Predicting Dynamics and Relative Free Energies of Binding of Four Inhibitors to Squalene-Hopene Cyclase

Single Step Perturbation *versus* Thermodynamic Integration with Squalene-Hopene Cyclase

Fabienne Schwab *

Advisors:
Bojan Žagrović
Wilfred F. van Gunsteren

Master Thesis

July 24, 2006

Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland
** Contact: fschwab@student.ethz.ch*

I would like to thank Bojan Žagrović, who was my patient advisor during the last four months, Daan Geerke and Clara Christ for their great help and Nathan Schmid who encouraged me to learn Perl programming.

ABSTRACT

The membrane protein squalene-hopene cyclase (SHC) is homologous to a human enzyme responsible for cholesterol formation. Protein-inhibitor binding was investigated to help find new proposals for anticholesteremic inhibitors. As a series of structurally similar inhibitors of SHC is known, the protein lends itself to relative free energy calculations *via* single step perturbation (SSP) and *via* thermodynamic integration (TI), whose results can be directly compared with experimental data. Both methods were realized and compared in this study.

SHC was simulated in water with GROMOSXX and the force field parameter set 45A3, once in complex with an inhibitor and once in an uncomplexed form. The analysis of these simulations showed that although the protein was not simulated in its native membrane environment, it exhibits significant stability with respect to secondary and tertiary structure. Furthermore, the dynamics of the protein, as measured by atomic-positional root mean square fluctuations, is in accordance with the experimental temperature factors. These findings justified further simulations and dynamics analysis of SHC in water. Finally, analysis of the fluctuations in the complexed and the uncomplexed protein structure affirmed suggestions for the entrance channel and revealed a novel proposition for the release channel.

The relative free energies of binding of four anticholesteremic inhibitors were calculated using TI and SSP. Comparison of the results of both methods showed that TI reproduces qualitatively the relative order of the experimental affinities, but SSP does not. However, it should be emphasized that relative free energy calculations using SSP could not be evaluated with sufficient stringency, due to the fact that in the limitated time-frame of this work, technical obstacles had to be overcome, mainly related to the size of the simulated system and the complexity of the setup for the SSP. As a consequence, the analysis was carried out on a data set of limited size, mandating that the results have to be interpreted with due caution. Since most of these challenges could be worked out during this study, the performance and analysis of MD simulations, TI, and SSP with SHC is now feasible in a much shorter time.

Furthermore, through SSP, some of the energetically most favoured conformations of four known inhibitors in complex with the SHC could be found. They deliver insight into the binding mode of the inhibitor.

CONTENTS

1. INTRODUCTION

The membrane protein squalene-hopene cyclase (SHC) catalyzes cyclisation from squalene to hopene or diplopterol in *Acyclobacillus acidocaldarius*. It is the procaryotic counterpart of the human enzyme responsible for cholesterol formation [1]. Because of the clear sequence homology, SHC is an attractive target in drug research to find new anticholesteremic inhibitors.
A series of structurally similar inhibitors and their inhibitor potencies (IC_{50} constants) of SHC is known [2]. This provides the ideal preconditions for relative free energy calculations *via* the methods single step perturbation (SSP) and *via* thermodynamic integration (TI).

1.1 Why Use Computer Simulation for Inhibitor Research?

Although the quest for novel inhibitors is crucial in drug research, especially in the field of cancer research, it is still laborious, expensive and time-consuming to find them. Most inhibitors are still found by trial-and-error or by evaluating thousands of assays.
Two main approaches exist to make the search for inhibitors more systematic:

1. Screen thousands of potential inhibitors (inhibitor libraries), for example by applying the method biology-oriented synthesis (BIOS).

2. Analyze the binding structures of existing protein-inhibitor complexes with X-ray diffraction or nuclear magnetic resonance spectroscopy to get more insight into the form of the inhibitor binding pocket.

A disadvantage of the first approach is the ignorance about the binding mode of the inhibitor. A drawback of the second approach is the lack of knowledge about the dynamics of the protein.
Combination with computer simulation enhances the efficiency of both methods, since it provides insight into the dynamics and the binding mode of the inhibitors that would otherwise stay inaccessible.

1.2 Goals of this Study

In the writing of the present thesis, focus has has been directed towards three parallel goals:

- Describe the scientific results of the molecular dynamics simulations of the protein squalene-hopene cylase (SHC) and for the calculation of the dynamics and relative free energies of binding of four inhibitors to SHC.
- Compare relative free energy calculations *via* two methods: single step perturbation (SSP) and thermodynamic integration (TI).
- Document the research process that has led to the above results, including methods, lessons learned, mistakes made and challenges overcome.

1.3 Squalene-Hopene Cyclase

Function of the SHC: High blood cholesterol levels greatly increase the risk of heart disease. The average person produces about 75 % of blood cholesterol in his or her liver, while only about 25 % is absorbed from food [3]. Thus, most approaches to reduce the blood cholesterol level aim to reduce the cholesterol biosynthesis in the liver: in other words, they aim to inhibit the enzymes that catalyze the cholesterol-producing reaction steps.
SHC is a monotopic[1] membrane protein that catalyzes a cationic cyclisation cascade from squalene to hopene or diplopterol in *A. acidocaldarius*. It is homologous to the human triterpene cyclase 2,3-oxidosqualenecyclase (OSC) which catalyzes a key step in cholesterol biosynthesis, the reaction of 2,3-oxidosqualene to lanosterol, a precursor of cholesterol and steroid hormones (fig. 1.1). It is shown that inhibitors of SHC are therefore potentially anticholesteremic drugs [1].

Structure, educt access and product release: The structure of SHC has been determined in crystal form at 2.0 Å resolution and refined to an R-factor of 15.3 % (R_{free} = 18.7 %) [1]. The pathways of the flexible educt squalene from the membrane interior to the acitve center cavity and of the rigid fused-ring product hopene in the reverse direction were investigated, but are still controversial. Educt access happens presumably through a nonpolar channel

[1] Integral membranes that submerge from one side into the nonpolar part of the phospholipid bilayer have been classified as monotopic [1].

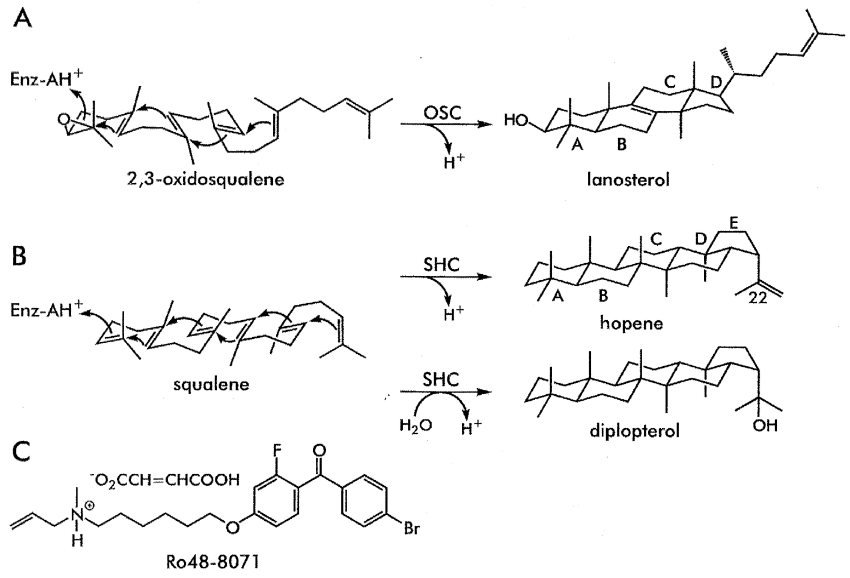


Fig. 1.1: Examples for the reactions catalyzed by triterpene cyclases. A: Bacterial SHC converts squalene to hopene or diplopterol. B: Human 2,3-oxidosqualene cyclase transforms squalene to the cholesterol precursor lanosterol. C: Example for a typical structure of one of the most binding inhibitors for SHC, Ro48.8071 [4].

(fig. 1.3) that contains a narrow constriction[2]. The constriction seems to be still mobile enough to permit the passage of the educt in an extended conformation. Product release *via* the same channel appears to be questionable, because these fused ring compounds are rigid and bulky and require a much larger opening of the channel [1].
The protein shows a highly lipophilic plateau around the opening of the entrance channel, where it could possibly bind to the membrane. In a model that has been proposed for SHC-binding to the membrane, the plateau region, presumabely plunged into the nonpolar membrane interior, consists of two alpha helices and three loops (about 15 % of the protein) [5], [6].
The dimeric structure of SHC in crystal form has been estabilshed by gel permeation. The binding interface of the dimer lies close to the lipophilic plateau plunged into the membrane. With its high polarity the dimer interface rather resembles crystal contacts. It is therefore conceivable that the

[2] From now on, the protein will be displayed like in fig. 1.3 such that the region of the protein which contains the amine site of the inhibitor will be called the "top" of the molecule.

SHC dimer is of a transient nature. *In vivo*, the nonpolar parts of the protein next to the interface could be occupied by phospholipids, which shield the polar interface center against water and thus promote the association.
SHC contains two subdomains, both with a very similar structure. This may be a result of a former gene duplication. Both subdomains consist of a $\alpha_6 - \alpha_6$ barrel and are highly stabilized through QW motifs (Q for glutamine, W for tryptophan) in the outer ring. Due to these barrel structures, SHC gains much stability [1], [5].

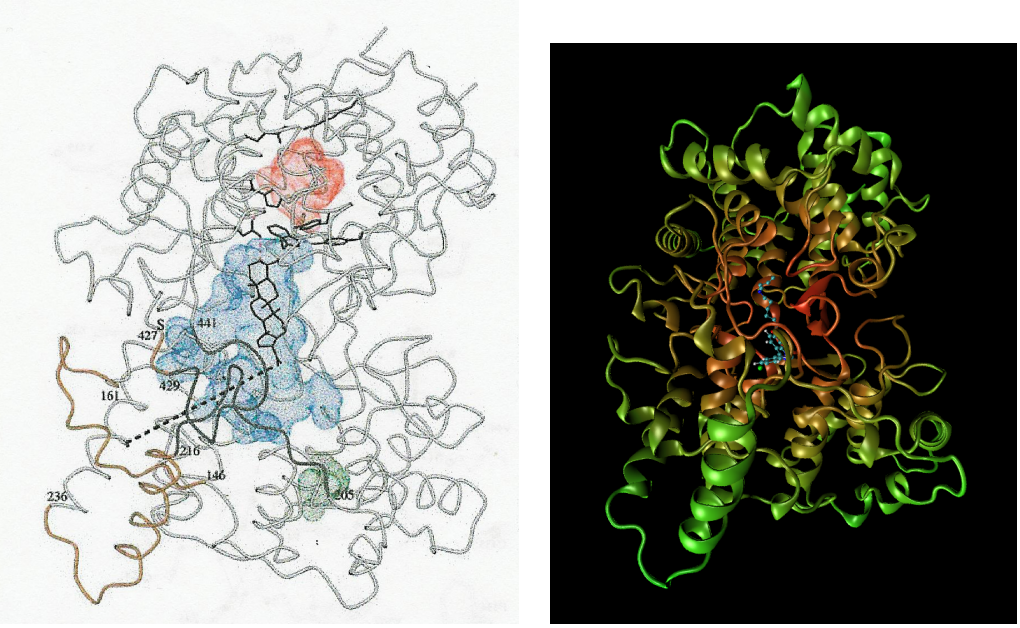


Fig. 1.2: Modeled SHC molecule. Left side: The black indicated loops show the supposed entrance channel and the interface of the protein to the second SHC in the dimer which ends up in the region that likely plunges into the membrane. The active cavity is indicated in blue, the other two cavities in red and green. The dotted line marks the supposed entrance channel [1]. Right side: SHC, visualized with vmd. Note the numerous α-helices of the protein.

Cavities: The protein contains three cavities (fig. 1.3): The main cavity is linked to the entrance channel for the substrate at the lipophilic plateau of the protein that directs into the membrane (it is assumed that the educt

squalene comes out of the nonpolar interior of the membrane directly into the protein). The next two small cavities are *vis-à-vis* of the big cavity in the middle of the two $\alpha_6 - \alpha_6$ barrels and face the water with respect to the second SHC protein in the dimer. Their role in the reaction is still unknown; apart from stabilizing the structure of the enzyme, they may have no function at all [1].

Reaction catalyzed by SHC: The reaction catalyzed by SHC is a cationic cyclisation cascade from squalene to hopene or diplopterol (fig. 1.1). It is a highly exergonic reaction; hopene formation releases about 200 kJ/mol. This exceeds by far the usual protein stabilisation energy of about 50 kJ/mol and provides a rationale for the high stability of the protein which is a consequence of the QW-motifs in the α-helices.
The cationic all-chair transition state is very sensitive to nucelophilic attack. It is therefore assumed that SHC plays mainly a protecting role to prevent access of water or deprotonation by a base [7].

Reactive center and transition state: Conformations of the side chains around the inhibitor are nearly identical in all experimentally estabilshed structures, indicating a rigid active center [6].
The 1200 $Å^3$ active site cavity is mainly nonpolar, but it has a highly polar patch at the top with the sequence motif DXDD (D for Aspartate, X for any amino acid). There, the initially protonating base is presumed. The second base that accepts a proton must be somewhere at the bottom of the cavity. The nonpolar rest of the reactive center is lined by aromatic amino acid residues which are supposed to stabilize the cationic transition state by cation-π interactions.
The position of squalene in the reactive center is still controversial. There is one model based on photoaffinity labeling [8] and a second model based on X-ray diffraction measurements [6]. The second model seems to be more likely because it explains the position of the inhibitor by the interactions of its protonated amino group with the amino acids catalyzing the protonation of squalene. Still, it may not be correct, because the same researchers found for the inhibitor a noncompetitive inhibition pattern which they explained with technical difficulties. The assay had to be performed in a biphasic system which may have caused artifacts in the measurement [6], [8].
The most important tasks for research are binding structure analysis and predictions to help find new proposals for anticholesteremic inhibitors [7].

