

CHARLES UNIVERSITY OF PRAGUE
FACULTY OF SCIENCE
DEPARTMENT OF BIOCHEMISTRY

Počítačové modelování interakcí typu enzym - substrát,
pomocí programu AutoDock 4

Computer modeling of enzyme-ligand interactions
using software package AutoDock 4

Bachelor Thesis

Supervisor: RNDr. Václav Martínek, Ph.D



Prague 2008

Bea Petrovičová

Declaration

I hereby declare that this bachelor thesis was elaborated individually, under the supervision of RNDr. Václav Martínek, Ph.D and that all used sources were cited properly.

Prague 22nd May 2008

.....

Acknowledgement

I would like to thank my supervisor RNDr. Václav Martínek, Ph.D for his scientific support.

Table of Contents

Abbreviations.....	5
Glossary.....	6
1 Introduction.....	7
1.1 Cytochromes P450.....	7
1.1.1 Occurrence.....	7
1.1.2 Family classification.....	8
1.1.3 Structural features.....	8
1.1.4 Mixed function oxidase.....	9
1.1.5 Substrate and function.....	9
1.1.6 X-ray crystal structures.....	10
1.1.7 Catalytic mechanism.....	11
1.1.8 Cytochromes P450 in human.....	12
1.1.9 CYP 1A subfamily in human.....	13
1.2 Molecular modeling.....	14
1.2.1 Computational algorithms overview.....	14
1.2.2 Molecular mechanics.....	15
1.2.3 Available docking packages.....	15
1.3 AutoDock software package.....	16
1.3.1 Grid map.....	16
1.3.2 Search algorithm.....	16
1.3.3 Scoring function.....	17
1.3.4 Accounting for solvent effects.....	18
1.3.5 Flexible docking in AutoDock 4.....	19
1.3.6 Algorithm calibration.....	19
1.3.7 AutoDockTools (ADT).....	19
2 Aims of the work.....	20
3 Methods.....	21
3.1 Flexible docking I.....	21
3.2 Flexible docking II.....	22
3.3 Flexible docking III.....	22
3.4 Rigid docking I.....	22
3.5 Rigid docking II.....	23
3.6 Implementation notes.....	25
3.6.1 Preparation of ligand file.....	25
3.6.2 Preparation of protein file.....	25
3.6.3 Execution of AutoDock calculation.....	27
3.6.4 Processing AutoDock results.....	27
3.7 Automated execution instructions.....	28
4 Results and discussion.....	29
4.1 Flexible docking I.....	29
4.2 Flexible docking II.....	32
4.3 Flexible docking III.....	36
4.4 Rigid docking I.....	39
4.5 Rigid docking II.....	39
4.6 Hardware requirements note.....	44
5 Conclusions.....	45
6 Bibliography.....	48
7 Appendix AutoDock 4 adaptations.....	51

Abbreviations

ACD - available chemical directory

ADT - auto dock tools

aNF - alpha-naphthoflavone

AMBER - assisted model building and energy refinement

CMC database - comprehensive medical chemistry database

CPU - central processing unit

CYP - cytochrome P450

ESP - electrostatic potential

FEP - free energy perturbation

GUI - graphical user interface

LRA - linear response approximation

MFO – cytochrome P450 mixed function oxidase complex

MD - molecular dynamics

MM - molecular mechanics

NCI - National Cancer Institute

OS - operating system

PDB - protein databank

QM - quantum mechanics

RMSD - root mean square distance

ZINC - zinc is not commercial (an recursive acronym)

Glossary

ab-initio quantum mechanics - a calculation based solely on first principles with no empirical input

activation of xenobiotic - an enzymatic process leading to increased toxicity of a chemical, it is a part of the first phase of a biotransformation of a xenobiotic

active-site - is a catalytic centre of an enzyme, that recognizes and binds a substrate

detoxication of xenobiotic - a enzymatic process resulting in decreased toxicity of a chemical, it is a part of the first phase of a biotransformation of a xenobiotic

intramolecular energy - energy contribution of non-bonding interactions between different parts of a molecule

intermolecular energy - energy contribution of weak interactions between a protein and a ligand

ligand - small molecule, typically either a substrate, an inhibitor or a medical drug

lead compound - an chemical compound with suitable pharmacological activity, which structure is used as a prime for search of new medical drug

source code - human readable form of computer program

xenobiotic – compound foreign to the organism

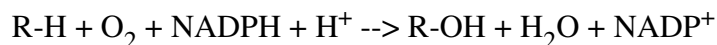
1 Introduction

1.1 Cytochromes P450

Cytochromes P450 (CYP) are ubiquitous haem proteins found in most multicellular organisms and many single cell organisms. The P450 descriptor consists of character, P denoting pigment, and 450, which is the wavelength of absorption maximum in nanometres that applies to a protein with carbon monoxide bound to a ferrous haem in the active-site of an enzyme [1].

CYP enzymes catalyse a wide range of oxygenation, oxidation, peroxidation and reduction reactions [2]. Among others, these reactions include the hydroxylation of saturated aliphatic compounds, epoxidation of double bonds, oxidations of aromatic compounds and oxidation of heteroatoms [3].

The following reaction summarizes typical reaction schema, carried out by CYP [2]:



1.1.1 Occurrence

It is assumed that all cytochromes P450 sprung from a single common predecessor archea gene that appeared about 1.5 billion years ago [2].

The number of genes present varies depending on the organism. See table 1-1.

Table 1-1 Estimated numbers of CYP genes in selected genomes
(adapted from [1,4])

Arabidopsis thaliana	286
Homo Sapiens	57
Caenorhabditis elegans	80
Drosophila melanogaster	83
Mycobacterium tuberculosis	20
Bacillus subtilis	7

Cytochrome P450 genes are often found in clusters. These clusters probably originated in gene duplication or amplification events. Such clusters are present in most organisms and typically contain up to fifteen CYP genes [1,4].

1.1.2 Family classification

Because of the large quantity of CYP genes (over 6500 gene sequences were determined until year 2007) classification into families and subfamilies was introduced. Family and subfamily membership is determined by homology in the primary structure of a protein [5,6]. A primary sequence identity of at least 40% is required for two genes to lie within the same family and over 55% for being classified within the same subfamily [1,5].

Each cytochrome is identified by an ID consisting of a CYP identifier, a family number, a subfamily character and an individual gene number, e.g.: CYP 1A1. Today the number of families exceed 200. Individual families are classified into groups called clans. A clan represents a group of cytochrome P450 families that emerged from single ancestor gene [4].

1.1.3 Structural features

Based on properties and function, cytochromes split into four distinct groups: Class I - Require FAD-containing NADPH-reductase and ferredoxin (low molecular weight iron-sulphur protein functioning as an electron carrier). Class II - Require only FAD-containing NADPH-reductase. Class III - Is self sufficient and requires no electron donor. Class IV - Exchange electrons directly with NADPH [1].

Common examples of CYP occurrence follow: Prokaryotic Class II - Often enable prokaryotes to metabolise xenobiotics as a source of energy. Eukaryotic Class I - In mammals is located in the inner membrane of mitochondria. It is involved in the biosynthesis of steroid hormones and vitamin D3. None of class I has been described in plants. Eukaryotic Class II - Is the most populated class of CYP enzymes. It is located in the outer side of the endoplasmic reticulum membrane and is very diverse in its function [1].

1.1.4 Mixed function oxidase

In recent decades, CYP enzymes have received attention for their involvement in the metabolism of drugs and foreign compounds (xenobiotics). Cytochrome P450 mixed function oxidase (MFO) is a multi-enzymatic plasma-membrane bound complex involved in I. phase of biotransformation of xenobiotics. MFO consists of NADPH-P450 reductase and cytochrome P450 that serves as a terminal oxidase. In some cases cytochrome b5 and cytochrome b5 oxidase are also part of the complex [2].

Exact 3D structure of the complex anchored in a membrane has not been determined yet, nevertheless water soluble domains of several CYP proteins are available in the PDB database.

1.1.5 Substrate and function

As the family of CYP enzymes is diverse and they are ubiquitous throughout nature, it is not surprising that their function is very diverse: Plant CYPs carry out biosynthesis of the blossom pigments, cinnamic acid, vitamin D, steroid hormones or prostaglandins. Another function of CYPs is the production of phytotoxins used in chemical defense as found in plants [1]. In contrast animal CYP enzymes take part in the detoxication and/or an activation of xenobiotics (phytotoxins, pollutants, steroids and medical drugs) [2].

Even though substrates of CYP enzymes are remarkably diverse, they metabolise almost exclusively hydrophobic compounds (including aromatic and poly-aromatic compounds such as sterols and PAHs). Hydrophobic compounds tend to concentrate in the membrane bilayer and adipose tissue. Eukaryotic CYPs are membrane bound. It is therefore expected that the access channel to active-site is oriented toward the plasma membrane [7].

1.1.6 X-ray crystal structures

The first 3D structures of CYP obtained were bacterial enzymes (P450_{cam}, P450_{BM3}) [8]. Until recently, no eukaryotic CYP X-ray structures were determined. In general, eukaryotic enzymes differ from the prokaryotic CYPs by being bound to a membrane bilayer. Incorporation into a membrane and the hydrophobic character of the membrane anchor of CYP proteins affect solubility and results in crystallization difficulties of eukaryotic CYPs [8].

However X-ray structures can be obtained for truncated recombinant CYP proteins without a hydrophobic anchor [9]. Such artificial CYP chimeras facilitate both gene expression and crystallization for purposes of X-ray crystallography [7].

The primary structure of CYPs is extremely variable with homology between two CYPs often as low as 20%. Nevertheless, comparison of available X-ray structures shows same overall folding pattern. Typical CYP structure consists of 12 alpha helices denoted A-L. Some beta-sheets are present as well [3]. See Figure 1-1.

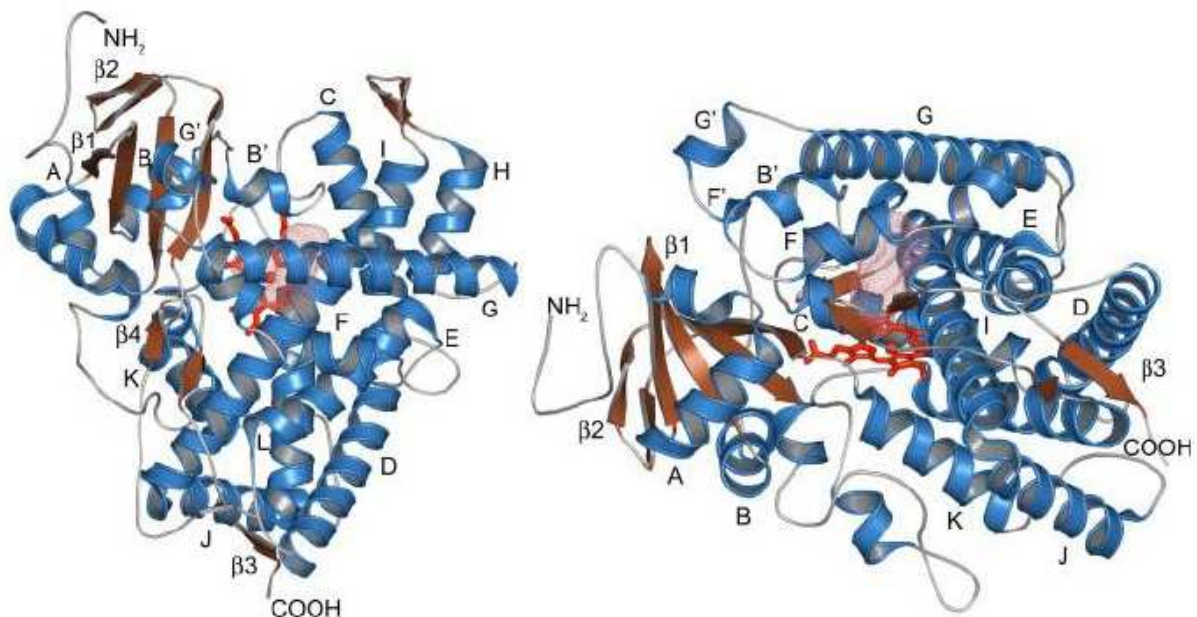


Figure 1-1 Overall folding of CYP 1A2.

Two perspectives are shown, beta-sheets in red, alpha-helices in blue.
(adapted from [9])

1.1.7 Catalytic mechanism

Main structural feature of CYP is the fero-protoporphyrin IX with the thiolate group of cysteine bound to central Fe atom of the haem. In a resting state a water molecule is the sixth ligand of the iron atom [2].

The basic outline of a reaction mechanism begins with four well understood steps [10]: See figure 1-2.

(1) Catalytic cycle begins by a substrate binding to the active-site of the enzyme. Water bound to the Fe atom in the resting state is displaced from the complex with the Fe atom of haem in a ferric(III) state. Displacement of water is accompanied by a transition in the Fe spin (from a high-spin $S=5/2$ to a low-spin $S=1/2$) resulting in the shift in a redox potential from -300 to -170 mV [3,10].

(2) Now the haem Fe atom is reduced to a ferrous(II) state by NADPH-P450 reductase, which transfers a single electron to CYP [10].

(3) Afterwards an O_2 molecule is bound to the Fe site of the ferrous enzyme [10].

(4) One more electron is transferred from NADPH-P450 reductase (or in some cases from cytochrome b5 reductase) resulting in 'activated oxygen species' [1,10].

After this step the catalytic mechanism is only partially understood. However it is broadly accepted that consecutive events lead to the formation of a hydroxylated form of haem (a radical specie) and a radical of a substrate. Recombination of these radicals leads to a hydroxylated substrate and the resting state of the enzyme with the haem Fe atom in a ferric(III) state [10].

