

Faculty of Science at Charles University in Prague
Department of Biochemistry



**Computer modelling of interactions of nitroaromatic
compounds with nitroreductases**

**Modelování interakcí enzymů redukujících cizorodé látky
s aromatickými nitrosloučeninami**

Bachelor's thesis

Ondrej Gutten

Supervisor: RNDr. Václav Martínek, Ph.D.

Prague 2008

Prohlašuji, že jsem tuto bakalářskou práci vypracoval na katedře biochemie PŘF UK samostatně pod vedením školitele RNDr. Václava Martínka, Ph.D. a všechny použité prameny jsem řádně citoval.

V Praze dne _____

Podpis

Acknowledgments

Foremost, I'd like to thank my supervisor, RNDr. Václav Martínek, Ph.D., for all the precious time and advice he provided.

I'd like to thank Marián Šuch for technical support and computer resources he kindly provided.

I'd like to thank MetaCentrum for granting me access to computational resources.

“Words are a heavy thing...they weigh you down. If birds talked,
they couldn't fly”

Contents

LIST OF ABBREVIATIONS.....	4
1 INTRODUCTION.....	5
1.1 THEORETICAL AND COMPUTATIONAL CHEMISTRY.....	5
1.1.1 <i>Characterization and limitations</i>	5
1.1.2 <i>Docking simulations and Autodock package</i>	6
1.1.2.1 Search algorithms.....	6
1.1.2.2 Scoring functions.....	8
1.2 DT-DIAPHORASE.....	10
1.2.1 <i>Basic information</i>	10
1.2.2 <i>Substrates and induction</i>	11
1.2.3 <i>Function</i>	11
1.2.4 <i>Mechanism of action</i>	12
1.3 DERIVATIVES OF BENZANTHRONE(BZ)–2-NITROBENZANTHRONE (2-NBA) AND 3-NITROBENZANTHRONE (3-NBA).....	13
1.3.1 <i>Sources and occurrence</i>	13
1.3.2 <i>Mutagenicity and metabolism</i>	14
2 AIMS.....	17
3 METHODS.....	18
3.1 INPUT DATA.....	18
3.2 PDBQT FILES.....	19
3.2.1 <i>Rigid part of the protein</i>	19
3.2.2 <i>Ligand file</i>	20
3.2.3 <i>Flexible part of the protein</i>	21
3.3 DOCKING PARAMETER FILES.....	22
3.3.1 <i>GPF file (grid parameter file)</i>	22
3.3.2 <i>DPF file (docking parameter file)</i>	23
3.4 RUNNING AUTOGRID AND AUTODOCK.....	24
3.5 METHODOLOGY OF SELECTING FLEXIBLE RESIDUES.....	24
3.6 RESULT ANALYSIS.....	27
4 RESULTS AND DISCUSSION.....	28
4.1 CRYSTALLOGRAPHIC POSITION OF A LIGAND VS. SHORT DOCKING.....	28
4.2 OPTIMIZING THE LENGTH OF A SIMULATION.....	30
4.3 ANALYSIS OF X-RAY STRUCTURES.....	32
4.4 RIGID DOCKING OF 2-NBA AND 3-NBA INTO THE D,T-DIAPHORASE.....	34
4.5 FLEXIBLE DOCKING OF 2-NBA AND 3-NBA INTO DT-DIAPHORASE.....	37
5 CONCLUSIONS.....	42
6 LITERATURE.....	45

List of abbreviations

3-ABA	Aminobenzanthrone
BZ	Benzanthrone
FAD/FADH ₂	Flavine Adenine Dinucleotide – oxidized/reduced form
NAD(P)/NAD(P)H	Nicotinamide Adenine Dinucleotide (Phosphate) – oxidized/reduced form
2-NBA	2-nitrobenzanthrone
3-NBA	3-nitrobenzanthrone
NBA _s	Nitrobenzanthrone _s
PAH	Polyaromatic Hydrocarbons
PDB	Protein Data Bank format
RMSD	Root Mean Square Deviation
SPDBV	Swiss-PDB Viewer package

1 Introduction

1.1 Theoretical and computational chemistry

1.1.1 Characterization and limitations

The goal for chemical theory (or any theory in general) is not only to describe the underlying mechanisms behind the observed reality but also to predict the results of related experiments. It means that mere existence of algorithms is not sufficient for a theory to be applicable, because existence of computational power able to solve problems in finite time is also required. This proved and still proves to be a significant problem in theoretical chemistry.

Analytical solutions of quantum mechanical problems are principally impossible even for simplest chemical species. Introduction of numerical methods and simplified models transfers part of the problem into the aforementioned necessity for computational potency. These simplifications, even though far less accurate, cut down the demand on time drastically. However it remains the main limitation in the course of calculation.

Thus, selection of specific method is governed not only by its appropriateness for given problem but also by availability of computational power and time. This fact determines the nature of development of this scientific approach. *In silico* methods have also advanced significantly in the past few decades thanks to the vast progress in the field. Even though, contemporary theoretical chemistry may not appear satisfactory, especially in the field of biochemistry where modelling of large and complex systems is required. However, the fact that the possibilities are to great extent delimited by hardware, which remains subject to unceasingly swift development, rather than by employed theory grants a great perspective to the field and many promises to the future.

1.1.2 Docking simulations and Autodock package

One part of molecular modelling deals with predicting the preferred mutual orientation of two molecules. These so called docking simulations find applications mainly in enzyme-substrate modelling.

The goal of finding the best mutual orientation of two individual species requires a suitable algorithm able to find the orientations to be considered, and a fitness function for assessing their quality. These issues are discussed in the following chapters, with focus on specific implementations used in Autodock program suite, one of the most successful docking programs currently available¹.

1.1.2.1 Search algorithms

Exact solution of global optimization problems is often impossible, especially in multi-dimensional spaces where searching the entire variable space for global minimum of a given function is extremely time consuming or even impossible in real time. Since finding the global minimum cannot be guaranteed, we instead focus on finding the best local minimum.

There are many approaches to solving this problem. Autodock v4.0² offers four different methods: local search, Monte-Carlo simulated annealing, genetic algorithm, and Lamarckian genetic algorithm.

Local search method explores only a small portion of a variable space. The process is stopped after a local minimum is reached. This is useful e.g. for fine-tuning results obtained by different search results, but is not usually appropriate for global optimization problems.

Monte-Carlo simulated annealing mimics the process of annealing used e.g. in metallurgy.

New candidate is chosen “close” to the previous one $x \rightarrow x + \Delta x$

The new solution is accepted with a probability that increases with the

improvement in the value of the studied function $f(x+\Delta x) - f(x)$

and on gradually decreasing *temperature* parameter^{3,4}. As a result, new individuals do not need to improve, especially in the beginning of the simulation, and the search is not limited to a small portion of variable space bounded by local slopes on fitness function hyperplane.

In cases relevant to this study, genetic algorithms were shown to be more effective than simulated annealing⁵. These methods implement the ideas of evolutionary biology⁶.

Genotype of a ligand is a string that represents its orientation, Cartesian coordinates and torsional angles. Its phenotype is a list of atoms' coordinates obtained by mapping of the genome string.

Genetic algorithm used in Autodock can be divided into four phases:

1. Mapping and fitness evaluation – Genome of each individual is translated into phenotype, which is then used by the scoring function to evaluate its fitness.
2. Selection – Individuals are ranked by their fitness and the top portion of the population is selected to produce offsprings
3. Crossover and mutation – crossover between genes of two random individuals is performed and can be followed, with certain probability, by mutation. This results in new genotype bearing characteristics of parents'.
4. Elitism – Produced offsprings are mapped and evaluated. Since population might have increased, depending on the number of offsprings produced in the third step, only top individuals are left to survive.

The algorithm is stopped after given number of generations or fitness evaluations.

The last method available is a Lamarckian genetic algorithm. Its name comes from the analogy to Jean Batiste de Lamarck's evolutionary theory, which states that characteristics acquired during individual's lifetime can become heritable traits.

The implementation of this idea could be described by two additional steps:

1. Local Search – New individual undergoes a local search, reaching a local minimum. This reflects the specific adaptation to environment.
2. Phenotype to genotype – Locally optimized fitness function is reflected back to genome string, which is then used to produce new offsprings. This represents the

heritability of characteristics adapted by individual.

Lamarckian genetic algorithm is therefore a hybrid search method providing both global search (provided by mutation and crossover) and local search and is often marked as GA-LS (genetic algorithm-local search).

Comparison of Monte-Carlo simulated annealing, genetic algorithm and Lamarckian genetic algorithm showed the latter to be the most effective for docking⁵.

1.1.2.2 Scoring functions⁷

The fitness, or scoring function, is problem-specific. In docking simulation it is designed to model the forces that govern the binding of a ligand with a protein.

Autodock estimates the free energy of binding of a ligand into a protein. This is done in two steps. Firstly, intramolecular energetics of transition from unbound to bound state is evaluated. Second step involves intermolecular energetics of transition from unbound to bound state and an estimate on loss of conformational entropy upon binding. Intermolecular energy of unbound system is considered zero. The entire evaluation is depicted in Fig. 1.

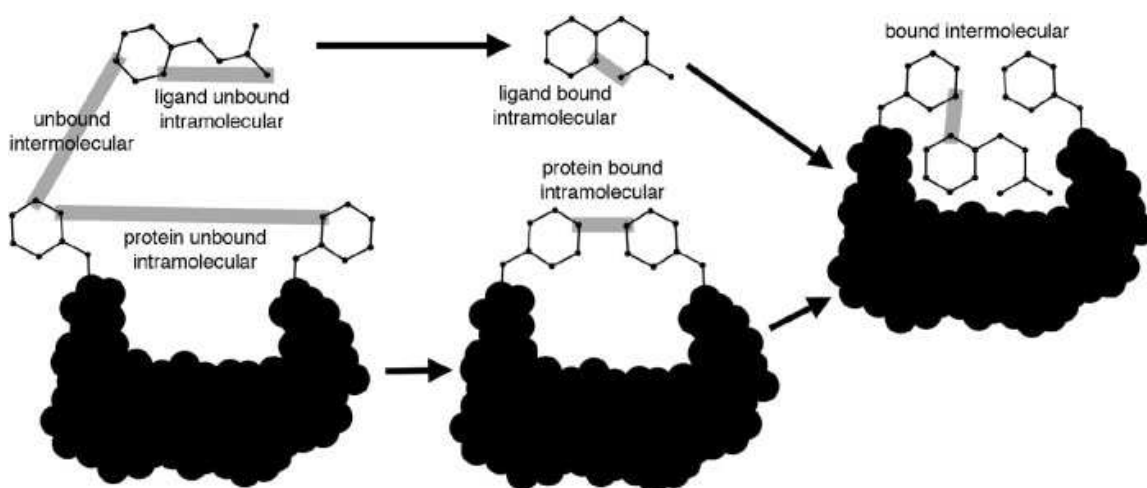


Figure 1:Evaluation of free energy of binding⁷. Each interaction term is represented by gray line and explained in Formula 1.

$$\Delta G = (V_{\text{bound}}^{\text{L-L}} - V_{\text{unbound}}^{\text{L-L}}) + (V_{\text{bound}}^{\text{P-P}} - V_{\text{unbound}}^{\text{P-P}}) + (V_{\text{bound}}^{\text{P-L}} - V_{\text{unbound}}^{\text{P-L}} + \Delta S_{\text{conf}})$$

Formula 1: Evaluation of free energy of binding. V-potential energy, L-ligand, P-protein.

In the output file, free energy of binding is decomposed into four contributions: intermolecular energy, internal energy, torsional free energy and unbound system's energy.

Each term in formula &1 is calculated using force field that calculates pair-wise interactions between atoms. All atoms except non-polar hydrogens (those bonded to carbon atoms) are included. Intramolecular energetics do not include 1-2, 1-3 and 1-4 interactions.

