

Novel inhibitors to novel targets in infectious diseases through structure-based virtual screening

Magnus Gäredal



UPPSALA
UNIVERSITET

Molecular Biotechnology Programme

Uppsala University School of Engineering

UPTEC X 08 054	Date of issue 2009-01
Author	Magnus Gäredal
Title (English)	Novel inhibitors to novel targets in infectious diseases through structure-based virtual screening
Title (Swedish)	
Abstract	<p>An approach using structure-based virtual screening to identify potential inhibitors for <i>Plasmodium falciparum</i> spermidine synthetase (<i>Pf</i>-SRM) and <i>Mycobacterium tuberculosis</i> 1-deoxy-d-xylulose 5-phosphate reductoisomerase (<i>Mt</i>-DXR) was developed. Starting from a database of 2.6 million compounds and applying this strategy, 26 and 30 potential inhibitors to <i>Mt</i>-DXR and <i>Pf</i>-SRM were suggested.</p>
Keywords	
Supervisors	Micael Jacobsson, iNovacia AB
Scientific reviewer	Anders Karlén, Uppsala University
Project name	Sponsors
Language	Security
English	
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages
	40
Department of medical Chemistry Box 574 S-75123 Uppsala	Biomedical Center Tel +46 (0)18 4710000
	Husargatan 3 Uppsala Fax +46 (0)18 555217

Novel inhibitors to novel targets in infectious diseases through structure-based virtual screening

Magnus Gäredal

Sammanfattning

Att framställa nya läkemedel är både väldigt dyrt och tar mycket tid i anspråk. En stor utmaning i framtagandet av nya antibakteriella mediciner är att hitta substanser – blivande läkemedel – som skadar centrala mekanismer i den sjukdomsalstrande organismen och samtidigt inte är för giftig mot värden (oftast människan) som ska ta läkemedlet.

Att med hjälp av datorbaserad screening ta fram substanser som har god sannolikhet att inhibera vitala delar hos en patogen organism är ett sätt att både snabba på och kostnadseffektivisera läkemedelsframställning. Det här exjobbet har dels syftat i att utarbeta en sådan screeningmetod, dels haft som mål att med denna metod ta fram 30 substanser som potentiellt skulle kunna vara utgångspunkter för utveckling av nya läkemedel mot sjukdomarna malaria och tuberkulos.

Malaria och tuberkulos orsakar enormt mänskligt lidande och död världen över. Varje år blir 500 miljoner människor allvarligt sjuka i malaria. De flesta dödsfallen orsakas av den mest aggressiva varianten av malaria – *Plasmodium falciparum* (*Pf*). Tuberkulos är i dagsläget en av de sjukdomar som orsakar flest dödsfall i världen, nästan 2 miljoner per år. Sjukdomen orsakas av bakterien *Mycobacterium tuberculosis* (*Mt*). För både malaria och tuberkulos är fallet att det främst är fattiga och i övrigt utsatta personer som drabbas svårast.

Den röntgenkristallografiskt bestämda strukturen av ett protein från vardera *Pf* och *Mt* har använts i detta projekt – *Plasmodium falciparum* spermidine synthetase (*Pf*-SRM) och *Mycobacterium tuberculosis* 1-deoxy-d-xylulose 5-phosphate reductoisomerase (*Mt*-DXR). Dessa proteiner har i tidigare studier visats vara essentiella för respektive organisms funktion.

En databas innehållande 2,6 miljoner olika substanser har sökts igenom för att hitta molekyler som binder till centrala delar av de aktuella proteinstrukturerna. En automatiserad filtrering med efterföljande dockningar har kombinerats med en visuell analys och selektion av substanser som anses lovande som potentiella inhibitorer. Målet har varit att välja 30 potentiella inhibitorer till vardera proteinstruktur, vilket har uppnåtts.

Resultatet från projektet har resulterat i en artikel publicerad i Journal of Medicinal Chemistry. I denna artikel har några av de substanser som togs fram genom detta projekt verifierats som inhibitorer till proteinet *Pf*-SRM genom experimentella studier.

Examensarbete 20p
Civilingenjörsprogrammet Molekylär bioteknik

Uppsala universitet december 2008

1. Introduction.....	2
1.1 Virtual screening.....	2
1.2 Aim of the project.....	3
2 Material and methods.....	3
2.1 Target structure files.....	3
2.2 Compound database.....	3
2.3 Schrödinger software suit.....	3
2.3.1 Maestro.....	4
2.3.2 Phase.....	4
2.3.3 Ligprep.....	4
2.3.4 Glide.....	4
2.4 Multiple active site correction (MASC).....	9
2.5 Visual analyzis.....	9
3 Experimental procedure.....	10
3.1 <i>Plasmodium falciparum</i> spermidine synthase (<i>Pf</i> -SRM).....	10
3.1.1 Target structure file.....	10
3.1.2 First pharmacophore search.....	10
3.1.3 Glide SP dock.....	11
3.1.4 Second pharmacophore search.....	11
3.1.5 Extracting the hit-compounds from the second pharmacophore search.....	11
3.1.6 Glide XP Dock.....	12
3.1.7 MASC.....	12
3.1.8 Final XP-docking and scoring.....	12
3.1.9 Visual analysis of top-scoring poses.....	12
3.2 <i>Mt</i> -DXR.....	13
3.2.1 Target structure file.....	13
3.2.1 First pharmacophore search.....	13
3.2.2 Glide SP dock.....	13
3.2.3 Second pharmacophore search.....	14
3.2.4 Glide XP Dock.....	14
3.2.5 MASC.....	14
3.2.6 Final visual analysis.....	14
4 Results.....	14
4.1 Targets.....	14
4.1.1 <i>Pf</i> -SRM.....	15
4.1.2 <i>Mt</i> -DXR.....	16
4.2 MASC.....	18
4.3 Virtual screening.....	18
4.3.1 <i>Pf</i> -SRM.....	18
4.3.2 <i>Mt</i> -DXR.....	27
5. Conclusion.....	35
6. Discussion.....	36

1. Introduction

Development of novel drugs is expensive and time consuming. The major challenge in the drug development process for antibacterials is to find targets crucial for the survival of the infectious organism of interest, and to find efficacious inhibitors of that target.

Computer-based methods for the identification of potential inhibitors and for the prediction of binding affinity are widely used in today's medicinal chemistry research. One of the more common methods used is referred to as "virtual screening" (VS), and can roughly be explained as an *in silico* evaluation of large libraries of compounds using computational methods.

This master degree project work presents a virtual screen to find potential inhibitors to two novel drug targets in the infectious diseases malaria and tuberculosis.

Malaria affects the lives of approximately 40% of the world's population and every year 500 million people become severely ill with malaria¹. Most of the deaths are caused by the most virulent strain of malaria-causing parasites – *Plasmodium falciparum* (*Pf*). Malaria strike mainly among poor people in countries already fighting poverty, and effective low-cost treatment against malaria is needed to end immense human suffering and death. Presently there are some anti-malarial drugs, but they are too expensive to be able to solve the problem, and drug-resistant strains of *Pf* are rapidly emerging and spreading. Hence, continuous research for new effective anti-malarial drugs is of great importance.

Tuberculosis is by now one of the diseases that causes most casualties throughout the world. According to the World Health Organization, Tuberculosis annually causes almost 2 million deaths². As in the case of malaria, mostly poor and vulnerable people are infected. Obviously, effective drugs against the bacteria causing tuberculosis, *Mycobacterium tuberculosis* (*Mt*) is urgently needed.

In the search for effective drugs against malaria and tuberculosis, metabolic pathways that differ between humans and *Pf* or *Mt* respectively should be the focus in order to do as much harm as possible to the infecting organism without too many negative effects for the human.

1.1 Virtual screening

Three important goals for computational chemistry can be identified in the drug discovery process: identifying novel ligands by virtual screening, predicting ligand-target binding affinities of novel ligands and predicting the binding modes of known active ligands³. All of these include the docking of ligands into target structures as well as scoring the ligand binding affinity. Docking of ligands into target structures means to fit small molecules into the region of interest in a target, often the catalytic site of a protein, by computer simulation. To evaluate the affinity of the potential binder, docking is followed by scoring. Scoring means to predict the free energy of binding of the ligand to the target, given a suggested binding pose.

The docking of a ligand includes several degrees of freedom. To make the docking calculation possible, the target usually is treated as a rigid body, called the "rigid body approximation" while the ligand usually is treated as a flexible molecule. The scoring that follows the sampling of the possible positions for the ligands can be performed by one or several scoring functions.

Many computer programs used for docking are quite successful in predicting the binding pose⁴. Scoring has however been shown to be very difficult. A successful prediction can be made by a specific scoring function for a specific target, however, no scoring function is very successful in

generally predicting ligand-target interactions⁵. Therefore, the ranking of binding affinities is one of the biggest challenges in the development of new docking programs and scoring functions¹.