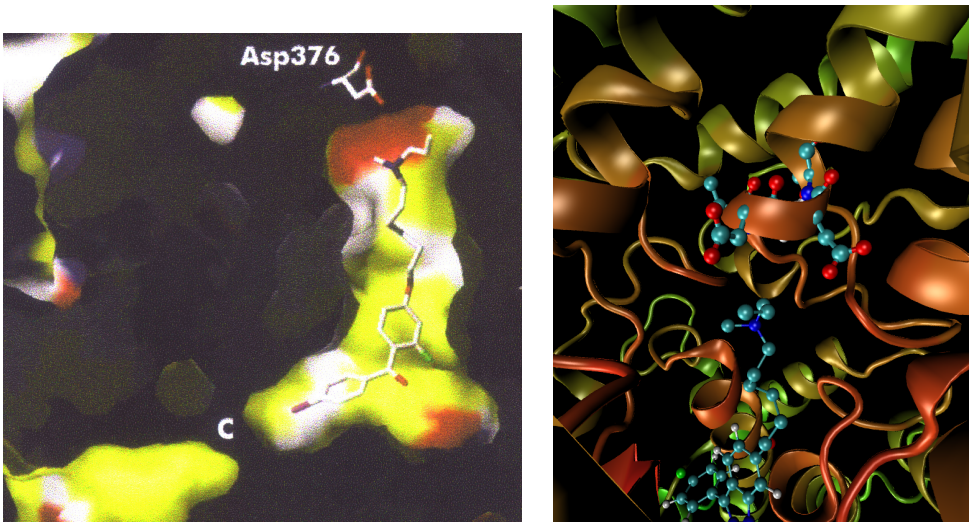


Fig. 1.3: Picture of the reactive cavity. Left side: Surface plot from [5]. Apolar sites are yellow, polar sites are red. The channel (C) to the reactive cavity is closed here, but is supposed to open readily. Right side: Modeled SHC. At the top in ball and stick representation the polar DVDD sequence that is supposed to initiate the cyclisation, at the bottom the inhibitor molecule. Note the short distance between the oxygen of the Asp376 and the amine of the ligand.

Differences between the two $\alpha_6 - \alpha_6$ barrels: The two $\alpha_6 - \alpha_6$-barrels show differences in their behaviour: the barrel that is situated closer to the complexed ligand stays more compact and better arranged than the barrel *vis-à-vis* that is supposed to be half plunged into the membrane interior. This may be an evolutionary consequence of the fact that the first barrel is closer to the reactive cavity and might therefore have to absorb most of the reaction energy of 200 kJ/mol of the hopene formation.

1.4 Inhibitors for SHC

Lenhart *et al.* [2] describe 28 oxidosqualene cyclase (OSC) inhibitors with a very similar structure bound to SHC (fig. 1.4). They mimic the transition

state of the cationic cyclisation. All of them contain an amine and an aromatic system, mostly built out of two linked phenyl rings substituted with one or two halogen atoms. Furthermore, an ether oxygen is present in all of the ligands.
The ligands nr. 9, 12, 13, 19 and 22 from Lenhart *et al.* were chosen to be simulated during this study, since they assemble the biggest set of ligands with the fewest structural differences. The experimental inhibitor potencies are listed in table 1.1 [2].

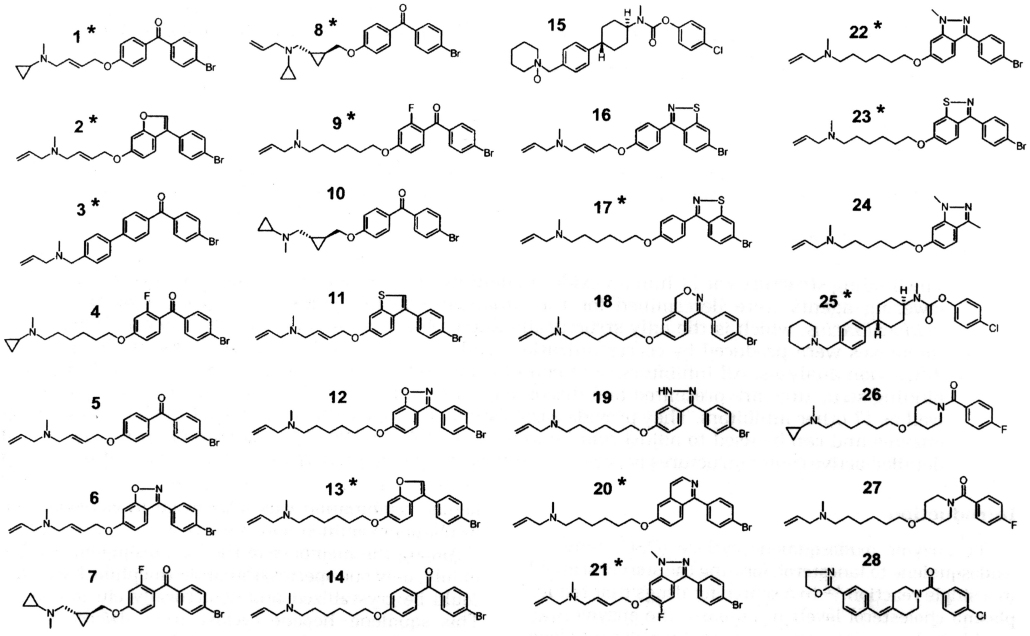


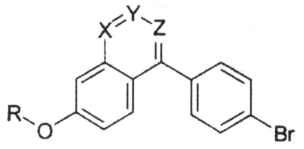
Fig. 1.4: Chemical formulas for 28 inhibitors. Binding structures with established crystal structure are marked with an asterisk (*) [2]. The ligands are ordered according to their affinity; ligand nr. 1 binds best.

1.5 Free Energy Calculations

Free energy can be seen as the driving force of all molecular processes. The Gibbs free enthalpy, $G(N, p, T)$, of a protein-inhibitor complex is calculated from the isothermal-isobaric partition function $Z(N, p, T)$,

$$G(N, p, T) = A(N, V, T) + pV = -k_B T ln Z(N, p, T). \tag{1.1}$$

N is the number of particles, p is the pressure, T is the temperature, V is the volume, k_B is the Boltzmann constant, and $A(N, V, T)$ is the Helmholtz



inhibitors[a]	atom[b] X	Y	Z	IC_{50}(SHC) (nM)	IC_{50}(OSC) (nM)
4	F	/	O	38	98
9*	F	/	O	60	6.5
12	O		N	75	380[c]
13*	O		CH	80	610[c]
14		/	O	96	5.4
18	CH_2	O	N	172	4.1
19	NH		N	180	39
20*	CH	CH	N	186	3.5
22*	*N*-Me		N	289	19.6
23*	S		N	306	2.9

Tab. 1.1: Inhibitor potencies for the ligands in fig. 1.4, bound to SHC and OSC. The errors lie between ±10% and ±15% . Some of the IC_{50} values of OSC were derived from a residual activity at a 100 nm inhibitor concentration and are therefore less accurate [2].

free energy of the system. $A(N, V, T)$ can be calculated from the canonical partition function $Z(N, V, T)$. Since in terms of simulation, calculation of the free energy would mean exploring the entire conformational space, it is not possible to calculate the absolute free energy of a system. Only free energy *differences* can be calculated [9].
Relative free energies of binding reflect the relative potencies of different inhibitors to bind to a protein. In computational chemistry, there are several methods to calculate these relative free energies. Two of these methods, thermodynamic integration (TI) and single step perturbation (SSP), are chosen for this study. TI [9] is a well estabilished method, whereas SSP [10] is still not much tested for the kind of ligands like SHC inhibitors. A detailed description of both methods will be given in the next chapter.

2. MATERIALS AND METHODS

The membrane protein SHC was simulated in water with GROMOSXX [11] and the force field parameter set 45A3 [12]. The initial structure of the SHC was taken from the protein data bank *www.rcsb.org/pdb*, code 1GSZ.

2.1 Calculation of Relative Free Energies of Binding

2.1.1 Single Step Perturbation

The relative free energy of binding $\Delta\Delta G_{AB}^{binding}$ between two ligands A and B was calculated over the thermodynamic cycle shown in fig. 2.1 over the difference of the free energy differences ΔG_{AR}^{water}, $G_{AR}^{protein}$, G_{BR}^{water}, and $G_{BR}^{protein}$ to a reference ligand R according to the formula

$$\Delta\Delta G_{AB}^{binding} = (\Delta G_{AR}^{protein} - \Delta G_{AR}^{water}) - (\Delta G_{BR}^{protein} - \Delta G_{BR}^{water}). \tag{2.1}$$

The free energy differences can be calculated *via* the *perturbation formula*

$$\Delta G_{AR} = G_A - G_R = -k_B T ln \langle e^{-(H_A - H_R)/k_B T} \rangle_R. \tag{2.2}$$

G_A and G_B are absolute free energies which can not be calculated for the reasons mentioned earlier. The ensemble average $\langle e^{-(H_A - H_R)/k_B T} \rangle_R$ is the time averaged mean of the exponential of the energy differences $H_A - H_R$ in an ensemble. To get these energy differences, only one single simulation of the reference ligand is needed. Out of this simulation comes a set of H_R's, i. e. the energies of the reference state in all the different conformations. After the simulation is done, the H_A's and H_B's can be calculated by putting the real ligands A and B at the position of the reference ligand, and calculating the H_A's and H_B's for all saved conformations of the simulation.
The calculation of a set of relative free energies out of a single simulation makes SSP attractive for inhibitor research, as long as some essential conditions are fulfilled:

1. The equation 2.2 can only be applied, if there is *enough overlap in the conformational* space of the reference ligand and the real ligands A and B (and C, D, E, ... if more than one $\Delta\Delta G^{binding}$ is calculated).

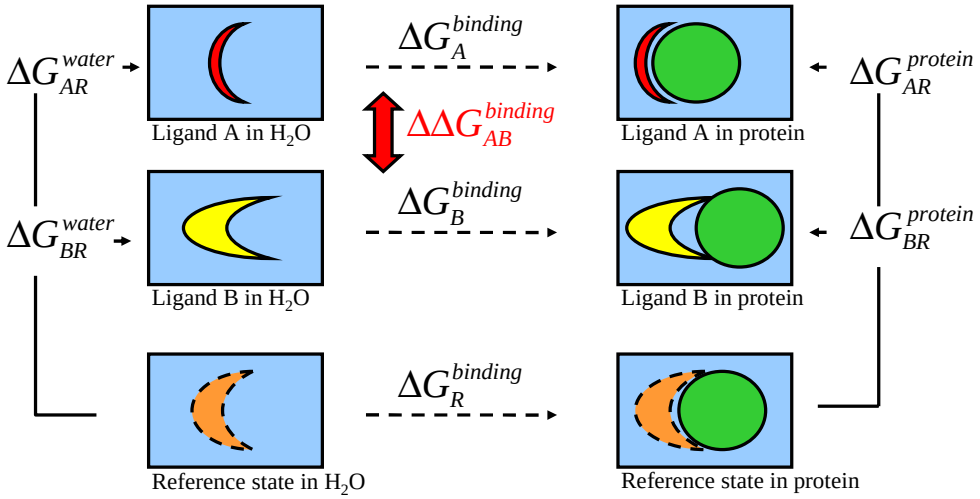


Fig. 2.1: A thermodynamic cycle to calculate $\Delta\Delta G_{AB}^{binding}$. The reference ligand must be able to adapt conformations of both ligands A and B. It contains therefore "soft" atoms which enhance the sampling of the simulation, indicated by the dotted lines.

2. Since *relative* free energies are calculated, the reference ligand can be unphysical. There is a lot of freedom in the design of a reference ligand. The only requirement is that point 1. is still valid.

3. *Softness*: To enhance the sampling, the reference ligand is made "soft", meaning, the atoms which differ from the real ligands have a modified non-bonded potential energy function $V^{nonb}(r_{ij}, \lambda)$, a *soft-core potential*, as shown in equation 2.3 and 2.4

$$V^{nonb}(r_{ij}, \lambda) = \lambda^n V^{LJC}(r_{ij}; B; (1-\lambda)) + (1-\lambda)^n V^{LJC}(r_{ij}; A; \lambda) \quad (2.3)$$

with

$$V^{LJC}(r_{ij}; X; (\lambda)) = \frac{1}{\alpha_{LJ}\lambda^2 C_{126}^X + r_{ij}^6} \times [\frac{C_{12}^X}{\alpha_{LJ}\lambda^2 C_{126}^X + r_{ij}^6} - C_6^X] + V^{Coulomb}, \quad (2.4)$$

where r_{ij} is the distance between the particles i and j, C_{12}, C_6, and C_{126} are the van der Waals parameters between i and j, and V^{LJC} is the contribution to the nonbonded potential from states A and B (denoted by the index $X = A, B$ in eq. 2.4). n is a parameter to change the

shape of the potential curve that is kept 1 in this study. $V^{coulomb}$ is the Coulombic potential energy. Like n, it plays a secondary role in this study, since the softness parameter for the Coulombic interaction is kept constant $\alpha_C = 0.2$. The scaling factors λ and α allow for a smooth change of real atoms (state A) to dummy atoms or other atoms (state B) and *vice versa*. They are responsible for the softness of the particles, since if α_{LJ} is chosen to be non-zero and λ is bigger than 0 or smaller than 1, the potential $V^{nonb}(r, \lambda)$ gets a finite value at $r = 0$. In other words, the particles behave less repulsive. In SSP, λ is kept 0.5, whereas α_{LJ} is chosen according to the desired softness for the reference ligand [11], [13], [10].

2.1.2 Thermodynamic Integration

Thermodynamic integration is a well established method for free energy difference calculation. It serves in this study to verify the less established SSP. In TI, the relative free energy of binding $\Delta\Delta G_{AB}^{binding}$ is calculated *via* the thermodynamic cycle in fig. 2.2 with the free energy differences ΔG_{AB}^{water} and $\Delta G_{AB}^{protein}$:

$$\Delta\Delta G_{AB}^{binding} = \Delta G_{AB}^{protein} - \Delta G_{AB}^{water}. \tag{2.5}$$

To get the ΔG_{AB}'s in water and in the protein, the system must be simulated at a set of different λ points (λ has the same meaning as in SSP, see previous subsection). The $\partial H/\partial\lambda$ ensemble averages at these λ points are calculated, that means the simulations at the different λ points must be long enough so that the energies are equilibrated. Finally, ΔG_{AB} is the integral area under the $\partial H/\partial\lambda$ ensemble average curve

$$\Delta G_{AB} = G(\lambda_A) - G(\lambda_B) = \int_{\lambda_B}^{\lambda_A} \langle \frac{\partial H}{\partial \lambda} \rangle_\lambda d\lambda. \tag{2.6}$$

2.2 Computational Details

Properties of the simulated systems: The simulated SHC contained 619 residues and 6'392 atoms. The volume of the rectangular water box was (8.5 nm)^3, 28'385 simple point charge (SPC) [14] water molecules were included. This gives a total atom number of the simulated system of 91'583 atoms. The total charge of the protein was -15 e, no counterions were added. The inhibitor concentrations were measured at pH 6, so His, Lys and Arg were protonated. In all the simulations except for the equilibration, an isothermal-isobaric (N,p,T) ensemble was simulated with $T = 300$ K and $p = 1$ atm.

