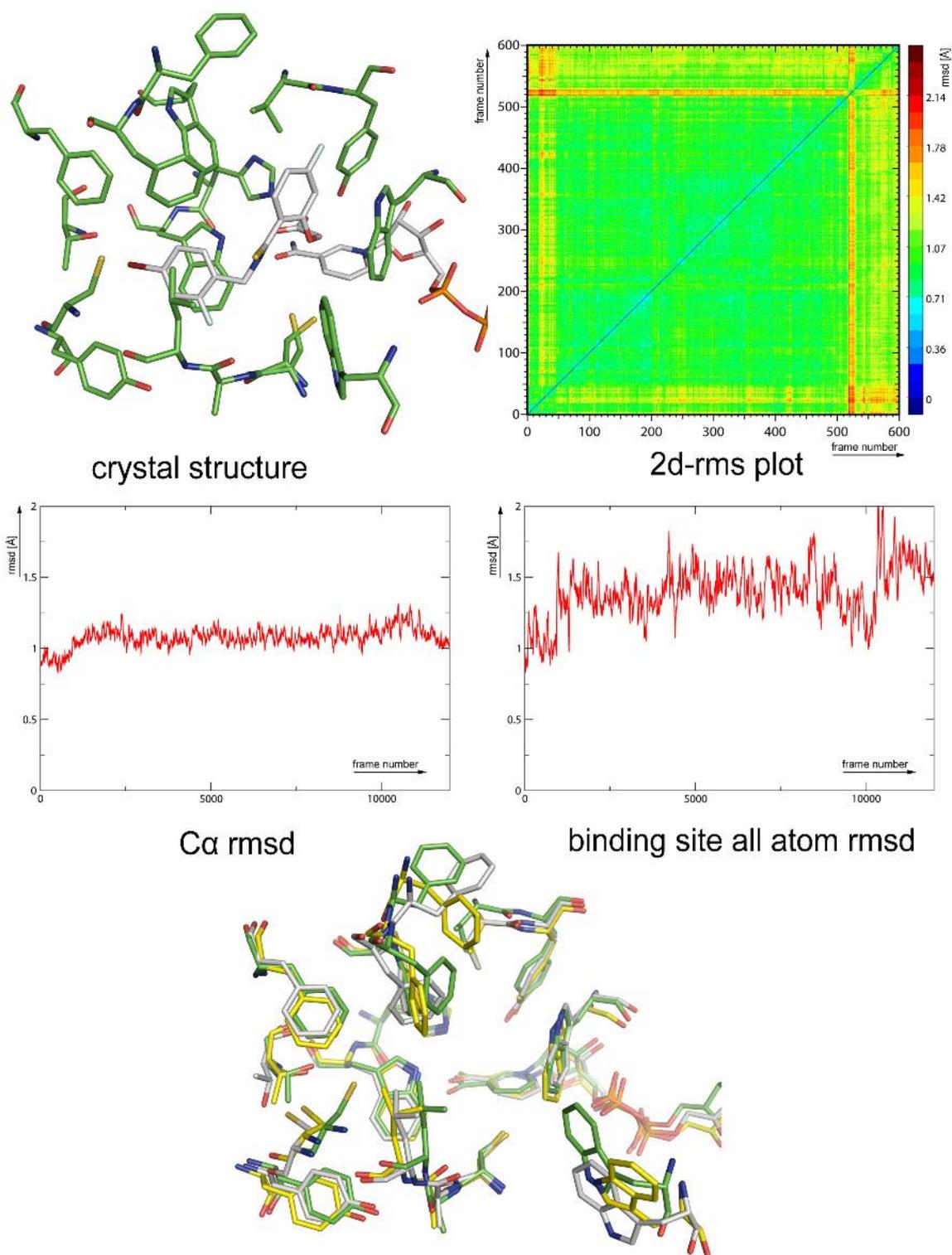


Figure 22: Overview over the MD simulation of the idd594 complex. Data are presented similarly to Figure 19. The snapshots representing the clustering are: 30 (green), 250 (silver), 525 (yellow)

The conformation describing the longest time span of the MD simulation is shown in silver. Thr 113 shows a flipped conformation of its side chain compared to the crystal structure, in which the oxygen of the side chain of Thr 113 forms an unusually short contact to the bromine moiety of the idd 594 ligand. The distance between these two atoms in the crystal structure is 2.97 Å, which is shorter than the sum of the van-der-Waals radii for an oxygen and a bromine. As expected the force field is not able to model this feature correctly, therefore the Thr 113 side chain moves away to avoid the short contact.

The conformation representing the first cluster is shown in green. Three interesting observations can be made for this cluster. The backbone of Ala 299 adopts a different conformation. Usually for this ligand the backbone nitrogen of this residue is oriented towards Tyr 309 in order to form an H-bond with its side-chain oxygen. Within this cluster, however, the nitrogen points towards the ligand which requires the observed flip of the backbone geometry in this region. The second feature is a rotation in the side-chain conformation of Cys 303. The third difference is the position of Trp 219 which is also involved in forming the most distinct cluster (shown in yellow) at the end of the simulation. In this case, the side chain rotates and presents its nitrogen atom towards the ligand. The more pronounced changes observed for this cluster are the movements of the side chains of Phe 121 and 122. These two residues adopt conformations similar to those observed in the previous simulations.

The results obtained from a previous simulation of the same protein-ligand complex carried out by Sotriffer et al.¹¹³ match qualitatively the results obtained for this study. One has to note, however, that the structural changes and the clustering of the conformations of the binding pocket is not as pronounced as found by Sotriffer et al.¹¹³. The main differences between the two simulations are the force field used to simulate the ligand and the starting structures. The general amber force field (GAFF)¹⁶⁰ was used in this study, whereas manual parameterization based on the Cornell et al. force field¹⁵⁸ was applied in the earlier simulation. In addition the previous simulation was based on a preliminary refinement of the ultra-high resolution complex structure of idd 594 whereas this study used the final structure as found in the PDB¹¹⁴. Therefore, a quantitative agreement between the two simulations is not to be expected.

MD of AR in complex with the Pfizer compound

When data about this compound were first published by Mylari et al.¹²⁵ of Pfizer, the authors claimed that this molecule was the first sub-nanomolar inhibitor of AR based on IC₅₀ values recorded from a kinetic assay. Free energies of binding derived from ITC experiments in our laboratory¹²⁶ indicated that binding of this compound to AR was significantly weaker than suggested by IC₅₀ determination. According to these measurements the compound is a low-micromolar binder. The Pfizer compound uses a pyridazinone moiety to anchor into the catalytic sub-pocket of the enzyme. The crystal structure¹²⁶ shows that it addresses the specificity pocket similar to idd 594.

The corresponding simulation summary is presented in Figure 23. The mean C α rmsd is 1.06 (\pm 0.09) Å. Again, this is a minor deviation along a 6 ns simulation, indicating a high structural integrity of the overall geometry of the complex. The rmsd for the residues comprising the binding pocket is 1.18 (\pm 0.14) Å.

In the beginning of the simulation low populated clusters representing short term fluctuations are observed. For the remainder of the simulation the system becomes less prone to changes in conformations and remains within one large cluster. Within the last nanosecond the binding pocket adopts a different conformations. However, the structural differences between this last cluster and the preceding snapshots are rather moderate.

The changes evolving in the first turbulent phase of the simulation are presented in the lower left part of Figure 23. Three regions of the binding pocket are responsible for the different states indicated by the cluster analysis. Phe 122 fluctuates between two distinct side-chain conformations, which is in line with the conformations seen in other simulations. Trp 219 shows flexibility in its backbone, which is part of the 'cofactor safety belt'. No pronounced changes are apparent for its side chain. The C-terminal loop region from Cys 298 to Leu 300 is the third flexible region. In the present case, no major changes in the main chain conformation are observed. Instead, the main chain moves as a whole within this region. Affected by these movements is the position of Tyr 309 which follows the fluctuations of the loop region to a certain degree. After this period of elevated flexibility in the binding pocket, the system adopts a conformation which is stable throughout the remaining part of the simulation. The picture in the lower right column of Figure 23 depicts the slight

changes occurring within the last nanosecond. The transition into the distinct cluster at the final end of the simulation is mainly caused by flips of the side chains of Cys 298, Cys 303, and Thr 113. All of these residues have shown conformational changes of their respective side-chains in other simulations.

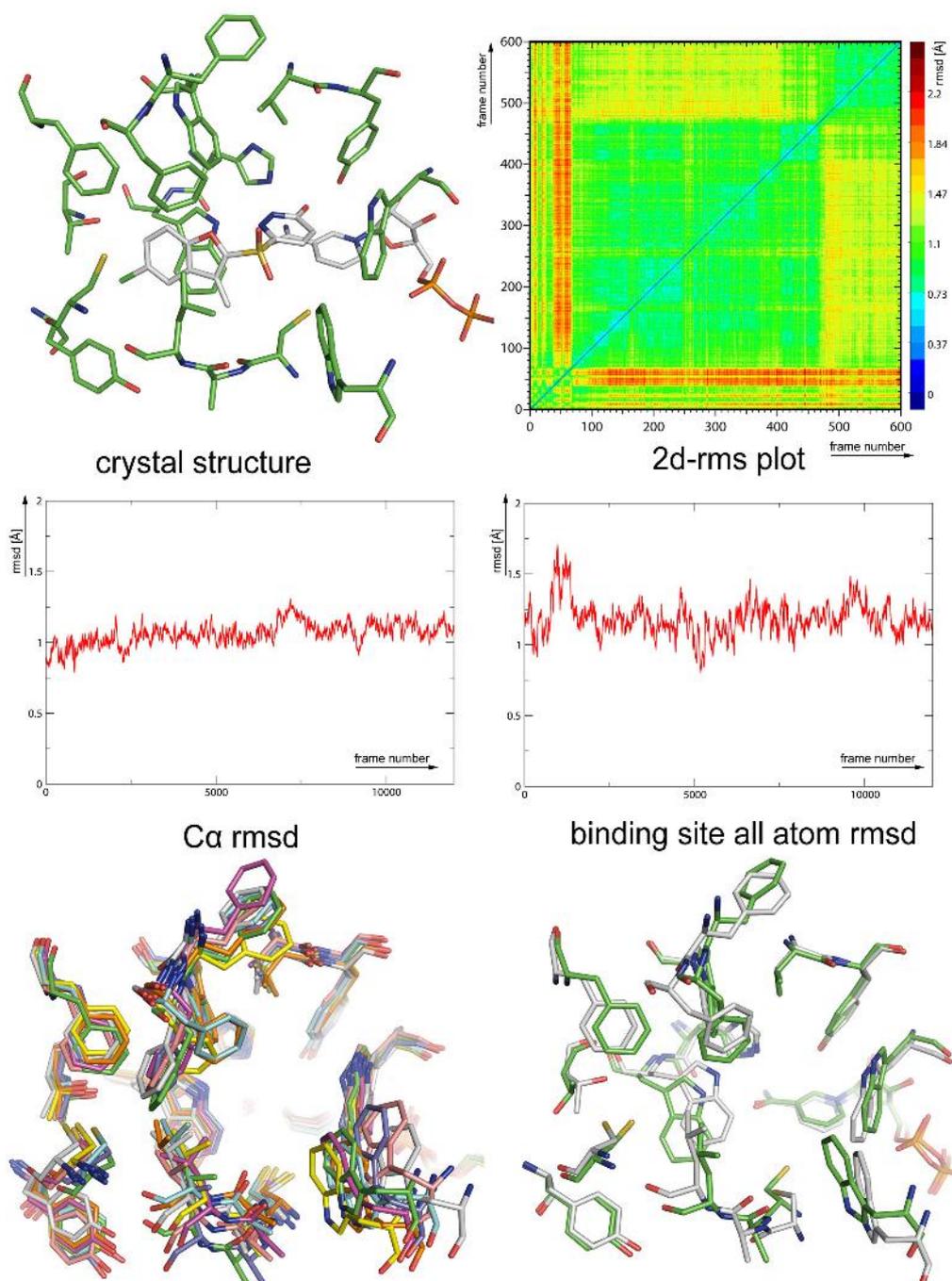


Figure 23: Overview over the MD simulation of the Pfizer compound complex. Data are presented similarly to Figure 19. Left: snapshots from first cluster showing high flexibility: 10 (green), 15 (light blue), 25 (magenta), 35 (yellow), 45 (salmon), 52 (silver), 60 (dark blue), 80 (orange) Right: snapshots from the remaining two main clusters: 300 (green), 550 (silver)

MD of AR in complex with the zopolrestat compound

A structure of the inhibitor zopolrestat in complex with AR has been solved in 1993 by Wilson et al.¹⁷⁹. However, for the pdb entry 1mar only coordinates of the ligand and the protein C α trace are deposited. Since zopolrestat shows an interesting selectivity profile between AR and aldehyde reductase, the structure has been re-determined in-house by Holger Steuber¹⁵³. Interestingly, dependent on the crystallization conditions, the backbone of Ala 299 and Leu 300 adopts different conformations. In one of the conformations a hydrogen bond is formed between the nitrogen of the aromatic moiety penetrating into the specificity pocket. In the other conformation this hydrogen bond is not formed. This behavior is described in more detail in chapter '*Expect the unexpected*' (page 163). The MD simulation reported here was started with the X-ray conformation lacking the hydrogen bond.

Zopolrestat exhibits a carboxylate anchor to bind to the catalytic sub-site and, similar to most potent AR inhibitors, it addresses the specificity pocket. It binds to the enzyme with a free energy of binding of $\Delta G = -11.3$ kcal/mol.

Also in this case the general fold of the enzyme is very well preserved. The average C α rmsd over the whole six ns trajectory is 1.10 (± 0.12) Å. For the binding site residues a mean rmsd value of 1.27 (± 0.11) Å is observed for all atoms.

In the 2d-rms plot of the binding site residues two small clusters in the beginning and the end can be detected which do not share high similarity neither with the remaining part of the simulation nor with each other. The rest of the simulation splits in two clusters. The first of these clusters is interrupted by a few snapshots which show differences compared to the rest of the simulation.

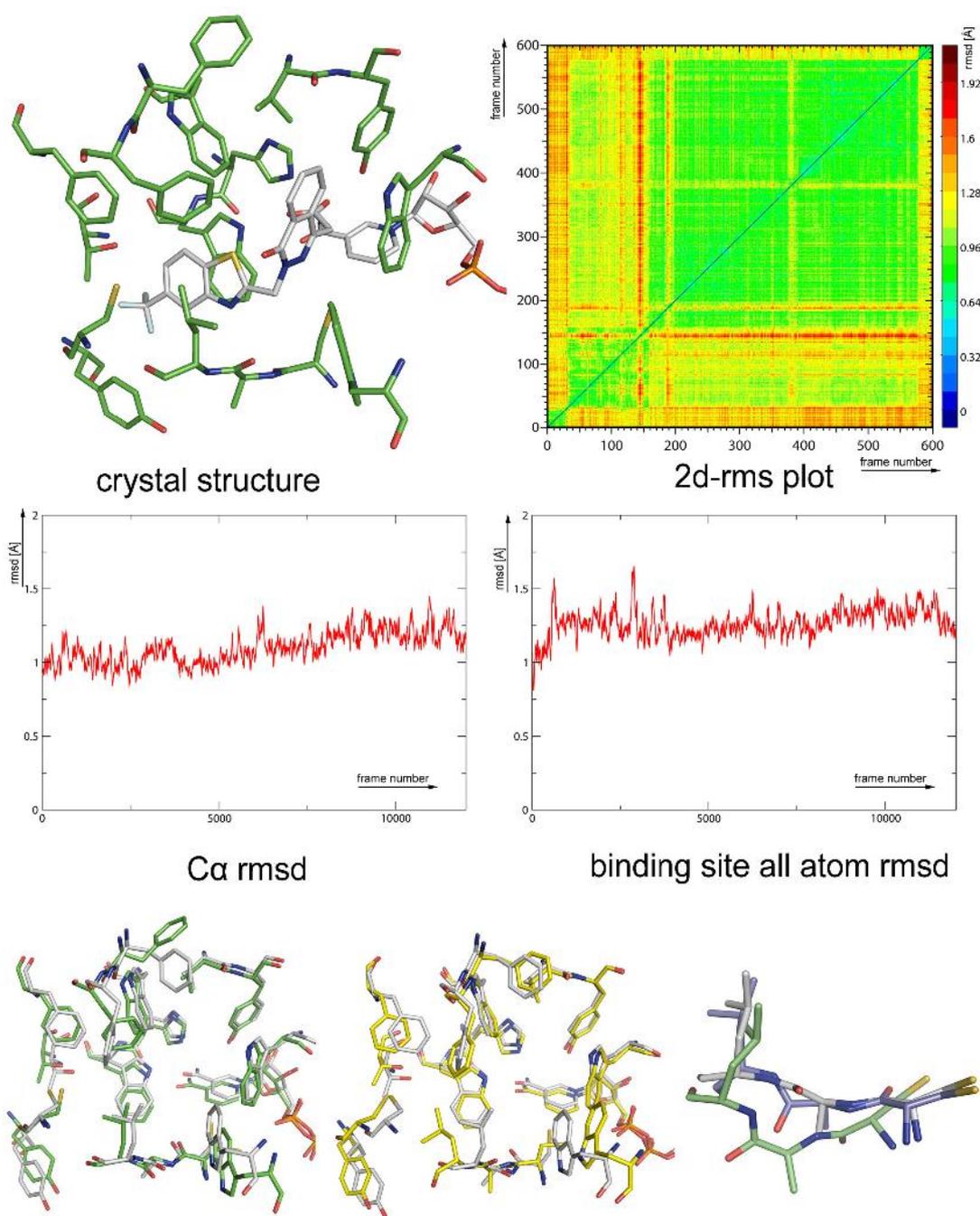


Figure 24: Overview over the MD simulation of the zopolrestat complex. Left: differences between the very first cluster present for the first 30 frames and the following conformation. Depicted snapshots: 20 (green), 100 (silver). Center: differences next to frame 145: 100 (silver), 145 (yellow). Right: mobility in the backbone: 300 (silver), 383 (blue), 590 (light green)

The green conformation in the lower left part of Figure 24 represents the first small cluster in the beginning of the simulation. A flip of the side chains of Phe 122, Cys 298 and 303 is responsible for the transition to the next larger cluster leading to the conformation shown in silver. Concerning the backbone conformation of Phe 122, it is interesting to see that for the majority of the simulation time the side chain remains in this flipped conformation.

The structural differences responsible for the distinct cluster around frame 145 is shown in the lower central part of Figure 24. The system moves from the conformation shown in silver towards the one depicted in yellow. Again both Cys 298 and 303 exhibit a flip of their side chains. In addition, Trp 219 shows larger movements. However, it is especially the side-chain of Leu 300 which shows large flexibility and adopts a unique orientation compared to the rest of the simulation.

Within the second half of the simulation the system remains quite stable. The little cluster around frame 380 and the small deviations at the end of the simulation are caused by movements of the Ala 299 - Leu 300 backbone. Compared to the conformations adopted in the majority of the snapshots of this simulation the backbone rotates by 90°. Within the final cluster this flip is accompanied by larger motions of this backbone region as a whole. Representatives for the different backbone conformations observed are given at the bottom on the right side of Figure 24.

MD of AR in complex with the 47d compound

The 47d compound is a carboxylate type AR inhibitor published in 1997 by Kotani et al.¹⁸⁰ whereas the corresponding structure was solved in-house. Due to its high similarity to zopolrestat, 47d addresses the specificity pocket in similar manner using the same moiety, a benzo-thiazol portion. The mean C α rmsd is 1.14 (\pm 0.15) Å, the rmsd of the binding pocket amino acids amounts to 1.15 (\pm 0.17) Å. Responsible for the increase in the C α rmsd for about one third of the simulation are significant movements in the loop region ranging from Phe 122 to Val 131. As shown in Figure 26, the C α of Glu 126 moves about \sim 9.9 Å. Despite being a flexible loop region in the other simulations, movements of this magnitude are not observed for any of the other simulations.

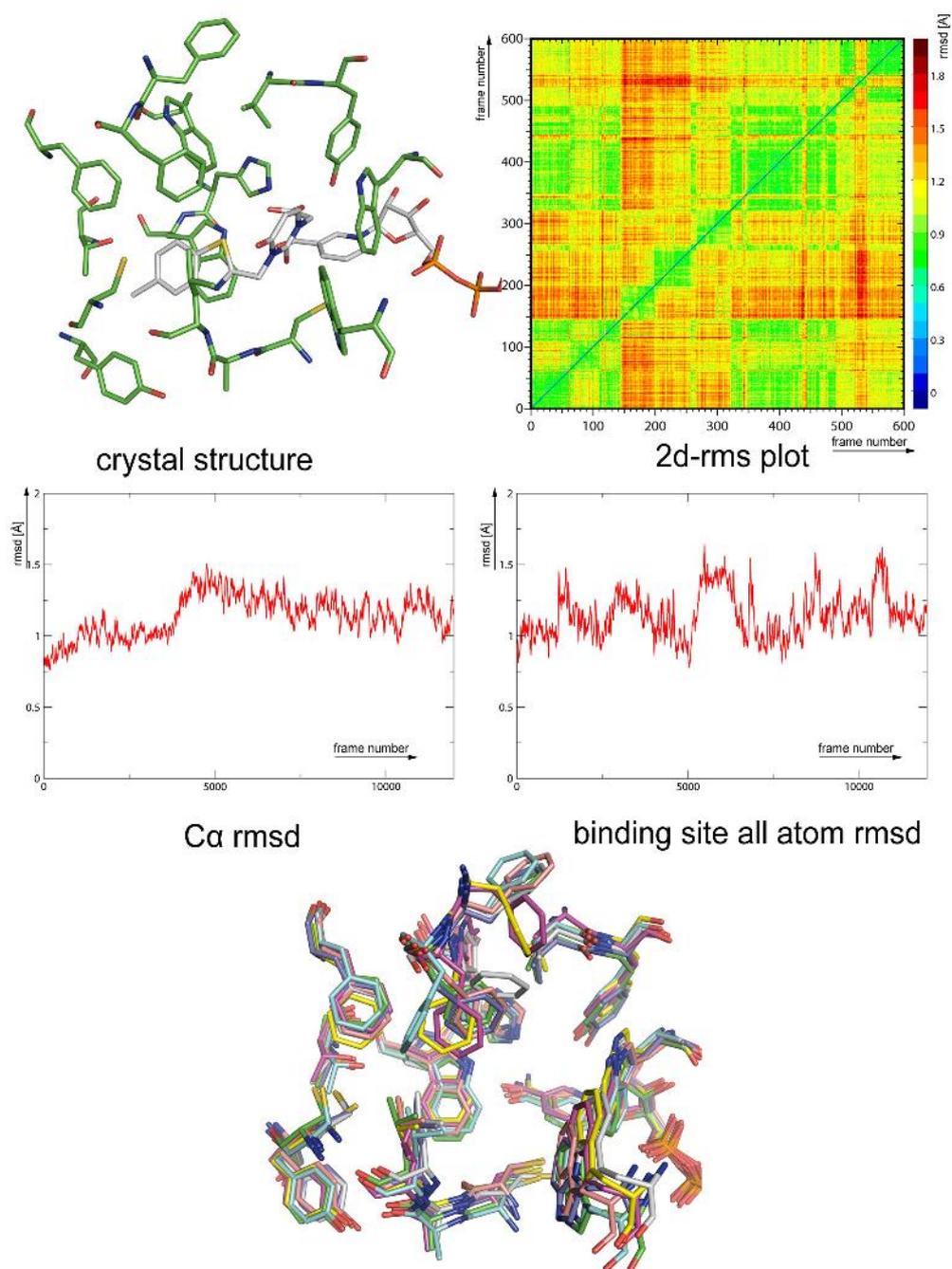


Figure 25: Overview over the MD simulation of the 47d complex. The following snapshots representing the clustering are given in the lower picture: 50 (green), 170 (light blue), 230 (magenta), 290 (yellow), 400 (salmon), 530 (silver), 580 (dark blue)

Analyzing the 2d-rms clustering in Figure 25 many transitions between small clusters are observed. The first 140 snapshots show cross-correlations with snapshots from the second part of the simulation. Between snapshots 140 and 320 three small clusters appear

which do not show high similarity with each other or the rest of the simulation. However, these three clusters are populated while at about the same time the pronounced movements of the 120s loop region mentioned above occurs. Additionally, close to frame 530 some snapshots emerge which depart from the rest of the simulation.

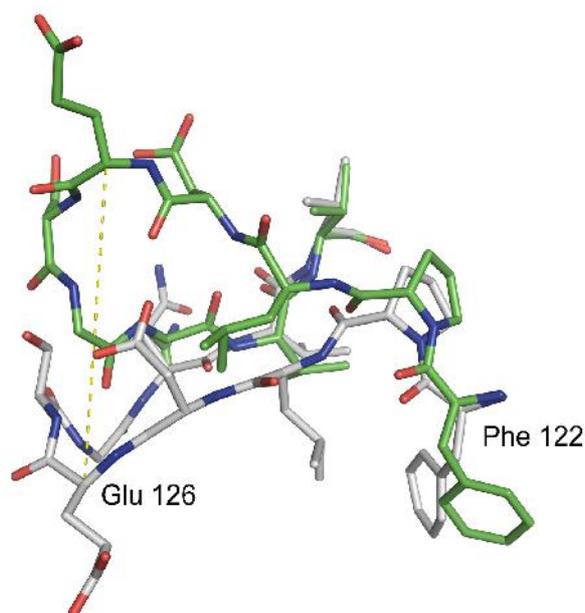


Figure 26: Movement in the flexible loop region following Phe 122 in sequence.

Since a large number of small clusters appears in this simulation a superposition of representative snapshots for many of these clusters is given in the lower part of Figure 25. The majority of the conformational transitions occur with Phe 121 and 122. Both adopt several distinct conformational states. Also Cys 303 shows a side-chain rotation and Trp 219 exhibits the same amount of flexibility as observed in other simulations. The backbone stretch from Ala 299 to Leu 300 remains relatively rigid. No side-chain shift of Cys 298 can be detected. These two observations can be rationalized by the fact that the inhibitor forms a hydrogen bond with the Leu 300 backbone nitrogen and places a carbonyl function into the region where a flipped Cys 298 side chain would be placed. Thus, features of the complexed inhibitor are likely to stabilize this otherwise flexible region.

MD of AR in complex with the JFD compound

The compound JFD was discovered as AR inhibitor by virtual screening as described by Kraemer et. al.¹¹⁵. As expected for a virtual screening hit which has not been further optimized, JFD has only moderate binding affinity ($\Delta G = -7.5$ kcal/mol) towards AR compared to the other inhibitors in this study. The compound addresses the specificity pocket in a similar way as the other inhibitors such as zopolrestat or 47d. However, as shown by the comparison of the available crystal structures, JFD induces less changes within the flexible backbone region next to Ala 299.

During the simulation the average C α rmsd is 1.21 (± 0.10) Å, and the binding-site residues experience the same value of 1.21 (± 0.12) Å. The evolution of these rmsd values with time is plotted in Figure 27. It is worth noticing that both rmsd values tend to fluctuate less as the simulation continues. This is also reflected in the 2d-rms plot of the binding site residues where almost all snapshots from the second part of the simulation coincide in one big and homogeneous cluster. Between frames 440 and 450 conformations are sampled which form 'cross-correlation' with the first section of the simulation.

In due course of the first three nanoseconds of the simulation two clusters can be detected. The image in the lower left part of Figure 27 depicts the changes associated with the transition from the first cluster (shown in green) to the conformation adopted in frame 150 (shown in silver). The backbone of Ala 299 rotates and the conformation of the Phe 122 side chain changes. In addition, the Leu 300 side chain flips almost 180° around its χ_2 angle. This conformation is maintained throughout the rest of the simulation time. The conformation shown in yellow represents frame 250 in the region between the two larger clusters. Here the Ala 299 backbone adopts yet another conformation and the Leu 300 side-chain does amplify this movement. The Phe 122 side chain flips back to its previous orientation.

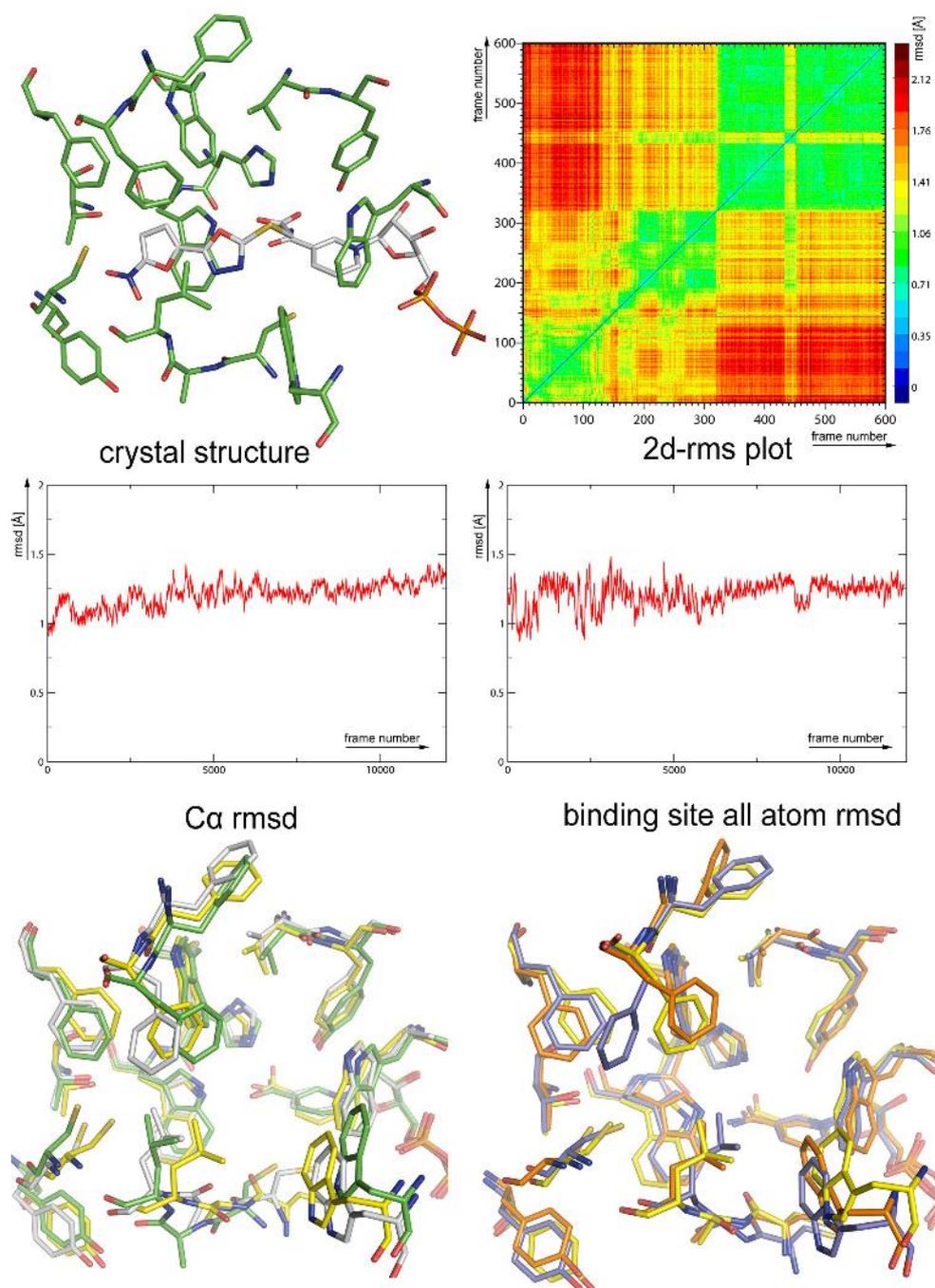


Figure 27: Overview over the MD simulation of the JFD compound complex. Left: snapshots representing changes in the first half of the simulation: 70 (green), 150 (silver), 250 (yellow). Right: changes occurring in the second part of the simulation: 250 (yellow), 400 (light blue), 445 (orange)

The lower right part of Figure 27 shows the transitions in the second part of the simulation. The conformation shown in yellow corresponds to the yellow conformation of the left

picture (frame 250). The transition to the large cluster dominating the second three nanoseconds results in consequence of the recursive side-chain flip of Phe 122. This flipped conformation is then retained for the rest of the simulation. Only for the small time frame close to frame 445 the Phe 122 swings back. Thus, Phe 122 predominantly adopts a conformation which deviates from the one observed in the crystal structure. The rest of the binding pocket is stable. Only Trp 219 and Cys 303 show some noticeable flexibilities.

Simulations of the holo enzyme

All previous simulations were carried out with a ligand present in the binding pocket of AR. Residual flexibility of the binding pocket residues could be observed in all of the cases. However, it appears likely that the conformational space available to the holo enzyme is larger than for any of the complexes, since bound ligands lock the binding pocket residues into certain conformational states.

To assess the conformational flexibility of uncomplexed AR, three different simulations were performed. To increase sampling, the simulations were started from three different crystal structures which represent the three known overall binding pocket conformations.

The geometry adopted in complex structures of sorbinil, tolrestat and idd 594 were chosen as archetypes for the three respective conformers. As indicated by the analysis of the available crystal structures, the sorbinil bound geometry is highly similar to the proposed holo form of AR. The protein conformation adopted in the idd 594 bound structure is representative for all ligands addressing the specificity pocket despite the minor differences observed in the Relibase⁺ analysis. Finally, tolrestat has shown to induce the most pronounced structural changes up to now in a unique fashion compared to all other structurally characterized ligands. In the following, these simulations will be named holo (sorbinil), holo (tolrestat), and holo (idd594), respectively.

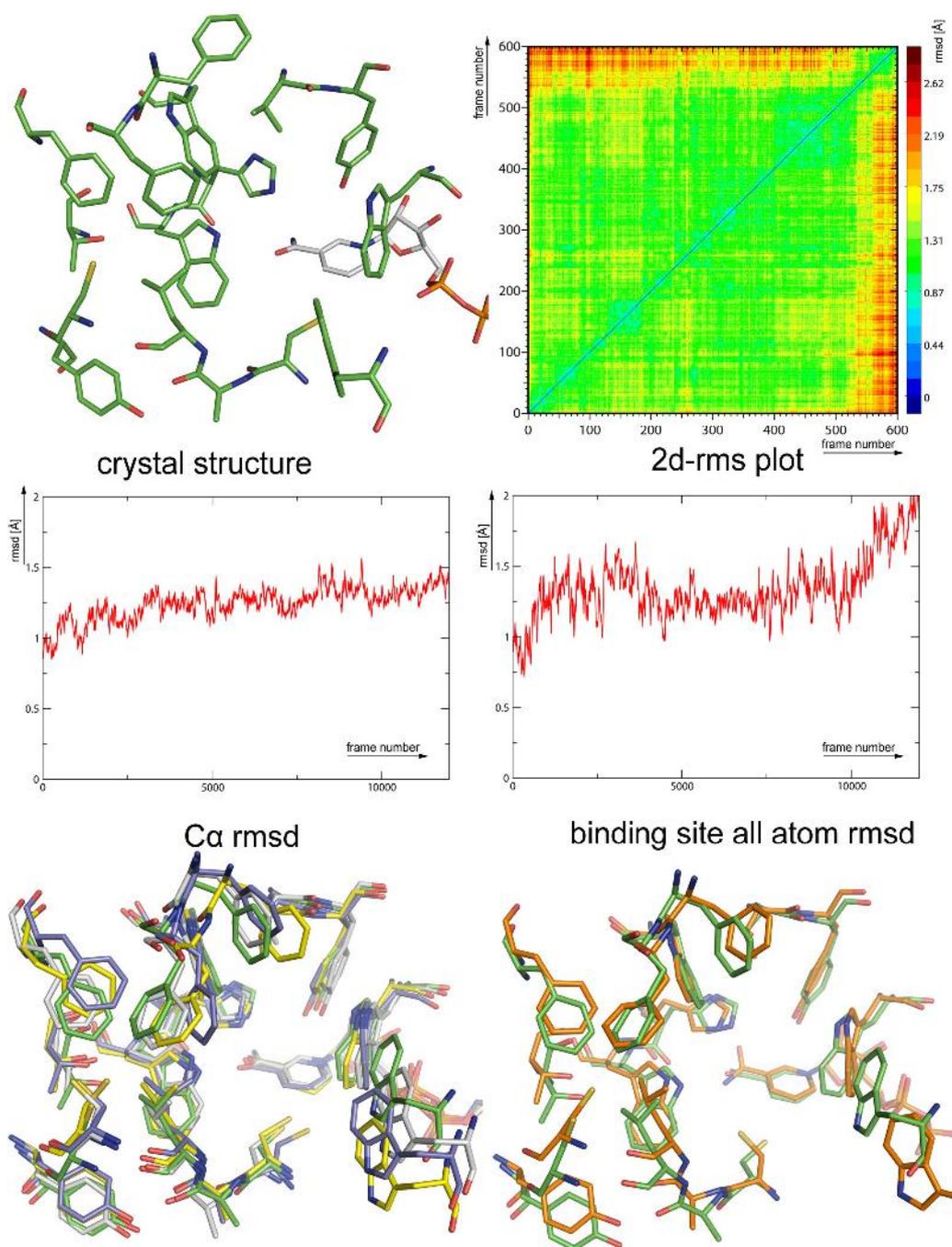


Figure 28: Overview over the MD simulation of the holo (sorbinil). Left: conformations taken from the first large homogeneous cluster: 100 (green), 200 (silver), 300 (yellow), 400 (blue). Right: structural differences between the large cluster and the one experienced at the end: 100 (green), 550 (orange)

Removing the ligand from a crystal structure prior to starting an MD of the produced holo enzyme introduces different levels of perturbations to the systems depending on the binding mode of the removed inhibitor.

In the case of sorbinil the perturbations are only minor. The mean rmsd for all C α atoms of the enzyme is 1.24 (\pm 0.12) Å and the corresponding rmsd for the binding site residues is 1.34 (\pm 0.23) Å. These values are slightly higher than in case of the complex simulations. However, on an absolute scale these deviations are still rather moderate.

Considering the 2d-rms plot in Figure 28, only one large cluster is apparent. This indicates that, in contrast to the complex simulation, no sudden transitions occur between single conformational states. For the last 500 ps of the simulation a small cluster which is different compared to the other snapshots is observed. The movements resulting in this distinct cluster are rearrangements of Trp 219. This residue is also responsible for the increasing rmsd values of the binding site atoms observed towards the end of the simulation. The large rms deviation between this cluster and the rest of the simulation and the resulting coloring of the 2drms plot makes it rather difficult to follow conformational changes occurring within the first 5.5 ns of the simulation.

The lower left part of Figure 28 shows conformations taken from the first large cluster of this simulation. Despite the fact that there is definitely motion in different regions of the binding pocket, especially for Trp 219 and Phe 122, no pronounced conformational changes occur. These changes lead to the clustering of snapshots which has been observed similarly for the complex simulations. The lower right part in Figure 28 superimposes one frame from the first cluster with a snapshot from the second cluster. Here a movement in the side-chain of Cys 298 is observed. More pronounced, however, are the movements of Trp 219 which is attached to the 'cofactor safety belt' loop region.

The next simulation under investigation is holo (tolrestat). In this case the C α rmsd is 1.25 (\pm 0.15) Å and the binding site rmsd is 1.51 (\pm 0.19) Å. The value for the binding pocket rmsd is higher than in the case of holo (sorbinil). However, this rmsd is determined with respect to the starting structure of the MD, which exhibits in case of tolrestat the specificity pocket in an open conformation. As can be seen from the 2d-rms plot of Figure 29 the clustering is not very pronounced as already seen in the previous simulation.

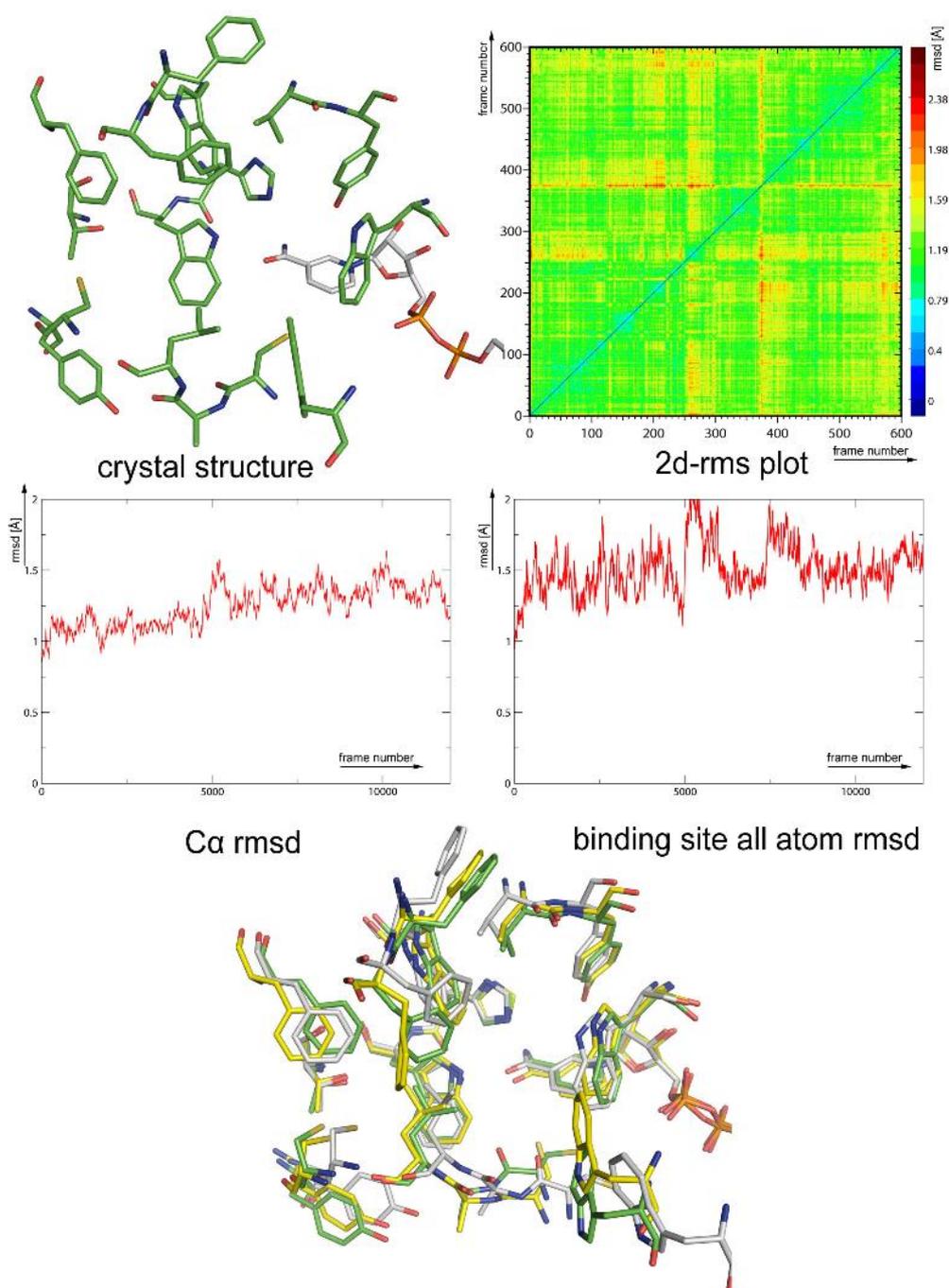


Figure 29: Overview over the MD simulation of the holo (tolrestat). Snapshots depicted in the lower part: 100 (green), 275 (silver), 373 (yellow)

Both the 2d-rms plot and the time course plot of the binding site rms in Figure 29 expose the same snapshots in terms of continuous fluctuations. Between frames 250 and 300 as well as around frame 375 differences compared to the rest of the simulation can be detected. The specificity pocket is in closed conformation right from the beginning of the simula-

tion. Hence, the changes to the closed form by transitions of Leu 300 and its surrounding backbone have already occurred during the minimization and equilibration phase. As shown in Figure 29 the transition from the first cluster (shown in green) to the cluster between frame 250 and 300 (shown in silver) is not directly caused by distinct conformational changes, but more a result of motions in different parts of the binding pocket and, again, especially of Trp 219. The situation is different for the transition to the cluster next to frame 375 (yellow conformation). Here a side-chain rotation of Phe 122 and a flip in the backbone region of Ala 299 are observed. Both being motions which have also been observed in the complex simulations.

The final holo simulation was started using the protein conformation of the idd594 complex as a starting point. The rmsd values for the C α atoms (1.16 (\pm 0.14) Å) and the binding site residues (1.56 (\pm 0.22) Å) are in agreement with the observations made for holo (tolrestat). Surprisingly, the clustering (see Figure 30) within the 2d-rms plot is much more pronounced compared to the other two holo simulations. Eight distinct clusters, which differ in their size, can be distinguished. A network of cross-correlations between the single clusters indicates that the system is fluctuating between different states.

Two binding pocket regions are responsible for the clustering: the Phe 122 side chain and the Ala 299 – Leu 300 loop region. The Phe 122 side chain flips between the two conformational states already seen in the other simulations (lower left part of Figure 30). A feature unobserved in the nine other simulations are the large movements of Leu 300. It fluctuates between conformations which either open or close the specificity pocket (lower right picture in Figure 30). Combinations of the conformations of these two flexible regions are responsible for the clustering and the cross-correlation among the different clusters. Obviously, removing the ligand from the open specificity pocket conformation introduces a large perturbation into the system, thus inducing elevated levels of flexibility into this part of the enzyme. Whether this is an artifact of the setup procedure or of any meaning in terms of dynamic properties of this part of the enzyme is difficult to decide *a priori*. The fact that these kinds of motions are not seen in the other two holo simulations suggests that the sampling is dependent on the selected start geometry. However, given a simulation duration of 6 ns, a fully converged sampling of the phase space of a system of this size cannot be expected.

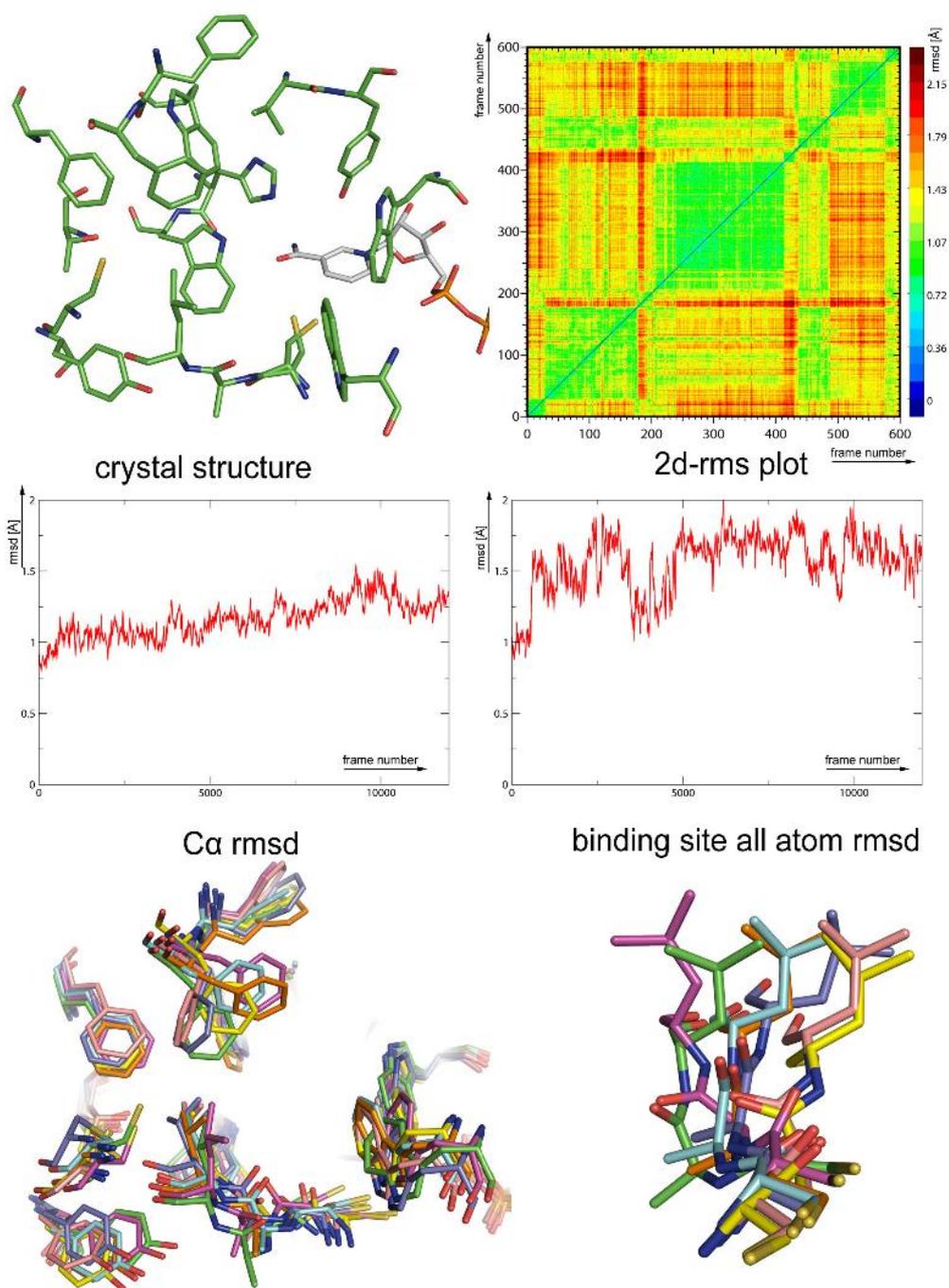


Figure 30: Overview over the MD simulation of holo (idd594). Left: snapshots representing the conformational sampling over the time course of the simulation: 20 (green), 100 (light blue), 185 (magenta), 300 (yellow), 420 (salmon), 550 (dark blue), 590 (orange) Right: snapshots depicting the elevated flexibility of Leu 300 and the neighboring backbone (the same snapshots are used).