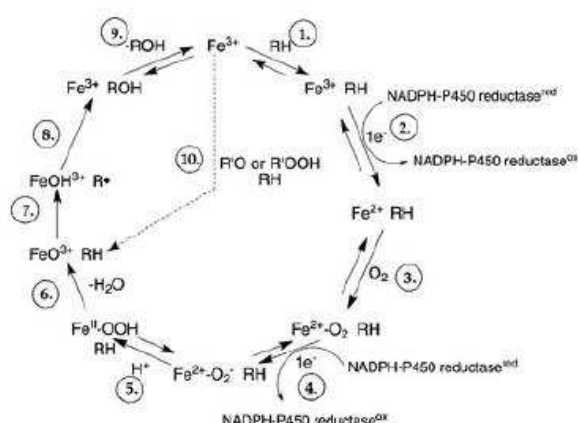


Figure 1-2 Catalytic cycle of Cytochrome P450.
(adapted from [10])

1.1.8 Cytochromes P450 in human

Human genome contains 57 CYP enzymes [11]. A large group of the CYP enzymes involved in the metabolism of xenobiotics is expressed in the liver CYP 1A2, 2A6, 2D6, 2C8, 2C9, 2C18, 2C19, 2E1, 3A4) [2]. Others are expressed in extrahepatic tissues (for example: kidneys, lungs, skin) CYP 1A1, 2B6, 2E1 and 3A4 [2]. Some enzymes such as 2E1 and 3A4 metabolise an especially wide range of substrates [12]. Multiple human enzymes are specialized in the metabolism of steroid hormones (CYP 7, 11A1, 11B1, 17, 19, 21A2). These CYPs have more regular expression patterns among the population, and they also possess higher substrate specificity than CYPs metabolising xenobiotics [12].

Conserved CYP enzymes with homologous substrate specificity are often observed in animals (for example CYP 1A1, 1A2, 2E1), but this is not always true. For this reason animals are not always reliable models of human CYP mediated metabolism [12].

Table 1-2 Some human CYP, localization, substrates inhibitors and inducers.
(adapted from [12])

CYP	localization	substrates	inhibitors	inducers / inducibility
1A1	lungs, other	polycyclic hydrocarbons	7,8-benzoflavone	cigarette smoking
1A2	liver, other	arylamines, arylamides	7,8-benzoflavone	cigarette smoking
2A6	liver	coumarin, 7-ethoxycoumarin		
2B6	liver, lungs	7-ethoxycoumarin	adamantane der.	barbiturates
2E1	liver, other	ethanol, vinyl monomers	4-methylpyrazole	ethanol
3A3	intestine	(steroids, cyclosporin,	gestodene	rifampicin, barbiturates
3A4	liver, other	aflatoxins, many other drugs)		
7	liver	cholesterol		
11A1	steroidogenic tissues	cholesterol		
11B1	steroidogenic tissues	11-deoxycortisol		
19	steroidogenic tissues	testosterone, androstenedione	several drugs	
21A2	steroidogenic tissues	17-hydroxyprogesterone		

1.1.9 CYP 1A subfamily in human

Three members of the CYP 1 family are present in humans. All three are considered to be involved in the activation of carcinogens but experimental evidence is limited. Their expression is regulated through the Ah receptor system, however the mechanism of activation is only partially understood. All of them are competitively inhibited by alpha-naphthoflavone (aNF) [13].

Two of the enzymes are members of the 1A subfamily. Both enzymes are similar in structure and in substrate specificity (a hydrophobic planar polycyclic compounds), but they differ in localization [2]. Genes for both enzymes contain seven exons localised on chromosome 15, and their homology is as high as 70% [13].

CYP 1A1 is expressed especially in the lungs with at least 100-fold inter-individual variability. Gene transcripts were also detected in several extrahepatal organs (skin, intestine, kidney, placenta ...). Although constitutive levels of CYP 1A1 in hepatic tissue are negligible, recent work conducted in our laboratory detected significant levels of CYP1A1 in human hepatic microsomal samples [28,29]. It is established that smoking is an inductor of CYP 1A1, and its involvement in lung carcinogenesis is being evaluated [2,8]. Benzo[a]pyrene is a known substrate of human CYP 1A1, and multiple oxidation products have been observed. Also, many polycyclic hydrocarbons and some heterocyclic aromatic amines are known substrates of CYP 1A1. It seems that involvement of CYP 1A1 in the metabolism of medical drugs is limited [13].

CYP 1A2 is expressed almost exclusively in the liver. It comprises approximately 10-15% of all CYPs present in hepatic tissues, with about 40-fold expression pattern variability. To date, at least thirteen different CYP 1A2 alleles have been described in humans. Among others, inductors of CYP 1A2 include smoking, charbroiled food, cruciferous vegetables and vigorous exercise. CYP 1A2 is known to carry out N3- demethylation of caffeine. N1- and N7- demethylation products were also observed with abundance decreasing in given order [14]. 17beta-estradiol and estrone are known endogenous substrates of CYP 1A2. As noted before, aNF is strong inhibitor of CYP 1A2. Several polycyclic aromatic hydrocarbons containing vinyl or ethynyl substituents are known to be inhibitors of human CYP 1A2 as well. Drug interactions of CYP 1A2 have been described, for example interaction with phenacetin potentially leading to methemoglobinemia [13,15].

1.2 Molecular modeling

The ability to accurately predict the interactions between a protein and a ligand (small molecule, typically either a substrate, an inhibitor or a medical drug) is of interest to chemists. A docking algorithm can be used to identify a ligand binding-site of a protein in cases where it is unknown. An estimation of the free energy of binding is often calculated by docking programs, however this should be viewed as a gross estimation only. More accurate results can be obtained using algorithms with implicit solvent models such as LRA and FEP [16,17].

Secondly, an effective algorithm (both fast and precise enough) can be used to search a directory of chemical compounds, ranking compounds based on their affinity to the active-site of an enzyme. Several 3D drug directories suitable for the purpose are available. For example: ACD - Available Chemical Directory, NCI database of drugs, ZINC database, CMC - Comprehensive Medical Chemistry database [18,19].

Such methodology is already being used by industry for virtual screening to identify lead compounds (chemical compounds with suitable pharmacological activity, which are used as primes in a search for new medical drugs), but might also assist in a search for compounds with a specific metabolic fate, or in the identification of potentially toxic compounds [16].

1.2.1 Computational algorithms overview

This overview is not meant to be complete or exact, but simply informative. Simple descriptions are provided only for the most broadly used algorithms.

Quantum mechanics (QM) algorithms are of two basic types: ab-initio and semi-empirical algorithms. The ab-initio QM can be briefly described as algorithms based on a solution of the Schroedinger equation. The semi-empirical QM reduces the overall complexity of calculation by limiting calculation to valence electron interactions. The complexity of QM calculations is substantial, and this restricts the use of QM to the evaluation of interactions of small molecules, therefore macromolecules can only be treated empirically [20].

1.2.2 Molecular mechanics

Empirical force-field based algorithms, also labelled molecular mechanics (MM) apply the laws of classical physics on an atomic scale. Atoms are idealized as charged points. Interactions are calculated in terms of the bond stretching, altering bond angles and the rotation around single bonds. Penalties are assigned for the deviation from optimal bond lengths, angles, and torsions based on simple functions such as Hooke's Law. Furthermore the contribution of non-bonding interactions is calculated as a sum of vdW and electrostatic terms. As dictated by the nature of calculation, such a model cannot accurately predict properties based on an electronic re-distribution within a molecule or between molecules. On the bright side MM is simple and therefore fast, and can be used to perform energy-minimisation of large systems (e.g.: nucleic acids, or proteins in a solution) [20].

An alternative methodology called molecular dynamics (MD) is also used for evaluation of interactions of macromolecules. The only difference between MM and MD is the fact that MM evaluates static conformations of molecules, while MD considers motion of molecules in the time. MD reuses the MM atomic model, plus it introduces an atomic mass and a momentum. Interactions between atoms are estimated based on application of Newton's laws of motion and the resulting trajectories are analysed [20].

Even with today's power of computers, the requirements of quantum mechanics are too huge to be used for routine scanning. Therefore alternative approaches have been explored. Molecular mechanics is among the most widely adopted algorithms used for modeling interactions of macromolecules with ligands [21].

MM uses search algorithms to generate conformations of enzyme-ligand interactions and an empirical force-field to rapidly evaluate the favourableness of conformations found [21,16,22].

1.2.3 Available docking packages

Numerous molecular mechanics software packages have been developed for the purpose of evaluation of macromolecule-ligand interactions. For example DOCK (Kuntz), GRID (Goodford), FlexX, GOLD, Glide, AMBER and AutoDock [18,19].

1.3 AutoDock software package

Each molecular mechanics package comprises several basic building blocks used together. A brief description of the main docking software components and concepts follows.

1.3.1 Grid map

Either a whole protein or its largest part is held at a constant position during the time of calculation. This speeds up the ranking step by creating a grid-map of a force-field. A grid-map is a common feature used in several docking packages.

AutoDock uses a set of grids with pre-calculated intensity of interaction (a map of interactions of atoms with a force-field). One grid-map is calculated in respect to every atom type present in a ligand, and optionally also for atom types present in flexible residues. Two more grid-maps are generated for electrostatic interactions (e) and for desolvation (d) [23,24].

1.3.2 Search algorithm

The variety of possible protein-ligand conformations is enormous (this is usually referred to as the conformational space). The complexity of conformational space increases with the number of degrees of freedom. A rigid ligand has six degrees of freedom. However most ligands being evaluated have structures that are not rigid and introduce further degrees of freedom. With the increasing number of degrees of freedom, the solution quickly becomes intractable. To tackle the problem, an efficient search algorithm is important to increase probability of finding the most favourable conformation [18,16].

AutoDock implements several search algorithms. The classic algorithm is the Monte Carlo Simulated Annealing. More recent approaches comprise the Genetic Search algorithm and Lamarckian Genetic Search algorithm [16].

The Genetic Search algorithm is based on the principles of Darwinian genetics. The conformation is considered as a genome, while the position of a ligand along with its orientation in space and angles of flexible parts are labelled as genes. Each resulting conformation of a protein-ligand complex is designated as an individual. Afterwards Structure

is randomly recombined based on a cross-over principle, resulting in a mixture of two individuals. The cross-over always occurs between whole genes, to prevent nonsensical combinations. Finally, the resulting conformation is subject to a random mutation based on a predefined probability [16].

The Lamarckian Genetic Search algorithm is based on the idea of Lamarckian genetics that a trait of environmental adaptation of an individual, gained during its lifetime, is passed down to its offspring. Even though this theory does not hold true for life on Earth, it is proven to be an efficient search method in conformation analysis. In this algorithm, the best ranked conformations gain the highest probability of becoming the basis for the next generation of conformations. The rest of the procedure is analogous to the Genetic Search algorithm [16].

It has been shown by Morris et al [16] that the Lamarckian Genetic Search algorithm is the most efficient and the Simulated Annealing the least efficient search algorithm available in the AutoDock software package.

1.3.3 Scoring function

The scoring function can be based either on an empiric forcefield or a theoretical formula of either classical physics or quantum mechanics [16,22].

It is clear that the quality of a scoring function is crucial, as it is the measure that discriminates between accepted and discarded structures of a protein-ligand complex. At the same time, the function has to be computationally plausible to allow a large number of evaluations [18,21].

As the proper conformation often appears among best results, it has been believed that a scoring function capable of correct ranking will be developed eventually. However accurate scoring has proven the most difficult part of docking [21].

The majority of currently used scoring functions estimate the free energy of binding of a protein-ligand complex. The conformation with the lowest free energy of binding is ranked as the best [18,22].

AutoDock uses an AMBER derived scoring function. Here AMBER stands for Assisted Model Building and Energy Refinement. Note that the acronym AMBER is used for both a software package and for a force-field function developed by Peter Kollman's group [16].

Variant of force-field functions used by AutoDock has following form [22]:

$$\Delta G = \Delta G_{intermol} + \Delta G_{internal} + \Delta S_{conf}$$

ΔG = free energy of protein–ligand complex formation

$\Delta G_{intermol}$ = ligand–protein interaction energy

$\Delta G_{internal}$ = energy of intramolecular interactions

ΔS_{conf} = compensation for loss of conformational freedom

For flexible docking, $\Delta G_{internal}$ encompasses interactions between the flexible parts of a ligand as well as for residuum-residuum interactions of flexible residues in a protein:

$$\Delta G_{internal} = \Delta G_{internal}^{ligand} + \Delta G_{internal}^{protein}$$

$\Delta G_{internal}^{ligand}$ = interactions between flexible parts of a ligand

$\Delta G_{internal}^{protein}$ = interactions between flexible residues of a protein

$\Delta G_{intermol}$ comprises three terms, each consisting of two evaluations. The difference between G of unbound and bound states ($V_{bound} - V_{unbound}$) of a ligand, unbound and bound states of a protein-ligand complex ($V_{bound} - V_{unbound}$) and bound and unbound states of a protein [22]:

$$\Delta G_{intermol} = (V_{bound}^{ligand} - V_{unbound}^{ligand}) + (V_{bound}^{protein-ligand} - V_{unbound}^{protein-ligand}) + (V_{bound}^{protein} - V_{unbound}^{protein})$$

Each of six evaluations has following form [22]:

$$V = \Sigma_{vdw} + \Sigma_{hbond} + \Sigma_{elect} + \Sigma_{solv}$$

$$\begin{aligned} \Sigma_{vdw} &= \text{dispersion/repulsion term} \\ \Sigma_{hbond} &= \text{hydrogen bonding forces} \\ \Sigma_{elect} &= \text{electrostatics forces term} \\ \Sigma_{solv} &= \text{desolvation contribution term} \end{aligned}$$

1.3.4 Accounting for solvent effects

When a ligand anneals with a receptor site, part of both the ligand and a protein is desolvated. There are two basic approaches for estimating the contribution of a desolvation. Either explicit solvent molecules are used in a calculation, or the amount of desolvation is estimated based on the buried surface area [22].

Explicit molecules of solvent greatly enhance the complexity of a calculation, therefore they are not suitable for use in a docking.

1.3.5 Flexible docking in AutoDock 4

It is well established that enzymes undergo an active-site reorganization after the binding of a ligand. This is commonly denoted as induced-fit. Contrary to this, most presently used docking applications treat the protein as a rigid body. In these programs only ligand is attributed flexibility.

In attempt to account for the effect of the induced-fit, AutoDock 4 allows a certain amount of flexibility to a protein molecule. For example, a user can enable rotation of a side chain of selected residues in the active-site of a macromolecule [23].

1.3.6 Algorithm calibration

In order to achieve a good agreement with reality, algorithms are calibrated on a large set of experimentally determined protein-ligand complexes. AutoDock 4 has been calibrated on several hundreds of enzyme-ligand structures with experimentally determined free energy of binding [24].