This master degree project combines two different methods of virtual screening; 3D pharmacophore filtering and ligand docking combined with scoring in the search of novel binders to the two targets. Since the scoring function can not be expected to be reliable enough by itself, it was combined with an extensive visual analysis of the docked poses. In addition, multiple active site correction (MASC)⁶ was used in an attempt to improve the scoring of the ligands.

1.2 Aim of the project

The aim of this project was to select 30 potential inhibitors for each of the targets described above which has exclusively been done by computational methods. The work can be seen as a starting point in the discovery and development of novel drugs to the diseases described above and has to be followed by many different experimental studies starting with determination of the substances enzyme inhibition capacity.

2 Material and methods

2.1 Target structure files

The study of the current proteins and examination and calculation of ligand-target interaction was made based on 3-dimensional X-ray structures. The pdb-files giving the atomic coordinates of the targets were fetched from the *RCSB Protein Data Bank* (www.pdb.org).

2.2 Compound database

The database containing all potential-inhibitor compounds used in this project work contained approximately 2.6 million non-redundant structures, compiled from 4 million commercially available screening compounds from 13 different suppliers. The compounds were filtered, removing those with a molecular weight larger than 550 kD or more than 10 rotatable bonds. The compounds were converted to 3D using LigPrep (see section 2.6), with one ionization state (the neutral state), one stereo isomer and one tautomer per structure. The final database that was used by Phase (see section 2.5) contained 2,640,712 compounds.

2.3 Schrödinger software suit

All calculations and visualizations in this project work have been made with software belonging to the Schrödinger software suit. The programs used are as follows:

Maestro ⁷	Graphical user interface (GUI) for all of Schrödinger's computational programs
Phase ^{8, 9}	Application for 3D-pharmacophore modelling and search of 3D-databases and files.
Ligprep ¹⁰	A collection of tools preparing all-atom 3D-structures from 2D- or 3D- structures
Glide ¹¹	Grid-based docking and scoring application

In the following section the programs as well as the theory and methodology for each calculation step will be presented.

2.3.1 Maestro

Maestro is the Graphical User Interface (GUI) for all the Schrödinger programs. Maestro contains tools for loading and storing chemical structures, for editing and manipulating them as well as building new molecules and for visualizing the calculation results for these structures.

2.3.2 Phase

Phase is a 3D pharmacophore modeling program. It can be used to construct 3D pharmacophores, and to search both 3D-databases and files for molecules that fit a pharmacophore hypothesis. Phase can also be used to prepare a 3D-database that includes pharmacophore information. The following description will focus on the creation and editing of a pharmacophore hypothesis and how the search is set up.

New hypotheses are created using a set of pharmacophore features. Phase uses the feature definitions to identify all the possible pharmacophore sites in a reference ligand, for example the ligand(s) present in a crystal structure. Features that can be identified are: hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic interactions (H), negative ionisable (N), positive ionisable (P), and aromatic ring (R). In addition to the pharmacophore sites, excluded volumes can be specified, to define volumes in 3D-space that compounds are forbidden to occupy. A pharmacophore hypothesis consists of several files, each of them containing information on the pharmacophore hypothesis. Manual editing of some of these files may be needed, for example by defining each pharmacophore feature as either mandatory or optional or by changing the coordinates of the features if they are supposed to be positioned somewhere else than on an atom in the reference ligand.

For each Phase run, an input file is created either by Maestro when starting the Phase run in the GUI or by manual creation of a text file. The input file contains all specific information necessary for a Phase run. Some crucial parameters of the input file are: the name of the pharmacophore to use for the phase run, name of the database or file to be searched, the minimum amount of pharmacophore sites in the pharmacophore to be matched to get a hit, and the name of the output file. The input file also defines if excluded volumes, given in an additional file, should be used in the phase run, as well as if the ligands should be treated as flexible or non-flexible. There is no way, though, in the version of Phase used in this work to prevent phase from translating and rotating the entire molecule during a search.

The pharmacophore hypothesis can be used for searching a file or a database for matches to the hypothesis. The matching conformation is saved if it gives rise to a hit during the search.

2.3.3 Ligprep

Ligprep is a collection of tools for preparing 3D-structures of large numbers of drug-like molecules. Ligprep can produce one single, low-energy 3D-structure for each molecule. It can also be used to produce several structures from each input compound with various stereochemistries, ring conformations, ionization states and tautomeric forms. If Ligprep is run together with the Schrödinger application Epik, a pKa prediction program, the tautomerization and ionisation stages are more efficiently adjusted as compared to when using only the default mode of Ligprep⁸.

2.3.4 Glide

Glide is a grid-based docking and scoring application. Glide is used when searching for favourable interactions between one or several ligands and a target receptor. Glide docking is performed by

using a series of hierarchical filters. The active site region is defined by a grid, calculated for a specific target receptor. The grid for the receptor is generated prior to the actual docking run. The target grids can be used for several docking runs as long as the target conditions are unchanged. The target structure file has to be prepared before generation of the grid.

2.3.4.1 Preparation of the target structure

Preparation of the target receptor includes several different steps. This preparation is performed in Maestro:

- Setting the correct bond orders and formal charges of the ligand(s) and other nonstandard residues in the pdb file.
- Running the command "protein preparation" in Maestro. This step includes definition of the reference ligand bound to the target structure file. If the structure file contains hydrogen atoms prior to protein preparation, these atoms has to be deleted before the protein preparation is run. Amino acid residues not participating in salt bridges and residues further away than a specified distance from any ligand atom are neutralized. For XP scoring (see section 2.3.4.5), a neutralization zone around the ligand is used instead of a sharp neutralization-border to get a smoother neutralization. This step is followed by addition of hydrogens to the protein and, if present, cofactors. The last step carries out a series of restrained minimizations of the protein-ligand complex, using Impact¹².
- After the protein preparation steps the protein is included as a new entry in the project table and is ready for docking studies.

2.3.4.2 Grid calculation

After target protein preparation the receptor grid is calculated. The grid describes the properties and shape of the part of the protein molecule where the ligands are going to be docked. The grids are constituted of "site points", and for each "site point" the properties corresponding to the terms of the scoring function are calculated. The Coulomb and van der Waals part of the grid is initially built using boxes of quite large size (typically 3,2 Å³). Depending of the distance of the box to the van der Waals surface of the protein, the box is then hierarchically refined into 1,6 Å³, 0,8 Å³ or 0,4 Å³ boxes.

Maestro is used when setting up the grid calculations. The parameters that have to be defined before the actual grid calculation are:

- Which reference ligand to be used in the calculations.
- The dimensions of the grids as well as for the core box. The grid dimension is the size of the box for which grids are calculated. In the center of the grid box a core box is found. The diameter midpoint for each docked ligand has to remain inside this box. The default size of the grids is the length of the reference ligand from the core box to the border of the grids.
- Definition of constraints (positional, hydrogen bond, metal or hydrophobic constraints) can be made. If grids with constraints have been calculated the use of the constraints are optional for each docking run.
- Selecting an output directory where the grid files are saved.

2.3.4.3 Ligand docking

The actual docking run can be performed using two different levels of accuracy

- Glide SP uses a soft scoring function allowing ligands to have some imperfections with regard to the interactions with the target.

- Glide XP gives severe penalties to poses that do not fit in the binding site having for example strongly polar groups not being adequately exposed to solvent.

Glide XP does a more precise docking but takes considerable more CPU time into account. This means that only quite small sets of ligands can be docked and scored in Glide XP mode. Large sets of ligands have to be docked by Glide SP. If better docking accuracy is required, the SP-dock can be followed by Glide XP docking for the top-scoring ligands. In the following, a brief overview of the Glide docking calculation workflow will be made. This will be followed by a more detailed description of the Glide XP scoring function, to give a concrete example of how scoring is performed.

2.3.4.4 Docking calculation procedure

In the initial calculation step, Glide produces a large number of different conformational states of each ligand to be docked to the target grid. Each ligand is divided into rotamer groups and a core region (which is the part of the ligand that's left when all rotamer groups are excluded). Every rotamer state of all the rotamer groups of a ligand are generated and enumerated. This is followed by a screening process where high-energy conformations or other conformation that are not suitable for receptor binding are eliminated. The screen is performed by evaluating the torsional energy of the different minima using a modified version of the OPLS-AA molecular mechanics force field.

Every rotamer state that is not rejected is, together with the core region, docked as a single object. An exhaustive search of possible positions and orientations are performed. The previously calculated grid is made up by "site points" of 2 Å distance, and the ligand is placed at each of these site points. For each placement, ligand atoms are positioned at a pre-specified selection of ligand diameter orientation. If the position causes too many steric clashes with the receptor the placement is skipped. If the orientation is accepted, the ligand is rotated about the ligand diameter and the resulting ligand-target interactions that the different positions of the ligand atoms gives rise to are scored by a scoring function.