Summarizing the results

Analyzing the individual MD simulations in a sequential order is a first step to trace the dynamical properties of the protein. However, considering multiple MD simulations from a comparative point of view it becomes difficult to keep track of all individual changes found in single simulations. Additionally, one of the aims of an MD study should be to correlate observations during MD simulations with evidence for dynamic behavior indicated in the crystal structures (summarized in Figure 31). To facilitate comparison of the results from the crystal structure analysis and the findings from the MD trajectories in terms of flexibility of different parts of the binding pocket, Figure 32 gives an overview over the structural changes observed for the 10 MD simulations discussed above. The information about the multiplicity of different conformational states adopted in different crystal structures is color coded in the 2d representation of the individual amino acids, whereas the summary of the observed flexibility within the MD simulations is represented by the histograms (see legend of Figure 32).

A good overall agreement can be observed between the results from MD and crystal structure analysis. The residues comprising the catalytic pocket (Trp 20, Val 47, Tyr 48, Trp 79, His 110) superimpose very well in the crystal structures (see Figure 31) and show only minor mobility during the MD simulations. This holds true for either the complexes as well as the different holo simulations. Thus, this region of the binding pocket is structurally very well conserved, independent of the type of ligand bound or whether no ligand is accommodated.

Trp 111 is a crucial residue within the binding pocket of AR. It can be classified as part of the catalytic pocket as well as part of the specificity pocket. It superimposes very well for all crystal structures and shows low fluctuations in all individual MD simulations. However, if no ligand is bound, this residues shows an increased level of flexibility. It seems that interactions with bound ligands stabilize the position of this residue. These interactions include a hydrogen bond to the polar head group of the inhibitor as well as an aromatic face-to-face stacking interaction with the part of the ligand addressing the specificity pocket.

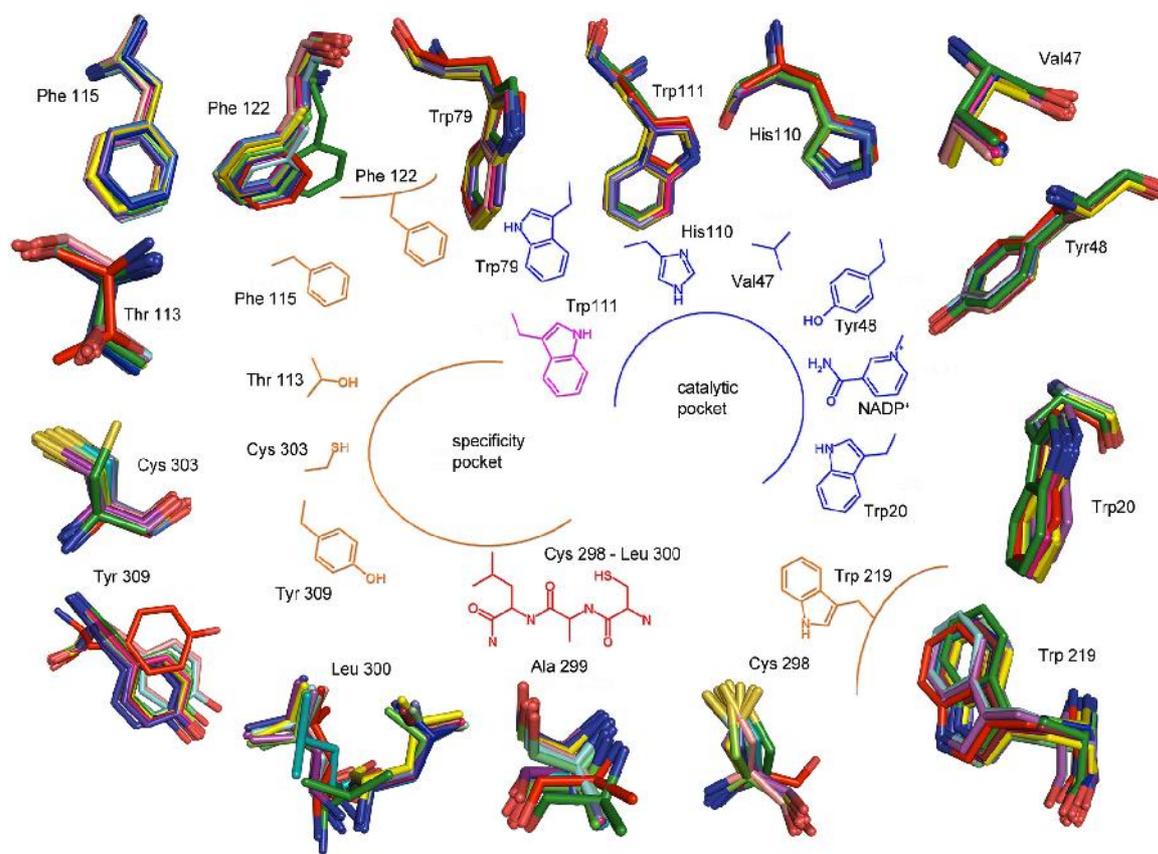


Figure 31: Summary of the crystallographically observed conformations of binding site residues. The shown superpositions are based on the Relibase⁺ superposition of binding site residues used for the crystallographic analysis. For each binding site residue all conformations from the data set used to analyze the conformational variability of each residue are shown.

Among the residues bordering the distant part of the specificity pocket, Phe 115 and Thr 113 show only minor movements. Phe 115 remains in position independent of the actual conformation adopted by the specificity pocket. For Thr 113 the side-chain torsion χ_1 persists in a small angular range over the course of the simulations. Only in case of the Pfizer compound the simulation reveals two distinct states of this torsion angle. This rotation is responsible for the elevated mobility indicated for Thr 113 in Figure 32.

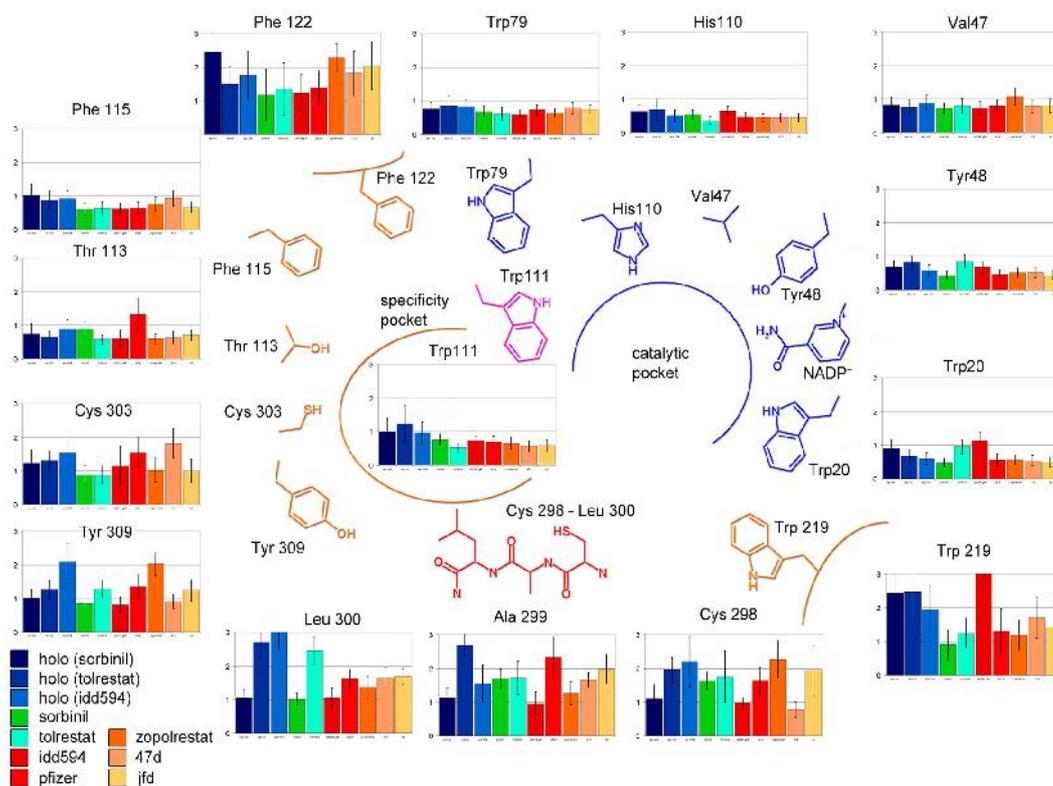


Figure 32: Summary of the flexibility of binding-site residues. Shown are the amino acids which comprise the binding pocket. The individual amino acids are color-coded with respect to their flexibility observed from the crystal structure analysis. Residues with little differences between the single crystal structures are shown in blue, residues with moderate flexibility are colored yellow and, finally, the most flexible parts of the binding pocket are colored in red. The histograms show the average rms deviations referring to the parent crystal structure and the standard deviation of the rmsd fluctuations as 'error bars'. Thus, large columns in the histogram indicate considerable differences compared to the crystal structure, whereas extensive 'error bars' indicate pronounced fluctuations.

Cys 303 and Tyr 309 show elevated levels of motion. Tyr 309 does not superimpose perfectly in the different crystal structures (see Figure 31). During the MD simulations it does not change its conformational state. The side-chain torsions are stable throughout the simulations. The reason for its elevated flexibility is the mobility of its main chain. This is not surprising since Tyr 309 is located at the C-terminal end of the protein chain, with Phe

315 being the final residue. The C-terminus of AR is often inappropriately defined in electron densities.

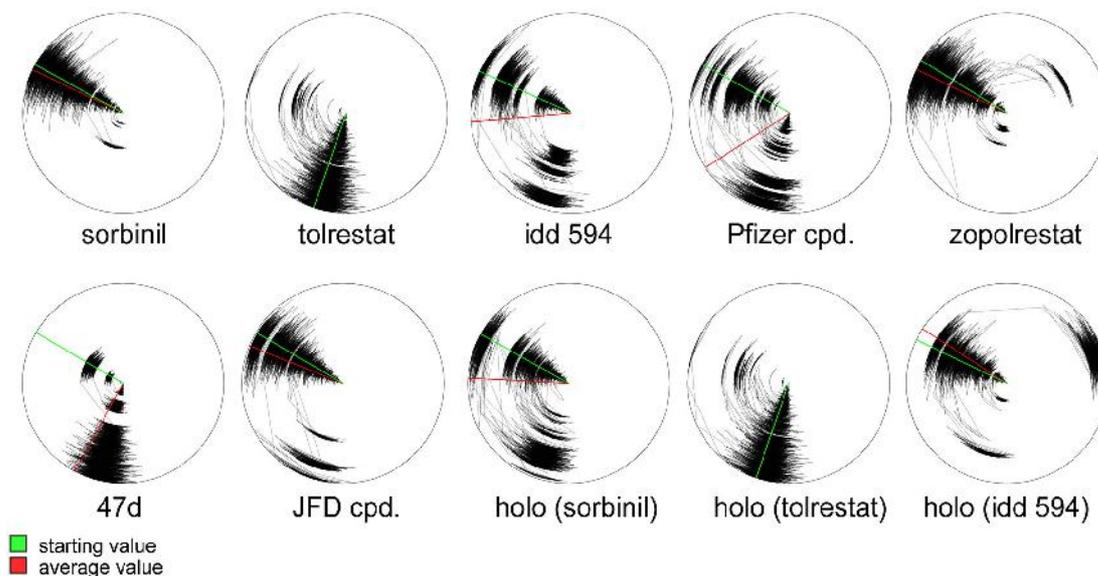


Figure 33: Dialsplots of Cys 303 χ_1 angle. In dialsplots the radius of the circle represents the time axis and the position on the circumference the value for the torsional angle itself. The red line indicates the mean value throughout the whole simulation, whereas the green line gives the value obtained from the corresponding crystal structure.

For Cys 303, however, the situation is different. This residue shows multiple conformers with respect to its χ_1 torsional angle. In every simulation two distinct states for this torsion are observed. They correspond to the two orientations observed in crystal structures. As indicated in Figure 33, the distribution between these two states varies from simulation to simulation. For the Pfizer compound and the 47d simulation the flipped conformation is present in the majority of the snapshots. For the zopolrestat complex simulation three states are sampled for a short period of time. This indicates that different inhibitors allow different amounts of flexibility in the distant part of the specificity pocket. However, no straightforward explanation can be given to rationalize these differences. The torsional changes of the side-chain are the explanation for the increased flexibility of this residue.

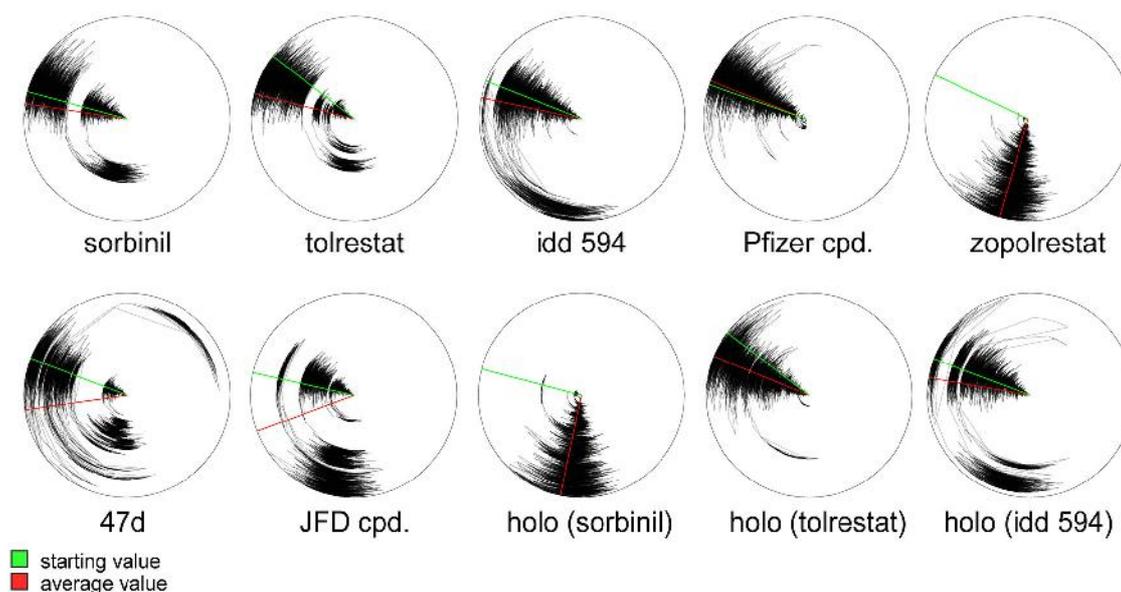


Figure 34: Dialsplots of the χ_1 angle of Phe 122 in all MD simulations. The data presentation is in analogy to Figure 33.

Phe 122 shows a much higher degree of flexibility in the MD simulations than one would anticipate simply considering the conformational scatter in the crystal structures. As can be seen from Figure 31 the side-chain of Phe 122 does move in the superimposed crystal structures to adapt to the bound ligand. However, these adjustments do not lead to major differences in the torsional angles of the side-chains. Concerning the fluctuations in the MD simulations the situation is different. The large rmsd values and standard deviations apparent in Figure 32 indicate already that the movements of this residue are substantially larger than the scatter indicated by the different crystal structures.

Figure 34 presents the χ_1 angle of Phe 122 of all MD simulations as dialsplots. For each of the simulations, two rotameric states of this dihedral angle are sampled. One corresponds to the value observed in the crystal structures (green line), the other corresponds to the flipped conformation already discussed for the individual MD trajectories. In the case of 47d and the holo (idd 594) simulations three distinct rotameric states are sampled. However, the population of the single states along the trajectory differs. For the simulations of the Pfizer compound and holo (tolrestat) the flipped conformation is only rarely sampled, whereas for zopolrestat and holo (sorbinil) the flipped conformation is retained for almost the whole simulation. In the other simulations transitions between both states occur with

varying frequency. In summary, this residue shows surprisingly high flexibility whether a ligand is bound or not.

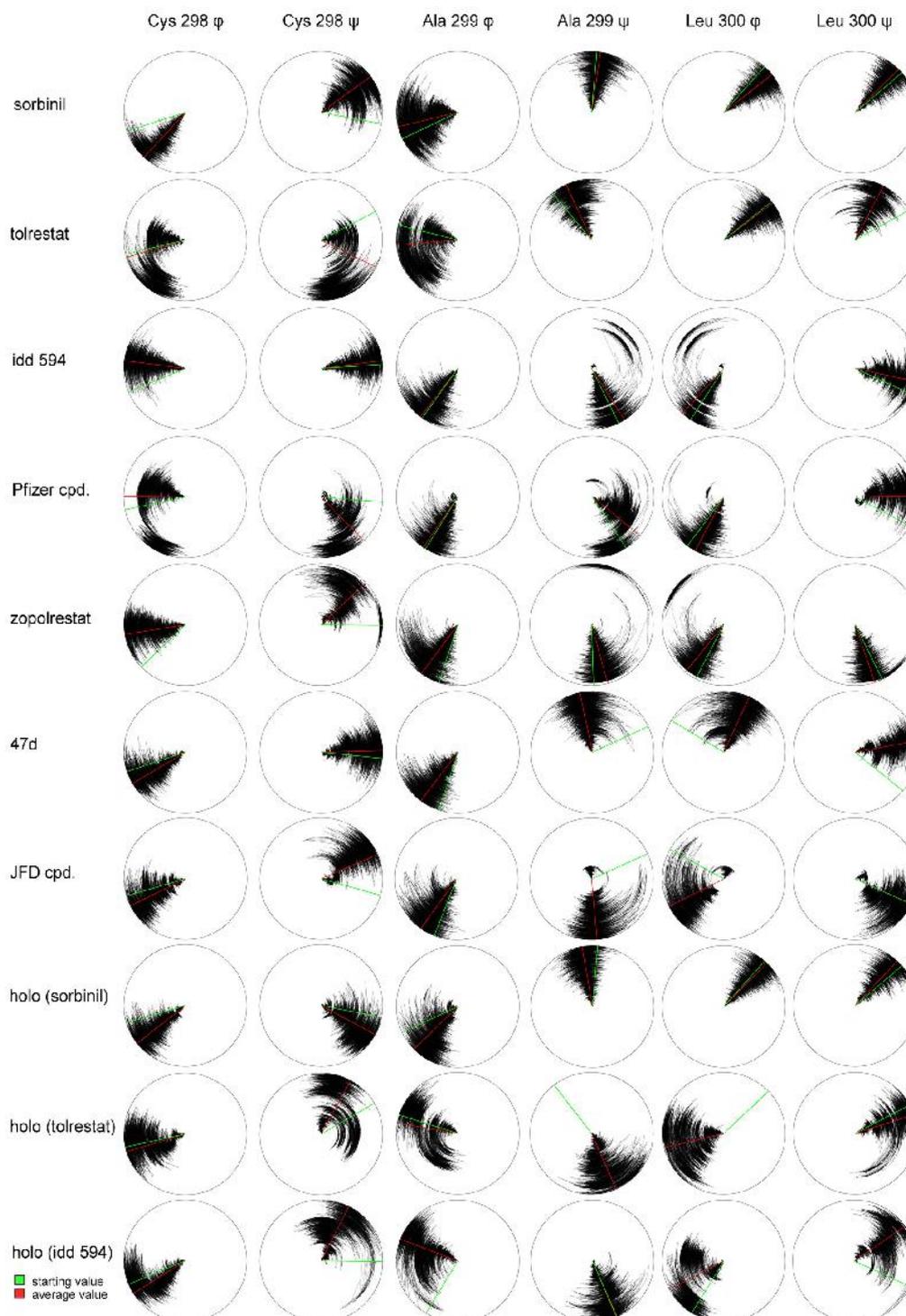


Figure 35: Dialsplots of the backbone torsions of the loop from Cys 298 to Leu 300. The data presentation is in analogy to Figure 33.

Another residue with deviating properties between MD simulations and observed data scatter in crystal structures is Trp 219. As apparent in Figure 31, the Trp 219 side and main chain superimpose well for all holo and complex structures. Neither large movements in torsional space nor significant overall movements of the whole side-chain can be monitored for this residue. However, as shown by Bohren et. al.³⁴, the C α atom of Trp 219 moves by as much as 17 Å between apo and holo forms of the enzyme. This mobility is essential for the function of the enzyme, since the 'cofactor safety belt' loop extending from residues 214 to 226 has to shift in order to enable the exchange of the cofactor after completion of the reaction cycle. Kinetic studies suggest that the exchange of NADP⁺ is the rate limiting step for one turn over of the enzyme³⁸⁻⁴⁰. Furthermore, considering the high affinity of the cofactor towards the enzyme ($K_d\text{NADP(H)} < 10\text{nM}^{181}$), it can be expected that movements of this magnitude occur on a time scale which is far beyond the scope of this study. However, the increased mobility of Trp 219 is an indication that the intrinsic flexibility of this part of the enzyme is to some extent covered by the simulations.

As shown by the analysis of the available crystal structures, the part of the enzyme showing the highest degree of flexibility is the loop region from Cys 298 to Leu 300. This small region determines the shape of the binding pocket and provokes the 'induced-fit' adaptations. Especially the backbone part of Ala 299 is essential for these adaptations. The high mobility of this region in the simulations has already been shown qualitatively for the individual simulations (see above). In order to perform a more quantitative analysis, Figure 35 plots the backbone torsion angles of the corresponding residues. These plots evidence pronounced flexibility for all considered torsion angles. Different states can be seen for the separate simulations, but also the overall picture of all simulations suggests similar findings. Mostly they are adopted in consequence of the different binding modes of the inhibitors. Generally speaking, the two outermost dihedrals (ϕ Cys 298, ψ Leu 300) tend to be less scattered than the inner ones. Considering the inner dihedrals it becomes obvious that the ψ angles change in correlated fashion with the corresponding ϕ angles of the following residue. In many cases the proceeding ϕ angle shows anti-correlated movements to the ψ angles or vice versa canceling out the total rotation of the chain. Thus, the overall side-chain orientations of the respective residues are retained, while the backbone itself is allowed to flip. Furthermore, it becomes evident that different inhibitors allow a different

degree of flexibility within this region. In the case of sorbinil, which binds to the closed form of the specificity pocket, the backbone shows the least amount of flexibility. This is also reflected in the holo (sorbinil) simulation. Thus, one can conclude that flexibility of this loop is reduced once the specificity pocket is closed. This finding is underlined by the available crystal structures. Figure 36 shows a superposition of all available crystal structures where the specificity pocket is in a closed conformation. For each of these examples the backbone adopts the same conformation. Obviously, one preferred conformation exists for the closed form of the specificity pocket, however, if the specificity pocket is opened conformational variability in this region is intensified.

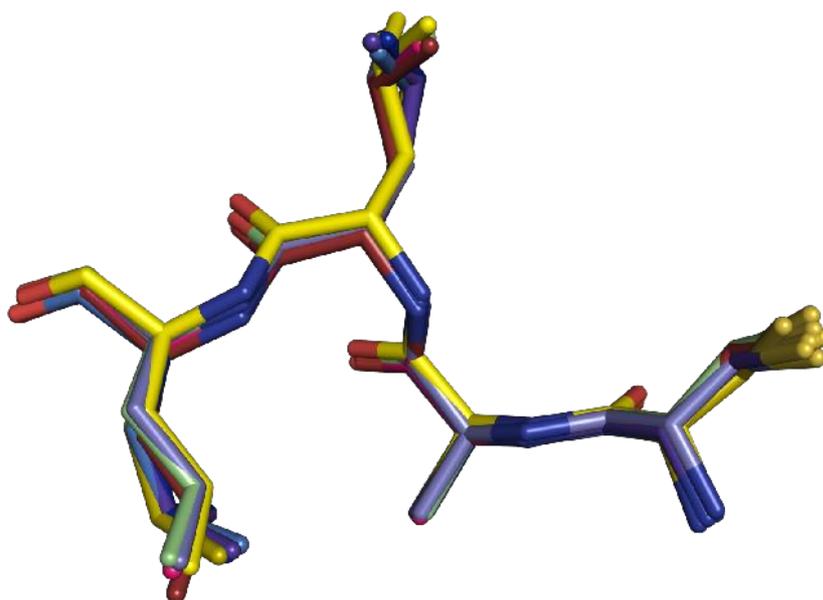


Figure 36: Superposition of all crystal structures which either have no ligand bound or where the ligand binds to the closed specificity pocket conformation. All structures superimpose very well and especially in the backbone there is almost no conformational variability to be registered.

The simulation of the tolrestat complex exhibits an unique behavior of this backbone stretch. The usually quite flexible torsions ψ Ala 299 and ϕ Leu 300 are exceptionally sta-

ble over the course of this simulation, whereas the two backbone torsions of Cys 298 and ϕ Ala 299 are the most flexible ones compared to all other complex simulations. It seems that the unique side-chain conformation of Leu 300 imposes different restraints on the backbone which do not allow the same degree of flexibility as observed with other inhibitors.

Analyzing the behavior of this backbone portion for the different holo simulations, it can be noticed that there are distinct differences depending on the starting geometry. As already mentioned, the conformations sampled for holo (sorbiniil) confirm the ones observed in the sorbiniil complex simulation and they are in agreement with the crystal structures.

However, for the cases of holo (tolrestat) and holo (idd 594) this part of the backbone is in a different conformation throughout the trajectories. Additionally, the observed flexibility is higher. Thus, it seems that on the time scale of several nanoseconds the sampling is different depending on the starting geometry, even if the movements involved are side-chain movements and small shifts of the backbone.

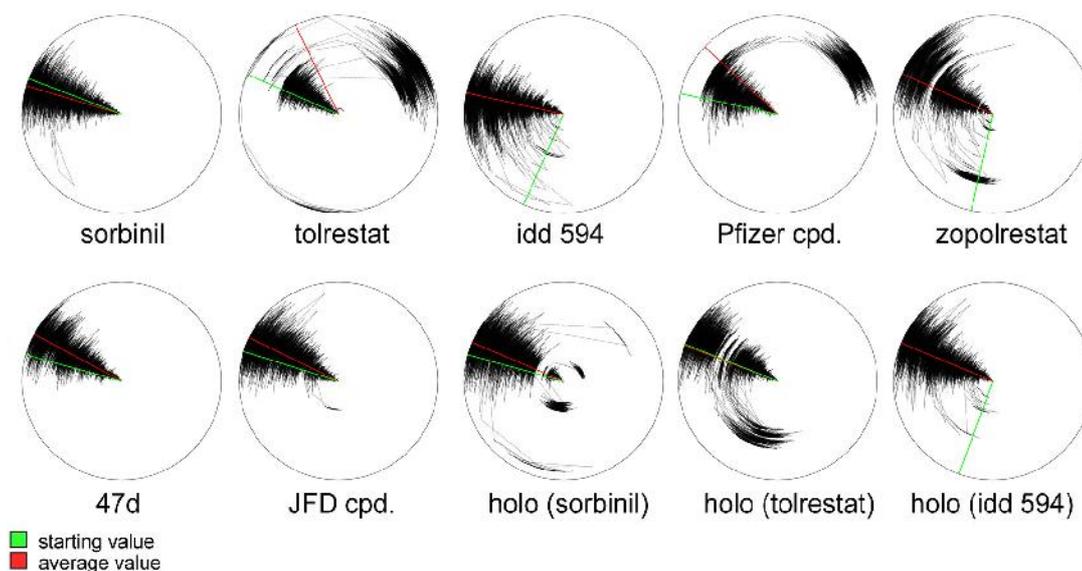


Figure 37: Dialsplots for the Cys 298 Chi 1 side-chain angle. The data presentation is in analogy to Figure 33.

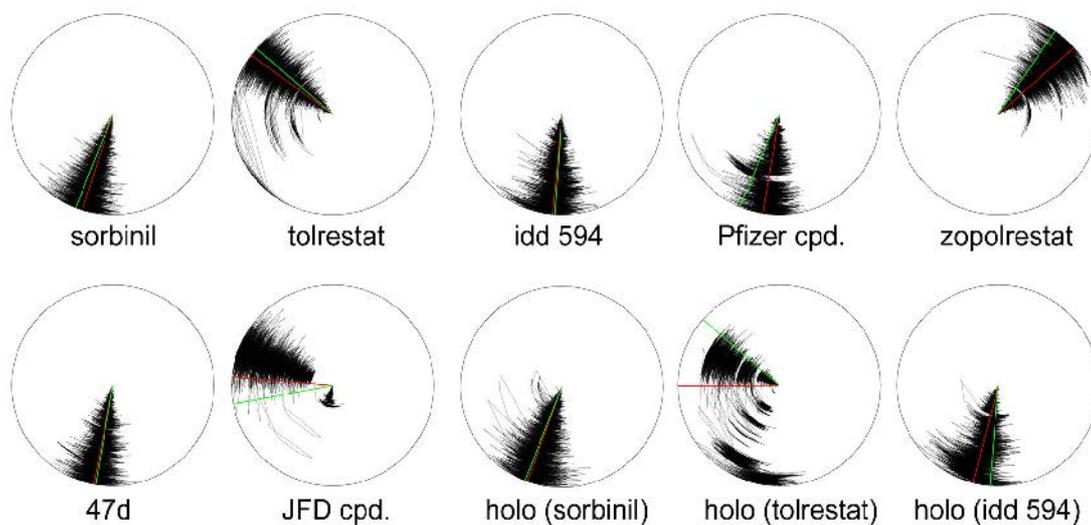
To further examine the flexibility of this region, it is instructive to also include the side chains into the analysis. For Cys 298 (Figure 37) it can be observed that the side chain shows variability in most simulations. Two different conformers are sampled in most of the cases (one in case of 47d, three for holo (sorbiniil)). This finding is in agreement with ob-

servations from crystal structures. For example, in the ultra-high resolution structure of PDB idd 594¹¹⁴ Cys 298 shows a double conformation. The decreased flexibility of this torsion angle in case of 47d can be rationalized with the fact that the ligand places a carbonyl group next to the spatial position of Cys 298, thus prohibiting a flip of this side chain.

The last interesting residue within this region is Leu 300. The two side-chain torsions of this residue are depicted in Figure 38. For most complex and holo simulations the χ_1 angle of Leu 300 is stable throughout the trajectories. Again tolrestat is the exception. Within both the complex as well as the holo simulation of tolrestat the χ_1 angle shows a considerable amount of flexibility. Again this behavior can be rationalized using findings from the crystal structure. Figure 39 A shows the inhibitor and the backbone region under consideration together with the experimental electron density from synchrotron data (1.08 Å resolution)¹²⁶. The 2FoFc density, contoured at 2σ , is shown in blue. At this contour level, all atoms of the inhibitor are well defined. This also holds true for most of the protein atoms in the region of the flexible C-terminal loop. However, the electron density for the side-chain atoms of Leu 300 is only poorly visible. This might indicate higher residual mobility or distribution over multiple states of this side chain.

For the χ_2 angle of Leu 300 different states are sampled. However, in this case the differences between individual states are not crucial for ligand binding. A rotation about the χ_2 angle corresponds to a rotation of the terminal branched hydrophobic moiety of a leucine combined with slight adjustments of the position of the side chain. Nevertheless, overall conserved properties are exposed towards a bound ligand. Also in this case, the MD findings are in agreement with crystallographic evidence. Figure 39 B shows a superposition of the structures of human (9tol, green) and porcine (1ah3, silver) AR in complex with tolrestat. One can clearly see the difference in the Leu 300 χ_2 angle. However, for a medium resolution structure such as 1ah3 (2.30 Å) it remains questionable whether the position of this side-chain can be determined undoubtedly. In addition, this rotation about the Leu 300 χ_2 torsion can also be registered comparing structures with different inhibitors bound. Thus, the flexibility observed in the MD simulations for different inhibitors is not unreasonable.

Leu 300 X1



Leu 300 X2

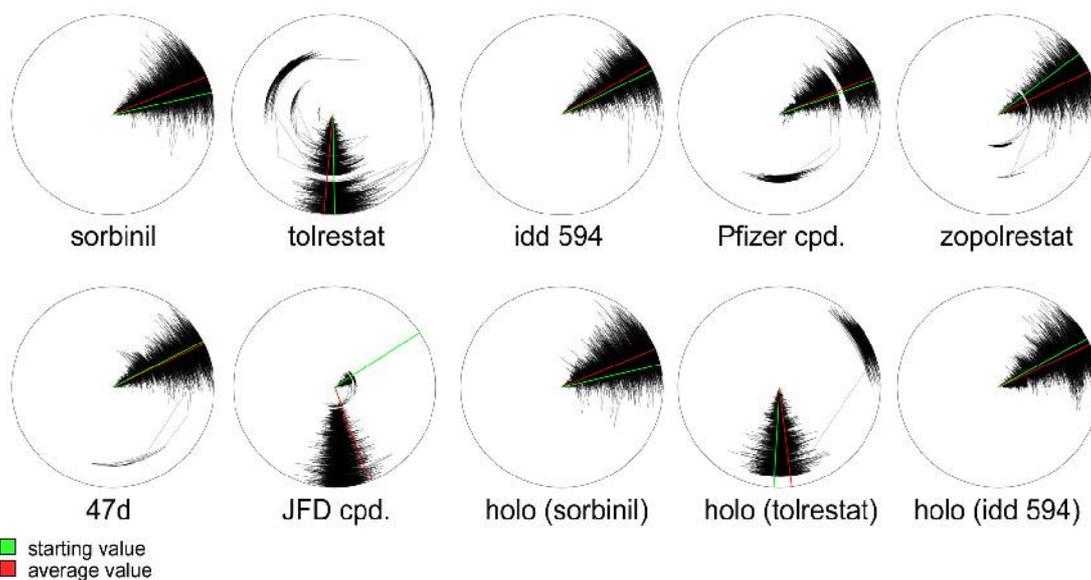


Figure 38: Dialsplots of the Chi1 and Chi2 torsional angles of Leu 300. The data presentation is in analogy to Figure 33.

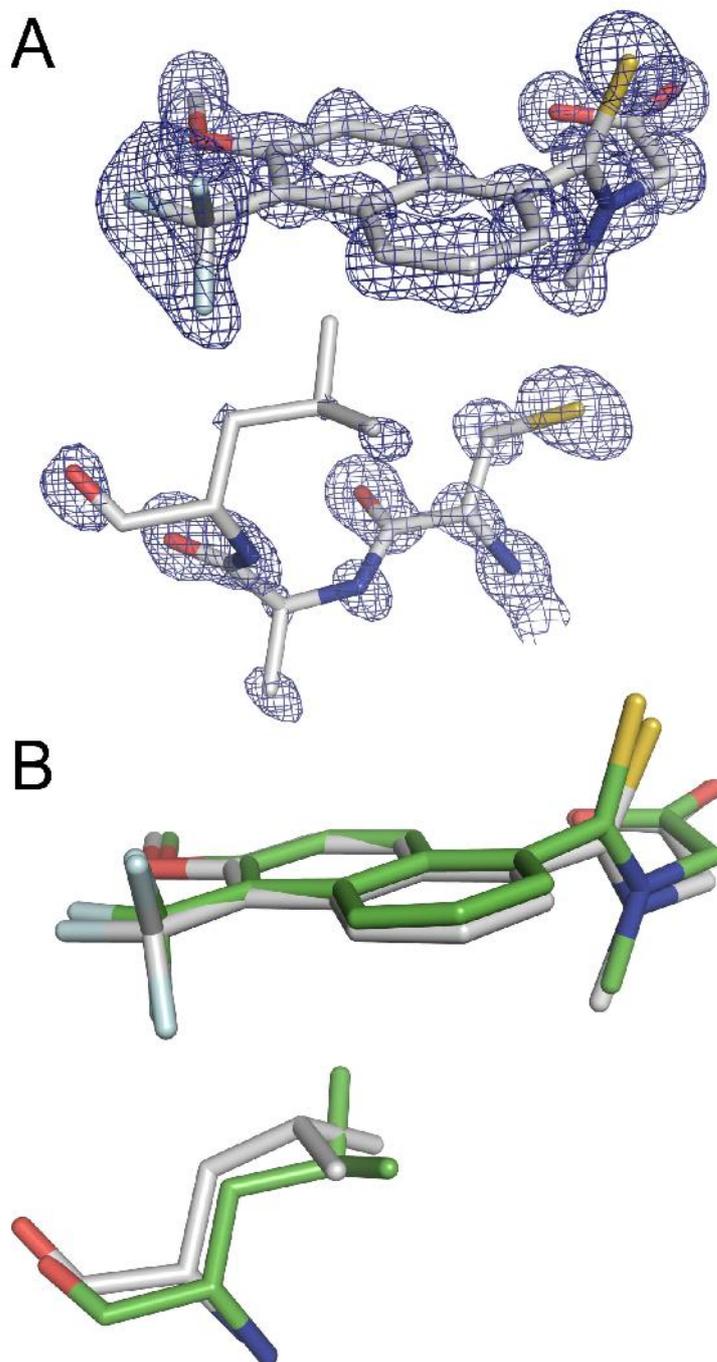


Figure 39: A shows the 2Fo-Fc electron density contoured at 2 sigma for the inhibitor tolrestat and the flexible C-terminal loop region. The side-chain atoms of Leu 300 are only poorly defined in the electron density, thus implying residual mobility of this side chain. B depicts the superposition of human (green) and porcine (silver) AR complex structure of tolrestat where a different side chain conformation caused by a rotation around the χ_2 has been assigned to the electron density.

Summarizing the previous analyses of the MD trajectories in general four regions in and next to the binding pocket are apparent showing considerable amount of mobility: Trp 219 and the cofactor safety belt (i), Phe 122 and adjacent residues (ii), the distant part of the specificity pocket comprising Cys 303 and Tyr 309 (iii), and the loop region responsible for the 'induced-fit' adaptations ranging from Cys 298 to Leu 300 (iv). In the latter case the observed flexibility is in agreement with conclusions drawn from the comparative analysis of the available crystal structures. Knowing that conformational variability observed in MD simulations and crystal structures cannot be compared on a quantitative scale, the flexibility of the first three portions is more pronounced in the MD simulations than expected from the superposition of the X-ray structures. Thus, the question arises whether the movements seen in the simulations corroborate experimental findings or whether they strongly exaggerate the situation due to artifacts in the applied force fields or insufficient relaxation of the geometries.

The facts that protein structures derived from diffraction experiments represent both spatial and temporal averages which are measured at rather low temperature has already been addressed above. Another factor which can take substantial influence on the conformations of the protein residues are contacts formed between individual molecules in the crystal packing. Due to the dense, symmetric and repetitive packing of molecules in crystals, contacts between symmetry-related protein molecules are formed. These contacts most likely do not occur if the protein is in solution under physiological conditions. Thus, it is instructive to study in more detail the properties for the different structures which serve as starting points for the MD simulations.

A crystal packing analysis is presented in Figure 40. It was carried out using the AreaIMol¹⁸² module of ccp4i¹⁸³. All parameter settings were kept at the implemented standard values. The probe radius for the water molecule was set to 1.4 Å. The seven protein-ligand complexes (sorbiniol, tolrestat, idd594, Pfizer compound, 47d, zopolrestat, JFD compound) which were used as input structures for the MD simulations were examined. The program computes the solvent accessible surface area (SASA) for each residue within the protein. Then the packing environment is generated and the SASA is recalculated for each residue. Atoms for which the SASA is smaller within the packing environment than for the isolated protein are indicated by blue spheres in Figure 40. Using this visualization tool, it

is possible to qualitatively assess which atoms are involved in crystal contacts. However, no quantitative conclusions about the type or strength of the involved interactions can be extracted from these diagrams.

For six of the seven complexes at least one of the ligand atoms is involved in contact formation to symmetry-related protein molecules. Only the JFD compound does not change its SASA due to crystal packing. In case of the Pfizer compound only one atom is affected. Two atoms are involved in 47d, whereas sorbinil, idd594, and zopolrestat show differences for three atoms. Tolrestat reduces its SASA at four ligand atoms in the packing. These findings give a first idea that packing effects occur next to the binding pocket which might influence the conformations and/or flexibilities of some of the binding-site residues.

Three regions next to the binding pocket should be considered in detail. The first is Trp 219 and the adjacent loop region, corresponding to the 'cofactor safety belt'. Only a few atoms of this loop are involved in crystal contacts. In fact, the temperature factors are relatively high for the entire loop. Trp 219 is the only residue which is located directly at the binding pocket. Interestingly, in six of the seven complexes at least one atom of Trp 219 is forming crystal contacts.

The second area of interest is adjacent to Phe 122. Glu 120, Phe 121, and Phe 122 form crystal contacts in all structures of the dataset. For Glu 120 and Phe 121, almost all atoms are affected, but also Phe 122 shows differences for main-chain and side-chain atoms in all cases. Many amino acids of the highly solvent-exposed loop preceding Phe 122 in sequence are strongly involved in crystal packing effects. However, as for the cofactor safety belt, the B-factors for the residues in this region are in general noticeably higher than for the rest of the protein. Therefore, the positions of the side chains are less accurately defined by the electron density obtained from the X-ray experiment.

Spatially adjacent to this loop is the third interesting part: the small loop to which Tyr 309 is attached. The side chain of Tyr 309 points towards the flexible loop next to Leu 300 which mainly determines the binding pocket conformation. Tyr 309 itself is not directly involved in crystal packing. However, the residues preceding Tyr 309 in sequence are affected in all cases.

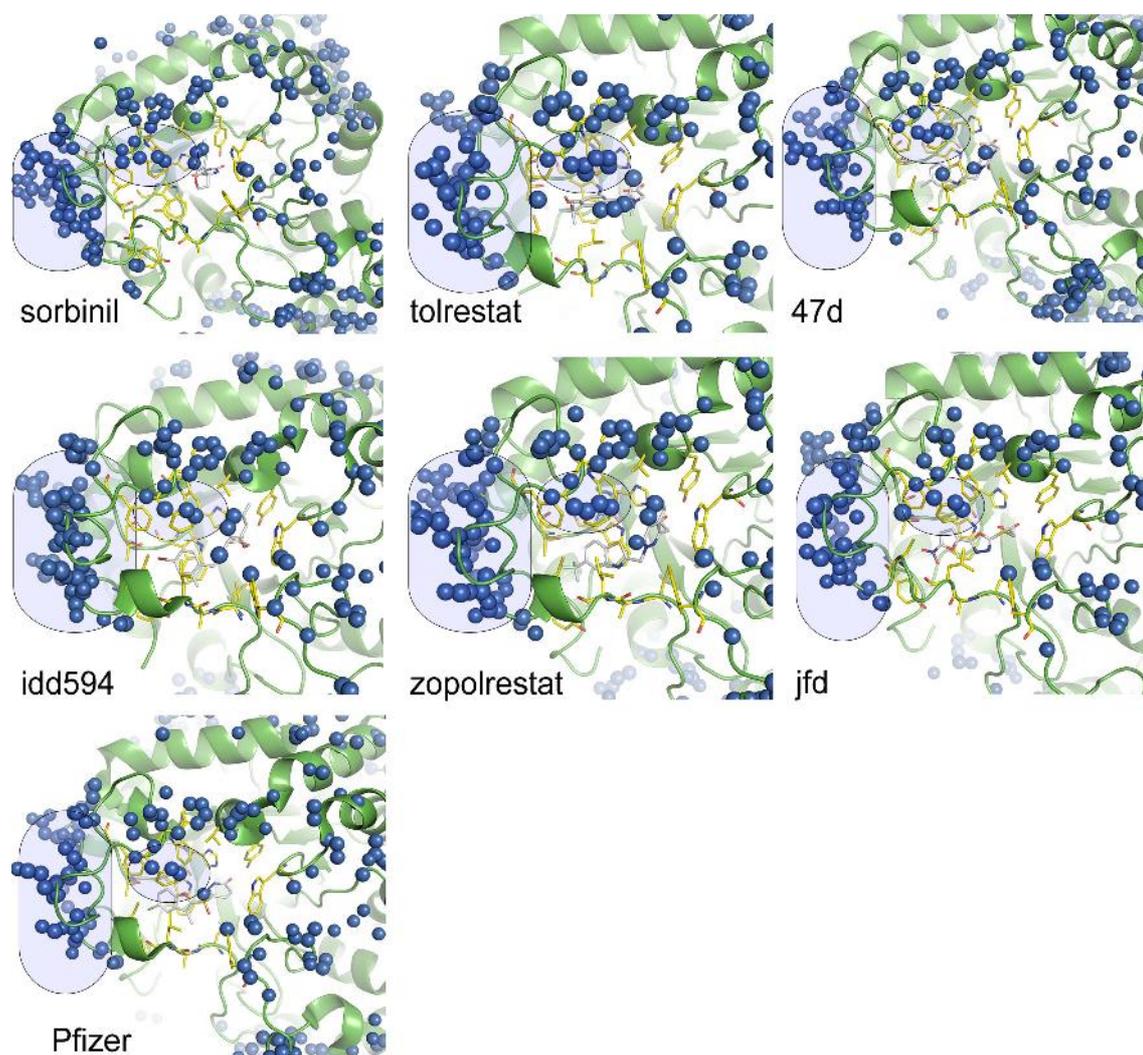


Figure 40: Visualization of crystal contacts formed by the complexes used as MD starting structures. Proteins are shown in cartoon view, the binding-side residues are shown as yellow sticks and the corresponding ligands are shown as silver sticks. Atoms reducing their solvent accessible surface area (SASA) in the crystal packing compared to with the isolated state are shown as blue spheres. Regions of special interest (see text) are enclosed by oval shapes.

The loop region neighboring Leu 300 is not involved in crystal packing. The only exception is the 47d complex where the CD2 atom of Leu 300 reduces its SASA due to crystal packing.

In conclusion, crystal packing clearly affects residues directly adjacent to the binding pocket. While it is difficult to assess their influence on flexibility and thus on the accessi-

ble conformational space of each individual amino acids, it cannot be ignored that some regions might be artificially constrained to evolve their inherent flexibility or they are possibly forced to adopt conformations which are not the preferred ones in solution.

Thus, there are good arguments to believe that the flexibilities noticed in the MD simulations may be larger and different from those observed in the crystal structures. Furthermore, they possibly reflect the fact that during MD simulations the protein is examined in a different environment than in the X-ray experiment.