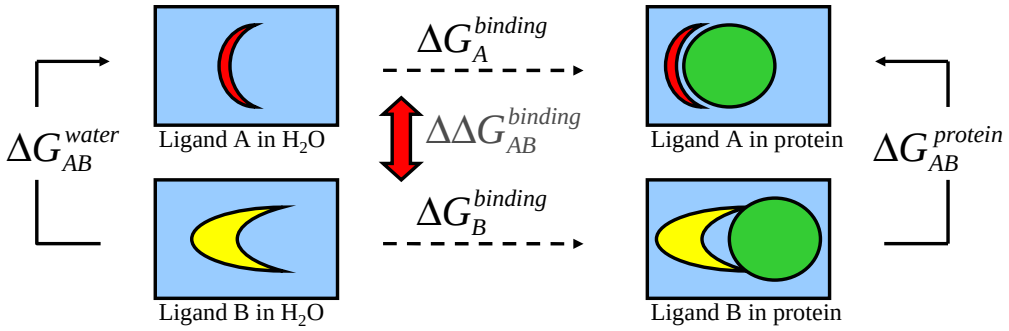


Fig. 2.2: The thermodynamic cycle to calculate $\Delta\Delta G_{AB}^{binding}$ *via* TI. No soft reference ligand is needed, but at the intermediate λ points, the molecules have also a soft-core potential.

The simulations were done with the GROMOSXX software, since grid based pairlist generation was used.

Topologies of the ligands: The molecular topologies of the protein and the ligand were built with the force field parameter set 45A3 [12]. Nonstandard force field parameters for angles and bonds are listed in the appendix, as well as the molecular building block files for the ligands.

Energy minimisation and equilibration prior to main simulation: The initial protein structure for every simulation was energy minimized *in vacuo*. The energy minimized structure was equilibrated in an NVT simulation in water for 120 ps in the following way: The first 20 ps at 60 K with a position restraint force constant of 25'000 $kJmol^{-1}nm^{-2}$, the second step for 20 ps at 120 K with a position restraint force constant of 20'000 $kJmol^{-1}nm^{-2}$, etc. until 300 K and 5'000 $kJmol^{-1}nm^{-2}$. The last 20 ps were carried out like an ordinary MD simulation, i.e. without restraints (force constant = 0 $kJmol^{-1}nm^{-2}$), at 300 K and with activated pressure coupling (NpT).

Computers and programs: The simulations were carried out on the C4 computers *gonzales* and *obelix* (*www.c4.ethz.ch*). Some less expensive simulations were done on the *penguin* cluster or interactively on *angmar* or

isengard.
The following GROMOS++ programs were used for the simulation and analysis: mkscript, maketop, comtop, prebbb, proion, ener, ene_ana, tstrip, rdf, rmsf, dssp, frameout, rmsd, probox, g962pdb, and vmdam.
Own analysis programs were programmed in perl and shell script, mainly to analyze the rmsf and the B-factors of the protein, to convert one-letter amino acid code of the protein, to analyze the TI, the SSP, etc. Mathematica [15] was used to calculate the relative free energy differences of binding of the SSP.
Visualisation of the protein was done with VMD and RASMOL.

3. RESULTS AND DISCUSSION

The energy minimisation *in vacuo* and the equilibration in water operated cleanly. The MD simulations, the TI and the SSP have been carried out and delivered the findings described in the next sections.

3.1 MD Simulations: Complexed and Uncomplexed SHC

Two molecular dynamics (MD) simulations, 2 ns long, with different initial velocities were performed for the SHC in complex with the inhibitor 22 of [2], and four times 2 ns MD simulations for the uncomplexed SHC in SPC [14] water. Two further simulations for the SHC in complex with the inhibitor 22 are in progress.

General behaviour of the protein in water: The protein structure does not change much during the simulation. As determined by DSSP [16], over 70 % of the residues stay during the whole simulation time to 100 % in the identical secondary structure category as in the experimental crystal structure (fig. 3.1). If not the experimental crystal structure, but the equilibrated structure of the protein in water was taken as a reference, the fraction of the residues staying in the same secondary structure is 74 %.

Stability of SHC: The stability of the protein is considerable. It was analyzed using root mean square deviation (RMSD) analysis. In both cases, the average RMSD value after equilibration for the backbone atoms with respect to the starting structure was 0.15 nm and for all atoms 0.25 nm. There was no significant difference between the RMSD of the complexed and the uncomplexed protein.

Flexibility of SHC: The flexibility of the protein was examined using root mean square fluctuations (RMSF). The calculated B-factors were compared with the B-factors of the X-ray structure from the protein data bank (3.2, 3.3,

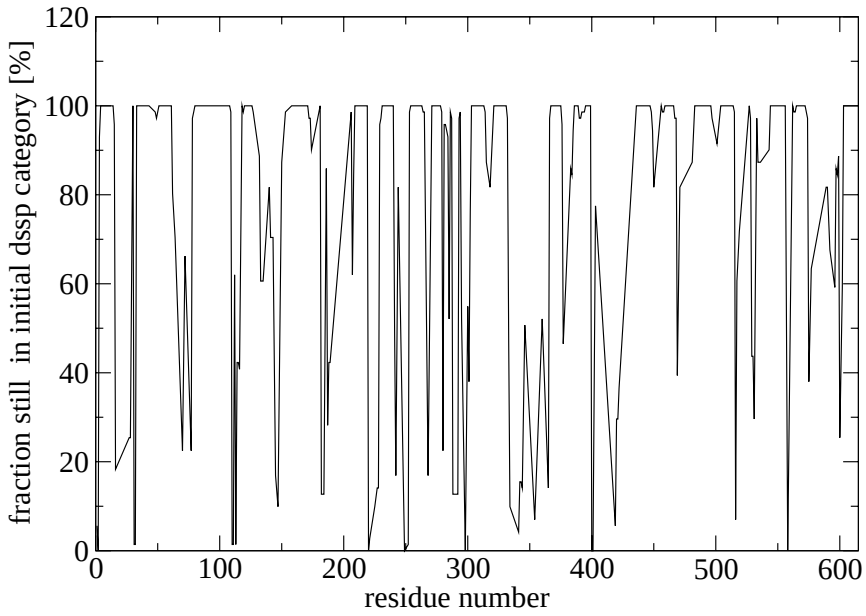


Fig. 3.1: Fraction of residues that stay to 100 % in the same secondary structure category (determined by DSSP) as in the experimental structure.

and 3.4). The plot 3.2 shows qualitative agreement with the experiment, but the calculated B-factors are lower than the ones derived from experiment. This finding is a consequence of the high uncertainty of the experimental values, and was already discussed in a previous study [17].

Most flexible parts: The atomic RMSF of the protein (fig. 3.3) show the most flexible parts of SHC. Interestingly, there is, apart from the assumed entrance channel loop (right alpha helix in the picture), a second higly flexible loop at the very bottom of the protein. This could be a starting point for the search for the release channel of the protein.

3.2 Calculated Relative Free Energies of Binding

For the four ligands nr. 12, 13, 19 and 22, $\Delta\Delta G_{AB}^{binding}$ were calculated and compared with the experimentally found $\Delta\Delta G_{AB}^{binding}(experiment)$, cal-

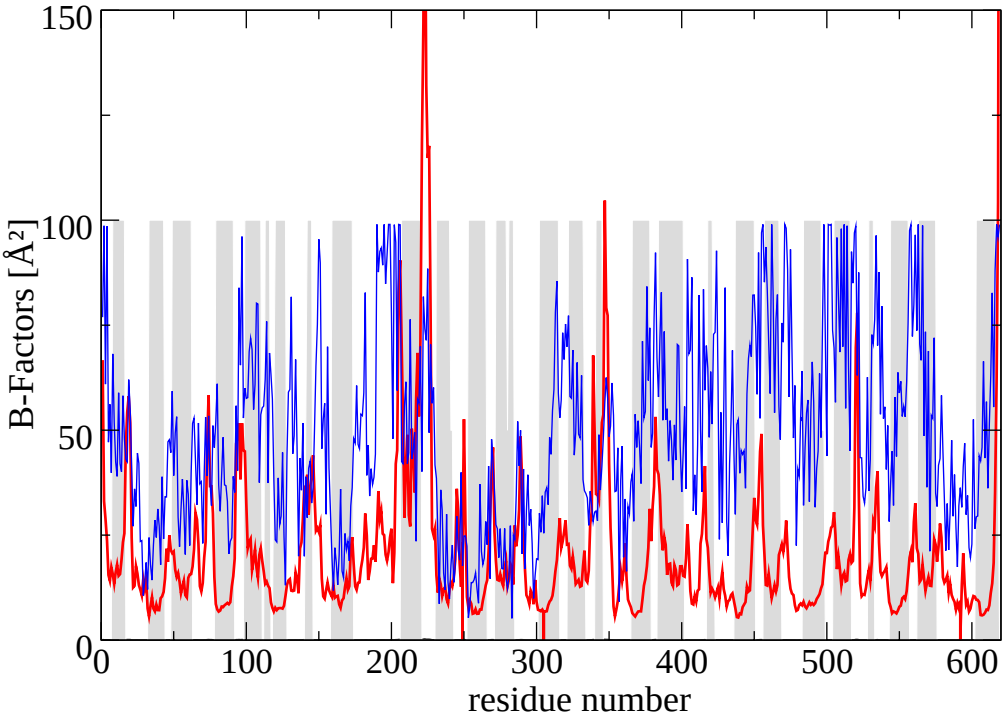


Fig. 3.2: B-Factors *versus* residue number of the protein. Blue: Crystal X-ray values. Red: Calculated from atomic RMSF values of the simulation. Grey indicated regions show the α-helices of the protein where the fluctuations are low.

culated from the IC_{50} values [2] *via* eq. 3.1

$$\Delta\Delta G_{AB}^{binding}(experiment) = -RTln\left(\frac{IC_{50}^{A}}{IC_{50}^{B}}\right)^{1}. \tag{3.1}$$

[1] The convention here is $\Delta\Delta G_{AB}^{binding} \equiv \Delta G_{A}^{binding} - \Delta G_{B}^{binding}$. The calculation of $\Delta\Delta G_{AB}^{binding}(experiment)$ is derived from Michaelis-Menten enzyme kinetics the following way: In a reaction $S+EI_X \leftrightarrows S+E+I_X \leftrightarrows SE+I_X$, where S is the substrate (squalene), E is the enzyme (SHC), I_X is the inhibitor of type $X = A, B$ and P is the product (hopene/diplopterol), the reaction constant for the competitive inhibition $S+E+I_X \leftrightarrows S+EI_X$ is $K_i^X = \frac{IC_{50}^X}{1+[S]^X/K_m^X}$ [18], where K_m^X is the Michaelis constant of the Enzyme and $[S]^X$ the concentration of the substrate. With the approximation $[S]^A/K_M^A \simeq [S]^B/K_M^B$, the experimental relative free energy of binding for the inhibitor is $\Delta\Delta G_{AB}^{binding}(experiment) = \Delta G_{A}^{binding} - \Delta G_{B}^{binding} = -RTln\left(\frac{K_i^A}{K_i^B}\right) = -RTln\left(\frac{\frac{IC_{50}^A}{1+[S]^A/K_m^A}}{\frac{IC_{50}^B}{1+[S]^B/K_m^B}}\right) \simeq -RTln\left(\frac{IC_{50}^A}{IC_{50}^B}\right)$.

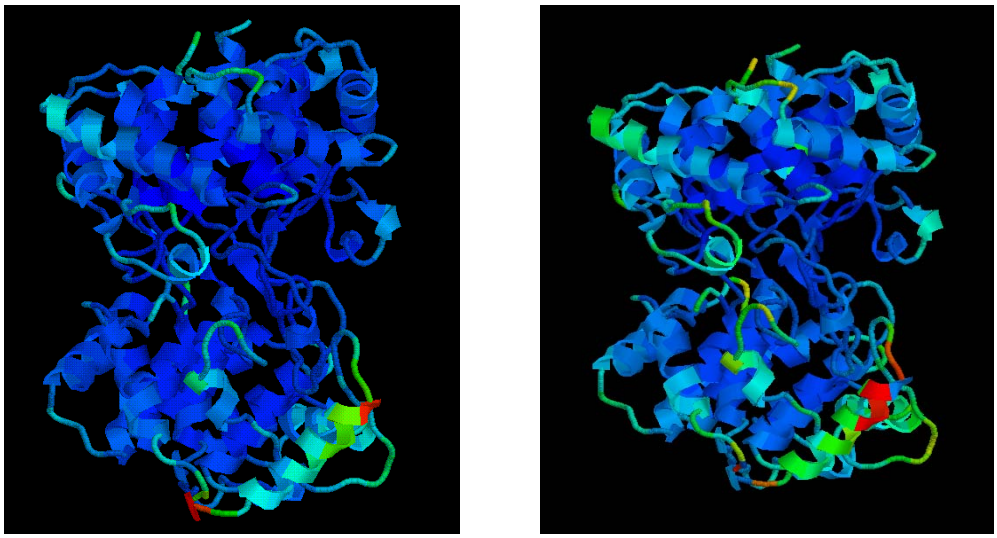


Fig. 3.3: Normalized atomic RMSF of the protein structure. The red regions are the most flexible, the blue regions are the most rigid. The relative flexibility is for the crystal structure (left) and for the simulation (right) the same, for example the loop on the right side of the protein, is in both cases one of the most fluctuating parts of the SHC. Qualitatively, the crystal structure is more flexible. Note the flexible region at the very bottom of the protein.