1.3.7 AutoDockTools (ADT)

AutoDock itself is a command-line program that has no graphical user interface (GUI). Preparing input for calculation comprises many steps that are difficult to perform without GUI tools.

ADT is an AutoDock comprehensive GUI suite. It facilitates all steps required to prepare input for AutoDock from Swiss-Prot protein databank (PDB) files.

2 Aims of the work

The primary aim of the study was to evaluate the reliability of the soft-soft (flexible) docking algorithm, recently implemented in the AutoDock 4 software package, with respect to its applicability to modeling interactions between hydrophobic ligand and cytochrome P450.

The second aim was to find the optimal input parameters and to modify the AutoDock program in an attempt to improve the convergence and reproducibility of the flexible docking calculations.

3 Methods

3.1 Flexible docking I

The PDB file 2hi4 of human CYP 1A2 was used as a model enzyme. The structure already contains co-crystallized aNF (inhibitor) in the active-site. The position of the ligand was used as a basis for calculation of a reference RMSD (root mean square distance). See figure 3-1.

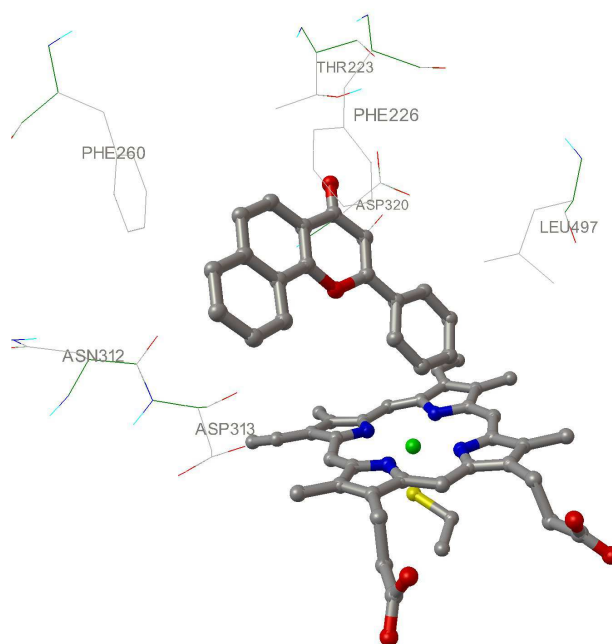


Figure 3-1 Ligand in a crystallographic structure.

In this docking, I attempted a first approach in order to create a maximally generic docking with the least customizations possible. The gasteiger charges were assigned automatically by ADT to all atoms in the protein and the ligand. The removal of four hydrogen atoms incorrectly generated by AutoDockTools on tetrapyrrole nitrogens was necessary. The next step was the merging of non-polar hydrogen atoms with adjacent heavy atoms. Afterwards, based on the proximity to aNF in the crystallographic structure, seven flexible residues were selected. The selected residues were: THR223; PHE226; PHE260; ASN312; ASP313; ASP320; LEU497; The grid-box center in the coordinate system of the macromolecule used in the crystallographic structure was set to $x=0.899$ $y=17.192$ $z=24.271$,

and the grid-box dimensions 60 x 50 x 50 Å were used.

The rest of the preparation was performed in an entirely generic way according to the official AutoDock 4 tutorial [23].

3.2 Flexible docking II

This session relies fully on the flexible docking I. However proper partial charge 0.998 is assigned to the central Fe atom of the haem with an overall haem charge -1.000. ADT assigned no partial charge to the Fe atom of the haem.

3.3 Flexible docking III

In the previous docking it was found that the program contains a bug that makes it difficult to compare results. For this reason, unmodified docking input for Flexible docking II was reused, while the code of the AutoDock program was modified.

The first modification was neglecting the correction to unbound state energy.

Another amendment was suppression of the attractive interactions between protein residues themselves, as they can not be predicted accurately without the explicit solvent model.

Source-code files with the modification are:

autodocksuite-4.0.1/src/autodock/calculateEnergies.cc

autodocksuite-4.0.1/src/autodock-4.0.1/eintcal.cc

Changes were recorded into Autodock/cyp1a2/flex5/autodocksuite-4.0.1-fix5.patch
For more details of application changes see appendix AutoDock 4 adaptations.

3.4 Rigid docking I

Was derived from the Flexible docking II however no flexibility was allowed to protein residues. The partial charge for the haem Fe atom 0.998 was used, with the overall haem charge -1.000.

3.5 Rigid docking II

Ligand geometries and partial charges were obtained via ab-initio methodology as implemented in the GAUSSIAN 03 program suite by Frisch et al. [25], using the Hartree-Fock level of theory in conjunction with the 6-31+G(d) basis set.

For this docking, electrostatic potential (ESP) partial charges were assigned to the ligand and Kollman charges were assigned to the protein. With the exception of Gasteiger partial charges retained for the haem. Details of ESP charges assigned to the aNF molecule are shown in Figure 3-2 and Table 3-1.

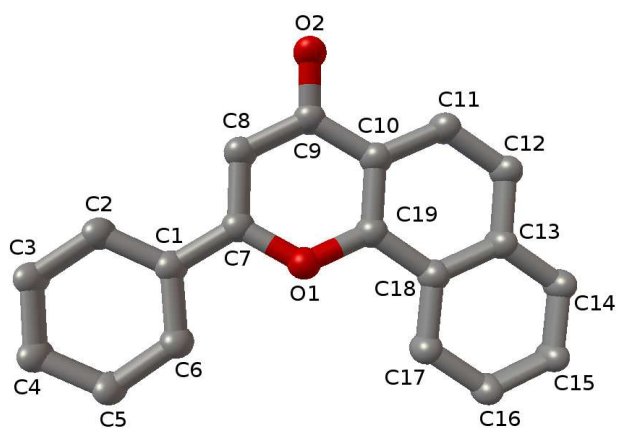


Figure 3-2 Alpha-naphthoflavone

Atom labels as assigned by AutoDockTools pdbqt file.

Table 3-1 Partial charges of alpha-naphthoflavone.

Provided by GAUSSIAN 03 program and those assigned by AutoDock.

atom id	partial charge of atom by ESP	partial charge of adjacent hydrogen atom by ESP	partial charge of atom with merged hydrogen by ESP	atom id	Gasteiger partial charges as set by AutoDock
C1	0.023		0.023	C1	0.023
C2	-0.153	0.155	0.001	C2	0.002
C3	-0.150	0.150	0.000	C3	0.000
C4	-0.125	0.149	0.024	C4	0.000
C5	-0.155	0.151	-0.004	C5	0.000
C6	-0.167	0.162	-0.005	C6	0.002
C7	0.427		0.427	C7	0.103
C8	-0.725	0.251	-0.475	C8	0.067
C9	0.945		0.945	C9	0.200
C10	-0.463		-0.463	C10	0.062
C11	-0.017	0.173	0.156	C11	0.005
C12	-0.461	0.207	-0.254	C12	0.000
C13	0.426		0.426	C13	0.002
C14	-0.372	0.189	-0.183	C14	0.000
C15	-0.076	0.150	0.074	C15	0.000
C16	-0.187	0.158	-0.029	C16	0.000
C17	-0.164	0.155	-0.009	C17	0.002
C18	-0.152		-0.152	C18	0.023
C19	0.498		0.498	C19	0.103
O1	-0.344		-0.344	O1	-0.307
O2	-0.659		-0.659	O2	-0.286

3.6 Implementation notes

First, I attempted to prepare docking of aNF on CYP 1A2 interactively. This proved problematic. Input data preparation consists of many steps using GUI. It is very hard to perform all steps in exactly the same fashion and in a deterministically reproducible way. Also, preparing AutoDock input files is usually an iterative task. After first attempt, improvements or fixes are incorporated and calculation is reiterated.

For these reasons, I prepared consequent dockings in a semi-automated fashion. The preparation of AutoDock input files were split into smaller self consistent sequences of steps that can be easily reproduced. Once a sequence was performed over GUI, it was stored as a raw python script.

Flexible Docking I is a third generation flexible docking, and its files are kept in ~/Autodock/cyp1a2/flex3 folder.

ADT Generated scripts are usually quite complex, encompassing several blocks of information, not strictly required. These contain ADT initialization steps and the 3D OpenGL viewer state. Therefore, from every raw script, I extracted only a short sequence of interest and stored it in a python automation script.

Splitting AutoDock input preparation into short sequences simplifies creation and maximizes reuse of automation scripts. Furthermore, it ensures that only desired improvements are incorporated into additional iterations, and leaves less room for mistakes.

3.6.1 Preparation of ligand file

3D structures available in crystallographic database are molecules with a fixed geometry, found in a crystal. However Autodock requires a ligand structure that has flexible bonds indicated. This allows evaluation of all possible conformations. Information on flexible bonds is kept in a PDBQT format that accounts for flexibility of a molecule.

3.6.2 Preparation of protein file

The structure from the crystallographic database contains the aNF ligand in the active-site. In order to be able to perform ligand docking, the ligand needs to be removed. The structure also comprises many water molecules. Leaving the water molecules fixed in space

does not reflect conditions in the liquid state, and wouldn't leave room for ligand docking. Both the original ligand and all water molecules need to be removed from the crystal structure obtained from the protein databank.

Note: AutoDock algorithm uses empiric method to estimate energy contributions arising from macromolecule and ligand desolvation, caused by surrounding water molecules [23].

Protein structure from protein databank does not contain hydrogen atoms, these have to be added using predictive algorithm. The algorithm fails on prediction of hydrogen atoms for a haem by predicting four extra hydrogens on central nitrogens of the tetrapyrrole and one in a side-chain. These misgenerated hydrogens need to be manually removed.

A central feature of flexible docking is the ability to account for interactions between a ligand and residues in the active-site. AutoDock implements a simple model of induced fit by allowing for the flexibility of protein residues. Adding flexible residues results in an increase of complexity of a problem, therefore flexible residues need to be carefully selected. Candidates for flexible residues were chosen by proximity to active center: THR223; PHE226; PHE260; ASN312; ASP313; ASP320; LEU497;

The last step of preparation of input is to prepare the parameter file for AutoDock. The parameter file contains information such as:

Maximum number of evaluations (ga_num_evals) denotes a maximum length of a search for a single resulting structure.

Maximum number of generations (ga_num_generations) is a maximum number of generations evaluated.

Calculation is finished when either the maximum number of generations or the maximum number of evaluations is reached (whichever criterion is met sooner).

Number of runs (ga_run) is the number of searches and therefore number of returned structures in a result-set.

RMSD tolerance (rmstol) is allowed tolerance for a grouping of results in a result-set into clusters of similar protein-ligand structures.

aNF is a relatively symmetric substrate with RMSD of mirrored structures < 2. For this reason RMSD tolerance was decreased to 1, to differentiate among mirrored structures.

3.6.3 Execution of AutoDock calculation

Initial conformation is selected randomly, afterwards small random movements of substrate and flexible ligands are generated and resulting conformation is evaluated. When a conformation with lower than original energy is found, it is accepted. If energy is higher than original, it is either accepted or refused on a random basis [16].

It is immediately clear that such an approach requires a very high number of evaluations to obtain the best possible conformations.

The nature of AutoDock calculation is statistical, therefore it is good practice to create a larger number of resulting structures and evaluate results in clusters.

Docking parameters are always a compromise between quality of results obtained and available computing resources. A matrix of dockings with these two AutoDock parameters was created, in order to determine optimal number of evaluations required.

Dockings differ in the **maximum number of evaluations** and the **number of runs**. Three different settings for each variable resulted in total of nine combinations (3 x 3).

To ensure reliability, the matrix was generated using sed commands invoked from automation make.

3.6.4 Processing AutoDock results

Results produced by AutoDock were in a range between 500 000 and 40 000 000 characters per calculation. ADT provides user interface for viewing of resulting protein-ligand structures. However raw data are not in a data format suitable for analysis in software such as OpenOffice Calc.

Manual evaluation of this number of results would probably take weeks for each result-set and would be very error prone. For this reason, a batch of awk scripts and an orchestrating Makefile was created, to post-process AutoDock data for OpenOffice. Transformation resulted into a CSV format, suitable for import into OpenOffice Calc. This automation heavily relied on linux make, awk and sed as well as several other linux commands. For more information on linux commands, awk and sed programming refer to [26] and [27].

3.7 Automated execution instructions

I reused ADT automation sequence with minor modifications five times in my calculations. Flexible Docking II and III as well as Rigid docking I and II are altered versions of this docking. The instructions are customized versions of the standard procedure to prepare input for AutoDock as described by [23].

4 Results and discussion

As a large number of structures was produced in every docking (up to 1000 structures per docking run), the identification of structures became important. Every structure was given a unique ID consisting of a code for docking type and a sequential number of docked structure.

The first character of ID describes the **number of runs** for a docking and therefore the number of results produced: S (small) = 10, M (medium) = 100, L (large) = 1000.

The second character of ID describes the **maximum number of evaluations** for a docking and therefore the quality of a docking: S (short) = 250 000, M (medium) = 2 500 000, L (long) = 25 000 000.

The first number is a sequence number of a structure and the ranking of a structure is denoted by the cluster number and the sequential number in a cluster. The cluster number denotes the ranking of the structure (first cluster is best).

For example ML7 2_1 means medium number of runs, long docking, seventh structure generated, and the first node in the second cluster.

4.1 Flexible docking I

Binding free energies obtained by this flexible docking were unusually high, ranging from -20.94 to -22.12 kcal/mol. See Table 4-1 a). The estimated free energies show unexpected tendency (resulting values of free energy are growing with increasing number of calculations). See Figure 4-1 a). In contrast, the intermolecular energies significantly improve as the number of evaluations increase from 250 000 to 2 500 000. See Figure 4-1 c).

AutoDock correctly placed aNF ligand into the active-site of CYP 1A2 as found in the crystallographic structure, however the ligand was in a flipped orientation. This is reflected in RMSD of docked ligand structures, between 1 and 2 Å. See Figure 4-1 b).

Also, from Figure 4-1 d) it is clear that structures with much lower RMSD were found, however they were not ranked among the best results. This indicates a problem with the scoring function or the forcefield model in general.