Implications for the binding pocket

Changes in torsion angles take predominantly local influences on the direct neighborhood of the involved residues. However, depending on the location and function of the affected residues, these changes can have dramatic effects on the total binding pocket. This is illustrated by the mobility of the Ala 299/Leu 300 backbone. Via local movements of the main chain, a new sub-pocket is created which alters the overall shape of the binding pocket quite dramatically. Considering the entire binding pocket, the question arises, how the flexibilities observed in the MD simulations can influence the overall shape and properties of the binding pocket of AR. Are pocket conformations generated which have not yet been seen in crystal structures? If so, can these pocket conformations be exploited for structure-based drug design projects, i.e. can they be addressed (and stabilized) by new inhibitors?

To answer the first question, it is necessary to analyze the effects that the four regions of elevated mobility (Trp 219, Phe 122, Tyr 309/Cys 303, and Ala 299/Leu 300) have on the overall morphology of the binding pocket.

Trp 219 borders the entry of the binding pocket and is a solvent-exposed residue. Fluctuations recorded during the MDs drive this residue away from the ligand positions as seen in the crystal structures. However, no additional sub-pocket is generated which possibly could be addressed by potential ligands. Therefore, this part of the binding pocket is probably not a good candidate to explore novel binding pocket conformations.

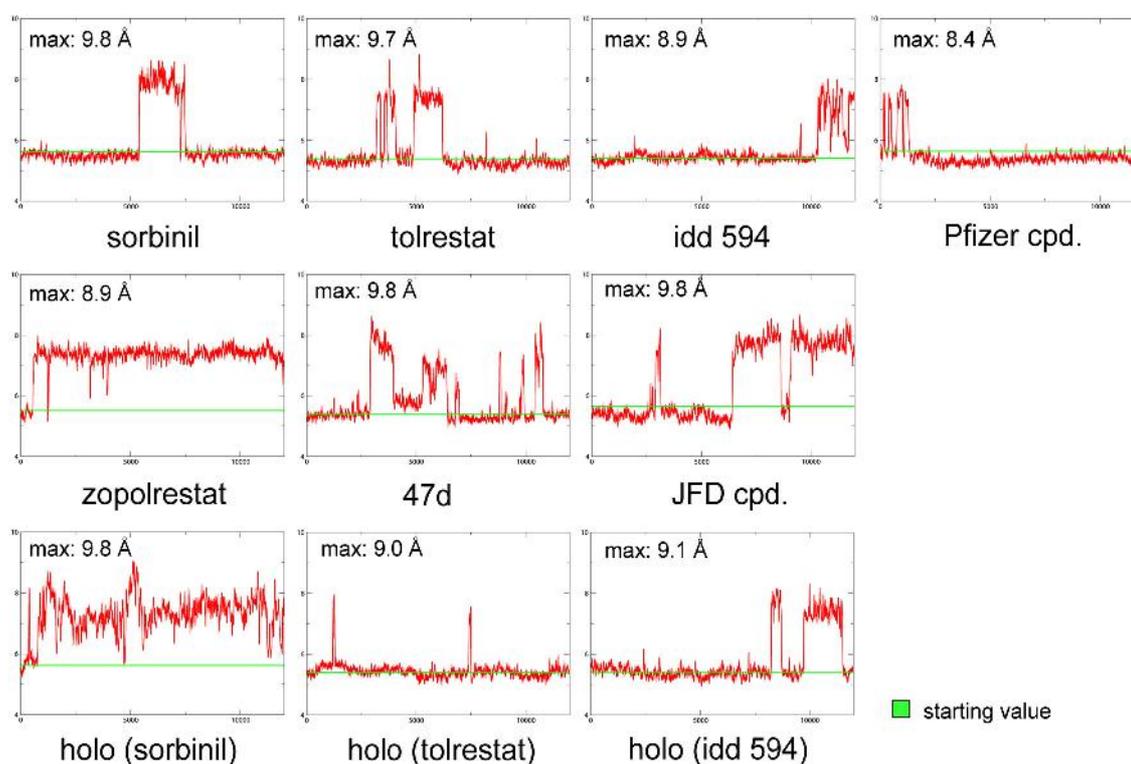


Figure 41: Time course of the distance between the ring centers of Trp 79 and Phe 122 for the individual trajectories. The value measured from the respective crystal structure is given as green line. Values are presented as running averages using a time window of 20 frames (10 ps). Thus, the local fluctuations are smoothed and short term transitions between subsequent frames are canceled out. Maximum distances observed in each simulation are given.

The situation is different for Phe 122. Once its χ_1 angle adopts the flipped orientation found in all simulations, an additionally expanded hydrophobic specificity pocket is produced directly above the ligand binding site. This new sub-pocket is bordered by Trp 79, Phe 115, and Phe 122. The distance between the ring centers of Trp 79 and Phe 122 is plotted in Figure 41 for all studied simulations. The distance between the ring centers of these residues in crystal structures is about 5.4 – 5.7 Å. In the flipped conformation this distance expands to maximally 9.8 Å. Figure 42 shows the binding pocket with open specificity pocket and Phe 122 in flipped conformation. Trp 79 closes up the specificity pocket and in flipped orientation Phe 122 gives additional space between Phe 122 and Trp 79.

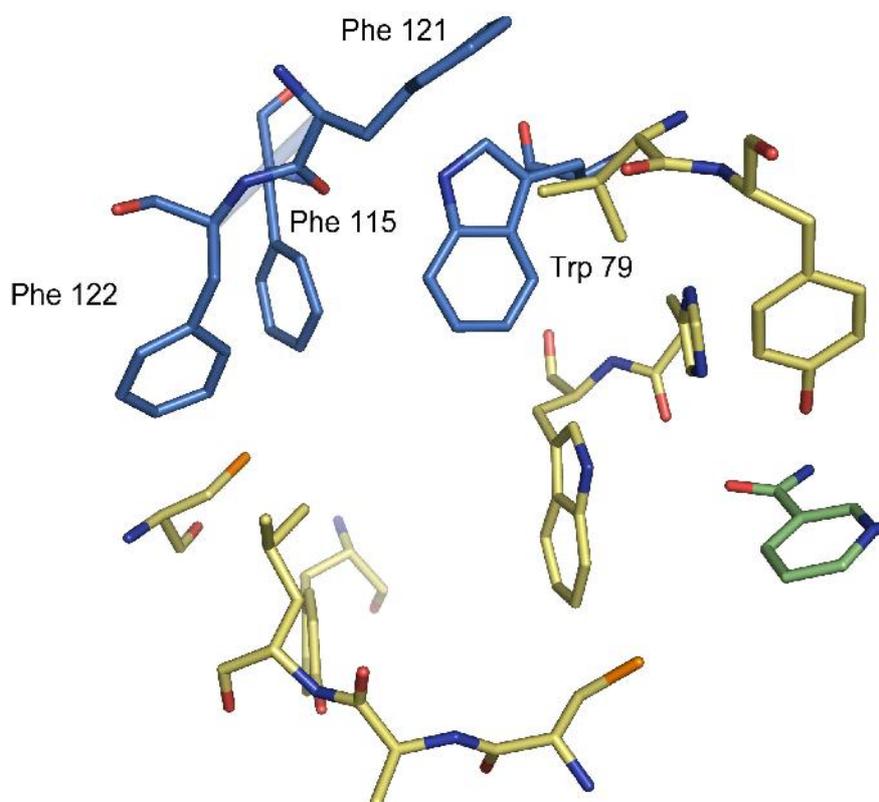


Figure 42: Binding pocket with open specificity pocket and Phe 122 in flipped conformation. The three residues aligning a small hydrophobic expansion of the pocket are shown in blue, the cofactor is depicted in green. All other binding site residues are shown in yellow.

The last two regions of interest, Tyr 309 and the C-terminal loop region around Ala 299 and Leu 300, have to be viewed as an assembly. As already mentioned, for Tyr 309 no major torsional flexibility can be detected. Instead, a rigid body type movement of the residue as a whole can be observed. Thus, it is instructive to analyze the distance between the OH group of Tyr 309 and the C α atom of Leu 300 to study whether concerted movements of these two residues alter the shape of the binding pocket.

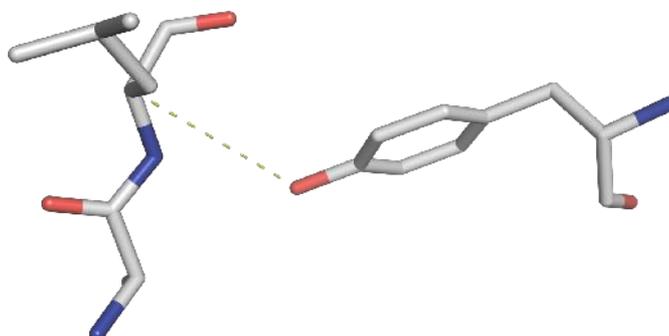


Figure 43: Distance between Leu 300 and Tyr 309. The structure is viewed from top where the ligand is located. The picture was taken from the Pfizer compound crystal structure.

Figure 43 shows the residues from viewed from within the binding pocket, where ligands are placed. If the distance between them is enlarged, additional space is created 'below' the ligand. In Figure 44 this distance is plotted over the course of the simulations. It can be seen that this distance shows great variations within some of the individual simulations, but also between different simulations. The movements of Leu 300 are in all cases larger than the ones observed for Tyr 309, thus, the fluctuations of Leu 300 within the single simulations play an important role for this distance. This partially explains the low variations in the course of the sorbinil and holo (sorbinil) simulations where movements of Leu 300 are rather moderate. However, the movements of Leu 300 do not fully explain the differences in the Leu 300 – Tyr 309 distance seen in Figure 44. For some of the simulations the distance increases to values as large as 10 Å. Tyr 309 partly confines the binding pocket below the respective ligand addressing the specificity pocket. Therefore, if Tyr 309 moves apart from the Leu 300 position, additional space is created, which is located below the part of the specificity pocket usually addressed by ligands.

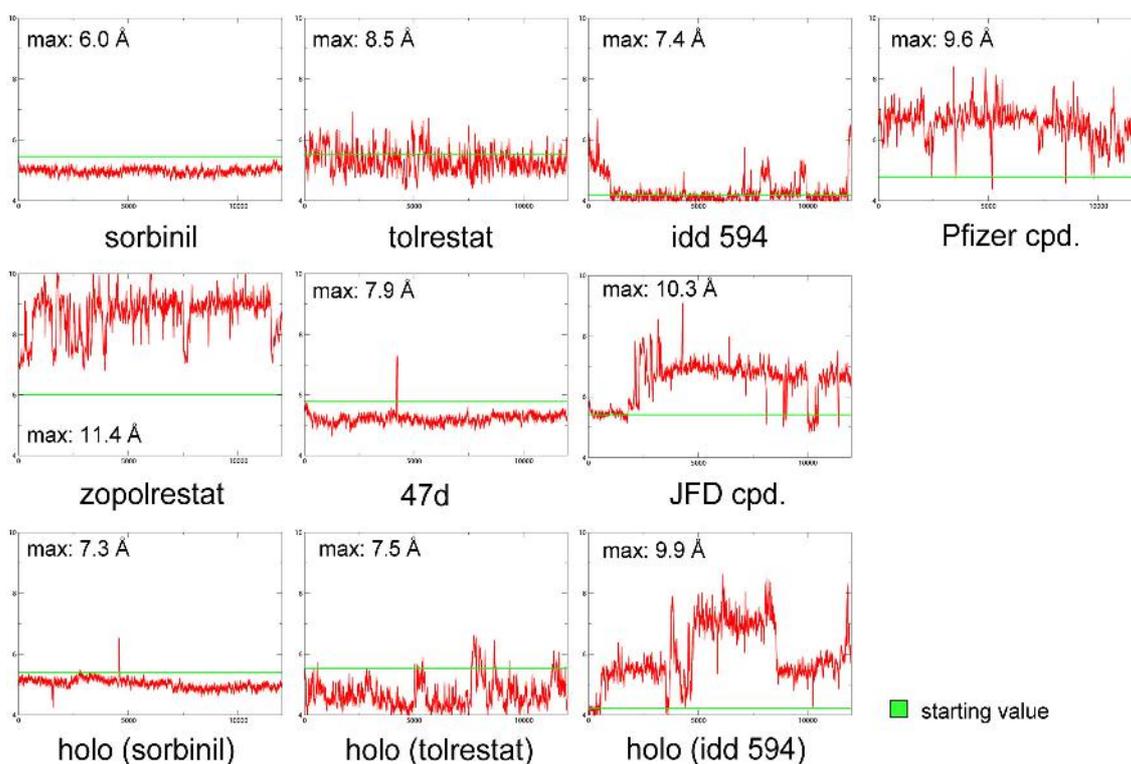


Figure 44: Plots of the distance between OH of Tyr 309 and Ca of Leu 300 for the individual MD simulations. The distance value measured from the crystal structure is indicated by a green line. As in Figure 41, values are presented as running averages using a time window of 20 frames (10 ps). The maximum values observed for each simulation are given.

In summary, two interesting changes of the AR binding pocket might create additional space to host putative ligands. Above the specificity pocket additional space is established *via* the rotation of the Phe 122 side-chain and below the specificity pocket the same is possible through concerted movements of Tyr 309 and Leu 300 relative to each other.

Addressing the new binding pocket conformations

Exploring the available conformational space of the AR binding pocket by comparative MD simulations has revealed potential new binding pocket conformations compared to the ones observed in crystal structures. However, no matter how careful the analysis of MD trajectories is carried out and how deliberate the results are interpreted, findings from MD simulations remain hypothetical as long as they are not supported by experimental observations.

One way to provide evidence for the suggested findings from MD simulations is to collect experimental results from crystal structures. For many of the findings discussed in the previous sections this has been done and it was shown that MD simulations are in good agreement with the observations from crystal structures. However, considering novel binding pocket conformers suggested by MD one runs into the problem, that there are no crystal structures to support the findings. Thus, the ultimate goal for such a study would be to find small molecules which bind to AR in a binding mode which stabilizes the binding pocket in one of the predicted new conformations. However, one will be faced with multiple challenges that need to be resolved. The first is to identify a snapshot from the MD simulation which is suitable for the subsequent docking and scoring step. The second problem is to find a suitable ligand scaffold which binds to AR with sufficient affinity. In addition, the scaffold has to allow ready accessible synthetic modifications at the positions which are suitable to address the putative new binding pocket conformations.

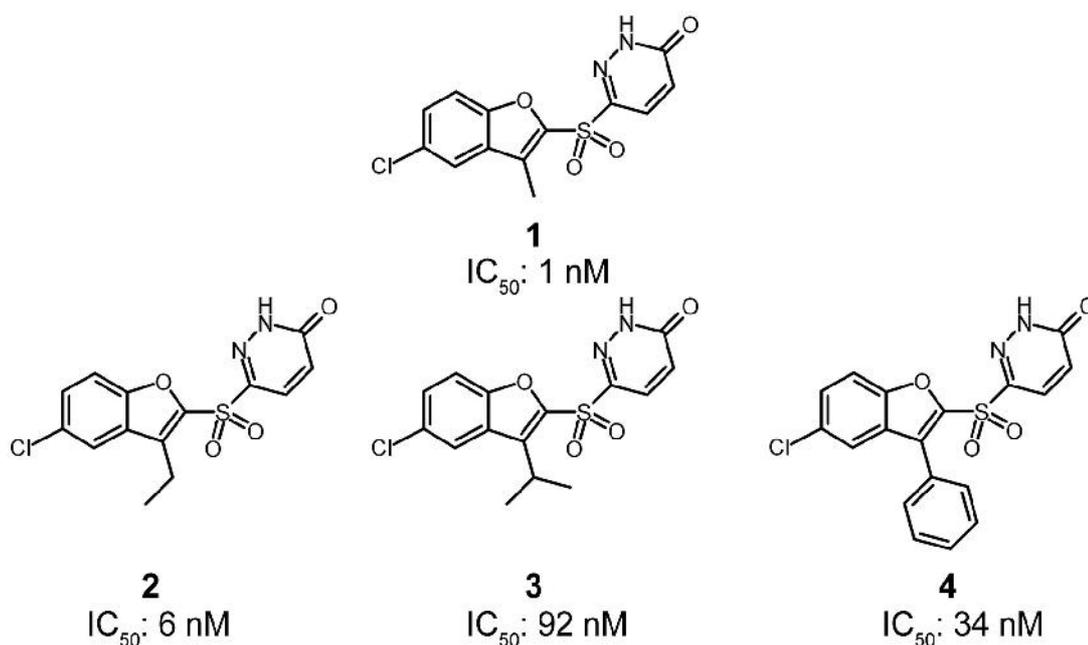


Figure 45: Molecules from the Pfizer inhibitor series published in 2005. IC_{50} values were taken from the publication by Mylari et al.¹⁸⁴.

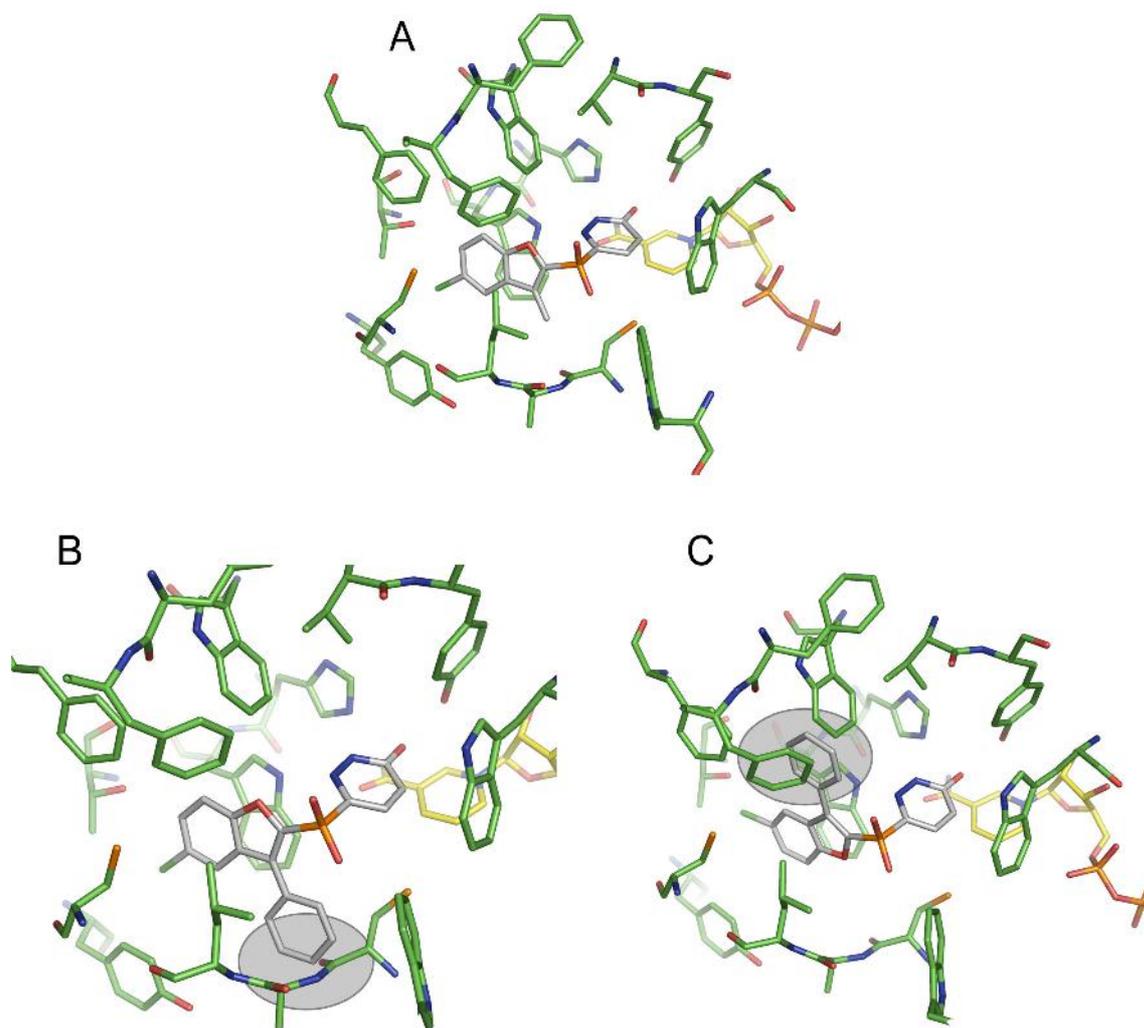


Figure 46: Crystal structure of the Pfizer compound **1** is shown in **A**. For **B**, the phenyl moiety of molecule **4** from Figure 44 was manually placed into the crystal structure of **1**. In **C**, a hypothetical flipped binding mode is shown, which was achieved by simply rotating the moiety addressing the specificity pocket. Regions where the hypothetical binding modes clash with the protein are shown in light gray.

In 2005 Pfizer published¹⁸⁴ additional information about the ligand series from which the Pfizer compound used in this study emerged as the most potent AR inhibitor. The paper disclosed additional ligands (Figure 45, molecules **2**, **3**, and **4**) which possess larger hydrophobic substituents in the position where the original Pfizer compound **1** exhibits a methyl group. As depicted in Figure 46 A, this group points towards the backbone of Cys

298 and Ala 299, a region which is known to be highly flexible. For molecules **2**, **3**, and **4** the published IC_{50} (Figure 45) are only slightly higher than for **1**. However, the binding modes of these ligands cannot be simply supposed as equal to compound **1**. In Figure 46 B, the phenyl moiety of **4** has been manually added to the crystal structure of **1**. As immediately obvious major clashes would be experienced between the phenyl moiety of the ligand and the protein backbone. As putative alternative binding mode a flipped conformation of the aromatic ring system of the ligand is suggested, which would place the phenyl ring in opposite direction. However, this would create major clashes with the residues Trp 79 and Phe 122, as shown in Figure 46 C. Thus, the Pfizer phenyl derivative **4** appears incompatible with the binding mode of the original Pfizer compound **1**.

To investigate the compatibility of **4** with the confirmed AR protein conformers, docking experiments were carried out using AutoDock¹²² following the 'in-situ' cross-docking approach^{124;168}.

The binding pockets of sorbinil, tolrestat and idd594 which are good representatives for the three major binding-pocket conformations, were used to compile the pasted grid. Hundred runs of the genetic algorithm were carried out using a population size of 50 and allowing for a maximum of 1500000 energy evaluations. These settings have proven to be successful in predicting the correct binding mode of the original Pfizer inhibitor using exactly the same grid setup¹⁶⁸. The 19 best-scored solutions address the idd 594 pocket which shows a conformation very similar to the pocket of **1**. However, as shown in Figure 47 the docking solutions suggest a reversed binding mode. The phenyl moiety points into the catalytic sub-pocket and the polar pyridazinon head group addresses the specificity pocket. Such a binding mode does not satisfy any of the well-established interactions patterns found for AR inhibitors and is indeed very unlikely according to current knowledge about AR. It also underlines that automated docking can be quite misleading and must be rigorously validated.

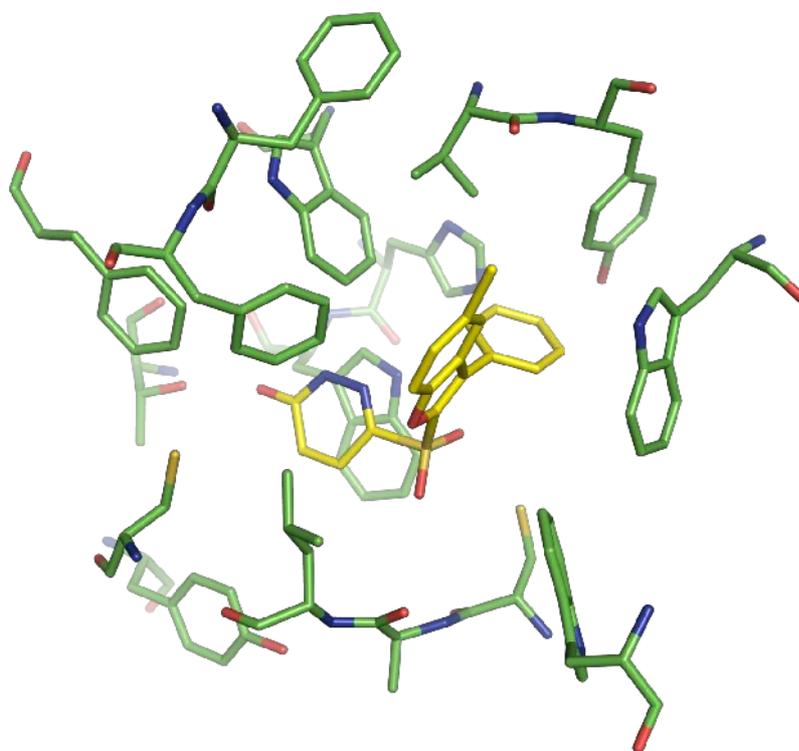


Figure 47: Proposed binding mode for the 19 best scored docking solution of the 'in-situ' cross docking carried out with the phenyl Pfizer compound into the binding pockets of sorbinil, tolrestat, and idd 594.

Not all solutions were found in the idd 594 pocket, the tolrestat and sorbinil pocket were also addressed. As expected, the docking poses in these pocket conformations were even less convincing than the ones in the idd 594 pocket. In summary, none of the docking poses found for **4** satisfies the known interaction patterns for binding to the enzyme.

As a next step **4** was docked into the crystal structure of compound **1**. The similarity between the two parent structures of the Pfizer compound and idd 594 in the crystalline phase are very high (Figure 48). Especially in the region of special interest, the flexible loop region next to Leu 300, the conformations are almost identical. Nevertheless different docking results are produced.

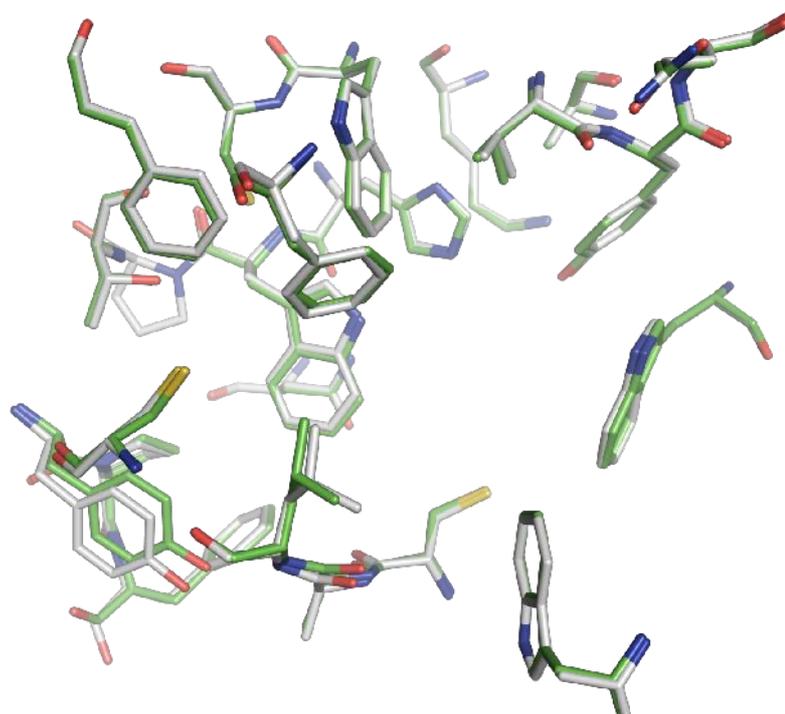


Figure 48: Superposition of the binding pockets of the Pfizer compound (silver) and idd 594 (green). Only subtle differences can be seen in the region around Leu 300.

The docking solutions on rank 1 to 5 are again not in agreement with present knowledge about reasonable binding modes in AR. The pyridazinon moiety is not placed into the catalytic pocket in a way to form essential hydrogen bonds with the protein. Figure 49 shows a docking solution which is found twice on position 6 and 7 out of 50 runs with the same settings as applied above. At first glance, this proposed solution is more in line with the expected binding mode of AR inhibitors. The pyridazinone moiety addresses the catalytic pocket and the aromatic moiety points into the specificity pocket. However, there are several issues which have to be considered while assessing this docking solution. In order to place the large phenyl moiety the sulfone group has to move upwards compared to the crystallographically studied methyl derivative. As a result, the 'upper' oxygen of the sulfone group and the furan ring show short contacts (2.9 and 2.8 Å) with the Phe 122 side chain. In addition, the phenyl ring shows short contacts with the Ala 299 main-chain oxygen (2.7 Å) and the side chain of Trp 219 (2.9 Å). The most severe clash occurs with the bulky sulfur atom of Cys 298, which is in only 2.8 Å distance. Thus, the docking solution

might propose a possible binding mode, but in order to accommodate the ligand, additional movements of the protein are necessary.

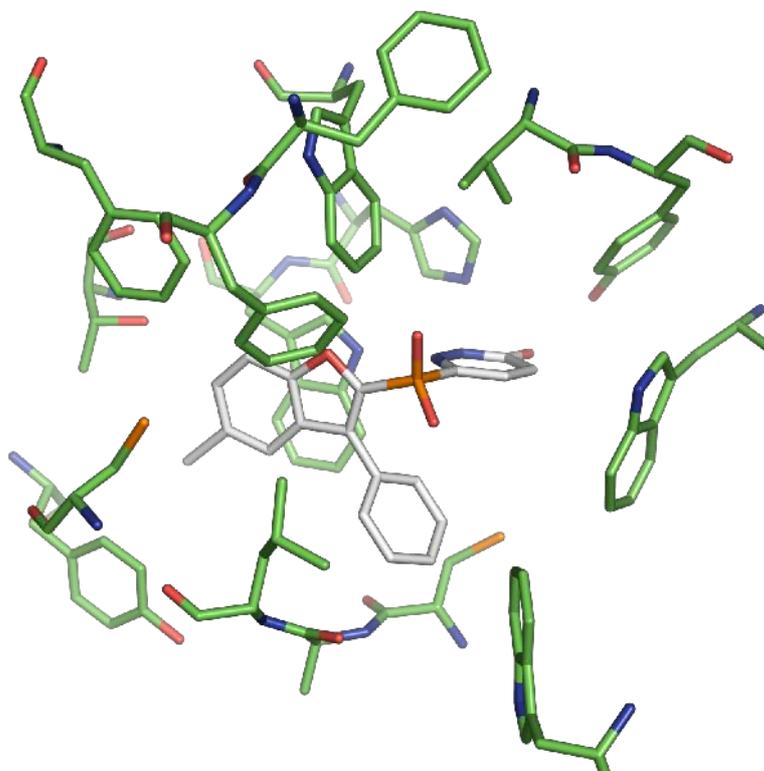
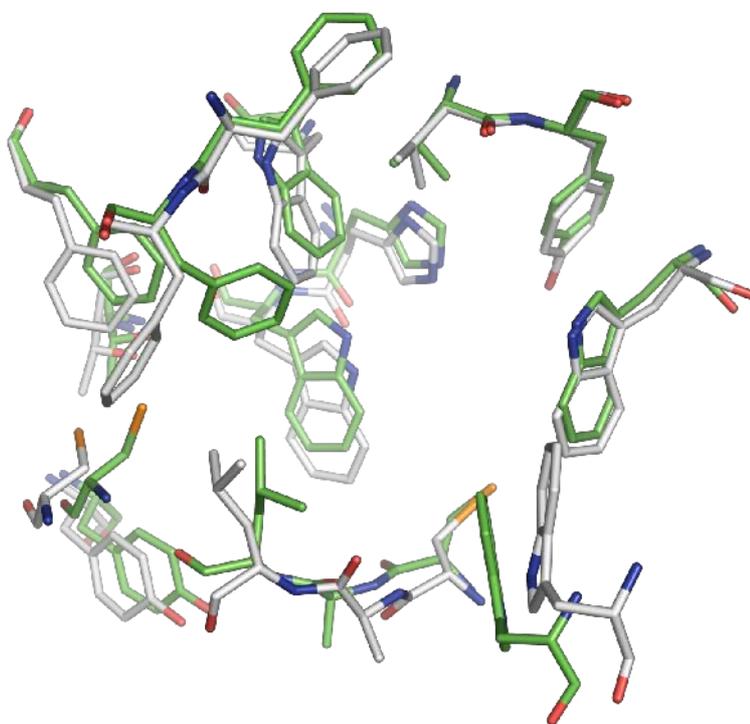


Figure 49: Rank 6 docking solution from docking the Pfizer phenyl derivative 4 into the crystal structure of the original Pfizer methyl derivative 1.

According to the different MD simulations, the region next to Leu 300 shows elevated flexibility. Thus, there might be the possibility that during the simulations conformations are sampled which provide more space to better accommodate **4** than the parent crystal structure of **1**.

Figure 50 shows a superposition of a snapshot from the trajectory performed on the crystalline complex of **1** together with the parent crystal structure. This snapshot shows primarily two interesting features: on the one hand, the side-chain of Phe 122 is in a flipped conformation and, on the other hand, the distance between Tyr 309 and Ala 299/Leu 300 is increased.



*Figure 50: Superposition of an MD snapshot (white) from a simulation with **1** and the parent crystal structure (green). Phe 122 is in flipped conformation and distance between Tyr 309 and Leu 300/ Ala 299 is enlarged.*

Using a pocket conformer generated by MD as input for a docking study with the aim to find a putative inhibitor for this new binding pocket conformation is a complex task. On the one hand, the selected snapshot should be significantly different from the conformations seen in known crystal structures. On the other hand, the used conformation should not exhibit obviously unrealistic features. Therefore, a snapshot from the complex trajectory of the original Pfizer methyl derivative **1** was selected by visual inspection. Taking the snapshot from the trajectory of the simulation of the complex with **1** has the advantage that large parts of the ligands are identical in both cases and similar behavior of the binding pocket can be supposed.

The selected MD snapshot was used for docking of **4** using AutoDock with the same settings as described above. Results are shown in Figure 51. In this solution the sulfur atom

of the sulfone moves downwards to place the phenyl ring between Ala 299 and Tyr 309. As a consequence, the pyridazinone moiety also slightly shifts its position, but all important hydrogen bonds between the polar head group and His 110, Trp 111, and Tyr 48 can still be formed. For this docking solution no short contacts between ligand and protein atoms occur.

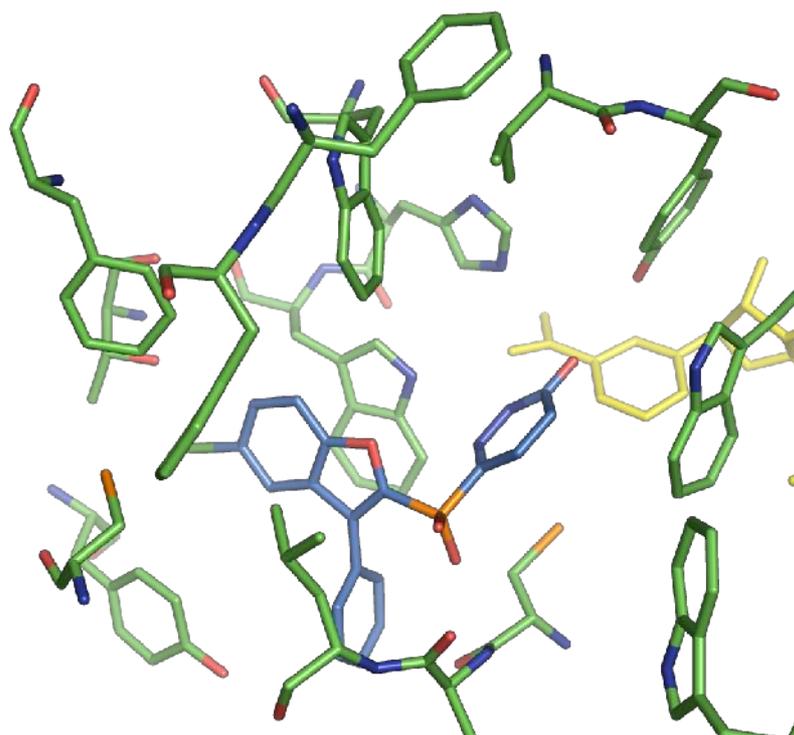


Figure 51: Solution from docking 4 into the MD snapshot derived from the complex simulation of 1.

Both solutions generated either from docking **4** into the crystal structure with **1** or into the MD derived snapshot, do not appear as perfect solutions. However, the latter solution based on the MD snapshot as input emerges as a plausible suggestion considering better the possible space provided by the conformational plasticity of the binding pocket.

Ligand **4** might be a candidate to stabilize the Ala 299/Leu 300 backbone in a conformation up to now unprecedented in crystal structures. However, a second flexibility feature has been seen in the MD simulations: this concerns the flip of the Phe 122 side chain. To lock this flipped conformation in one orientation a ligand would be required that places a hydrophobic moiety into this region. However, in literature no such ligand has been de-

scribed as AR inhibitor. As an alternative a ligand of known binding mode can be substituted in a way that the available space next to Phe 122 is sufficiently addressed and the desired flip is enforced.

To suggest a synthetically accessible candidate, a benzodiazepine scaffold (Figure 52 5) was chosen, which offers the possibility to introduce different substitution patterns in the two positions addressing both new putative binding pocket features: the Phe 122 side-chain flip and the Ala 299/Leu 300 main-chain movement. Benzodiazepines have shown in previous studies to inhibit AR with moderate affinity¹⁸⁵. However, no crystal structure of a suitable benzodiazepine inhibitor bound to AR is yet available.

Figure 52 shows the proposed molecules. Molecules **6** to **9**, as well as **10** and **11** are series of molecules with the same skeleton but linker chains of deviating length to connect the seven-membered ring and the carboxylate function. All molecules were docked into the structure of **1** and the corresponding snapshot from the MD simulation of **1** using AutoDock with standard settings as described above. No suitable docking solutions could be found for docking into the crystal structure, whereas docking into the MD snapshot led to promising results. The docked conformations of the molecules exhibiting the longer linker chain showed better scored docking modes. However, adding the linker chain introduces a stereo center into **5-9** and **12**. The docking solutions unanimously suggest that the S-enantiomer should better satisfy the stereochemical requirements to bind to AR.

After visual inspection the best docking results were obtained for **6** and **12** which differ only by an oxygen in the linker chain. As a representative, the best scored docking solution for **6** into the snapshot used for docking the Pfizer phenyl derivative **4** is presented in Figure 53. As can be seen, the carboxylic acid head group addresses the catalytic site, whereas the benzodiazepine scaffold binds to the specificity pocket and two of the phenyl rings address the new binding pocket features.

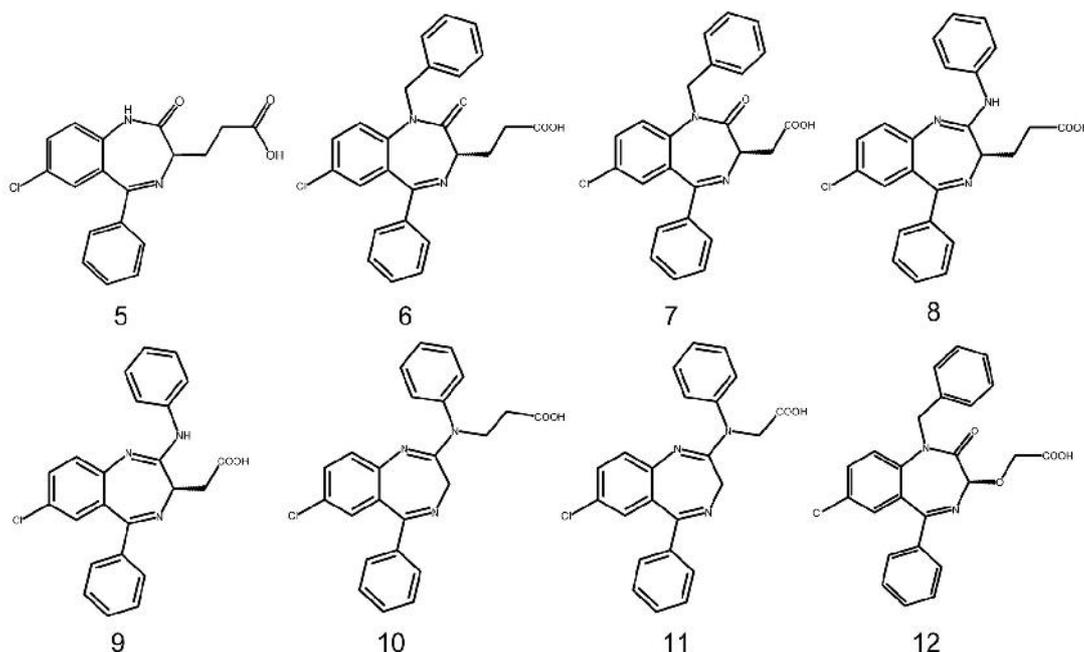


Figure 52: Proposed benzodiazepine-type AR inhibitors which possibly address the expanded binding pocket of AR.

The only molecule synthesized and tested so far is the racemate of **5** which shows an IC_{50} value of $\sim 700 \mu M$. However, this value has to be regarded with some caution, since the assay is not designed to measure inhibition at this low affinity range. Furthermore, due to the low amount of substance (~ 1 mg) which was obtained from synthesis and the thus only achievable low molar concentrations applied under the assay conditions, the standard deviation for this IC_{50} determination is of approximately the same size as the value itself. In addition, the compound was produced as trifluoro-acetate salt and it cannot be ruled out whether the weak inhibition is caused by the anion rather than **1** itself.

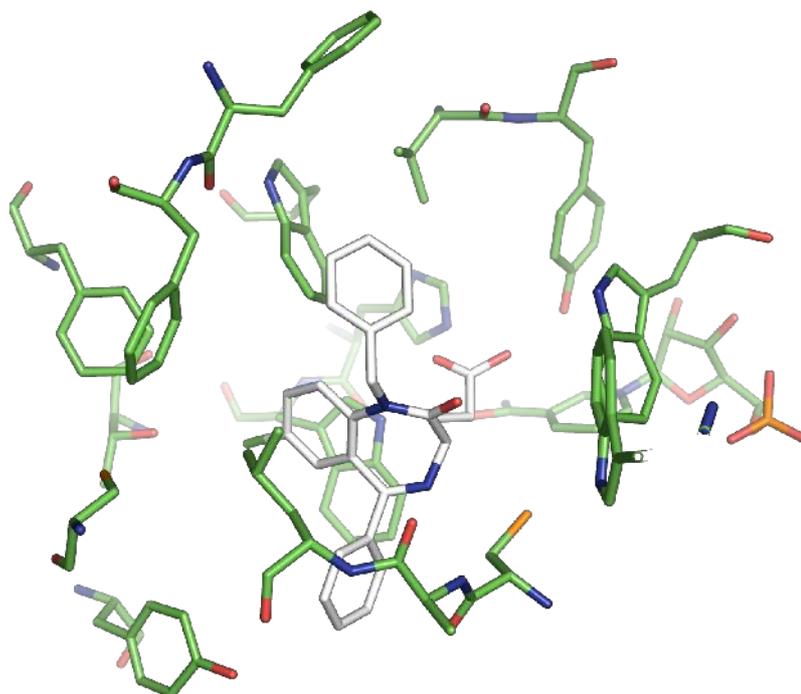


Figure 53: Docking solution of a benzodiazepine like molecule in an MD snapshot of the Pfizer compound complex simulation.

Upon summarizing this manuscript, a further crystal structure of AR in complex with a benzodiazepine-type inhibitor (Figure 54) was solved¹⁸⁶. This compound was published by Klebe et al. as a result from a virtual screening. In this complex the χ_1 angle of Phe 122 is rotated to almost to the flipped orientation to partly match the conformation seen in the MDs, thus this is the first experimental indication that this Phe 122 orientation is indeed possible. However, the crystal structure does not perfectly match the conformations seen in the MD simulations. The flip of the Phe 122 side chain does not result in the suggested opening of an additional sub-pocket in this region. Instead the Phe 122 side chain stills forms contacts with Trp 79. Crystals suitable for structure determination could only be obtained using cocrystallization techniques, soaking the compound into preformed crystals failed. Due to the different space group (P1 for cocrystallization compared to P21 usually obtained from soaking experiments) the crystal packing in the neighborhood of Phe 122 is different. This gives an indication that indeed crystal packing effects might be determinant in restricting Phe 122 in the conformation usually seen in crystal structures.

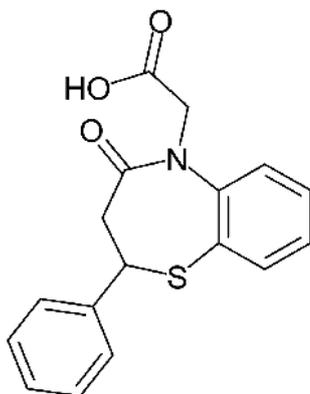


Figure 54: Benzodiazepine-type inhibitor²³⁴ which induces a conformational change of Phe 122 similar to the ones observed in the MD simulations.

However, despite promising docking results and initial indications from a crystal structure, the experimental validation for the new binding pocket conformers indicated by MD simulations is still to be provided and remains as a goal for the future of the project.

Summary and Conclusions

The conformational space available to the enzyme AR has been investigated by comparative crystal structure analysis and ten different MD simulations (seven different complexes plus three additional simulations of the holo enzyme starting from different pocket conformations). Compared a previous analysis of crystal structures of AR, many new high quality structures have been deposited with the PDB and several in-house structures have been determined. The analysis of the available crystal structures included only structures of the human enzyme, any mutated species have been discarded. Based on these crystal structures, the overall properties of the binding pocket of AR have been analyzed. The binding pocket can be divided into two sub-pockets: the catalytic and the specificity pocket. The catalytic sub-pocket shows almost no flexibility and the spatial positions of the individual residues are very well conserved between the distinct crystal structures. However, AR exhibits an 'induced-fit' binding mechanism to accommodate structurally deviating ligands. The specificity pocket can either be opened in two major conformations or entirely closed. Only a few residues of a flexible loop region next to Ala 299/Leu 300 are involved to produce the different pocket conformations.

A more detailed analysis of the structures exhibiting the same overall binding mode and accessing one of the three major protein conformers showed that despite the high similarities of the structures there are subtle differences between the individual structures which make them incompatible to each other in the majority of the cases. In light of docking attempts these findings are rather obstructive results. They provide the important message that docking has to consider conformational adaptability of the protein to produce successfully relevant binding modes. The protein conformation adopted with the inhibitor tolrestat has shown to be unique and requires the largest amount of adaptations compared to the holo structure.

The comparative analysis of the ten MD simulations showed an overall good agreement with the findings from the crystal structure analysis in terms of conformational variability in different parts of the binding pocket. The residues within the catalytic pocket showed only minor thermal motions and do not indicate conformational flexibility. For residues where fluctuations have been recorded in the MDs, this flexibility was in agreement with the crystal structure analysis in the majority of cases. Phe 122, Trp 219, and Tyr 309 exhibited conformational changes in the MD simulations which are not observed in the crystal structures. In case of Phe 122 and Tyr 309 this was to some degree attributed to contacts formed between symmetry-related molecules in the crystal-packing environment, which probably imposes additional restraints on these regions compared to the simulations which are carried out under conditions to reflect the properties in solution. For Trp 219 huge motions can be seen between crystal structures of the apo and holo form of AR. Thus, these motions indicated by MD are likely to be reasonable also with respect to the required opening of the cofactor binding pocket upon exchange of NADP⁺/NADPH.

The movements of Phe 122 on the one hand, and the combined movements of Tyr 309 and Ala 299/ Leu 300 suggested binding pocket conformations which have not yet been observed in crystal structures. The docking experiments carried out to find small molecule inhibitors which stabilize these new binding pocket conformations led to reasonable binding modes. However, the experimental validation is still an ongoing project.