A small number of the best refined ligand poses are energy minimized on the grids using the pre-calculated OPLS-AA van der Waals and electrostatic energies. The energy minimization is performed using standard three-dimensional interpolation methods. Due to the hierarchical multigrid strategy explained above, sufficient accuracy in regions where the ligand and target come into contact is ensured. The three to six lowest energy poses are subjected to a Monte Carlo procedure to examine nearby torsional minima. The minimized poses are then rescored using either SP or XP scoring.

The output of a Glide docking run is a ligand structure output file and a text file containing a table of ranked poses, scores, and score components.

2.3.4.5 The XP scoring function

In a Glide docking calculation, the binding affinities are estimated by the scoring function. Glide, as well as other high-throughput docking applications, uses the "rigid receptor approximation", meaning that the target is treated as a rigid body. The approximation is necessary since the computational time otherwise would be untenable. However, this can give false negatives, compounds that indeed bind to the target experimentally, but for which the docking program cannot identify a high-scoring pose due to the receptor having the wrong conformation. The problem is partly solved by scaling down the van der Waals radii of the nonpolar ligand atoms by 0.8. This does not solve situations, though, where the binding conformation of the protein involves larger movements compared to the protein confirmation used for docking, such as different rotamers of certain active-site side-chains. This has been taken into account in the

visual analysis, where compounds receiving a poor score have been accepted as potential hits if it looks as if the low score is due to steric clashes with a seemingly flexible side-chain.

XP Glide docking begins with a SP dock. Yet, this dock diverts from a regular SP dock by producing a greater diversity of docked structures. The SP dock has to produce at least one properly docked ligand structure for the XP docking to proceed. The XP dock then divides rigid parts of the molecule into a set of "anchors". A better scoring pose is then attempted to be found from each anchor. Various positions of each anchor are chosen, and from these positions individual side chains are grown and then complete molecules are selected by combining high-scoring fragment poses. The candidate structures are minimized, followed by the XP-specific part of the scoring function.

The XP Glide scoring function is made up of the following terms:

XP GlideScore = E_{coul} + E_{vdW} + $E_{\text{hyd_enclosure}}$ + $E_{\text{phobic_pair}}$ + $E_{\text{hb_pair}}$ + $E_{\text{hb_nn_motif}}$ + $E_{\text{hb_cc_motif}}$ + E_{PI} + E_{desolv} + $E_{\text{ligand_strain}}$. Each term is described below.

E_{coul} and E_{vdW} – The Coloumb and van der Waals interaction energies

These energies are calculated according to standard molecular mechanics definitions¹³.

E_{coul} is the contributing interaction energy due to the charges of the atoms. Each charged atom pair contributes with Coloumb energies given by $(4\pi\epsilon_0)^{-1}(q_1 * q_2) * r^{-2}$ where ϵ_0 is the electrical permittivity of space, q_1 and q_2 the charges of the two atoms and r the distance between them. In the calculation of the coulomb interaction energy the net ionic charge of atoms with a formal charge is reduced by approximately 50% to make the interaction energy a better predictor of binding according to empirical data¹⁴.

E_{vdW} gives the van der Waals interacting energy between two atoms, and is given by the formula $4\epsilon * ((\sigma/r)^{12} - (\sigma/r)^6)$, where ϵ is the energy scale (could be seen as the lowest potential energy possible) and σ is the distance of closest possible approach between the two atoms.

$E_{\text{hyd_enclosure}}$ – Hydrophobic interactions (hydrophobic enclosure) energies

The characterization and scoring of hydrophobic atoms and their contributing hydrophobic score are made in several steps. The hydrophobic interacting atoms are divided into "connected groups" according to a set of rules defining lipophilic atoms and connected groups. For each connected group, the group score is calculated as $\sum_j S_j(r) * S_j(a)$, where $S_j(r)$ is a radial energy term and $S_j(a)$ is an angular term. The procedure is described below.

For each lipophilic ligand atom, the closest lipophilic protein atom is selected, and a vector is drawn between the two atoms. This vector is the "anchor" of that ligand atom. When the anchor has been set, vectors are drawn between the lipophilic ligand atoms and all lipophilic protein atoms within a distance of 3 Å plus the sum of the vdW radii of the ligand and the protein atom, approximately 6 Å for regular protein and ligand atoms. Each lipophilic atom within that distance is counted as a lipophilic contact. The radial contact score $S(r)$ is calculated for each lipophilic ligand atom simply by the number of contacts with protein lipophilic atoms. The angle between the anchor vector and all other vectors are calculated. If the angle between the anchor and another vector is below 90 degrees, the lipophilic protein atom is estimated to be on the "same side" as the anchor atom and gives 0 in angular score $S(a)$. If the angle is between 90 and 180 degrees, the angle is given an angular score between 0.6 and 1.0. The energy of a connected group is calculated by the sum of radial scores multiplied by angular scores; $\sum_j S_j(r) * S_j(a)$. The overall score for a group is the sum of all atom scores belonging to that group. The penalty for one single group is maximum 4,5 kcal/mol (even if the sum is higher), due to empirical determination. The total hydrophobic enclosure term $E_{\text{hyd_enclosure}}$ is given by the negative sum of all connected group energies.

$E_{\text{phobic_pair}}$ Hydrophobic atom-atom pair energy term

$E_{\text{phobic_pair}}$ is defined as in ChemScore: the hydrophobic energy term is calculated for all hydrophobic ligand and receptor atoms and is defined as $\sum f(r_{ij})$ (where i and j refer to hydrophilic atoms; chlorine, bromine and iodine atoms which are not ions; sulphurs which are not acceptor or polar types; carbons which are not polar type). For $r_{ij} <$ the sum of the atomic vdW-radii plus 0.5 Å, f is 1.0. Between this value and the sum of the atomic vdW-radii plus 3.0 Å, the function is a linear ramp down to 0.0, and for $r_{ij} >$ the sum of the atomic vdW-radii plus 3.0 Å f is assigned the value 0.0. The term is a quite low-range energy term, representing the displacement of water in hydrophobic regions.

$E_{\text{hb_pair}}$, $E_{\text{hb_nn_motif}}$ and $E_{\text{hb_cc_motif}}$ Hydrogen bond motif and special neutral-neutral and charged-charged hydrogen bond motifs

The $E_{\text{hb_pair}}$ -term is defined as in ChemScore: $\sum g_1(\Delta r) g_2(\Delta \alpha)$ where $g_1(\Delta r) = 1$ if $\Delta r \leq 0.25$ Å; $(1 - (\Delta r - 0.25))$ if $0.25 \text{ Å} < \Delta r \leq 0.65$ Å; 0 if $\Delta r > 0.65$ Å and $g_2(\Delta \alpha) = 1$ if $\Delta \alpha \leq 30^\circ$; $(1 - (\Delta \alpha - 30)/50)$ if $30^\circ < \Delta \alpha \leq 80^\circ$; 0 if $\Delta \alpha > 80^\circ$. Glide XP identifies three different kinds of hydrogen bonds: charged-charged, charged-neutral and neutral-neutral. The default values assigned are 1.0 kcal/mol for neutral-neutral, 0.0 kcal for charged-charged and (as a result of an interpolation of the other two hydrogen-bond-types, experimentally verified) 0.5 kcal/mol for neutral-charged.

When geometry deviates from an ideal hydrogen-bonding term, Glide XP has two special hydrogen bonding motifs with additional increments of binding affinity. These are described for neutral-neutral and charged-charged hydrogen bonding. According to the authors, no motif for neutral-charged hydrogen bond interaction is found and neither defined which is why this type of hydrogen bond is excluded.

$E_{\text{hb_nn_motif}}$ identifies special motifs commonly crucial for ligand binding to an active site, where a hydrogen bond is found in an otherwise hydrophobic region. Both the ligand and the target atoms have to fulfil special binding properties. A bond is determined to be a special neutral-neutral binding motif if the hydrogen bond is considered to be in a hydrophobically constrained environment. An example of a suitable group would be a planar nitrogen in an aromatic ring binding, for example to a protein N-H backbone group in a region surrounded by hydrophobic atoms. Such a hydrogen bond reward is 1.5 kcal/mol.

Pair-correlated hydrogen bonds are defined as a pair of ligand atoms separated by not more than one rotatable bond participating in hydrogen bonds to the target in such a hydrophobic region as described above. Several restrictions of the pair that can be made are defined. If all criteria are fulfilled, the hydrogen bonding pair is given a 3.0 kcal/mol-reward if the sum of the hydrophobic enclosure score for the participating atoms is above cut-off.