Table 4-1 Results of flexible docking I.

a) Estimated Free Energy of Binding [kcal/mol]

runs	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
10	-21.32 [SS1 1_1]**	-21.96 [SM10 1_1]	-22.12 [SL5 1_1]
100	-21.3 [MS66 1_1]	-21.89 [MM18 1_1]	-21.31 [ML30 1_1]
1000	-20.94 [LS333 1_1]	-21.26 [LM462 1_1]	-

b) Reference RMSD [Å]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	0.53 [SS1 1_1]	0.49 [SM10 1_1]	1.58 [SL5 1_1]
100	1.67 [MS66 1_1]	1.64 [MM18 1_1]	1.59 [ML30 1_1]
1000	1.01 [LS333 1_1]	1.58 [LM462 1_1]	-

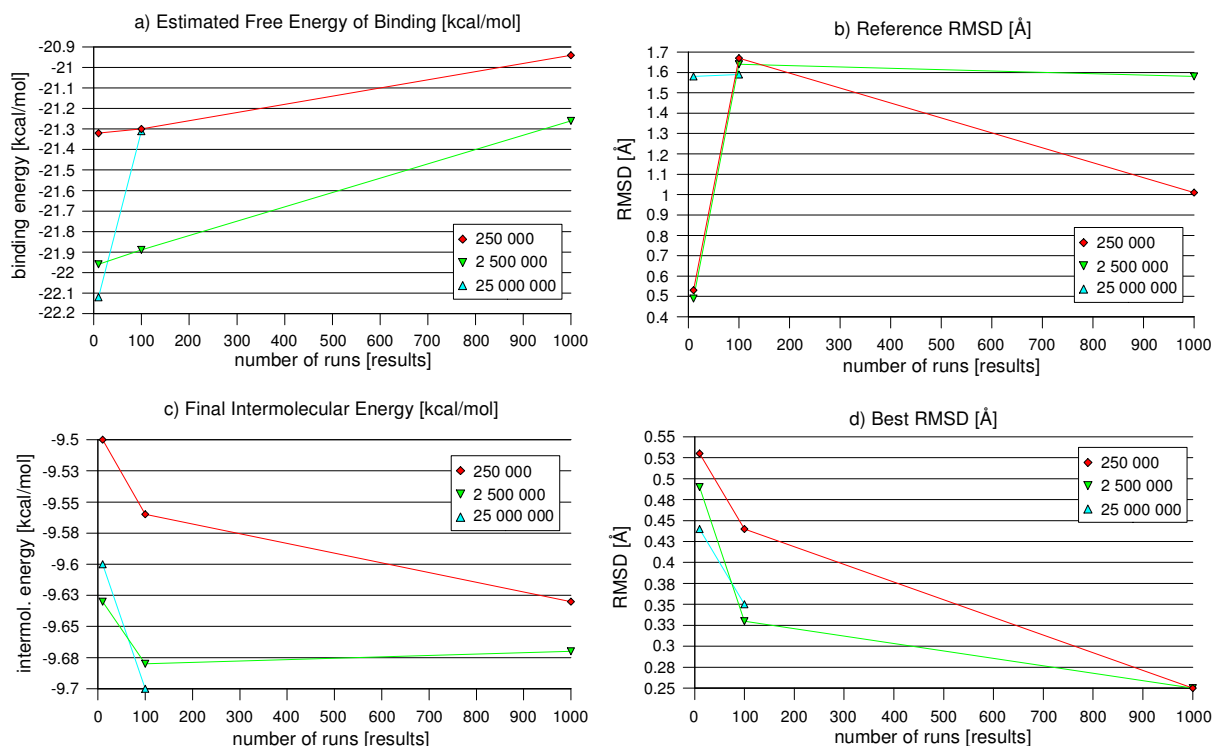
c) Final Intermolecular Energy [kcal/mol]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	-9.5 [SS1 1_1]	-9.63 [SM10 1_1]	-9.6 [SL5 1_1]
100	-9.56 [MS66 1_1]	-9.68 [MM18 1_1]	-9.7 [ML30 1_1]
1000	-9.63 [LS333 1_1]	-9.67 [LM462 1_1]	-

d) Best RMSD [Å]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	0.53 [SS1 1_1]	0.49 [SM10 1_1]	0.44 [SL6 3_1]
100	0.44 [MS53 3_19]	0.33 [MM39 3_1]	0.35 [ML39 2_8]
1000	0.25 [LS929 3_21]	0.25 [LM579 2_24]	-

* Maximum number of evaluations. ** Unique structure ID and position in cluster.

**Figure 4-1 Results of flexible docking I.**

Trends for varying number of runs are shown on X axis and variable number of evaluations is presented in different colors.

Table 4-2 Results of flexible docking I.

All shown data is for number of runs = 100.

a) Estimated Free Energy of Binding [kcal/mol]

cluster	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
1	-21.3 [MS66 1_1]**	-21.89 [MM18 1_1]	-21.31 [ML30 1_1]
2	-20.99 [MS50 2_1]	-21.88 [MM22 2_1]	-21.28 [ML7 2_1]
3	-20.73 [MS44 3_1]	-21.87 [MM39 3_1]	-21.27 [ML9 3_1]

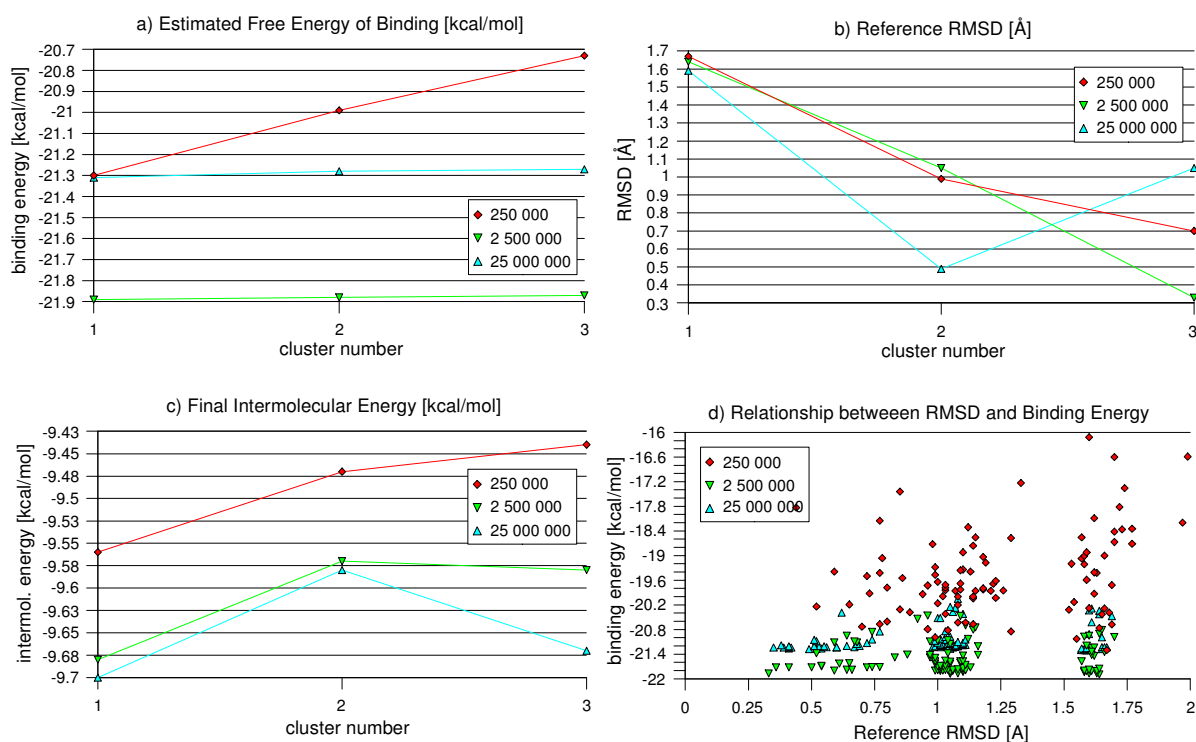
b) Reference RMSD [$\text{\AA}$]

cluster	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
1	1.67 [MS66 1_1]	1.64 [MM18 1_1]	1.59 [ML30 1_1]
2	0.99 [MS50 2_1]	1.05 [MM22 2_1]	0.49 [ML7 2_1]
3	0.7 [MS44 3_1]	0.33 [MM39 3_1]	1.05 [ML9 3_1]

c) Final Intermolecular Energy [kcal/mol]

cluster	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
1	-9.56 [MS66 1_1]	-9.68 [MM18 1_1]	-9.7 [ML30 1_1]
2	-9.47 [MS50 2_1]	-9.57 [MM22 2_1]	-9.58 [ML7 2_1]
3	-9.44 [MS44 3_1]	-9.58 [MM39 3_1]	-9.67 [ML9 3_1]

* Maximum number of evaluations. ** Unique structure ID and position in cluster.

**Figure 4-2 Results of flexible docking I.**

Displayed data is for number of runs = 100 only. Trends for first three cluster heads are shown on X axis and for varying number of evaluations by different colors.

Estimated free energies of individual structures within a single calculation are very close in case of 2 500 000 and 25 000 000 evaluations. Figure 4-2 d). However results for 250 000 evaluations vary considerably. An interpretation of this result is that 2 500 000 evaluations is a sufficient number for the flexible docking, however 250 000 is not, as most of structures found are not close to the structure with the lowest free energy.

Many potential protein-ligand complexes are generated during every docking run (between 10 and 1000). But most of them represent nearly the same structure, with only slight shift in coordinates. To group similar structures a clustering analysis is performed. Table 4-2 a), b), c) and Figure 4-2 a), b), c) summarize data of the best three clusters. For 2 500 000 and 25 000 000 evaluations, these three clusters represent three distinct ligand orientations showing nearly the same binding free energies, varying < 0.5 kcal/mol, therefore they should be treated as equivalent results.

I tried to address the problem with the forcefield model by evaluating electrostatic potential in the vicinity of ligand. The most apparent discrepancy found was the partial charge of the iron atom of the haem cofactor. This partial charge was incorrectly assigned by ADT as zero. This issue was addressed in the next generation of flexible docking.

4.2 Flexible docking II

Input parameters in flexible docking II were the same as in flexible docking I, however proper partial charge 0.998 was assigned to the central Fe atom of the haem. Therefore, overall charge of the haem residue changed from -2 to -1.

Nevertheless, results of flexible docking II reveals that the Fe atom charge plays little role in the AutoDock algorithm and very similar results were obtained. See Figures 4-3, 4-4 and Tables 4-3, 4-4. This is consistent with the fact that the electrostatic term plays a minor role in the AutoDock scoring function.

Differences between individual docking runs are caused by a bug in calculation of the lowest energy of unbound state. Energy shifts between different dockings are clearly visible in Figure 4-4 d). Similar results of the free energy of binding cause random selection of the best result, which becomes the seed of the first cluster. This results in large differences in the RMSD, relative to the reference structure from the crystal. See Table and Figure 4-3 b).

Table 4-3 Results of flexible docking II.

a) Estimated Free Energy of Binding [kcal/mol]

runs	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
10	-21.34 [SS1 1_1]**	-21.39 [SM1 1_1]	-21.91 [SL5 1_1]
100	-21.22 [MS81 1_1]	-21.25 [MM46 1_1]	-22.05 [ML79 1_1]
1000	-21.63 [LS876 1_1]	-21.2 [LM216 1_1]	-

b) Reference RMSD [$\text{\AA}$]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	1.66 [SS1 1_1]	1.59 [SM1 1_1]	0.39 [SL5 1_1]
100	1 [MS81 1_1]	1.6 [MM46 1_1]	1.64 [ML79 1_1]
1000	1 [LS876 1_1]	1.64 [LM216 1_1]	-

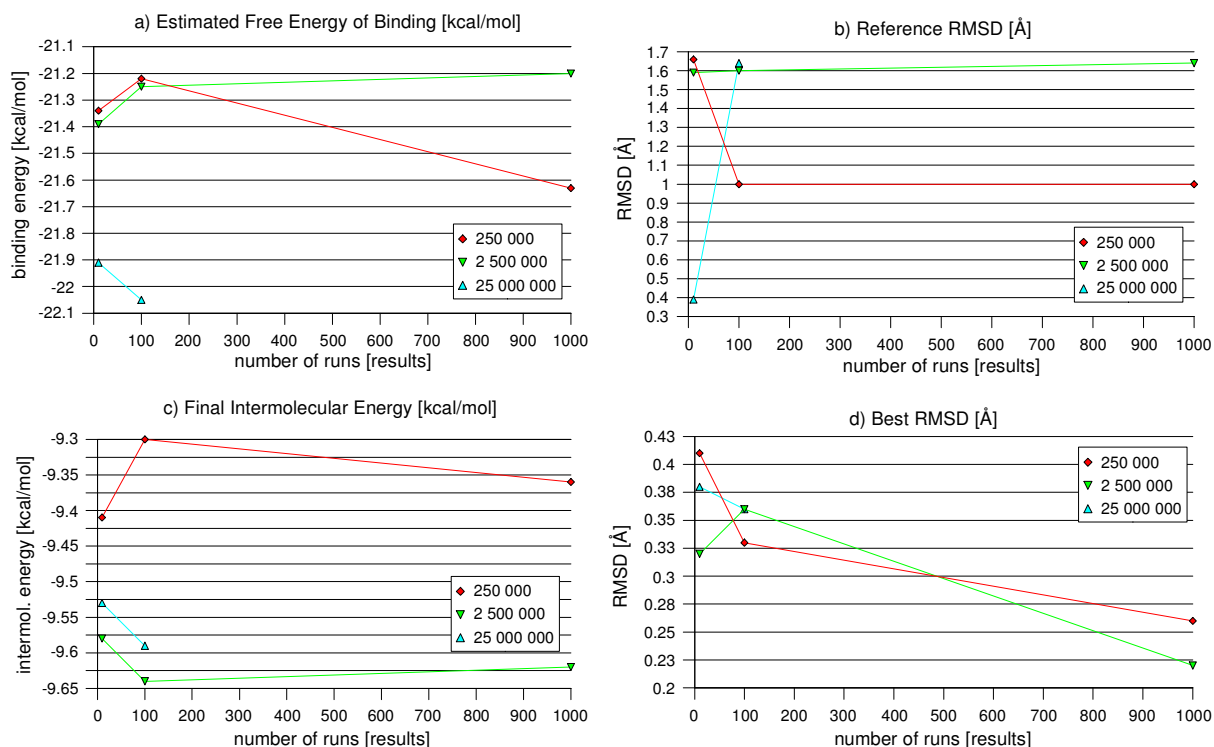
c) Final Intermolecular Energy [kcal/mol]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	-9.41 [SS1 1_1]	-9.58 [SM1 1_1]	-9.53 [SL5 1_1]
100	-9.3 [MS81 1_1]	-9.64 [MM46 1_1]	-9.59 [ML79 1_1]
1000	-9.36 [LS876 1_1]	-9.62 [LM216 1_1]	-

d) Best RMSD [$\text{\AA}$]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	0.41 [SS9 3_2]	0.32 [SM5 3_2]	0.38 [SL4 1_2]
100	0.33 [MS70 3_2]	0.36 [MM42 3_14]	0.36 [ML18 2_8]
1000	0.26 [LS949 3_57]	0.22 [LM565 2_100]	-

* Maximum number of evaluations. ** Unique structure ID and position in cluster.