Evaluating MM-PBSA in case of a flexible binding pocket: the AR test case

Introduction

Scoring functions for assessing protein-ligand binding affinities suffer from many deficiencies. They lack sound physico-chemical fundamentals and are inevitably based on many assumptions and simplifications. However, for many cases scoring of protein-ligand complexes works reasonably well and many drug-design projects greatly benefited from their performance. The main advantage of scoring functions are their extremely low computational costs. Scoring a protein-ligand complex is a matter of seconds. Hence, they can be applied to large datasets and are therefore valuable at the stages of a drug development project where computational methods are typically used: finding of initial hits and their optimization to useful leads.

Throughout a structure-based drug design project scoring functions are typically faced with two different challenges: selecting a near-native binding geometry out of set of decoy conformations and predicting the binding free energy of a putative ligand. Concerning the discrimination between decoy docking poses and a near native binding geometry, recent scoring functions perform rather well. Up to 87% success rate¹⁸⁷ can be achieved using as criterion to define a near-native docking pose a threshold of 2 Å rmsd with respect to the crystal structure. Concerning the prediction of binding constants for putative ligands, the accuracy of scoring functions has on the mean reached about one pK_i unit deviation from the experimental value. Usually, regression-based scoring functions¹⁸⁸ perform best for this task. The reason for this is that affinity information is included in the process of deriving these functions.

However, scoring functions fail in many cases for a variety of reasons. On the one hand, a dependence of the success rate on the type of protein can be registered. On the other hand, more fundamental problems arise which prohibit scoring functions to succeed in every case. From thermodynamics it is known that the formation of protein-ligand complexes is an equilibrium process. Assessing the involved energies based only on a single rigid structure of the bound state is not an adequate model for the process. Furthermore, impor-

tant contributions to the free energy of binding of a ligand to a protein are often neglected or only rudimentary included in scoring functions. Among them are (de-)solvation and entropic effects as well as conformational adaptations of protein and ligand upon complexation. The latter is especially important for proteins which are known to exhibit an 'induced-fit' binding mechanism for different ligands. In many cases the only energy contribution to the free energy of binding considered by scoring functions is the interaction energy between protein and ligand. Sometimes an energy term which covers the conformational energy of the ligand is also included in the calculations. However, the conformational energy of the protein is rarely included, likely on purpose since there is no method available to even estimate this term within a readily available time frame. However, it is reasonable to assume that different binding pocket conformations differ in their free energies, and this fact would have to be considered when calculating binding affinities.

Methodology

Within the last decades several approaches to more rigorously calculate free energies of binding have been developed based on different levels of physical theory. These methods provide in many cases reliable results, however, at the cost of significantly increased computational demands. Among these methods are: free energy perturbation (FEP)¹⁸⁹, thermodynamic integration (TI)¹⁹⁰, ligand interaction energy (LIE) approaches¹⁹¹, λ -dynamics¹⁹², and MM-PBSA.

The MM-PBSA method introduced by Kollman and Case¹⁹³ has probably gained the greatest interest in recent years. Its largest advantage over many of the other methods is probably its broad applicability. MM-PBSA combines force-field derived molecular mechanics (MM) energies, a continuum solvent Poisson-Boltzmann (PB) model for modeling polar solvation effects, and a solvent-accessible surface area (SA) dependent nonpolar solvation term to obtain the absolute free energy G of a system. The binding free energy of a complex is computed using equation (1) (see Methods for details):

$$\Delta G_{binding} = \langle G_{complex} \rangle - \langle G_{receptor} \rangle - \langle G_{ligand} \rangle . \quad (1)$$

The angle brackets indicate averaging over an ensemble of different conformations. The corresponding trajectories are usually derived using explicit-solvent molecular dynamics (MD) simulations. In recent years, several publications have emerged, showing excellent

agreement between calculated and experimental binding affinities. The scope of these studies covers protein-protein^{194,195}, protein-ligand¹⁹⁶⁻²⁰³, RNA/DNA-ligand interactions²⁰⁴, protein folding²⁰⁵, or comparison of the stability of different DNA/RNA conformations^{206,207}. As expected for such a broad spectrum of studies, these publications differ in many of the applied settings and parameters. Although there is no need to parameterize the method for every system under investigation, there are many settings can be adjusted in an MM-PBSA study.

Setup and workflow

The calculation and assessment of binding free energies using MM-PBSA can be roughly divided into four steps:

- 1) finding suitable starting structures for generating the ensemble;
- 2) sampling the conformational space of the system;
- 3) carrying out the actual MM-PBSA calculations;
- 4) evaluating the results.

To obtain starting geometries for the MD simulations several approaches are possible. The most reliable method is to start from an experimentally derived X-ray structure for each system. This implies solving a crystal structure for each complex plus one structure for the apo/holo enzyme. A second method is to propose the starting geometries via computational methods. Starting from an X-ray structure of the protein, the protein-ligand complex structures are generated via computational methods such as docking, or, alternatively, the ligands are placed manually into the binding pocket. If no structure of the target protein is available, a modeled structure obtained from homology modeling is also a possible alternative. The reliability, however, decreases dramatically when moving from experimental structures to homology models and the quality of the latter strongly depend on the degree of homology between template and candidate structure. Of course, one has to distinguish between different types of studies. In many cases the only reasonable way is to start from a modeled complex structure. However, for test systems which serve as examples to evaluate a method, the only reasonable way is to start with crystal structures.

The second step in the MM-PBSA work flow, the calculation of the ensemble trajectories, introduces many more possible variations. Among them, the most important are:

- a) preparation of input files;
- b) choosing different simulation models and parameters;
- c) carrying out the simulations.

The settings and parameters chosen here can have substantial impact on the results of MM-PBSA calculations. The preparation of protein structures for MD simulations is straightforward with most MD simulation packages. Consistent parameters and charges are available for every MD force field suitable to simulate proteins. The challenging part is the setup of the ligand(s) and/or cofactors. Despite some recent force field and program developments¹⁶⁰, not every ligand structure can be setup in a consistent and automated fashion, due to the high structural diversity in small molecules.

To establish consistent partial charges for the individual atoms of the ligand imposes one of the two most challenging tasks. For the Amber force field¹⁵⁸, the RESP¹⁶² method is the most widely accepted approach to derive charges, which are consistent with the parameterization of the force field itself. Therefore, the electrostatic field around the ligand is calculated using ab-initio methods. After that, partial charges are assigned to the individual atoms in a restrained iterative procedure. The target function of the optimization is the ab-initio electrostatic potential. The entire procedure is time consuming, not fully automated and requires access to an ab initio program, such as GAUSSIAN¹⁶¹. In the Amber program suite²⁰⁸ methods are available via the antechamber module; they try to produce similar charges in shorter times without having to perform high-level ab-initio calculations. While usually providing similar results, they introduce additional inconsistencies into the setup procedure.

The second problem of ligand preparation is the assignment of consistent force field parameters. Especially problematic is the assignment of parameters for dihedral angles, which usually have the highest impact on the accuracy of a simulation. Furthermore, manual parameterization of ligands is a tedious process, prohibiting application to large datasets.

Besides the force field parameters themselves, there are various other parameters and models to be chosen when preparing and setting up a system for MD calculations. Assignment of the solvation procedure is one of these critical steps. Most MM-PBSA studies use explicit solvent models, where the solvent is represented by separate molecules. Alternatively, implicit water models can be used, where the solvent is represented by a continuous medium with certain properties. The almost exclusively used model for generating MD trajectories used for MM-PBSA is the TIP3P¹⁶⁴ water model. To add water molecules to the system different approaches exist. The best choice is to embed the solute in a box of pre-equilibrated water molecules with a sufficiently large margin of several Å around the solute. To avoid edge effects, methods such as periodic boundary conditions can be utilized. A computationally less expensive solution is to use only a small cap of water molecules around the active site atoms of the solute. Solvent effects for the remaining part of the system are modeled via a continuum approach. To prevent explicit water molecules from penetrating into the area of implicit solvent, an additional force is added counteracting this behavior, but introducing a further non-physical component to the system. Problematic is also the treatment of the transition between the explicit water molecules and the implicit solvation model. The computationally cheapest approach is to choose an implicit solvent model for the whole system. However, this approach is rarely applied in the context of MM-PBSA.

Besides the treatment of the solvent, an appropriate setup of how to model the atoms of the solute is required. To save computational time, restraining some atoms in position is a commonly used strategy for MM-PBSA trajectories. To achieve this goal, artificial forces are introduced, which lack any kind of physical justification.

The treatment of non-bonded interactions and especially the electrostatic interactions is another important factor. For van-der-Waals interactions simple cutoff-based schemes are used. In principle, these schemes could also be applied to model electrostatic interactions at low computational costs. However, as electrostatic interactions decline very slowly with the interatomic distances, these arbitrarily defined cut-offs result in non-negligible artifacts. Nowadays, Ewald summation methods (PME)¹⁶⁷ are an elegant solution to circumvent most of these problems. However, not all MM-PBSA studies have used PME to cover electrostatic effects.

Besides assigning parameters and choosing a model, the procedure used to equilibrate the system and to run the simulations is of importance. In most cases, the system is first subjected to an energy minimization, followed by a short MD simulation in the pico-second time range to equilibrate the system.

Finally, the simulation period is crucial for obtaining meaningful results since representative ensembles for equilibrium thermodynamic properties can only be expected if a sufficient time window is selected. Given the current computer power, simulation times in the low nano-second range are feasible for most simulations of macromolecular systems. As such periods are too short to record large-scale movements, e.g., of domains, analyses of systems governed by such large-scale motions have to be interpreted with sufficient caution.

Once the trajectories are calculated, the next step is to carry out the MM-PBSA calculations. Although in contrast to other free energy methods such as LIE¹⁹¹, no empirical parameterization is necessary, there are multiple options available to set up an MM-PBSA run.

The most fundamental question is, whether the snapshots for complex, protein, and ligand are taken from one single trajectory or whether they are derived via separate simulations. If conformational adaptation processes of either the protein or the ligand occur upon binding, taking the snapshots from separate trajectories is the only way to consider these effects. By using only one trajectory for all species involved, their corresponding conformational spaces cannot be sampled independently. Hence, one of the main advantages of the method compared to single-structure scoring functions is lost: the ability to handle appropriately 'induced-fit' effects. At first sight, the main advantage of the single-trajectory approach are the lower computational costs. If a series of calculations is anticipated for the same protein, however, this advantage diminishes, since the reference trajectory for the uncomplexed protein needs to be calculated only once and can later on be re-used for every calculation. Furthermore, as pointed out by Pearlman in his seminal study²⁰³, the calculation of the trajectories is not the only time-consuming step in an MM-PBSA study. The computation of the electrostatic contribution to the solvation energy via the Poisson-Boltzmann equation is also a computationally expensive process. In case of single-trajectory

sampling it needs to be re-done for every receptor trajectory, but for the separate trajectory approach the receptor trajectory needs to be processed only once. To reduce the computational costs required to solve the PB equation, methods using the Generalized Born (GB) approximation have also been used for MM-PB(GB)SA studies and compared with PB results²⁰⁹⁻²¹¹.

A further computationally very expensive part of an MM-PBSA study is the estimation of the entropic contribution of the solute to the free energy of binding. Several different approaches are available to calculate this term²¹²⁻²¹⁵: normal mode analysis, quasi-harmonic analysis and the quasi-Gaussian approach. Only the first two are of relevance for biological macromolecules. The main difference between the two methods is that the matrix of the atomic fluctuations is in the first case the result of a normal mode analysis and only in the second (i.e. quasi-harmonic analysis) it is directly obtained from the MD trajectories. The advantage of quasi-harmonic analysis is that not only the harmonic contribution is taken into account. However, one will face inadequacies once the overall sampling of the different microstates is poor, a situation that is quite likely to occur with large biological systems such as protein-ligand complexes. Due to the very high computational costs, the calculation of entropic contributions is often omitted in MM-PBSA studies. For the sake of a relative ranking within a series of related ligands, this approximation is normally justified by the assumption that in these cases the entropic effects should be comparable in magnitude and, thus, negligible. However, exactly the same assumptions are made in using empirical scoring functions, thus one important relief to be recovered by sophisticated MM-PBSA calculations will be canceled. The number of snapshots used to estimate the entropic contribution is significantly lower than for the remaining calculations. In addition, the calculations are often restrained to a spherical region around the binding pocket. The calculation of entropic effects is probably the most crucial and least satisfying part of MM-PBSA.

The final step in an MM-PBSA study is the evaluation of the results. The procedure depends on the aim of the study and the available data. Ideally, in test studies the computed results are compared to experimental data, available as K_i or ΔG values. For relative ranking, IC_{50} values are also acceptable. However, since they are dependent on the exact assay conditions they should be treated with caution.

Literature overview

Having these points in mind, a short summary of some selected publications dealing with MM-PBSA shall be discussed briefly. Srinivasan et al.²⁰⁶ introduced the MM-PBSA method in its current form, although the basic approach had been described earlier²¹⁶⁻²²². The authors used MM-PBSA to computationally analyze the relative stability of the A and B form of DNA, RNA, and modified DNA sequences depending on different salt conditions. The study is based on previous investigations²²³ where the authors were able to show that MD simulations starting with either of the possible starting geometries converged to the corresponding experimentally preferred conformation. The obtained results were in good qualitative agreement with experimental findings. Since MM-PBSA provided interpretable results in terms of partitioning the interaction energy into individual contributions, the authors were able to rationalize the different preferential helix formations depending on varying conditions.

Massova et al.¹⁹⁴ expanded the scope of the method to investigating protein-protein interactions and introduced an extension to the method: computational alanine scanning. In this study, the binding of the oncoprotein Mdm2 to the tumor suppressor protein p53 was analyzed. Again, the calculated binding free energies were in reasonable qualitative agreement with experiment, independent whether single or separate trajectory sampling was applied. Starting from the wild-type, all 11 residues of a peptide derived from the Mdm2-binding stretch of p53 were computationally mutated and the effects of these mutations on the binding free energy were analyzed. Instead of re-calculating the MD trajectory for each mutant, the mutations were performed in-place using only the wild-type trajectory. Along the binding peptide three hot-spot residues were known from experiment. Computationally mutating each of these consecutively resulted in large losses of the corresponding binding free energy. Thus, Massova et al. were able to show that the method is capable of identifying key residues for protein-protein interactions.

Gohlke et al. published two papers^{209,210} presenting a study of protein-protein interactions using Ras–Raf and Ras–RalGDS complexes. Alongside these studies the authors presented some seminal methodological improvements and validations of the method. Different implicit solvent methods (Poisson-Boltzmann^{224,225}, Generalized Born²²⁶), as well as normal mode and quasi-harmonic analysis to estimate the entropy of binding were com-

pared. Using MD simulations of unprecedented 10 ns length, the convergence of MM-PBSA calculations was evaluated and a systematic survey of the effects of different parameters and settings was carefully carried out. Furthermore, an improved implementation to decompose the free energy on a per-residue basis was introduced. Via comparing energetic and entropic contributions of interface and non-interface residues as well as investigating dynamic and conformational changes of hotspot residues upon complexation the driving forces and dynamics of protein-protein interactions were elucidated.

Luo et al.¹⁹⁵ took the approach one step further and used MM-PBSA to study protein-protein interactions of structures derived by homology modeling. They studied the binding of NRG-1 β and some of its mutant peptides to ErbB3 and ErbB4 receptors. MM-PBSA's computational alanine scanning method was used to map the contribution of each residue involved in the interaction. The accordingly obtained interaction maps were in good agreement with data from experimental alanine scanning studies. Thus, the authors conclude that the homology models used were reliable and could be consulted to guide further experiments.

In 2006 Beà et al.²²⁷ used a combined MD and MM-PBSA study to investigate the thermodynamic driving forces of the chelate effect in cyclodextrin dimers. Five ns of MD trajectories were collected, which, due to the smaller nature of the system, provided extensive sampling of the conformational space available to the system. Hence, the authors were able to present a good agreement between the experimental finding and the computational predictions that, in this case, the complexation is an enthalpy-driven process.

Kuhn and Kollman were the first to extend the method to protein-ligand interactions. They applied MM-PBSA to a set of nine ligands binding to Avidin¹⁹⁶, which had been used previously²²⁸ to successfully calculate binding free energies using the LIE¹⁹¹ method. The range of binding free energies covered ~ 15 kcal/mol, and an impressive correlation coefficient of $r^2=0.92$ was achieved using MM-PBSA. This was superior to the correlation of $r^2=0.55$ obtained using the LIE method. Considering the individual contributions to the free energy of binding the authors proposed that the main driving force of complex formation in that case were favorable non-polar interactions. However, polar groups which did not find

adequate bonding partners in the protein compared to the uncomplexed situation in water contributed significantly unfavorably to the free energy of binding.

In recent years, a rising interest in applying MM-PBSA to protein-ligand interactions could be observed in literature. Wang et al.¹⁹⁷ successfully ranked a congeneric set of 12 HIV-1 RT inhibitors. In addition, the authors were the first to use MM-PBSA to assess the correct binding mode in a docking experiment of a small molecule to a protein. They were able to correctly predict the binding mode of efavirenz to HIV-1 RT. Huo et al.¹⁹⁸ ranked a set of seven Cathepsin-D inhibitors with a correlation coefficient of $r^2=0.96$ using single trajectory MM-PBSA. Gouda et al.²⁰⁴ used the method to predict RNA-ligand interactions. Absolute binding free energies as well as a relative ranking for a set of theophylline and five structurally close analogs were calculated and compared with the results from TI calculations. As expected for a congeneric series of ligands, TI performed better ($r^2=0.98$) than MM-PBSA ($r^2=0.51/0.82$ for different setups). However, one has to note that the TI calculations were carried out using only a subset of four molecules from the dataset. According to the authors, the inadequate treatment of solvation effects especially in the first water shell by the PB model, is responsible for the inferior, nevertheless still reasonable, performance of MM-PBSA compared to TI. Additionally, the first attempt was made to estimate the energy needed to convert the conformation predominantly present in solution into the conformation adapted in complex with a ligand. The calculated values for this conformational energy were estimated too high to be reasonable. The authors explain the deviations by errors in the force field as well as in the continuum solvent model. Suenaga et al.¹⁹⁹ examined interactions between the SH2 domain of Grb2 and different ErbB phosphotyrosyl peptides. For their data set of five peptides a correlation coefficient of $r^2=0.92$ was achieved. However, the calculated absolute values for the binding free energies were at least 50 kcal/mol off from the experimental values, thus limiting the success to a relative ranking of the ligands. Laitinen et al.²⁰⁰ used FEP and MM-PBSA to analyze the affinity and specificity of steroid binding to an anti-estradiol antibody. Both methods were able to reproduce the relative binding affinities of the steroids in good agreement with experimental values.

Recently, Steinbrecher et al.²⁰² further investigated the use of MM-PBSA for structure-based drug design. They proposed a multi-step scoring approach using human neutrophil

elastase as model case. MM-PBSA was successfully used as an intermediate filter to select docking poses for further processing using TI as a more sophisticated method.

In another paper Kuhn et al.²⁰¹, who first applied MM-PBSA to address protein-ligand interactions, stimulated current interest for the method, especially in an industrial environment. In this contribution, MM-PBSA was evaluated with regard to different applications, relevant for the industrial drug discovery process: virtual screening, de novo design, and ranking of related ligands. Within their well-designed study, the authors used different datasets for the different application areas. For ranking related ligands and virtual screening hits MM-PBSA performed for most datasets better than a combination of traditional docking and scoring. In case of de-novo design, MM-PBSA significantly supported prioritization and ranking of suggestions created by the de-novo program Skelgen^{229;230}. Additionally, they evaluated the MM-PBSA performance, whether multiple MD-generated snapshots or only single energy-minimized structures were used as input for the calculations. The latter was found to be equally powerful while saving considerable amount of computational effort, thus, implying that no additional information is obtained using multiple snapshots. Consequently, the computationally expensive process of performing MD simulations for each complex could be discarded. Hence, the most valuable use of MM-PBSA was found to be as a force-field based scoring function with an additional term considering solvation effects.

Pearlman²⁰³ published a critical MM-PBSA validation study in which the performance of the method to rank inhibitors according to their experimental IC₅₀ values was evaluated in comparison to a set of different scoring functions. Furthermore, data from a previous study²³¹ applying thermodynamic integration and the one-window free energy grid (OWFEG) method²³² to the same dataset was used to compare the performances of the different methods. As test case a congeneric series of 16 p38 MAP kinase inhibitors was selected. All ligands comprised the same scaffold with only small variations in their decorations. The IC₅₀ values ranged from 36 nM to 1.9 μM, covering roughly two orders of magnitude. Within this study, several variables and parameters were modified and the results compared: single as well as separate trajectories sampling was carried out for either one or five ns using varying levels of positional restraints on the system. Outer atoms were restrained in all cases, atoms within a 12 Å sphere of the ligand were either free to move or

restraint positionally. MM-PBSA performed worse than TI and OWFEG and only slightly better than most of the applied scoring functions in terms of ranking the ligands according to their affinity. By investigating the effects of different parameters and settings, it was shown that leaving the protein atoms near the binding pocket unrestrained was inferior to further damping the motions of these atoms. A combination of deficiencies of the force fields and inadequate sampling if the atoms near the binding pocket are free to move is assumed to be responsible for this finding. No conclusive picture was obtained concerning simulation length and evaluating separate or single trajectory sampling. However, the best overall performance was obtained for the combination of 5 ns of separate trajectory sampling using additional positional restraints. From this, the author concludes that MM-PBSA is able to make reasonable predictions if given enough care and computer resources in terms of sampling. However, the author emphasizes that MM-PBSA is not cost-effective for the system. Considering the enormous amount of computational time spent to obtain the results, the performance of MM-PBSA is disappointing. Furthermore, he points out that multiple trajectories sampling should be the method of choice, instead of allegedly saving time by extracting all snapshots using one trajectory only.

In summary, MM-PBSA has been applied with reasonable success in most cases, but also drawbacks and failures of the method have been experienced. In general, the diversity of the small molecule ligands used in the studies is larger than what can be handled using more rigorous techniques such as free energy perturbation or thermodynamic integration. However, in many of the cases the ligands under investigation shared a common scaffold and the binding modes to the target proteins were similar. Thus, it can be concluded that the conformations of the proteins in the complexed states are similar to those in the unbound state. This is demonstrated by the fact that the majority of the studies used the single trajectory approach where both states are considered equal and all internal energy contributions cancel out. Some studies applied also the separate trajectory approach, but the authors had to realize that noise and inaccuracies introduced by independent sampling in most cases outweighs the advantages. Furthermore, due to the high computational costs, especially when using large datasets of molecules, the MD sampling requirements are in most cases short. In addition, the applied protocols which restrain most atoms of the protein with respect to their positions were generally rather strict.

SupplTable 1 and SupplTable 2 of the Supplementary Material summarize the settings used in the studies cited above. Usually, the production length of the MD simulation for MM-PBSA studies dealing with datasets of protein-ligand complexes is in the hundred picosecond range, making it difficult to achieve sufficient sampling for macromolecular systems. The movement of protein atoms is often confined to a discrete sphere around the binding site, thus keeping the rest of the protein rigid. Simultaneously, the presence of explicit solvent water molecules is often limited to a spherical region around the ligand. By committing these simplifications it is implicitly assumed that no adaptation processes occur during complexation, since only then it is valid to assume that the generated trajectories are suitable to provide additional information compared to a single protein geometry scoring. Starting geometries for the complexes are in none of the cases exclusively obtained from X-ray crystallography. Instead, many of the input structures used for the MD simulations are modeled via docking, manual placements or homology modeling, thus introducing additional structural assumptions and uncertainties. Combined with the short sampling time this provokes the concern, whether the system is given enough time to fully relax and sample states of a thermodynamic equilibrium. In addition, entropic effects are omitted in many cases. Again, this might be justified with negligible differences between molecules exhibiting similar binding modes and proteins which show no significant differences between complexed and free form. A more practical reason for neglecting entropic effects are the large computational costs accompanied with estimating the entropy of a system by means of normal mode or quasi-harmonic analysis.

Despite its apparent success in many cases, the question remains how the method performs if applied to a system where most of the usual implicit assumptions and simplifications are not valid: a protein which exhibits a pronounced induced-fit adaptations upon binding a set of highly diverse ligands. In order to apply MM-PBSA to such a system, one has to increase the sampling and eliminate as many assumptions as possible. This means using only crystal structures as starting geometries and sampling times at least in the nanosecond range for each complex. Furthermore, no restraints should be applied to the system and the parameters both for the MD simulations and the MM-PBSA calculations are must be set to maximize accuracy instead of minimizing computational costs.

The test system: Aldose Reductase

Aldose Reductase (AR) is an ideal test case for such a study, since it is an enzyme which exhibits an 'induced-fit' binding mechanism with significant differences between individual binding-pocket conformations. Structurally, it is a very well characterized enzyme with many high-resolution crystal structures available^{114,126}. The enzyme comprises an α/β -TIM barrel fold with only little structural variations between different structures¹⁷. The binding pocket of AR itself can be divided into two sub-pockets: a catalytic or anion pocket and a hydrophobic specificity pocket. The cofactor as well as the residues involved in catalysis are located around the catalytic pocket (see Figure 55 A). As expected for a catalytic sub-site, very little flexibility is observed within that part of the binding pocket (see Figure 55 B). The specificity pocket, however, can accommodate different ligands by adjusting its overall shape. To accomplish this adaptation the protein utilizes only a few flexible amino acids. The opening and closing of the specificity pocket is regulated *via* a leucine residue (Leu 300), which acts as a gate-keeper residue. The crucial parameter to record the state of the specificity pocket is the distance between Leu 300 and Trp 111. Besides side-chain flexibility of Leu 300, backbone movements of the loop region next to this gate-keeper residue are involved in defining the conformation of the specificity pocket. Since no large-scale movements are involved, multi-nanosecond MD trajectories should suffice to sample the different AR pocket conformations adequately.

All AR ligands observed in crystal structures bind to one of three specificity pocket conformations. Figure 55 B and C show representative structures for these three conformations. Figure 55 B shows a superposition of the binding pocket based on an all-residue C α atom fit. Residues confining the catalytic pocket superimpose very well, whereas flexibility is mostly confined to the small loop region ranging from Val 297 to Leu 300. The left image in Figure 55 C depicts the closed form of the binding pocket (shown in green), as observed in complex with the inhibitor sorbinil (**1**). The central image (blue) represents the open pocket in a conformation observed in complex with tolrestat (**2**). In this conformation the Leu 300 side chain adopts a unique conformation compared to all other crystal structures. The right image (red) represents the binding pocket conformation seen with many of the high-affinity ligands. The distance between Leu 300 and Trp 111 is increased, the

specificity pocket is opened and a hydrophobic moiety of the ligand penetrates into the hydrophobic cleft.

AR belongs to the family of the aldo-keto reductases. It catalyzes the reduction of different aldoses and aldehydes to the corresponding alcohols using NADPH as cofactor. Even though the physiological function of the enzyme is not fully understood, a role as a general detoxifying enzyme for miscellaneous aldehydes is discussed¹¹. Thus, AR accepts a broad variety of substrates¹⁴. It is the first and rate-limiting enzyme of the sorbitol pathway. Within this pathway, AR is responsible for converting glucose into sorbitol. Under physiological conditions the sorbitol pathway is not utilized to metabolize glucose. Under diabetic conditions, however, up to 30% of the excessive glucose is processed via AR^{8;9;144}. As a consequence, NADPH is depleted, inducing oxidative stress in the cells¹⁴⁶. Additionally, sorbitol accumulates within the cells, causing increased osmotic stress. A variety of the late-onset diabetic complications have been attributed to glucose processing by AR^{145;149}. Accordingly, AR has become a target for the development of anti-diabetic drugs¹⁴⁸. The enzyme has proven to prolongate diabetic complications in model systems¹⁴⁷ as well as in clinical studies⁸⁸. However, most inhibitors have failed so far in clinical trials for various reasons^{78;148}.

Reflecting the broad variety of substrates accepted by the enzyme, inhibitors binding to AR are chemically diverse. To match the dual nature of the binding pocket, AR inhibitors exhibit a polar head group which addresses the catalytic subpocket. The second structural element of many AR inhibitors is a hydrophobic moiety, which addresses the specificity pocket. Ultra-high resolution crystal structures as well as neutron scattering experiments²³³ suggest, that AR inhibitors bind with deprotonated polar head groups.

AR has been targeted successfully in structure-based design projects²³⁴ using either combined virtual screening and docking approaches^{115;235-237} or de-novo design methods²³⁸. However, these studies were limited by applying the corresponding methods to one specific conformation of the binding pocket.

AR is a threefold challenge for MM-PBSA calculations: a) small molecules binding to the enzyme are chemically diverse, b) inhibitors presumably bind in a charged form, which

has been described earlier¹⁹⁷ to cause problems and inaccuracies, and c) ligands bind to different protein conformations.

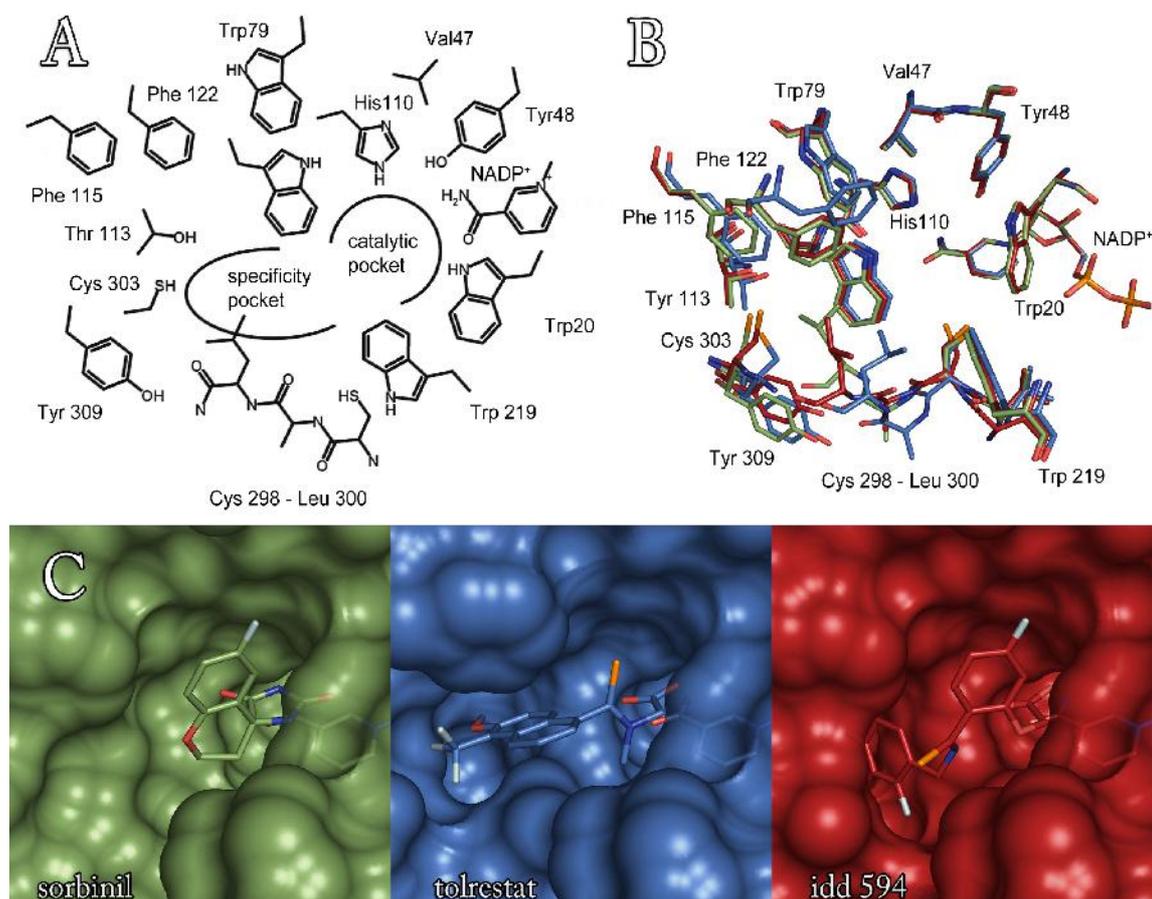


Figure 55: Overview over the three known AR binding-pocket conformations. *A* shows a schematic representation of the residues around the binding pocket. *B* shows a superposition of the same residues. Flexible and rigid parts of the binding pocket can be distinguished. The binding pocket conformation of sorbiniil is shown in green, the blue structure represents the conformation seen in the tolrestat complex and the red structure depicts the pocket conformation observed with idd 594. *C* shows a surface representation of the same three crystal structures. The color coding is the same as in *B*.

Materials and Methods

Data set

The seven ligands used in this study are depicted in Table 12. Representatives for the three main AR conformations observed in crystal structures, as well as for the most potent polar head groups (carboxylic acid, hydantoin, pyridazinone) binding to the catalytic sub-pocket of AR are included in the dataset.

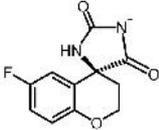
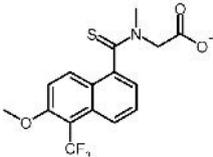
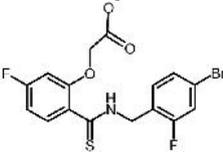
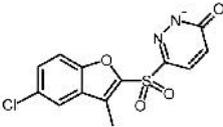
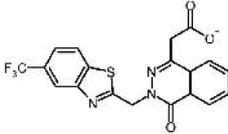
The crystal structures of human AR in complex with sorbinil **1**, tolrestat **2** (pdbcode: 2fzd)²³⁹, the Pfizer compound **4** (pdbcode: 1z89)¹²⁶, zopolrestat **5** (pdbcode: 2fz9)²³⁹, 47d **6**, as well as the JFD compound **7** were determined in our lab. For the idd 594 (**3**) complex input coordinates were taken from pdb entry 1us0¹¹⁴. Complex structures of AR with sorbinil, tolrestat, and zopolrestat have been previously deposited with the PDB. However, the structures with sorbinil (pdbcode: 1ah0)¹²⁰ and tolrestat (pdbcode: 1ah3)¹²⁰ were derived from porcine AR and the structure with zopolrestat (pdbcode: 1mar)¹⁷⁹ contains only the C α trace of the protein.

For all protein-ligand complexes ITC measurements were carried out to determine the free energy of binding, thus providing a consistent set of high quality affinity data. The corresponding values are also given in Table 12. Although the narrow range in affinity of 3.8 kcal/mol is not ideal for assessing the predictive power of MM-PBSA in terms of ranking related ligands, the quality of the input structures and the thermodynamic measurements are unprecedented for evaluating MM-PBSA results. However, one has to note that the initial selection of the dataset was based on published IC₅₀ data, since no ITC data were available at that time. Using these values, the affinity range of the dataset is from 0.84 nM for the Pfizer compound **4** to 4.1 μ M for the JFD compound **7**, thus spanning almost four orders of magnitude. The main difference arises from the deviating values obtained for the Pfizer compound which shows the best IC₅₀ value in literature^{125;184}, but only a moderate binding free energy in the ITC experiments¹²⁶.

ITC measurements

Experimental ΔG values were obtained from Isothermal Titration Calorimetry using an MCS ITC-instrument from MicroCal Inc.(Northampton, USA). Details about the experimental setup and data analysis have been published elsewhere¹²⁶. All measurements were

carried out at 298 K and a pH of 8.0. Data were analyzed using the ORIGIN software (MicroCal Inc.) using a single-site binding model. Experimental heats of the protein-inhibitor titration were corrected for heats of dilution by subtracting the corresponding data of a blank titration (inhibitor into buffer). Standard Gibbs free energy values were calculated using the equation $\Delta G^0 = -RT \ln K_b$, where $R = 8.3144 \text{ J mol}^{-1} \text{ K}^{-1}$, and K_b is the binding constant.

Ligand	Formula	IC ₅₀ (literature) [nM]	ΔG_{ITC} [kcal/mol]	Rank ΔG_{ITC}
sorbinil (1)		150-1000 ²⁴⁰ , 1500 ¹¹⁹ , 2000 ^{181;241}	-9.1	5
tolrestat (2)		35 ¹¹⁹ , 100 ^{181;241}	-9.2	4
idd 594 (3)		30 ¹¹⁴	-9.8	2
Pfizer compound (4)		0.8 ¹²⁵ , 1 ¹⁸⁴	-8.2	6
zopolrestat (5)		3 ²⁴² , 60 ¹⁸¹	-11.3	1

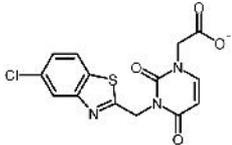
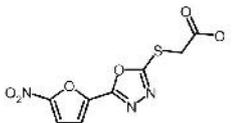
Ligand	Formula	IC ₅₀ (literature) [nM]	ΔG _{ITC} [kcal/mol]	Rank ΔG _{ITC}
47d (6)		6 ¹⁸⁰	-9.3	3
JFD compound (7)		4100 ¹¹⁵	-7.5	7

Table 12: Data set used for the MM-PBSA study. In the first two columns the names of the compounds and chemical formulas are given. Column three provides the IC₅₀ values as found in the cited literature. In column four the free energy of binding is given as obtained from ITC experiments (see Materials and Methods). In column five the rank order of the compounds is given as obtained from the ITC derived free energy of binding.

Molecular Dynamics Simulations

Starting geometries for each protein-ligand complex were taken from the corresponding crystal structures. Starting geometries for the simulations of the holo enzymes were derived by extracting the ligands sorbinil, tolrestat and idd 594 from their corresponding crystal structure. All preparation steps and simulations were carried out with the AMBER 8.0¹⁵⁷ suite of programs, using the Cornell et al.¹⁵⁸ force field in the variant of the parm99 parameter set¹⁵⁹. Parameters for the ligand were taken from the GAFF force field¹⁶⁰ and assigned using ANTECHAMBER. Parameters for the cofactor NADP⁺ were taken from a previous study¹¹³. Atomic charges for both the ligand and the cofactor were calculated by fitting to the HF/6-31G* electrostatic potential using the conformation observed in the crystal structure. The corresponding ab-initio calculations were performed with GAUSSIAN98¹⁶¹, the restrained electrostatic potential fit with the ANTECHAMBER program using the RESP method^{162;163}. Hydrogen atoms were added to the protein using Amber templates. The simulated protein thus consisted of 316 amino acids or 5071 atoms, plus the cofactor with 73 atoms and the ligand atoms. The crystal structure was subjected to a short initial energy minimization in vacuo using 20 steps of steepest descent followed by 180

steps conjugate gradient minimization. After addition of two sodium counterions to ensure neutrality (placed at minima of the electrostatic potential), the systems were solvated in a box of TIP3P¹⁶⁴ water molecules, which resulted in box sizes of $\sim 80\text{\AA} \times 65\text{\AA} \times 75\text{\AA}$ and a total of ~ 9000 water molecules. The MD simulations were then started by heating the solvent to 300K over a period of 20 ps and cooling to 100 K over a period of 5 ps, keeping the solute fixed. After this procedure, the entire system was gradually brought to 300 K over a period of 25 ps. The simulation was then carried on for 6050 ps under constant temperature and pressure (NPT), applying periodic boundary conditions. The temperature was kept constant by coupling to a heat bath through the Berendsen algorithm¹⁶⁵. Pressure was adjusted by isotropic position scaling using a Berendsen-like algorithm. Covalent bonds to hydrogen atoms were constrained by the SHAKE¹⁶⁶ algorithm and a time step of 2 fs was used. A cutoff of 8 Å was applied to the van der Waals interactions, while the electrostatics were treated by the Particle Mesh Ewald method¹⁶⁷. The simulations were carried out with the SANDER MD module of AMBER 8.0. For analysis, energy data were saved every 10 time steps, solvent and solute coordinates every 0.5 ps. All results presented refer to the 6.0 ns trajectories which exclude the first 100 ps required for temperature adjustment and equilibration.

MM-PBSA calculations

MM-PBSA is based on the assumption that the free energy G_A of a system A can be represented as a sum of different terms. According to equation (2) G_A can be dissected into a gas-phase contribution and a solvation energy term:

$$G_A = G_{gas} + G_{solv} \quad (2)$$

The energy for the gas-phase contribution G_{gas} can be approximated via

$$G_{gas} = E_{gas} - TS_{solute} \quad (3)$$

where E_{gas} is the Molecular Mechanics (MM) energy directly available from the employed force field. All water molecules as well as counterions are stripped off the snapshots prior to the MM-PBSA calculations. The terms for van-der-Waals interactions, internal energies, and electrostatic contributions of the solute are calculated without applying any cutoff to avoid artifacts at the cutoff distance. S_{solute} is the rotational, translational, and vibrational entropy of the solute estimated using normal mode analysis (NMA)²¹². T is the absolute

temperature in Kelvin at which the simulations took place. G_{solv} from equation (2) is further dissected into an electrostatic as well as a nonpolar contribution (equation (4)).

$$G_{solv} = G_{elec} + G_{nonpolar} \quad (4)$$

The electrostatic contribution G_{elec} is computed using a finite difference solution to the Poisson-Boltzmann (PB) equation^{224,225}. Delphi II²⁴³ is used as PB solver applying dielectric constants of $\epsilon=1$ for the protein and 80 for the solvent. A cubic lattice with a grid spacing of 0.5 Å is 80% filled with the solute. The dielectric boundary is taken as the solvent accessible surface defined by a probe atom with a radius of 1.4 Å. The radii for the PB calculations are taken from the PARSE parameter set²⁴⁴. Missing radii for halogens were added using Amber Parm99 force field parameters. Atomic charges for protein, cofactor and ligands were the same as in the corresponding MD simulations described above. A salt concentration of 150 mM was used for the calculation and 300 iteration steps were performed. To test the calculations for convergence, PB energies for a subset of the data set were evaluated using 1000 iterations, leading to the same results as in the case of 300 iterations (data not shown).

The nonpolar contribution to the solvation free energy was calculated using equation (5). MOLSURF was used to compute the solvent accessible surface (SASA) as implemented in the MM-PBSA module of AMBER 8.

$$\Delta G_{nonpolar} = \gamma \frac{\text{kcal}}{\text{mol } \text{Å}^2} \times \text{SASA} + b \frac{\text{kcal}}{\text{mol}} \quad (5)$$

In agreement with literature, b and γ were set to 0.92 and 0.00542, respectively.

The solute entropy is estimated using the NMODE module of Amber 8. All parameters were kept at their standard values as provided with Amber 8. A maximum of 1000 cycles of minimization were applied using a 0.1 kcal/mol convergence criterion for the energy gradient. Despite the high computational demands, the normal mode calculations were carried out for all atoms of the complex using a distance-dependent dielectric constant of $\epsilon=4R_{ij}$, with R_{ij} being the distance between two atoms.

Finally, the binding free energy of a protein-ligand complex is computed using equation (6)

$$\Delta G_{binding} = \langle G_{complex} \rangle - \langle G_{receptor} \rangle - \langle G_{ligand} \rangle \quad (6)$$

where the angle brackets indicate averaging over an ensemble of conformations. To calculate G_{gas} (cf. eq. 2,3) snapshots were extracted every 10 ps from the generated MD trajectories, resulting in an ensemble of 600 snapshots. Due to the considerably higher computational costs of normal mode analysis calculations, the time period for the estimation of absolute entropies was chosen to be 300 ps, leading to an ensemble of 20 snapshots. The time gap of at least 10 ps ensures that only uncorrelated snapshots are used for energy calculation.

To derive snapshots for complex, protein, and ligand two distinct approaches were applied. The more rigorous method is to generate the corresponding snapshots from separate trajectories, i.e. the degrees of freedom for protein and ligand as well as complex are simulated independently. The second possibility is to extract the snapshots for all entities from a single trajectory. In this case, no independent sampling is performed and the resulting snapshots for complex, protein, and ligand are highly correlated. At first glance, the obvious advantage of using only a single trajectory are the lower computational costs. Nevertheless, as pointed out by Pearlman²⁰³, the time savings are not significant when applying MM-PBSA to a series of ligands binding to the same protein, as done in this contribution.

Calculation of Predictive Indices (PI)

To assess the ability of a scoring method in terms of ranking a set of ligands in a more quantitative manner Pearlman and Charifson²³¹ introduced the concept of the predictive index (PI). The PI for a given series of experimental and calculated affinities is derived according to equation 7 as a sum over all pairwise ligand comparisons.

$$PI = \frac{\sum_{j>i} \sum_i \omega_{ij} C_{ij}}{\sum_{j>i} \sum_i \omega_{ij}} \quad (7)$$

with

$$\omega = |E(j) - E(i)| \quad (8)$$

and

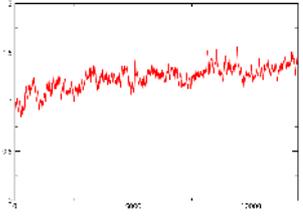
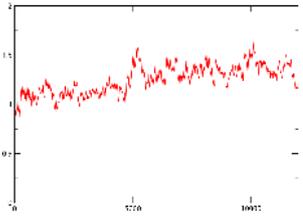
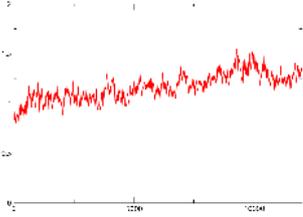
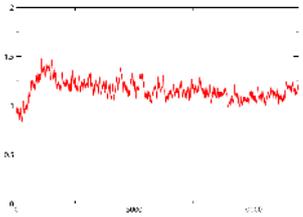
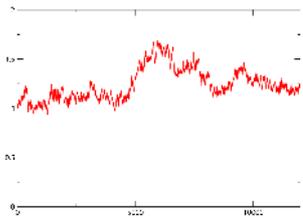
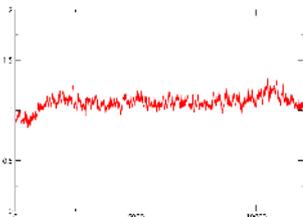
$$C_{ij} = \begin{cases} 1 & \text{if } [E(j) - E(i)] / [P(j) - P(i)] > 0 \\ -1 & \text{if } [E(j) - E(i)] / [P(j) - P(i)] < 0 \\ 0 & \text{if } [P(j) - P(i)] = 0 \end{cases} . \quad (9)$$

$P(i)$ is the binding energy in kcal/mol calculated from MM-PBSA assigned to ligand i and $E(i)$ is the corresponding binding free energy from the ITC experiment. A PI of 1 indicates that for any given pair of molecules, the predicted ranking of these molecules is correct. A PI of -1 indicates that the pairwise rankings are wrong in every case. A PI of 0 is expected for a series of random predictions.

Results and Discussion

MD trajectories

As known from previous MD studies¹¹³ of the system, AR is an enzyme with a compact globular fold that shows overall high structural integrity during the simulations. Table 13 summarizes the C α carbon root mean square deviation (rmsd) values of the different simulations used in this study. All values are measured against the energy-minimized starting crystal structure. Absolute rmsd values tend to be higher in the second part of the trajectories by a small amount and some of the plots in Table 13 show a slight drift during the simulation towards higher values. However, the mean C α movements are rather modest with none of them being larger than 1.4 Å for the second part of the trajectory. In addition, the accompanying standard deviations move towards slightly smaller values. This might indicate that the system has reached an equilibrated state with respect to the force field and simulation conditions, which does not necessarily correspond to the local minimum conformation of the starting crystal structure.