Four types of interactions are considered:

$$V = W_{vdw} \sum_{ij} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{\text{hbound}} \sum_{ij} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{\text{elec}} \sum_{ij} \frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} + W_{\text{sol}} \sum_{ij} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)}$$

Formula 2: Autodock potential energy function.

vdw – Van der Waals' interaction. A and B parameters were taken from Amber force field and simplified for atom types used.

hbound – Hydrogen bonding. E(t) function reflects the geometry of a hydrogen bond. Parameter C and D were assigned values that give the bond maximum strength at values obtained from experiments.

elec – Coulomb's law is used to calculate the electrostatic interactions.

sol – Solvation energy. Amount of solvation is evaluated from the energy of transporting an atom from fully hydrated to fully buried state (Si) and an estimate on amount of desolvation when ligand is docked. (Vi).

W – these are scaling parameters, calibrated to fit experimental data.

1.2 DT-Diaphorase

1.2.1 Basic information

DT-Diaphorase (NAD(P)H quinone acceptor oxidoreductase; NQO1; menadione oxidoreductase; EC 1.6.99.2) is a cytoplasmatic flavoprotein with minor activity in mitochondria and cytoplasmatic reticulum⁸. Its name abbreviates obsolete names of its cofactors DPNH (reduced nicotinamide adenine dinucleotide-NADH) and TPNH (reduced triphosphopyridine nucleotide also called nicotinamide adenine dinucleotide phosphate - NADPH). It has been known for a little over 50 years⁸.

Structure and sequence of human, mouse, and rat DT-Diaphorase has been identified and found to be highly homological, despite varying catalytic efficiency⁹. DT-diaphorase is a homodimeric enzyme with head-to-tail arrangement of chains. The structural chain consists of 274 aminoacid residues¹⁰⁻¹³. Dimeric enzyme non-covalently binds two flavine adenine dinucleotides (FAD), which act as mediators of hydrogen transfer and form part of the active sites. The active site is situated on the contact of the two subunits.

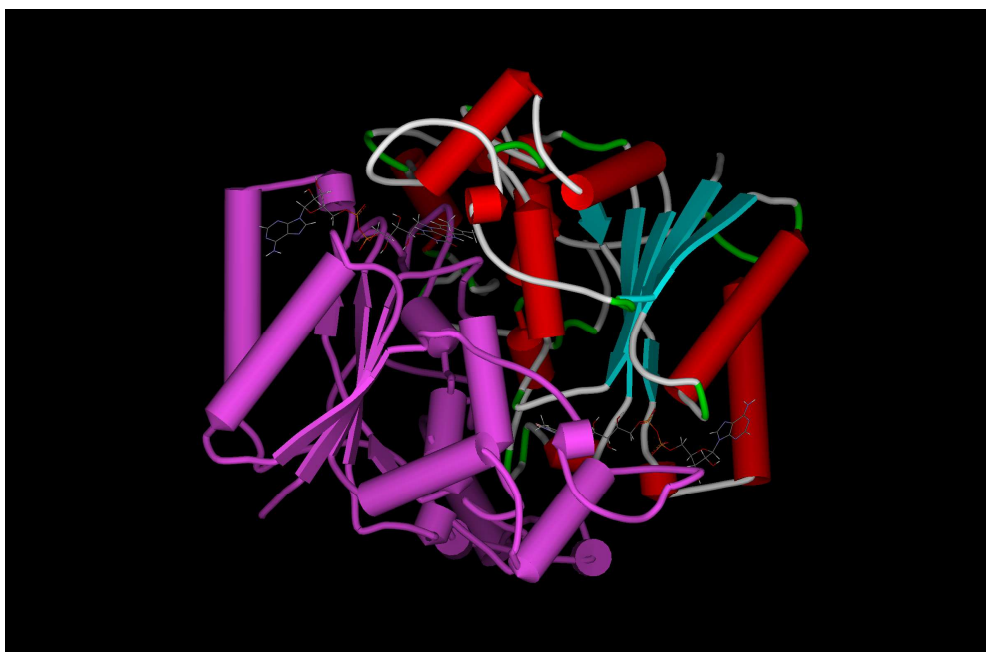


Figure 2: Human DT-diaphorase¹⁰: Ribbon representation – One entire chain is purple. The other is coloured according to secondary structure: alpha helices are red, beta sheets are light blue, turns are green. Coenzymes are drawn in sticks representation.

Two naturally occurring modifications of human DT-diaphorase have been discovered¹⁴. First of these mutations causes exchange of proline at position 187 for serine. This Pro187Ser mutant has been expressed and its catalytic properties have been studied. Its activity towards studied compounds (menadione, dichlorophenolindophenol, prodrug CB1954) is significantly lower when compared to wild type enzyme⁹.

1.2.2 Substrates and induction

DT-diaphorase catalyzes reduction of quinones and quinonoid compounds to their hydroquinonoid analogues¹⁵.

It is a highly non-specific enzyme. Its substrates include e.g. coenzyme Q¹⁶, compound involved in electron transport chain, menadione¹⁷, a vitamin K₂ precursor participating in process of blood coagulation, and others. The enzyme is also capable of reducing other substances, such as some cytostatics (e.g. certain mitomycins, anthracyclines, and aziridiny benzoquinones⁹) and xenobiotics (namely nitroaromatic compounds and azodyes)^{18,19}.

Expression of this enzyme is induced by numerous xenobiotics, including phenolic antioxidants, lipophilic azodyes, polycyclic aromatic hydrocarbons, coumarine, flavonoids²⁰.

1.2.3 Function

Among other functions, resulting from its ability to reduce some of the molecules native to the organism, DT-diaphorase is believed to play role in natural prevention of cancerogenesis³. This belief is supported by a study, which shows significantly higher percentage of specific NQO1 gene mutants among studied sample of kidney tumour patients when compared to normal population⁵². This Pro187Ser mutant enzyme was shown to have only 10% of wild-type enzyme's activity towards studied substrates⁹.

DT-diaphorase has been shown to have generally higher activity in tumour cells²¹. Combined with its ability to activate certain cytostatic compounds by catalyzing their reduction, this enzyme is a potential target for drug design aimed against tumours.

1.2.4 Mechanism of action

The enzyme uses ping-pong bi-bi mechanism for reducing substrates²².

It's capable of using both NADH and NADPH as electron donors with almost identical effectivity⁸. In the first step, NAD(P)H is oxidized by transfer of two hydrogens to FAD coenzyme. The first product, NAD(P) is released and is replaced by a second substrate. The latter is then reduced by accepting two hydrogens from FAD, which regains its oxidized form. Reduced second substrate is released, regenerating enzyme²³.

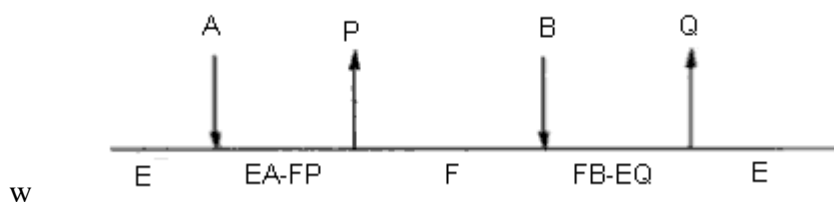


Figure 3: Schematic illustration of bi-bi ping-pong reaction mechanism:

E-enzyme; F-modified enzyme; A-first substrate; P-first product; B-second substrate; Q-second product²³.

Studies show that His161 residue might play important role in charge relay mechanism^{22,9}.

Multiple studies show that during reduction of quinones, no accumulation of semiquinone intermediate occur^{22,24,25}. Combined with other results, suggested mechanism for the enzyme operates with simultaneous transfer of both electrons and is therefore believed to play role in natural prevention of carcinogenesis.

Presence of semiquinones increases oxydative stress, due to spontaneous reaction with molecular oxygen, which yields regenerated quinone and superoxide radical species. In addition, semiquinones themselves are unstable radical species and are also potentially mutagenic.

1.3 Derivatives of benzanthrone(BZ)–2-nitrobenzanthrone (2-NBA) and 3-nitrobenzanthrone (3-NBA)

1.3.1 Sources and occurrence

Many nitro derivatives of polyaromatic hydrocarbons (PAH) are known mutagens²⁶ and many are suspected human carcinogens. One of the strongest mutagens known is 3-NBA²⁷. Parental BZ was not found mutagenic at tested doses²⁸. Several other NBAs were studied and also found to possess mutagenic properties but to a lesser extent²⁹.

Exposure to these NBAs is common, especially to population in urban areas. They are present in atmosphere and adsorb to airborne particles²⁷. Concentrations of 3-NBA are highest at the source of its formation, due to suggested isomerisation of 3-NBA to 2-NBA that occurs at atmospheric conditions^{30,31}. Concentration of 2-NBA in samples of atmosphere was found to be up to 70 times higher than that of 3-NBA³¹. Even though 2-NBA was not shown to be nearly as mutagenic as 3-NBA, it's now studied for its higher abundance

A case study among underground salt mine workers, occupationally exposed to diesel exhausts, discovered 3-ABA, a metabolite of 3-NBA, in their urine³².

Studies showed presence of 3-NBA in rainwater and surface soil, indicating its further transport from atmosphere^{33,34}.

3-NBA is produced during combustion of fossil fuels²⁷. 3-aminobenzanthrone (3-ABA), metabolite of 3-NBA, is used in colouration of microporous films, as a textile disperse dye and fluorescent pigment^{35,36,37}.

1.3.2 Mutagenicity and metabolism

Mutagenicity was proved in multitude of studied NBAs with measure strongly depending on the position of nitro-group²⁹. Many of them are suspected human carcinogens. 3-NBA was capable of forming DNA adducts in more kinds of human cell lines than 2-NBA³⁸. This indicates differences in bioactivation of each compound.

Nevertheless, correlation between exposure and formation of tumours was shown, especially in lungs and bladder and kidney, which suffer from highest exposure to NBA and its metabolites³⁹.

In addition, some NBAs are suspected of promoting formation of reactive oxygen species, which may induce the formation of a tumour⁴⁰.

Mutagenicity of nitroaromates is known to be linked to their metabolism. Nitro group is reduced in several steps to amino group. One of the intermediates, the hydroxylamine group, is unstable and forms nitrenium or carbenium ions. These are reactive electrophiles which can easily attack DNA bases and create covalent adducts⁴¹.

Hydroxylamines can be acetylated or sulphonated by the corresponding acetyltransferases and sulphotransferases. Formed species is also unstable and forms aforementioned electrophiles.

Density functional calculations suggest that nitrenium ion formed from 3-NBA (see Fig. 4) is more stable than nitrenium ion formed from 2-NBA, possibly providing it with more time to reach nucleus, while 2-NBA-derived ion is more likely to react with the molecules in immediate vicinity of ion's creation⁴².

Proposed metabolism for 3-NBA is shown in Fig. 4. DT-diaphorase was identified as main enzyme responsible for reduction of 3-NBA⁴³. On the other hand, reduction of 2-NBA by DT-diaphorase was not detected⁴⁴. Metabolism of 2-NBA has not been sufficiently explored, however, proposed mechanisms are analogical (see Fig. 5)⁴⁵.