**Figure 4-3 Results of flexible docking II.**

Trends for varying number of runs are shown on X axis and variable number of evaluations is presented in different colors.

Table 4-4 Results of flexible docking II.

All shown data is for number of runs = 100.

a) Estimated Free Energy of Binding [kcal/mol]

cluster	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
1	-21.22 [MS81 1_1]**	-21.25 [MM46 1_1]	-22.05 [ML79 1_1]
2	-21.17 [MS50 2_1]	-21.2 [MM47 2_1]	-22.01 [ML47 2_1]
3	-21.13 [MS19 3_1]	-21.18 [MM96 3_1]	-21.99 [ML90 3_1]

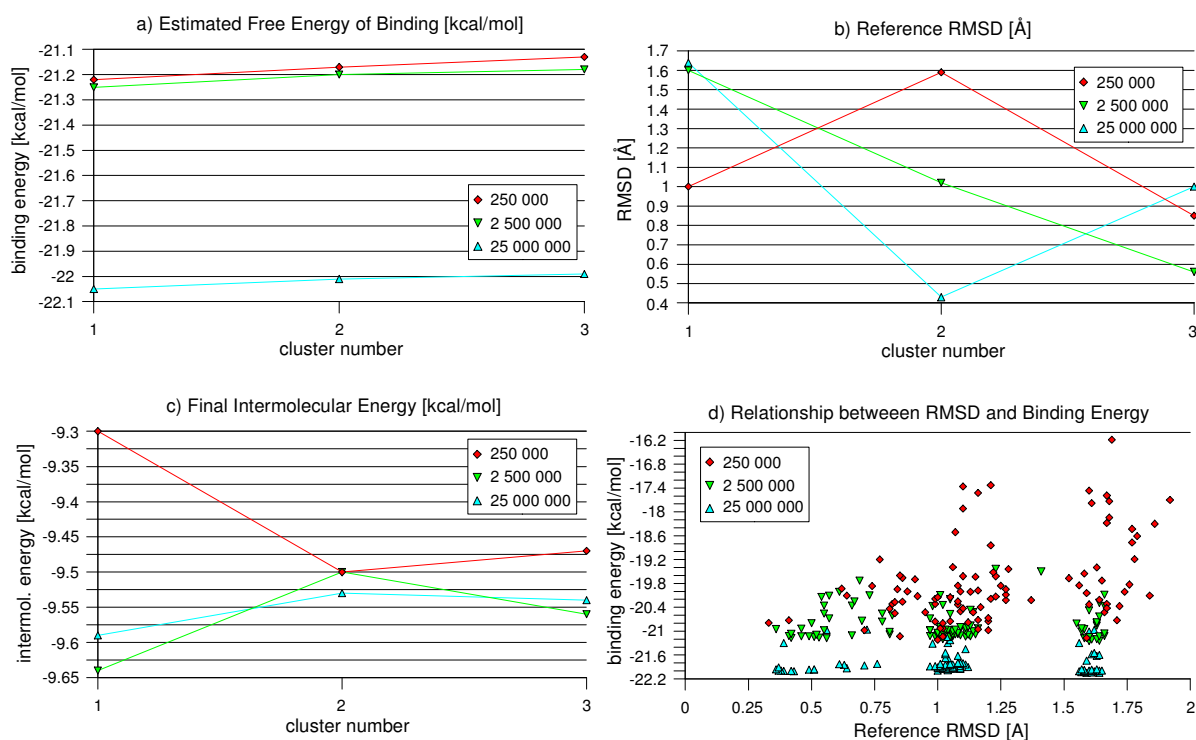
b) Reference RMSD [$\text{\AA}$]

cluster	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
1	1 [MS81 1_1]	1.6 [MM46 1_1]	1.64 [ML79 1_1]
2	1.59 [MS50 2_1]	1.02 [MM47 2_1]	0.43 [ML47 2_1]
3	0.85 [MS19 3_1]	0.56 [MM96 3_1]	1 [ML90 3_1]

c) Final Intermolecular Energy [kcal/mol]

cluster	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
1	-9.3 [MS81 1_1]	-9.64 [MM46 1_1]	-9.59 [ML79 1_1]
2	-9.5 [MS50 2_1]	-9.5 [MM47 2_1]	-9.53 [ML47 2_1]
3	-9.47 [MS19 3_1]	-9.56 [MM96 3_1]	-9.54 [ML90 3_1]

* Maximum number of evaluations. ** Unique structure ID and position in cluster.

**Figure 4-4 Results of flexible docking II.**

Displayed data is for number of runs = 100 only. Trends for first three cluster heads are shown on X axis and for varying number of evaluations by different colors.

The major problem of flexible docking I and II is the error affecting their predicted free energies of binding. The energy of the unbound state for both the protein and the ligand is estimated prior to the calculation of bound states. Unbound state evaluation is performed only once for each program run and length of this calculation is insufficient. This results in random values of the unbound state – differing approximately 1 kcal/mol, being deducted from resultant free energies. See Table 4-5 and Figure 4-5.

Table 4-5 Variable energy of unbound states for Flexible docking I and II.

a) Estimated free energy of unbound state for flexible docking I [kcal/mol]

runs	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
10	-1.380 [SS]**	-1.098 [SM]	-1.303 [SL]
100	-1.416 [MS]	-1.194 [MM]	-1.888 [ML]
1000	-1.861 [LS]	-1.884 [LM]	-

a) Estimated free energy of unbound state for flexible docking II [kcal/mol]

runs	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
10	-1.121 [SS]**	-1.779 [SM]	-1.303 [SL]
100	-1.224 [MS]	-1.917 [MM]	-1.216 [ML]
1000	-1.242 [LS]	-2.017 [LM]	-

* Maximum number of evaluations. ** Calculation ID.

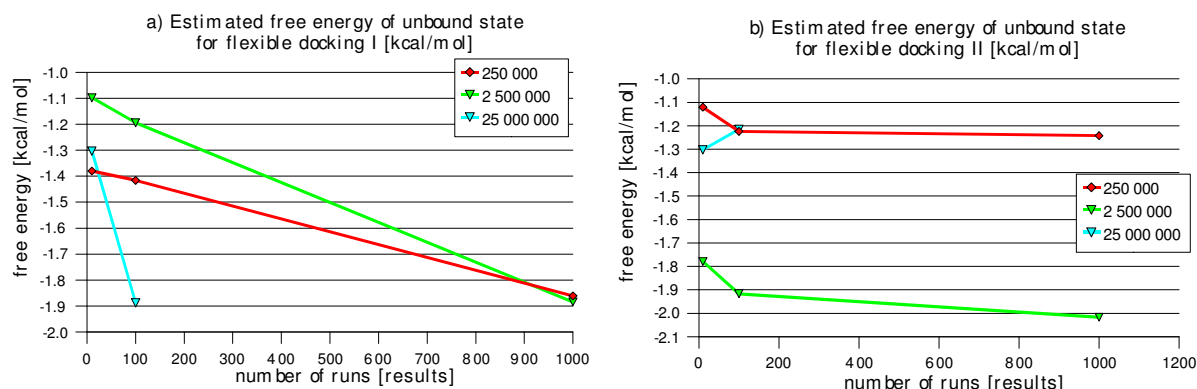


Figure 4-5 Variable energy of unbound states for Flexible docking I and II.

The second reason for problematic convergence of flexible docking I and II can be the overestimated contribution of intramolecular interactions between residues in the active-site. The sum of contributions of these seven residues is larger than the intermolecular component of ligand-protein interactions. For this reason the free energy of binding is dominated by the less precise internal energy of the protein. This led AutoDock into optimising binding free energy via optimisation of the internal energy of the protein, instead of improving the protein-ligand interaction energy.

4.3 Flexible docking III

To resolve the issues described above, the source code of AutoDock 4.0.1 was altered in order to attempt to improve the reliability of the calculation.

In order to eliminate the random error introduced by correction to unbound state, the energy of the unbound state for both the protein and the ligand were removed from the intermolecular part of the AutoDock scoring function, resulting in following form:

$$\Delta G_{intermol} = V_{bound}^{ligand} + V_{bound}^{protein-ligand} + V_{bound}^{protein}$$

The second issue related to high contribution of protein intramolecular energy, was solved by neglecting the negative values of intramolecular interactions. Positive (repulsive) intramolecular interaction energies were kept to exclude all energetically unfavourable structures with overlapping protein residues:

$$\Delta G_{internal} = \Delta G_{internal}^{ligand} + \Delta^+ G_{internal}^{protein}$$

$$\Delta G_{internal}^{ligand} = \text{interactions between flexible parts of a ligand}$$

$$\Delta^+ G_{internal}^{protein} = \text{unfavourable interactions between flexible residues of a protein}$$

It is evident that the first algorithm fix, which removed compensation for the lowest energy of the unbound state, resulted in better comparability of results, see Figure 4-7 c), as all results from different dockings converge toward similar lowest energies. Predicted binding energies are improving with increasing number of evaluations, see Figure 4-6 a).

A difference in best intermolecular energies was found for short calculations and long calculations – long calculation actually have worse results, indicating that the fix to AutoDock was not sufficient. See Figure 4-7 a).

Ranking in a clustering step was fixed, however ranking in a search was not, and this caused the best intermolecular energy to be overridden by the intramolecular energy. This supports the hypothesis that the incorporation of intramolecular energy into the calculation does not improve overall results. A simple workaround is to perform many short calculations. Due to the statistical nature of the calculations, a few best conformations will then be found.

Table 4-6 Results of flexible docking III.

a) Final Intermolecular Energy [kcal/mol]

runs	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
10	-9.76 [SS7 1_1]**	-9.7 [SM7 1_1]	-9.66 [SL3 1_1]
100	-9.88 [MS81 1_1]	-9.72 [MM80 1_1]	-9.72 [ML48 1_1]
1000	-9.87 [LS114 1_1]	-9.78 [LM964 1_1]	-

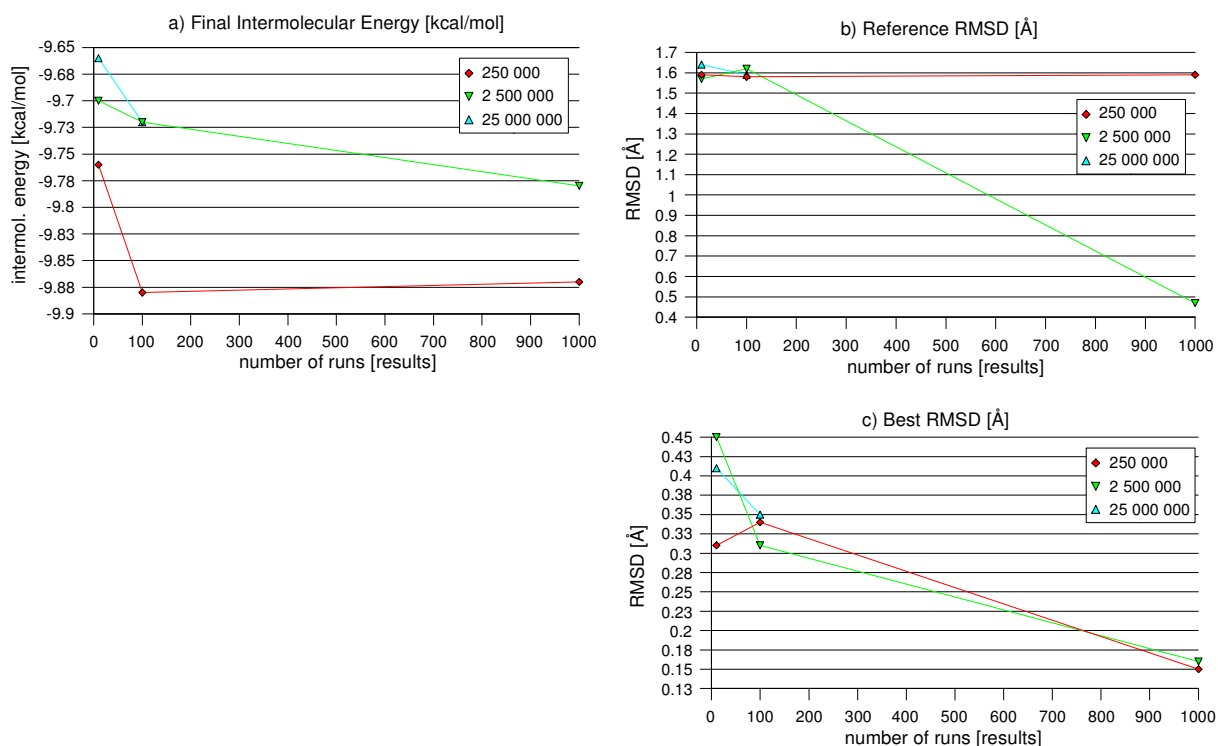
b) Reference RMSD [Å]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	1.59 [SS7 1_1]	1.57 [SM7 1_1]	1.64 [SL3 1_1]
100	1.58 [MS81 1_1]	1.62 [MM80 1_1]	1.59 [ML48 1_1]
1000	1.59 [LS114 1_1]	0.47 [LM964 1_1]	-

c) Best RMSD [Å]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	0.31 [SS9 2_2]	0.45 [SM3 3_1]	0.41 [SL9 3_2]
100	0.34 [MS51 3_5]	0.31 [MM24 3_25]	0.35 [ML90 3_15]
1000	0.15 [LS808 2_57]	0.16 [LM353 1_180]	-

* Maximum number of evaluations. ** Unique structure ID and position in cluster.