	C α RMSD [\AA]			Plot
	0-3 ns	3-6 ns	0-6 ns	
Holo (sorbiniil)	1.17 (± 0.12)	1.31 (± 0.08)	1.24 (± 0.12)	
Holo (tolrestat)	1.15 (± 0.13)	1.35 (± 0.09)	1.25 (± 0.15)	
Holo (idd 594)	1.07 (± 0.10)	1.26 (± 0.10)	1.16 (± 0.14)	
hAR sorbiniil	1.20 (± 0.12)	1.12 (± 0.07)	1.16 (± 0.10)	
hAR tolrestat	1.17 (± 0.17)	1.32 (± 0.13)	1.24 (± 0.17)	
hAR idd 594	1.05 (± 0.08)	1.10 (± 0.07)	1.07 (± 0.08)	

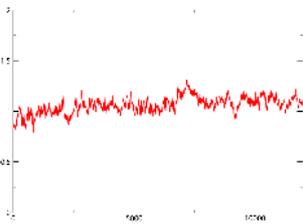
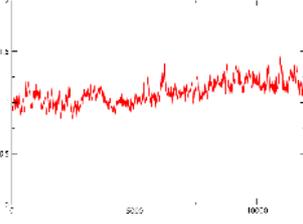
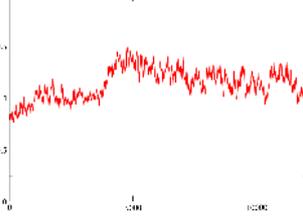
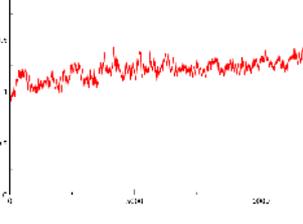
	C α RMSD [\AA]			
hAR Pfizer	1.02 (± 0.08)	1.11 (± 0.07)	10.6 (± 0.09)	
hAR zopolrestat	1.02 (± 0.08)	1.18 (± 0.09)	1.10 (± 0.12)	
hAR 47d	1.10 (± 0.18)	1.18 (± 0.10)	1.14 (± 0.15)	
hAR JFD	1.16 (± 0.10)	1.26 (± 0.07)	1.21 (± 0.10)	

Table 13: C α rmsd values for the MD trajectories used in the study measured against the corresponding energy minimized starting structures. Values are given for the first and second part of the trajectory as well as for the whole trajectory. The values in parenthesis indicate the standard deviations of the respective values. In the final column the time course of the C α rmsd is plotted: snapshots are plotted on the x-axis, the rmsd in \AA is indicated on the y-axis.

Sampling the reference state

To compare the binding free energy of ligands which bind to different binding-pocket conformations and in order to examine the energetics involved in the corresponding protein and ligand adaptation processes upon complex formation, it is essential to sufficiently sample the uncomplexed state of the protein and the ligands. For small molecule ligands it can be assumed that the conformational ensemble generated with a multi-nanosecond MD covers a sufficiently large phase space to provide thermodynamically meaningful results. For

proteins, however, this cannot be implied. The reliability of the derived receptor energies is of particular importance since it is used to derive the binding free energy for every protein-ligand complex under investigation.

To improve sampling of the uncomplexed state, three distinct simulations starting from different protein conformations were conducted. Representative structures for the three main binding-pocket conformations were chosen as templates: sorbinil, tolrestat, and idd 594. The sorbinil complex represents the binding pocket with the specificity pocket in a closed conformation; it shows only slight deviations compared to a holo structure deposited in the PDB (pdbcode: 1ads¹⁹). The tolrestat and idd 594 complexes represent the two main binding-pocket conformations with an open specificity pocket.

Removing the ligands from the corresponding complex structures *prior* to starting an MD simulation perturbs the system by different amounts, and the trajectories will certainly take a different time course. However, the energies derived from simulations with varying starting conformations should converge to a common value, given enough sampling time. The question is whether 6 ns sampling are sufficient to ensure this assumption. Before analyzing the calculated MM-PBSA energies it is instructive to examine whether or not there are significant differences in the sampled binding-pocket conformations regarding the three simulations.

Without performing a full structural analysis of the three MD simulations, one example shall be discussed here. The molecular dimension which indicates whether the specificity pocket is opened or closed is the distance between the side chains of Trp 111 and Leu 300. This distance is plotted for the three holo simulations in Figure 56. As can be seen from the green lines in the plots the starting values for the corresponding simulations differ as expected depending on whether the specificity pocket is open or closed. For major parts of the simulation time the distances are similar for all three simulations and correspond to the closed specificity pocket conformation. Therefore, one can conclude that most of the time the specificity pocket is indeed closed throughout the three simulations. However, differences can be seen between the simulations. Holo (sorbinil) shows hardly any variations of the distance. Possibly this arises from the fact that this structure was least perturbed upon removal of the ligand. Holo (tolrestat) shows stronger movements with respect to this dis-

tance. However, the only simulation which intermediately returns into the opened specificity pocket is holo (idd 594). Interestingly, these open conformations do not occur solely at the beginning of the simulation, but is rather revisited after a few nanoseconds half way through the simulation. These differences might already indicate that there is (still) a dependence of the sampling on the starting geometry even though it simply expresses as a local difference in this case. Possibly it indicates that such systems would require a much longer sampling time to achieve fully equilibrated state not getting trapped by local hysteresis effects. The question remains whether these differences are also reflected in the MM-PBSA derived energy values for the simulations.

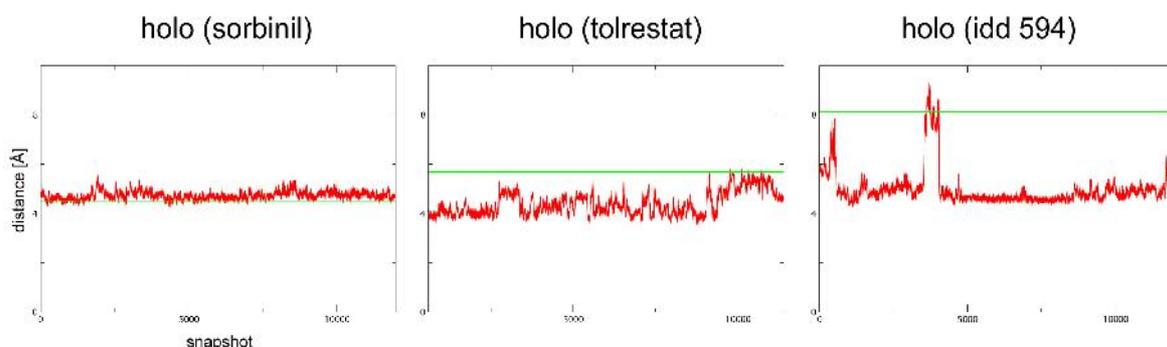


Figure 56: Distance between of Leu 300 CG and the Trp 111 ring center. The start values from the crystal structures are depicted as green line.

Table 14 lists the MM-PBSA energies of the holo trajectories derived for different starting geometries. In addition, the receptor energies derived by the single trajectory approach are given. To obtain these values, the ligand is removed from the single snapshots after the MD simulation has been carried out (see above). In the remainder of this study these trajectories will be called 'holo like' or receptor trajectories, since they do not represent a true sampling of the conformational space of the holo enzyme, but are treated as such for the MM-PBSA calculations in the single trajectory approach.

Analyzing the values for the first three nanoseconds, it becomes evident that for holo (sorbini) and holo (tolrestat) are equal in their energies whereas holo (idd 594) deviates by 12 kcal/mol from the two former ones. Also for the 'holo-like' receptor trajectories all simulations carried out with ligands addressing the specificity pocket (all apart from sorbini)

coincide in a window of 14 kcal/mol with respect to their receptor energies. The same holds true if one compares the values for the holo and the 'holo-like' trajectories. Only the receptor/sorbinil trajectory is off by more than 30 kcal/mol. This is especially puzzling since sorbinil does not address the specificity pocket. Therefore, in this case conformational differences concerning the binding pocket between receptor/sorbinil and holo (sorbinil) should be minimal. This observation appears alarming and possibly indicates that changes in the MM-PBSA energy caused by local changes of the binding pocket may be overwhelmed by larger deviations in other parts of the enzyme (for example loop movements). Looking at the second part of the simulations, it can be recognized that almost all energy values are lower than in the first part. Only the values for receptor/zopolrestat and receptor/sorbinil become more positive. Despite the surprising offset of the receptor/sorbinil energy in the first 3 ns with respect to the other receptor trajectories, this energy value matches with the remaining ones for the second half is in line with most of the other values. The energy scatter across the trajectories of the holo simulations increase to 33 kcal/mol. The energies of holo (sorbinil) are the most stable, since the second part differs only by 10 kcal/mol from the first, whereas for holo (idd 594) this difference amounts to 31 kcal/mol.

However, there is one aspect which should be addressed. In Table 14 the standard error of the mean is given for each energy as an estimate for the uncertainty in the mean values. In analogy to other studies²¹⁰, the standard error is calculated by dividing the standard deviation of the distribution by the number of data points in the evaluation (300 frames for the first and second part, 600 frames for the whole trajectory). Usually the standard deviations of the energy values are in the range of 50 kcal/mol for most of the simulations, thus the fluctuations of the energies obtained for single snapshots are larger than suggested by the standard error of the mean.

<i>System</i>	energy 0-3ns [kcal/mol]		energy 3-6ns [kcal/mol]		energy 0-6ns [kcal/mol]	
	mean	stde	mean	stde	mean	stde
Holo (sorbinil)	-8776	3.1	-8786	2.9	-8781	2.1
Holo (tolrestat)	-8776	3.0	-8803	2.9	-8789	2.1
Holo (idd 594)	-8788	3.1	-8819	2.8	-8804	2.2
Holo (mean)	-8780	1.8	-8803	1.7	-8791	1.3
Receptor/sorbinil	-8820	3.0	-8808	3.1	-8814	2.1
Receptor/tolrestat	-8778	3.0	-8801	2.8	-8789	2.1
Receptor/idd 594	-8787	3.3	-8806	2.9	-8797	2.2
Receptor/Pfizer compound	-8782	3.2	-8802	2.6	-8792	2.1
Receptor/zopolrestat	-8782	2.9	-8778	2.8	-8780	2.0
Receptor/47d	-8788	2.8	-8802	2.9	-8795	2.0
Receptor JFD	-8774	2.8	-8787	3.0	-8780	2.1

Table 14: MM-PBSA energy values derived for different MD trajectories without the entropic contribution. Total values for the whole trajectory as well as values for the first and second part of the simulation are presented. Furthermore the standard error of the mean is given which is defined as $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ where σ is the standard deviation and n is number of frames in the sample²⁴⁵. For the trajectories of Holo (sorbinil), Holo (tolrestat), and Holo (idd 594) the starting geometries were obtained by extracting the bound ligands from the corresponding complex structures before starting the MD simulations. The remaining receptor trajectories were generated by removing the ligand from the complex solely for MM-PBSA calculations, i.e. the ligands were present throughout the generation of the trajectories.

The question remains, whether the three holo simulations represent an adequate sampling which can be used for the MM-PBSA calculations. Considering the mean values for the entire trajectories (last column in Table 14) it can be seen that the maximum difference in energy between the three simulations is 23 kcal/mol. On a relative scale this corresponds to a deviation of 0.25%, which is low considering the deficiencies and differences in sampling already discussed. However, on an absolute scale 23 kcal/mol are more than twice the expected binding energies for the ligands and, therefore, non-negligible. Thus, the problem remains which energy value to select as reference state of the uncomplexed pro-

tein? Depending on this reference value, all calculated binding energies will be shifted accordingly. Hence, hardly any reasonable and meaningful results can be expected on an absolute scale for the separate-trajectories approach. To derive energies which are at least useful on a relative scale (i.e. for ranking), the energies derived from the three holo simulations were averaged and assumed to be representative for the reference state.

An idea which was first introduced by Gouda et al.²⁰⁴ is to estimate the conformational energy associated with the adaptation of the protein to the ligand. This is achieved by subtracting the receptor energy derived from the holo simulation from the energy of the corresponding holo-like simulation. In the study of Gouda et al. the derived energies were too high to be reasonable. Transferring this idea to the current data set, it becomes evident that the energies for the complex derived holo-like receptor trajectories are in many cases lower than the ones from the holo simulations. This would result in negative energies for the conformational changes, which appears quite unreasonable. However, this finding is in agreement with results from Gohlke et al.²¹⁰, who observed a similar behavior for the RAS-RAF protein-protein complex. The origin of this phenomenon is not understood, but supposedly it indicates problems and inaccuracies in the implicit solvent representation of trajectories generated with explicit solvent molecules.

MM-PBSA results

The results for the MM-PBSA calculations in terms of ranking the ligands according to their binding free energies are presented in Table 15. In this table the predictive indices (PI) for different MM-PBSA variants (single and separate trajectories) as well as different sampling times (energy minimized crystal structure, 200 ps sampling, 6 ns sampling) are given. Furthermore, the calculated ΔG value as well as two different subsets of MM-PBSA terms (all but the entropy term, vdW term only) are compared with respect to their predictive power for this dataset.

Overall, the obtained PI values for the present data set are rather disappointing. For many settings, the predictions are worse than random, as indicated by a negative PI. Usually, the best PI is to be expected for the calculations using the longest sampling (6 ns) and separate trajectories for complex, receptor, and ligand. However, the PI for these settings is 0.15, which is only marginally above a completely random prediction. Interestingly, this

value is in (coincidental) agreement with a PI of 0.16 which was obtained in the study of Pearlman²⁰³ for his dataset using similar settings: 5ns of sampling, no additional positional restraints on atoms near the binding site, and separate trajectories. If the results of the time-consuming normal mode calculations are left out, the PI decreases slightly to 0.09. However, it is questionable whether it is meaningful to discuss slight changes of a PI which basically indicates a random prediction. For the single-trajectories sampling approach the results are worse. For 6 ns of sampling the PI of -0.29 implies a prediction which is worse than random. Neglecting the entropy term leads to a PI of -0.44, again a change which is impossible to interpret.

Considering the PIs for 6 ns of sampling, the question arises what might be responsible for these disappointing results. Why has MM-PBSA been applied to a variety of systems with great success, but is not capable to predict the present dataset successfully? To obtain some ideas what might cause the failure of the method it is instructive to analyze the individual contributions to the MM-PBSA energy. These detailed results are presented for single-trajectory sampling in Table 16 and for separate-trajectory sampling in Table 17. Considering the calculated ΔG values in both tables, the computed variations between the distinct complexes are much larger than the experimentally determined values. The experimental deviations between strongest and weakest binder is 3.8 kcal/mol. However, this difference is 19 kcal/mol for the single trajectory variant and 35.6 kcal/mol for separate trajectories, respectively. Also the calculated absolute values are significantly off in many cases. For separate trajectories this has already been discussed in the previous section, but also for single trajectories this is observed. For both sampling approaches positive binding energies are obtained, suggesting that ligand binding parallels an endergonic process. Despite being rather puzzling, this behavior has been observed previously in other studies^{196;199;203}.

Regarding the individual energy terms, it can be noticed that the main driving force favorable for binding is the van der Waals term of the force field. It is compensated by the electrostatic forces, which are the summed contributions of the force field electrostatics and the electrostatic contribution to the solvation free energy computed via PB calculations. The suggestion that the van der Waals term would be the main driving force for

binding is in line with many previous studies^{194-200;204;210;227} where either the authors explicitly describe a similar behavior or it can be concluded from the provided tabulated values.

sampling time	Predictive Index				
	single snapshot	200 ps		6 ns	
comparison		single traj	separate traj	single traj	separate traj
ΔG_{exp} vs. $\Delta G_{\text{MM-PBSA}}$	-0.36	-0.71	-0.18	-0.29	0.15
ΔG_{exp} vs. $E_{\text{TOT, MM-PBSA}}$	0.17	-0.42	-0.13	-0.44	0.09
ΔG_{exp} vs. $E_{\text{vdW, MM-PBSA}}$	0.62	0.74	0.34	0.52	-0.34

Table 15: Predictive Indices for different comparisons. ΔG_{exp} is the experimental free energy of binding derived from ITC experiments. $\Delta G_{\text{MM-PBSA}}$ is the free energy of binding from MM-PBSA including all terms, $E_{\text{TOT, MM-PBSA}}$ is the energy calculated from MM-PBSA neglecting the entropic contribution, finally, $E_{\text{vdW, MM-PBSA}}$ is the energy from MM-PBSA using only the vdW force field term to estimate the binding energy. Two different sampling periods are considered: 6 ns and 200 ps as well as the variant where only the energy minimized starting structure (single snapshot) is used for the MM-PBSA calculations. Predictive indices are calculated using equations 7-9.

The overall unfavorable contribution of the electrostatic term result from the fact that the favorable force field electrostatics are overwhelmed by the unfavorable PB energies (which represent the interactions with the surrounding solvent). One explanation for this might be that for a system such as AR, where a charged ligand is desolvated upon complexation, the electrostatic interactions of the protein environment cannot compensate for the loss of the solvation shell – at least in the computational simulations.

Considering only the vdW contributions from Table 16 and Table 17, the PI for the single-trajectory cases rises to 0.52 if the entire 6 ns trajectories are evaluated. In case of separate trajectories, however, the prediction based the vdW term only is not nearly as good, but rather worse than random. How can such a behavior be explained? If using single trajectory sampling the conformations of complex, protein, and ligand are the same for every snapshot. Therefore, all vdW terms cancel mutually out, except for the additional contacts between ligand and protein which are newly experienced upon the complex formation. This cancellation is not necessarily given for separate trajectories, where complex, protein, and ligand are sampled independently.

	Sorbinil	Tolrestat	IDD594	Pfizer	Zopolrestat	47d	jfd
0-6ns							
ΔV_{mm}	-123.6	-102.0	-109.5	-118.0	-141.8	-133.4	-152.1
vdW	-30.5	-36.5	-42.2	-43.1	-50.3	-45.1	-37.1
Ele	15.7	23.9	23.4	18.4	31.3	24.4	14.4
$\Delta\Delta G_{solv}$	105.1	84.5	85.7	88.7	117.2	107.8	125.3
E_{tot}	-18.6	-17.5	-23.8	-29.3	-24.6	-25.5	-26.8
T Δ S	-17.1	-23.6	-18.4	-16.4	-15.2	-26.2	-17.1
ΔG	-1.4	6.1	-5.4	-12.9	-9.4	0.7	-9.8
Rank vdW	7	6	4	3	1	2	5
Rank E_{tot}	6	7	5	1	4	3	2
Rank ΔG	5	7	4	1	3	6	2
0-200ps							
ΔV_{mm}	-122.1	-100.9	-113.1	-113.2	-133.1	-144.1	-212.5
vdW	-31.6	-37.2	-42.2	-40.8	-49.4	-45.3	-36.7
Ele	16.5	24.8	25.8	19.4	33.4	27.0	18.1
$\Delta\Delta G_{solv}$	103.3	83.5	91.5	87.2	111.7	120.9	116.6
E_{tot}	-18.9	-17.4	-21.5	-26.0	-21.4	-23.2	-22.7
T Δ S	5.8	0.6	-39.6	-6.0	-9.3	-39.1	2.9
ΔG	-24.7	-18.0	18.1	-20.0	-12.2	15.9	-25.5
Rank vdW	7	5	3	4	1	2	6
Rank E_{tot}	6	7	4	1	5	2	3
Rank ΔG	2	4	7	3	5	6	1

Table 16: Summary of the results of single trajectory MM-PBSA calculations. All energies are given in kcal/mol. ΔV_{mm} : energy difference arising from molecular mechanics force field; vdW: binding energy contribution from force field van-der-Waals term; Ele: sum of force field and PB electrostatics; $\Delta\Delta G_{solv}$: binding free energy component from solvation free energy; E_{tot} : sum of ΔV_{mm} and $\Delta\Delta G_{solv}$; T Δ S solute entropic energy calculated via normal mode analysis, sum of E_{tot} and T Δ S. Rank E_{tot} , Rank vdW and Rank ΔG rank order of ligands according to the different energies.

	Sorbinil	Tolrestat	IDD594	Pfizer	Zopolrestat	47d	jfd
0-6ns							
ΔV_{mm}	-117.3	-105.0	-43.8	-169.7	-141.7	-130.8	-157.1
vdW	-60.2	-35.8	-54.9	-72.3	-65.9	-52.0	-67.1
Ele	31.9	34.7	39.2	40.3	67.4	40.4	52.3
$\Delta\Delta G_{solv}$	76.1	90.6	15.5	141.6	128.6	102.2	144.4
E_{tot}	-41.3	-14.4	-28.3	-28.1	-13.1	-28.6	-12.8
$T\Delta S$	-17.5	-26.3	-17.5	-28.3	-17.8	-23.5	-18.6
ΔG	-23.7	11.9	-10.8	0.2	4.7	-5.1	5.9
Rank vdW	4	7	5	1	3	6	2
Rank E_{tot}	1	5	3	4	6	2	7
Rank ΔG	1	7	2	4	5	3	6
200ps							
ΔV_{mm}	-238.0	-138.7	-185.8	-95.8	-240.8	-197.7	-249.7
vdW	-59.5	-38.4	-60.7	-73.0	-65.7	-63.5	-56.1
Ele	48.4	32.0	46.5	59.0	51.3	51.4	40.5
$\Delta\Delta G_{solv}$	200.1	138.4	173.2	81.0	211.8	171.8	145.6
E_{tot}	-38.0	-0.3	-12.6	-14.8	-28.9	-25.9	-31.0
$T\Delta S$	6.0	-1.5	-2.1	-1.8	-10.4	-28.2	-2.7
ΔG	-44.0	1.2	-10.5	-13.1	-18.5	2.4	-28.2
Rank vdW	5	7	4	1	2	3	6
Rank E_{tot}	1	7	6	5	3	4	2
Rank ΔG	1	6	5	4	3	7	2

Table 17: Summary of the results of separate trajectory MM-PBSA calculations using the averaged 18 ns trajectory as reference. All energies are given in kcal/mol. ΔV_{mm} : energy difference arising from molecular mechanics force field; vdW: binding energy contribution from force field van-der-Waals term; Ele: sum of force field and PB electrostatics; $\Delta\Delta G_{solv}$: binding free energy component from solvation free energy; E_{tot} : sum of ΔV_{mm} and $\Delta\Delta G_{solv}$; $T\Delta S$ solute entropic energy calculated via normal mode analysis, sum of E_{tot} and $T\Delta S$. Rank E_{tot} , Rank vdW and Rank ΔG rank order of ligands according to the different energies.

Compared to previous studies, however, the hypothesis that the vdW term alone is a good indicator for the binding free energy cannot be confirmed. Huo et al.¹⁹⁸ found for a dataset of seven peptido-mimetic inhibitors of Cathepsin-D with common scaffold a decrease of the correlation coefficient r from 0.98 to 0.16 once all terms except the vdW term are neglected. In the study of Wang et al.¹⁹⁷, a decrease in the PI from 0.79 to 0.22 for the

ranking of a series of closely related TIBO-type inhibitors for HIV-1 protease has been observed. For a congeneric series of seven biotin analogs plus HABA in the study from Kuhn et al.¹⁹⁶ one can see a decrease in predictive power going from $PI=0.99$ to $PI=0.56$.

In the present case, however, the deficiencies to capture the electrostatics of binding using the PB approach seems to be most problematic. Wang et al.¹⁹⁷ encountered a similar problem with electrostatics as they observed that for charged ligands the calculation of binding free energy does not produce correct results as for neutral species. The reason for this finding might be that the uncharged and charged form of a ligand molecules are separated by a large change in the computed solvation free energies. This arises the question, whether such effects can be suitably handled by the MM-PBSA approach.

As mentioned in the introduction, an important factor to obtain thermodynamically meaningful results is the sampling time. With respect to this aspect, previous MM-PBSA studies show large variations. The studies which successfully applied the method to medium or larger sized data sets of ligands for a protein usually used sampling times in the low hundreds picosecond range or even less (see SupplTable 1 and SupplTable 2 in Supplementary Material). Probably this was mostly due to practical reasons since longer sampling requires longer computer simulation, either for generating the trajectories or for performing MM-PBSA calculations. However, the performance of the method should increase with longer sampling times in two ways. On the one hand, more snapshots can be used for the MM-PBSA calculations, which improves the statistical significance of the derived energy values; on the other hand, the larger the generated ensemble is the more meaningful it becomes from a thermodynamic point of view. In order to assess the influence of different sampling times two additional simulation lengths are compared to the full-length 6 ns sampling: the first 200 ps of every trajectory and the single energy-minimized starting structure.

	Sorbinil	Tolrestat	IDD594	Pfizer	Zopolrestat	47d	jfd
ΔV_{mm}	-102.8	-95.1	-98.0	-109.3	-133.3	-131.1	-126.8
vdW	-34.3	-45.7	-52.7	-50.4	-60.4	-55.7	-46.2
Ele	22.7	38.0	38.4	29.3	40.7	34.5	31.5
$\Delta\Delta G_{solv}$	87.5	82.5	78.6	83.7	108.3	105.2	108.1
E_{tot}	-15.3	-12.6	-19.4	-25.6	-25.0	-25.9	-18.7
$T\Delta S$	18.4	-25.6	-32.0	-13.0	-31.6	-36.2	-24.9
ΔG	-33.7	13.0	12.6	-12.6	6.6	10.3	6.2
Rank vdW	7	6	3	4	1	2	5
Rank E_{tot}	6	7	4	2	3	1	5
Rank ΔG	1	7	6	2	4	5	3

Table 18: MM-PBSA energies derived from the energy minimized input structure. All energies are given in kcal/mol. ΔV_{mm} : energy difference arising from molecular mechanics force field; vdW: binding energy contribution from force field van-der-Vaals term; Ele: sum of force field and PB electrostatics; $\Delta\Delta G_{solv}$: binding free energy component from solvation free energy; E_{tot} : sum of ΔV_{mm} and $\Delta\Delta G_{solv}$; $T\Delta S$ solute entropic energy calculated via normal mode analysis, sum of E_{tot} and $T\Delta S$. Rank E_{tot} , Rank vdW and Rank ΔG rank order of ligands according to the different energies.

A period of 200 ps were chosen to represent the short sampling times usually applied in previous protein-ligand MM-PBSA studies of larger scale. The single energy minimized structure was included since Kuhn et al.²⁰¹ found that for their data sets almost no enhanced predictive power once going from single energy-minimized structures to ensembles generated via MD simulations. The results for the 200 ps sampling are included in Table 16 and Table 17, the corresponding values for using only the energy minimized starting structure are summarized in Table 18. The general trends with respect to the main driving forces for the free energy of binding remain similar for these two calculations. The vdW term is still the main source of binding free energy and electrostatics contribute unfavorably to the binding free energy. Considering the corresponding PIs in Table 15, the trend that solely the vdW term reflects best the binding free energies of the ligands is also indicated. Including the electrostatic terms decreases the performance significantly. The vdW term of the single energy-minimized structures alone, gives a good PI of 0.62. Sampling for a short 200 ps time increases the PI to 0.74, but going to the full 6ns sampling decreases the PI to 0.52. It seems that a single structure is sufficient to rank the ligands according to their

affinity, provided one correctly assumes that only the vdW terms have to be taken into account. Expanding the simulation beyond 200 ps does not increase performance under these assumptions, at least with respect to the vdW energies.

Assessing the reliability of the entropic contribution to the free energy of binding is difficult. Due to the low number of snapshots used for the calculations the statistical significance can hardly be evaluated. In the case of 6 ns of sampling using separate trajectories including the entropic term does improve the PI slightly (Table 15). For the single snapshots, however, including the entropic contribution decreases the prediction significantly.

As a summary, the best performance was achieved considering rather arbitrarily only the vdW term using either short sampling or only a single energy minimized structure.

One important point concerning the calculations of PI should be mentioned which is of importance to evaluate the results from this study. Due to the pairwise comparisons of all ligands and the weighting factor ω which accounts for the difference in the experimental binding free energy of each ligand pair, the predictive index is strongly dependent on ranking the extrema of the dataset correctly. Ranking the extrema incorrect means that all comparisons including these ligands fail. Additionally, since the experimental free energy difference between the ligand pair is in most cases larger if either the best or the worst ligand is included in the pairwise comparison these rankings are assigned to comparably high weighting factors. This is especially important for a small dataset of only seven ligands where many of the ligands are similar in their experimental binding free energies except for the two extrema. The overall devastating performance of the single-trajectory approach can partly be explained by this phenomenon. The Pfizer ligand, which appears on sixth rank according to the ITC experiments obtains the best binding energy by MM-PBSA, whereas the JFD compound which shows the lowest affinity in the ITC experiments is predicted second best binder by MM-PBSA (cf. Table 16).

However, the case of the Pfizer compound is interesting since according to the published IC_{50} this compound would be the tightest binder within in the dataset^{125;184}. Using published IC_{50} values and calculating a PI against the ITC data, one can end up with values between ~ 0.2 and ~ 0.5 for 6 ns of single-trajectory sampling depending on which of the

published IC_{50} values are taken. This emphasizes the need to take only IC_{50} into account once they are measured in a consistent fashion.

Summary and Conclusions

In the present study, MM-PBSA has been applied to an experimentally very well characterized data set of seven ligands binding to AR. For all protein-ligand complexes X-ray structures as well as high quality thermodynamic ITC data are available. The sampling times were set to 6 ns and all atoms of the system were allowed to move without any positional restraints. All parameters for the MDs were chosen to guarantee state-of-the-art simulations. Separate trajectories were sampled for the holo enzyme starting from three different crystal structures. Each trajectory remains stable over the whole simulation time as shown by the low C_{α} atom movements.

The ranking of the data set according to the calculated binding free energy, however, has not been successful. The experimental binding free energies correlate best with the calculated vdW term. Comparison of different sampling times did not show significant improvements with increasing computational efforts. The energy-minimized crystal structures provide results which in principle would not justify the enormous computational costs associated with the generation of the nanosecond trajectories and the calculations of the single MM-PBSA terms. Comparing $\Delta G_{\text{MM-PBSA}}$ for single and separate-trajectories sampling, the latter seems to reveal slightly better results, but on a significance level which is not really apart from a random prediction. In fact, single trajectory sampling is in most cases worse than a random prediction, indicating systematic errors which prevents the results from improving with increasing level of sophistication of the theory involved.

Despite being a theoretically more solid method compared to empirical scoring functions, the simplifications and assumptions inherent in the method apparently do not perform well in case of the present dataset of charged ligands binding to a protein performing 'induced-fit' binding adaptations. An additional challenge is the moderate difference in the experimental binding free energies of the ligands in the data set. Combining results from rather different approaches (force field, PB, normal mode analysis), which work surprisingly well for other datasets, does not produce reasonable results in the present case. Pre-

sumably many errors mutually cancel out, if similar ligands are compared which bind with similar binding modes to one protein, but do not compensate each other if the binding modes are as different as in the present case. It might well be that much longer sampling times are required in order to cover the differences in the thermodynamic ensembles of ligands binding to different protein conformers. However, with elongated sampling time, errors in the force field as well as in the PB and normal mode calculations reveal increasing importance.

Supposedly, one has to be careful in generalizing the current findings. MM-PBSA has been proven its value in many other studies and has been used to rank ligand data sets successfully. Nevertheless results from MM-PBSA calculations are still strongly dependent on the data set and the applied parameters and settings. Kuhn et al.²⁰¹, for example, obtained PIs between 0.15 and 0.86 for three different data sets using the GAFF force field and MD sampling, indicating a strong dependence on the data set under investigation. Pearlman²⁰³ presented in his study PIs ranging from -0.04 to 0.51 for the same data set using different sampling times and approaches. In contrast to other empirical methods, which are computationally cheaper, it is difficult to decide *a priori* whether a data set is suited for the method or which sampling approach is the most promising one. A significant amount of computational time has to be spent before one can estimate whether MM-PBSA is promising for a system under investigation. Automatizing the usage of MM-PBSA would be beneficial for the final users. Improving the results by modifying some of the parameters is also difficult and time consuming and only possible in post-modeling studies not in a predictive analysis. Adjusting the radii or the dielectric constants for the PB calculations might be promising concepts to improve results, but is probably out of scope for routine drug design projects.

In summary, MM-PBSA can be a valuable tool for a drug design project according to previous studies, however this study suggests difficulties with systems for which it should be the method of choice, i.e. a protein with a flexible binding pocket and a heterogeneous data set of ligands.

Paper	Kuhn et al. ²⁰¹	Gouda et al. ²⁰⁴	Kuhn 2000 ¹⁹⁶	Luo et al. ¹⁹⁵	Wang et al. ¹⁹⁷	Suenaga et al. ¹⁹⁹	Laitinen et al. ²⁰⁰
biological system	avidin/p38 map kinase	theophylline-binding RNA aptamer	avidin	EGF in complex with NRG-1 β	HIV-1 RT to TIBO/efavirenz	mod. peptides binding to Grb2	steroid binding to antiestradial antibody
starting geometries	manual modelling	NMR plus modelled ions	biotin derivatives modelled	homology model	docking	X-ray for protein, complexes modelled	17 β -estradiol-antibody X-ray, derivatives modelled
MD simulations							
program	Amber 7	Amber 6	Amber 5	Amber 7	Amber 5	Amber 6	Amber 6
force fields	Parm94/Gaff/MAB	Cornell et al.	Parm94	Parm99	Parm99	parm96	extended parm96
ligand charges	AM1-BCC	RESP	RESP	-	RESP	RESP	RESP
solvation model	TIP3P, 24 Å sphere a. BS	TIP3P, 11 Å box	TIP3P, 20 Å sphere a. lig.	TIP3P, 10 Å box,	20/30 Å diameter water cap at BS	TIP3P, 15 Å sphere around protein#	TIP3P, 25 Å cap around lig center
restraints	12 Å sphere	-	18 Å a. lig.	-	20/30 Å form lig. center	-	12 Å around ligand
nb cutoff	12 Å	?	dual (12/17 Å)	8 Å	9/12 Å	?	12 Å
electrostatics	nb cutoff	PME	nb cutoff	PME	nb cutoff	?	nb cutoff
equilibr. length	150 ps	75 ps	90-180 ps	40 ps	100/300 ps	700 ps	350 ps
prod. sim. lengt	50 ps	1 ns	300 ps	1100/900 ps	200 ps	300 ps	700 ps
MM-PBSA calculations							
single/separate trajectories	single	single	single	single	single	separate	single
atomic radii	n.a.	PARSE	PARSE	PARSE	PARSE	PARSE	PARSE
dielec. const. (solv/solute)	80/1	80/1	80/1	80/1	?	80/4	80/1
PBSA solver	MEAD	Delphi II	Delphi	Delphi	Delphi*	MEAD	Delphi
no of snapshots	10	50	50	100	100	30	50
time btw. snapshots (ps)	5	10	6	4	2	10	14
entropy snapshots	10	5	6	10	20	?	0
entropy restraints [Å] around lig.	6	-	-	-	10	?	-
Results							
exp. validation	yes	yes	yes	yes	yes	yes	yes

* only residues within 30 Å (40 Å) were used for PBSA calculations # no periodic boundary conditions were used

SupplTable 1: Overview over the MD and MM-PBSA settings used in literature studies. Data were extracted from the original papers. If specific data was not given in the original publication a “?” is put into the table.

Paper	Steinbrecher et al. ²⁰²	Srinivasan et al. ²⁰⁶	Massova et al. ¹⁹⁴	Huo et al. ¹⁹⁸	Pearlman D A ²⁰³	Gohlke et al. ²¹⁰	Beà et al. ²²⁷
biological system	hNeutrophil Elastase	DNA, RNA and modified DNA	p53 in complex MDM2	cathepsin D	p38 map kinase	RAS-RAF complex	cyclodextrin host-gues compl.
starting geometries	Docking	modelled	x-ray	modelling	modelling	X-ray, NMR	-
MD simulations							
program	Amber 8	Amber 4.1§	Amber	Amber	Amber 5	Amber 7	Amber 7
force fields	ff03/GAFF	parm94	parm94	Parm94/99/gaff	Cornell et al./Charmm Parm22	Cornell et al.	parm94/99
ligand charges	RESP	-	-	RESP	ESP	RESP	RESP
solvation model	TIP3P, 12 Å box	TIP3P	TIP3P, 10 Å box	TIP3P, 22 Å cap	TIP3P, 16 Å cap	TIP3P	TIP3P, 10 Å box
restraints	-	-	-	16 Å around ligand	12 Å, centered on ligand	-	-
nb cutoff	?	9 Å§	8 Å	12 Å	8 Å#	9 Å	9 Å
electrostatics	?	PME	PME	?	nb cutoff	PME	PME
equilibration length	200 ps	variable	40 ps	30 ps	variable (200ps-1.4ns)	2 ns	575 ps
production sim. length	2 ns	400ps-1.5ns	400 ps	450 ps	1ns or 5ns	10 ns	5 ns
MM-PBSA calculations							
single/sep. traj	? (single)	-	both	single	both	both	separate
atomic radii	?	PARSE	PARSE	PARSE	PARSE	PARSE/BONDI	?
dielec. const. (solv/solu)	?	80/1	80/1	80/1	80/1	80/1	?
PBSA solver	pbsa	Sanner algorithm	Delphi	Delphi	Delphi II	Delphi	pbsa
no of snapshots	200	100	100	300 (?)	100 or 500	500	500
time btw. snapshots (ps)	10	4-15 ps	4	1.5 ps (?)	10 ps	20 ps	10 ps
entropy snapshots		?	1 (X-ray structure)	?	10	50	51
entropy restraints [Å] around lig.		?	-	8 Å around ligand	12 Å centered on ligand	-	-
Results							
exp. validation		yes	yes	yes	yes	yes	yes

§ values taken from Cheatham et al., JMB²²³ # values taken from Pearlman et al, JMedChem²³¹

SupplTable 2: Overview over the MD and MM-PBSA settings used in literature studies (continued). Data were extracted from the original papers. If specific data was not given in the original publication a “?” is put into the table.

Expect the Unexpected while working with AR

Introduction

During the last four decades protein crystallography emerged as an essential tool to elucidate the binding geometry and interaction patterns experienced in protein-protein or protein-small molecule ligand complexes. Especially as prerequisite for drug design accurate crystal structure determinations are of utmost importance^{138;139}. Novel leads are either discovered experimentally by screening existing compound libraries (high-throughput screening) or computationally by screening virtual libraries against a given target²⁴⁶. For the latter strategy, X-ray crystallography is essential to provide the structure of the target protein and to analyze its characteristics with respect to ligand binding.

However, as the following study will show, even in the case of successful crystal structure determination a unique and definite answer with respect to the binding mode resembling the most relevant in-vivo situation is not necessarily given. One important explanation for this complexity results from protein flexibility with respect to either side-chain or main-chain atoms localized in loop regions. On the one hand adaptability is an important feature and prerequisite for biological function^{98;139;246}, however, on the other hand, it provides special challenge for inhibitor design strategies as flexibility complicates a reliable prediction of the binding geometry adopted by potential small molecule leads²⁴⁷. Even though atomic B-factors give some evidence for local mobility, and, when refined anisotropically, information about its directionality, it has often been discussed whether crystal structures only represent a static frozen-in picture of proteins, thereby underestimating their dynamic nature^{138;248}.

In the first part of this study, the dynamic properties of a protein-ligand complex are analyzed using a molecular dynamics simulation based on the binding geometry observed in a crystal structure determined by Holger Steuber in our laboratory. Characteristic changes of the binding geometry, observed by MD simulations, prompted us to collect multiple data sets of crystals obtained by different soaking or cocrystallization conditions in order to investigate whether the MD-indicated flexibility is reflected by different states in crystalline phase.

By doing so, distinct protein-ligand complexes were obtained, which emphasize the agreement between experimentally observed and computationally predicted adaptability of the investigated binding pocket.

An intricate interplay between flexibility of some amino-acid side-chains and cooperative packing effects caused by symmetry-equivalent molecules enable binding of more than one ordered ligand within and near the binding pocket. Such unexpected binding mode observed by X-ray crystallography is reported in the second part of this study.

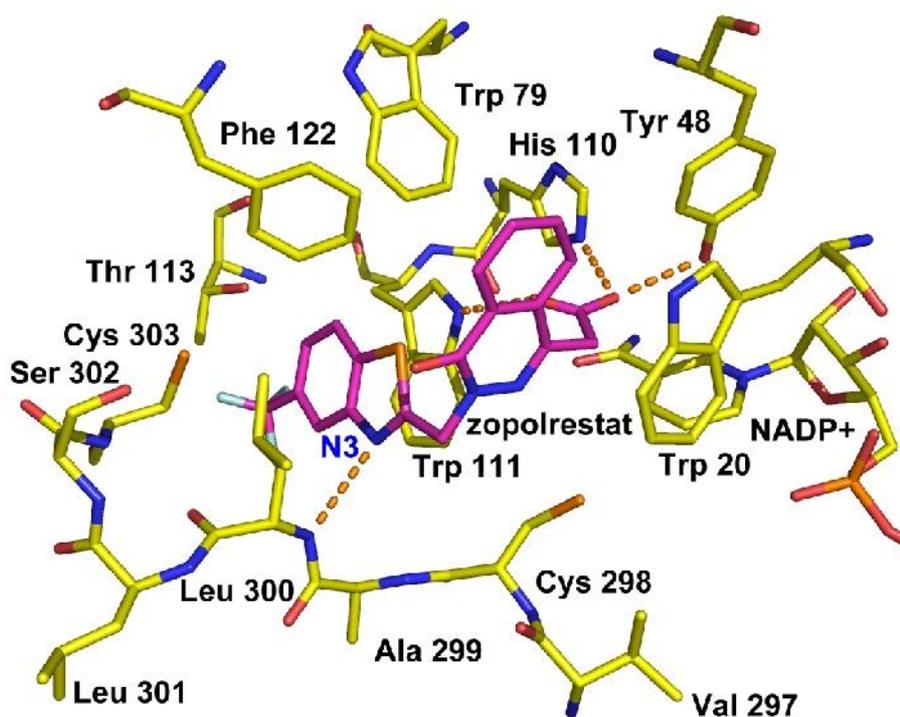


Figure 57: Refinement model of the crystal structure obtained after one day soaking of the ligand zopolrestat into AR crystals (1d). Residues comprising the binding pocket are shown as sticks colored in yellow, the ligand is shown in magenta. H-bond interactions are represented in orange. The benzothiazole moiety penetrates the specificity pocket formed by Trp 111, Phe 122, and Leu 300. The phthalazinone system intercalates between Trp 20, Trp 79, and Phe 122, whereas the carboxylic head group penetrates into the catalytic cleft. Note the formation of the H-bond donated by Leu 300 NH to the ligand's N3.

Results and Discussion

For the first part of this study human AR is complexed with zopolrestat (Figure 58, **1**), a potent carboxylic-acid type inhibitor formerly investigated in clinical trials. The structure was obtained from preformed crystals after they have been soaked for one day with the ligand (**1d**, Figure 57). It was refined to a resolution of 1.48 Å and shows the following binding mode: The benzothiazole moiety of the inhibitor occupies the specificity pocket by penetrating between Leu 300 and Trp 111, whereas its phthalazinone part intercalates with Phe 122, Trp 111 and Trp 20.

Remarkably, Trp 20 Nε1 forms a contact of 3.1 Å to the 5'-carbon atom of the phthalazinone system which is exceptionally short for a hydrophobic van der Waals contact. The carboxyl group of zopolrestat is probably deprotonated and forms electrostatic interactions to the positively charged nicotinamide moiety of the cofactor. Furthermore, it participates in a charge-assisted H-bond network to Tyr 48, His 110, and Trp 111. Importantly for the following, the nitrogen atom of the benzothiazole moiety accepts an H-bond from the

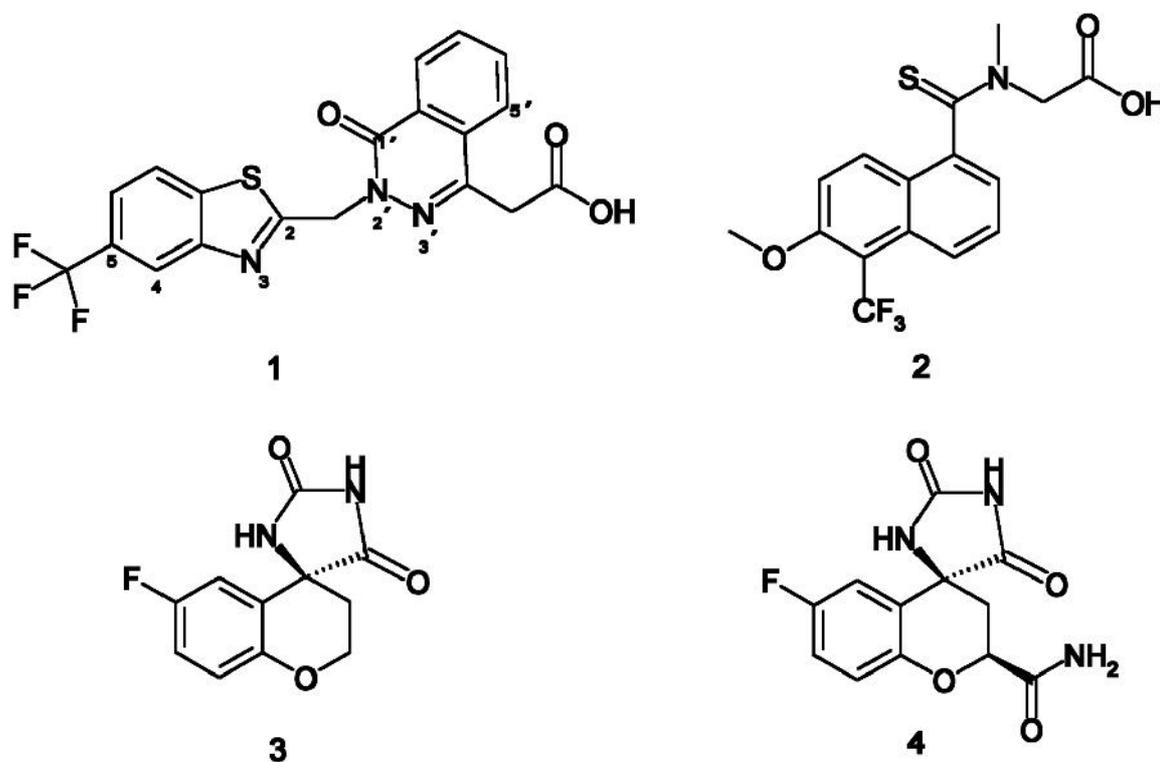


Figure 58: Chemical formulae of the aldose reductase inhibitors zopolrestat (**1**), tolrestat (**2**), sorbinil (**3**), and fidarestat (**4**).

backbone NH group of Leu 300 facilitated by an appropriate orientation of the amide bond linking Ala 299 to Leu 300. The electron density observed for **1d** gave no evidence for any split conformation relevant for the binding site residues.

A similar binding geometry for the same complex obtained by cocrystallisation was described by Wilson et al.¹⁷⁹. However neither structure factors nor complete protein coordinates, but rather the C α trace are deposited with pdb entry 1mar.

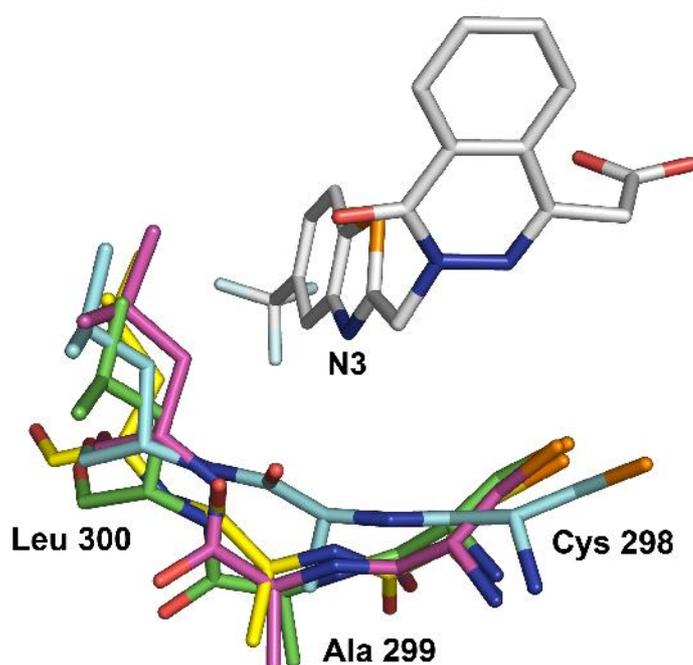


Figure 59: Selected conformational snapshots obtained from the MD simulation of the AR zopolrestat complex are represented for the residues Cys 298, Ala 299, and Leu 300. These snapshots suggest enhanced mobility in this region: while the conformations shown in green or magenta enable H-bond formation to the ligand's N3, this H-bond is ruptured in the conformations coloured in light blue or yellow. The inhibitor as observed in co_10 is represented as grey sticks after superimposing this crystal structure with the MD snapshots using a Ca-fit.