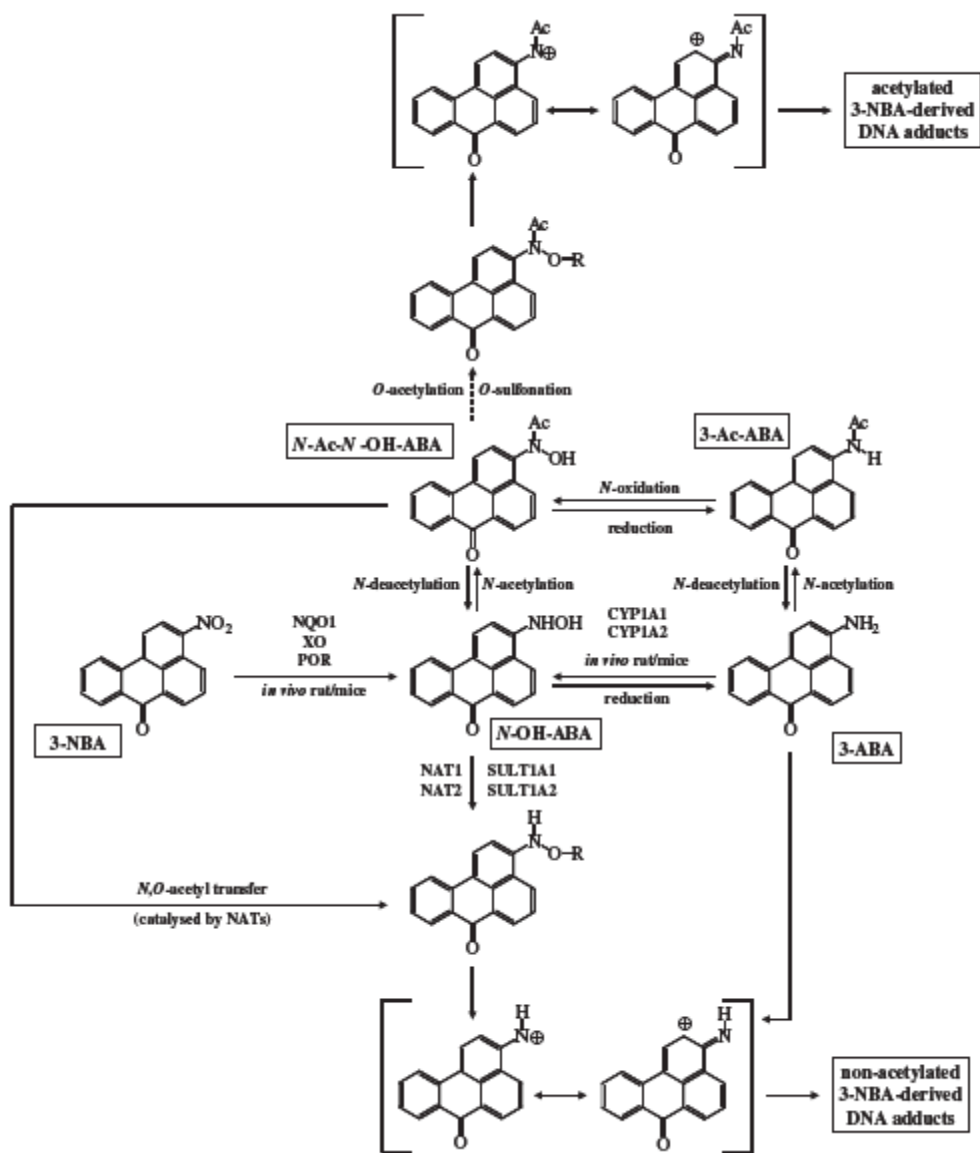


Figure 4: Proposed metabolism of 3-NBA⁴¹ (edited): NQO1 – DT-diaphorase, XO – xanthine peroxidase, POR – P450 oxidoreductase, CYP – P450 cytochromes, NAT – N,O-acetyltransferases; SULT – sulphotransferases; R = COCH₃ or SO₃H.

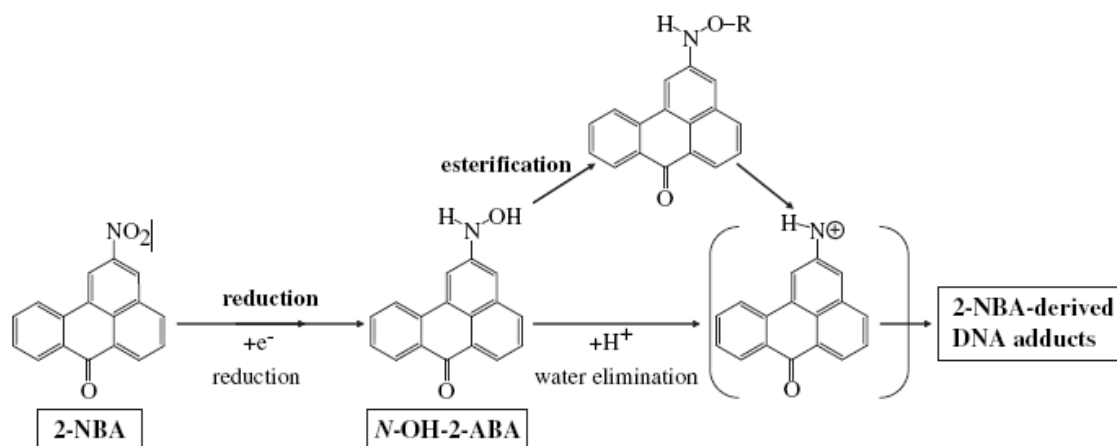


Figure 5: Proposed metabolic activation of 2-NBA⁴²(edited): R = COCH₃ or SO₃H.

As illustrated in Fig. 4, metabolic changes of 3-NBA are catalyzed by number of enzymes. First metabolic step to activation is reduction, which is in large part carried out by DT-diaphorase. Understanding the differences in affinity of this enzyme towards the studied NBAs could help to explain differences in their metabolism and contribute to evaluation of overall risk they pose to humankind.

2 Aims

Primary aim was to evaluate the potential of flexible (or soft-soft) docking algorithms recently implemented in the Autodock 4.0 program package, and compare it to already established rigid (hard-soft) docking approach, that is implemented in the same program suite. Specifically, we want to assess the reproducibility and reliability of these two docking approaches for model system consisting of 2-NBA and 3-NBA, and DT-diaphorase with varying size and shape of the originally co-crystallized ligand.

Another aim is to contribute to our understanding of interactions of this biologically important reductase with environmental pollutants and potent carcinogens 3-NBA and 2-NBA.

3 Methods

3.1 Input Data

Coordinates of proteins were downloaded from protein databank⁴⁶. Since multiple structures of DT-diaphorase are available, those to be used for docking had to be selected.

Choice of PDB file was affected by various requirements. Human DT-diaphorase was of interest, therefore only human isoforms of the enzyme were selected.

Two other factors were included in the selection:

1) Good resolution implies better representation of real protein, thus increasing the precision of the simulation. For this reason structures with higher resolution were preferred.

2) Ligand bound by protein in the crystal structure. Even in flexible docking the part of the protein allowed to move is relatively small. In addition, since only sidechains of aminoacid residues can move, but movements of backbone resulting from binding of substrate cannot be accounted for. Therefore, it is important to choose X-ray structures containing ligands most similar to the ones to be docked.

To examine the influence of original substrate on results of docking, structures with various substrates and resolutions were selected: 1D4A, 1H66, 1H69, 1KBQ, 2F1O

All of the used structures were overlaid by Magic Fit function implemented in Swiss-PDB Viewer (SPDBV) program. This is the least square method minimizing root mean square deviation (RMSD) of corresponding atoms present in two PDB files. Due to generally large flexibility of sidechains, only backbone atoms were used for fitting.

Geometries of 2-NBA and 3-NBA and partial charges of 3-NBA were obtained via ab-initio methodology as implemented in GAUSSIAN 03 program suite⁴⁸ using the Hartree-Fock level of theory in conjunction with the 6-31+G(d) basis set. Output was converted to pdbqt (see later) format using MGL-Tools program⁴⁷.

Since differences in interaction between protein and the ligand due to the various positions of nitro group on BZ were to be examined, these calculations were carried out for 3-NBA only. The other ligand was optimized using the same parameters as those used for 3-NBA and converted to pdbqt format using charges consistent with those calculated for 3-NBA. Thus, electrostatic interaction between protein and ligand should be roughly the same for both ligands used.

MGL-Tools - v1.4.6 for Linux and 1.5.0 for Windows operating system were used for preparing most of the input for simulations. Protein and ligand pdbqt files and configuration files for autogrid (gpf) and autodock (dpf) are required for docking performed by Autodock.

3.2 PDBQT files

Pdbqt file is essentially a PDB file with charges, atom types and selected torsions.

3.2.1 Rigid part of the protein

MGL-Tools package offers computation of two types of charges – Gasteiger and Kollman. Each has different requirements on starting structure.

Algorithm for Gasteiger charges requires all hydrogens to be present, and since x-ray diffraction method is unable to detect hydrogens, the PDB files contain no hydrogens and they should be added (e.g. using MGL-Tools) and incorrectly placed hydrogens should be corrected manually.

Adding hydrogens is performed by

➤ **Edit → Hydrogens → Add**

and then in the dialog window

- **All hydrogens**
- **Trust bond order information – no;**

Gasteiger charges are then computed by

➤ **Edit → Charges → Compute Gasteiger**

Since Autodock works with polar hydrogens only, non-polar hydrogens need to be

removed by merging their charge into heavy atoms to which they are bound. This can be done either automatically when selecting macromolecule for Autogrid (see chapter 3.3.1) or immediately by

- **Edit → Hydrogens → Merge non-polar**

Alternatively, Kollman charges can be used. They require only polar hydrogens to be present.

- **Edit → Hydrogens → Add**

and then in the dialog window

- **Polar only • trust bond order information – no**

It is worth noting that Kollman charges for aminoacid residues are read from a library. If some part of the protein is not recognized or present in the library (specifically coenzyme in this case), it is assigned zero charges. Therefore, correction is necessary, for example by calculating Gasteiger charges.

Autodock atom types are added automatically when selecting macromolecule for Autogrid (see chapter 3.3.1) or by:

- **Edit → Atoms → Assign AD4 types**

File needs to be saved in PDBQT format:

- **File → Save → Write PDBQT**

3.2.2 Ligand file

Ligand used for optimization of docking parameters was extracted from PDB file. Again, hydrogens had to be added.

- **File → Read Molecule → *select PDB file***

Again, only polar hydrogens are required

- **Edit → Add-Hydrogens; polar only, trust bond order info – no**

Program is then told to recognize current structure as ligand by selecting it in the dialog window

- **Ligand → Input → Choose**

If the original file contains necessary hydrogens it can be loaded directly via Ligand tab menu

➤ **Ligand → Input → Open**

Now, torsions need to be set. All bonds excepts bonds in cycles and bonds between atoms of which at least one is bonded only to non-polar hydrogens can be rotated. This includes even guanidium and amide bonds.

Whether a bond should be allowed to rotate is set in

➤ **Ligand → Torsion Tree → Choose Torsions**

However, since ligands modelled in this study were relatively small, all of their rotatable bonds were allowed to rotate.

File is saved by:

➤ **Ligand → Output → Save PDBQT**

3.2.3 Flexible part of the protein

Autodock4 offers feature unavailable in previous version - option to rotate some of the bonds in the sidechains of protein with same limitations applied as for ligand. The choice is optional.

Protein has to be selected:

➤ **Flexible Residues → Input → Choose/Open Macromolecule**

Then, sidechains to be rotated are chosen by selecting them, e.g.:

➤ **Select → Select from string**

followed by

➤ **Flexible Residues → Choose torsions in selected residues**

This hides the rest of the molecule, leaving only selected sidechains visible. Default setting shows nonrotatable bonds in red colour, rotatable bonds in green and disabled rotatable bonds in purple.

Bonds to be rotated are un/selected.

Selection is saved by writing rigid and flexible parts of the protein as two separate files with Pdbqt extension.

- **Flexible Residues – Output – Save PDBQT**

3.3 Docking parameter files

3.3.1 GPF file (grid parameter file)

It is good to remove any objects loaded in the application to avoid confusion, especially if we are preparing flexible docking.

Macromolecule is initialized by

- **Grid → Macromolecule → Open**

This loads the molecule. Besides, if the selected file has charges already assigned it asks whether they should be kept or recalculated (if so, Gasteiger charges are computed). If no charges are set this is done automatically. All hydrogens need to be present, otherwise the computed charges are incorrect!

Non-polar hydrogens are then automatically merged and their charges summed up to heavy atoms bonded to them; Autodock atom types are assigned and, if the structure is not yet saved in Pdbqt format, user is asked to do so.