3.2.1 Results of the Thermodynamic Integration Simulations

Simulations: Six thermodynamic integrations were performed, three in protein from ligand 22 to ligand 12, from ligand 22 to ligand 13, from ligand 22 to ligand 19, and the same three perturbations of the ligands in water. The choice of the ligand resulted from the choice of ligands in the SSP, since the TI was meant as a verification. Every λ point was equilibrated for 100 ps and only the following 400 ps were used in the calculation. Some of the λ points in water were simulated longer than 400 ps to increase the sampling. The $\partial H/\partial\lambda$ curves of the simulations are shown in tables 3.1 and 3.2. $\Delta\Delta G_{AB}^{binding}$ was calculated according to eq. 2.5 and 2.6 in table 3.3. [2]

Comparison with the experimental $\Delta\Delta G_{AB}^{binding}$ in tab. 3.6 shows that the relative order of the values could be reproduced, i. e. $\Delta\Delta G_{22,12}^{binding} < \Delta\Delta G_{22,13}^{binding} < \Delta\Delta G_{22,19}^{binding}$. The TI therefore fulfilled the requirements. Still, there are two

[2] Note: the integral has to be evaluated from B to A because of the used convention $\Delta X_{AB} \equiv X_A - X_B$, i. e. $\Delta G_{AB} = \int_{\lambda_B}^{\lambda_A} \langle \frac{\partial H}{\partial \lambda} \rangle_\lambda d\lambda$

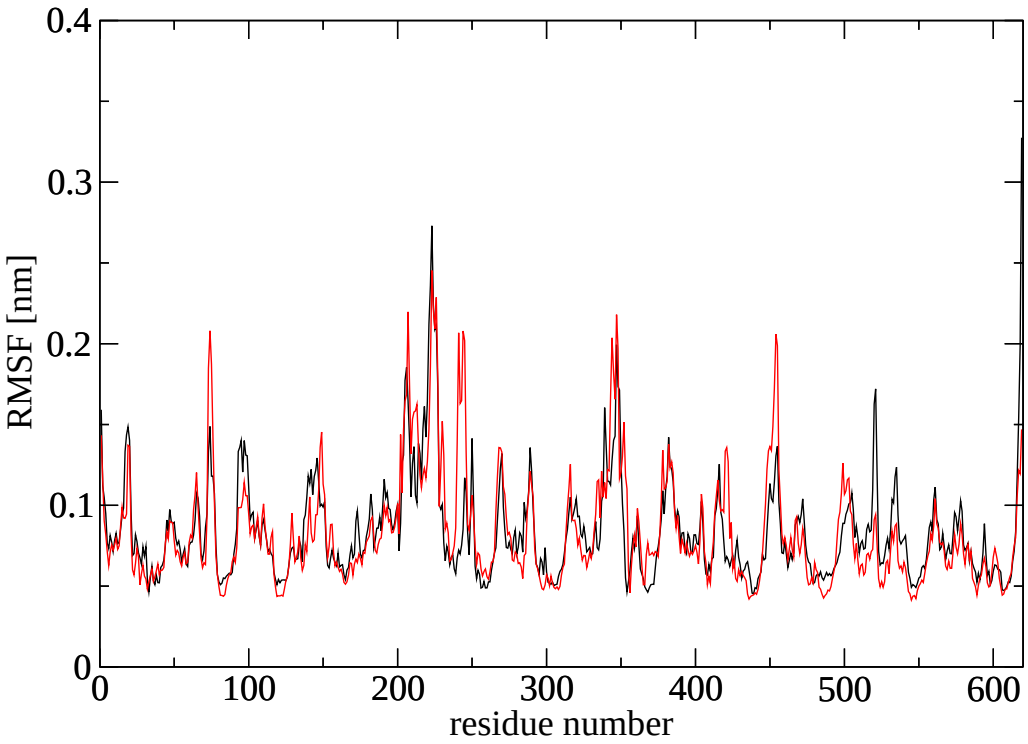


Fig. 3.4: Red: Atom-positional RMSF of the uncomplexed protein. Black: protein in complex with ligand 22. Highest differences in RMSF are around residue 200 and residue 400. They are visualized on the protein structure in 3.5.

reasons why the result should be taken with reserve:

1. The $\Delta\Delta G^{binding}_{22,19}$ is positive, meaning, in disagreement with the literature, that ligand 22 has a higher affinity than ligand 19, and might be a consequence of the higher kinetic energy of the environment atoms in water.

2. $\Delta\Delta G^{binding}_{22,12}$ and $\Delta\Delta G^{binding}_{22,13}$ are almost equal in experiment. In this TI, they differ by over 5 kJmol^{-1}. Still, the result can be true, because of the high error in experiment ($\pm 15\%$).

Accuracy of the results: Apparently, the free energy differences are not very accurate. The uncertainty arises mainly from the TI's in the protein. There, 400 ps are for certain λ points not enough to get a well equilibrated

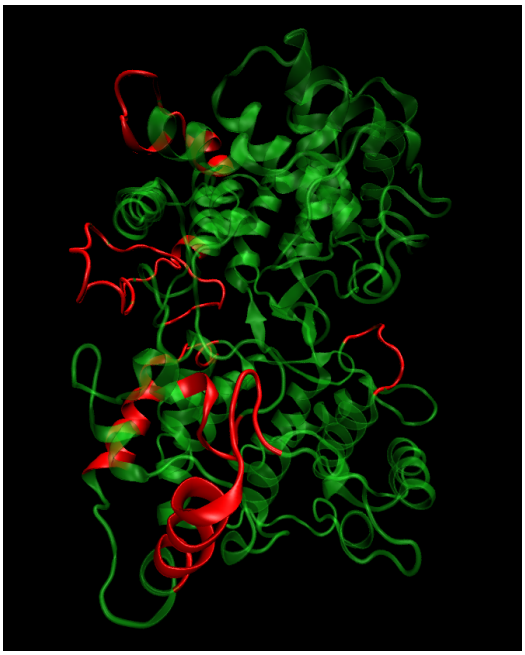


Fig. 3.5: Red: Loops with highest differences in atomic-positional RMSF between complexed and uncomplexed form of the protein, derived from the substraction of the curves from fig 3.4. Note the differences in the entrance channel loop.

$\partial H/\partial\lambda$ ensemble. This discrepancy is shown in fig. 3.6. It demonstrates an example for a set of well equilibrated $\partial H/\partial\lambda$ set and a suboptimal $\partial H/\partial\lambda$ set with a clear bias.
Attempts to smoothen the $\partial H/\partial\lambda$-curve by taking only the last 200 ps of the ensemble could not enhance the result. For example, $\Delta G_{22,19}^{protein}$ would be -5.19 $\pm$ 1.8 $kJmol^{-1}$ instead of -6.4 $\pm$ 1.4 $kJmol^{-1}$, if only the last 200 ps were taken into account.

3.2.2 Results of the Single Step Perturbation

The relative free energies of binding between ligand 12, 13, 19 and 22 have been calculated. For the calculation, a total of 8 ns of MD simulation with the reference ligand at $\lambda = 0.5$ was carried out, two times 2 ns with a softness of $\alpha_{LJ} = 1.8$ and two times 2 ns with a softness of $\alpha_{LJ} = 2.2$. The softness $\alpha_{LJ} = 1.51$, used in former studies [13], was not used, as 1) this value was

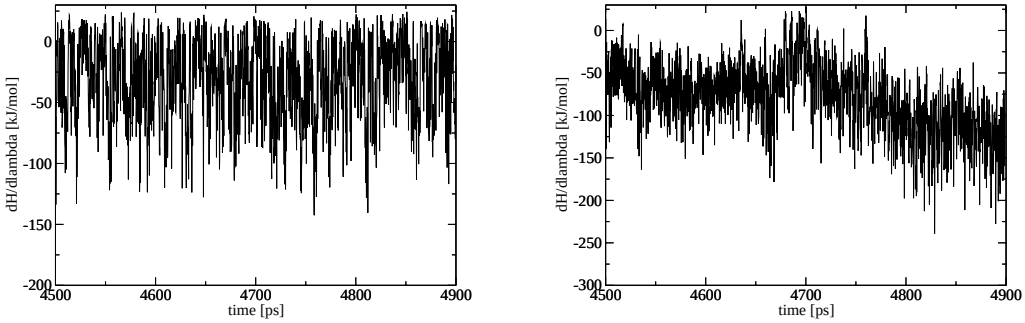


Fig. 3.6: Comparison between perturbation of ligand 22 into ligand 12 at λ=0.45 once in water and once in the protein. Left plot: Perturbation in water. Right plot: Perturbation in the protein.

used in water and not in protein environment, 2) it was used with other van der Waals parameters C_{12}, C_{12} and C_{126}, and 3) own tests showed that a higher α_{LJ} is more appropriate for the present system.

Choice of the ligands: The choice of this series of ligands resulted from the mutual similarity of the ligands, since in all other possible perturbations of ligands from [2], the geometry of the ligands, meaning the angle between the two phenyl rings, the connectivity (the five-membered ring is not always there), and atom bond distances would have changed significantly. After dissatisfying tests with bonds changed to distance restraints, λ-scaled force constants for bond angles, and omitted dihedral energy terms, it was found that the set up of a working reference state for ligands with different geometry demands a thorough investigation of the bonds, angles, and atoms to be perturbed, in order to get stringent results. Hence, instead of using a complex reference state, a series of ligands with simple perturbations, i. e. altering atom types and an altering few bond types and bond angle types, was chosen. The reference ligand is shown in fig. 3.7.

Relative free energy differences of binding: The differences in relative free energy of binding were calculated 3.6. In table 3.4, the differences in free energy for a single step perturbation with a soft $\alpha_{LJ} = 2.2$ are listed, in table 3.5 are the differences in free energy for a harder alpha $\alpha_{LJ} = 1.8$.

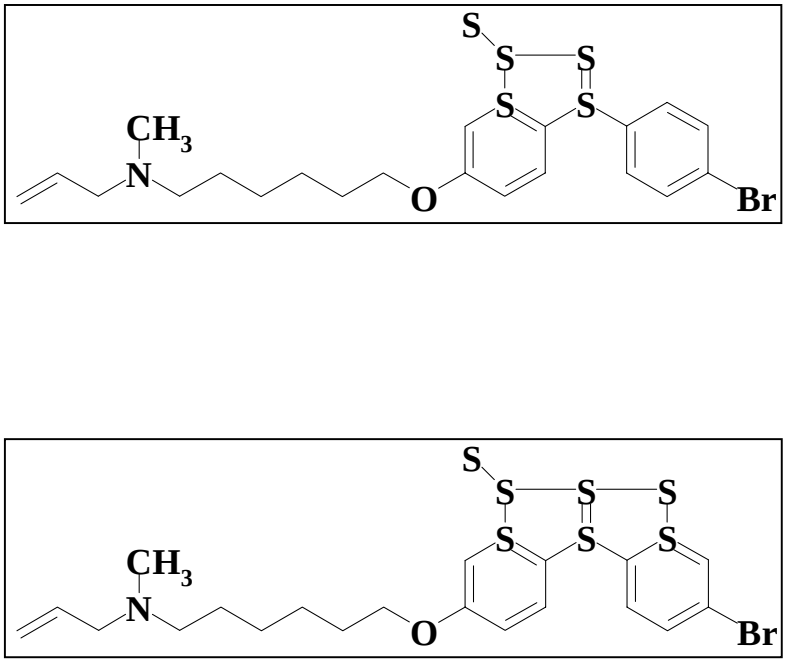


Fig. 3.7: Reference ligand used for the SSP. The ligand is identical with ligand 22 [2], but has soft atoms at the indicated sites (S). The upper ligand was used in this SSP. The lower ligand is a suggestion for a more complex reference ligand for use in future studies.

Compared with the experimental values in 3.6, the calculated values are negative, too, and in the same order of magnitude, but they are not in the same relative order. In other words, the ranking for the softer α_{LJ} is $\Delta\Delta G^{binding}_{22,13} > \Delta\Delta G^{binding}_{22,12} > \Delta\Delta G^{binding}_{22,19}$, respective $\Delta\Delta G^{binding}_{22,13} > \Delta\Delta G^{binding}_{22,19} > \Delta\Delta G^{binding}_{22,12}$ for the harder α_{LJ}. This would mean that ligand 19 has a higher affinity than in experiment. [3]
These inaccuracies may be a result of the low number of conformations used for the calculation of the free energy differences in the protein. Only 416 conformations per simulation, i. e. every 5 ps of the total 2080 ps, were saved due to the limited disk space (see section"Technical Findings"). The values in water are better equilibrated (10'100 saved conformations, every 2 ps), but still, they could be improved by more sampling, too. The equilibration in protein and water is visualized in fig. 3.8.

[3] Due to the remarkabley negative results of the SSP, the program dg_ener was veryfied with Mathematica [15] and found to yield the same result.

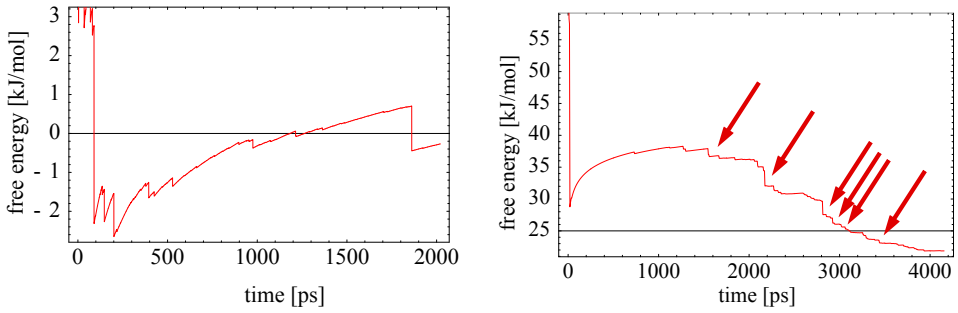


Fig. 3.8: Development of ΔG_{AR} for ligand 12 in the SSP with $\alpha = 2.2$. Left: ΔG_{AR}^{water}, right: $\Delta G_{AR}^{protein}$. The arrows indicate the conformations used to visualize the lowest energy conformation in fig. 3.2.2.

Choice of α: Comparison of the SSP results with different softness parameters showed a significant dependence of the end result. The differences in free energy of binding are for the softer SSP reference ligand $\sim$ 10kJmol^{-1} higher. They seem to be better equilibrated because of the less repulsive behaviour of the atoms of the reference ligand, see also the RDF plots 3.9. The choice of α_{LJ} has a considerable influence on the end result of a SSP. This is another evidence that the $\Delta\Delta G_{AB}^{bindinb}$ are not yet equilibrated values due to insufficient sampling.
The choice of α has the biggest influence on $\Delta G_{22,R}^{protein}$ of ligand 22, which is the only ligand with a methyl group. This may be the explanation why the relative free energies of binding for the softer reference ligand are more accurate: there, the sampling is better. In water, the sampling is less dependent from α because the kinetic energy of the environment atoms is higher.

Changing the reference ligand into a real ligand in water versus changing the reference lingand into a real ligand in the protein: Compared with other literature, it seems to be more difficult to perturb the reference ligand in water. A good equilibration needs many more conformations in water. This is in agreement with the literature [19].