However, is this binding geometry observed as frozen-in situation in our crystal structure also relevant under conditions that allow to evolve dynamic behavior? In order to ob-

tain some insights into the latter properties, MD simulations of the described complex were carried out.

Analysis of rms deviations to detect families of conformations together with the visual inspection of the corresponding frames (Figure 59) clearly indicate the presence of several distinct states differing in their orientation of the loop residues Ala 299 and Leu 300.

Conformation 1 resembles the geometry observed in the crystal structure which served as starting template. It shows the described H-bond between Leu 300 NH and the inhibitor's N3. However, conformation 2 displays a situation where this H-bond is ruptured as a result of a flip in the backbone of Ala 299. Thus, a conformation is obtained which is similar to one observed in the ultra-high resolution crystal structure of human AR in complex with the inhibitor IDD 594¹¹⁴.

Additionally, Figure 59 represents two intermediates between the conformations 1 and 2. The dominant conformation throughout the MD is the one where no hydrogen bond between Leu 300 backbone and the inhibitor is formed. Nevertheless, is this indicated mobility an artefact of our MD simulation or does it suggest enhanced flexibility in this region?

A single crystal structure can hardly provide evidence for such behaviour. Accordingly, we were stimulated to perform further crystallisation experiments. In particular, the influence of different crystallization conditions (soaking period and cocrystallisation) was addressed. In total nine independent data sets for crystal structures were collected.

After soaking for two hours, no appropriate electron density for zopolrestat could be detected in the Fo-Fc difference map. Instead, a citrate molecule picked up from the crystallisation buffer occupied the active site. The specificity pocket is closed, as indicated by short van der Waals contacts between the side-chains of Leu 300 and Trp 111. The backbone carbonyl group of Ala 299 likely accepts an H-bond from Tyr 309 OH through a 2.5 Å distance. The corresponding NH group of Leu 300 points towards the binding pocket and donates an H-bond to the solvent-exposed carboxylate of the citrate. As several AR citrate complexes have already been described and deposited in the PDB, this citrate complex is not further described here.

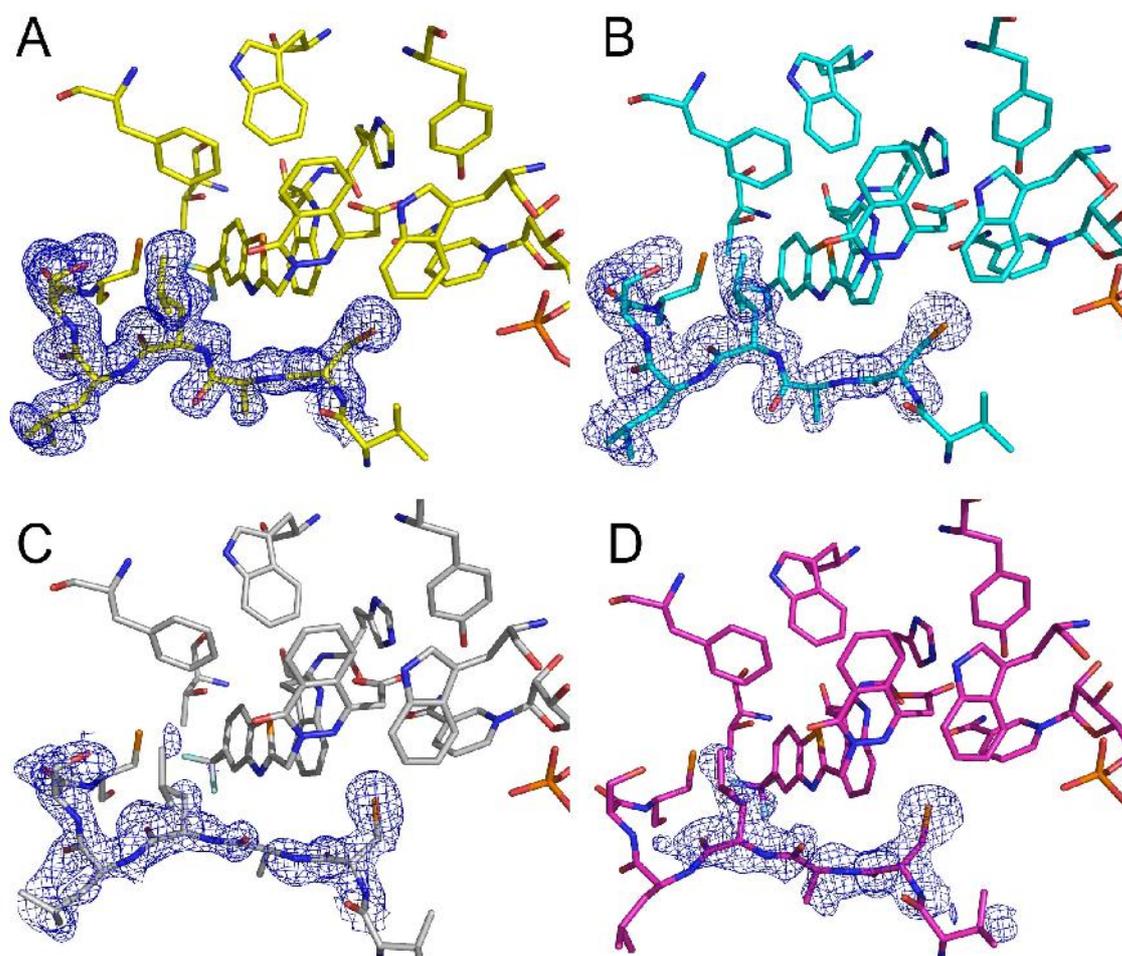


Figure 60: Representation of the binding pocket occupied by the inhibitor zopolrestat. Fo-Fc difference densities observed after excluding the corresponding residues from the model are shown contoured at 2.0σ A) The electron density shows the conformation of the loop residues of 1d. B) The electron density observed for 3d_a confirms the loop conformation in Fig. 1 and Fig. 4a. c) Representation of the loop conformation found in a third crystal soaked for three days (3d_c). A backbone flip of the Ala 299-Leu 300 peptide bond accompanied by a rupture of the H-bond to the ligand is observed in this crystal structure. Additionally, a gap in the electron density within the backbone suggests enhanced mobility. d) Binding pocket representation obtained from crystals soaked for six days (6d_a). The difference density is poorly defined for some parts of the C-terminal loop, but it clearly shows a flip similar to the conformation found in 3d_c as well as rupture of the density near Ala 299.

Expanding the soaking period to three or even six days introduced remarkable signs of flexibility and even conformational changes. This behaviour is in agreement with the observations made in the MD simulations.

Two of three crystal structures based on data collected after a soaking period of three days (**3d_a**, **3d_b**) exhibit a break of the $F_o - F_c$ electron density at $> 2.8 \sigma$ level between the $C\alpha$ -atom and the carbonyl group of Ala 299, thereby providing some evidence for extended flexibility present in this region.

However, the conformations found in these two structures (Figure 60 b) resembles the one observed after short soaking (**1d**), thus, the H-bond donated by the Leu 300 amide group to N3 of the ligand is maintained.

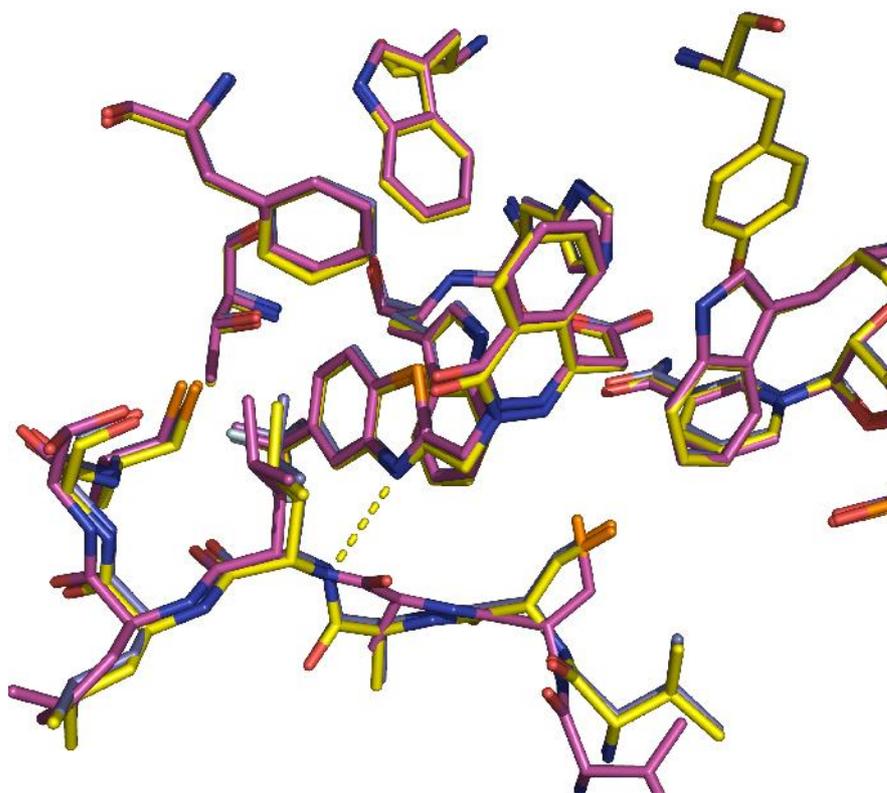


Figure 61: Superposition of the refinement models obtained for co_10 (blue), 6d_a (magenta) and 1d (yellow). This representation clearly shows 1d and co_10 being virtually identical, whereas in 6d_a the C-terminal loop adopts a different conformation.

Surprisingly, in the third crystal structure obtained from a crystal soaked for three days (**3d_c**) a backbone flip of the amide bond next to Ala 299, accordingly accompanied by a rupture of the respective H-bond to the ligand is observed.

To assess the relevance of our refinement model, the Fo-Fc difference electron density was examined as obtained from excluding the residues 298-302 from the model (Figure 60 c). Conformational changes and a possible distribution over multiple conformational states is suggested by the poorly defined electron density next to Ala 299 in **3d_c**. No significant electron density is observed for Leu 300 C γ , C δ , and C ϵ , indicating high flexibility for this side-chain.

Both crystal structures determined after a soaking period of six days (**6d_a** and **6d_b**) showed a similar binding geometry as **3d_c**, confirming the occurrence of the backbone flip observed for Ala 299. Whereas the Ala 299 ϕ angle is nearly unchanged in all structures ($-167.8 \pm 6.6^\circ$), the ψ angle is rotated by 120° from $66.7 \pm 2.5^\circ$ in **1d**, **3d_a**, and **3d_b** to $-175.2 \pm 6.9^\circ$ in **3d_c**, **6d_a**, and **6d_b**. Hence, the flip of the Ala 299 carbonyl group is facilitated,

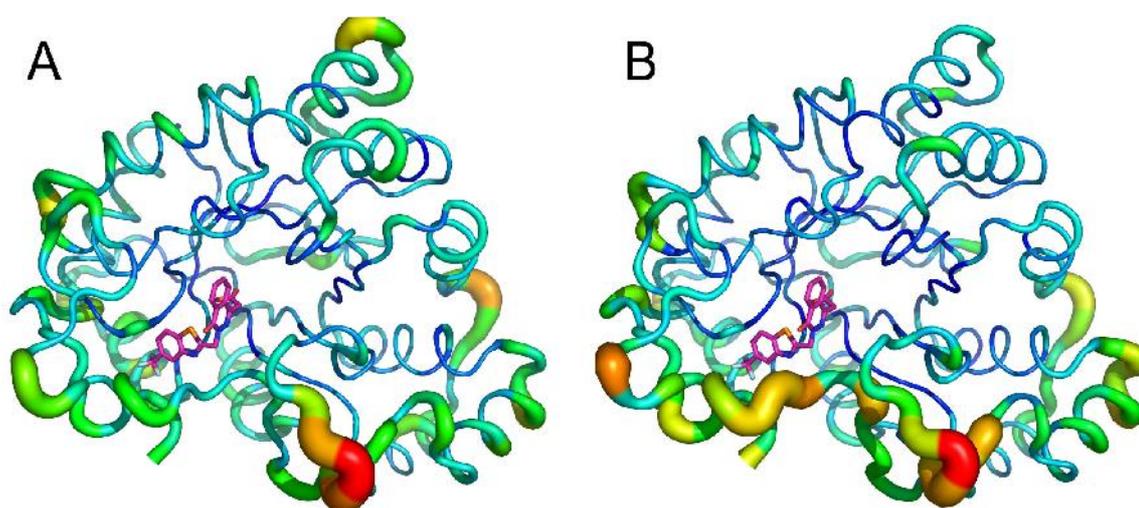
In order to examine which of the observed binding geometries is obtained carrying out cocrystallisation experiments, two additional structures of the same complex were determined. One set, **co_1**, was collected using a crystal grown within one day, and a second data set, **co_10**, after crystals were exposed for ten days to the mother liquid. In contrast to long-term soaking, for both structures obtained by cocrystallisation the electron density clearly indicates the conformation resembling that in **1d**, without any ill-defined electron density next to Ala 299.

As summary, shown in Figure 61, long-term ligand soaking of preformed crystals induces conformational changes which are not observed after short-term soaking only or *via* cocrystallisation.

Figure 62 visualises these changes by representing the TIM-barrel as tube diagram coloured by B-factors. In particular, within the ligand-binding C-terminal loop a remarkable increase of the B-factors is observed when comparing, e.g., the crystal structures **co_10** and **6d_a**. Despite these representations are not based on normalized B-factors, the Wilson

B-factor for the **co_10** and **6d_a** data sets are 17.7 and 16.9 Å², respectively. This suggests that B-factors derived from model refinement are to some degree comparable.

To evaluate whether the enhanced mobility observed by the MD simulation is dependent on the corresponding starting conformation, a second MD run was performed. The **3d_c** structure was used as initial model, where the H-bond between the inhibitor's N3 and Leu 300 NH is not formed. Under these conditions, the clusters representing the conformation missing this hydrogen bond are higher populated. However, the backbone flip, which triggers the H-bond formation is also observed. Thus, the enhanced flexibility in this region, suggested by the MD simulation, is not dependent on the starting geometry.



*Figure 62: TIM-barrel of AR represented as tube, emphasizing the local mobility with respect to the refined B-factors. The tube is colored by B-factor: blue regions correspond to low temperature factors, whereas green, yellow and red color characterize regions of subsequently increasing B-factor. Additionally, gain of temperature factor is represented by increasing radius of the tube. The inhibitor zopolrestat is shown in magenta. **A** represents the corresponding tube representation for **co_10**, in **B**) the one for **6d_a** is given. Note the remarkable gain of local mobility within the C-terminal loop region lining the ligand binding pocket observed in **6d_a** compared to **co_10**. These comparison suggests that extended soaking periods provoke increasing mobility with respect to distinct regions represented by higher B-factors.*

In the second part of this investigation, the incorporation of several ligand molecules into the same complex depending on the crystallization conditions will be reported. The binding of the additional ligands is accompanied by cooperative changes of residues participating in crystal packing. The corresponding crystal structures have been determined as part of this thesis (Appendix B).

The structure of tolrestat (Figure 58, 2) has been determined previously in complex with porcine AR (pdbcode: 1ah3) to a resolution of 2.30 Å by Urzhumtsev et al.¹²⁰ in 1997 (Figure 63).

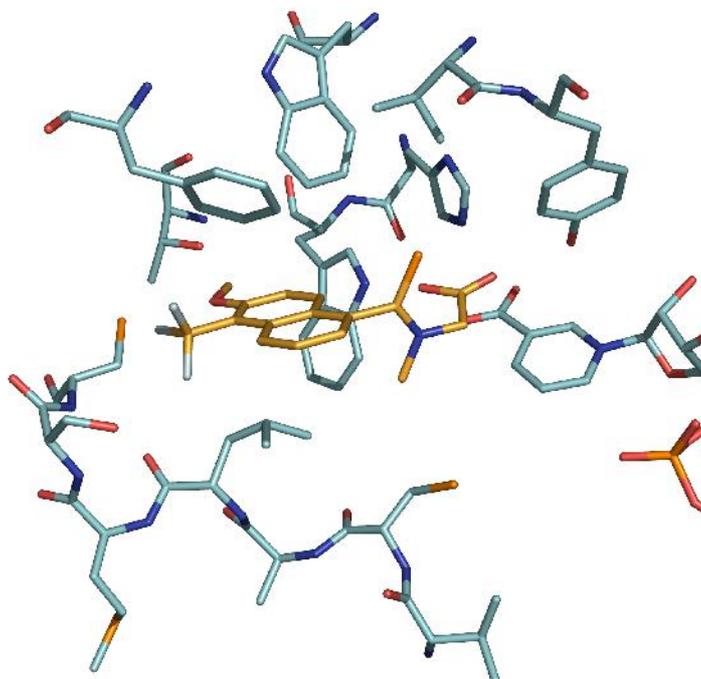


Figure 63: Binding mode of tolrestat in pig AR as described in literature¹²⁰. The carboxylic-acid head group occupies the catalytic pocket, whereas the naphthyl moiety forms hydrophobic contacts to Trp 111, Phe 122, and Leu 300. Note the kinked conformation of Leu 300 allowing the aromatic system of the ligand to adopt an edge-to-face orientation.

In order to elucidate the crystal structure with the corresponding human enzyme, crystals of the complex were obtained by soaking for three days. The crystal structure was refined to 1.48 Å. Data collection and refinement statistics are given in Appendix B (page 197). The overall structure is very similar to the complex with porcine AR, showing a Cα rmsd of 0.37 Å.

Surprisingly, the final structure comprises four bound ligand molecules in and next to the binding pocket (Figure 64). One ligand, L1 adopts the same binding mode as the single ligand in the deposited pig aldose reductase structure (0.1 Å rmsd).

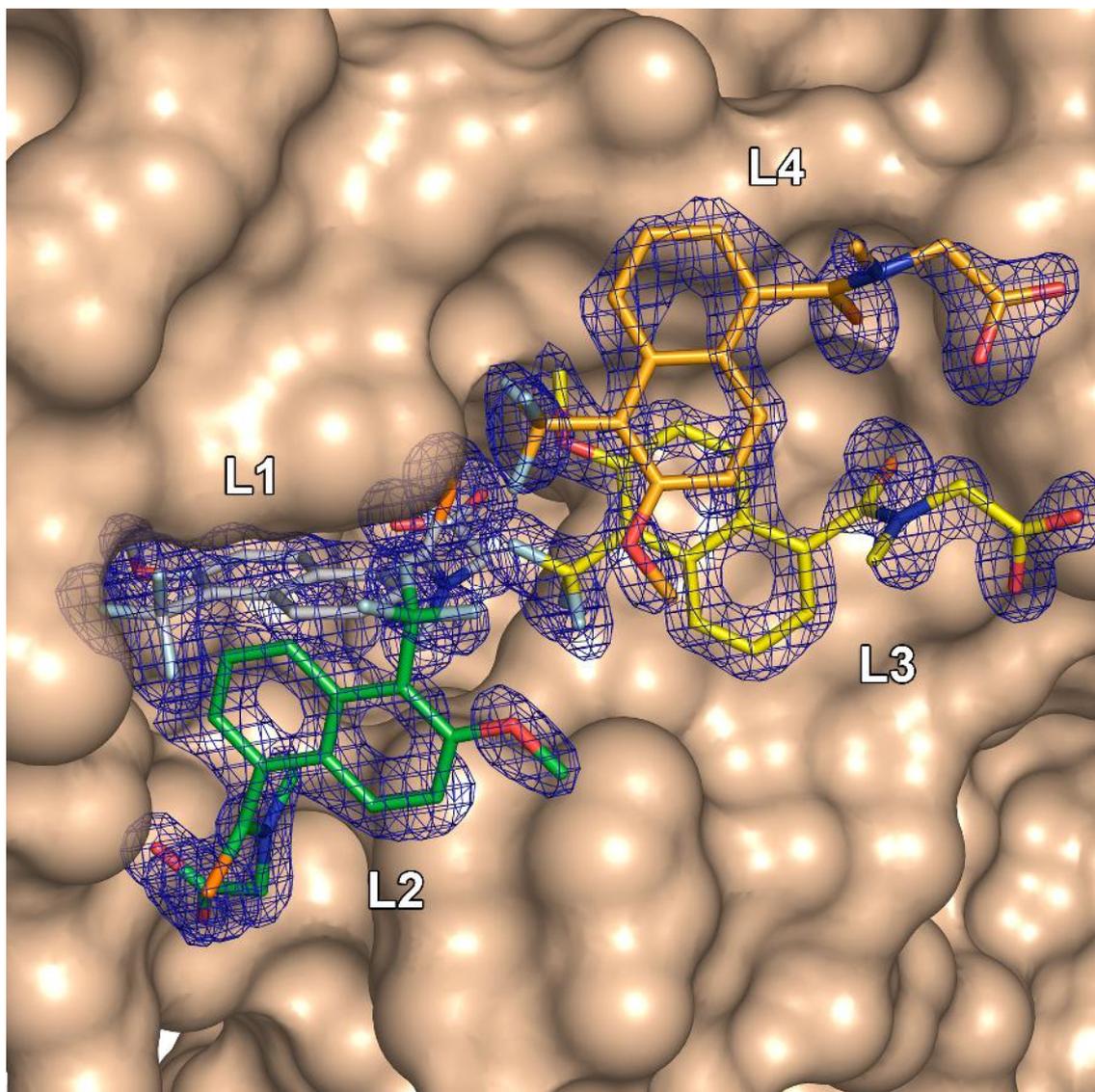


Figure 64: Representation of ALR2 in complex with four tolrestat molecules, Fo-Fc densities are colored in blue and contoured at 1.5σ . Ligand L1 occupies the binding pocket as observed in the 1:1 complex with the enzyme. However, in the present complex, a second ligand, L2 is placed in front of the binding site of L1, forming an edge-to-face interaction with each other. Two additional tolrestat molecules, L3 and L4, occupy a binding cleft formed by crystal contacts.

A second tolrestat molecule, L2, is placed in front of the binding pocket perpendicular to L1. The carboxylate group of this molecule forms two hydrogen bonds with the backbone nitrogen atoms of Leu 301 and Ser 302. The naphthalin moiety adopts T-shaped hydrophobic π - π stacking geometry with respect to the corresponding part of ligand L1.

The two additional ligands L3 and L4 are placed in a small pocket created by crystal contacts. They form hydrophobic stacking interactions with each other, most likely further stabilizing their binding mode. The carboxylate moiety of L3 forms a salt bridge to Lys 194 of a symmetry equivalent molecule, while the carboxylate group of L4 accepts an H-bond from Asn 292 ND2 of the same symmetry equivalent molecule.

In order to further analyze how the protein adapts to the additional ligands, two further datasets using crystals with soaking times of two hours and four days were collected, keeping all other parameters constant. The long-term soaked crystal was measured at a synchrotron beamline to improve data quality and resolution of the putative multi-ligand complex structure. As expected the structure determined from crystals of a short-term soak comprises only one ligand. It is placed in the binding pocket as already found in singly complexed tolrestat crystal structures (Figure 63). Surprisingly, also the crystal structure derived from the four-day soak reveals only one bound ligand, thus leaving the four ligand-bound structure as an unique observation.

The most striking difference between the crystals is a change in the length of the b axis of the unit cell of about 2 Å, probably one feature resulting from the incorporation of the additional ligands. Screening for further crystals with a similar expansion of the cell dimensions was not successful as suggested by data collected from five other crystals soaked for three days.

A close comparison of the one and four ligand-bound structure reveals how the additional ligands are accommodated in a cleft formed between the protein molecules (Figure 65). The interactions experienced between the ligands and the protein (see above) are accompanied by additional changes.

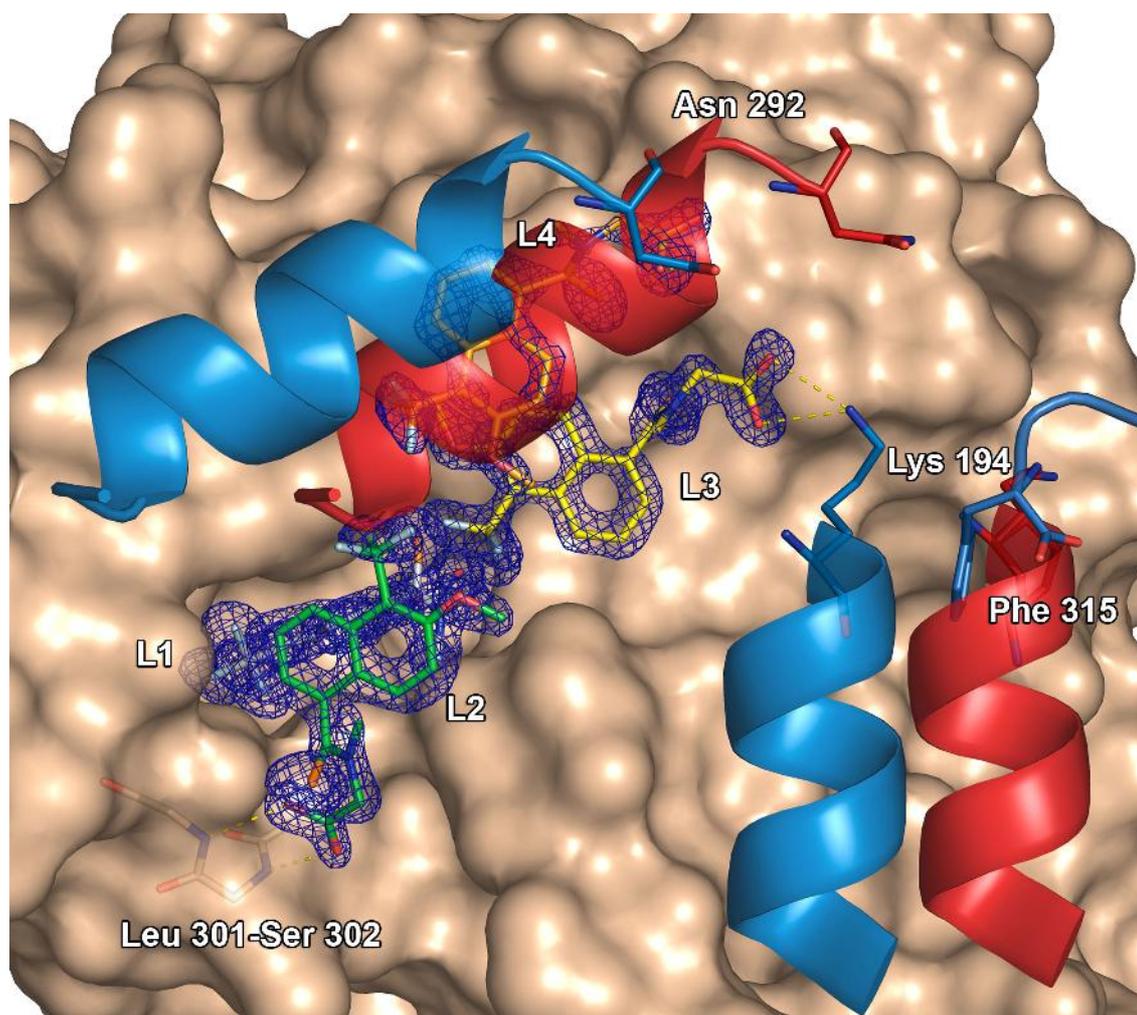


Figure 65: Interactions of the additional ligands with ALR2 and its symmetry related molecules. Fo-Fc densities of the inhibitor molecules are coloured in blue and contoured at 1.5σ . The carboxylate moiety of L3 forms a salt bridge to Lys 194 of one symmetry equivalent, while the carboxylate group of L4 accepts an H-bond from Asn 292 ND2 of this symmetry equivalent. The adoption of the additional ligands requires significant changes of the protein structure: The helix formed by the residues Ser 282 - Tyr 291 performs a Ca shift of about 7.0 \AA RMSD away from the binding site, whereas the helix Lys 195 - Lys 203 moves 6.2 \AA towards L3. For comparison these helices representing the conformation observed in the four-ligand complex are colored in blue, whereas the spatial location of these helices known from the one-ligand structure is shown by red ribbons.

The C-terminal residues Glu 313 – Phe 315, in the one-ligand structure significantly disordered and difficult to localize in electron density maps, become better ordered upon in-

corporation of the additional ligands. The C-terminal carboxylate oxygens at Phe 315 form an interaction with the backbone nitrogen of His 163, while the phenyl moiety is involved in a π - π stacking interaction with the imidazole moiety of the latter residue. The residues which directly interact with L3 and L4 adopt different conformations to facilitate these interactions.

In order to expand the pocket to accommodate L3 and L4 the helix formed by the residues Ser 282 – Tyr 291 performs a C α shift of about 7.0 Å rmsd. The spatial location of this helix adopted in the one-ligand structure would clash with both, L3 and L4.

Additionally, the helix comprised of residues Lys 194 to Lys 202 moves 6.2 Å towards L3 in order to facilitate the salt bridge between Lys 194 and the carboxylate group of L3. During this process the otherwise disordered residue Lys 194 becomes ordered. The space created by the motion of the helix is occupied by the C-terminal residue Phe 315, which also becomes ordered (see above).

These features clearly indicate that remarkable conformational changes apart from the induced-fit adaptations of the binding pocket take place in order to bind the additional ligands. Even though the occurrence of two ligands, both located within the binding pocket, have been reported for a double mutant of AR²⁴⁹, such a phenomenon as described in this study was never observed before in AR crystals.

Conclusions

Several conclusions can be drawn from our studies. It is widely assumed that the same binding mode is found for a protein-ligand complex crystallized from the same mother liquid under constant conditions.

This immediately provokes the question under which conditions the most relevant geometry resembling the in-vivo situation may be obtained in order to develop meaningful ligand design hypotheses. Even though an explanation for the results of this study cannot be given, it provides a caveat to structural studies. Usually, crystals are grown under conditions that turn out to be successful. The best diffracting crystal is subjected to data collection and the data are evaluated. Once the structure has been determined successfully, usual-

ly no further analysis of this structure in dependence of the crystal growth conditions is performed. This is a quite acceptable strategy, in particular, since the systematic re-evaluation of a structure can be rather tedious.

For the present case study, in total 17 complete data sets had to be collected, 12 of those were subsequently refined. However, as indicated, the relevance of the resulting structures with respect to drug design can be different. Furthermore, such multiple structure determinations can indicate the intrinsic flexibility of some portions of a protein that are in particular influenced and modified upon ligand binding. The present multiple structure determinations have been combined with MD simulations to generate a conclusive picture among the individual frozen-in snapshots seen in the different crystal structures.

Appropriate handling of protein flexibility in structure-based drug design remains one of the major challenges. Even though a generally applicable protocol is not yet available, this study suggests that more relevant hypotheses can be produced if crystal growth and protein-ligand complex generation is performed under varying conditions and crystal structure analysis is combined with MD simulations. The latter is particularly important if only one crystal structure is available.

Visual inspection and thorough comparison of multiple crystal structures and NMR studies complemented by the analysis of frames generated by MD simulations might help to estimate, in which regions enhanced protein mobility can affect ligand binding.

Summary and Outlook

Protein flexibility is an important, but still barely understood phenomenon. It is recognized as a crucial factor to be included in any design of small-molecule protein inhibitors⁹⁷⁻⁹⁹.

In the first part of this work the relevance of considering protein flexibility in structure-based drug design was highlighted. Docking experiments were carried out for two inhibitors which were originally designed¹¹⁸ to mimic the unique binding mode of the AR inhibitor tolrestat^{119;120}. Instead of using only the complex structure of tolrestat as template, the three known binding-pocket conformations of AR were considered as putative pocket conformations to be addressed by the ligands.

Separate docking simulations were carried out for each combination of pocket and ligand using AutoDock 3.0¹²² with its default scoring function. For both ligands the obtained energy scores indicated that the tolrestat binding pocket was unlikely to be the preferred pocket conformation. Instead, AutoDock suggested that the ligands adopt a binding mode similar to the inhibitor sorbinil^{120;240} where the specificity pocket is in a closed conformation.

After the crystal structures became available this hypothesis could be verified. In both X-ray structures the specificity pocket was closed and the inhibitors adopted binding modes which are most compatible with the one of sorbinil.

However, the predictions were by no means perfect. One of the compounds induced a conformational change in the catalytic sub-site of the AR binding pocket. Previously this part of the enzyme had always been regarded as being relatively rigid. Since no crystal structure appropriate as template for the docking experiments was available at the time they were carried out, there was no chance to correctly predict the exact binding mode of this compound. The other compound did not induce an unexpected conformational change. However, with this molecule the importance of considering water during docking could be emphasized. The region of the binding pocket where AutoDock placed one hydrophilic side chain of the inhibitor was occupied by three water molecules in the corresponding crystal structure.

Thus, it could be shown that, despite the limitations of the available methods, it is recommendable to include multiple pocket conformations into the docking process.

In the second part of this thesis a new method to simplify the tedious process of docking to multiple targets was evaluated in the context of protein flexibility: in-situ cross-docking¹²⁴. With this method, instead of performing sequential docking experiments of multiple ligands into multiple protein structures, several protein conformations can be addressed in one docking run. Hence, the number of docking experiments is reduced significantly and the time needed for setting up the docking runs, carrying out the actual calculations, and analyzing the results is shortened. If docking a large compound library to a set of multiple protein conformations this can significantly save computing time.

During the docking process with AutoDock the protein is represented by a set of grids with preassigned energy values. These grids are derived by placing probe atoms on each intersection of a regularly spaced grid and calculating the interaction energy for that probe with the entire protein. Therefore, each set of grids represents exactly the pocket conformation it was derived from. The main idea of 'in-situ' cross-docking is to paste multiple of these grids together in a linear fashion, thus providing a joined 'super-grid' which represents all the protein conformations in one grid.

Using representatives of the three main binding-pocket conformations of AR, it could be shown in a first step that for the three inhibitors, whose complex protein structures were used to generate the single grids, the correct binding mode was found. In a second step it was demonstrated that ligands which adopt binding modes similar to one of the three representative ligands, but whose complex crystal structures were not used to calculate the grids could also be docked correctly. As a last step the binding mode of an inhibitor¹²⁵ for which at that time no crystal structure was available could be predicted correctly. 'In-situ' cross-docking proposed a binding mode similar to idd 594¹¹⁴ which addresses the specificity pocket. The crystal structure solved subsequently verified the predicted binding mode. The best-scored docking solution and the ligand in the crystal structure deviated by only 0.49 Å rmsd.

In the next part of this thesis the flexibility of the AR binding pocket was examined in detail. To provide insight into this issue the available crystal structures were superimposed and analyzed using Relibase⁺ in a first step.

In agreement to previous studies¹¹³ it was shown that with respect to the binding pocket flexibility is limited to only a handful amino acids close to the specificity pocket. It was elucidated how the enzyme performs its 'induced-fit' binding mechanism. Only a small C-terminal loop region is utilized to facilitate most of the adaptation process necessary to accommodate different ligands.

In addition, a detailed comparison of the binding modes of different ligands addressing the specificity pocket was carried out. This analysis was enabled by the availability of many high-quality crystal structures of human, wild-type AR available. The analysis exceeds previous ones significantly with respect to the number and the quality of the crystal structures considered. It was found that despite similar binding modes of many inhibitors, there are subtle differences between the corresponding binding-pocket conformations.

To further explore the conformational space available to the AR binding pocket, multiple MD simulations were carried out. While crystal structure analysis is a powerful tool to analyze binding pocket features determined under certain experimental conditions, MD simulations are the method of choice to computationally examine the dynamic properties of a protein. In total, ten MD simulations (seven AR complexes plus three different holo simulations) were carried out. All simulations were started from crystal structure geometries and the parameters were set according to the current state of the art. Two of the crystal structures required to perform the MD simulation project were solved in the context of this thesis.

MD simulations produce vast amounts of data, which have to be post-processed and analyzed. This task can easily become impracticable if multiple trajectories have to be considered. To facilitate comparative analysis, an MD database (MDDDB) based on Python and MySQL has been developed. It allows to store results from multiple simulations. The powerful query features of the database helped to analyze different aspects from the same trajectory on the one hand and enabled comparison of features from different MD simulations on the other.

The MDDB tool developed as part of this study has built-in functionalities to automatically analyze distances, different rmsd measures as well as torsion angles. These features have proven useful in the comparative analysis of the ten MD simulations described in this thesis. However, there is room for future developments, since there are many more aspects of an MD simulations which are of putative interest depending on the aim of the study carried out.

The ten MD simulations mentioned above were analyzed with respect to the dynamic behavior of the binding pocket. Good overall agreement between the results from the MD simulations and the crystal structure analysis was found. Residues which exhibited elevated levels of flexibilities in the MD simulations showed in most cases also differences between the single crystal structures. However, a few residues showed unexpected behavior in the MD simulations: Phe 122, Trp 219, and Tyr 309.

In an attempt to explain this observation, an analysis of the effects of crystal packing on the binding-pocket residues was carried out. Under the special conditions of the tightly packed crystal environment, additional contacts are formed which are not necessarily present in solution. These contacts impose additional restraints on a protein, thus presumably altering its dynamical behavior. The results from the packing analysis showed that Phe 122 and Tyr 309 are involved in such contacts, thus providing a putative explanation why the conformations observed in the MD simulations are not seen in the crystal structures. Trp 219 is attached to a loop region which is also involved in crystal packing. The flexibility seen in the MD simulations is reasonable, since it is important for the biological function of the enzyme that this loop region is flexible. After each turn-over of the catalytic reaction the loop has to move as a prerequisite.

Further analysis of the flexibility of the binding pocket revealed that through the movement of Phe 122 an additional hydrophobic sub-pocket above the specificity pocket is opened. Additionally, the combined movements of Tyr 309 and the C-terminal loop region create extra space in the lower part of the AR binding pocket. Therefore, the next step was to search for small molecules which are suitable to induce or stabilize the binding pocket in one of the novel conformations observed in the MD simulations.

In the literature an interesting compound from Pfizer was presented¹⁸⁴. Docking studies using an MD snapshot as protein template suggested that the bulky phenyl group of the ligand may potentially address the additional space generated in the lower part of the binding pocket. However, the crystal structure which may confirm or disprove this hypothesis could not be obtained in the time frame of this study, thus this has to be clarified in future investigations.

To address the space generated by the Phe 122 movements different approaches were considered. A benzodiazepine scaffold was chosen, due to its easy synthetic availability and variability. Docking a small library of benzodiazepines with different side-chain decorations helped to prioritize the molecules for synthesis. One compound could be synthesized so far in the group of Prof. Schlitzer. The measured IC_{50} value indicates only very moderate binding in the high μM range. Also this part of the project leaves room for future investigations. If one of the proposed molecules shows considerable affinity towards AR, the determination of a crystal structure would be desirable to verify the binding mode proposed by the docking experiments. If none of the benzodiazepines is sufficiently active for crystallization experiments, the idea of modifying an AR inhibitor with known binding mode to address the putative new pocket conformation, provides an alternative concept. This strategy could not be followed in this study, after considering the elaborate chemistry needed to produce appropriately altered AR inhibitors to address this region of the pocket.

In a further project, the generated MD trajectories were used to energetically analyze the process of 'induced-fit' adaptation. The method MM-PBSA¹⁹⁴ was chosen for this purpose, since at first glance it seemed to be ideally suited. It calculates the binding free energies for a protein-ligand complex from ensembles of MD-generated conformations rather than using only one single rigid structure. By incorporating separate trajectories for the complexed and uncomplexed states of the protein as well as the ligand it should be possible to capture the interactions as well as the conformational energies involved in the 'induced-fit' binding process. Compared to other more rigorous methods to calculate binding free energy differences, it is applicable to larger data sets of diverse molecules. In addition, it had been shown in numerous previous studies^{194-202;204;206;207;210;223;227} that it can be a valuable tool to analyze various aspects of structure-based drug design and related topics. Up to date only

one study has been published²⁰³ where the method provides results of reduced predictability.

However, despite thorough testing of different MM-PBSA variants, binding free energy predictions were, at best, only slightly better than random. Considering the enormous amount of computational power required to perform the necessary calculations and the remarkable time needed to analyze the results, MM-PBSA did not turn out to be a cost-effective method to predict binding free energies for the dataset of AR inhibitors used in this study. Since a correct prediction of the binding free energies is a prerequisite for further interpretation in order to estimate the conformational energies involved in 'induced-fit' binding, this issue could not be further investigated.

Although many variations of the method have already been tested, a few additional points could possibly be considered in the future. One issue is the systematic parameterization of a consistent set of radii for the calculations. In the recently released new version of Amber²⁵⁰ this has been done, and it might be of interest to investigate the influences on the results. A second topic would be to analyze the effect of different dielectric constants on the PB calculations with respect to the predictions. However, keeping the overall rather disappointing results in mind it remains questionable whether significant improvements could be achieved to justify the high computational costs.

In the final section of this thesis a study was presented where in the first part aspects of flexibility of the C-terminal loop of AR were examined. In a combined study using MD simulations and multiple crystal structures, determined in the thesis of Holger Steuber, it was shown that there are clear differences between individual crystal structures of the same protein-ligand complex in this region of the enzyme. The MD simulation part of this study was performed as part of this thesis. A nice agreement between the observations made in the MD and multiple crystal structures derived from different experimental crystallization conditions was found. This is one of the rare examples which emphasizes the fact that protein flexibility might be underestimated by protein crystallography.

In the second part of this section the unexpected occurrence of multiple ligands in and close to the binding pocket of AR was described. The determination of the corresponding crystal structures was part of this work. Four tolrestat molecules occupied the space in and

next to the binding pocket. The binding of the additional ligands was accompanied by structural changes in large parts of the protein which had not been observed to this extent before.

In summary, this has analyzed many aspects of protein flexibility using AR as a model system. The importance of incorporating protein flexibility into docking experiments was shown. A new method to include multiple protein structures into the docking process has been presented. To characterize the binding-pocket flexibility of AR an exhaustive comparative analysis of the available crystal structures was carried out. To further explore the AR binding-pocket flexibility multiple MD simulations were carried out and analyzed using a database tool especially developed for this purpose. The analysis suggests two new putative binding-pocket conformations. Subsequently, initial trials were made to stabilize these new conformations using small molecules. In addition, energetic analysis of 'induced-fit' binding was carried out, thereby evaluating MM-PBSA for flexible proteins. Furthermore, the flexibility of parts of the binding pocket was examined in a joined MD/X-ray study. Again, excellent agreement was detected between the two approaches. Finally, the binding of a set of multiple inhibitor molecules to AR provides further insights into flexibility since large movements and adaptations take place which are totally unexpected. AR proves to be a valuable test system to investigate different aspects of protein flexibility.

Zusammenfassung

Die Flexibilität von Proteinen und deren Bindetaschen ist ein wichtiges, aber bis zum heutigen Tag nur schlecht verstandenes Phänomen. Insbesondere für das strukturbasierte Ligandendesign ist eine Berücksichtigung dieser Plastizität von enormer Bedeutung.

Aldose Reductase (AR) ist das erste und geschwindigkeitsbestimmende Enzym des sog. „Sorbitol-Stoffwechselweges“. Es reduziert unter Verbrauch von NADPH als Kofaktor verschiedene Aldehyde zu den entsprechenden Alkoholen. Um diese Aufgabe effizient durchführen zu können, zeigt seine Bindetasche eine intrinsische Flexibilität. Diese geht im wesentlichen von einem kleinen Bereich der Bindetasche aus, dessen Bewegungen es erlauben, eine sog. „Spezifitätstasche“ zu öffnen oder zu schließen. Dadurch wird es dem Enzym erlaubt, eine ganze Reihe von strukturell verschiedenen Aldehyden als Substrate akzeptieren zu können.

Der erste Teil der Arbeit dokumentiert ein Beispiel, anhand dessen sich die Auswirkungen einer flexiblen Bindetasche auf die Zuverlässigkeit von Dockingexperimenten abschätzen lassen. Die Experimente wurden für zwei Moleküle durchgeführt, die ursprünglich dafür konzipiert wurden, den bis zum heutigen Tag einzigartigen Bindungsmodus des AR Inhibitors Tolrestat zu kopieren. Anstatt die Dockingexperimente jedoch ausschließlich mit der Bindetaschenkonformation des Tolrestat-Komplexes durchzuführen, wurden beide Inhibitoren unter Verwendung des Programms AutoDock 3.0 einzeln in die drei bekannten Hauptbindetaschenkonformationen der AR gedockt.

Für keines der beiden Moleküle sagte die in AutoDock implementierte Bewertungsfunktion einen tolrestatähnlichen Bindungsmodus als besonders günstig voraus. Im Gegenteil, die abgeschätzten Bindungsenergien waren jeweils in der Tolrestat-Bindetasche die schlechtesten. Vielmehr wurde ein Bindungsmodus ähnlich zu dem von Sorbinil vorgeschlagen, bei dem die Spezifitätstasche sich in geschlossenem Zustand befindet.

Diese Hypothese konnte durch die später erhaltenen Kristallstrukturen der jeweiligen Komplexe bestätigt werden. In beiden Fällen ist die Spezifitätstasche geschlossen und die erhaltenen Bindungsmodi waren am ehesten mit dem von Sorbinil zu vergleichen.

Trotz dieser zunächst erfreulichen Ergebnisse waren die Vorhersagen des Dockingprogramms alles andere als perfekt. Eines der beiden Moleküle verursachte eine strukturelle Veränderung in einem Teil der Bindetasche, der bis dahin als rigide betrachtet wurde. Da unter den bekannten Kristallstrukturen kein geeignetes Templat für den gefundenen Bindungsmodus vorhanden war, war eine exakte Vorhersage des genauen Bindungsmodus dieses Moleküls nicht möglich. Im Gegensatz dazu induzierte das zweite Molekül keine ungewöhnliche Bindetaschenkonformation. Hier war es vielmehr so, dass die Stellen in der Bindetasche, in die AutoDock die Sauerstoffe einer hydrophilen Seitenkette des Moleküls platzierte, in der Kristallstruktur von drei Wassermolekülen eingenommen werden.

Diese Beispiele illustrieren auf der einen Seite, dass die vorhandenen Methoden, um mit Proteinflexibilität und dem Einfluss von Wasser auf den Bindungsmodus umzugehen, noch stark verbesserungswürdig sind. Auf der anderen Seite hat sich aber auch gezeigt, dass es empfehlenswert ist, mehrere bekannte Bindetaschenkonformationen beim Docken von kleinen Molekülen zu berücksichtigen.

Im zweiten Teil der Arbeit wurde ein kürzlich neu vorgestelltes Dockingverfahren im Zusammenhang mit Proteinflexibilität getestet. Hierbei handelt es sich um das sog. „in-situ cross docking“. Anstatt Dockingexperimente in mehrere Bindetaschen sequentiell durchzuführen, können bei diesem Verfahren multiple Proteinkonformationen simultan adressiert werden. Dadurch wird zum einen die Anzahl der notwendigen Rechnungen reduziert, zum anderen wird auch die Auswertung solcher „cross docking“ Experimente erleichtert. Betrachtet man eine größere Bibliothek von kleinen Molekülen, die man an mehrere Bindetaschenkonformation docken möchte, so stellt sich ein signifikanter Zeitgewinn ein.

Das Verfahren basiert darauf, dass das Protein während des Dockens, um Rechenzeit einzusparen, in Form eines Satzes von vorberechneten Energiegittern repräsentiert wird. Diese werden dadurch erhalten, dass verschiedene Sondenatome auf den Schnittpunkten eines definierten Gitters platziert werden, welches die Bindetasche umschließt. Von diesen Gitterpunkten aus wird die Wechselwirkungsenergie der jeweiligen Sondenatome mit dem gesamten Protein berechnet und abgespeichert. Eine Konsequenz dieser Methode ist, dass ein Satz von Energiegittern genau die Proteinkonformation repräsentiert, aus der er

abgeleitet wurde. Die Idee des „in-situ cross docking“ ist nun, mehrere solcher Gitter aneinander zu reihen und somit eine Art „Super-Gitter“ zu erzeugen, das mehrere Proteinkonformationen gleichzeitig repräsentiert.