$E_{\text{hb_cc_motif}}$

Five different types of special charged-charged hydrogen bond motifs that signal enhanced binding properties have been identified. Examples of such motifs are the number of waters surrounding the protein part of a salt bridge, the number of charged-charged hydrogen bonds made by the charged ligand group and the binding of zwitterion ligands. The five different kinds of special charged-charged hydrogen bond motifs generate rewards from 0.0 to 4.7 kcal/mol.

E_{PI} and some other terms

A number of other terms have been investigated. These include terms rewarding pi-stacking interactions (E_{PI}), rewarding of halogen atoms placed in hydrophobic regions and some empirical corrections rewarding the binding affinity or smaller ligands compared to larger ones. These parameterizations have mostly been made based on limited experimental data and are not fully mature yet. The terms are relatively small compared to other ones.

E_{desolv} – Water scoring; rapid docking of explicit waters

The purpose of this term is to limit the number of false positives by penalizing desolvation inadequacies. To high-scoring poses from the initial XP-run, 2.8 Å spheres (representing water molecules) are added. An inadequately solvated polar or charged ligand or protein group results in desolvation penalties. Additionally, each water molecule is probed to search for unusual hydrophobic contacts. If the number of such contacts exceeds a given limit, a penalty score is given. Another use of the addition of water is to determine whether the special-hydrogen bonding described previously should be assigned.

$E_{\text{ligand_strain}}$ – Contact penalties

The penalizing of strain energy is one of the single most difficult components to calculate in the scoring function. The problem arises since the ligand has to adjust to fit into an imperfect rigid cavity since induced-fit effects are not calculated. Given these limitations of perfection, this penalizing term is made up of two different functions. One counts the intra-molecular heavy atoms contacts below approximately 2.2 Å. The pose is rejected if there are four or more such contacts. The other assembles the contacts into groups and gives a penalty according to the range of contacts, the size of the contacting groups and the position of contacting groups (it is more difficult to penalize peripheral contacting groups than central ones).

2.4 Multiple active site correction (MASC)

When scoring ligand docking, one problem is ligand-dependent scoring properties, which means that the scoring function gives a ligand false good or bad score due to molecular properties of the ligand without having anything to do with the ligand-target interactions. The result is that some ligands often get a good score for a wide variety of targets, independent of if they are active for the target or not¹⁵.

MASC¹⁶ is a simple statistical correction, trying to reduce this scoring problem. In the first step, binding score to the target of interest is calculated. The second step includes calculation of binding scores of each ligand to a set of different targets, MASC-targets. For the most accurate MASC-score, the MASC-targets should have as wide range as possible. The average score from the MASC-targets is calculated for each ligand. MASC-score is then calculated by subtracting the average score from the MASC-targets from the score of the target of interest. The resulting MASC-score for the target of interest is then intended to be corrected from the ligand-dependent part of the scoring, and has been shown to be useful to some extent¹⁷.

2.5 Visual analysis

A visual analysis and judgment of the binding pose makes the selection of compounds more refined. The visual analysis makes it possible to combine personal experience with the computer-based scoring.

A tool in the visual analysis is generation of molecular surfaces as well as hydrophobic and hydrophilic surfaces for the target. This is done using the Maestro application "Generate surfaces". The surfaces can be used to facilitate examination of how the docking poses are positioned in the target active site, and how well the hydrophilic and hydrophobic parts of the docked ligands align with the hydrophilic and hydrophobic parts of the target. The surfaces can be turned on and off, and several surfaces can be displayed at the same time.

The surface generated in Maestro is a Connolly surface¹⁸. The calculation is done by rolling an imaginary ball representing a solvent molecule over the assembly of spheres consisting of the van der Waals radii of the target. The surface is placed where the outside of the "ball" is in contact with the van der Waals radii. The surface that is generated represents the boundary between the solvent-accessible and solvent-free regions of the target.

Hydrophobic and hydrophilic potentials are calculated using a distance-dependent, dielectric formulation analogous to Goodford's GRID algorithm¹⁹. The van der Waals energy together with the direction and magnitude of the electric field are computed for a probe, considered to interact with all atoms of the receptor site within a defined distance, centered at each grid-point in the target receptor. Hydrophobic and hydrophilic "volumes" are then defined by defining an energy-threshold for the hydrophobic and hydrophilic volumes respectively.

In addition to the surfaces, hydrogen bonds between the ligands and target can be displayed. The default criterion for an interaction to be displayed as a hydrogen bond in Maestro is a maximum distance of 2.50 Å, a minimum donor angle of 120° and a minimum acceptor angle of 90°. In the visual analysis it sometimes is obvious that an interaction is a hydrogen bond even though it not is displayed as one, and in these cases the interaction is better considered as a hydrogen bond. It also exist interactions that would be hydrogen bonding if not the target rigid body approximation had been made. This aspect is important to be aware of in the visual analysis, since it affects the ligand position score.

3 Experimental procedure

In the following section, a detailed description of the experimental procedure will be given for each target.

3.1 *Plasmodium falciparum* spermidine synthase (*Pf*-SRM)

3.1.1 Target structure file

The pdb-file used for *Plasmodium falciparum* Spermidine synthetase (*Pf*-SRM) was 2I7C²⁰, consisting of three asymmetric positioned subunits. The PDB-file was edited by removing two of the subunits - the B and C chains, leaving the A chain in 2I7C to be used in the calculations.

3.1.2 First pharmacophore search

Two different pharmacophore hypotheses where created from two different fragments of the ligand structure in 2I7C: *srmhyp1* (six features, hydrogen bond acceptors and donors and one aromatic ring) and *srmhyp2* (two features, one positively ionisable and one hydrophobic feature). To these hypotheses excluded volumes were added. The pharmacophore hypotheses were created in Maestro by 'Application – Phase – Edit Hypothesis'. Three of the features were made obligatory and three optional by manual editing of the *srmhyp1.mask* – file according to the coordinates for the different features found in the *srmhyp1.xyz* file. For *srmhyp2*, both of the features were required and no manual editing of the files was needed. Except for the manual editing mentioned above, default settings were used.

In this phase-run, different rotamer states for each compound were tested since the screened compounds were treated as flexible. The Phase run was in "database search" mode, started from the command line by '\$SCHRODINGER/phase_dbsearch *jobname*'.

The Phase-filtering using the pharmacophore hypothesis *srmhyp1* was meant to be run on four different processors over the network. However, some difficulties occurred when the Phase-filtering should be run. The job stopped after a while and stopped over and over again when being restarted. Unfortunately neither we nor the Schrödinger support were able to find an explanation. The search was finally made possible by splitting the Phase input file into four different files using a Perl-script (*split_phase.pl*). The script copied every fourth line in the input

file to four new files. Each subjob was started separately on the four processors. When a job stopped, the log-file from the subjob was studied, and the processed compounds were deleted from the **_phase.inp*-file. The subjob input-file was copied to another name and all names in that file were changed according to the new jobfile name. The subjob was started again with the new jobfile as input file. The hits were read into a Maestro-project and exported to one single hitfile.

To make the filtering using the pharmacophore hypothesis *srmhyp2* a bit smoother than for the previous filtering, a script-based queue-system was set up. Since the Phase-filtering stopped after processing between 60000 and 300000 compounds, the Phase-run was divided into 200 different subjobs with about 13000 compounds each. Splitting of the subset file into 200 sub-subset files was made using the script *split_phase.pl*. A list-file with the names of the sub-subset files was generated by the command *ls -l subset*_phase.inp > subsetlistname.lst*. Input files corresponding to the subset files were created by running the script *make_phase_inp_files.pl*. To be able to run the script, an input file with the right settings except for the subset file names was needed as input file. The second step in the preparation was to create one list-file for each processor that was going to be used for the Phase-run. The present job was run on two processors. Despite of all this preparation, the job was unexpectedly interrupted since the disc was full, but this problem was easily solved.

3.1.3 Glide SP dock

The hit-compounds from the first pharmacophore search were expanded by Ligprep. Each of the Phase filtering results were separately used as input structures for Glide SP docking runs. The SP dockings were run using grids sized 8x12x14 Å without constraints. The SP-docking was set to save at most 9 poses for each structure, and “twisted non-planar amide bonds” was set as “not allowed”.

3.1.4 Second pharmacophore search

Predicted binding conformations from the SP dock were subjected to a new Phase pharmacophore filtering run. In this run, the docked ligands were set as non-flexible. This Phase-job was run in “file search mode” since the Phase run was screening a file instead of a database. The run was started by the command ‘\$SCRODINGER/phase_fileSearch jobname’. The same two different pharmacophore hypotheses used in the first pharmacophore search were used in this Phase-run as well. Because of the much smaller amount of ligand structures screened by the second pharmacophore search compared to the first one, this Phase-job required much less CPU-time.