Now, map types need to be set. Electrostatic and desolvation maps are required for every docking and are calculated automatically. In addition, one map for each atom type in flexible domain is required. There are two ways they can be set – either by directly listing them

- **Grid – Set Map Types – Directly**

or by selecting ligand and flexible residues, either as objects in the interface or as files:

- **Grid – Set Map Types – Open/Choose Ligand/FlexRes**

The space in which the coordinates of flexible domain can be varied is defined by box.

Its center and dimensions are set in

➤ **Grid – Gridbox**

It is essential that all flexible parts of the protein are within the box. Otherwise, no error is announced, but the calculations are incorrect and value of energies rise to large positive values.

Some other parameters can be saved, however, these were kept default. Settings are saved via:

➤ **Grid – Output – Save GPF**

3.3.2 DPF file (docking parameter file)

Macromolecule and ligand files are selected through corresponding menu tabs

- **Docking – Macromolecule – Set Rigid Filename**
- **(Docking – Macromolecule – Set Flexible Residues Filename)**
- **Docking – Ligand – Open/Choose Ligand**

Some parameters for ligand can be set in

➤ **Docking – Ligand – Ligand Parameters.**

Default settings were kept for all but one simulation, where predefined starting position of ligand was desired.

Next, settings defining the type of the simulation and its basic characteristics are set.

Genetic algorithm was used exclusively in this work.

➤ **Docking – Search Parameters – Genetic algorithm**

A widget allowing setting of parameters of the algorithm appears. Default settings were used, except for **Number of GA runs** and **Maximum number of evals**. These two parameters were varied in range 10-1000 and 250,000-25,000,000 respectively.

It is important to note that value, that defines the maximum number of runs in Autodock is 256. In order to perform simulations with higher number of runs, source code has to be modified and recompiled.

The constant can be found in file **constants.h** at line **103**.

Other settings that influence the course of docking can be changed in

➤ **Docking** → **Docking Parameters**

However, none of these were changed.

3.4 Running Autogrid and Autodock

Both Autogrid and Autodock can be run through **Run** menu tab or directly from terminal window by commands:

```
autogrid4 -p path_to_gpf_file -l path_to_output_file
```

```
autodock4 -p path_to_dpf_file -l path_to_output_file
```

While MGL-Tools can be run under Microsoft Windows operating system, Autodock itself cannot, requiring a different platform (e.g.: Linux).

3.5 Methodology of selecting flexible residues

The key to effective simulation lies in optimizing its parameters to provide satisfactory results with as little computational power and time required as possible. This task includes not only minimizing time required for the docking to converge to local energy minimum, but also minimizing its dimensions and number of variables to the lowest values. The latter factors include number of rotatable bonds and the dimensions of the box where these bonds and the ligand itself are allowed to move.

Active site of DT-diaphorase has been already localized by previous studies¹⁰⁻¹³. The enzyme dimer (chain A and chain C were used for all PDB structures) contains two virtually identical active sites. Only one of them was used for docking (active site closer to Tyr128C residue).

SPDBV program was used to select residues, that will be treated as flexible. Previously selected and overlaid structures (1D4A,1H66,1H69,1KBQ,2F1O) were closely examined.

Choice of flexible residues was determined by relative variability of their position and orientation in examined PDB structures and their proximity to active site. Sidechains showing high variability in their orientation among different structures are the ones most easily influenced by factors that differ in these structures – most notably the ligand present. Moreover, if they are close to space that is likely to be occupied by ligand, it seems incorrect to constrain them by keeping them static.

An important fact to take into consideration is that Autodock works with polar hydrogens only, and that it cannot rotate double bonds. Also, rotating methyl groups is pointless, as the hydrogens bound to the terminal carbon are not explicitly present during the calculation. This excludes residues such as glycine, alanine and proline as candidates for flexibility and restricts number of potentially rotatable bonds in most of other residues. On the other hand, no hydrogens are present in the original PDB files. **Since the position of polar hydrogens is generated artificially (and differently each time!) it is reasonable to allow their rotation.**

Taking all this facts into consideration, 14 rotatable bonds in 8 aminoacidic residues (Phe106A, Met154A, Tyr155A, His194A, Tyr126C, Tyr128C, Phe178C ,Phe232C). To minimize time requirements for docking not all of the rotatable bonds in the listed residues were left to rotate but only those with highest relative variability (see Fig. 6 and 8).

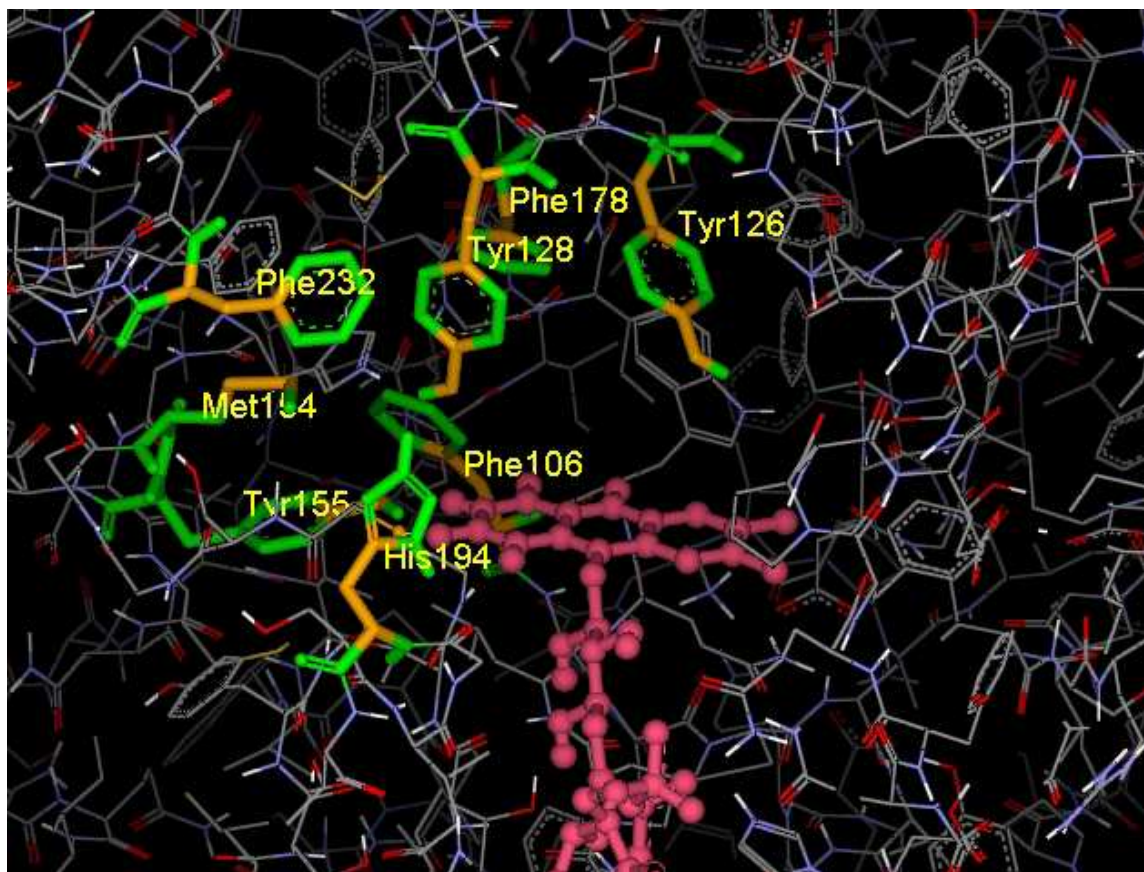


Figure 6.: Residues and bonds used in flexible docking. Only polar hydrogens are displayed. Residues used in flexible domain are represented in bold. Orange – flexible bonds; Green – rigid bonds. Pink – FADH₂.

Dimensions of the box, that defines the space where docking takes place, were determined from maximum and minimum values of coordinates of nonflexible atoms that represent boundaries of active site. Since some of the flexible sidechains could move behind these points, the edges of the box were extended slightly behind their potential reach.

Since active site has different boundaries in rigid and flexible docking, also the box has different dimensions.

Grid used in flexible docking had center at (18.550; -12.085; 7.576) and had dimensions of (50; 72; 48) with 0.375 spacing.

Grid used in rigid docking had center at (20.000; -10.628; 6.250) and had dimensions of (34; 58; 40) with 0.375 spacing.

3.6 Result analysis

Output of Autodock's dockings are written into a DLG file, which contains, among other, information about energy and coordinates of discovered orientations. These results were further processed using MGL-Tools. Pictures were created using programs VMD⁴⁹, SPDBV⁵⁰ and Accelrys DS Viewerlite⁵¹.

4 Results and discussion

4.1 Crystallographic position of a ligand vs. short docking

First docking was carried out to show, whether force-field implemented in Autodock is capable of recognizing experimentally confirmed ligand orientation as an optimal orientation for the modelled system.

For this purpose, 1H66 structure without original ligand was used as a macromolecule.

Its original ligand was isolated to separate PDB file and processed to pdbqt format keeping the maximum possible number of rotatable bonds (= four). Both rigid and flexible dockings were carried out with 10 runs and 2,500,000 energy evaluations for each. The initial position of the ligand was set to match the position in original X-ray structure. If Autodock's scoring function would recognize this orientation as a local minimum, ligand should not be displaced significantly from the original coordinates. The ligand displacement measured as RMSD relative to the crystallographic orientation was used as a measure of difference between actual minimum and the minimum recognized by the scoring function.

All docking simulations successfully placed ligand into position resembling its original crystallographic coordinates (see Fig. 7). The only significant difference was systematical translation of the ligand. Nevertheless, if we consider the simplicity of Autodock's force field, RMSD values obtained from both rigid and flexible dockings are very low. Such a small deviation could be caused either by imprecisions in X-ray ligand density fit or by inaccuracy of applied force field.

All of the final orientations were grouped in one cluster (RMSD tolerance = 1.0). Moreover, values of RMSD referenced to the position of ligand found in crystal were 0.23-0.61 for flexible docking, and 0.50-0.83 for rigid docking, see Table 1.

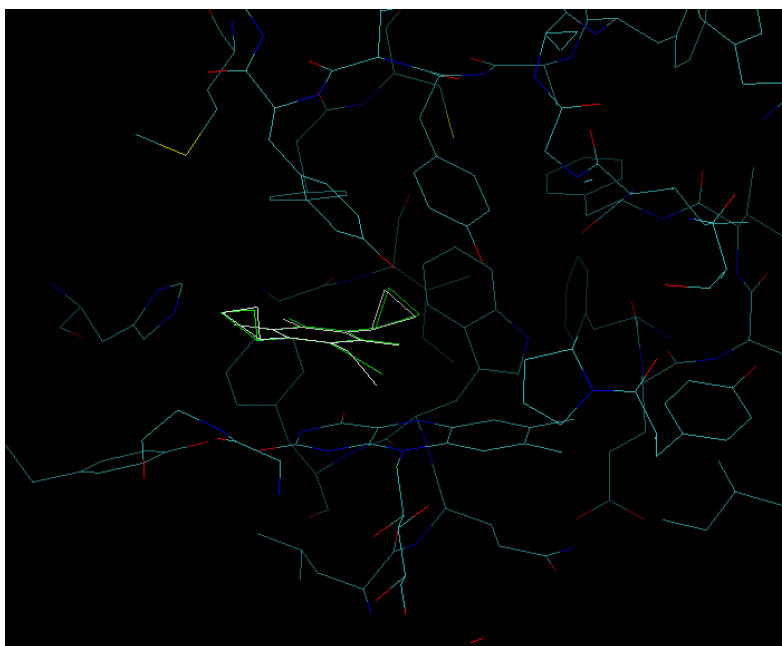


Figure 7: Comparison of crystal and docking orientations. Crystal – white, docking with the lowest free energy of binding (no.7 from flexible docking) - green

Table 1: RMSD values of dockings referenced to crystal coordinates, ranked by free energy of binding

run	flexible		rigid	
	Free E of binding [kcal/mol]	RMSD	Free E of binding [kcal/mol]	RMSD
1	-14.53	0.56	-6.16	0.54
2	-14.72	0.23	-6.17	0.55
3	-14.84	0.3	-6.12	0.55
4	-14.75	0.52	-6.16	0.67
5	-14.68	0.39	-6.16	0.62
6	-14.79	0.61	-6.16	0.82
7	-15.07	0.28	-6.16	0.66
8	-15.01	0.33	-6.16	0.55
9	-14.67	0.50	-6.16	0.50
10	-14.40	0.27	-6.16	0.83

4.2 Optimizing the length of a simulation

In computational chemistry, the time of simulation is just as important as the accuracy of its results, we need to optimize this ratio, which is likely to be similar for resembling macromolecules and closely related ligands, with same degree of torsional freedom.