**Figure 4-6 Results of flexible docking III.**

Trends for varying number of runs are shown on X axis and variable number of evaluations is presented in different colors.

Table 4-7 Results of flexible docking III.

All shown data is for number of runs = 100.

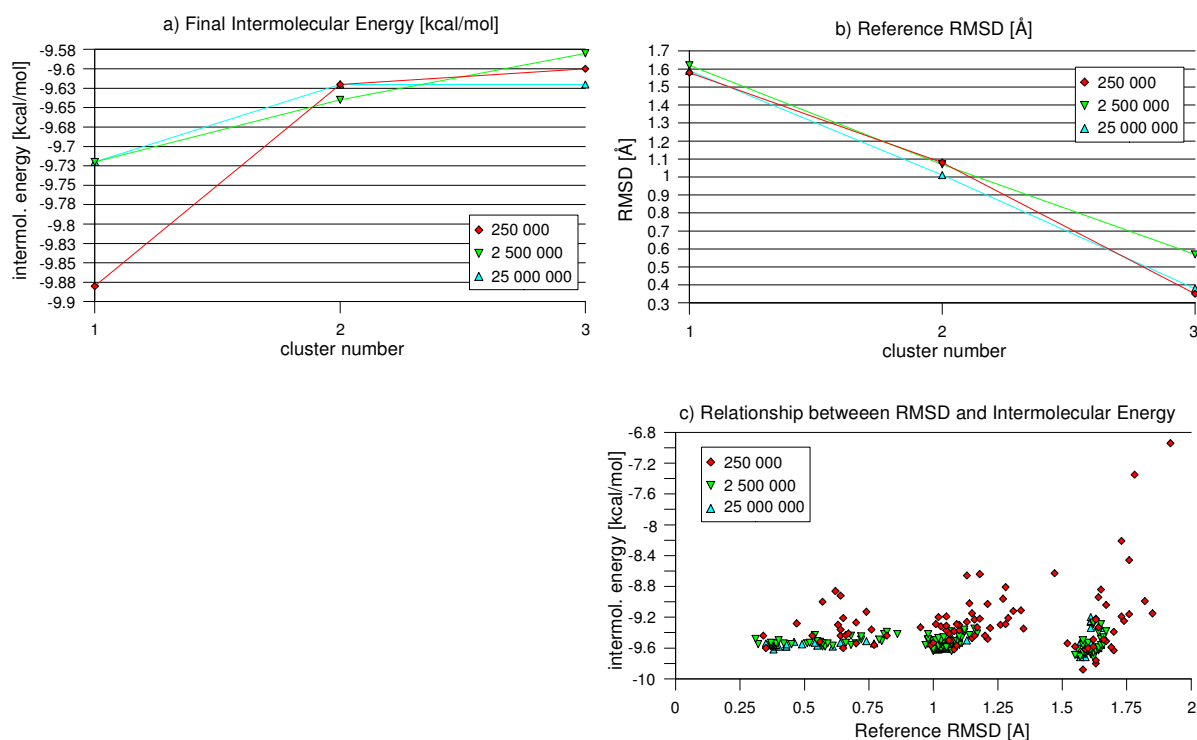
a) Final Intermolecular Energy [kcal/mol]

cluster	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
1	-9.88 [MS81 1_1]**	-9.72 [MM80 1_1]	-9.72 [ML48 1_1]
2	-9.62 [MS35 2_1]	-9.64 [MM67 2_1]	-9.62 [ML17 2_1]
3	-9.6 [MS91 3_1]	-9.58 [MM21 3_1]	-9.62 [ML92 3_1]

b) Reference RMSD [$\text{\AA}$]

cluster	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
1	1.58 [MS81 1_1]	1.62 [MM80 1_1]	1.59 [ML48 1_1]
2	1.08 [MS35 2_1]	1.07 [MM67 2_1]	1.01 [ML17 2_1]
3	0.35 [MS91 3_1]	0.57 [MM21 3_1]	0.38 [ML92 3_1]

* Maximum number of evaluations. ** Unique structure ID and position in cluster.

**Figure 4-7 Results of flexible docking III.**

Displayed data is for number of runs = 100 only. Trends for first three cluster heads are shown on X axis and for varying number of evaluations by different colors.

Important result is an improvement in match between structures predicted by flexible docking III and crystallographic structure of evaluated protein-ligand complex. Although, the first cluster RMSD was 1.6 Å, the third cluster, with energy just 0.1 kcal/mol higher, was well matching the experimental structures with RMSD value 0.38 Å. See Table and Figure 4-7 b).

Experimental evidence suggests that CYP 1A2 produces multiple products from single precursor. For example in case of CYP 1A2 mediated metabolism of caffeine three hydroxylation products were detected [14]. Possibly several aNF orientations might be present at room temperature in the complex with CYP 1A2.

4.4 Rigid docking I

Unlike flexible dockings, rigid dockings do not suffer from a bug in the calculation of the unbound state. See Figures 4-8, 4-9 and Tables 4-8, 4-9. This is in agreement with expectations, as rigid docking is available for several AutoDock versions and is already a well established algorithm. However the problem being processed is much simpler, because input protein structure was based on crystal structure of CYP 1A2-aNF complex. Therefore, the enzyme was already optimized for the aNF and optimal orientation of residues in active-site are already present in the crystal structure.

4.5 Rigid docking II

Results of rigid docking II exhibit similar tendencies as results of rigid docking I. The one exception was the increase in values of the binding free energy from approximately -10 to -7 kcal/mol.

Using GAUSSIAN ESP partial charges does not improve results. See Figures 4-10, 4-11 and Tables 4-10, 4-11.

In case of this docking results were highly reproducible with matching clusters and energies. The result with RMSD as low as 0.48 was found among first three clusters. The first three clusters were also the most populated clusters.

Table 4-8 Results of rigid docking I.

a) Estimated Free Energy of Binding [kcal/mol]

runs	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
10	-10.31 [SS7 1_1]**	-10.37 [SM1 1_1]	-10.38 [SL7 1_1]
100	-10.36 [MS95 1_1]	-10.39 [MM62 1_1]	-10.39 [ML78 1_1]
1000	-10.38 [LS341 1_1]	-	-

b) Reference RMSD [Å]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	1.56 [SS7 1_1]	1.56 [SM1 1_1]	1.56 [SL7 1_1]
100	1.55 [MS95 1_1]	1.56 [MM62 1_1]	1.56 [ML78 1_1]
1000	1.56 [LS341 1_1]	-	-

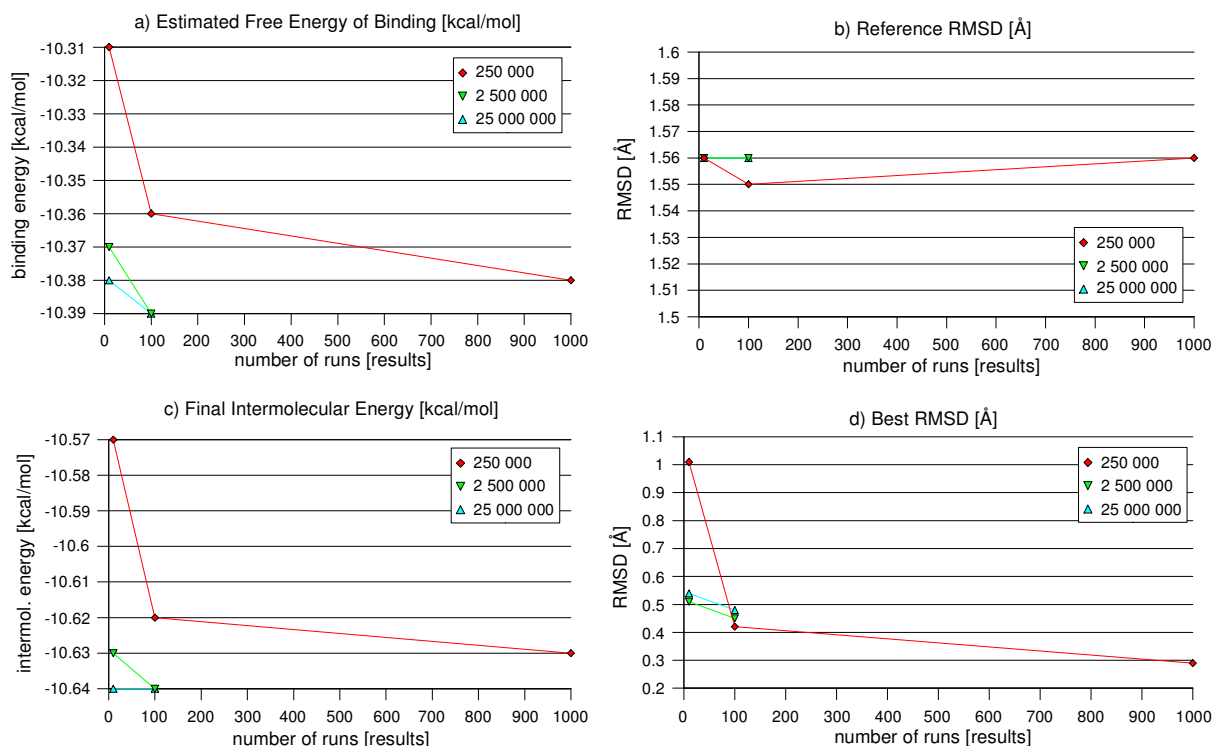
c) Final Intermolecular Energy [kcal/mol]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	-10.57 [SS7 1_1]	-10.63 [SM1 1_1]	-10.64 [SL7 1_1]
100	-10.62 [MS95 1_1]	-10.64 [MM62 1_1]	-10.64 [ML78 1_1]
1000	-10.63 [LS341 1_1]	-	-

d) Best RMSD [Å]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	1.01 [SS4 2_1]	0.51 [SM10 2_1]	0.54 [SL3 2_1]
100	0.42 [MS46 2_3]	0.45 [MM68 2_9]	0.48 [ML7 2_11]
1000	0.29 [LS207 2_39]	-	-

* Maximum number of evaluations. ** Unique structure ID and position in cluster.

**Figure 4-8 Results of rigid docking I.**

Trends for varying number of runs are shown on X axis and variable number of evaluations is presented in different colors.

Table 4-9 Results of rigid docking I.

All shown data is for number of runs = 100.

a) Estimated Free Energy of Binding [kcal/mol]

cluster	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
1	-10.36 [MS95 1_1]**	-10.39 [MM62 1_1]	-10.39 [ML78 1_1]
2	-10.23 [MS23 2_1]	-10.25 [MM84 2_1]	-10.25 [ML31 2_1]
3	-10.19 [MS49 3_1]	-10.2 [MM15 3_1]	-10.21 [ML95 3_1]

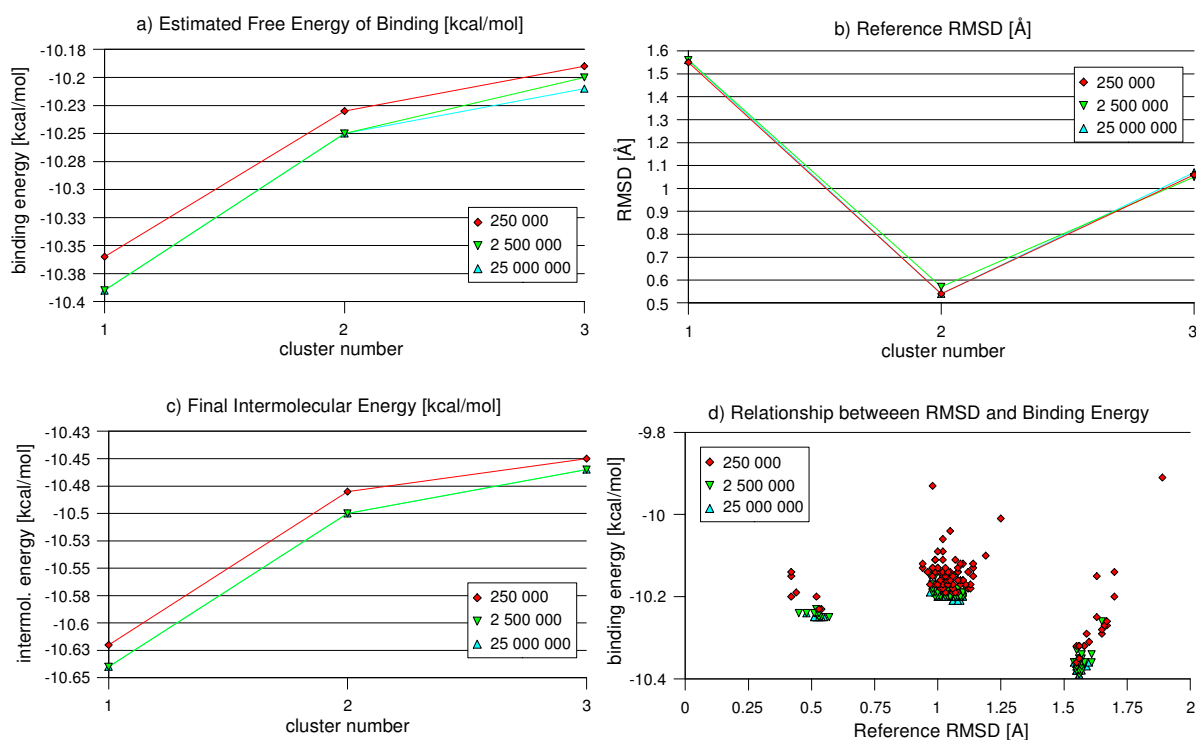
b) Reference RMSD [$\text{\AA}$]

cluster	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
1	1.55 [MS95 1_1]	1.56 [MM62 1_1]	1.56 [ML78 1_1]
2	0.54 [MS23 2_1]	0.57 [MM84 2_1]	0.54 [ML31 2_1]
3	1.06 [MS49 3_1]	1.05 [MM15 3_1]	1.07 [ML95 3_1]

c) Final Intermolecular Energy [kcal/mol]

cluster	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
1	-10.62 [MS95 1_1]	-10.64 [MM62 1_1]	-10.64 [ML78 1_1]
2	-10.48 [MS23 2_1]	-10.5 [MM84 2_1]	-10.5 [ML31 2_1]
3	-10.45 [MS49 3_1]	-10.46 [MM15 3_1]	-10.46 [ML95 3_1]

* Maximum number of evaluations. ** Unique structure ID and position in cluster.