Most stable conformations for ligand 12 in protein: The six overlaid conformations of ligand 12 in the single step perturbation with $\alpha_{LJ} = 2.2$ that

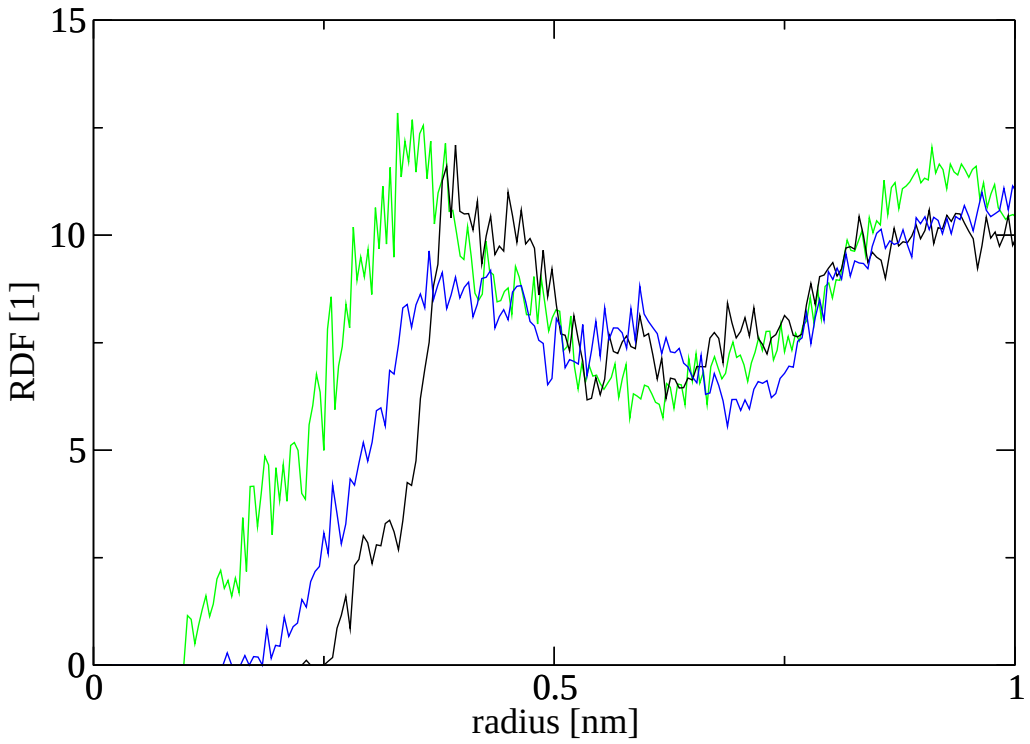


Fig. 3.9: Radial distrubution function (RDF) of the surrounding protein atoms, calculated for the methyl carbon atom of the reference state in the SSP. λ was all the time 0.5. Black: no softness from alpha, means $\alpha = 1.0$, blue: $\alpha = 2.2$, green: $\alpha = 2.4$. The alpha of the methyl group was chosen to be 0.2 higher than the rest of the soft atoms to enhance the sampling, as this group is in ligand 19, 13, and 12 not present.

contribute most to $\Delta G_{AR}^{protein}$ have been visualized in 3.2.2.
The important part of the inhibitor seems to be the top with the amine. There, strong interactions hold the inhibitor in its place. In nature, there might be even stronger interactions because of electrons in the Hückel system of the cyclopropene and the double bond. The lower part of the inhibitor is more mobile, which may allow more structural diversity in future inhibitor design.

Narrow range of IC_{50}'s: The drawback of the chosen ligand series is the small range of the involved IC_{50} values. The highest IC_{50} value is 75 nM for ligand 12, the lowest IC_{50} value is 289 nM for ligand 22. This yields a maximal

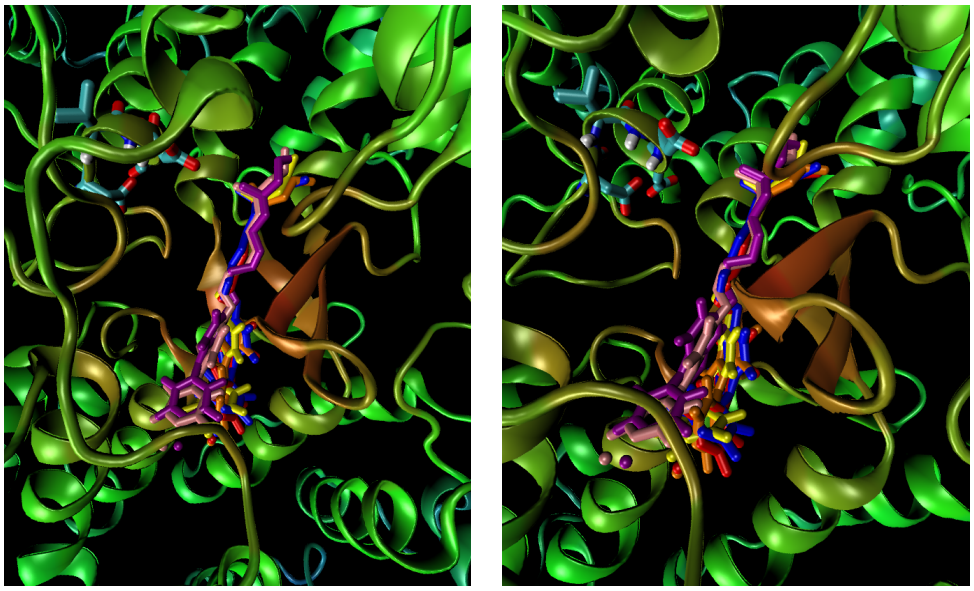


Fig. 3.10: Picture of the reactive cavity. The six overlaid conformations that contribute most to $\Delta G_{AR}^{protein}$ of ligand 12 in the single step perturbation with $\alpha_{LJ} = 2.2$. Left side: The residue with the amine stays next to Asp376, what confirms the proposal of [6] that this may be the protonating residue. Right side: The lower part of the ligand seems to be quite mobile.

relative difference in free energy of binding of -3.5 kJmol^{-1}, which is a small value, even if there are studies that reached this accuracy [20]. Values much smaller than kT are hardly reproducable by SSP or TI. Due to the confidence interval of $\pm$ 15 % of the experiment, the smallest $\Delta\Delta G_{AB}^{binding}(experiment)$ between ligand 12 and 13 of -0.2 kJmol^{-1} is an uncertain value. It is also not sure if ligand 12 has really a higher affinity than ligand 13. For example in OSC, the two IC_{50} values differ much more.

TI versus SSP: Comparison of SSP with the TI shows that the latter method is more accurate, but computationally more costly, too. Both methods produce deviation from the experimental values in the same order of magnitude. Economically seen, it would therefore be more efficient to go for the SSP for future calculations, in particular because of the additional structural information, but the SSP needs further tests to get better agreement with the literature.

3.3 Technical Findings

Execution of simulations demanded to overcome several technical obstacles:

1. *Lack of disk space*: This was the main technical challenge. Still, it is not completely worked out. For example, with a disk quota of 8 gigabytes on an account of *gonzales*, the only possible way to save trajectories was to save them every 5.0 ps, and to remove the data from the account daily. This produced a lot of manual time-consuming data file transfer, checks, zipping, and removal of solvent. Furthermore, the simulation speed would allow to get for the investigated SHC system over 1 ns of simulation in three days. But since the disk space is limited, it is hardly feasible to save enuogh data to yield accurate $\Delta\Delta G_{AB}^{binding}$ with a system of almost 100'000 atoms.

2. *slow simulation speed*: after tests on obelix, igc, and gonzales, the latter seems to yield the fastest simulation of all the computers. The MPI implementation of the MD program was used in order to use parallel processors. The optimal simulation speed was found out with tests (fig: 3.11). Furthermore, the use of grid- based pairlist generation enhanced the speed, too. Totally, a speed-up from a few ps to 470 ps/day was reached.

3. *Queuing:* The above simulation speed could not be realized due to numerous server crashes (mean: 1/week) and exited job files. Due to the queuing system, some simulations were waiting for three days until the first job started. By distributing the jobs over different queues with 8, 16 and 32 processors, the queuing time could be shortened.

4. *Simulation set up:* Creation of working simulation set ups for four different computation resources with different queuing systems: *gonzales*, *obelix*, *penguin* and *angmar/isengard* was for all the computers different and needed a lot of testing.

5. *Bugs:*

   - Certain versions of the GROMOS++ program mkscript were often defective. Primarily the set up of simulations for the TI with subfolders needed great many corrections. Only one makescript version worked error-free. A goal for future studies could be to make the set up of the simulations easier by writing for example a program that compresses the data automatically after the simulation, removes the solvent, checks if the simulation is successfully

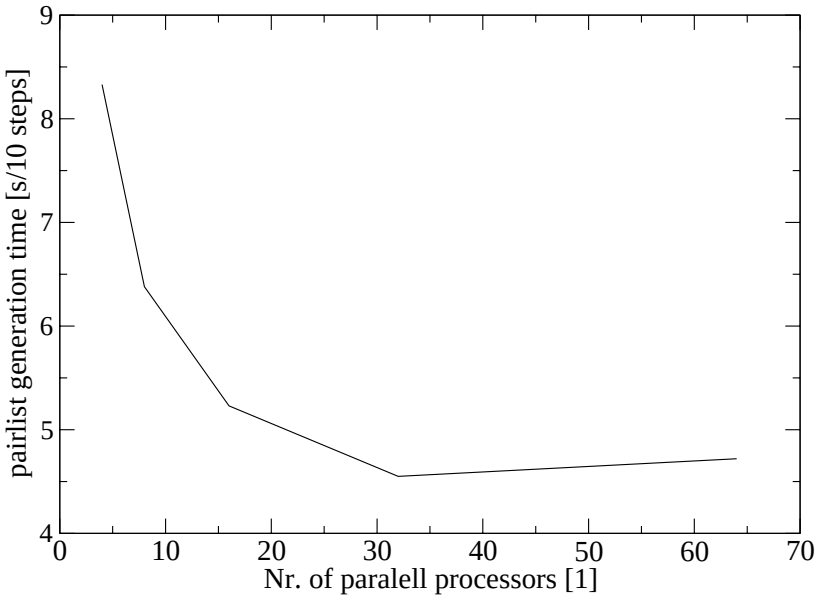
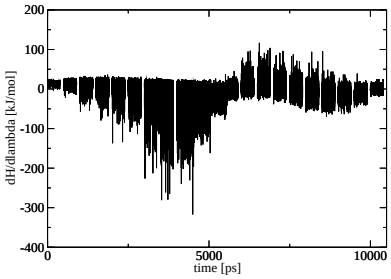
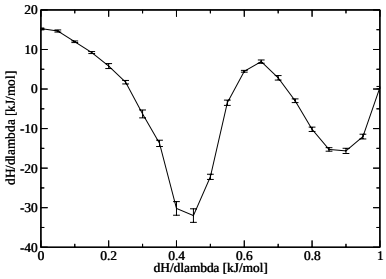
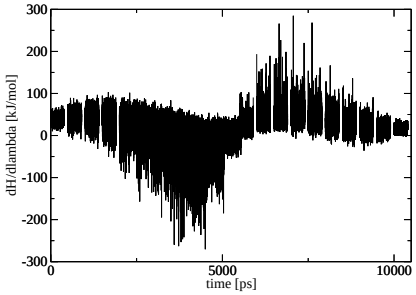
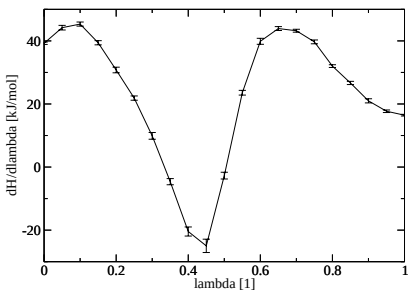
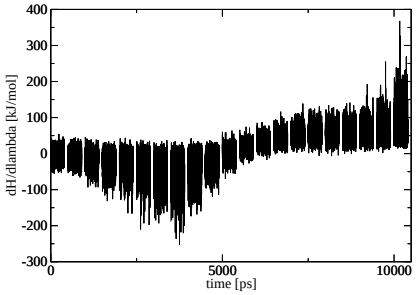
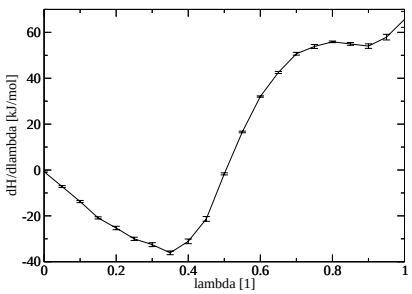


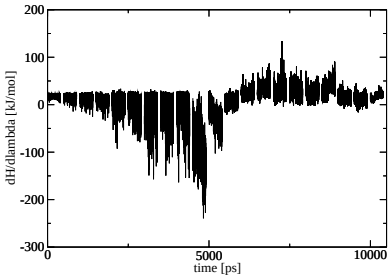
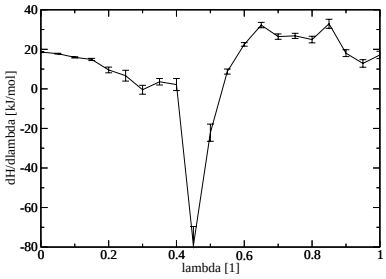
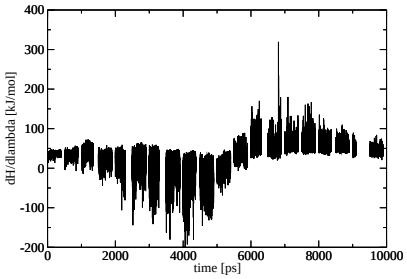
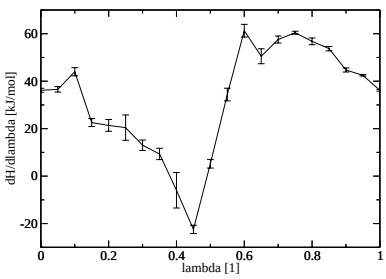
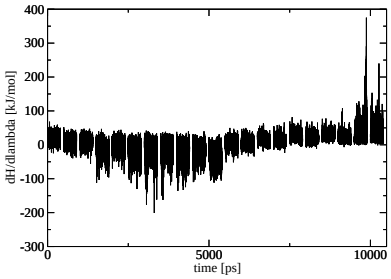
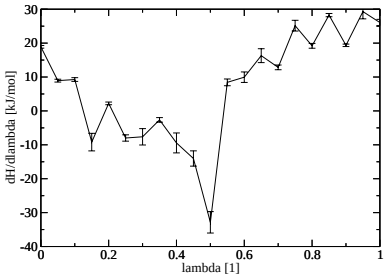
Fig. 3.11: Simulation speed optimisation. By changing from 4 to 32 processors, the simulation speed could be improved from 120 to 250 ps per day.

completed and resubmits the simulations to the server in case of an exited job. This would save time-consuming checks and data file transfers which have to be done manually at the moment.