Among many input parameters used in docking simulation the following two are the most important and most significantly affect the accuracy and calculation length – **the “number of GA runs”** and the **“maximum number of evaluations”**. Therefore, extensive series of dockings were carried out in order to find their optimal values.

1H66 structure with its original ligand was used. Gasteiger charges were assigned to both macromolecule and the ligand. Ligand was kept fully flexible. Both rigid and flexible dockings were carried out.

All of the parameters were kept at default values, with exception of **number of GA runs** and **maximum number of evaluations**.

In Table 2 we can see that the free energy of binding fluctuates significantly, showing no tendencies throughout rows or columns. Examining the individual components that contribute to free energy of binding, we can see that the unbound system’s energy is responsible for these fluctuations. Its value is calculated once at the beginning of the simulation and used for all dockings contained within given simulation. As a result, free energies of binding are not comparable among parallel simulations. For this reason, we used intermolecular energies, which does not contain the unbound system energy contribution and, hence, is not subject to random error.

The estimate of optimal values of the **number of GA runs** together with **maximum number of evaluations** was based on intermolecular energy. For rigid docking simulation, these values are 10 and 2,500,000, respectively (see Table 2). However, genetic algorithms are stochastic and doing low number of runs can result in generating no appropriate initial populations. Therefore, 100 was chosen as a value of **number of GA runs** to be used in rigid dockings. Even with ten-fold increase in number of runs, simulation is still shorter than two hours.

Table 2: Rigid dockings with different number of runs and evaluations. Free energy of binding and intermolecular energies are shown.

#of evals/runs	Energy	10	50	100	250	1000
250,000	Intermolecular	-7.19	-7.20	-7.22	-7.22	-7.23
	Free E of binding	-6.41	-6.14	-6.20	-6.26	-6.16
2,500,000	Intermolecular	-7.24	-7.23	-7.23	-7.24	-7.24
	Free E of binding	-6.18	-6.19	-6.15	-6.20	-6.16
25,000,000	Intermolecular	-7.24	-7.24	-7.24	/	/
	Free E of binding	-6.18	-6.19	-6.23		

The situation is less clear for flexible dockings. Increase in either of the two factors lead to significant improvement in predicted free energy of binding. However, simulations with 2,500,000 evaluations and 1,000 runs (or 25,000,000 evaluations and 100 runs) already require more than 75 times more than rigid docking with chosen settings. The idea to continue in increasing the parameters was therefore dismissed.

As stated in the previous paragraph, low number of runs can result in poor results even with high number of evaluations. The relevance of this issue is further demonstrated by fluctuations in best free energy of binding found (see Table 3). Because of this, docking with 2,500,000 evaluations and 1,000 runs was preferred to that with 25,000,000 evaluations and 100 runs, which is roughly equivalent in terms of computational cost.

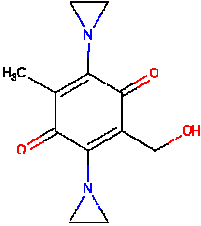
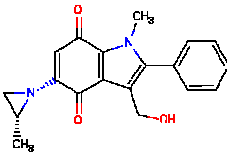
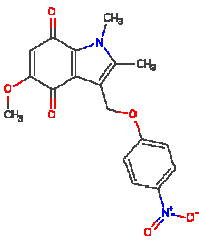
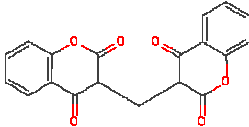
Table 3: Flexible dockings with different number of runs and evaluations. Free Energy of binding and intermolecular energies are shown.

#of evals/runs	Energy	10	50	100	250	1000
250,000	Intermolecular	-4.36	-5.40	-5.83	-5.98	-6.30
	Free E of binding	-12.93	-13.83	-13.80	-14.24	-14.48
2,500,000	Intermolecular	-6.97	-6.72	-6.69	-6.52	-7.06
	Free E of binding	-15.03	-15.53	-15.57	-15.2	-15.65
25,000,000	Intermolecular	-6.56	-6.96	-7.14	/	/
	Free E of binding	-14.55	-15.82	-15.77		

4.3 Analysis of X-ray structures

Each DT-diaphorase structure used in this work was cocrystallized with different ligand. These ligands differ significantly from each other in number of atoms (0-28) as well as in their shape (see Table 4). Their orientation towards FAD's isoalloxazine ring is shown in Fig \$8.

Table 4: Information about original ligands in used PDB structures of DT-diaphorase

PDB code	original ligand systematic name	Structural formula	Number of heavy atoms in original ligand
1D4A	none	-	0
1H66	2,5-diaziridinyl-3-hydroxyl-6-methyl-1,4-benzoquinone		17
1H69	2,3,5,6,tetramethyl-p-benzoquinone		24
1KBQ	methoxy-1,2-dimethyl-3-(4-nitrophenoxy)methyl)indole-4,7-dione		25
2F1O	3,3'-methylenebis(4-hydroxy-2H-chromen-2-one)		26

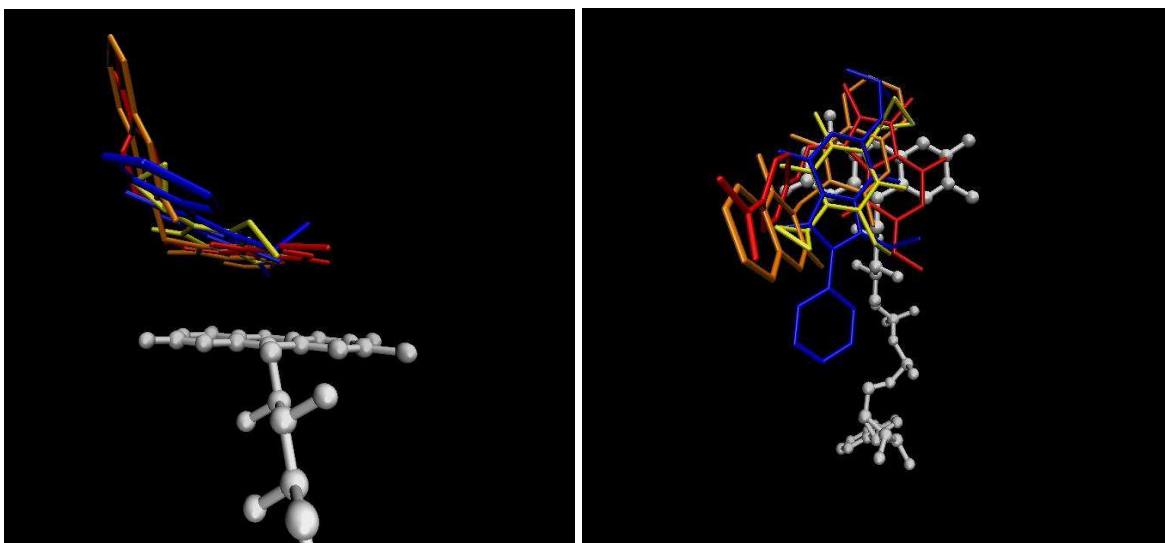


Figure 8: Differences in size and orientations of ligands in original crystal structures. (1H69 – blue, 1KBQ – red, 2F10 – orange, 1H66 – yellow, FAD coenzyme – white, balls and sticks)

In agreement with induced fit theory the different spatial requirements of the ligands induce conformational variation of enzyme's active site. This causes relative displacement of aminoacid sidechains, mainly of His194A, Tyr128C, Phe232C (central and left part of Fig. 9) Slight shifts of enzyme's backbone were observed, most notably of residues Phe232:C and His194:A (see left part of Fig. 9). None of these changes can be accounted for in rigid dockings. Fig. 9 illustrates that significant conformational changes of DT-diaphorase's residues occur during ligand's binding. Therefore we aimed to investigate their impact on the results of docking and decide, if possible, which of the presented structures is most suitable for docking of 2-NBA and 3-NBA into DT-diaphorase.

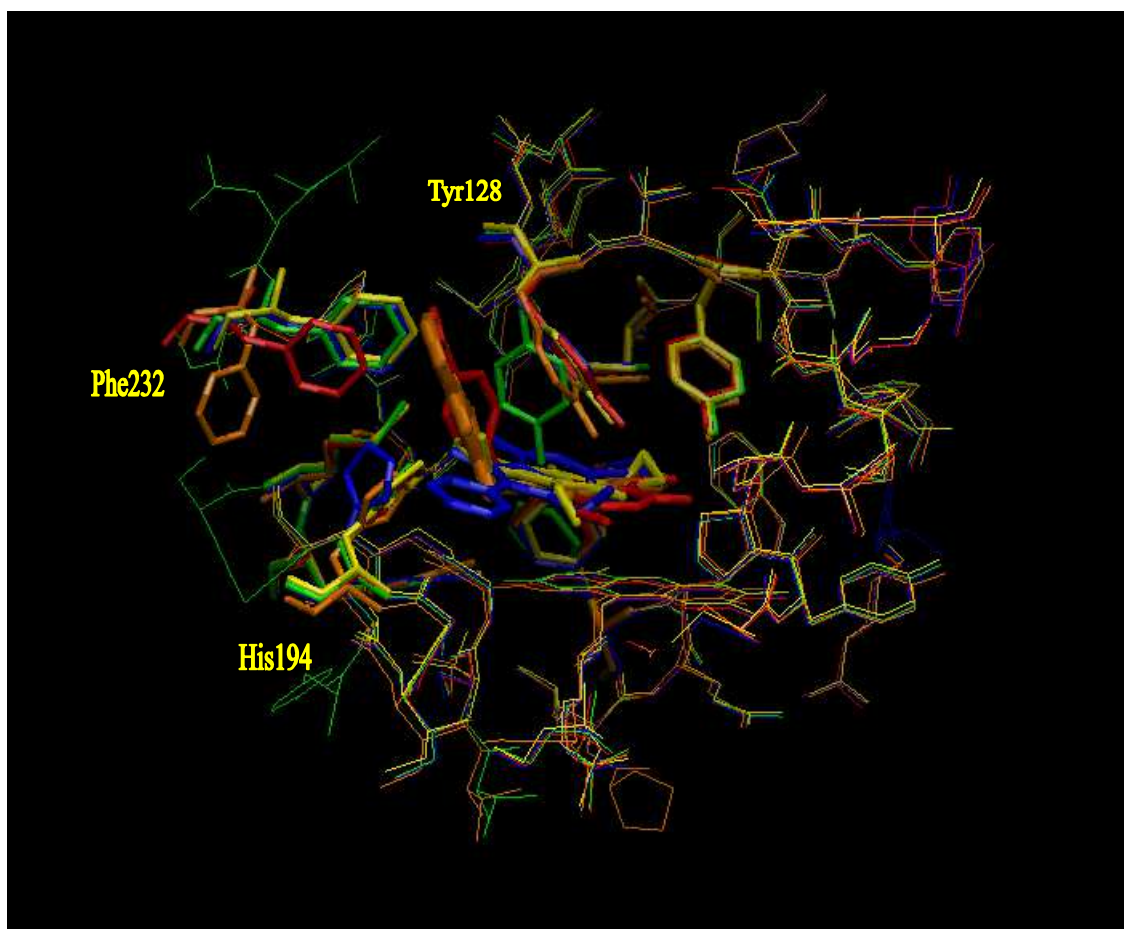


Figure 9: Active site of DT-diaphorase as obtained from various X-ray studies. (1H69 – blue, 1KBQ – red, 2F1O – orange, 1H66 – yellow, 1D4A – green; ligands – bold sticks, most flexible residues – sticks, rest – wires)

4.4 Rigid docking of 2-NBA and 3-NBA into the D,T-diaphorase

Rigid dockings were carried out at 2,500,000 energy evaluations and 100 runs.