**Figure 4-9 Results of rigid docking I.**

Displayed data is for number of runs = 100 only. Trends for first three cluster heads are shown on X axis and for varying number of evaluations by different colors.

Table 4-10 Results of rigid docking II.

a) Estimated Free Energy of Binding [kcal/mol]

runs	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
10	-7.04 [SS6 1_1]**	-7.09 [SM1 1_1]	-7.12 [SL8 1_1]
100	-7.11 [MS63 1_1]	-7.12 [MM11 1_1]	-7.12 [ML15 1_1]
1000	-7.1 [LS18 1_1]	-7.12 [LM230 1_1]	-

b) Reference RMSD [Å]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	1.58 [SS6 1_1]	1.54 [SM1 1_1]	1.54 [SL8 1_1]
100	1.56 [MS63 1_1]	1.55 [MM11 1_1]	1.54 [ML15 1_1]
1000	1.56 [LS18 1_1]	1.55 [LM230 1_1]	-

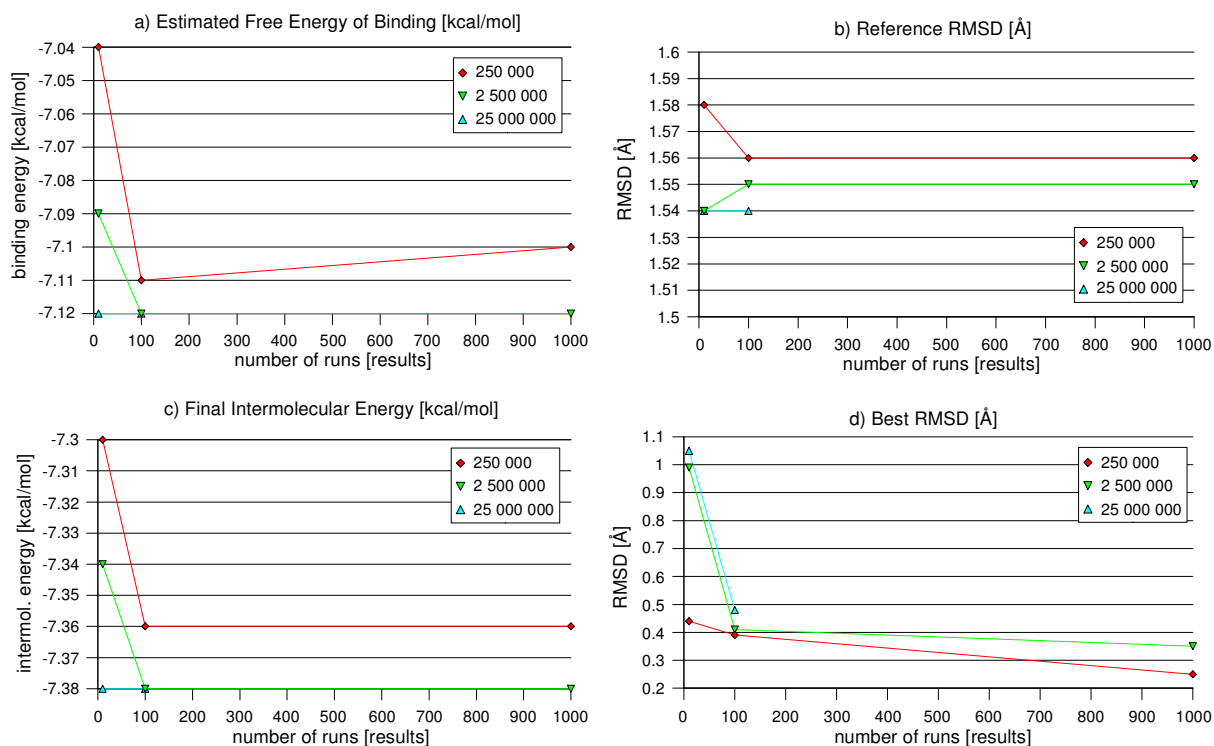
c) Final Intermolecular Energy [kcal/mol]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	-7.3 [SS6 1_1]	-7.34 [SM1 1_1]	-7.38 [SL8 1_1]
100	-7.36 [MS63 1_1]	-7.38 [MM11 1_1]	-7.38 [ML15 1_1]
1000	-7.36 [LS18 1_1]	-7.38 [LM230 1_1]	-

d) Best RMSD [Å]

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	0.44 [SS4 2_1]	0.99 [SM2 2_8]	1.05 [SL2 2_7]
100	0.39 [MS16 3_10]	0.41 [MM16 3_6]	0.48 [ML22 3_9]
1000	0.25 [LS676 3_74]	0.35 [LM667 3_77]	-

* Maximum number of evaluations. ** Unique structure ID and position in cluster.

**Figure 4-10 Results of rigid docking II.**

Trends for varying number of runs are shown on X axis and variable number of evaluations is presented in different colors.

Table 4-11 Results of rigid docking II.

All shown data is for number of runs = 100.

a) Estimated Free Energy of Binding [kcal/mol]

cluster	250 000 evaluations*	2 500 000 evaluations	25 000 000 evaluations
1	-7.11 [MS63 1_1]**	-7.12 [MM11 1_1]	-7.12 [ML15 1_1]
2	-6.93 [MS22 2_1]	-6.93 [MM13 2_1]	-6.93 [ML23 2_1]
3	-6.91 [MS93 3_1]	-6.92 [MM36 3_1]	-6.92 [ML84 3_1]

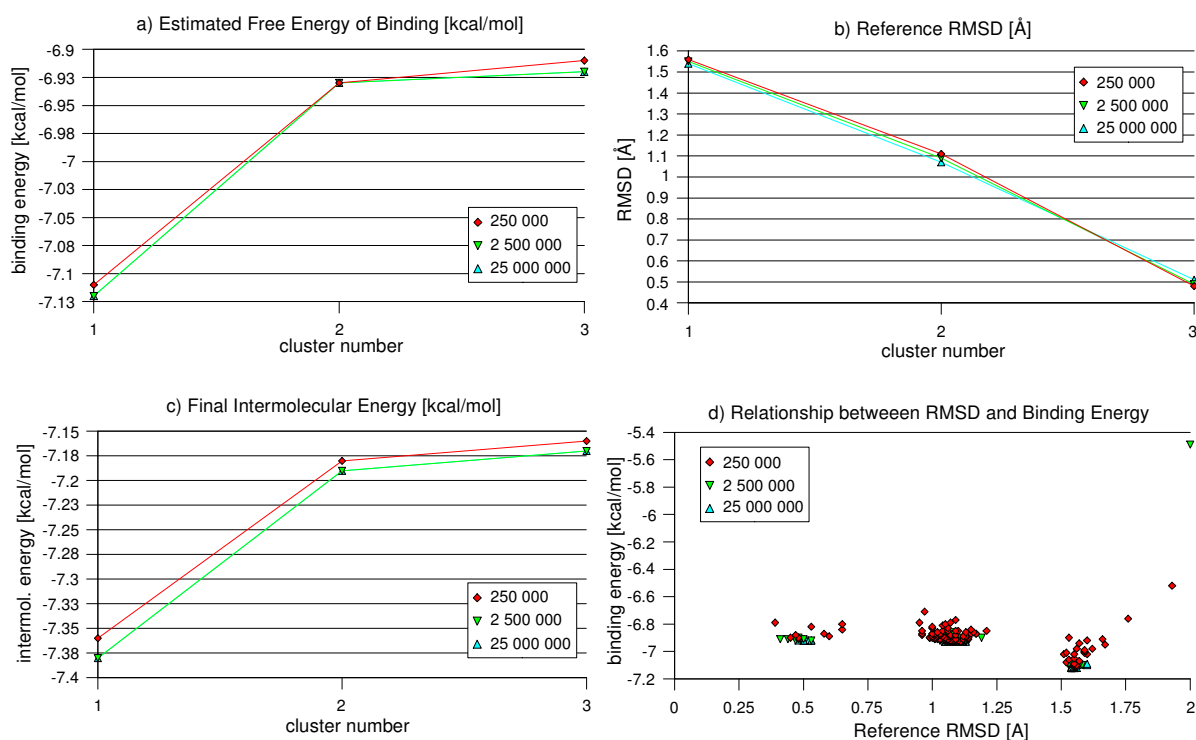
b) Reference RMSD [$\text{\AA}$]

cluster	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
1	1.56 [MS63 1_1]	1.55 [MM11 1_1]	1.54 [ML15 1_1]
2	1.11 [MS22 2_1]	1.09 [MM13 2_1]	1.07 [ML23 2_1]
3	0.48 [MS93 3_1]	0.49 [MM36 3_1]	0.51 [ML84 3_1]

c) Final Intermolecular Energy [kcal/mol]

cluster	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
1	-7.36 [MS63 1_1]	-7.38 [MM11 1_1]	-7.38 [ML15 1_1]
2	-7.18 [MS22 2_1]	-7.19 [MM13 2_1]	-7.19 [ML23 2_1]
3	-7.16 [MS93 3_1]	-7.17 [MM36 3_1]	-7.17 [ML84 3_1]

* Maximum number of evaluations. ** Unique structure ID and position in cluster.

**Figure 4-11 Results of rigid docking II.**

Displayed data is for number of runs = 100 only. Trends for first three cluster heads are shown on X axis and for varying number of evaluations by different colors.

4.6 Hardware requirements note

As mentioned before, the quality of the results depend on the maximum number of evaluations. However with the increasing number of evaluations, the requirements for calculation time become excessive. Tables 4-12 summarizes CPU time requirements for rigid docking and 4-13 for flexible docking.

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	1 min SS	9 min SM	81 min SL
100	9 min MS	87 min MM	18 hours ML
1000	91 min LS	18 hours LM	-- LL

Table 4-12 Calculation times for rigid docking.

Applies for Intel(R) Xeon(TM) CPU 3.20GHz with 60 x 50 x 50 Å grid-box, and 1 degree of freedom in a substrate.

runs	250 000 evaluations	2 500 000 evaluations	25 000 000 evaluations
10	7 min SS	1 hod SM	11 hours SL
100	57 min MS	10 hours MM	91 hours ML
1000	11 hours LS	91 hours LM	702 hours LL

Table 4-13 Calculation times for flexible docking.

Applies for Intel(R) Xeon(TM) CPU 3.20GHz with 60 x 50 x 50 Å grid-box, 7 flexible residues in macromolecule and 1 degree of freedom in a substrate.

One of the calculations was completed on two different systems with similar central processing units (CPU). The resulting times are quite surprising. The two systems are seemingly very similar, as both are running 32bit (i686) RHEL operating systems (OS) with 2.6.9 kernel. See Table 4-14. However on System A calculation took over 91 hours while on System B it only took 33 hours. The only apparent difference between the two is the size of the CPU cache.

Parameters	System A	System B
Model name	Intel(R) Xeon(TM) CPU 3.20GHz	Intel(R) Xeon(TM) CPU 3.00GHz
Cache size	1024 KB	4096 KB
Operating System	32bit (i686) RHEL 2.6.9 kernel	32bit (i686) RHEL 2.6.9 kernel

Table 4-14 Hardware and OS comparison.

5 Conclusions

All dockings were successful in placing substrate to the active-site with RMSD relative to reference crystallographic structure > 2 . However several structures with similar free energy of binding were usually found. The majority of conformations was almost always placed into three distinct clusters. See Figure 5-1. Predicted binding free energy differences among these three binding modes is very low (< 0.15 kcal/mol). This suggest that all these conformations can naturally occur with similar probability. The third cluster with binding energy -9.58 kcal/mol corresponds to the experimentally observed crystallographic structure of aNF-CYP1A2 complex, see Figure 5-4.

This study confirmed suitability of flexible docking algorithms implemented in AutoDock 4 for evaluation of interaction between CYPs and their ligands. Optimal input parameters, python automation scripts and necessary source code modifications were suggested and successfully evaluated.

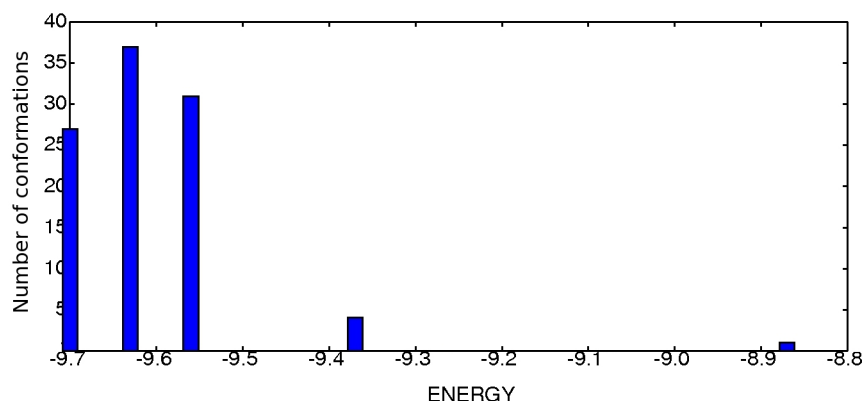


Figure 5-1 Sample of result clustering.

Flexible docking III, MM. Note that majority of results is grouped into first three clusters.

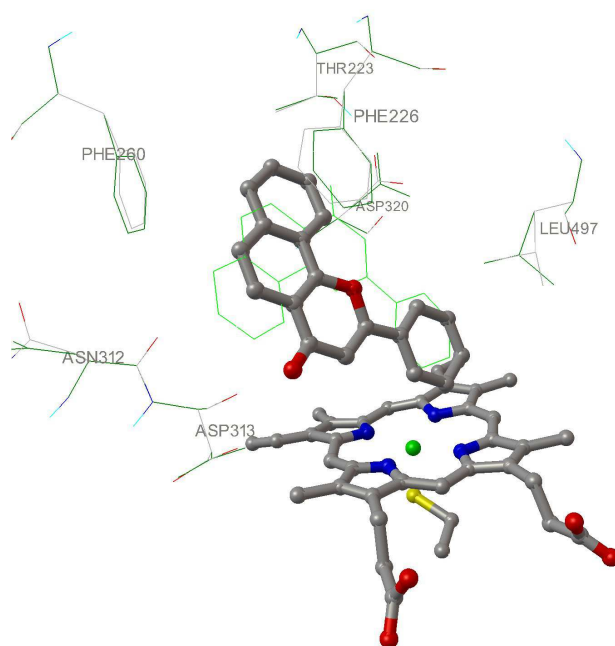


Figure 5-2 Docked ligand first node of best cluster.
Flexible docking III, MM 1_1. Reference structure from crystal is shown in green.