3.1.5 Extracting the hit-compounds from the second pharmacophore search

The compounds giving rise to hits in the pharmacophore search should in the next step be scored by XP-docking and MASC. Since the hits from the pharmacophore filtering potentially contained several poses of the same compound a non-redundant (in other words, a file with each compound name only present once) list of compound was needed.

A list of the compounds was made using a couple of Perl scripts. The first script (*sd_extract_names*) used the result-file from the second pharmacophore search as input parameter and produced a list-file (**.lst*) with the poses found in the result poses from the Phase run. The next script (*make_non_redundant_list*) generated a non-redundant list-file from the recently generated list-file. Finally, the compounds present in the non-redundant list-file were

extracted from the the result of the first pharmacophore search, using the script *sd_extract_records_list.pl*.

The extracted compounds were expanded by Ligprep. The expanded result-file was then subjected to the two parallel final scoring calculations by Glide XP and MASC. Each of the scoring calculations was made separately for the hits originating from the two different pharmacophore hypotheses.

3.1.6 Glide XP Dock

The grids used for the Glide XP dock were identical to the grids used for the prior SP docking run with the dimensions 8x12x14 Å, but calculated with "neutralization zone around the ligand".

The XP-dock was intended to yield the 9 top-scoring poses for each compound structure. However just one pose for each compound was added to the result-files. During visual analysis it is important to study different poses of each structure to be able to judge the reliability of the scoring poses. Thus it was decided to make a rough visual selection of the ligand poses followed by a second XP-dock; in this second docking also the compounds chosen from the MASC scoring were included.

3.1.7 MASC

The MASC-scoring of the ligands was made as described in "Material and methods". Grids for the MASC targets were of the dimensions 12x12x12 Å. After being docked into the grids of 217C, the output docking positions were renamed according to the name of the positions in the output file using the script *sd_rename_with_attribute.pl*. When being renamed, the ligands were docked to the MASC target structure grids. In all dockings performed in calculation of the MASC-score, Glide SP-docking was used.

As in the case of the XP-docking result, the visual analyses of the best-ranked ligand poses were quite allowing, since the chosen compounds were about to be subject to a final XP-dock.

3.1.8 Final XP-docking and scoring

The chosen compounds from the rough visual analyses of the XP and MASC-scoring were merged to one single file. A list-file containing all the selected compounds was created. The selected compounds were extracted from the result of the first pharmacophore search by the pearl-script described above. The hit-files from the initial phase-filtering were merged into one single file as well, and the selected compounds from the visual analysis were extracted according to the created list using the pearl-script *sd_extract_records.pl*.

The same grids were used as for the past XP-dock. Maximum 9 poses for each compound structure were saved, giving the possibility to compare different poses of the same ligand structure. The scoring function was set to penalize twisted amide bonds in the docked ligands.

3.1.9 Visual analysis of top-scoring poses

To facilitate the examination of the docking poses, the ligand found in the crystal structure was displayed together with the target structure. By displaying the docking pose in "tubes" and the crystal structure ligand in "wire", the comparison was easily made without risking any confusion between the two structures.

The ligand structure was examined with aspect to internal interactions, repellations and energetically unfavorable poses. Other features taken into account were improbable stereochemistry (such as cis-amide bonds) and to some extent chemical properties of the ligand (for example if it was easily hydrolyzed). The ligand position relating to the target and ligand-target interactions were other aspects of great importance in the visual analysis.

The order of the analysis was as follows: The text file was sorted by increasing glide-score, and the ligand poses was analysed in ranking order. Since the visual analyses leads to selection or rejection of a compound but not a pose, all different poses of the same compound had to be analysed together. Thus, analysis of the compounds was made according to the best ranked pose of each compound.

3.2 *Mt*-DXR

3.2.1 Target structure file

The PDB-structure used for *Mycobacterium tuberculosis* 1-deoxy-d-xylulose 5-phosphate reductoisomerase (*Mt*-DXR), was 2JCZ²¹. The crystal structure consists of two asymmetric positioned subunits. The pdb-file was edited by removing the B-chain, leaving the A-chain to be used in the work.

3.2.1 First pharmacophore search

For *Mt*-DXR, one single pharmacophore hypothesis was used, hereby referred to as *dxrhyp*. The pharmacophore hypothesis was created in Maestro by 'Application – Phase – Edit Hypothesis'. Excluded volumes were added as spatial constraints.

Some changes of the pharmacophore features were made by manual editing of the pharmacophore hypothesis coordinate file *dxrhyp.xyz*. The editing was made by labelling the residues surrounding the actual pattern with their 3D-coordinates in Maestro, and changing the coordinates in *dxrhyp.xyz* according to that information. Editing of the hydrophobic pharmacophore radius was made by setting the radii of the hydrophobic pharmacophore features to 2.5 Å in the *dxrhyp.tol-file*. The two hydrogen bonds were set as required and the two hydrophobic features as optional in *dxrhyp.mask* according to the feature coordinates in *dxrhyp.xyz*. In the input-file for the Phase-job, minimum features to be matched was set to 3, meaning that the two hydrogen bonds and one hydrophobic interaction had to be fulfilled for a compound to be chosen as a hit. In the filtering, the compounds were treated as flexible, thus several rotamer states for each compound structure were screened.

The Phase run was performed using the script-based queue-system described for *Pf*-SRM. The job was run on three processors. After filtering, the three hit-files were merged together.

3.2.2 Glide SP dock

The result from the first pharmacophore search was expanded by Ligprep. The ligand structures were subject to a Glide SP docking run. The set of ligands to be docked in Glide SP was split into four equally sized files. An old SP input file was used as a template for making one input file for each subjob. Each job was locally and manually started for each processor by the command '\$SCHRODINGER/impact -i jobname_*.inp -HOST'. Docked poses involving twisted amide bonds were set as forbidden, and a maximum number of 9 poses per ligand structure was saved for

each structure. The grids used for the SP dock had the core dimensions 14x14x14 Å. The outer dimensions of the box were set making it possible for ligands of 25 Å to dock.

3.2.3 Second pharmacophore search

Ligand structures predicted to bind to the target were subjected to a second pharmacophore search. The same pharmacophore hypothesis was used once again. However, this search was run in "File search mode", with the ligands treated as non-flexible. The second pharmacophore search was run on four processors. The output from the previously performed SP dock, divided in four subjobs, was used as input for the second pharmacophore search. The Phase runs were set up by using an old fileSearch input file as a template. The file was manually edited according to the input- and pharmacophore hypothesis file names. The filtering took a long time since the number of ligands in the four files was altogether about 700000.

The result of the second pharmacophore search was used as template for extracting the hits from the result of the first pharmacophore search in the same procedure as described in "3.1.5 Extracting the hit-compounds from the second pharmacophore search". For a description of the procedure, see this section above

3.2.4 Glide XP Dock

The result from the second pharmacophore search was used as input in a final docking run, using Glide XP. The result from the four subsearches in the second pharmacophore search was expanded by ligprep. The four expanded sets of ligand structures were used as input for four parallel Glide XP Docking runs performed separately on one processor each.

The output of the four subjobs were merged by running the Schrödinger utility Glidesort from the command line; *\$SCRODINGER/utilities/glidesort -o outfile.mae -r outfile.rept -nonrecep infile1.mae infile2.mae infileN.mae*.

3.2.5 MASC

The MASC scoring was performed as described in "Material and methods" and the MASC section of Pf-SRM.

3.2.6 Final visual analysis

Visual analysis was performed on the compounds resulting in good-scoring poses in the MASC and the XP scoring. The procedure of the visual analysis was the same as described in the section "Final visual analysis" for Pf-srm.

4 Results

4.1 Targets

Targets to be used in a virtual screening approach to find potential inhibitors of infectious diseases have to fulfill several properties. They have to be central in the metabolism of the organism causing the disease, and differ enough from any central metabolic pathway in the host organism not to cause the host too much harm. To be able to use the target in virtual screening and docking, the structure has to be solved at a fairly high resolution.

4.1.1 *Pf*-SRM

Spermidine, a polyamine present in all living organisms, is involved in many cellular processes. Inhibition of spermidine synthesis correlates with blocking of cell growth and is therefore an attractive target for both cancer therapy and parasite infection treatment. The spermidine synthesis differs between mammals and *Plasmodium falciparum*²² and is therefore a very interesting pathway to target in the search for new anti-malarial drugs.