With the optimal parameters researched (see chapter 4.2), the comparison of effects of choice of protein structures with different resolutions and original ligands could finally be approached. Rigid dockings of 2-NBA and 3-NBA into all five crystallographic structures of DT-diaphorase were performed and analyzed.

The lowest estimated free energy of binding and intermolecular energy for both ligands was obtained with 1H69 structure (see Table 5). Its original ligand, duroquinone

(see Table 4), bears the closest resemblance to NBAs when compared to other original ligands. It has 24 heavy atoms (NBA has 21) and, similarly to NBA, consists of rings that lie in a single plane.

Other ligands are either significantly smaller (1H66) or are not planar (1KBQ, 2F1O) (see Fig. 8).

As we can see in Fig. 10, substrates usually situate themselves right above the isoalloxazine ring of the cofactor. Such orientation is well stabilized by large Van der Waals contact between FADH₂ and the planar ligand.

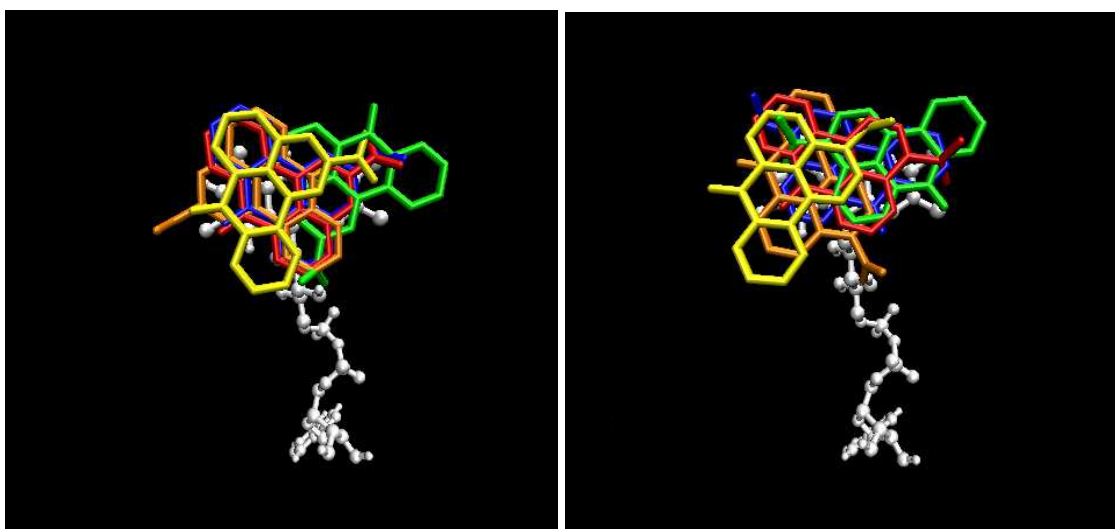


Figure 10: Best orientations found by rigid docking with various protein structures (1H69 – blue, 1KBQ – red, 2F1O – orange, 1H66 – yellow, 1D4A – green, FAD coenzyme – white, balls and sticks; 2-NBA – on the left, 3-NBA – on the right)

If we look at the intermolecular energies obtained we can say they are roughly similar for all starting structures (-8.45 to -8.88 for 2-NBA and -8.25 to -8.95 for 3-NBA), with the single exception of 1D4A, for which significantly weaker interactions were predicted (see Table 5).

This is probably related to the absence of the cocrystallized ligand and therefore by different shape of the active site cavity.

The most significant difference between 1D4A (crystal with no ligand) and the other structures is in position of Tyr128's sidechain. Tyr128 residue hangs right above the isoalloxazine ring of the coenzyme (see Fig. 9). This sidechain sterically hinders the proper orientation of a ligand, forcing it to occupy tight position next to the tyrosine residue,

resulting in significant decrease of VdW interactions of the ligand. The decrease is reflected by weaker intermolecular energy in comparison to structures with cocrystallized ligand. In those structures, Tyr128 is already displaced and allows the ligand to get into more suitable position. See Fig. 11 and 8.

Other, less obvious differences in position of sidechains or even protein's backbone can also contribute to differences among the results of the dockings.

Table 5: Evaluation of best individuals found by rigid dockings. Choosing structure other than 1H69 as referent for RMSD calculation does not yield any result smaller than 1.

ligand	structure	Estimated Free energy of binding	Intermolecular energy	RMSD
2-NBA	1D4A	-5.25	-5.52	6.290
	1H66	-8.18	-8.45	4.205
	1H69	-8.61	-8.88	0.000
	1KBQ	-8.53	-8.79	0.384
	2F1O	-8.19	-8.46	6.558
3-NBA	1D4A	-4.87	-5.24	4.642
	1H66	-7.90	-8.30	5.773
	1H69	-8.60	-8.95	0.000
	1KBQ	-8.55	-8.86	5.466
	2F1O	-7.97	-8.25	5.452

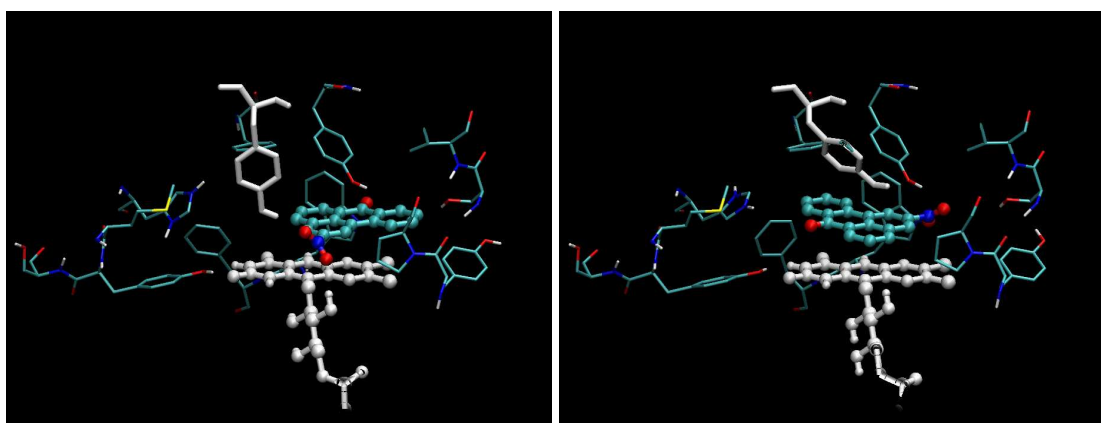


Figure 11: Comparison of 2-NBA orientations found by rigid docking into structures with and without cocrystallized ligand. Left: 1D4A + 2-NBA Right: 1H69+2-NBA. FADH₂ and Tyr128 are white.

Among the different starting PDB structures, the orientations of best individuals of docked NBAs differed significantly (see Fig. 10). As a result, large variability in RMSD was observed (see Table 5).

Original cocrystallized ligand affects the conformation of enzyme's active site to a degree that alters the outcome of rigid docking. Since use of the same protein does not yield similar results, rigid docking cannot be considered a reliable method for determining the best orientation of a ligand.

Hypothesis of induced-fit states that both ligand and the enzyme are spatially deformed in complex. In rigid docking, ligand can be deformed, if its bonds are rotatable, but the protein cannot, which makes the outcome of a docking rather questionable. This problem should be partially compensated in flexible docking, where at least some residues are allowed to react to the presence of a ligand.

4.5 Flexible docking of 2-NBA and 3-NBA into DT-diaphorase

Flexible dockings were carried out at 2,500,000 energy evaluations and 1000 runs.

Flexible dockings of 2-NBA and 3-NBA into all five crystallographic structures of DT-diaphorase were performed and results were analysed.

Flexible domain is shown in Fig. 6.

The best docked individuals elicit intermolecular energies ranging from -8.00 to -8.73 kcal/mol for 2-NBA and from -7.90 to -8.21 kcal/mol for 3-NBA. These values are comparable to those found in rigid docking calculations.

Best results were acquired with 1H69, 1KBQ and 2F1O structures, their intermolecular energies being very close to each other (see Table 6).

Flexible dockings of NBAs into 1H69 retain good results achieved in rigid dockings. In addition, 1KBQ and 2F1O structures, whose original ligands are similar in size to NBAs

but whose 3D structure is quite bulky, now show results which are very close to 1H69.

The averaged intermolecular energy predicted for 2-NBA is 0.4 kcal/mol more negative than corresponding value for 3-NBA. This does not contradict the experimental fact that 2-NBA is worse substrate for DT-diaphorase than 3-NBA⁴⁴ because the metabolic rate is not determined only by the binding affinity but also by reaction activation barrier, which was not determined by these calculations.

Table 6: Evaluation of best individuals found by flexible dockings. Choosing structure other than 1H69 as referent for RMSD calculation does not yield any result smaller than 1.

ligand	PDB structure	Intermolecular energy	RMSD
2-NBA	1D4A	-8.00	5.914
	1H66	-8.23	6.385
	1H69	-8.71	0.000
	1KBQ	-8.73	0.359
	2F1O	-8.73	4.671
3-NBA	1D4A	-7.90	0.297
	1H66	-8.03	6.190
	1H69	-8.21	0.000
	1KBQ	-8.14	5.749
	2F1O	-8.17	0.883

Docking of both ligands into 1D4A was improved by ca. 2.6 kcal/mol of intermolecular energy. Even though it still shows the worst results of flexible docking among the PDB structures used, the difference in intermolecular energy from other PDB structures evaluated was lower than 0.8 kcal/mol. This improvement implies that selected flexible residues do play a major role during docking and, most likely, also during actual enzyme-substrate interaction. However, differences in position of residues other than those present in flexible domain or even backbone displacements could affect the docking results.

Comparison of the best dockings shows two similar orientations for 2-NBA (1H69 and

1KBQ) and three similar orientations for 3-NBA (1D4A, 1H69 and 2F1O). Use of different structures of the same protein in flexible docking produces somewhat similar results, showing the increased reproducibility of this type of docking.

The best individuals failed to propose one orientation as the most probable. We shall now take a look at other than best-energy clusters, which could also hold a key for linking results of separate dockings.

The intermolecular energy values of these alternative clusters are shown in Tables 7 and 8.

Table 7: Intermolecular energies of best and close to best clusters found by flexible dockings of 2-NBA: Clusters with intermolecular energy up to 0.5 kcal/mol lower than best cluster of given docking are listed.

cluster rank \ PDB	1	2	3	4	5
flexible-1D4A	-8.00	-7.56	-7.52		
flexible-1H66	-8.23	-8.11			
flexible-1H69	-8.71	-8.57	-8.37	-8.23	
flexible-1KBQ	-8.73	-8.64	-8.40	-8.31	
flexible-2F1O	-8.73	-8.59	-8.46		

Comparison of these results shows that most of these 16 low energy clusters of 2-NBA, can be well described by 3 representative ligand orientations (see Fig. 12).

Table 8: Intermolecular energies of best and close to best clusters found by flexible dockings of 3-NBA: Clusters with intermolecular energy up to 0.5 kcal/mol lower than best cluster of given docking are listed.

cluster rank \ PDB	1	2	3	4	5
flexible-1D4A	-7.90	-7.71	-7.52	-7.46	
flexible-1H66	-8.03	-7.92	-7.77	-7.62	
flexible-1H69	-8.21	-8.04	-8.00	-7.91	-7.75
flexible-1KBQ	-8.14	-7.93	-7.92	-7.75	-7.73
flexible-2F1O	-8.17	-7.71	-7.71	-7.68	

Similarly, most of these 22 low energy clusters of 3-NBA, are well described by 2 representative ligand orientations (see Fig. 13).