- The program rmsf crashed, if too many trajectory files were analyzed. It is not known if this is a matter of the size of the investigated system or of the specified file number.
- In GROMOSXX, the angle type for perturbed angles in the block PERTBANGLE03 in the molecular perturbation topology file has to be specified the other way around compared to all the other blocks. One has to specify *first* the angle type of the state B and *then* the angle type of the state A. This is only true for the PERTBANGLE03. If the types are specified like in the other blocks, this leads to the error message "non-existing bond angle perturbed" in the output file.

	$\partial H/\partial\lambda$ in water	averaged $\partial H/\partial\lambda$ in water
A		
B		
C		

Tab. 3.1: $\partial H/\partial\lambda$-curves of the TI's in water. A: Perturbation of ligand 22 to ligand 12. B: Ligand 22 to ligand 13. C: Ligand 22 to ligand 19. Left row: Simulation time for every $\partial H/\partial\lambda$ ensemble. The time axis was chosen to visualize the length of the sampling for a certain λ point. For example, the $\partial H/\partial\lambda$ values from λ=0.8 are the values between 8000 and 8500 ps. Right row: ensemble average of each λ point.

	$\partial H/\partial\lambda$ in protein	averaged $\partial H/\partial\lambda$ in protein
A		
B		
C		

Tab. 3.2: $\partial H/\partial\lambda$-curves of the TI's in protein. A: Perturbation of ligand 22 to ligand 12. B: Ligand 22 to ligand 13. C: Ligand 22 to ligand 19. Left row: Simulation time for every $\partial H/\partial\lambda$ ensemble. The time axis was chosen to visualize the length of the sampling for a certain λ point. For example, the $\partial H/\partial\lambda$ values from λ=0.8 are the values between 8000 and 8500 ps. Right row: ensemble average of each λ point.

A, B	$\Delta G_{AB}^{protein}$ [kJmol^{-1}]	ΔG_{AB}^{water} [kJmol^{-1}]	$\Delta\Delta G_{AB}^{binding}$ [kJmol^{-1}]
22, 12	-9.6 ± 2.1	4.92 ± 0.63	-14.5 ± 2.7
22, 13	-32.1 ± 2.1	-22.71 ± 0.80	-9.4 ± 2.9
22, 19	-6.4 ± 1.4	-11.55 ± 0.73	5.3 ± 2.1

Tab. 3.3: Result of the thermodynamic integrations: Differences in free energies in the protein, in water, and the resulting relative free energy of binding. The confidence intervals are added as a measure of the estimated error of the ensemble averages and not representative as real confidence interval. The real error is not known.

ligand A	$\Delta G_{AR}^{protein}$ [kJmol^{-1}]	ΔG_{AR}^{water} [kJmol^{-1}]	$\Delta G_{AR}^{binding}$ [kJmol^{-1}]
12	21.9	-0.3	22.2
13	40.7	26.2	14.6
19	643.3	619.5	23.8
22	18.9	5.4	13.5

Tab. 3.4: Differences in free energy for the single step perturbation with softer alpha $\alpha_{LJ} = 2.2$.

ligand A	$\Delta G_{AR}^{protein}$ [kJmol^{-1}]	ΔG_{AR}^{water} [kJmol^{-1}]	$\Delta G_{AR}^{binding}$ [kJmol^{-1}]
12	24.0	-5.4	29.4
13	34.2	20.7	13.5
19	643.7	616.3	27.4
22	7.6	1.2	6.4

Tab. 3.5: Differences in free energy for the single step perturbation with the harder $\alpha_{LJ} = 1.8$.

A, B	$\Delta\Delta G_{AB}^{binding}$ [kJmol^{-1}] $\alpha_{LJ} = 1.8$	$\Delta\Delta G_{AB}^{binding}$ [kJmol^{-1}] $\alpha_{LJ} = 2.2$	$\Delta\Delta G_{AB}^{binding}$ [kJmol^{-1}] experiment
22, 12	-23.8	-8.7	-3.5
22, 13	-7.1	-1.1	-3.3
22, 19	-21.0	-10.3	-1.2

Tab. 3.6: Results from the SSP with a harder $\alpha_{LJ} = 1.8$, calculated from values of 3.5, and a softer $\alpha_{LJ} = 2.2$, calculated from values of 3.4. Experimental $\Delta\Delta G_{AB}^{binding}$ are calculated from the IC_{50} values of [2] with eq. 3.1.

4. CONCLUSIONS

SHC was simulated in water, once in form of a protein-inhibotor complex, and once in uncomplexed form. In both cases, the protein shows high stability. The examinaton of the atom-positional RMSF of the protein showed agreement with literature [1] and [6]. Therefore, much of the behaviour of SHC seems to be reproducible in water.
The supposed entrance channel from the literature showed in the MD simulation a significantly higher flexibility. A second hihly flexible region, found at the bottom of the protein and suspected to provide information about a release channel, could be the starting point for further examinations of SHC in water.
Relative free binding energies of four anticholesteremic inhibitors were calculated using themodynamic integration and single step perturbation. Both methods yielded $\Delta\Delta G_{AB}^{binding}$'s in the same order of magnitude as the experiment, although only a small number of conformations was analyzed. Comparison of the obtained results showed that TI reproduces qualitatively the relative order of the experimental affinities, but SSP does not, presumabely due to insufficient sampling. The reason for the lack of sampling was the limitated time-frame of this study and the technical obstacles which had to be overcome, mainly by reason of the size of the simulated system and the complexity of the setup for the SSP. Therefore, the results of both methods have to be interpreted with due caution.
The setup of the simulations on four different queuing systems, the elimination of bugs in the programs, the creation of shell scripts for the setup of the simulations and Mathematica notebooks and perl programs for the analysis had to be managed, as well as the handling of big data files. The simulation speed was optimized the maximal rate to get results in the specified timeframe. Moreover, it was necessary to perform several test simulations with GROMOS96 and GROMOSXX, before a suitable system to simulate could be found. Since most of these challenges could be worked out during this study, the performance of MD simulations, TI, and SSP with SPC is now feasible in a much shorter time. A simulation setup for maximimal simulation speed within the available computation resources on the IGC and C4 clusters (*penguin*, *obelix*, *gonzales*) is now avilable. With some more disk

space, usable valuable could be produced within a few weeks. Only then could reliable testing of SSP be carried out.
For future studies with big proteins, it would be meaningful to automatize the analysis and checks during the simulations, the data file shuffling over different computers and the resubmitting of exited jobs in a program.
A goal for a future study could be the testing of the SSP with a more complex reference ligand which allows to extend the series of investigated ligands.

BIBLIOGRAPHY

[1] K. U. Wendt, A. Lenhardt, and G. E. Schulz. The Structure of the Membrane Protein Squalene-Hopene Cyclase at 2.0 Å Resolution. *J. Mol. Biol.*, 286:175–187, 1999.

[2] Alexander Lenhart, Dirk J. Reinert, Johannes D. Aebi, Henrietta Dehmlow, Oliver H. Morand, and Georg E. Schulz. Binding Structures and Potencies of Oxidosqualene Cyclase Inhibitors with the Homologous Squalene-Hopene Cyclase. *J. Med. Chem.*, 46:2083–2092, 2003.

[3] Harvard School of Public Health. Fats and Cholesterol. *www.hsph.harvard.edu/nutritionsource/fats.html*, 2006.

[4] B. Seckler and K. Poralla. Characterisation and partial purification of squalene-hopene cyclase from *Bacillus acidocaldarius*. *Biochim. Biophys. Acta*, 881:356–363, 1986.

[5] K. Ulrich Wendt, Karl Poralla, and Georg E. Schulz. Structure and Function of a Squalene Cyclase. *Science*, 277:1811–1815, 1997.

[6] Alexander Lenhart, Wilhelm A. Weihofen, Axel E. W. Pleschke, and Georg E. Schulz. Crystal Structure of a Squalene Cyclase in Complex with the Potential Anticholesteremic Drug Ro48-8071. *Chemisty & Biology*, 9:639–645, 2002.

[7] K. Ulrich Wendt, Georg E. Schulz, Elias J. Corey, and David R. Liu. Enzyme Mechanisms for Polycyclic Triterpene Formation. *Angew. Chem. Int. Ed.*, 39:2812–2833, 2000.

[8] T. Dang, I. Abe, Y. F. Zheng, and G. D. Prestwich. The binding site for an inhibitor of squalene-hopene cyclase determined using photoaffinity labeling and molecular modeling. *Chem. Biol.*, 6:333–341, 1999.

[9] Wilfred F. van Gunsteren, Xavier Daura, and Alan E. Mark. Computation of Free Energy. *Helvetica Chimica Acta*, 85:3113–3129, 2002.

[10] Chris Oostenbrink and Wilfred F. van Gunsteren. Single-Step Perturbations to Calculate Free Energy Differences from Unphysical Reference States: Limits on Size, Flexibility, and Character. *J. Comput. Chem.*, 24:1730–1739, 2003.

[11] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I.P. Tironi. *Biomolecular Simulation: The GROMOS96 manual and user guide.* BIOMOS, b.v., ETH Zürich, Laboratory of Physical Chemistry, ETH Zentrum, CH-8092 Zürich, Switzerland, 1996.

[12] Lukas D. Schuler, Xavier Daura, and Wilfred F. van Gunsteren. An Improved GROMOS96 Force Field for Aliphatic Hydrocarbonds in The Condensed Phase. *Journal of Computational Chemistry*, 22(11):1205–1218, 2001.

[13] Heiko Schäfer, Wilfred F. van Gunsteren, and Alan E. Mark. Estimating Relative Free Energies from a Single Ensemble: Hydration Free Energies. *J. Comput. Chem*, 20(15):1604–1617, 1999.

[14] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and B. Pullman (ed.) J. Hermans. *Intermolecular Forces.* D. Reidel Publishing Company, Dordrecht, 1981.

[15] Wolfram Research. Wolfram Information Center. *library.wolfram.com*, 2006.

[16] W. Kabsch and C. Sander. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen Bonded and Geometrical Features. *Biopolymers*, 22:2577–2637, 1983.

[17] P. H. Hünenberger, A. E. Mark, and W.F. van Gunsteren. Fluctuation and Cross-Correlation Analysis of Protein Motions Observed in Nanosecond Molecular Dynamics Simulations. *J. Mol. Biol.*, 252:492–503, 1995.

[18] R. Blume. Quantitative Enzymkinetik: Die Alkoholdehydrogenase (ADH); 1. Charakteristik, kinetische Analyse und Substratspezifität. *Praxis der Naturw.*, 27:65–78, 1978.

[19] Chris Oostenbrink and Wilfred F. van Gunsteren. Free energies of ligand binding for structurally diverse compounds. *PNAS*, 102(19):6750, 6754 2005.

[20] Chris Oostenbrink and Wilfred F. van Gunsteren. Free Energies of Binding of Polychlorinated Biphenyls to the Estrogen Receptor From a Single Simulation. *PROTEINS: Structure, Funcion and Bioinformatics*, 54:237–246, 2004.

[21] Chris Oostenbrink and Wilfred F. van Gunsteren. Efficient Calculateion of Many Stacking ond Pairing Free Energies in dDNA from a Few Molecular Dynamics Simulations. *Chem. Eur. J.*, 11:4340–4348, 2005.

[22] Jožica Dolenc, Chris Oostenbrink, Jože Koller, and Wilfred F. van Gunsteren. Molecular dynamics simulations and free energy calculations of netropsin and distamycin binding to an AAAAA DNA binding site. *Nucleic Acids Research*, 33(2):725–733, 2005.

[23] Chris Oostenbrink, Daniel Juchli, and Wilfred F. van Gunsteren. Amine Hydration: A United-Atom Force-Field Solution. *ChemPhysChem*, 6:1800–1804, 2005.

[24] Chris Oostenbrink, Alessandra Villa, Alan E. Mark, and Wilfred F. van Gunsteren. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.*, 25:1656–1676, 2004.

APPENDIX

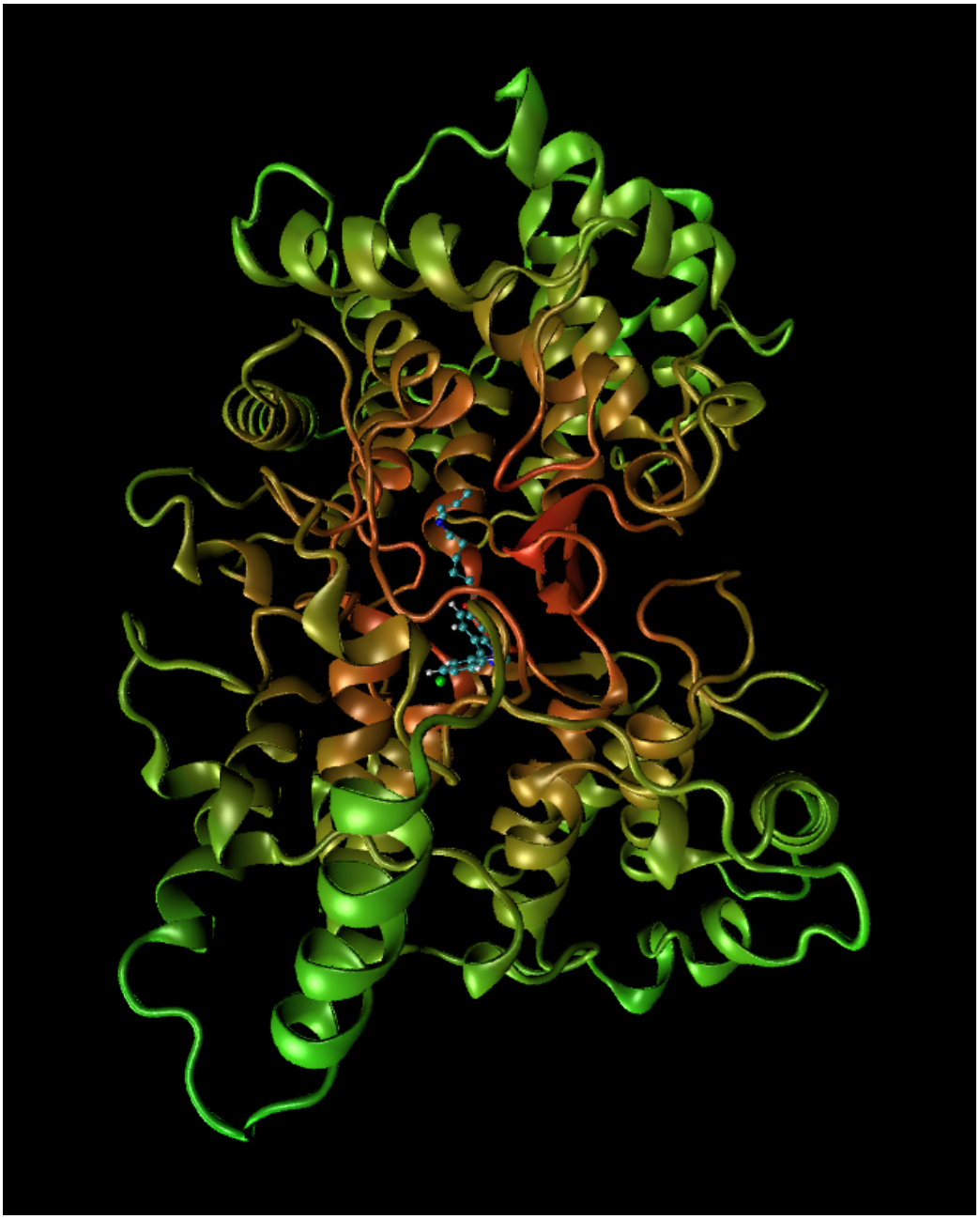


Fig. .1: Squalene-hopene cyclase in NewCartoon representation. The colour changes according to the distance of the inhibitor from red to green.