A known inhibitor of *Pf* spermidine synthetase (*Pf*-SRM), the enzyme responsible for the final step in spermidine synthesis in *Pf* is shown to block parasite development²³. However, little research has been done on this novel target of anti-malarial drugs. The structure of *Pf*-SRM used in this work has the Protein Data Bank code 217C²⁴, where the protein is crystallized in complex with adoDATO (S-adenosyl-1,8-diamino-3-thiooctane), a transition state analogue containing both substrate and product moieties (*fig 1*). The binding pocket of 217C, with adoDATO drawn in tubes, is shown in *fig 2*

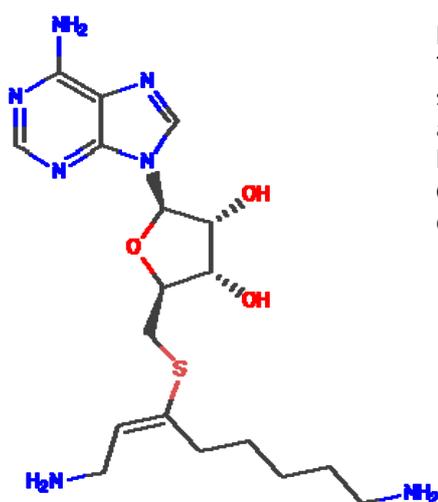


Figure 1
The 2-dimensional structure of adoDATO, the ligand in the crystal structure of 217C.

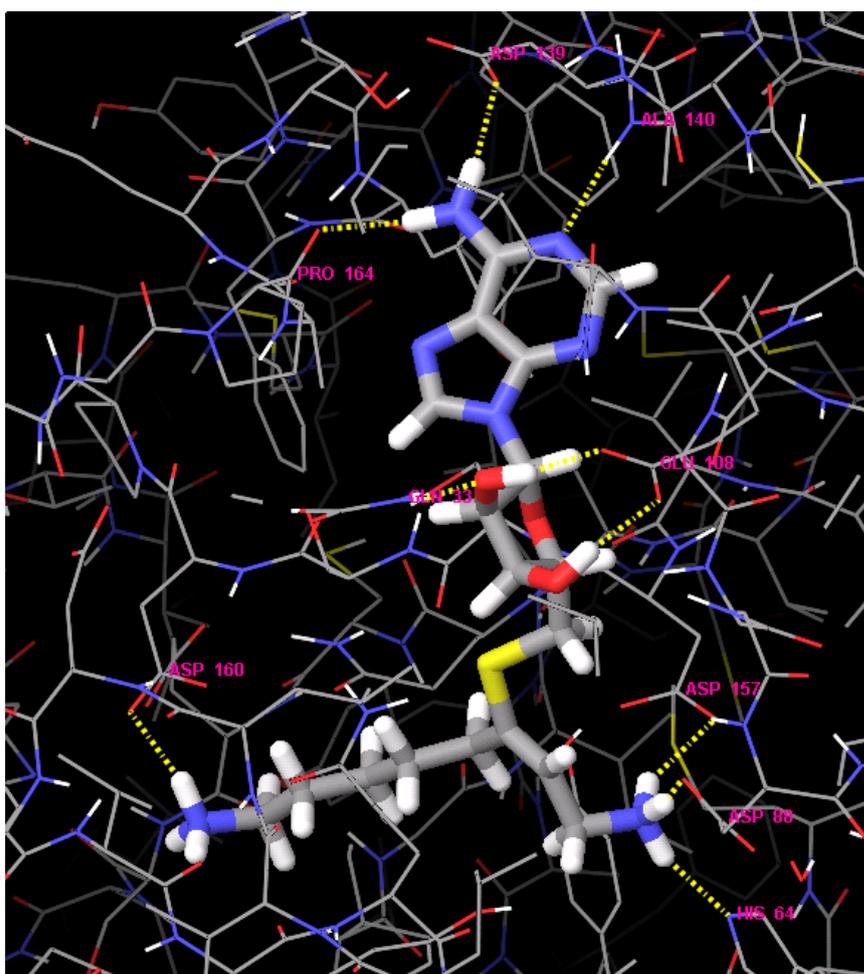


Figure 2
The crystal structure of the binding pocket in 2I7C. The hydrogen bonds between the aa-residues in the protein and the ligand (adoDATO) are marked with yellow dotted lines. The aa-residues participating in the hydrogen bonds are labelled with name and number in pink characters.

4.1.2 *Mt*-DXR

Isopentenyl diphosphate is a precursor of various isoprenoids, essential in all living organisms. In mammals this diphosphate is produced in the mevalonate pathway. In protozoa, plants and many bacteria, including *Mt*, it is produced by a different route; the 2-*C*-methylerythritol 4-phosphate (MEP) pathway. The second step in this pathway is a NADPH-dependent rearrangement and reduction of 1-deoxy-D-xylulose 5-phosphate by the enzyme 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR). Since the enzyme is absent in humans, it is an interesting drug target in several infectious protozoa and bacteria where *Mycobacterium tuberculosis* is a very good example. The essential function of DXR in several organisms including *Mt* has been demonstrated^{17,25}, and the substance fosfidomycin (FOM) has been shown to specifically inhibit the protein²⁶. Thus, the protein is an interesting target for novel drugs against tuberculosis.

The structure of *Mt*-DXR has recently been solved, and it has the PDB-code 2JCZ. In the structure, *Mt*-DXR is crystallized in complex with the known inhibitor FOM²⁷ (3-[formyl(hydroxy)amino]propylphosphonic acid, C₄H₁₀NO₅), manganese (Mn²⁺) and the cofactor NADPH (dihydro-nicotinamide-adenine-dinucleotide-phosphate, C₂₁H₃₀N₇O₁₇P₃) (*fig 3*). The binding pocket of 2JCZ, with FOM, NADPH and the manganese ion drawn in tubes, is shown in *fig 4*.

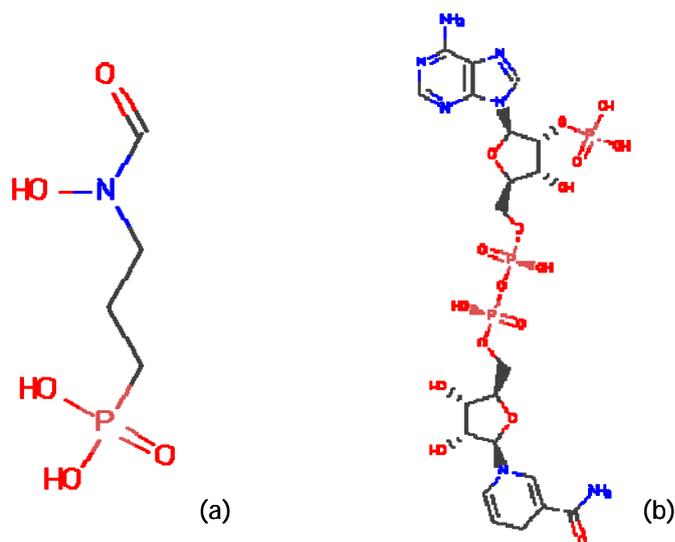


Figure 3
The ligands co-crystallized with DXR in 2JCZ; FOM (a) and NADPH (b)

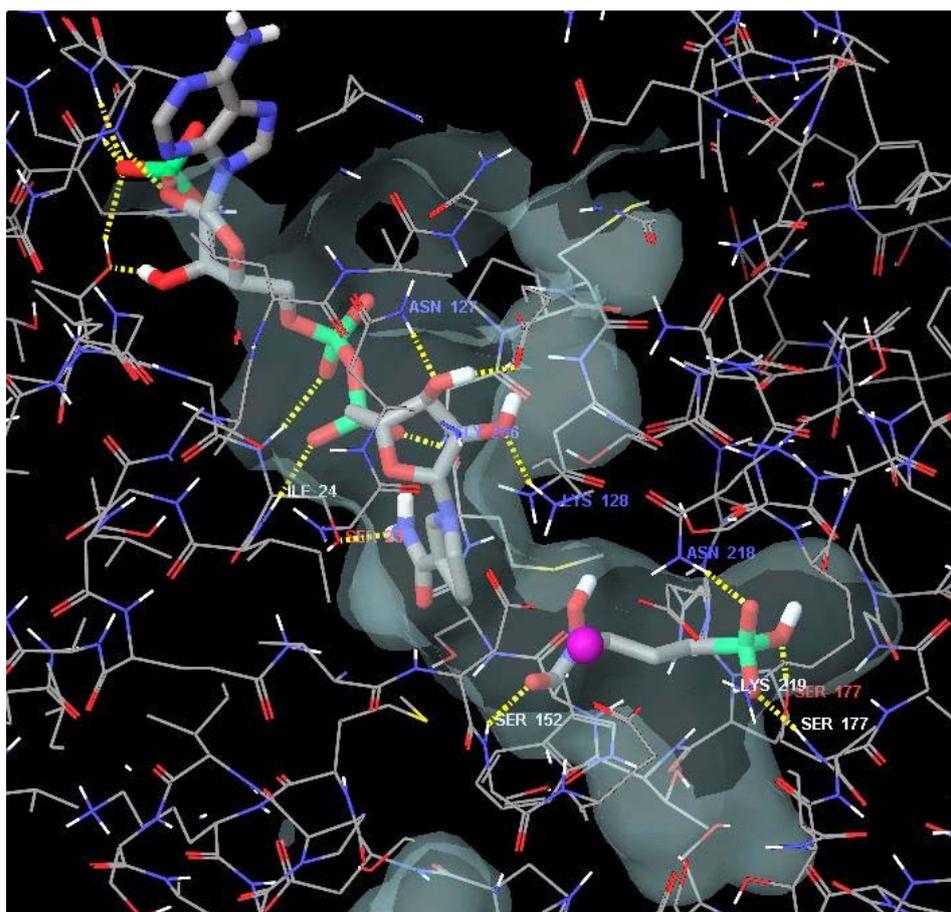


Figure 4
The crystal structure of the binding pocket in 2JCZ. The hydrogen bonds between the aa-residues in the protein and the ligand and cofactor are marked with yellow dotted lines, and the name and number of some interacting aa-residues are given in the figure. The manganese-ion is drawn as a magenta-coloured sphere. The surface of the protein, meaning the border between solvent-exposed and non-solvent exposed areas, is shown as a grey shadow.