Best and close to best clusters found by rigid docking were also examined (data not shown) but no representative structures were determined. These results generally show much higher variability than the results of flexible docking. This means that Autodock's flexible docking algorithm was able to find results with higher reproducibility and is, thus, less sensitive to the variation of active site than rigid docking.

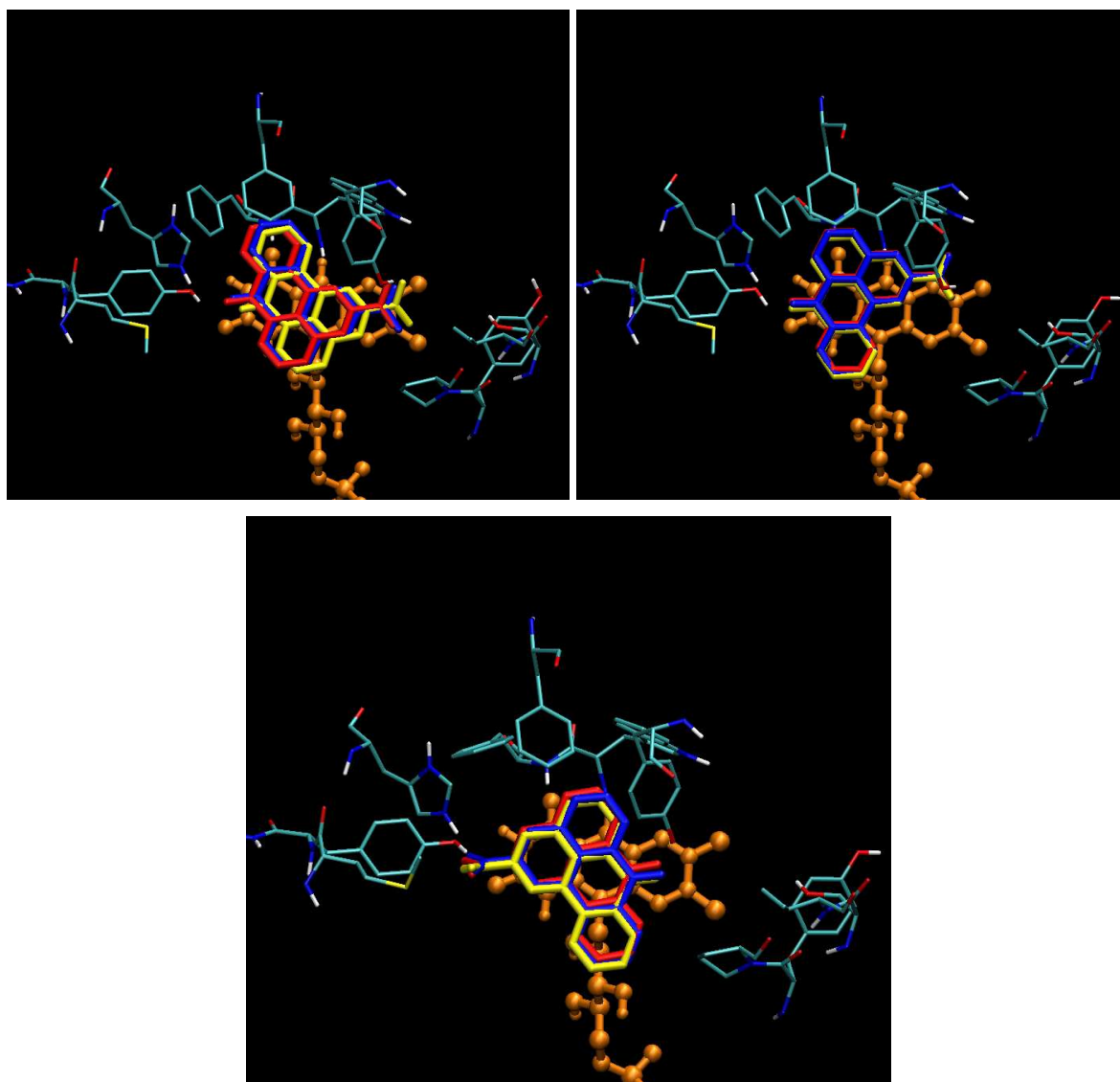


Figure 12 (see next page): Representative results of flexible docking of 2-NBA:

Legend: Colour: PDBcode-ClusterRank; see Table 7 for more info.

Upper left: Blue:1H69-1, Red:1KBQ-1, Yellow-1KBQ-4, Orange:FADH₂

Upper right: Blue:1H66-2, Red:1H69-2, Yellow-1KBQ-2, Orange:FADH₂

Middle: Blue:1H66-1, Red:1KBQ-3, Yellow-2F1O-3, Orange:FADH₂

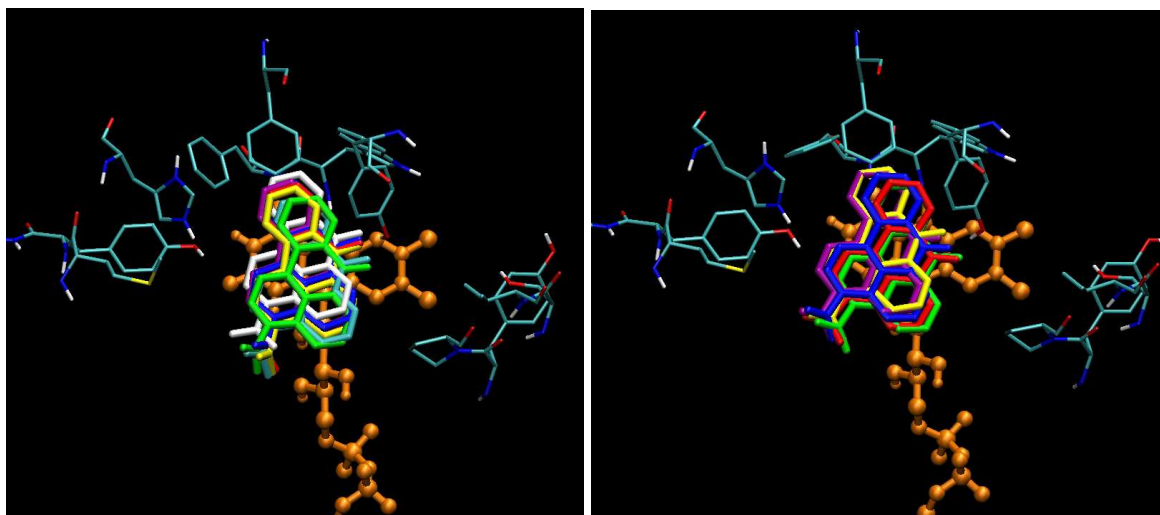


Figure 13: Representative results of flexible docking of 3-NBA:

Legend: Colour: PDBcode-ClusterRank; see Table 8 for more info.

On the left: Blue:1H69-1, Red:1D4A-1, Yellow:1H66-2, Green:1H66-3,

Purple:1KBQ-2, Cyan-1KBQ-5, White:2F1O-1, Orange:FADH₂

On the right: Blue:1H66-4, Red:1D4A-2, Yellow:1H69-2, Green:1H69-5,

Purple:1KBQ-4, Orange:FADH₂

5 Conclusions

Well-established classical hard-soft approach to ligand docking (rigid docking) is considered to be suitable for prediction of ligand binding modes and energies to rigid proteins or for docking of a ligand structurally resembling the original cocrystallized ligand. Unfortunately, a typical enzyme, that metabolizes drugs or other xenobiotics, is able to bind wide spectrum of ligands of varying shape and size. In addition, the number of crystallographic structures data available for these enzymes is usually limited. Utilizability of rigid docking under these conditions is very limited. Thus, new approaches taking into account also enzyme's flexibility are required.

Such approach, implemented in Autodock v.4.0, was successfully tested in this study. Fortunately, several crystallographic structures of human DT-diaphorase (NQO1) are available in PDB database. These carry substrates of varying shape and size. Hence, this enzyme is an ideal candidate for testing abilities of the new flexible docking implementation, and for evaluation of reproducibility of classical rigid docking approach of non-cocrystallized ligands.

Rigid docking of model substrates (2-NBA and 3-NBA) to various structures of human DT-diaphorase showed to be heavily dependant on the ligand originally cocrystallized with the enzyme. Overlaid structures of different crystallographic structures showed its high flexibility, as the active site adjusts to the original ligand (Fig. 9). Rigid docking into such structure is limited to the free space originally occupied by the original ligand. This results in major discrepancy among outcomes of rigid dockings (Fig. 10). Thus, we can conclude that rigid docking cannot be considered a reproducible and reliable method for modelling the system studied in this work.

Employment of the soft-soft (flexible) docking algorithms have proved to be worth of increased computational cost. Both substrates were docked successfully into the binding site cavity above the isoalloxazine ring of FADH₂, occupying the same space as original ligands. In addition, resulting substrate orientations were much less sensitive to the variations inside the enzyme's active site. Analysis of results of flexible dockings

identified three main structural motives representing the most preferable binding modes of 2-NBA in human DT-diaphorase's active site (Fig. 14). Predicted intermolecular energies for these representative structures were very close, therefore all of them should be considered to be results that do occur in the model, although with unequal yet relevant probability. Calculated intermolecular energies are -8.73 kcal/mol for 1KBQ-1; -8.64 kcal/mol for 1KBQ-2 ; -8.46 kcal/mol for 2F1O-3 and probabilities inferred from them are 40%; 35% and 20%, respectively.

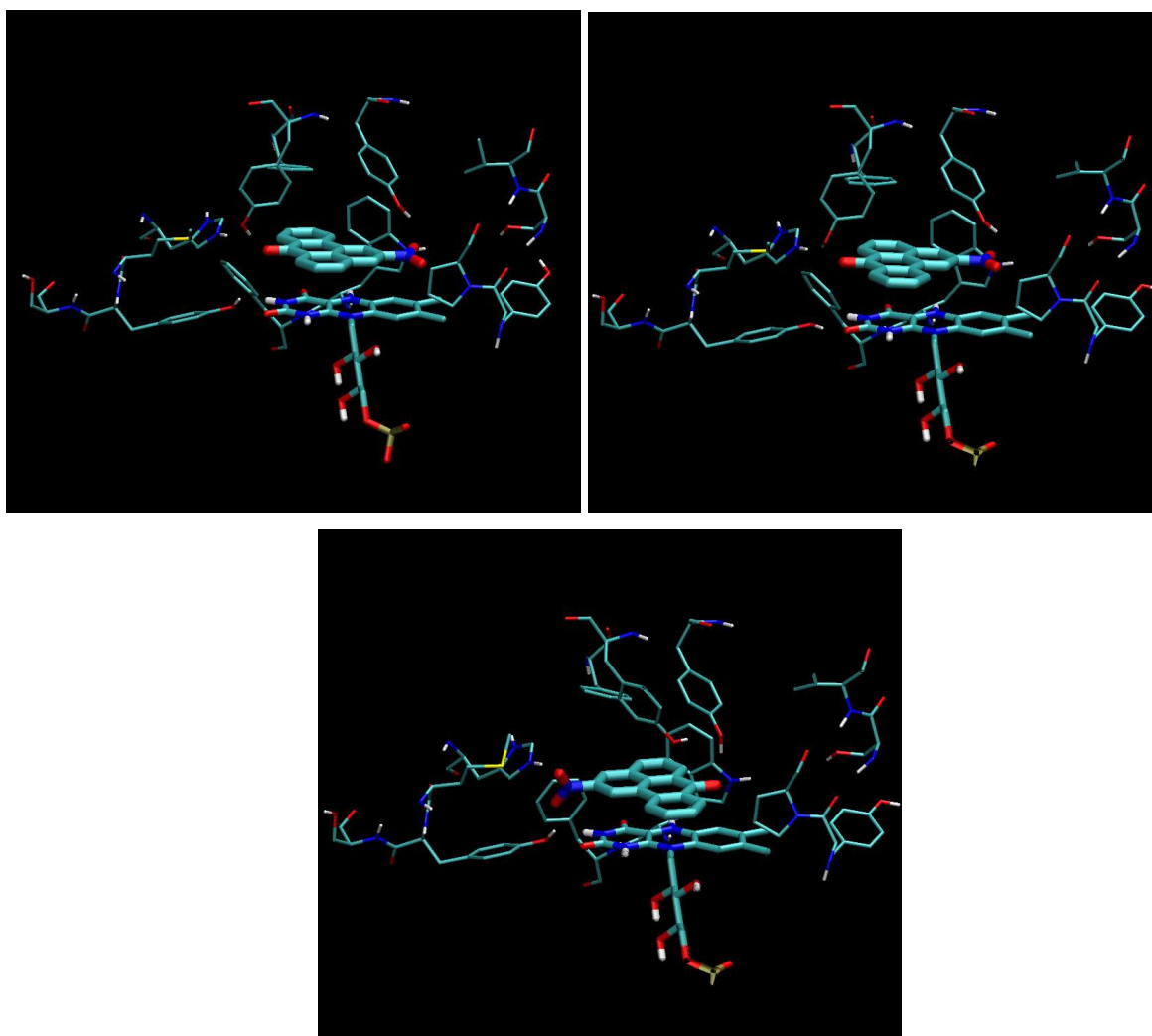


Figure 14: Representative orientations of 2-NBA found by flexible dockings. Upper left: 1KBQ-1; Upper right: 1KBQ-2; Middle: 2F1O-3