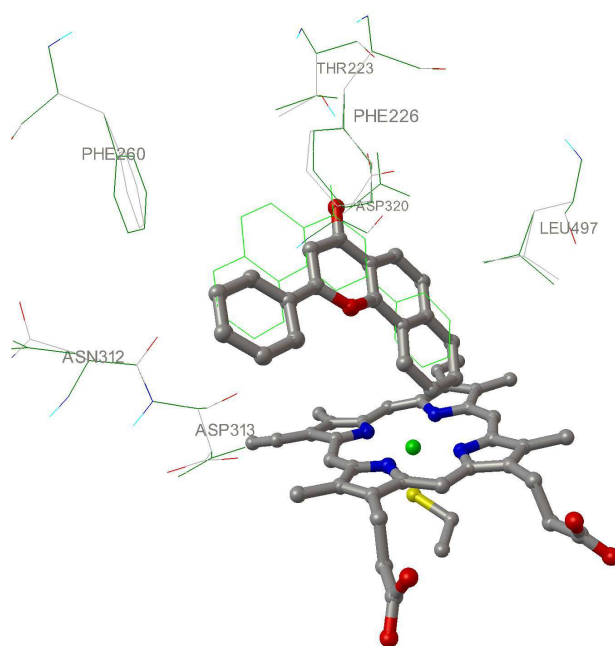


Figure 5-3 Docked ligand first node of second cluster.
Flexible docking III, MM 2_1. Reference structure from crystal is shown in green.

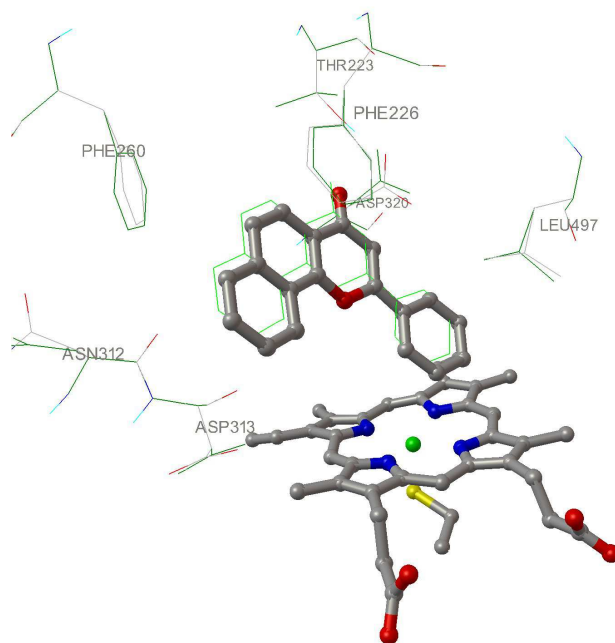


Figure 5-4 Docked ligand first node of third cluster.
Flexible docking III, MM 3_1. Reference structure from crystal is shown in green.

6 Bibliography

- [1] D. Werck-Reichhart, R. Feyereisen Cytochromes P450: a success story, *Genome Biology* **1**(6), 2000 reviews3003.1-3003.9, ISSN 1465-6906
- [2] M. Stiborová, J. Hudeček, P. Hodek, E. Frei Význam cytochromů P450 pro lidské zdraví, *Chem. Listy* **93**, 229-237, 1999
- [3] B. Meunier, S. P. Visser, S. Shaik Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes, *Chem. Rev.* 2004 DOI 10.1021/cr02443g
- [4] D. R. Nelson, Cytochrome P450 and the Individuality of Species, *Arch. Bioch. Bioph.* **369**, 1, 1-10, 1999 DOI abbi.1999.1352
- [5] G. D. Szklarz, J. R. Halpert Molecular Basis of P450 Inhibition and Activation, Implications for Drug Development and Drug Therapy, *Drug Met. Dispo.* **26**, 12, 1998, DOI 0090-9556/98/2612-1179-1184
- [6] D. C. Lamb, M. R. Waterman, S L Kelly, F. P. Guengerich, Cytochromes P450 and drug discovery, *Curr. Opin. Biotechnol.* **18**, 1-9, 2007, DOI 10.1016/j.copbio.2007.09.010
- [7] P. Hodek, M Karabec, M. Šulc, B. Sopko, S. Smrček, V. Martinek, J. Hudeček, M. Stiborová Mapping of cytochrome P450 2B4 substrate binding sites by photolabile probe 3-azidiamantane: Identification of putative substrate access regions, *Arch. Bioch. Bioph.* **468**, 82-91, 2007, DOI 10.1016/j.abb.2007.09.017
- [8] G. D. Szklarz, M. D. Paulsen Molecular Modeling of Cytochrome P450 1A1: Enzyme-Substrate Interactions and Substrate Binding Affinities, *J. Biomol. Struct. Dyn.* **20**, 2, 2002 ISSN 0739-1102
- [9] S. Sansen, J. K. Yano, R. L. Reynald, G. A. Schoch, K. J. Griffin, C. D. Stout, E. F. Johnson Adaptations for the Oxidation of Polycyclic Aromatic Hydrocarbons Exhibited by the Structure of Human P450 1A2, *J. Biol. Chem.* **282**, 19, 2007 DOI 10.1074/jbc.M611692200
- [10] F. P. Guengrerich Mechanisms of Cytochrome P450 Substrate Oxidation: MiniReview *J. Biochem. Mol. Toxicol.* **21**, 163-168, 2007, DOI 10:1002/jbt.20174
- [11] D. Nelson Cytochrome P450 Homepage, University of Tennessee
(Web resource <http://drnelson.utmem.edu/CytochromeP450.html> downloaded 22. 5. 2007)
- [12] F. P. Guengerich Characterization of human cytochrome P450 enzymes, *FASEB J.* **6**, 745-748, 1992, DOI 0892-6638/92/0006-745
- [13] P. R. Ortiz de Montellano Cytochrome P450 Structure, Mechanism, and Biochemistry, Third Edition, ISBN 0-306-48324-6

- [14] W. Tassaneeyakul, D. J. Birkett, M. E. McManus, M. E. Veronese, T. Andersson, R. H. Tukey, J. O. Miners Caffeine metabolism by human hepatic Cytochromes P450: Contributions of 1A2, 2E1 and 3A isoforms, *Bioch. Pharm.* **47**, 1767-1776, 1994
DOI 0006-2952(94)E0050-U
- [15] R. Delgoda A. C. G. Westlake Herbal Interactions Involving Cytochrome P450 Enzymes A Mini Review, *Toxicol. Rev.* **4**, 23, 2004 DOI 1176-2551/04/0004-0239
- [16] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson Automated Docking Using a Lamarckian Genetic Algorithm and An Empirical Binding Free Energy Function, *J. Comput. Chem.* **19**, 1639-1662, 1998, CCC 0192-8651 / 98 / 141639-24
- [17] U. Bren, V. Martínek, J. Florian Free energy simulations of uncatalyzed DNA replication fidelity: structure and stability of T.G and dTTP.G terminal DNA mismatches flanked by a single dangling nucleotide. *J. Phys Chem B.* **21**, 110, 2006 10557-66
- [18] I. Halperin, B. Ma, H. Wolfson, R. Nussinov Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions, *Prot. Struct. Funct. and Gen.* **47**, 409-443, 2002
- [19] D. Chen, G. Menche, T. D. Power, L. Sower, J. W. Peterson, C. H. Schein Accounting for Ligand-Bound Metal Ions in Docking Small Molecules on Adenylyl Cyclase Toxins, *Prot. Struct. Funct. and Bioinf.* **67**, 593-605, 2007, DOI 10.1002/prot.21249
- [20] A. R. Leach Molecular Modelling, Principles and Approaches, Second Edition, Pearson - Prentice Hall ISBN 0-582-38210-6
- [21] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case T. E. Cheatham Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models, *Acc. Chem. Res.* **33**, 889-897, 2000, DOI 10.1021/ar000033j
- [22] R. Huey, G. M. Morris, A. J. Olson, D. S. Goodsell Software News and Update, A Semiempirical Free Energy Force Field with Charge-Based Desolvation, *J. Comput. Chem.* **28**, 1045-1152, 2007, DOI 10.1002/jcc.20634
- [23] R. Huey, G. M. Morris, Using AutoDock 4 with ADT: A Tutorial, 22 May 2007 (Software Manual available as PDF, <http://autodock.scripps.edu/faqs-help/tutorial/> downloaded 30.11.2007)
- [24] G. M. Morris, D. S. Goodsell, R. Huey, S. Halliday, R. Belew, A. J. Olson User's Guide, AutoDock, Automated Docking of Flexible Ligands to Receptors, Version 3.0.5, 2001 (Software Manual available as PDF, <http://autodock.scripps.edu/faqs-help/manual/> downloaded 30.11.2007)
- [25] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. A. Cheeseman, J. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant and others. 2003. Gaussian 03. Gaussian, Inc. Pittsburgh PA.

- [26] D. J. Barrett Linux, kapesní přehled, Computer Press ISBN 80-251-0838-4
(Translated from: Daniel J. Barrett Linux Pocket Guide O'Reilly&Associates 2004)
- [27] H. Herold awk&sed, příručka pro dávkové zpracování textu, Computer Press
ISBN 80-251-0309-9 (Translated from: Helmut Herold awk&sed, Addison-Wesley DE 2004)
- [28] M. Stiborová, V. Martínek, H. Rýdlová, P. Hodek, H.H. Schmeiser, E. Frei Sudan I Is a Potential Carcinogen for Humans: Evidence for Its Metabolic Activation and Detoxication by Human Recombinant Cytochrome P450 1A1 and Liver Microsomes, *Cancer Res.* **2002**, 62, 5678-5684
- [29] M. Stiborova, V. Martinek, H. Rydlova, T. Koblas, P. Hodek Expression of cytochrome P450 1A1 and its contribution to oxidation of a potential human carcinogen 1-phenylazo-2-naphthol (Sudan I) in human livers, *Cancer Letters* **2005**, 220, 145-154

7 Appendix AutoDock 4 adaptations

This appendix contains AutoDock program source-code modifications.

Patch flex5/autodocksuite-4.0.1-fix5.patch

```
diff -Naur autodocksuite-4.0.1/src/autodock-4.0.1/calculateEnergies.cc autodocksuite-4.0.1/src/autodock-4.0.1/calculateEnergies.cc
--- autodocksuite-4.0.1/src/autodock-4.0.1/calculateEnergies.cc      2007-04-27 08:01:47.000000000 +0200
+++ autodocksuite-4.0.1/src/autodock-4.0.1/calculateEnergies.cc      2008-03-31 13:35:08.000000000 +0200
@@ -106,6 +106,9 @@
        info, B_outside?SOME_ATOMS_OUTSIDE_GRID:ALL_ATOMS_INSIDE_GRID,
        ignore_inter, elec, emap,
        NULL, NULL);
+    if(eb.e_intra_moving_fixed_rec < 0.0) {
+        eb.e_intra_moving_fixed_rec = 0.0;
+    }
+
    if (ntor > 0) {
@@ -158,7 +161,7 @@
    eb->e_intra_rec = 0.0;          // receptor intramolecular energy = (2) + (5)

    eb->e_torsFreeEnergy = torsFreeEnergy;          // empirical torsional free energy penalty
-    eb->e_unbound_internal_FE = unbound_internal_FE; // computed internal free energy of the unbound state
+    eb->e_unbound_internal_FE = 0.0; //unbound_internal_FE; // computed internal free energy of the unbound
state
    eb->deltaG = 0.0;          // estimated change in free energy upon binding
}
diff -Naur autodocksuite-4.0.1/src/autodock-4.0.1/constants.h autodocksuite-4.0.1/src/autodock-4.0.1/constants.h
--- autodocksuite-4.0.1/src/autodock-4.0.1/constants.h      2007-04-27 08:01:48.000000000 +0200
+++ autodocksuite-4.0.1/src/autodock-4.0.1/constants.h      2008-03-31 13:59:27.000000000 +0200
@@ -100,7 +100,7 @@
#define MAX_RECORDS 128 /* Maximum number of PDBQ records =~1.5*MAX_ATOMS */
#define MAX_NONBONDS 16384 /* 1048576 Maximum number of non-bonds in Small Molecule.*/
#else
-#define MAX_RUNS 256 /* Maximum number of runs. */
+#define MAX_RUNS 1024 /* Maximum number of runs. */
#define MAX_ATOMS 2048 /* Maximum number of atoms in Small Molecule. */
#define MAX_RECORDS 2048 /* Maximum number of PDBQ records =~1.5*MAX_ATOMS */
#define MAX_NONBONDS 524288 /* 1048576 Maximum number of non-bonds in Small Molecule.*/
diff -Naur autodocksuite-4.0.1/src/autodock-4.0.1/eintcal.cc autodocksuite-4.0.1/src/autodock-4.0.1/eintcal.cc
--- autodocksuite-4.0.1/src/autodock-4.0.1/eintcal.cc      2007-04-27 08:01:48.000000000 +0200
+++ autodocksuite-4.0.1/src/autodock-4.0.1/eintcal.cc      2008-03-31 13:33:08.000000000 +0200
@@ -302,6 +302,9 @@
    } else if (nb_group == INTRA_RECEPTOR) { // [2]
        // intramolecular energy of receptor
        nb_group_energy[INTRA_RECEPTOR] = total_e_internal - nb_group_energy[INTRA_LIGAND] -
nb_group_energy[INTER];
+    if (nb_group_energy[INTRA_RECEPTOR] < 0.0) {
+        nb_group_energy[INTRA_RECEPTOR] = 0.0;
+    }
+
    }

    } // nb_group -- intra lig, inter, intra rec
```

I modified source, and created the patch to keep record of a changes made:

```
$ diff -Naur autodocksuite-4.0.1-orig/ autodocksuite-4.0.1/ > autodocksuite-4.0.1-fix5.patch
```

Note: To apply patch unzip autodocksuite-4.0.1-src.tar.gz to /tmp, then copy autodocksuite-4.0.1-fix5.patch to /tmp and execute following commands:

```
$ cd /tmp
```

```
$ patch -p0 <autodocksuite-4.0.1-fix5.patch
```