Um die Methode in Bezug auf Proteinflexibilität zu validieren, wurde ein vereinigtes Gitter aus repräsentativen Strukturen der drei bekannten Hauptbindetaschenkonformationen erzeugt. Dann konnte zunächst gezeigt werden, dass die Bindungsmodi von Inhibitoren korrekt berechnet wurden, deren Komplexstrukturen in die Berechnung der Gitter eingeflossen waren. In einem zweiten Schritt konnten die Bindungsmodi von Inhibitoren reproduziert werden, welche ähnliche Bindetaschenkonformationen induzieren wie die ausgewählten repräsentativen Strukturen, deren Komplexstrukturen aber nicht zur Berechnung der Gitter verwendet wurden. Als letztes wurde schließlich mit dieser Methode ein Bindungsmodus für einen damals neuen AR Inhibitor vorgeschlagen. Diese Vorhersage konnte später anhand einer Kristallstruktur verifiziert werden, wobei die Abweichung (rmsd) zwischen dem vorhergesagten und dem experimentell bestimmten Bindungsmodus nur 0,49 Å betrug.

Der nächste Teil dieser Arbeit untersucht detailliert das Ausmaß der Flexibilität der einzelnen Aminosäuren der AR Bindetasche. Dafür wurde zunächst eine Analyse der vorhandenen Kristallstrukturen mit Hilfe des Programms Relibase⁺ durchgeführt.

In Übereinstimmung mit vorangegangenen Studien konnte gezeigt werden, dass lediglich einige wenige Aminosäuren für den Hauptteil der Anpassung an verschiedene Liganden verantwortlich sind. Es wurde im Detail beleuchtet, wie der „induced-fit“ Anpassungsmechanismus der Bindetasche funktioniert. Dabei zeigte sich, dass hauptsächlich eine kleine Schleifenregion, welche die Bindetasche von einer Seite begrenzt, dafür benutzt wird.

Zusätzlich wurde ein genauer Vergleich der Bindungsmoden verschiedener Liganden durchgeführt, welche alle die Spezifitätstasche adressieren. Diese Auswertung wurde ermöglicht durch die große Anzahl an qualitativ hochwertigen Komplexstrukturen der humanen AR, die in den letzten Jahre veröffentlicht oder intern gemessen wurden. Der verwendete Datensatz ist signifikant größer und qualitativ hochwertiger als alle bisherigen. Als

Ergebnis konnte festgehalten werden, dass trotz großer Ähnlichkeiten zwischen den einzelnen Bindungsmoden doch zum Teil subtile Unterschiede festzustellen sind.

Um den für das Enzym zugänglichen Konformationsraum weiter abzudecken, wurden zehn Molekulardynamik (MD) Simulationen durchgeführt. Für alle Simulationen wurden als Startpunkte die entsprechenden Kristallstrukturen gewählt. Von den verwendeten Strukturen wurden zwei im Rahmen dieser Arbeit gelöst. Die Einstellungen und Parameter für die Simulationen wurden so gewählt, dass sie dem aktuellen Stand der Technik entsprechen.

Die unmittelbare Herausforderung vor die man von MD Simulationen gestellt wird, sind die ungeheure Mengen an Daten, die durch das Verfahren generiert werden. Diese müssen nach Ende der Simulation aufbereitet und analysiert werden. Beim Vergleich der Daten von mehreren Simulationen stößt man schnell an die Grenzen des Handhabbaren. Um eine solche Analyse dennoch zu ermöglichen, wurde als Teil dieser Arbeit eine MD Datenbank (MDDDB) entwickelt. Diese enthält Routinen, um automatisiert verschiedene Größen aus unterschiedlichen MD Simulation analysieren und vergleichen zu können. Diese Datenbank hat sich als sehr hilfreich bei der vergleichenden Analyse der einzelnen MD Simulationen herausgestellt.

Die bereits erwähnten zehn MD Simulationen wurden in Bezug auf das dynamische Verhalten der einzelnen Aminosäuren innerhalb der Bindetasche hin analysiert. Dabei wurde eine gute generelle Übereinstimmung zwischen den Ergebnissen der Simulationen und der Kristallstrukturanalyse festgestellt. Aminosäuren, die erhöhte Flexibilität in den Simulationen zeigten, waren in der Regel auch in mehreren unterschiedlichen Konformationen in den Kristallstrukturen zu finden. Dennoch verhielten sich einige Aminosäuren in den Simulationen anders, als man es zunächst nach der Analyse der Kristallstrukturen erwartet hätte. Diese waren im einzelnen: Phe 122, Trp 219 und Tyr 309.

Um eine Erklärung für dieses unerwartete Verhalten zu finden, wurde eine Analyse der Auswirkungen der Kristallpackung auf die Aminosäuren der Bindetasche durchgeführt. Die eng gepackte, symmetrische Umgebung in einem Kristall generiert zusätzliche Kontakte, die in Lösung nicht notwendigerweise existieren. Diese Kontakte können strukturelle Umlagerungen bedingen oder das dynamische Verhalten des Proteins beeinflussen. Die

Packungsanalyse zeigte, dass sowohl Phe 122 als auch Tyr 309 an solchen Packungskontakten beteiligt sind. Dies könnte eine mögliche Begründung dafür sein, warum in den MD Simulationen teilweise Konformationen auftreten, die noch nie in Kristallstrukturen beobachtet wurden. Auch Trp 219 ist in einer Schleifenregion lokalisiert, die an der Kristallpackung beteiligt ist. Die in den Simulationen für diesen Rest beobachtete erhöhte Flexibilität ist sinnvoll, da die Schleifenregion um Trp 219 wichtig ist für die biologische Funktion des Enzyms: Sie öffnet und schließt die Kofaktorbindetasche. Nach jedem katalytischen Zyklus muss der Kofaktor ausgetauscht werden, daher ist eine intrinsische Flexibilität dieser Region notwendig.

Bei der genaueren Analyse der Bewegungen innerhalb der Bindetasche fiel auf, dass durch die Konformationsänderung von Phe 122 eine zusätzliche kleine Subtasche geöffnet wird. Weiterhin erzeugen die konzertierten Bewegungen von Tyr 309 und der Schleifenregion um Leu 300 zusätzlichen Raum im „unteren“ Teil der Bindetasche. Daher war der nächste Schritt, geeignete Liganden zu identifizieren, welche eine oder beide dieser neuen Bindetaschenkonformationen stabilisieren und damit experimentell zugänglich machen können.

Ein erster interessanter Ligand konnte in der Literatur identifiziert werden. Dockingexperimente, die mit einer aus einer MD Simulation extrahierten Proteinkonformation durchgeführt wurden, lassen vermuten, dass eine sterisch anspruchsvolle Phenyl-Gruppe dieses Liganden in den zusätzlich generierten Raum im „unteren“ Teil der Bindetasche ragt. Innerhalb dieser Arbeit konnte von dem entsprechenden Komplex keine Kristallstruktur erhalten werden.

Um die neue Subtasche zu adressieren, die durch die Bewegung von Phe 122 entsteht, wurde eine andere Herangehensweise gewählt. Ein Benzodiazepin-Grundgerüst wurde als Ausgangspunkt für eine kleine Bibliothek von Molekülen mit unterschiedlichen Substitutionsmustern gewählt. Benzodiazepine zeichnen sich durch ihre leichte synthetische Zugänglichkeit und Variabilität aus. Durch Docking wurde eine Vorauswahl und Priorisierung der zu synthetisierenden Moleküle getroffen. Bis zum Ende dieser Arbeit konnte im Arbeitskreis von Prof. Dr. Schlitzer eines dieser Moleküle hergestellt werden. Dieses zeigte eine moderate Bindung an das Enzym. Eine Verbesserung der Affinität und das an-

schließende Lösen einer Kristallstruktur verbleiben als zukünftige Ziele bei der Weiterführung des Projektes.

Ein weiterer Teil dieser Dissertation beschäftigt sich mit der Fragestellung, ob MD Simulationen genutzt werden können, um die an der „induced-fit“ Anpassung beteiligten Energien zu quantifizieren. Zu diesem Zweck wurde die Methode MM-PBSA gewählt, da sie auf den ersten Blick geeignet dafür erschien. MM-PBSA berechnet die freie Bindungsenergie eines Komplexes aus MD Trajektorien anstatt nur eine einzige Kristallstruktur dafür zu verwenden. Durch die Verwendung von unabhängigen Trajektorien für den Komplex, den Apo-Zustand des Proteins und den freien Liganden sollten theoretisch sowohl die Interaktions- als auch die beteiligten Konformationsenergien erfasst werden. Ein weiterer Vorteil ist die breitere Anwendbarkeit von MM-PBSA im Vergleich zu anderen Methoden der theoretischen Chemie zur Berechnung von freien Bindungsenergien. Weiterhin findet man in der Literatur eine ganze Reihe von Studien, die MM-PBSA erfolgreich für die unterschiedlichsten Aspekte des rationalen Wirkstoffdesigns einsetzen.

Im Falle der AR hat sich MM-PBSA trotz intensiver Tests verschiedener Varianten als nicht zuverlässig herausgestellt. Das berechnete Ranking der verwendeten Liganden war bestenfalls leicht besser als komplett zufallsbasierte Vorhersagen. Bedenkt man den enormen Rechenaufwand, sowie die nicht unerhebliche Zeit, die für die Analyse der Ergebnisse notwendig ist, bleibt festzustellen, dass diese in keinem Verhältnis zu den erhaltenen Ergebnissen stehen. Da bereits die Vorhersage der freien Bindungsenergien scheiterte, verboten sich alle weiterführenden Analysen in Bezug auf Konformationsenergien.

Obwohl viele Varianten der Methode evaluiert wurden, verbleiben noch ein paar weitere Punkte die weiter untersucht werden könnten. Einer dieser Punkte ist sicherlich eine systematische Parametrisierung der für die Rechnungen verwendeten Atomradien. In der vor kurzem veröffentlichten neuesten Version des Amber Programmpaketes wurde dies durchgeführt und eine Untersuchung der Auswirkungen dieser neuen Radien auf die Ergebnisse wäre sicher interessant. Weiterhin wäre eine systematische Analyse der Auswirkungen unterschiedlicher Dielektrizitätskonstanten auf die Ergebnisse von Interesse. Betrachtet man aber die in diesem Fall sehr ernüchternden Ergebnisse der Methode,

so ist es fraglich, ob ausreichende Verbesserungen zu erzielen sind, die den hohen Rechenaufwand rechtfertigen.

Im letzten Abschnitt dieser Arbeit wird eine zweiteilige Studie präsentiert. Der erste Teil beschäftigt sich mit einem weiteren Aspekt der Flexibilität des bereits erwähnten Schleifenbereiches, der für die „induced-fit“ Adaptionen verantwortlich ist. In einer kombinierten Studie aus MD Simulationen und mehreren Kristallstrukturen, die von Holger Steuber für seine Dissertation angefertigt wurden, konnte gezeigt werden, dass zwischen mehreren Strukturen des gleichen Protein-Ligand-Komplexes signifikante Unterschiede zu sehen waren. Eine exzellente Übereinstimmung zwischen den Befunden der Kristallstrukturen und den MD Simulationen wurde gefunden.

Im zweiten Teil dieses Abschnittes geht es um die unerwartete Bindung mehrerer Ligandmoleküle in und in der Nähe der Bindetasche. Die entsprechenden Kristallstrukturen wurden als Teil dieser Arbeit gelöst. Insgesamt vier Tolrestat-Moleküle konnten deutlich in der Elektronendichte lokalisiert werden. Die Bindung dieser zusätzlichen Liganden ging mit konformationellen Änderungen in weiten Teilen des Proteins einher, die in dieser Form vorher noch nicht beobachtet wurden.

Insgesamt hat sich diese Arbeit mit vielen Aspekten der Proteinflexibilität beschäftigt. Dabei wurde AR als Modellsystem verwendet. Die Bedeutung der Proteinflexibilität für das Docken von kleinen Molekülen wurde gezeigt. Anschließend wurde eine neue Docking-Methode evaluiert, mit der sich mehrere Proteinkonformationen simultan adressieren lassen. Um das Ausmaß der Flexibilität von AR zu charakterisieren, wurden Kristallstrukturanalysen und vergleichende MD Simulationen durchgeführt. Zur Analyse der Simulationen wurde ein eigens dafür entwickeltes Datenbanktool verwendet. Durch diese Analyse zeigten sich zwei neue potentielle Bindetaschenkonformationen. Anschließend wurden erste Versuche unternommen, diese Konformationen durch geeignete Liganden zu stabilisieren. Weiterhin wurde versucht, den Prozess der „induced-fit“ Bindung energetisch zu beschreiben. Dabei wurde die Methode MM-PBSA an einem Protein mit flexibler Bindetasche evaluiert. In einer kombinierten Studie aus Röntgenstrukturen und MD Simulationen wurde im letzten Teil der Arbeit auf spezielle Aspekte der Proteinflexibilität näher

eingegangen. Dabei wurde exzellente Übereinstimmung zwischen Experiment und Simulation erzielt.

Insgesamt hat sich diese Arbeit am Beispiel der AR mit einer Reihe von Aspekten der Proteinflexibilität beschäftigt. Dabei hat sich AR als wertvolles Testsystem herausgestellt, um solche Phänomene mit unterschiedlichen Methoden zu untersuchen.

Appendix A

The Molecular Dynamics Database (MDDB)

Introduction

The primary results of a Molecular Dynamics (MD) simulation are the positions of all atoms of the system under investigation over the course of the simulation time. At regular time intervals defined by the user the positions of the atoms are stored in a file. A single set of coordinates belongs to a certain instant in time, called a snapshot. The series of snapshots collected during the simulation is called the trajectory of this simulation. Using appropriate visualization software such as PyMol²⁵¹ or VMD²⁵², MD trajectories can be loaded and visualized. If loaded sequentially, the individual snapshots create a movie of the movements occurring within the system. While visualizing the trajectory is a valuable tool to get a qualitative overview of the development of the system, other methods are required to analyze the simulation in a more quantitative manner. Depending on the goal of the study, the analysis of torsion angles, distances between atoms or groups of atoms as well as different rms values are normally required to investigate the flexibility of the protein binding pocket based on a MD trajectory.

Similar to other MD package, the Amber program suite¹⁵⁷ contains routines which allow post-processing of trajectories to extract this kind of information from the trajectory: *ptraj* and *carnal*. Among other aspects, these programs allow the measurement of distances, rms values, and torsion angles. Both programs store their results in simple text files, but use slightly different file formats. If analyzing a trajectory with *a priori* knowledge about the most interesting parts of the system every parameter of interest can be measured in one run. However, while examining MD simulations in a rather exploratory way without any predefined structural parameters of interest, the problem arises that for every parameters investigated at a later stage, the entire trajectory has to be post-processed repeatedly. Even with most recent computer hardware this can be a tedious process for a multi-nanosecond trajectories with stored water positions. In addition, there is the problem of finding the most interesting parts of the simulation, thus many features have to be extracted from the trajectories. To perform this, the different parameters need to be plotted for graphical dis-

play and statistically analyzed to obtain mean values and standard deviations. Furthermore, functionalities are needed to compare and sort different parameters according to their statistical values.

In addition to the coordinates for all atoms, further parameters describing the state of the system are written to a second file, which contains information about different energy values. For example, the total, potential, electrostatic, kinetic, and van-der-Waals energies, but also the pressure and volume are monitored throughout the simulation. Examining these values is also of great importance to ensure the system remains within normal conditions. Again, these parameters have to be extracted from the trajectory, plotted and statistically analyzed.

In a comparative MD analysis between multiple simulations all these problem scale linearly with the number of trajectories under investigation. Moreover, there is the simple technical difficulty of how to compare most efficiently the same parameters obtained from different simulations. Usually this requires storage of all plots and statistical informations in separate files. With increasing numbers of simulations this becomes increasingly cumbersome. Thus, there is a need for a more ordered storage system, which allows easy comparisons of both parameters from the same MD simulation and between different trajectories: a molecular dynamics database (MDDDB).

Database and program architecture

MDDDB is based on a *MySQL* backend machine (version 4.1.2, www.mysql.com). To browse and query the data, *phpMyAdmin* (version 2.6.2, www.phpmyadmin.net) is used as web-interface. The post-processing of the trajectories, the data insertion into the database, and the analysis and plotting of the data is automatically handled using *Python* (version 2.3.5, www.python.org). Plots are generated using *Gnuplot* (version 4.0 patchlevel 0, www.gnuplot.info) for distance, rms, and dihedral plots. Torsdialplots are generated using the *pst-col* package from *latex* (version 3.14, www.latex-project.org). The post-processing of the trajectories itself is handled by *ptraj* from the Amber suite.

Data content

Meta data about every simulation, such as the name of the MD simulation and the location of the topology and trajectory files, is stored in the table *md*. Data from Amber topology

files required for MDDB data analysis are included in the database. These informations are stored in the tables *atom* and *residue*. The information contained within the energy 'trajectory' of an Amber run is completely added to the database and plotted. The results are stored in the table *frames_analysis*. Since these energy data are especially important for the equilibration phase of an MD simulation, energy files of this phase are separately analyzed and the corresponding results are stored in the table *frames_initial_analysis*. Concerning structural data relevant to analyze the trajectory, MDDB is able to handle the following parameters from an MD simulation:

- **torsion angles:** all torsions of the protein are automatically calculated and added to the database. Dihedral angles for ligands and cofactors need to be manually specified. The results are stored in the table *dihedral_analysis*.
- **distances:** results are stored in the table *distances_analysis*.
- **rms values:** MDDB can handle rms values for single atoms, complete residues or ligands as well as the entire protein. The results are stored in the *rms_atom_analysis*, *rms_residue_analysis*, and *rms_all_analysis* tables respectively. Rms values can be automatically added to the database using *ptraj*. Additionally, there is the possibility to pre-calculate rms values using *carnal* and subsequently add them to the database.

Using MDDB

The general work flow can be dissected into three steps:

1. preparation of input files and adding of missing group definitions to *group_def* table;
2. execution of the program to add the data, the corresponding statistical parameters and the plots to the database;
3. analysis of the data.

In a first step the config files and the program script need to be adjusted to the current MD simulation (see the manual for technical details). Most adjustments needed (MD name, locations of the topology and trajectory files, ...) are self-explanatory.

In the next step the data are added to the database by executing the program. Depending on the trajectory size and the number of data sets to be added the runtime of the program can vary. However, in most cases the program should accomplish the job within a few hours.

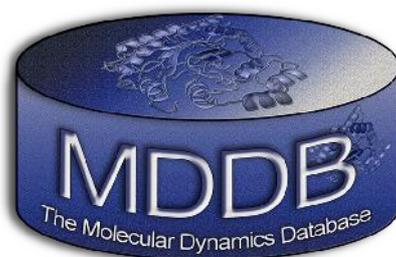


Figure 66: The MDDDB logo.

Once the data are added, the web-interface can be used to analyze them. Through the web-interface the data can be browsed and standard sql queries can be submitted to the database. The results from these queries can be sorted according to different criteria by simply clicking on the corresponding column.

MDDDB enables an efficient and time-saving way to cope with the enormous amount of data produced by multiple MD simulations. Instead of having to handle literally thousands of plots and text files with mean values and standard deviations, the user can simply browse through the data, sort them and identify exactly those parts of the simulations which are of highest interest to him.

Appendix B

Additional crystal structures for MD simulations

For every ligand in the MD data set complex crystal structures with human AR were needed to guarantee the highest possible quality of the starting geometries for the MD simulations. In order to complete the dataset, complex crystal structures for the AR inhibitors sorbinil and tolrestat were solved in the context of this work. The unexpected findings in the four-ligand tolrestat complex structure have already been discussed in greater detail in chapter 'Expect the unexpected' (page 163). For the sorbinil complex structure no unusual observations were made. Therefore, the description and analysis of the sorbinil and the synchrotron derived-tolrestat complex will be limited to a Relibase⁺ analysis of the differences between the binding pockets of the corresponding human (9sbi, 9tol) and porcine (1ah0¹²⁰, 1ah3¹²⁰) structures. The protocol used for superposition and analysis is analogous to the comparative crystal structure analysis (page 38) and based on the same settings and parameters.

Structure Determination

Cloning, expression, purification, and crystallization of Aldose Reductase as well as structure determination has been done according to procedures published earlier^{114;253;254}.

Data were collected at 100 K using crystals previously exposed to a solution of 40 % (m/V) PEG 6000 in 50 mM di-ammonium hydrogen citrate at pH 5 as cryoprotectant. The data sets were determined with a RIGAKU copper rotating anode (Molecular Structure Cooperation) at 50 kV, 90 mA as X-ray source and a R-AXIS IV++ image plate system. For each frame the exposure time and oscillation range were set to 5 min and 0.5°, respectively. The data set of the tolrestat complex at 1.08 Å resolution was collected at the beamline X06SA at SLS, Switzerland.

All data processing and scaling was performed using the HKL2000 package²⁵⁵. The coordinates of human AR (pdb code 1e13²⁵⁶) were used for initial rigid-body refinement of the protein atoms followed by repeated cycles of conjugate gradient energy minimization, simulated annealing, and B-factor refinement using the CNS program package²⁵⁷. Refine-

ment at later stages was performed with the program SHELXL²⁵⁸. Here, at least 20 cycles of conjugate gradient minimization were performed with default restraints on bonding geometry and B-values. Five percent of all data were used for R_{free} calculation. Amino-acid side chains were fitted into sigma A-weighted 2Fo-Fc and Fo-Fc electron density maps using O²⁵⁹. After the first refinement cycle water molecules and, subsequently, cofactor and ligand were located in the electron density and added to the model. Restraints were applied to bond lengths and angles, chiral volume, planarity of aromatic rings and van der Waals contacts. Multiple side-chain conformations were built if an appropriate electron density was observed and maintained during the refinement. The threshold for keeping the less populated side-chain conformation in the model was set to 10% occupancy. During the last refinement cycles, riding H atoms were introduced without using additional parameters. The final models were validated using PROCHECK²⁶⁰. Data collection, unit cell parameters, and refinement statistics are given in Table 19.

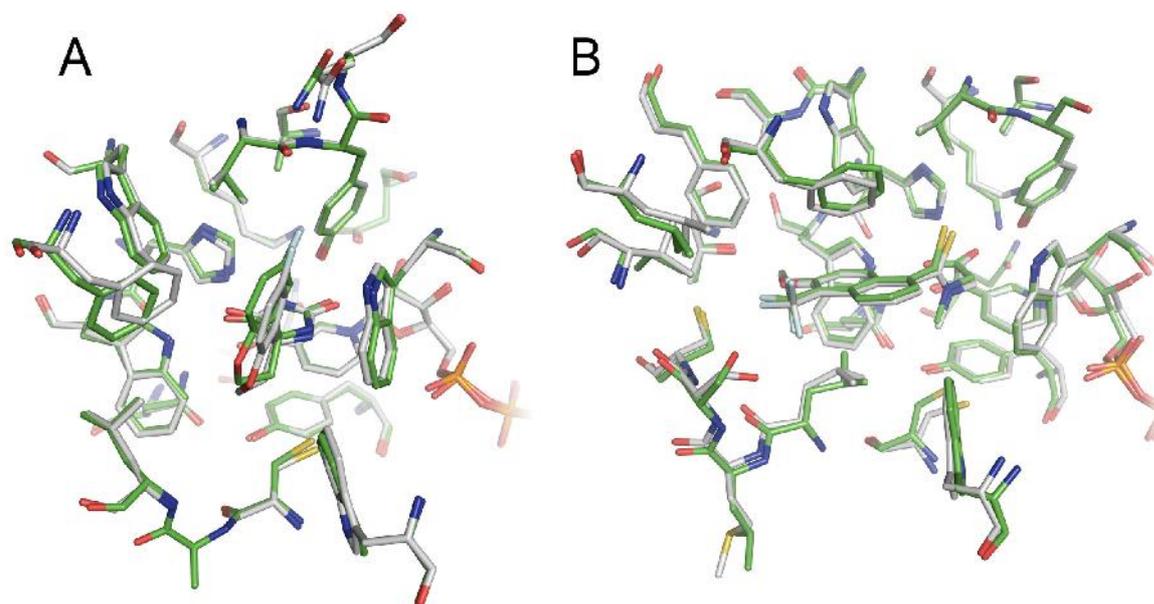


Figure 67: Superpositions of human (green) and porcine (silver) AR. **A** shows the superposition for the inhibitor sorbinil, **B** for the inhibitor tolrestat.

The sorbinil complex structure

The porcine (1ah0) and human (9sbi) sorbinil complex structures match with an overall $C\alpha$ rmsd value of 0.62 Å. The binding-site rmsd using all residues in a 6 Å sphere around the ligand is 0.21 Å. These low rms deviations already indicate that the two structures super-

impose very well. No $C\alpha$ or side-chain center movements can be observed. The ligand volume overlap between the ligands in the two structures is 96%. The only clash that can be measured is between one atom of the phenyl moiety of the ligand in the human structure and the CZ atom of Phe 122 in the porcine structure. As can be seen in Figure 67 A, Phe 122 shows the largest deviation for all binding-site residues. This is not surprising, since in position 121 there is a proline in the porcine structure, whereas in the human enzyme a Phe is located at this position. Therefore, the largest structural perturbations are to be expected in this region. However, the impact on the binding-pocket geometry is rather moderate.

The tolrestat complex structure

Since the structure with four molecules bound has already been discussed in chapter 'Expect the unexpected' (page 163), it will be omitted here and only the synchrotron structure will be compared to the porcine PDB complex.

The overall $C\alpha$ rmsd between the human and porcine tolrestat complex is 0.59 Å, whereas the binding pocket residues superimpose with an rmsd of 0.37 Å. Compared to the rmsds of the sorbinil structures the overall value is in the same range, whereas the binding pocket deviation is slightly higher. However, the tolrestat binding pocket is larger than the sorbinil pocket, since only tolrestat addresses the specificity pocket.

Three $C\alpha$ movements above the threshold are observed. Two of the involved residues (Leu 124, Val 130) are located at the upper end of the specificity pocket and are attached to the flexible loop region following Phe 122 in sequence. The third residue, Trp 219, is also part of a loop region. Therefore, it is questionable whether these are important structural changes or inherent fluctuations for a loop region. The same is true for the single movement of the Val 130 side chain: no direct impact on the binding pocket can be observed from this movement (see Figure 67 B). The volume overlap between the two ligands is 99%, which indicates a virtually perfect match of the corresponding binding modes.

	sorbinil	tolrestat four ligands	tolrestat synchrotron
Data collection and Processing			
No. of crystals used	1	1	1
Wavelength [Å]	1.5418	1.5418	0.97920
Space group	P 2 ₁	P 2 ₁	P 2 ₁
Unit cell parameters			
a, b, c [Å]	49.4, 66.8, 47.4	49.2, 68.1, 46.7	49.5, 66.7, 47.1
β [°]	91.8	92.8	92.5
Diffraction data			
Resolution range [Å]	50-1.5 (1.53 -1.5)	50-1.5 (1.53 -1.5)	30-1.08 (1.1 - 1.08)
Unique reflections	68182 (2947)	43 631 (1 920)	123 403 (5 449)
R(I) _{sym} [%]	6.0 (20.0)	8.3 (22.5)	5.2 (8.7)
Completeness [%]	91.9 (84.1)	88.2 (78.7)	94.5 (83.3)
Redundancy	1.6 (1.7)	2.4 (2.3)	2.8 (1.9)
I/σ(I)	17.6 (5.8)	10.2 (3.9)	17.9 (8.9)
Refinement			
Resolution range used in refinement [Å]	8-15	20-1.5	10-1.08
Reflections used in refinement (work/free)	42 909 / 1942	41 439 / 2 163	116 227/6 130
Final R values for all reflections (work/free) [%]	15.5 / 18.9	16.9 / 22.0	11.2/13.9
Final R values for reflections with F > 4 σ (work/free) [%]	15.2 / 18.3	16.2 / 21.0	11.0/13.6
Protein residues	316	316	316
Coenzyme	1	1	1
Inhibitor	1	4	1
Water molecules	408	344	350

	sorbinil	tolrestat four ligands	tolrestat synchrotron
RMSDs			
Bonds [Å]	0.012	0.011	0.017
Angles [°]	2.6	2.3	2.4
Ramachandran plot			
Most favoured [%]	90.3	91.7	90.2
Additionally allowed [%]	9.7	8.3	9.8
Generously allowed [%]	-	-	-
Mean B factors [Å²]			
Protein	11.4	19.4	11.7
NADP ⁺	6.9	15.0	5.5
Inhibitor	18.1	26.1	8.2
Water molecules	21.7	29.1	21.8

Table 19: Crystallographic parameters for the three AR complex crystal structures solved in context of this work. The structure labeled 'tolrestat four ligands' corresponds to the one discussed in chapter 'Expect the unexpected' (page 163) where in total four ligand molecules were found near the binding site, whereas 'tolrestat synchrotron' corresponds to the second structure mentioned in the same chapter.

Usage of crystal structures from collaboration partners

The determinations of some crystal structures used as input geometries for the MD simulations (Pfizer compound¹²⁶, 47d, JFD compound), the structures used to validate the docking experiments presented in the first chapter as well as the zopolrestat bound structures (1d, 3d_a, 3d_b, 3d_c, 6d_a, 6d_b, co_10)¹⁵³ of AR described in the chapter 'Expect the unexpected' are not part of this thesis. They were determined by Holger Steuber, PhD student in the Klebe group, as part of his PhD project. All other structures used in this thesis are either taken from the PDB¹⁵⁴ or were solved as part of this thesis (sorbiniol, tolrestat, tolrestat with four bound ligands).

References

- [1] HERS H.G. The mechanism of the transformation of glucose in fructose in the seminal vesicles. *Biochim Biophys Acta* (1956) **22** pp.202-203.
- [2] JEDZINIAK J.A., YATES E.M. & KINOSHITA J.H. Lens polyol dehydrogenase. *Exp Eye Res* (1973) **16** pp.95-104.
- [3] KINOSHITA J.H., FUTTERMAN S., SATOH K. & MEROLA L.O. Factors affecting the formation of sugar alcohols in ocular lens. *Biochim Biophys Acta* (1963) **74** pp.340-350.
- [4] CAO D., FAN S.T. & CHUNG S.S. Identification and characterization of a novel human aldose reductase-like gene. *J Biol Chem* (1998) **273** pp.11429-11435.
- [5] CARR I.M. & MARKHAM A.F. Molecular genetic analysis of the human sorbitol dehydrogenase gene. *Mamm Genome* (1995) **6** pp.645-652.
- [6] MAEKAWA K., TANIMOTO T., OKADA S., SUZUKI T., SUZUKI T. & YABE-NISHIMURA C. Analysis of gene expression of aldose reductase and sorbitol dehydrogenase in rat Schwann cells by competitive RT-PCR method using non-homologous DNA standards. *Brain Res Brain Res Protoc* (2001) **8** pp.219-227.
- [7] AIDA K., IKEGISHI Y., CHEN J., TAWATA M., ITO S., MAEDA S. & ONAYA T. Disruption of aldose reductase gene (Akr1b1) causes defect in urinary concentrating ability and divalent cation homeostasis. *Biochem Biophys Res Commun* (2000) **277** pp.281-286.
- [8] CHENG H.M., GONZALEZ R.G. The effect of high glucose and oxidative stress on lens metabolism, aldose reductase and senile cataractogenesis. *Metabolism* (1986) **35** pp.10-14.
- [9] GONZALES R.G., BARNETT P., AGUAYO J., CHENG H.M., CHYLACK L.T.J. Direct measurement of polyol pathway activity in the ocular lens. *Diabetes* (1984) **33** pp.196-199.
- [10] KINOSHITA J.H. & NISHIMURA C. The involvement of aldose reductase in diabetic complications. *Diabetes Metab Rev* (1988) **4** pp.323-337.
- [11] PETRASH J.M. All in the family: aldose reductase and closely related ald-keto reductases. *Cell Mol Life Sci* (2004) **61** pp.737-749.
- [12] SRIVASTAVA S.K., RAMANA K.V. & BHATNAGAR A. Role of aldose reductase and oxidative damage in diabetes and the consequent potential for therapeutic options. *Endocr Rev* (2005) **26** pp.380-392.
- [13] SRIVASTAVA S., RAMANA K.V., TAMMALI R., SRIVASTAVA S.K. & BHATNAGAR A. Contribution of aldose reductase to diabetic hyperproliferation of vascular smooth muscle cells. *Diabetes* (2006) **55** pp.901-910.

- [14] YABE-NISHIMURA C. Aldose reductase in glucose toxicity: a potential target for the prevention of diabetic complications. *Pharmacol Rev* (1998) **50** pp.21-33.
- [15] CHUNG S.S.M., HO E.C.M., LAM K.S.L. & CHUNG S.K. Contribution of polyol pathway to diabetes-induced oxidative stress. *J Am Soc Nephrol* (2003) **14** p. S233-6.
- [16] CHUNG S.S.M. & CHUNG S.K. Aldose reductase in diabetic microvascular complications. *Curr Drug Targets* (2005) **6** pp.475-486.
- [17] EL-KABBANI O., WILSON D.K., PETRASH M. & QUIOCHO F.A. Structural features of the aldose reductase and aldehyde reductase inhibitor-binding sites. *Mol Vis* (1998) **4** p. 19.
- [18] RONDEAU J.M., TETE-FAVIER F., PODJARNY A., REYMANN J.M., BARTH P., BIELLMANN J.F. & MORAS D. Novel NADPH-binding domain revealed by the crystal structure of aldose reductase. *Nature* (1992) **355** pp.469-472.
- [19] WILSON D.K., BOHREN K.M., GABBAY K.H. & QUIOCHO F.A. An unlikely sugar substrate site in the 1.65 Å structure of the human aldose reductase holoenzyme implicated in diabetic complications. *Science* (1992) **257** pp.81-84.
- [20] TAKAHASHI M., FUJII J., TESHIMA T., SUZUKI K., SHIBA T. & TANIGUCHI N. Identity of a major 3-deoxyglucosone-reducing enzyme with aldehyde reductase in rat liver established by amino acid sequencing and cDNA expression. *Gene* (1993) **127** pp.249-253.
- [21] KADOR P.F., KINOSHITA J.H., STRIBLING D., BRITAIN D.R., MIRLESS D.J., SENNITT C.M. Rat lens aldose reductase and polyol production: a reply. *Biochem Journal* (1987) **247** pp.495-496.
- [22] DEL CORSO A., COSTANTINO L., RASTELLI G., BUONO F. & MURA U. Aldose reductase does catalyse the reduction of glyceraldehyde through a stoichiometric oxidation of NADPH. *Exp Eye Res* (2000) **71** pp.515-521.
- [23] PENNING T.M. Molecular endocrinology of hydroxysteroid dehydrogenases. *Endocr Rev* (1997) **18** pp.281-305.
- [24] MADORE E., HARVEY N., PARENT J., CHAPDELAINE P., AROSH J.A. & FORTIER M.A. An aldose reductase with 20 alpha-hydroxysteroid dehydrogenase activity is most likely the enzyme responsible for the production of prostaglandin f2 alpha in the bovine endometrium. *J Biol Chem* (2003) **278** pp.11205-11212.
- [25] RAMANA K.V., DIXIT B.L., SRIVASTAVA S., BALENDIRAN G.K., SRIVASTAVA S.K. & BHATNAGAR A. Selective recognition of glutathiolated aldehydes by aldose reductase. *Biochemistry* (2000) **39** pp.12172-12180.
- [26] DIXIT B.L., BALENDIRAN G.K., WATOWICH S.J., SRIVASTAVA S., RAMANA K.V., PETRASH J.M., BHATNAGAR A. & SRIVASTAVA S.K. Kinetic and structural characterization of the glutathione-binding site of aldose reductase. *J Biol Chem* (2000) **275** pp.21587-21595.

- [27] SRIVASTAVA S., SPITE M., TRENT J.O., WEST M.B., AHMED Y. & BHATNAGAR A. Aldose reductase-catalyzed reduction of aldehyde phospholipids. *J Biol Chem* (2004) **279** pp.53395-53406.
- [28] SRIVASTAVA S., WATOWICH S.J., PETRASH J.M., SRIVASTAVA S.K. & BHATNAGAR A. Structural and kinetic determinants of aldehyde reduction by aldose reductase. *Biochemistry* (1999) **38** pp.42-54.
- [29] WERMUTH B. & MONDER C. Aldose and aldehyde reductase exhibit isocorticosteroid reductase activity. *Eur J Biochem* (1983) **131** pp.423-426.
- [30] MATSUURA K., DEYASHIKI Y., BUNAI Y., OHYA I. & HARA A. Aldose reductase is a major reductase for isocaproaldehyde, a product of side-chain cleavage of cholesterol, in human and animal adrenal glands. *Arch Biochem Biophys* (1996) **328** pp.265-271.
- [31] VANDER JAGT D.L., HASSEBROOK R.K., HUNSAKER L.A., BROWN W.M. & ROYER R.E. Metabolism of the 2-oxoaldehyde methylglyoxal by aldose reductase and by glyoxalase-I: roles for glutathione in both enzymes and implications for diabetic complications. *Chem Biol Interact* (2001) **130-132** pp.549-562.
- [32] VANDER JAGT D.L., ROBINSON B., TAYLOR K.K. & HUNSAKER L.A. Reduction of trioses by NADPH-dependent ald-keto reductases. Aldose reductase, methylglyoxal, and diabetic complications. *J Biol Chem* (1992) **267** pp.4364-4369.
- [33] VANDER JAGT D.L., KOLB N.S., VANDER JAGT T.J., CHINO J., MARTINEZ F.J., HUNSAKER L.A. & ROYER R.E. Substrate specificity of human aldose reductase: identification of 4-hydroxynonenal as an endogenous substrate. *Biochim Biophys Acta* (1995) **1249** pp.117-126.
- [34] BOHREN K.M., BROWNLEE J.M., MILNE A.C., GABBAY K.H. & HARRISON D.H.T. The structure of Apo R268A human aldose reductase: hinges and latches that control the kinetic mechanism. *Biochim Biophys Acta* (2005) **1748** pp.201-212.
- [35] BOHREN K.M., GRIMSHAW C.E., LAI C.J., HARRISON D.H., RINGE D., PETSKO G.A. & GABBAY K.H. Tyrosine-48 is the proton donor and histidine-110 directs substrate stereochemical selectivity in the reduction reaction of human aldose reductase: enzyme kinetics and crystal structure of the Y48H mutant enzyme. *Biochemistry* (1994) **33** pp.2021-2032.
- [36] CZODROWSKI P., SOTRIFFER C.A., KLEBE G. Manuscript in Preparation. (2006) .
- [37] STEUBER H., KLEBE G. Manuscript in Preparation. (2006) .
- [38] GRIMSHAW C.E., SHAHBAZ M. & PUTNEY C.G. Mechanistic basis for nonlinear kinetics of aldehyde reduction catalyzed by aldose reductase. *Biochemistry* (1990) **29** pp.9947-9955.
- [39] KUBISESKI T.J., HYNDMAN D.J., MORJANA N.A. & FLYNN T.G. Studies on pig muscle aldose reductase. Kinetic mechanism and evidence for a slow conformational change upon coenzyme binding. *J Biol Chem* (1992) **267** pp.6510-6517.

- [40] GRIMSHAW C.E., BOHREN K.M., LAI C.J. & GABBAY K.H. Human aldose reductase: rate constants for a mechanism including interconversion of ternary complexes by recombinant wild-type enzyme. *Biochemistry* (1995) **34** pp.14356-14365.
- [41] VANDER JAGT D.L., STANGEBYE L.A., HUNSAKER L.A., EATON R.P. & SIBBITT W.L.J. Purification of aldose reductase from human placenta and stabilization of the inhibitor binding site. *Biochem Pharmacol* (1988) **37** pp.1051-1056.
- [42] PETRASH J.M., HARTER T.M., DEVINE C.S., OLINS P.O., BHATNAGAR A., LIU S. & SRIVASTAVA S.K. Involvement of cysteine residues in catalysis and inhibition of human aldose reductase. Site-directed mutagenesis of Cys-80, -298, and -303. *J Biol Chem* (1992) **267** pp.24833-24840.
- [43] SRIVASTAVA S., CHANDRASEKAR B., BHATNAGAR A. & PRABHU S.D. Lipid peroxidation-derived aldehydes and oxidative stress in the failing heart: role of aldose reductase. *Am J Physiol Heart Circ Physiol* (2002) **283** p. H2612-9.
- [44] SRIVASTAVA S., LIU S.Q., CONKLIN D.J., ZACARIAS A., SRIVASTAVA S.K. & BHATNAGAR A. Involvement of aldose reductase in the metabolism of atherogenic aldehydes. *Chem Biol Interact* (2001) **130-132** pp.563-571.
- [45] DAS B. & SRIVASTAVA S.K. Activation of aldose reductase from human tissues. *Diabetes* (1985) **34** pp.1145-1151.
- [46] SRIVASTAVA S.K., ANSARI N.H., HAIR G.A., AWASTHI S. & DAS B. Activation of human erythrocyte, brain, aorta, muscle, and ocular tissue aldose reductase. *Metabolism* (1986) **35** pp.114-118.
- [47] SRIVASTAVA S.K., HAIR G.A. & DAS B. Activated and unactivated forms of human erythrocyte aldose reductase. *Proc Natl Acad Sci U S A* (1985) **82** pp.7222-7226.
- [48] GRIMSHAW C.E., SHAHBAZ M., JAHANGIRI G., PUTNEY C.G., MCKERCHER S.R. & MATHUR E.J. Kinetic and structural effects of activation of bovine kidney aldose reductase. *Biochemistry* (1989) **28** pp.5343-5353.
- [49] GRIMSHAW C.E. & LAI C.J. Oxidized aldose reductase: in vivo factor not in vitro artifact. *Arch Biochem Biophys* (1996) **327** pp.89-97.
- [50] DEL CORSO A., BARSACCHI D., CAMICI M., GARLAND D. & MURA U. Bovine lens aldose reductase: identification of two enzyme forms. *Arch Biochem Biophys* (1989) **270** pp.604-610.
- [51] GRIMSHAW C.E. Chromatographic separation of activated and unactivated forms of aldose reductase. *Arch Biochem Biophys* (1990) **278** pp.273-276.
- [52] DEL CORSO A., DAL MONTE M., VILARDO P.G., CECCONI I., MOSCHINI R., BANDITELLI S., CAPIELLO M., TSAI L. & MURA U. Site-specific inactivation of aldose reductase by 4-hydroxynonenal. *Arch Biochem Biophys* (1998) **350** pp.245-248.

- [53] CAPPIELLO M., VOLTARELLI M., GIANNESI M., CECONI I., CAMICI G., MANAO G., DEL CORSO A. & MURA U. Glutathione dependent modification of bovine lens aldose reductase. *Exp Eye Res* (1994) **58** pp.491-501.
- [54] CHANDRA A., SRIVASTAVA S., PETRASH J.M., BHATNAGAR A. & SRIVASTAVA S.K. Modification of aldose reductase by S-nitrosoglutathione. *Biochemistry* (1997) **36** pp.15801-15809.
- [55] CHANDRA A., SRIVASTAVA S., PETRASH J.M., BHATNAGAR A. & SRIVASTAVA S.K. Active site modification of aldose reductase by nitric oxide donors. *Biochim Biophys Acta* (1997) **1341** pp.217-222.
- [56] SRIVASTAVA S.K., RAMANA K.V., CHANDRA D., SRIVASTAVA S. & BHATNAGAR A. Regulation of aldose reductase and the polyol pathway activity by nitric oxide. *Chem Biol Interact* (2003) **143-144** pp.333-340.
- [57] SRIVASTAVA S., DIXIT B.L., RAMANA K.V., CHANDRA A., CHANDRA D., ZACARIAS A., PETRASH J.M., BHATNAGAR A. & SRIVASTAVA S.K. Structural and kinetic modifications of aldose reductase by S-nitrosothiols. *Biochem J* (2001) **358** pp.111-118.
- [58] MORRE D.M., LENAZ G. & MORRE D.J. Surface oxidase and oxidative stress propagation in aging. *J Exp Biol* (2000) **203** pp.1513-1521.
- [59] KINOSHITA J.H., BARBER G.W., MEROLA L.O. & TUNG B. Changes in the levels of free amino acids and myo-inositol in the galactose-exposed lens. *Invest Ophthalmol* (1969) **8** pp.625-632.
- [60] LEE A.Y., CHUNG S.K. & CHUNG S.S. Demonstration that polyol accumulation is responsible for diabetic cataract by the use of transgenic mice expressing the aldose reductase gene in the lens. *Proc Natl Acad Sci U S A* (1995) **92** pp.2780-2784.
- [61] NISHIZUKA Y. Intracellular signaling by hydrolysis of phospholipids and activation of protein kinase C. *Science* (1992) **258** pp.607-614.
- [62] ISHII H., TADA H. & ISOGAI S. An aldose reductase inhibitor prevents glucose-induced increase in transforming growth factor-beta and protein kinase C activity in cultured mesangial cells. *Diabetologia* (1998) **41** pp.362-364.
- [63] KEOGH R.J., DUNLOP M.E. & LARKINS R.G. Effect of inhibition of aldose reductase on glucose flux, diacylglycerol formation, protein kinase C, and phospholipase A2 activation. *Metabolism* (1997) **46** pp.41-47.
- [64] KAMIYA H., NAKAMURA J., HAMADA Y., NAKASHIMA E., NARUSE K., KATO K., YASUDA Y. & HOTTA N. Polyol pathway and protein kinase C activity of rat Schwannoma cells. *Diabetes Metab Res Rev* (2003) **19** pp.131-139.
- [65] WOLF B.A., WILLIAMSON J.R., EASOM R.A., CHANG K., SHERMAN W.R. & TURK J. Diacylglycerol accumulation and microvascular abnormalities induced by elevated glucose levels. *J Clin Invest* (1991) **87** pp.31-38.
- [66] HA H., YU M.R., CHOI Y.J. & LEE H.B. Activation of protein kinase c-delta and c-epsilon by oxidative

- stress in early diabetic rat kidney. *Am J Kidney Dis* (2001) **38** p. S204-7.
- [67] BURG M.B. Molecular basis of osmotic regulation. *Am J Physiol* (1995) **268** p. F983-96.
- [68] TERUBAYASHI H., SATO S., NISHIMURA C., KADOR P.F. & KINOSHITA J.H. Localization of aldose and aldehyde reductase in the kidney. *Kidney Int* (1989) **36** pp.843-851.
- [69] BAGNASCO S.M., MURPHY H.R., BEDFORD J.J. & BURG M.B. Osmoregulation by slow changes in aldose reductase and rapid changes in sorbitol flux. *Am J Physiol* (1988) **254** p. C788-92.
- [70] COWLEY B.D.J., FERRARIS J.D., CARPER D. & BURG M.B. In vivo osmoregulation of aldose reductase mRNA, protein, and sorbitol in renal medulla. *Am J Physiol* (1990) **258** p. F154-61.
- [71] KANEKO M., CARPER D., NISHIMURA C., MILLEN J., BOCK M. & HOHMAN T.C. Induction of aldose reductase expression in rat kidney mesangial cells and Chinese hamster ovary cells under hypertonic conditions. *Exp Cell Res* (1990) **188** pp.135-140.
- [72] HENRY D.N., DEL MONTE M., GREENE D.A. & KILLEN P.D. Altered aldose reductase gene regulation in cultured human retinal pigment epithelial cells. *J Clin Invest* (1993) **92** pp.617-623.
- [73] FERRARETTO A., NEGRI A., GIULIANI A., DE GRADA L., FUHRMAN CONTI A.M. & RONCHI S. Aldose reductase is involved in long-term adaptation of EUE cells to hyperosmotic stress. *Biochim Biophys Acta* (1993) **1175** pp.283-288.
- [74] KOLB N.S., HUNSAKER L.A. & VANDER JAGT D.L. Aldose reductase-catalyzed reduction of acrolein: implications in cyclophosphamide toxicity. *Mol Pharmacol* (1994) **45** pp.797-801.
- [75] LEE E.K., REGENOLD W.T. & SHAPIRO P. Inhibition of aldose reductase enhances HeLa cell sensitivity to chemotherapeutic drugs and involves activation of extracellular signal-regulated kinases. *Anticancer Drugs* (2002) **13** pp.859-868.
- [76] EL-KABBANI O., RUIZ F., DARMANIN C. & CHUNG R.P. Aldose reductase structures: implications for mechanism and inhibition. *Cell Mol Life Sci* (2004) **61** pp.750-762.
- [77] DE LA FUENTE J.A. & MANZANARO S. Aldose reductase inhibitors from natural sources. *Nat Prod Rep* (2003) **20** pp.243-251.
- [78] HAMADA Y. & NAKAMURA J. Clinical potential of aldose reductase inhibitors in diabetic neuropathy. *Treat Endocrinol* (2004) **3** pp.245-255.
- [79] HANDELSMAN D.J. & TURTLE J.R. Clinical trial of an aldose reductase inhibitor in diabetic neuropathy. *Diabetes* (1981) **30** pp.459-464.
- [80] FAGIUS J. & JAMESON S. Effects of aldose reductase inhibitor treatment in diabetic polyneuropathy - a clinical and neurophysiological study. *J Neurol Neurosurg Psychiatry* (1981) **44** pp.991-1001.