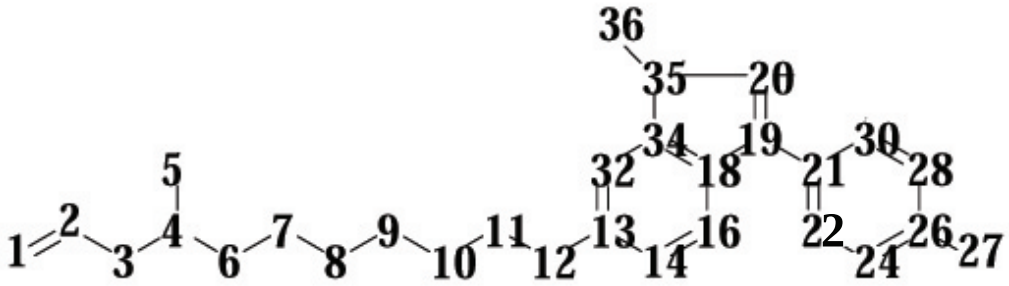


Fig. .2: Numbering of the molecular topologies. The missing numbers belong to the hydrogens on the six-membered rings.

Bond	Type code	Force const. [kJ/mol^{-1}nm^{-4}]	Ideal bond length [nm]
C - Br	48	2.3900e+06	0.190000
C - F	49	2.3900e+06	0.136000
C - N	50	8.7100e+06	0.147000 [23]
Bond angle	Type code	Force const. [kJmol^{-1}]	Ideal angle [degrees]
CH3 - N - CH3	47	425.00	109.50 [23]

Tab. .1: Nonstandard force field parameters for the real and unphysical ligands studied.

Building block for ligand 12:

```
MTBUILDBLSOLUTE
# building block version 29/06/2006 fschwab ligand 12
# RNME
LI12
# number of atoms, number of preceding exclusions
# NMAT,NLIN
   36    0
# preceding exclusions
#ATOM                                MAE MSAE
# atoms
#ATOM ANM  IACM MASS        CGMICGM MAE MSAE
    1 C1     13    4    0.00000   0   2    2    3
    2 C2     12    3    0.00000   1   2    3    4
    3 C3     13    4    0.21000   0   3    4    5    6
    4 N4      8   14   -0.63000   0   3    5    6    7
    5 C5     14    5    0.21000   0   1    6
    6 C6     13    4    0.21000   1   2    7    8
    7 C7     13    4    0.00000   0   2    8    9
    8 C8     13    4    0.00000   0   2    9   10
    9 C9     13    4    0.00000   0   2   10   11
   10 C10    13    4    0.00000   1   2   11   12
   11 C11    13    4    0.18000   0   2   12   13
   12 O12     3   16   -0.36000   0   7   13   14   15   16   32
   33   34
   13 C13    11   12    0.18000   1   9   14   15   16   17   18
   32   33   34   35
   14 C14    11   12   -0.10000   0   8   15   16   17   18   19
   32   33   34
   15 H15    17    1    0.10000   1   4   16   17   18   32
   16 C16    11   12   -0.10000   0   8   17   18   19   20   21
   32   34   35
   17 H17    17    1    0.10000   1   3   18   19   34
   18 C18    11   12    0.00000   1  10   19   20   21   22   30
   32   33   34   35   36
   19 C19    11   12    0.36000   0  12   20   21   22   23   24
   28   30   31   32   34   35   36
   20 N20     8   14   -0.36000   1   7   21   22   30   32   35
   36   34
   21 C21    11   12    0.00000   1  11   22   23   24   25   26
   28   29   30   31   34   35
   22 C22    11   12   -0.10000   0   8   23   24   25   26   27
   28   30   31
   23 H23    17    1    0.10000   1   4   24   25   26   30
   24 C24    11   12   -0.10000   0   6   25   26   27   28   29
   30
   25 H25    17    1    0.10000   1   3   26   27   28
   26 C26    11   12    0.08700   0   5   27   28   29   30   31
   27 BR27   31   80   -0.08700   1   3   28   29   30
```

```
   28 C28    11   12   -0.10000   0   3   29   30   31
   29 H29    17    1    0.10000   1   2   30   31
   30 C30    11   12   -0.10000   0   1   31
   31 H31    17    1    0.10000   1   0
   32 C32    11   12   -0.10000   0   4   33   34   35   36
   33 H33    17    1    0.10000   1   2   34   35
   34 C34    11   12    0.36000   0   2   35   36
   35 O35     3   16   -0.36000   0   1   36
   36 C36    19    5    0.00000   1   0
# bonds
#  NB
   38
#  IB   JB  MCB
    1    2   12
    2    3   26
    3    4   50
    4    5   50
    4    6   50
    6    7   26
    7    8   26
    8    9   26
    9   10   26
   10   11   26
#    10
   11   12   17
   12   13   12
   13   14   15
   13   32   15
   14   15    3
   14   16   15
   16   17    3
   16   18   15
   18   19    9
   18   34   15
#    20
   19   20    9
   19   21   15
   20   35    9
   21   22   15
   21   30   15
   22   23    3
   22   24   15
   24   25    3
   24   26   15
   26   27   48
#    30
   26   28   15
   28   29    3
   28   30   15
```

```
   30   31    3
   32   33    3
   32   34   15
   34   35    9
   35   36   50
# bond angles
# NBA
   55
#  IB   JB   KB  MCB
    1    2    3   26
    2    3    4   14
    3    4    5   47
    3    4    6   47
    5    4    6   47
    4    6    7   14
    6    7    8   14
    7    8    9   14
    8    9   10   14
    9   10   11   14
#    10
   10   11   12   14
   11   12   13   11
   12   13   14   26
   12   13   32   26
   14   13   32   26
   13   14   15   24
   13   14   16   26
   15   14   16   24
   14   16   17   24
   14   16   18   26
#    20
   17   16   18   24
   16   18   19   38
   16   18   34   26
   19   18   34    6
   18   19   20    6
   18   19   21   36
   20   19   21   36
   19   21   22   26
   19   21   30   26
   22   21   30   26
#    30
   19   20   35    6
   21   22   23   24
   21   22   24   26
   23   22   24   24
   22   24   25   24
   22   24   26   26
   25   24   26   24
```

```
   24   26   27   26
   24   26   28   26
   27   26   28   26
#    40
   26   28   29   24
   26   28   30   26
   29   28   30   24
   21   30   28   26
   21   30   31   24
   28   30   31   24
   13   32   33   24
   13   32   34   26
   33   32   34   24
   18   34   32   26
#    50
   18   34   35    6
   32   34   35   38
   34   35   36   36
   20   35   34    6
   20   35   36   36
# NIDA
    32
#  IB   JB   KB   LB  MCB
    4    3    5    6    2
   13   14   32   12    1
   14   13   16   15    1
   16   14   18   17    1
   18   16   34   19    1
   34   18   32   35    1
   32   13   34   33    1
   14   13   32   34    1
   13   32   34   18    1
   32   34   18   16    1
#    10
   14   16   18   34    1
   13   14   16   18    1
   16   14   13   32    1
   19   18   34   35    1
   19   18   20   21    1
   35   20   34   36    1
   18   34   35   20    1
   19   20   35   34    1
   18   19   20   35    1
   20   19   18   34    1
#    20
   21   22   30   19    1
   22   21   24   23    1
   24   22   26   25    1
   26   24   28   27    1
```

```
   28   26   30   29    1
   30   21   28   31    1
   30   21   22   24    1
   21   22   24   26    1
   22   24   26   28    1
   24   26   28   30    1
#    30
   21   30   28   26    1
   22   21   30   28    1
# dihedrals
# NDA
   11
#  IB   JB   KB   LB  MCB
    1    2    3    4   20
    2    3    4    5   19
    3    4    6    7   19
    4    6    7    8   17
    6    7    8    9   17
    7    8    9   10   17
    8    9   10   11   17
    9   10   11   12   17
   10   11   12   13   12
   11   12   13   32    2
#    10
   20   19   21   30    1
END
```

Building block for ligand 13:

```
MTBUILDBLSOLUTE
# building block version 29/06/2006 fschwab ligand 13
# RNME
LI13
# number of atoms, number of preceding exclusions
# NMAT,NLIN
   36    0
# preceding exclusions
#ATOM                                MAE MSAE
# atoms
#ATOM ANM  IACM MASS        CGMICGM MAE MSAE
    1 C1     13    4    0.00000   0   2    2    3
    2 C2     12    3    0.00000   1   2    3    4
    3 C3     13    4    0.21000   0   3    4    5    6
    4 N4      8   14   -0.63000   0   3    5    6    7
    5 C5     14    5    0.21000   0   1    6
    6 C6     13    4    0.21000   1   2    7    8
    7 C7     13    4    0.00000   0   2    8    9
    8 C8     13    4    0.00000   0   2    9   10
    9 C9     13    4    0.00000   0   2   10   11
```

```
   10 C10    13    4    0.00000   1   2   11   12
   11 C11    13    4    0.18000   0   2   12   13
   12 O12     3   16   -0.36000   0   7   13   14   15   16   32
   33   34
   13 C13    11   12    0.18000   1   9   14   15   16   17   18
   32   33   34   35
   14 C14    11   12   -0.10000   0   8   15   16   17   18   19
   32   33   34
   15 H15    17    1    0.10000   1   4   16   17   18   32
   16 C16    11   12   -0.10000   0   8   17   18   19   20   21
   32   34   35
   17 H17    17    1    0.10000   1   3   18   19   34
   18 C18    11   12    0.00000   1  10   19   20   21   22   30
   32   33   34   35   36
   19 C19    11   12    0.00000   0  12   20   21   22   23   24
   28   30   31   32   34   35   36
   20 C20    16    3    0.00000   1   7   21   22   30   32   35
   36   34
   21 C21    11   12    0.00000   1  11   22   23   24   25   26
   28   29   30   31   34   35
   22 C22    11   12   -0.10000   0   8   23   24   25   26   27
   28   30   31
   23 H23    17    1    0.10000   1   4   24   25   26   30
   24 C24    11   12   -0.10000   0   6   25   26   27   28   29
   30
   25 H25    17    1    0.10000   1   3   26   27   28
   26 C26    11   12    0.08700   0   5   27   28   29   30   31
   27 BR27   31   80   -0.08700   1   3   28   29   30
   28 C28    11   12   -0.10000   0   3   29   30   31
   29 H29    17    1    0.10000   1   2   30   31
   30 C30    11   12   -0.10000   0   1   31
   31 H31    17    1    0.10000   1   0
   32 C32    11   12   -0.10000   0   4   33   34   35   36
   33 H33    17    1    0.10000   1   2   34   35
   34 C34    11   12    0.36000   0   2   35   36
   35 O35     3   16   -0.36000   0   1   36
   36 C36    19    5    0.00000   1   0
#  NB
   38
#  IB   JB  MCB
    1    2   12
    2    3   26
    3    4   50
    4    5   50
    4    6   50
    6    7   26
    7    8   26
    8    9   26
    9   10   26
```

```
   10   11   26
#    10
   11   12   17
   12   13   12
   13   14   15
   13   32   15
   14   15    3
   14   16   15
   16   17    3
   16   18   15
   18   19    9
   18   34   15
#    20
   19   20    9
   19   21   15
   20   35    9
   21   22   15
   21   30   15
   22   23    3
   22   24   15
   24   25    3
   24   26   15
   26   27   48
#    30
   26   28   15
   28   29    3
   28   30   15
   30   31    3
   32   33    3
   32   34   15
   34   35    9
   35   36   50
# NBA
   55
#  IB   JB   KB  MCB
    1    2    3   26
    2    3    4   14
    3    4    5   47
    3    4    6   47
    5    4    6   47
    4    6    7   14
    6    7    8   14
    7    8    9   14
    8    9   10   14
    9   10   11   14
#    10
   10   11   12   14
   11   12   13   11
   12   13   14   26
```

```
   12   13   32   26
   14   13   32   26
   13   14   15   24
   13   14   16   26
   15   14   16   24
   14   16   17   24
   14   16   18   26
#    20
   17   16   18   24
   16   18   19   38
   16   18   34   26
   19   18   34    6
   18   19   20    6
   18   19   21   36
   20   19   21   36
   19   21   22   26
   19   21   30   26
   22   21   30   26
#    30
   19   20   35    6
   21   22   23   24
   21   22   24   26
   23   22   24   24
   22   24   25   24
   22   24   26   26
   25   24   26   24
   24   26   27   26
   24   26   28   26
   27   26   28   26
#    40
   26   28   29   24
   26   28   30   26
   29   28   30   24
   21   30   28   26
   21   30   31   24
   28   30   31   24
   13   32   33   24
   13   32   34   26
   33   32   34   24
   18   34   32   26
#    50
   18   34   35    6
   32   34   35   38
   34   35   36   36
   20   35   34    6
   20   35   36   36
# NIDA
    32
#  IB   JB   KB   LB  MCB
```

```
    4    3    5    6    2
   13   14   32   12    1
   14   13   16   15    1
   16   14   18   17    1
   18   16   34   19    1
   34   18   32   35    1
   32   13   34   33    1
   14   13   32   34    1
   13   32   34   18    1
   32   34   18   16    1
#    10
   14   16   18   34    1
   13   14   16   18    1
   16   14   13   32    1
   19   18   34   35    1
   19   18   20   21    1
   35   20   34   36    1
   18   34   35   20    1
   19   20   35   34    1
   18   19   20   35    1
   20   19   18   34    1
#    20
   21   22   30   19    1
   22   21   24   23    1
   24   22   26   25    1
   26   24   28   27    1
   28   26   30   29    1
   30   21   28   31    1
   30   21   22   24    1
   21   22   24   26    1
   22   24   26   28    1
   24   26   28   30    1
#    30
   21   30   28   26    1
   22   21   30   28    1
# dihedrals
# NDA
   11
#  IB   JB   KB   LB  MCB
    1    2    3    4   20
    2    3    4    5   19
    3    4    6    7   19
    4    6    7    8   17
    6    7    8    9   17
    7    8    9   10   17
    8    9   10   11   17
    9   10   11   12   17
   10   11   12   13   12
   11   12   13   32    2
```

```
#    10
   20   19   21   30    1
END
```