4.2 MASC

The theory of multiple active site correction (MASC) is described in Material and Methods above. Targets used for calculation of ligand MASC-scores were the same as the targets used in Jacobsson 2006: 1EVE, 1G49, 1IQE, 1KE7, 1L7F, 1PXX, 1RX3 and 3ERD. A summary of the eight targets is given in *fig 5*.

PDB Code	1EVE	1G49	1IQE	1KE7
Common name	Acetylcholinesterase	Matrix metalloproteinase 3	Human coagulation factor xa	Cyclin-dependent kinase 2
Classification	Serine hydrolase	Hydrolase	Hydrolase	Transferase
Source	<i>Torpedo californica</i>	<i>Homo sapiens</i>	<i>Homo sapiens</i>	<i>Homo sapiens</i>
Resolution	2.50	1.90	2.90	2.00
Crystal ligand	-	carboxylic acid based inhibitor	-	LS3
Primary citation	Kryger, G (1999)	Natchus, M. G. (2000)	Matsusue, T. (tbp)	Bramson, H. N. (2001)
PDB Code	1L7F	1PXX	1RX3	3ERD
Common name	Influenza virus neuraminidase	Cyclooxygenase 2	Dihydrofolate reductase	Human estragon receptor alpha
Classification	Hydrolase	Oxidoreductase	Oxidoreductase	Nuclear receptor
Source	Avian influenza virus	<i>Mus musculus</i>	<i>Escherichia coli</i>	<i>Homo sapiens</i>
Resolution	1.80	2.90	2.20	2.03
Crystal ligand	BCZ	DIF	NAP, MTX	DES
Primary citation	Smith, B. J. (2002)	Rowlinson, S.W. (2003)	Sawaya, M. R. (1997)	Shiau, A. K. (2008)

Figure 5 The eight different targets used in the calculation of MASC-scores. A brief summary of the characteristics of the targets is given in the figure together with the PDB-codes of the crystal structures.

In the MASC-scoring, each of the ligands were first SP-docked to 217C and 2JCZ respectively, and then to the eight targets given above. The MASC-score was calculated by subtracting the average score from the MASC-target dockings from the score of the docking to 217C/2JCZ.

4.3 Virtual screening

As described in the "Experimental procedure", the virtual screening was performed separately for *Pf*-SRM and *Mt*-DXR. Thus, the result from the different stages of the virtual screening process will be described separately for each target. The results are described in the order given for each target in the section "Experimental procedure".

4.3.1 *Pf*-SRM

The crystal structure 217C was examined to find features probable to be crucial for ligand-target interaction. The 3D-structure was aligned with other spermidine synthetase structures to find conserved features in the ligand-target interaction. The crystal structures compared to 217C, and the species of origin, were: 2HTE (*Plasmodium falciparum*), 1XJ5 (*Arabidopsis thaliana*), 1ZDZ (*Homo sapiens*), 1JQ3 (*Thermotoga maritima*) and 2B2C (*Caenorhabditis elegans*).

The binding pocket was divided into five different regions according to the identified binding properties of each region. The regions were named as follows: adenine-part, sugar site, active site, dc-pocket and substrate pocket (*fig 6*).

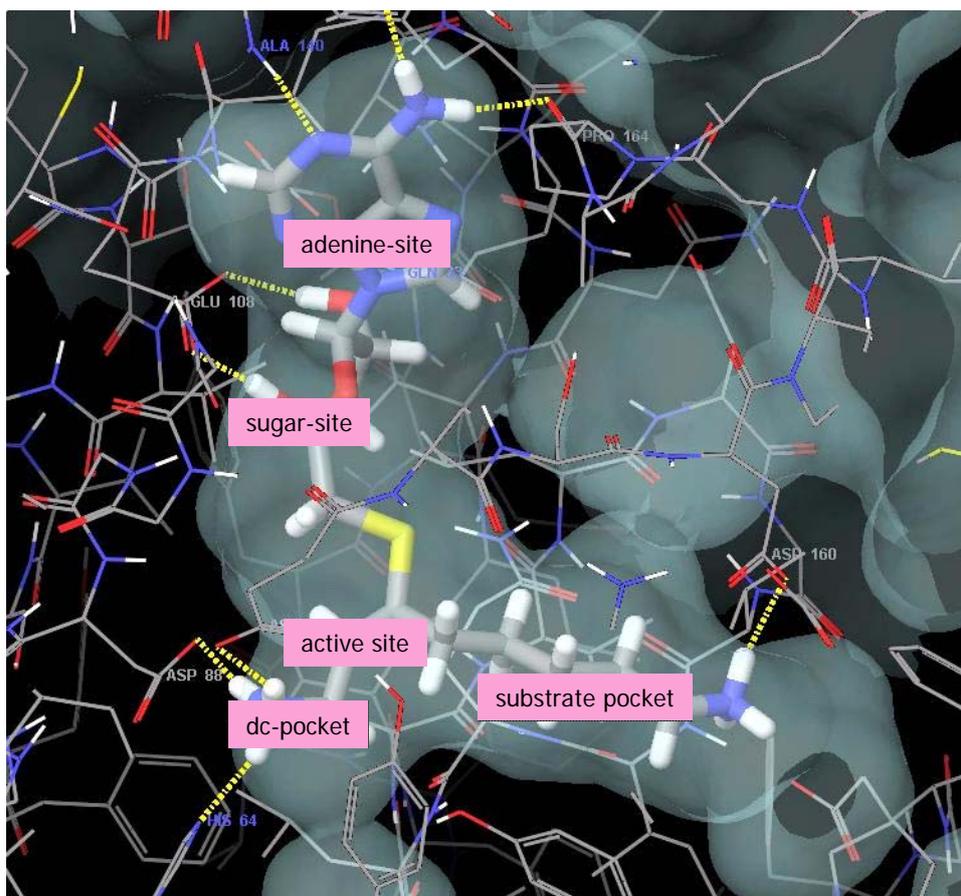


Figure 6
The five different parts of the adoDATO-binding pocket identified according to the binding features and the properties in the target function. The names of the different parts of the binding pocket are shown in the figure. The surface of the protein, meaning the border between solvent-exposed and non-solvent exposed areas, is shown as a grey shadow.

According to the ligand-target interactions identified as crucial in ligand binding, two sets of pharmacophore hypothesis^a were made:

Srmhyp1: The hypothesis focused on the upper part of the binding pocket, i.e. the adenine-part and sugar site. In this hypothesis, the aromatic feature and one of the hydrogen bond acceptors in the adenine-part as well as one of the hydrogen bond donors in the sugar site was set as required, while one hydrogen bond donor in the adenine part as well as one hydrogen bond donor and one acceptor in the sugar site were set as optional. The six pharmacophore features in *srmhyp1* are visualized in *fig 7*. To give a hit in the pharmacophore search, four of the six pharmacophore features had to be fulfilled. The search was narrowed with spatial restraints, since "excluded volumes" were added to the pharmacophore hypothesis (*fig 8*).

Srmhyp2: The other pharmacophore hypothesis focused on ligand-target interaction in the "substrate pocket". The hypothesis required two ligand features to give a hit: a positive charge in the end of the pocket, occupied with a NH_3^+ -group in the crystal structure and a hydrophobic part in the middle of the pocket, in the crystal structure occupied with a hydrophobic carbon-chain (*fig 9*). "Excluded volumes" were added as spatial restraints to narrow the pharmacophore search (*fig 10*).

^a See "Material and methods" and "Experimental procedure" for a more detailed description of the theory and experimental approach of a database screen with a pharmacophore hypothesis.