- [81] JENNINGS P.E., NIGHTINGALE S., LE GUEN C., LAWSON N., WILLIAMSON J.R., HOFFMAN P. & BARNETT A.H. Prolonged aldose reductase inhibition in chronic peripheral diabetic neuropathy: effects on microangiopathy. *Diabet Med* (1990) **7** pp.63-68.
- [82] NICOLUCCI A., CARINCI F., GRAEPEL J.G., HOHMAN T.C., FERRIS F. & LACHIN J.M. The efficacy of tolrestat in the treatment of diabetic peripheral neuropathy. A meta-analysis of individual patient data. *Diabetes Care* (1996) **19** pp.1091-1096.
- [83] SIMA A.A., GREENE D.A., BROWN M.B., HOHMAN T.C., HICKS D., GRAEPEL G.J., BOCHENEK W.J., BEG M. & GONEN B. Effect of hyperglycemia and the aldose reductase inhibitor tolrestat on sural nerve biochemistry and morphometry in advanced diabetic peripheral polyneuropathy. The Tolrestat Study Group. *J Diabetes Complications* (1993) **7** pp.157-169.
- [84] FOPPIANO M. & LOMBARDO G. Worldwide pharmacovigilance systems and tolrestat withdrawal. *Lancet* (1997) **349** pp.399-400.
- [85] GREENE D.A. & SIMA A.A. Effects of aldose reductase inhibitors on the progression of nerve damage. *Diabet Med* (1993) **10 Suppl 2** p. 31S-32S.
- [86] AREZZO J.C., KLIOZE S.S., PETERSON M.J. Efficacy and safety results of a phase II multicenter study of the aldose reductase inhibitor zopolrestat in patients with peripheral symmetrical diabetic polyneuropathy. *Diabetes* (1996) **45 Suppl. 2** p. 276A.
- [87] GREENE D.A., AREZZO J.C. & BROWN M.B. Effect of aldose reductase inhibition on nerve conduction and morphometry in diabetic neuropathy. Zenarestat Study Group. *Neurology* (1999) **53** pp.580-591.
- [88] HOTTA N., TOYOTA T., MATSUOKA K., SHIGETA Y., KIKKAWA R., KANEKO T., TAKAHASHI A., SUGIMURA K., KOIKE Y., ISHII J. & SAKAMOTO N. Clinical efficacy of fidarestat, a novel aldose reductase inhibitor, for diabetic peripheral neuropathy: a 52-week multicenter placebo-controlled double-blind parallel group study. *Diabetes Care* (2001) **24** pp.1776-1782.
- [89] ISHII A., KOTANI T., NAGAKI Y., SHIBAYAMA Y., TOYOMAKI Y., OKUKADO N., IENAGA K. & OKAMOTO K. Highly selective aldose reductase inhibitors. 1. 3-(Arylalkyl)-2,4,5-trioximidazolidine-1-acetic acids. *J Med Chem* (1996) **39** pp.1924-1927.
- [90] KURONO M., FUJIWARA I. & YOSHIDA K. Stereospecific interaction of a novel spirosuccinimide type aldose reductase inhibitor, AS-3201, with aldose reductase. *Biochemistry* (2001) **40** pp.8216-8226.
- [91] HOHMAN T.C., MENG X.U., TSE S. The ARI-509 development team: enhanced biochemical efficacy of the novel aldose reductase inhibitor, ARI-509, in human tissues. *Diabetes* (1996) **45 Suppl. 2** p. 5A.
- [92] GOTO Y., HOTTA N., SHIGETA Y., SAKAMOTO N. & KIKKAWA R. Effects of an aldose reductase inhibitor, epalrestat, on diabetic neuropathy. Clinical benefit and indication for the drug assessed from the results of a

- placebo-controlled double-blind study. *Biomed Pharmacother* (1995) **49** pp.269-277.
- [93] Hotta N., Sakamoto N., Shigeta Y., Kikkawa R. & Goto Y. Clinical investigation of epalrestat, an aldose reductase inhibitor, on diabetic neuropathy in Japan: multicenter study. Diabetic Neuropathy Study Group in Japan. *J Diabetes Complications* (1996) **10** pp.168-172.
- [94] Nakayama M., Nakamura J., Hamada Y., Chaya S., Mizubayashi R., Yasuda Y., Kamiya H., Koh N. & Hotta N. Aldose reductase inhibition ameliorates pupillary light reflex and F-wave latency in patients with mild diabetic neuropathy. *Diabetes Care* (2001) **24** pp.1093-1098.
- [95] Baba M., Kimura K., Suda T. & Yagihashi S. Three-year inhibition of aldose reductase on development of symptomatic neuropathy in diabetic patients. *J Peripher Nerv Syst* (2006) **11** pp.176-178.
- [96] Eisenmesser E.Z., Millet O., Labeikovsky W., Korzhnev D.M., Wolf-Watz M., Bosco D.A., Skalicky J.J., Kay L.E. & Kern D. Intrinsic dynamics of an enzyme underlies catalysis. *Nature* (2005) **438** pp.117-121.
- [97] Rauh D., Klebe G. & Stubbs M.T. Understanding protein-ligand interactions: the price of protein flexibility. *J Mol Biol* (2004) **335** pp.1325-1341.
- [98] Teague S.J. Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* (2003) **2** pp.527-541.
- [99] Wong C.F. & McCammon J.A. Protein flexibility and computer-aided drug design. *Annu Rev Pharmacol Toxicol* (2003) **43** pp.31-45.
- [100] Claussen H., Buning C., Rarey M. & Lengauer T. FlexE: efficient molecular docking considering protein structure variations. *J Mol Biol* (2001) **308** pp.377-395.
- [101] Osterberg F., Morris G.M., Sanner M.F., Olson A.J. & Goodsell D.S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* (2002) **46** pp.34-40.
- [102] Ferrari A.M., Wei B.Q., Costantino L. & Shoichet B.K. Soft docking and multiple receptor conformations in virtual screening. *J Med Chem* (2004) **47** pp.5076-5084.
- [103] Broughton H.B. A method for including protein flexibility in protein-ligand docking: improving tools for database mining and virtual screening. *J Mol Graph Model* (2000) **18** p. 247-57, 302-4.
- [104] Sherman W., Beard H.S. & Farid R. Use of an induced fit receptor structure in virtual screening. *Chem Biol Drug Des* (2006) **67** pp.83-84.
- [105] Radestock S., Bohm M. & Gohlke H. Improving binding mode predictions by docking into protein-specifically adapted potential fields. *J Med Chem* (2005) **48** pp.5466-5479.
- [106] Alonso H., Bliznyuk A. & Gready J. Combining docking and molecular dynamic simulations in drug

design. *Med Res Rev* (2006) .

[107] KNEGTEL R.M., KUNTZ I.D. & OSHIRO C.M. Molecular docking to ensembles of protein structures. *J Mol Biol* (1997) **266** pp.424-440.

[108] LEACH A.R. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol* (1994) **235** pp.345-356.

[109] WEI B.Q., WEAVER L.H., FERRARI A.M., MATTHEWS B.W. & SHOICHET B.K. Testing a flexible-receptor docking algorithm in a model binding site. *J Mol Biol* (2004) **337** pp.1161-1182.

[110] CARLSON H.A. & McCAMMON J.A. Accommodating protein flexibility in computational drug design. *Mol Pharmacol* (2000) **57** pp.213-218.

[111] MEAGHER K.L. & CARLSON H.A. Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as a test case. *J Am Chem Soc* (2004) **126** pp.13276-13281.

[112] BARRIL X. & MORLEY S.D. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J Med Chem* (2005) **48** pp.4432-4443.

[113] SOTRIFFER C.A., KRAMER O. & KLEBE G. Probing flexibility and "induced-fit" phenomena in aldose reductase by comparative crystal structure analysis and molecular dynamics simulations. *Proteins* (2004) **56** pp.52-66.

[114] HOWARD E.I., SANISHVILI R., CACHAU R.E., MITSCHLER A., CHEVRIER B., BARTH P., LAMOUR V., VAN ZANDT M., SIBLEY E., BON C., MORAS D., SCHNEIDER T.R., JOACHIMIAK A. & PODJARNY A. Ultrahigh resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 Å. *Proteins* (2004) **55** pp.792-804.

[115] KRAEMER O., HAZEMANN I., PODJARNY A.D. & KLEBE G. Virtual screening for inhibitors of human aldose reductase. *Proteins* (2004) **55** pp.814-823.

[116] STEUBER H., KLEBE G. Structure of JFD and BTB compounds. *Manuscript in preparation* (2006) .

[117] RAREY M., KRAMER B., LENGAUER T. & KLEBE G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* (1996) **261** pp.470-489.

[118] DA SETTIMO F., PRIMOFIORE G., LA MOTTA C., SARTINI S., TALIANI S., SIMORINI F., MARINI A.M., LAVECCHIA A., NOVELLINO E. & BOLDRINI E. Naphtho[1,2-d]isothiazole acetic acid derivatives as a novel class of selective aldose reductase inhibitors. *J Med Chem* (2005) **48** pp.6897-6907.

[119] SESTANI K., BELLINI F., FUNG S., ABRAHAM N., TREASURYWALA A., HUMBER L., SIMARD-DUQUESNE N. & DVORNIK D. N-[5-(trifluoromethyl)-6-methoxy-1-naphthalenyl]thioxomethyl]- N-methylglycine (Tolrestat), a potent, orally active aldose reductase inhibitor. *J Med Chem* (1984) **27** pp.255-256.

- [120] URZHUMTSEV A., TETE-FAVIER F., MITSCHLER A., BARBANTON J., BARTH P., URZHUMTSEVA L., BIELLMANN J.F., PODJARNY A. & MORAS D. 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** pp.601-612.
- [121] WROBEL J., DIETRICH A., GORHAM B.J., SESTANI K. Conformationally Rigid Analogues of Aldose Reductase Inhibitor, Tolrestat. Novel Syntheses of Naphtalene-Fused γ -, δ -, and ϵ -Lactams. *J Org Chem* (1990) **55** pp.2694-2702.
- [122] MORRIS G.M., GOODSSELL D.S., HALLIDAY R.S., HUEY R., HART W.E., BELEW R.K. & OLSON A.J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J Computational Chemistry* (1998) **19** pp.1639-1662.
- [123] STEUBER H., ZENTGRAF M., LA MOTTA C., DA SETTIMO F., HEINE A., KLEBE G. Manuscript in preparation. (2006) .
- [124] SOTRIFFER C.A. & DRAMBURG I. "In situ cross-docking" to simultaneously address multiple targets. *J Med Chem* (2005) **48** pp.3122-3125.
- [125] MYLARI B.L., ARMENTO S.J., BEEBE D.A., CONN E.L., COUTCHER J.B., DINA M.S., O'GORMAN M.T., LINHARES M.C., MARTIN W.H., OATES P.J., TESS D.A., WITHBROE G.J. & ZEMBROWSKI W.J. A highly selective, non-hydantoin, non-carboxylic acid inhibitor of aldose reductase with potent oral activity in diabetic rat models: 6-(5-chloro-3-methylbenzofuran-2-sulfonyl)-2-H-pyridazin-3-one. *J Med Chem* (2003) **46** pp.2283-2286.
- [126] STEUBER H., ZENTGRAF M., PODJARNY A., HEINE A. & KLEBE G. High-resolution crystal structure of aldose reductase complexed with the novel sulfonyl-pyridazinone inhibitor exhibiting an alternative active site anchoring group. *J Mol Biol* (2006) **356** pp.45-56.
- [127] FOKKENS J. & KLEBE G. A simple protocol to estimate differences in protein binding affinity for enantiomers without prior resolution of racemates. *Angew Chem Int Ed Engl* (2006) **45** pp.985-989.
- [128] Fokkens J. **Dissertation**. Philipps-Universität Marburg (2005)
- [129] Obst U. **Dissertation**. ETH Zürich (1997)
- [130] OBST U., BANNER D.W., WEBER L. & DIEDERICH F. Molecular recognition at the thrombin active site: structure-based design and synthesis of potent and selective thrombin inhibitors and the X-ray crystal structures of two thrombin-inhibitor complexes. *Chem Biol* (1997) **4** pp.287-295.
- [131] CASE D.A., PEARLMAN D.A., CALDWELL J.W., CHEATHAM III T.E., ROSS W.S., SIMMERLING C.L., DARDEN T.A., MERZ K.M., STANTON R.V., CHENG A.L., VINCENT J.J., CROWLEY M., TSUI V., RADMER R.J., DUAN Y., PITERA J., SEIBEL G.L., SINGH U.C., WEINER P K, KOLLMAN P.A. **Amber 6**, University of California, San Fran-

cisco, 1999.

[132] TRIPOS INC. **Sybyl 7.1**, , 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA, .

[133] GASTEIGER J., MARSILI M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* (1980) **36** pp.3219-3228.

[134] KLEBE G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today* (2006) **11** pp.580-594.

[135] VAN DONGEN M., WEIGELT J., UPPENBERG J., SCHULTZ J. & WIKSTROM M. Structure-based screening and design in drug discovery. *Drug Discov Today* (2002) **7** pp.471-478.

[136] KLEBE G. Recent developments in structure-based drug design. *J Mol Med* (2000) **78** pp.269-281.

[137] ROWLAND R.S. Using X-ray crystallography in drug discovery. *Curr Opin Drug Discov Devel* (2002) **5** pp.613-619.

[138] BLUNDELL T.L., JHOTI H. & ABELL C. High-throughput crystallography for lead discovery in drug design. *Nat Rev Drug Discov* (2002) **1** pp.45-54.

[139] DAVIS A.M., TEAGUE S.J. & KLEYWEGT G.J. Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. *Angew Chem Int Ed Engl* (2003) **42** pp.2718-2736.

[140] ACHARYA K.R. & LLOYD M.D. The advantages and limitations of protein crystal structures. *Trends Pharmacol Sci* (2005) **26** pp.10-14.

[141] SMYTH M.S. & MARTIN J.H. x ray crystallography. *Mol Pathol* (2000) **53** pp.8-14.

[142] ISIN E.M. & GUENGERICH F.P. Kinetics and thermodynamics of ligand binding by cytochrome P450 3A4. *J Biol Chem* (2006) **281** pp.9127-9136.

[143] Brandon C., Tooze J. **Introduction to Protein Structure**, 2. Edition, Garland Publishing, 1999.

[144] OYAMA T., MIYASITA Y., WATANABE H. & SHIRAI K. The role of polyol pathway in high glucose-induced endothelial cell damages. *Diabetes Res Clin Pract* (2006) .

[145] BROWNLEE M. Biochemistry and molecular cell biology of diabetic complications. *Nature* (2001) **414** pp.813-820.

[146] OBROSOVA I.G. Increased sorbitol pathway activity generates oxidative stress in tissue sites for diabetic complications. *Antioxid Redox Signal* (2005) **7** pp.1543-1552.

[147] YAGIHASHI S., YAMAGISHI S.I., WADA RI R., BABA M., HOHMAN T.C., YABE-NISHIMURA C. & KOKAI Y. Neuropathy in diabetic mice overexpressing human aldose reductase and effects of aldose reductase inhibitor. *Brain* (2001) **124** pp.2448-2458.

- [148] MIYAMOTO S. Recent advances in aldose reductase inhibitors: potential agents for the treatment of diabetic complications. *Expert. Opin. Ther. Patents* (2002) **12** pp.621-631.
- [149] OBROSOVA I.G., PACHER P., SZABO C., ZSENGELLER Z., HIROOKA H., STEVENS M.J. & YOREK M.A. Aldose reductase inhibition counteracts oxidative-nitrosative stress and poly(ADP-ribose) polymerase activation in tissue sites for diabetes complications. *Diabetes* (2005) **54** pp.234-242.
- [150] BERGNER A., GUNTHER J., HENDLICH M., KLEBE G. & VERDONK M. Use of Relibase for retrieving complex three-dimensional interaction patterns including crystallographic packing effects. *Biopolymers* (2001-2002) **61** pp.99-110.
- [151] GUNTHER J., BERGNER A., HENDLICH M. & KLEBE G. Utilising structural knowledge in drug design strategies: applications using Relibase. *J Mol Biol* (2003) **326** pp.621-636.
- [152] HENDLICH M., BERGNER A., GUNTHER J. & KLEBE G. Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions. *J Mol Biol* (2003) **326** pp.607-620.
- [153] STEUBER H., ZENTGRAF M., GERLACH C., HEINE A., SORTRIFFER C.A., KLEBE G. Expect the Unexpected - Backbone flips and a multi-ligand complex observed in Aldose Reductase Crystal Structures Reveal Surprising Challenges for Drug Designers. *J Mol Biol* (2006) **360** p. in press.
- [154] BERMAN H.M., WESTBROOK J., FENG Z., GILLILAND G., BHAT T.N., WEISSIG H., SHINDYALOV I.N. & BOURNE P.E. The Protein Data Bank. *Nucleic Acids Res* (2000) **28** pp.235-242.
- [155] STUBBS M.T., REYDA S., DULLWEBER F., MOLLER M., KLEBE G., DORSCH D., MEDERSKI W.W.K.R. & WURZIGER H. pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design. *Chembiochem* (2002) **3** pp.246-249.
- [156] SINGH R., WHITE M.A., RAMANA K.V., PETRASH J.M., WATOWICH S.J., BHATNAGAR A. & SRIVASTAVA S.K. Structure of a glutathione conjugate bound to the active site of aldose reductase. *Proteins* (2006) **64** pp.101-110.
- [157] CASE D.A., DARDEN T.A., CHEATHAM III T.E., SIMMERLING C.L., WANG J., DUKE R.E., LUO R., MERZ K.M., WANG B., PEARLMAN D.A., CROWLEY M., BROZELL S., TSUI V., GOHLKE H., MONGAN J., HORNAK V., CUI G., BEROZA P., SCHAFMEISTER C., CALDWELL J.W., ROSS W.S. & KOLLMAN P.A. **AMBER 8**, University of California, San Francisco, , 2004.
- [158] CORNELL W.D., CIEPLAK P., BAYLY C.I., GOULD I.R., MERZ JR. K.M., FERGUSON D.M., SPELLMEYER D.C., FOX T., CALDWELL J.W., KOLLMAN P.A. A 2nd generation force-field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* (1995) **117** p. 5179.
- [159] WANG J.M., CIEPLAK P., KOLLMAN P.A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?. *J Comp Chem* (2000)

21 pp.1049-1074.

[160] WANG J., WOLF R.M., CALDWELL J.W., KOLLMAN P.A. & CASE D.A. Development and testing of a general amber force field. *J Comput Chem* (2004) **25** pp.1157-1174.

[161] FRISCH M.J., TRUCKS G.W., SCHLEGEL H.B., SCUSERIA G.E., ROBB M.A., CHEESEMAN J.R., ZAKRZEWSKI V.G., MONTGOMERY J.A., STRATMANN R.E., BURANT J.C., DAPPRICH S., MILLAM J.M., DANIELS, A D, KUDIN, K N, STRAIN, M C, FARKAS O., TOMASI J., BARONE V., COSSI M., CAMMI R., MENNUCCI B., POMELLI C., ADAMO C., CLIFFORD S., OCHTERSKI J., PETERSSON G.A., AYALA P.Y., CUI Q., MOROKUMA K., MALICK D.K., RABUCK A.D., RAGHAVACHARI K., FORESMAN J.B., CIOSLOWSKI J., ORTIZ J.V., STEFANOV B.B., LIU G., LIASHENKO A., PISKORZ P., KOMAROMI I., GOMPERS R., MARTIN R., FOX D.J., KEITH T., AL-LAHAM M.A., PENG C.Y., NANAYAKKARA A., GONZALEZ C., CHALLACOMBE M., GILL P.M.W., JOHNSON B.G., CHEN W., WONG M.W., ANDRES J.L., HEAD-GORDON M., REPLOGLE E.S., POPLER J.A. **Gaussian98**, Gaussian, Inc., Pittsburgh, PA, 1998.

[162] BAYLY C.I., CIEPLAK P., CORNELL W.P. & KOLLMAN P.A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - The RESP Model. *J Phys Chem* (1993) **97** p. 10269.

[163] CORNELL W.P., CIEPLAK P., BAYLY C.I., KOLLMAN P.A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J Am Chem Soc* (1993) **115** p. 9620.

[164] JORGENSEN W.L., CHANDRASEKHAR J., MADURA J.P., IMPEY R.W., KLEIN M.L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* (1983) **79** p. 926.

[165] BERENDSEN H.J.C., POSTMA J.P.M., VAN GUNSTEREN W.F., DI NOLA A., HAAK J.R. Molecular dynamics with coupling to an external bath. *J Chem Phys* (1984) **81** p. 3684.

[166] RYCKAERT J.P., CICCOTTI G., BERENDSEN H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* (1977) **23** p. 327.

[167] DARDEN T., YORK D., PEDERSEN L. Particle mesh Ewald. An $N * \log(N)$ method for Ewald sums in large systems. *J Chem Phys* (1993) **98** p. 10089.

[168] ZENTGRAF M., FOKKENS J., SOTRIFFER C.A. Addressing protein flexibility and ligand selectivity by "in-situ cross-docking". *ChemMedChem* (2006) **1** p. submitted.

[169] CARPER D.A., HOHMAN T.C. & OLD S.E. Residues affecting the catalysis and inhibition of rat lens aldose reductase. *Biochim Biophys Acta* (1995) **1246** pp.67-73.

[170] GRIMSHAW C.E., BOHREN K.M., LAI C.J. & GABBAY K.H. Human aldose reductase: pK of tyrosine 48 reveals the preferred ionization state for catalysis and inhibition. *Biochemistry* (1995) **34** pp.14374-14384.

[171] LUQUE I. & FREIRE E. Structural stability of binding sites: consequences for binding affinity and al-

- losteric effects. *Proteins* (2000) **Suppl 4** pp.63-71.
- [172] HOMANS S.W. NMR spectroscopy tools for structure-aided drug design. *Angew Chem Int Ed Engl* (2004) **43** pp.290-300.
- [173] PELLECCIA M., BECATTINI B., CROWELL K.J., FATTORUSSO R., FORINO M., FRAGAI M., JUNG D., MUSTELIN T. & TAUTZ L. NMR-based techniques in the hit identification and optimisation processes. *Expert Opin Ther Targets* (2004) **8** pp.597-611.
- [174] SILLERUD L.O. & LARSON R.S. Nuclear magnetic resonance-based screening methods for drug discovery. *Methods Mol Biol* (2006) **316** pp.227-289.
- [175] VILLAR H.O., YAN J. & HANSEN M.R. Using NMR for ligand discovery and optimization. *Curr Opin Chem Biol* (2004) **8** pp.387-391.
- [176] KEMPF J.G. & LORIA J.P. Protein dynamics from solution NMR: theory and applications. *Cell Biochem Biophys* (2003) **37** pp.187-211.
- [177] KAY L.E. NMR studies of protein structure and dynamics. *J Magn Reson* (2005) **173** pp.193-207.
- [178] KARPLUS M. & MCCAMMON J.A. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* (2002) **9** pp.646-652.
- [179] WILSON D.K., TARLE I., PETRASH J.M., QUIOCHO F.A. Refined 1.8 Å structure of human aldose reductase complexed with the potent inhibitor zopolrestat. *Proc Natl Acad Sci* (1993) **90** pp.9847-9851.
- [180] KOTANI T., NAGAKI Y., ISHII A., KONISHI Y., YAGO H., SUEHIRO S., OKUKADO N. & OKAMOTO K. Highly selective aldose reductase inhibitors. 3. Structural diversity of 3-(arylmethyl)-2,4,5-trioxoimidazolidine-1-acetic acids. *J Med Chem* (1997) **40** pp.684-694.
- [181] EHRIG T., BOHREN K.M., PRENDERGAST F.G. & GABBAY K.H. Mechanism of aldose reductase inhibition: binding of NADP⁺/NADPH and alrestatin-like inhibitors. *Biochemistry* (1994) **33** pp.7157-7165.
- [182] LEE B., RICHARDS F.M. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* (1971) **55** pp.379-400.
- [183] The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* (1994) **50** pp.760-763.
- [184] MYLARI B.L., ARMENTO S.J., BEEBE D.A., CONN E.L., COUTCHER J.B., DINA M.S., O'GORMAN M.T., LINHARES M.C., MARTIN W.H., OATES P.J., TESS D.A., WITHBROE G.J. & ZEMBROWSKI W.J. A novel series of non-carboxylic acid, non-hydantoin inhibitors of aldose reductase with potent oral activity in diabetic rat models: 6-(5-chloro-3-methylbenzofuran-2-sulfonyl)-2H-pyridazin-3-one and congeners. *J Med Chem* (2005) **48** pp.6326-6339.

- [185] EVCIMEN N.D., PEKINER B., SUZEN S. & BUYUKBINGOL E. The inhibitory effect of benzodiazepine derivatives on the bovine lens aldose reductase enzyme. *Biochem Mol Biol Int* (1998) **45** pp.381-387.
- [186] Koch C., Steuber H., Heine A., Klebe G..**unpublished results**.
- [187] VELEC H.F.G., GOHLKE H. & KLEBE G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J Med Chem* (2005) **48** pp.6296-6303.
- [188] SOTRIFFER C.A. SFC Score. *Manuscript in preparation* (2006) .
- [189] KIRKWOOD J. G. Statistical Mechanics of Fluid Mixtures. *J. Phys. Chem.* (1935) **3** pp.300-313.
- [190] KOLLMAN P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem Rev* (1993) **93** pp.2395-2417.
- [191] AQVIST J., MEDINA C. & SAMUELSSON J.E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng* (1994) **7** pp.385-391.
- [192] GUO Z., BROOKS III C.L., KONG X. Efficient and Flexible Algorithm for Free Energy Calculations Using the I-Dynamics Approach. *J Phys Chem B* (1998) **102** pp.2032-2036.
- [193] KOLLMAN P.A., MASSOVA I., REYES C., KUHN B., HUO S., CHONG L., LEE M., LEE T., DUAN Y., WANG W., DONINI O., CIEPLAK P., SRINIVASAN J., CASE D.A. & CHEATHAM T.E.3. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res* (2000) **33** pp.889-897.
- [194] MASSOVA I. & KOLLMAN P.A. Computational Alanine Scanning to Probe Protein-Protein Interactions: A novel Approach to evaluate Binding Free Energies. *J Am Chem Soc* (1999) **121** pp.8133-8143.
- [195] LUO C., XU L., ZHENG S., LUO X., SHEN J., JIANG H., LIU X. & ZHOU M. Computational analysis of molecular basis of 1:1 interactions of NRG-1beta wild-type and variants with ErbB3 and ErbB4. *Proteins* (2005) **59** pp.742-756.
- [196] KUHN B. & KOLLMAN P.A. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J Med Chem* (2000) **43** pp.3786-3791.
- [197] WANG J., MORIN P., WANG W., KOLLMAN P.A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of Efavirenz by docking and MM-PBSA. *J Am Chem Soc* (2001) **123** pp.5221-5230.
- [198] HUO S., WANG J., CIEPLAK P., KOLLMAN P.A. & KUNTZ I.D. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. *J Med Chem* (2002) **45**

pp.1412-1419.

- [199] SUENAGA A., HATAKEYAMA M., ICHIKAWA M., YU X., FUTATSUGI N., NARUMI T., FUKUI K., TERADA T., TAJI M., SHIROUZU M., YOKOYAMA S. & KONAGAYA A. Molecular dynamics, free energy, and SPR analyses of the interactions between the SH2 domain of Grb2 and ErbB phosphotyrosyl peptides. *Biochemistry* (2003) **42** pp.5195-5200.
- [200] LAITINEN T., KANKARE J.A. & PERAKYLA M. Free energy simulations and MM-PBSA analyses on the affinity and specificity of steroid binding to antiestradiol antibody. *Proteins* (2004) **55** pp.34-43.
- [201] KUHN B., GERBER P., SCHULZ-GASCH T. & STAHL M. Validation and use of the MM-PBSA approach for drug discovery. *J Med Chem* (2005) **48** pp.4040-4048.
- [202] STEINBRECHER T., CASE D.A., LABAHN A. A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. *J Med Chem* (2006) **49** pp.1837-1844.
- [203] PEARLMAN D.A. Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J Med Chem* (2005) **48** pp.7796-7807.
- [204] GOUDA H., KUNTZ I.D., CASE D.A. & KOLLMAN P.A. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers* (2003) **68** pp.16-34.
- [205] LEE M.R., DUAN Y. & KOLLMAN P.A. Use of MM-PB/SA in estimating the free energies of proteins: application to native, intermediates, and unfolded villin headpiece. *Proteins* (2000) **39** pp.309-316.
- [206] SRINIVASAN J., CHEATHAM T.E., CIEPLAK P., KOLLMAN P.A., CASE D.A. Continuum solvent studies of the stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J Am Chem Soc* (1998) **120** pp.9401-9409.
- [207] SRINIVASAN J., MILLER J., KOLLMAN P.A. & CASE D.A. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J Biomol Struct Dyn* (1998) **16** pp.671-682.
- [208] CASE D.A., CHEATHAM, III, T E, DARDEN T., GOHLKE H., LUO R., MERZ K.M., ONUFRIEV A., SIMMERLING C., WANG B. & WOODS R. The Amber biomolecular simulation programs. *J. Computat. Chem.* (2005) **26** pp.1668-1688.
- [209] GOHLKE H., KIEL C. & CASE D.A. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J Mol Biol* (2003) **330** pp.891-913.
- [210] GOHLKE H. & CASE D.A. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J Comput Chem* (2004) **25** pp.238-250.
- [211] LEE M.C., YANG R. & DUAN Y. Comparison between Generalized-Born and Poisson-Boltzmann meth-

- ods in physics-based scoring functions for protein structure prediction. *J Mol Model (Online)* (2005) **12** pp.101-110.
- [212] BROOKS B.R., JANEZIC D., KARPLUS M. Harmonic Analysis of Large Systems. *J Comp Chem* (1995) **16** pp.1522-1553.
- [213] JANEZIC D., BROOKS B.R. Harmonic Analysis of Large Systems. II. comparison of different protein models. *J Comp Chem* (1995) **16** pp.1543-1553.
- [214] JANEZIC D., VENABLE R.M., BROOKS B.R. Harmonic analysis of large systems. III. comparison with molecular dynamics. *J Comput Chem* (1995) **16** pp.1554-1566.
- [215] ROCCATANO D., AMADEI A., APOL M.E.F. & NOLA, A D, BERENDSEN, H J C Application of the quasi-Gaussian entropy theory to molecular dynamics simulations of Lennard-Jones Fluids. *J Chem Phys* (1998) **109** pp.6358-6363.
- [216] SMITH K.C. & HONIG B. Evaluation of the conformational free energies of loops in proteins. *Proteins* (1994) **18** pp.119-132.
- [217] YANG A.S. & HONIG B. Free energy determinants of secondary structure formation: II. Antiparallel beta-sheets. *J Mol Biol* (1995) **252** pp.366-376.
- [218] YANG A.S. & HONIG B. Free energy determinants of secondary structure formation: I. alpha-Helices. *J Mol Biol* (1995) **252** pp.351-365.
- [219] YANG A.S., HITZ B. & HONIG B. Free energy determinants of secondary structure formation: III. beta-turns and their role in protein folding. *J Mol Biol* (1996) **259** pp.873-882.
- [220] ÖSAPAY K., YOUNG W., BASHFORD D., BROOKS, C L , III & CASE D.A. Dielectric continuum models for hydration effects on peptide conformational transitions. *J Phys Chem* (1996) **100** pp.2698-2705.
- [221] BASHFORD D., CASE D.A., CHOI C., GIPPERT G.P. A Computational Study of the Role of Solvation Effects in Reverse Turn Formation in the Tetrapeptides APGD and APGN. *J Am Chem Soc* (1997) **119** pp.4964-4971.
- [222] DEMCHUK E., BASHFORD D., GIPPERT G., CASE D.A. Thermodynamics of a reverse turn motif. Solvent effects and side-chain packing. *J Mol Biol* (1997) **270** pp.305-317.
- [223] CHEATHAM T.E.3. & KOLLMAN P.A. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J Mol Biol* (1996) **259** pp.434-444.
- [224] WARWICKER J., WATSON H.C. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *J Mol Biol* (1982) **157** pp.671-679.
- [225] SHARP K., HONIG B. Electrostatic Interactions in Macromolecules: Theory and Applications. *Ann Rev*

Biophys Biophys Chem (1990) **19** pp.301-332.

[226] STILL W.C., TEMPczyk A., HAWLEY R.C. & HENDRICKSON T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* (1990) **112** pp.6127-6129.

[227] BEA I., GOTSEV M.G., IVANOV P.M., JAIME C. & KOLLMAN P.A. Chelate effect in cyclodextrin dimers: a computational (MD, MM/PBSA, and MM/GBSA) study. *J Org Chem* (2006) **71** pp.2056-2063.

[228] WANG J., DIXON R. & KOLLMAN P.A. Ranking ligand binding affinities with avidin: a molecular dynamics-based interaction energy study. *Proteins* (1999) **34** pp.69-81.

[229] TODOROV N.P. & DEAN P.M. Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J Comput Aided Mol Des* (1997) **11** pp.175-192.

[230] TODOROV N.P. & DEAN P.M. A branch-and-bound method for optimal atom-type assignment in de novo ligand design. *J Comput Aided Mol Des* (1998) **12** pp.335-349.

[231] PEARLMAN D.A. & CHARIFSON P.S. Are free energy calculations useful in practice? A comparison with rapid scoring functions for the p38 MAP kinase protein system. *J Med Chem* (2001) **44** pp.3417-3423.

[232] PEARLMAN D.A. Free energy grids: a practical qualitative application of free energy perturbation to ligand design using the OWFEG method. *J Med Chem* (1999) **42** pp.4313-4324.

[233] BLAKELEY M., MITSCHLER A., HAZEMANN I., MEILLEUR F., MYLES D. & PODJARNY A. Comparison of hydrogen determination with X-ray and neutron crystallography in a human aldose reductase-inhibitor complex. *Eur Biophys J* (2006) .

[234] KLEBE G., KRAMER O. & SOTRIFFER C. Strategies for the design of inhibitors of aldose reductase, an enzyme showing pronounced induced-fit adaptations. *Cell Mol Life Sci* (2004) **61** pp.783-793.

[235] IWATA Y., ARISAWA M., HAMADA R., KITA Y., MIZUTANI M.Y., TOMIOKA N., ITAI A. & MIYAMOTO S. Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. *J Med Chem* (2001) **44** pp.1718-1728.

[236] EL-KABBANI O., RAMSLAND P., DARMANIN C., CHUNG R.P. & PODJARNY A. Structure of human aldose reductase holoenzyme in complex with statil: an approach to structure-based inhibitor design of the enzyme. *Proteins* (2003) **50** pp.230-238.

[237] RASTELLI G., FERRARI A.M., COSTANTINO L. & GAMBERINI M.C. Discovery of new inhibitors of aldose reductase from molecular docking and database screening. *Bioorg Med Chem* (2002) **10** pp.1437-1450.

[238] IWATA Y., NAITO S., ITAI A. & MIYAMOTO S. Protein structure-based de novo design and synthesis of aldose reductase inhibitors. *Drug Des Discov* (2001) **17** pp.349-359.

[239] STEUBER H., ZENTGRAF M., GERLACH C., SOTRIFFER C.A., HEINE A., KLEBE G. Expect the Unexpected or

Caveat for Drug Designers: Multiple Structure Determinations Using Aldose Reductase Crystals Treated under varying Conditions. *J Mol Biol* (2006) **360** p. submitted.

[240] SARGES R., BORDNER J., DOMINY B.W., PETERSON M.J. & WHIPPLE E.B. Synthesis, absolute configuration, and conformation of the aldose reductase inhibitor sorbinil. *J Med Chem* (1985) **28** pp.1716-1720.

[241] BARSKI O.A., GABBAY K.H., GRIMSHAW C.E. & BOHREN K.M. Mechanism of human aldehyde reductase: characterization of the active site pocket. *Biochemistry* (1995) **34** pp.11264-11275.

[242] MYLARI B.L., LARSON E.R., BEYER T.A., ZEMBROWSKI W.J., ALDINGER C.E., DEE M.F., SIEGEL T.W. & SINGLETON D.H. Novel, potent aldose reductase inhibitors: 3,4-dihydro-4-oxo-3-[[5-(trifluoromethyl)-2-benzothiazolyl] methyl]-1-phthalazineacetic acid (zopolrestat) and congeners. *J Med Chem* (1991) **34** pp.108-122.

[243] HONIG B., NICHOLLS A. Classical Electrostatics in Biology and Chemistry. *Science* (1995) **268** pp.1144-1149.

[244] SITKOFF D., SHARP K.A., HONIG B. Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models. *J Phys Chem* (1994) **98** pp.1978-1988.

[245] Papula L. **Mathematik für Ingenieure und Naturwissenschaftler**, Band 3, Vieweg, 2001.

[246] SHOICHET B.K. Virtual screening of chemical libraries. *Nature* (2004) **432** pp.862-865.

[247] KITCHEN D.B., DECORNEZ H., FURR J.R. & BAJORATH J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* (2004) **3** pp.935-949.

[248] SCHMIDT A. & LAMZIN V.S. Veni, vidi, vici - atomic resolution unravelling the mysteries of protein function. *Curr Opin Struct Biol* (2002) **12** pp.698-703.

[249] HARRISON D.H., BOHREN K.M., PETSKO G.A., RINGE D. & GABBAY K.H. The alrestatin double-decker: binding of two inhibitor molecules to human aldose reductase reveals a new specificity determinant. *Biochemistry* (1997) **36** pp.16134-16140.

[250] CASE D.A., DARDEN T.A., CHEATHAM III T.E., SIMMERLING C.L., WANG J., DUKE R.E., LUO R., MERZ K.M., PEARLMAN D.A., CROWLEY M., WALKER R.C., ZHANG W., WANG B., HAYIK S., ROITBERG A., SEABRA G., WONG K.F., PAESANI F., WU X., BROZELL S., TSUI V., GOHLKE H., YANG L., TAN C., MONGAN J., HORNAK V., CUI G., BEROZA P., MATHEWS D.H., SCHAFMEISTER C., ROSS W.S., KOLLMAN P.A. **Amber 9**, University of California, San Francisco, , 2006.

[251] DELANO W.L. The PyMol Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA.

[252] HUMPHREY W., DALKE A. & SCHULTEN K. VMD: visual molecular dynamics. *J Mol Graph* (1996) **14** p.

33-8, 27-8.

[253] LAMOUR V., BARTH P., ROGNIAUX H., POTERSZMAN A., HOWARD E., MITSCHLER A., VAN DORSSELAER A., PODJARNY A. & MORAS D. Production of crystals of human aldose reductase with very high resolution diffraction. *Acta Crystallogr D Biol Crystallogr* (1999) **55** (Pt 3) pp.721-723.

[254] EL-KABBANI O., DARMANIN C., SCHNEIDER T.R., HAZEMANN I., RUIZ F., OKA M., JOACHIMIAK A., SCHULZEBRIESE C., TOMIZAKI T., MITSCHLER A. & PODJARNY A. Ultrahigh resolution drug design. II. Atomic resolution structures of human aldose reductase holoenzyme complexed with Fidarestat and Minalrestat: implications for the binding of cyclic imide inhibitors. *Proteins* (2004) **55** pp.805-813.

[255] OTWINOWSKI Z., MINOR W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* (1997) **276** pp.307-326.

[256] CALDERONE V., CHEVRIER B., VAN ZANDT M., LAMOUR V., HOWARD E., POTERSZMAN A., BARTH P., MITSCHLER A., LU J., DVORNIK D.M., KLEBE G., KRAEMER O., MOORMAN A.R., MORAS D. & PODJARNY A. The structure of human aldose reductase bound to the inhibitor IDD384. *Acta Crystallogr D Biol Crystallogr* (2000) **56** (Pt 5) pp.536-540.

[257] BRUNGER A.T., ADAMS P.D., CLORE G.M., DELANO W.L., GROS P., GROSSE-KUNSTLEVE R.W., JIANG J.S., KUSZEWSKI J., NILGES M., PANNU N.S., READ R.J., RICE L.M., SIMONSON T. & WARREN G.L. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* (1998) **54** (Pt 5) pp.905-921.

[258] SHELDRIK G.M. & SCHNEIDER T. SHELXL: High Resolution Refinement. *Methods Enzymol* (1997) **277** pp.319-343.

[259] JONES T.A., ZOU J.Y. & COWAN S.W. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* (1991) **47** (Pt 2) pp.110-119.

[260] LASKOWSKI R., MACARTHUR M., MOSS D., THORNTON J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* (1993) **26** pp.283-291.

Publications arising from this work

Articles

Zentgraf M., Steuber H., Sotriffer C., Heine A., Klebe G.,
How Reliable are Current Docking Approaches for Structure-based Drug Design?
Angewandte Chemie, accepted

Zentgraf M., Fokkens J., Sotriffer C.,
Addressing protein flexibility and ligand selectivity by 'in-situ cross-docking'
ChemMedChem, in press

Zentgraf M., Steuber H., Klebe G., Sotriffer C.,
Extending charted space: comparative MD simulations of Aldose Reductase
manuscript in preparation

Zentgraf M., Steuber H., Klebe G., Sotriffer C.,
Evaluating MM-PBSA in case of a flexible binding pocket: the AR test case
manuscript in preparation

Steuber H., Zentgraf M., Podjarny A., Heine A., Klebe G.,
High-resolution Crystal Structure of Aldose Reductase Complexed with the Novel Sulfonyl-pyridazinone Inhibitor Exhibiting an Alternative Active Site Anchoring Group
JMB, 356(1), 45-56, **2006**

Steuber H., Zentgraf M., Gerlach C., Sotriffer C., Heine A., Klebe G.,
Expect the Unexpected or Caveat for Drug Designers: Multiple Structure Determinations Using Aldose Reductase Crystals Treated under varying Soaking and Cocrystallization Conditions
JMB, in press, **2006**

Oral Communications

Zentgraf M., Sotriffer C., Klebe G.,

Charting Aldose Reductase' conformational space: comparative analysis of MD simulations

University of California, San Francisco, USA, **2006**

Zentgraf M., Sotriffer C., Steuber H., Klebe G.,

Thermodynamic characterization of Aldose Reductase inhibitors by comparative MD simulations

Open Mike Session, Gordon Research Conference on Computer Aided Drug Design, Tilton, NH, USA, **2005**

Zentgraf M., Sotriffer C., Klebe G.,

Simultaneous Crossdocking to Aldose Reductase Conformers using AutoDock

DPhG Doktorandentagung, Freudenstadt-Lauterbad, Germany, **2004**

Zentgraf M., Sotriffer C., Klebe G.,

"in-situ" Crossdocking to Aldose Reductase Conformers using AutoDock

meeting with BASF AG, Ludwigshafen, Germany, **2004**

Zentgraf M., Sotriffer C., Klebe G.,

"in-situ" Crossdocking to multiple Aldose Reductase Conformers using AutoDock

meeting with BiosolveIT GmbH, Sankt Augustin, Germany, **2004**

Posters

Zentgraf M., Sotriffer C., Steuber H., Klebe, G.,

Thermodynamic characterization of Aldose Reductase Inhibitors by comparative MD simulations

Gordon Research Conference on Computer Aided Drug Design, Tilton, NH, USA, **2005**

Zentgraf M., Sotriffer C., Klebe G.,

“In-situ” cross-docking into different protein conformers of Aldose Reductase

High Resolution Drug Design Meeting, Bischberg-Strasbourg, France, **2004**

Zentgraf M., Sotriffer C., Klebe G.,

Addressing protein flexibility: Simultaneous cross-docking into different protein conformers of Aldose Reductase

Third Aventis [I]-lab workshop on Chemical Biology, Eberbach Monastery, Germany, **2004**

Acknowledgments

Numerous people have contributed to this work in one way or another. Although I cannot list them all, I am grateful to all of them. Especially, I would like to thank the following people:

- *Prof. Dr. Gerhard Klebe* for the opportunity to do my PhD thesis under his supervision and for the many things I learned from him. I feel privileged to have worked in his group.
- *Dr. Christoph Sotriffer* for his never-ending support concerning all aspects of my thesis and more. I have never seen a person so skillful and passionate concerning his work and at the same time so patient and helpful with other people's problems and issues.
- My girlfriend, *Isabell Engelhardt*, for her love, patience, and invaluable support throughout my entire PhD time.
- My parents, *Christa and Bernhard Zentgraf*, to which I am grateful for so many things that I do not even try to list them here. This PhD thesis would never exist without their continuous support.
- *Katrin Silber* for a great time in A 203. She has set the highest possible standard for every future office mate.
- *Nils Weskamp* for many stimulating discussions about science and things we could never sell as science at all.
- *Holger Steuber* for the nice and fruitful collaboration on the aldose reductase project.
- *Michael Eisenmann* from the group of *Prof. Dr. Martin Schlitzer* for synthesizing molecules for the aldose reductase project.
- *Dr. Peter Block* and *Hans Velec* for helping me to find my way through the jungle of programming and for introducing me to the wonderful world of system administration.

-
- *Dr. Andreas Heine* and *Dr. Peter Haebel* for their help and supervision on my trip to the realm of solving and refining X-ray protein structures.
 - *Angela Scholz* for keeping the AG Klebe world turning every day.
 - The whole *AG Klebe* group for the nice and stimulating working atmosphere. In many ways it is a assembly of truly exceptional people.
 - The Graduiertenkolleg '*Proteinfunktion auf atomarer Ebene*' for funding parts of this work and for some nice seminar trips to Hirschegg.

Curriculum vitae

- Full name:** Matthias Christian Bernd Zentgraf
- Date of birth:** October, 21 1976
- Place of birth:** Fulda, Germany
-
- 2003- 2006** PhD student at the Institute of Pharmaceutical Chemistry, Philipps-Universität Marburg, in the group of *Prof. Dr. Gerhard Klebe, Research group for drug design & X-ray crystallography*
-
- 2003-2005** Fellow of the Graduiertenkolleg "*Protein function at the atomic level*"
-
- 2002** State examination and licensure as pharmacist
-
- 2001-2002** Obligatory practical training student in pharmacy (research and trade)
- First half (public pharmacy): "Kosmas Apotheke", Cologne.
Second half (research): Department for galenics, Bayer AG, Leverkusen
-
- 1997-2001** Studies of pharmacy, Philipps-Universität, Marburg
-
- 1987-1996** Rabanus-Maurus Schule (Domgymnasium), Fulda
- Abitur (majors: chemistry and english)
-
- 1983-1987** Rauschenbergschule, Petersberg

Erklärung

Ich versichere, dass ich meine Dissertation

Characterization of Binding Pocket Flexibility of Aldose Reductase

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlich Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 04. 10. 2006