Building block for ligand 19:

```
MTBUILDBLSOLUTE
# building block version 29/06/2006 fschwab ligand 19
# RNME
LI19
# number of atoms, number of preceding exclusions
# NMAT,NLIN
   36    0
# preceding exclusions
#ATOM                                MAE MSAE
# atoms
#ATOM ANM  IACM MASS        CGMICGM MAE MSAE
    1 C1     13    4    0.00000   0   2    2    3
    2 C2     12    3    0.00000   1   2    3    4
    3 C3     13    4    0.21000   0   3    4    5    6
    4 N4      8   14   -0.63000   0   3    5    6    7
    5 C5     14    5    0.21000   0   1    6
    6 C6     13    4    0.21000   1   2    7    8
    7 C7     13    4    0.00000   0   2    8    9
    8 C8     13    4    0.00000   0   2    9   10
    9 C9     13    4    0.00000   0   2   10   11
   10 C10    13    4    0.00000   1   2   11   12
   11 C11    13    4    0.18000   0   2   12   13
   12 O12     3   16   -0.36000   0   7   13   14   15   16   32
   33   34
   13 C13    11   12    0.18000   1   9   14   15   16   17   18
   32   33   34   35
   14 C14    11   12   -0.10000   0   8   15   16   17   18   19
   32   33   34
   15 H15    17    1    0.10000   1   4   16   17   18   32
   16 C16    11   12   -0.10000   0   8   17   18   19   20   21
   32   34   35
   17 H17    17    1    0.10000   1   3   18   19   34
   18 C18    11   12    0.00000   1  10   19   20   21   22   30
   32   33   34   35   36
   19 C19    11   12    0.36000   0  12   20   21   22   23   24
   28   30   31   32   34   35   36
   20 N20     8   14   -0.36000   1   7   21   22   30   32   35
   36   34
   21 C21    11   12    0.00000   1  11   22   23   24   25   26
   28   29   30   31   34   35
   22 C22    11   12   -0.10000   0   8   23   24   25   26   27
   28   30   31
   23 H23    17    1    0.10000   1   4   24   25   26   30
```

```
   24 C24    11   12   -0.10000   0   6   25   26   27   28   29
   30
   25 H25    17    1    0.10000   1   3   26   27   28
   26 C26    11   12    0.08700   0   5   27   28   29   30   31
   27 BR27   31   80   -0.08700   1   3   28   29   30
   28 C28    11   12   -0.10000   0   3   29   30   31
   29 H29    17    1    0.10000   1   2   30   31
   30 C30    11   12   -0.10000   0   1   31
   31 H31    17    1    0.10000   1   0
   32 C32    11   12   -0.10000   0   4   33   34   35   36
   33 H33    17    1    0.10000   1   2   34   35
   34 C34    11   12    0.01000   0   2   35   36
   35 N35     8   14   -0.20000   0   1   36
   36 H36    18    1    0.19000   1   0
#  NB
   38
#  IB   JB  MCB
    1    2   12
    2    3   26
    3    4   50
    4    5   50
    4    6   50
    6    7   26
    7    8   26
    8    9   26
    9   10   26
   10   11   26
#    10
   11   12   17
   12   13   12
   13   14   15
   13   32   15
   14   15    3
   14   16   15
   16   17    3
   16   18   15
   18   19    9
   18   34   15
#    20
   19   20    9
   19   21   15
   20   35    9
   21   22   15
   21   30   15
   22   23    3
   22   24   15
   24   25    3
   24   26   15
   26   27   48
```

```
#    30
   26   28   15
   28   29    3
   28   30   15
   30   31    3
   32   33    3
   32   34   15
   34   35    9
   35   36    2
# NBA
   55
#  IB   JB   KB  MCB
    1    2    3   26
    2    3    4   14
    3    4    5   47
    3    4    6   47
    5    4    6   47
    4    6    7   14
    6    7    8   14
    7    8    9   14
    8    9   10   14
    9   10   11   14
#    10
   10   11   12   14
   11   12   13   11
   12   13   14   26
   12   13   32   26
   14   13   32   26
   13   14   15   24
   13   14   16   26
   15   14   16   24
   14   16   17   24
   14   16   18   26
#    20
   17   16   18   24
   16   18   19   38
   16   18   34   26
   19   18   34    6
   18   19   20    6
   18   19   21   36
   20   19   21   36
   19   21   22   26
   19   21   30   26
   22   21   30   26
#    30
   19   20   35    6
   21   22   23   24
   21   22   24   26
   23   22   24   24
```

```
   22   24   25   24
   22   24   26   26
   25   24   26   24
   24   26   27   26
   24   26   28   26
   27   26   28   26
#    40
   26   28   29   24
   26   28   30   26
   29   28   30   24
   21   30   28   26
   21   30   31   24
   28   30   31   24
   13   32   33   24
   13   32   34   26
   33   32   34   24
   18   34   32   26
#    50
   18   34   35    6
   32   34   35   38
   34   35   36   35
   20   35   34    6
   20   35   36   35
# NIDA
    32
#  IB   JB   KB   LB  MCB
    4    3    5    6    2
   13   14   32   12    1
   14   13   16   15    1
   16   14   18   17    1
   18   16   34   19    1
   34   18   32   35    1
   32   13   34   33    1
   14   13   32   34    1
   13   32   34   18    1
   32   34   18   16    1
#    10
   14   16   18   34    1
   13   14   16   18    1
   16   14   13   32    1
   19   18   34   35    1
   19   18   20   21    1
   35   20   34   36    1
   18   34   35   20    1
   19   20   35   34    1
   18   19   20   35    1
   20   19   18   34    1
#    20
   21   22   30   19    1
```

```
   22   21   24   23    1
   24   22   26   25    1
   26   24   28   27    1
   28   26   30   29    1
   30   21   28   31    1
   30   21   22   24    1
   21   22   24   26    1
   22   24   26   28    1
   24   26   28   30    1
#    30
   21   30   28   26    1
   22   21   30   28    1
# dihedrals
# NDA
   11
#  IB   JB   KB   LB  MCB
    1    2    3    4   20
    2    3    4    5   19
    3    4    6    7   19
    4    6    7    8   17
    6    7    8    9   17
    7    8    9   10   17
    8    9   10   11   17
    9   10   11   12   17
   10   11   12   13   12
   11   12   13   32    2
#    10
   20   19   21   30    1
END
```

Building block for ligand 22:

```
MTBUILDBLSOLUTE
# version 14/06/2006 fschwab reference ligand for ti, ligand 22
# RNME
LI22
# number of atoms, number of preceding exclusions
# NMAT,NLIN
   36    0
# preceding exclusions
#ATOM                                MAE MSAE
# atoms
#ATOM ANM  IACM MASS        CGMICGM MAE MSAE
    1 C1     13    4    0.00000   0   2    2    3
    2 C2     12    3    0.00000   1   2    3    4
    3 C3     13    4    0.21000   0   3    4    5    6
    4 N4      8   14   -0.63000   0   3    5    6    7
    5 C5     14    5    0.21000   0   1    6
    6 C6     13    4    0.21000   1   2    7    8
```

```
    7 C7     13    4    0.00000   0   2    8    9
    8 C8     13    4    0.00000   0   2    9   10
    9 C9     13    4    0.00000   0   2   10   11
   10 C10    13    4    0.00000   1   2   11   12
   11 C11    13    4    0.18000   0   2   12   13
   12 O12     3   16   -0.36000   0   7   13   14   15   16   32
   33   34
   13 C13    11   12    0.18000   1   9   14   15   16   17   18
   32   33   34   35
   14 C14    11   12   -0.10000   0   8   15   16   17   18   19
   32   33   34
   15 H15    17    1    0.10000   1   4   16   17   18   32
   16 C16    11   12   -0.10000   0   8   17   18   19   20   21
   32   34   35
   17 H17    17    1    0.10000   1   3   18   19   34
   18 C18    11   12    0.00000   1  10   19   20   21   22   30
   32   33   34   35   36
   19 C19    11   12    0.36000   0  12   20   21   22   23   24
   28   30   31   32   34   35   36
   20 N20     8   14   -0.36000   1   7   21   22   30   32   35
   36   34
   21 C21    11   12    0.00000   1  11   22   23   24   25   26
   28   29   30   31   34   35
   22 C22    11   12   -0.10000   0   8   23   24   25   26   27
   28   30   31
   23 H23    17    1    0.10000   1   4   24   25   26   30
   24 C24    11   12   -0.10000   0   6   25   26   27   28   29
   30
   25 H25    17    1    0.10000   1   3   26   27   28
   26 C26    11   12    0.08700   0   5   27   28   29   30   31
   27 BR27   31   80   -0.08700   1   3   28   29   30
   28 C28    11   12   -0.10000   0   3   29   30   31
   29 H29    17    1    0.10000   1   2   30   31
   30 C30    11   12   -0.10000   0   1   31
   31 H31    17    1    0.10000   1   0
   32 C32    11   12   -0.10000   0   4   33   34   35   36
   33 H33    17    1    0.10000   1   2   34   35
   34 C34    11   12    0.20000   0   2   35   36
   35 N35     8   14   -0.20000   0   1   36
   36 C36    14    5    0.00000   1   0
#  NB
   38
#  IB   JB  MCB
    1    2  12
    2    3  26
    3    4  50
    4    5  50
    4    6  50
    6    7  26
```

```
    7    8   26
    8    9   26
    9   10   26
   10   11   26
#    10
   11   12   17
   12   13   12
   13   14   15
   13   32   15
   14   15    3
   14   16   15
   16   17    3
   16   18   15
   18   19    9
   18   34   15
#    20
   19   20    9
   19   21   15
   20   35    9
   21   22   15
   21   30   15
   22   23    3
   22   24   15
   24   25    3
   24   26   15
   26   27   48
#    30
   26   28   15
   28   29    3
   28   30   15
   30   31    3
   32   33    3
   32   34   15
   34   35    9
   35   36   50
# NBA
   55
#  IB   JB   KB  MCB
    1    2    3   26
    2    3    4   14
    3    4    5   47
    3    4    6   47
    5    4    6   47
    4    6    7   14
    6    7    8   14
    7    8    9   14
    8    9   10   14
    9   10   11   14
#    10
```

```
   10   11   12   14
   11   12   13   11
   12   13   14   26
   12   13   32   26
   14   13   32   26
   13   14   15   24
   13   14   16   26
   15   14   16   24
   14   16   17   24
   14   16   18   26
#    20
   17   16   18   24
   16   18   19   38
   16   18   34   26
   19   18   34    6
   18   19   20    6
   18   19   21   36
   20   19   21   36
   19   21   22   26
   19   21   30   26
   22   21   30   26
#    30
   19   20   35    6
   21   22   23   24
   21   22   24   26
   23   22   24   24
   22   24   25   24
   22   24   26   26
   25   24   26   24
   24   26   27   26
   24   26   28   26
   27   26   28   26
#    40
   26   28   29   24
   26   28   30   26
   29   28   30   24
   21   30   28   26
   21   30   31   24
   28   30   31   24
   13   32   33   24
   13   32   34   26
   33   32   34   24
   18   34   32   26
#    50
   18   34   35    6
   32   34   35   38
   34   35   36   36
   20   35   34    6
   20   35   36   36
```

```
# NIDA
    32
#  IB   JB   KB   LB  MCB
    4    3    5    6    2
   13   14   32   12    1
   14   13   16   15    1
   16   14   18   17    1
   18   16   34   19    1
   34   18   32   35    1
   32   13   34   33    1
   14   13   32   34    1
   13   32   34   18    1
   32   34   18   16    1
#    10
   14   16   18   34    1
   13   14   16   18    1
   16   14   13   32    1
   19   18   34   35    1
   19   18   20   21    1
   35   20   34   36    1
   18   34   35   20    1
   19   20   35   34    1
   18   19   20   35    1
   20   19   18   34    1
#    20
   21   22   30   19    1
   22   21   24   23    1
   24   22   26   25    1
   26   24   28   27    1
   28   26   30   29    1
   30   21   28   31    1
   30   21   22   24    1
   21   22   24   26    1
   22   24   26   28    1
   24   26   28   30    1
#    30
   21   30   28   26    1
   22   21   30   28    1
# dihedrals
# NDA
   11
#  IB   JB   KB   LB  MCB
    1    2    3    4   20
    2    3    4    5   19
    3    4    6    7   19
    4    6    7    8   17
    6    7    8    9   17
    7    8    9   10   17
    8    9   10   11   17
```

```
    9   10   11   12   17
   10   11   12   13   12
   11   12   13   32    2
#    10
   20   19   21   30    1
END
```

Perturbation topology for the SSP:

```
TITLE
single step perturbation of ligssp in water , harder alpha
for the softer alpha, add 0.2 to the alphas (ALJ) below
END
PERTATOM03
# number of perturbed atoms
   5
#
#   NR RES NAME IAC(A) MASS(A)  CHARGE(A)
  IAC(B) MASS(B) CHARGE(B)   ALJ  ACRF
  6411 620  C19     11  12.011       0.36     11  12.011
  0.36   1.8  0.2
  6412 620  N20      8  14.0067     -0.36      8  14.0067
  -0.36   1.8  0.2
  6426 620  C34     11  12.011       0.20     11  12.011
   0.20   1.8  0.2
  6427 620  N35      8  14.0067     -0.20      8  14.0067
  -0.20   1.8  0.2
  6428 620  C36     14  15.035       0.00     14  15.035
   0.00   2.2  0.2
END
```