Similarly, results of flexible dockings of 3-NBA show two main structural motives that represent the most preferable binding modes of this substrate to human DT-diaphorase (Fig. 15). Close intermolecular energies predicted for both structures suggest similar

probability of their occurrence in the real systems. Calculated intermolecular energies are -8.21 kcal/mol for 1H69-1 and -8.04 kcal/mol for 1H69-2, and therefore both binding modes occur with almost equal probability. The RMSD of these two orientations (Fig. 15) is only 1.056, which may indicate that 3-NBA prefers conformation that tilts between them.

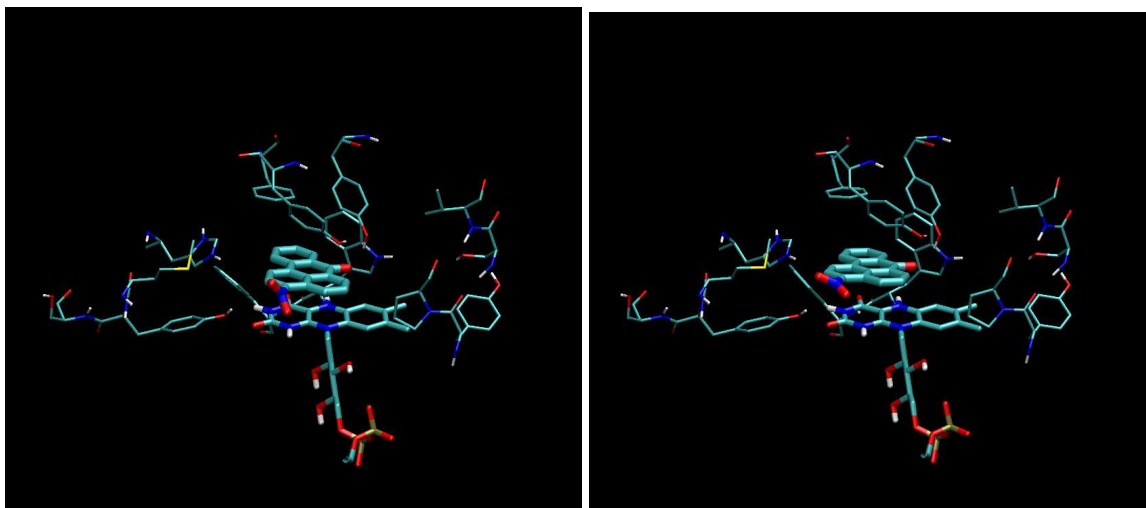


Figure 15: Representative orientations of 3-NBA found by flexible dockings:
Left: 1H69-1; Right: 1H69-2

Although calculated energies indicate that both studied ligands fit well into the DT-diaphorase's active site, 3-NBA shows slightly lower affinity than 2-NBA, 0.4 kcal/mol on average. 3-NBA is known to be a better substrate for DT-diaphorase than 2-NBA, and this result suggests that the difference in metabolic rate is not caused by their different affinity toward the enzyme.

6 Literature

1. Sousa,S.F., Fernandes,P.A., Ramos,M.J.; *Proteins* **65**,15-26, 2006
2. Morris,G.M., Goodsell,D.S., Huey,R., Olson,A.J.; Autodock v4.00, The Scripps Research Institute, 1991-2007
3. http://en.wikipedia.org/wiki/Simulated_annealing
4. Vacek,J.; Computer simulations of biomacromolecules, Course at Charles University 2007-2008
5. Morris,M.G., Goodsell,S.D., Halliday,S.R., Huey,R., Hart,E.W., Belew,K.R., Olson,J.A.; *J Comput Chem* **19**, 1639-1662, 1998
6. Ross,B.J.; *Practical Handbook of Genetic Algorithms (vol.3)*, 1-16, CRC Press, 1999
7. Huey,R., Morris,M.G., Olson,J.A., Goodsell,S.D.; Wiley Interscience 1.2.2007, www.interscience.wiley.com
8. Ernster,L., Navazio,F.; *Acta Chem Scand* **12**, 595, 1958
9. Chen,S., Wu,K., Knox,R.; *Free Radic Biol Med* **29**, 276-284, 2000
10. Faig,M., Bianchet,M.A., Talalay,P., Chen,S., Winski,S., Ross,D., Amzel,L.M.; *Proc Natl Acad Sci USA* **97**, 3177, 2000
11. Faig,M., Bianchet,M.A., Winski,S., Hargreaves,S., Moody,C.J., Hudnott,A.R., Ross,D., Amzel,L.M.; *Structure (Camb)* **9**, 659, 2001
12. Winski,S., Faig,M., Bianchet,M.A., Siegel,D., Swann,E., Funk,K., Duncan,M.W., Moody,C.J.,Amzel,L.M.,Ross,D.; *Biochemistry* **40**, 15135, 2001
13. Yosef,S.; Unpublished data; Found at <http://www.pdb.org/pdb/explore/explore.do?structureId=2F1O>
14. <http://atlasgeneticsoncology.org/Genes/NQO1ID375.html>
15. Prochaska,H.J., Talalay,P.; *Academic Press, Ltd.*, New York 195-211, 1991
16. Beyer,E.R., Segura-Aguilar,J., di Bernardo,S., Cavazzoni,M., Fato,R., Fiorentini,D., Galli,C.M., Setti,M., Landi,L., Lenaz,G.; *Proc Natl Acad Sci USA* **93**, 2528-2532, 1996
17. Wallin,R.; *Biochem J* **236**, 685-693, 1986
18. Boland, M.P., Knox, R.J., Roberts, J.J.; *Biochem Pharmacol* **41** , 867-875, 1991
19. Huang, M.T., Miwa,G.T., Cronheim, N., Lu, A.Y.H.; *J Biol Chem* **254**, 11223-11227, 1979
20. de Long,J.M., Prochaska,J.H., Talalay,P.; *Proc Natl Acad Sci USA* **83**, 787-791, 1986
21. Ross,D., Beall,H.D., Siegel,D., Traver,R.D., Gustafson,D.L.; *Br J Cancer* **74**, (Suppl.XXVII), S1, 1996
22. Li,R., Bianchet,A.M., Talalay,P., Amzel,M.L.; *Proc Natl Acad Sci USA* **92**, 8846-

- 8850, 1995
23. Frey,P.A., Northrop,D.B.; *Enzymatic mechanisms*, IOS Press, 1999
 24. Iyanagi,T., Yamazaki,I.; *Biochim Biophys Acta* **216**, 282-294, 1970
 25. Tedeschi,G., Chen,S., Massey,V.; *J Biol Chem* **270**, 1198-1204, 1995
 26. Purohit,V., Basu,A.K.; *Chem Res Toxicol* **13**, 673 - 692, 2000
 27. Enya,T., Suzuki,H., Watanabe,T., Hirayama,T., Hisamatsu,Y.; *Environ Sci Technol* **31**, 2772-2776, 1997
 28. Salamone,M.F., Heddle,J.A., Katz,M.; *Environ Int* **2**, 37-43, 1979
 29. Takamura-Enya,T., Suzuki,H., Hisamatsu,Y.; *Mutagenesis* **21**, 399-404, 2006
 30. Phouongphouang,P.T., Arey, J; *Atmos Environ* **37**, 3189-3199, 2003
 31. Tang,N., Taga,R., Hattori,T., Tamura,K., Toroba,A., Kizu,R., Hayakawa,K.; *Anal Sci* **20**, 119-123, 2004
 32. Seidel,A., Dahmann,D., Krekeler,H. and Jacob,J.; *Int J Hyg Environ Health* **204**, 333-338, 2002
 33. Murahashi,T., Iwanaga,E., Watanabe,T., Hirayama,T.; *J Health Sci* **49**, 386-389, 2003
 34. Murahasi,T., Watanabe,T., Otake,S., Hattori,Y., Takamura,T., Wakabayashi,K., Hirayama,T.; *J Chromatogr A* **992**, 101-107, 2003
 35. Grabchev,I., Moneva,I., Kozlov,A., Elyashevich,G; *Mat Res Innovat* **4**, 301-305, 2001
 36. Grabchev,I., Bojinov,V., Moneva,I.; *Dyes Pigments* **48**, 143-150, 2001
 37. Kirilova,E.M., Kalnina,I., Kirilov,G.K., Meirovics,I.; *J Fluoresc* **18**, 645-648, 2008
 38. Nagy,E., Shuichi,A., Takamura-Enya,T., Zeisig,M., Moller,L.; *Mutagenesis* **22**, 135-145, 2007
 39. Bieler,C.A., Cornelius,M.G., Klein,R., Arlt,V.M., Wiessler,M., Phillips,D.H., Schmeiser,H.H.; *Int J Cancer* **116**, 833-838, 2005
 40. Nagy,E., Johansson,C., Zeisig,M., Möller,L.; *J Chromatogr B Analyt Technol Biomed Life Sci* **827**, 94-103, 2005
 41. Arlt,V.M.; *Mutagenesis* **20**, 399-410, 2005
 42. Arlt,V.M., Glatt,H., Gamboa da Costa,G., Reynisson,J., Takamura-Enya,T., Phillips,H.D.; *Toxicol Sci* **98**, 445-457, 2007
 43. Arlt,V.M., Stiborová,M., Henderson,C.J., Osborne,M.R., Bieler,C.A., Frei,E., Martínek,V., Sopko,B., Wolf,C.R., Schmeiser,H.H., et al.; *CancerRes* **65**, 2644-2652, 2005
 44. Sborník Abstraktu; *Chemické Listy* **101**, 927-985, 2007
 45. Nagy,E.; *Diploma thesis*, Karolinska institutet, 2006
 46. <http://www.rcsb.org/>
 47. Sanner,M.F.; *J Mol Graphics Mod* **17**, 57-61, 1999

48. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Montgomery,Jr.,J.A., Vreven,T., Kudin,K.N., Burant,J.C. and others; Gaussian 03, Gaussian,Inc. 2003
49. Humphrey,W., Dalke,A, Schulten,K.; *J Molec Graphics* **14**, 33-38, 1996
50. Guex,N. Peitsch,M.C.; *Electrophoresis* **18**, 2714-2723, 1997
51. Accelrys Software Inc; Accelrys DS Visualizer 1.7, 2005-2006
52. Schulz,W.S., Krummeck,A., Rosinger,I., Eickelmann,P., Neuhaus,C., Ebert,T., Schmitz,-Drager,B.J., Sies,H.; *Pharmacogenetics* **7**, 235-239, 1977