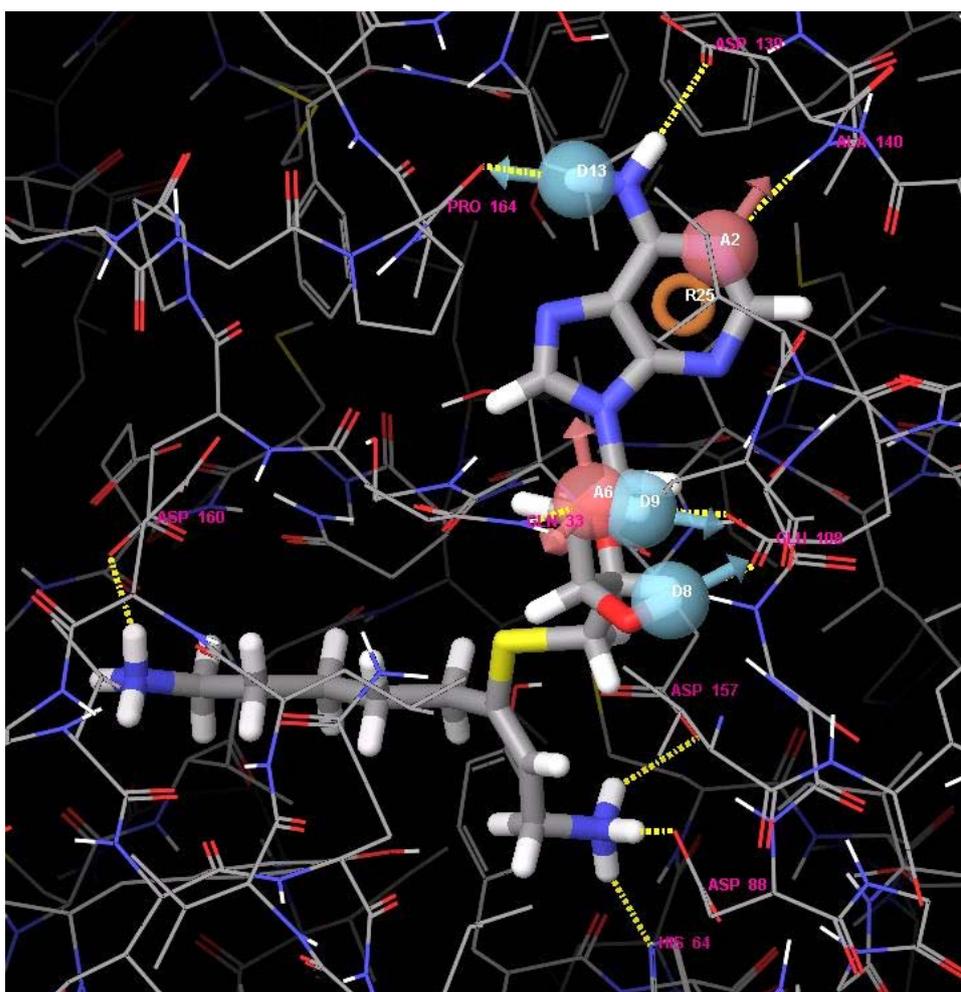


Figure 7
The pharmacophore hypothesis *srmhyp1*, using six different target – ligand interaction features: 3 hydrogen bond donors (blue with label 'D'), 2 hydrogen bond acceptors (red with label 'A') and one aromatic feature (yellow circle with label 'R').

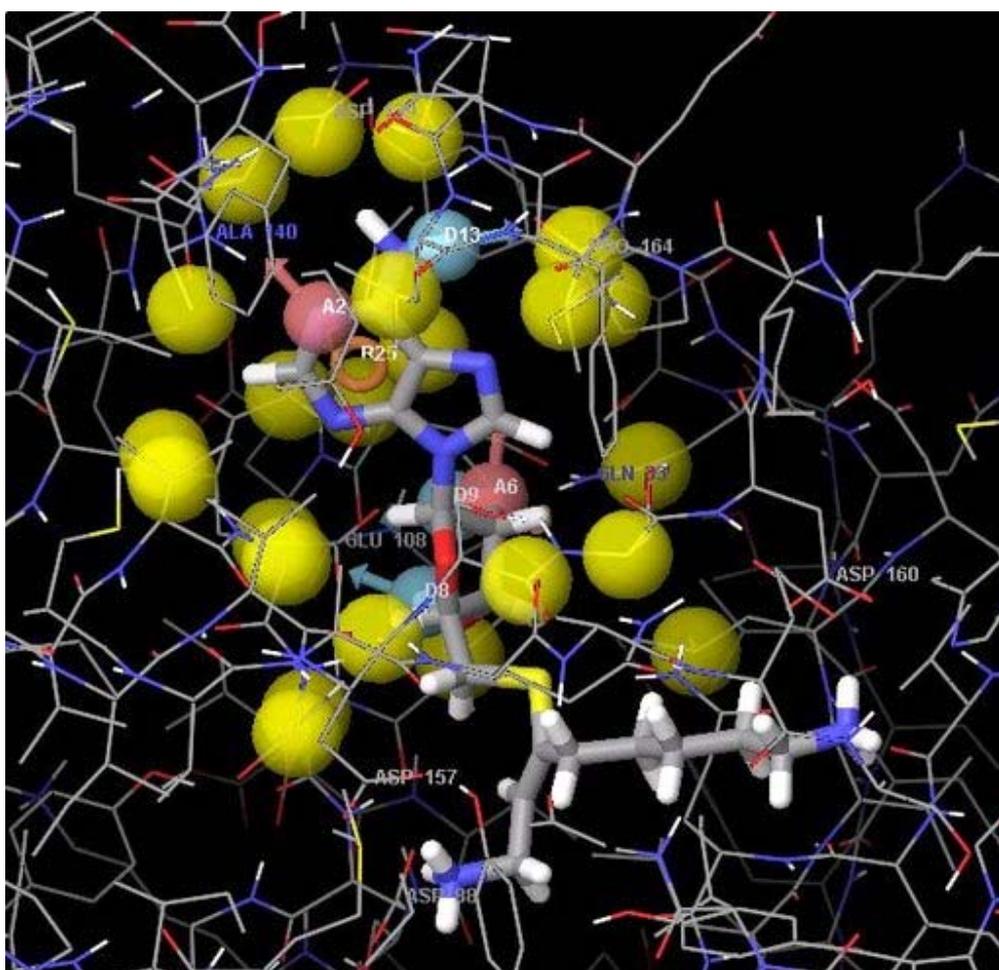


Figure 8 "Excluded volumes", displayed as yellow spheres, complementing the pharmacophore hypothesis *srmhyp1* as spatial restraints.

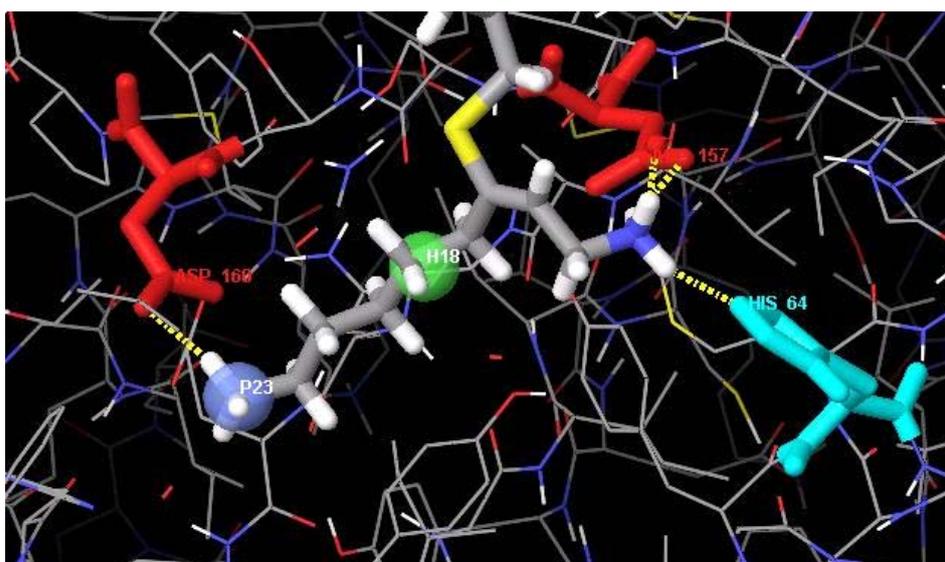


Figure 9
The pharmacophore hypothesis *srmhyp2*, located in the substrate-pocket. The two features is one positive charge (blue sphere labelled 'P') and one hydrophobic feature (green sphere labelled 'H'). The aa-residues interacting with the ligand are drawn in tubes with aspartic acid coloured red and histidine coloured cyan.

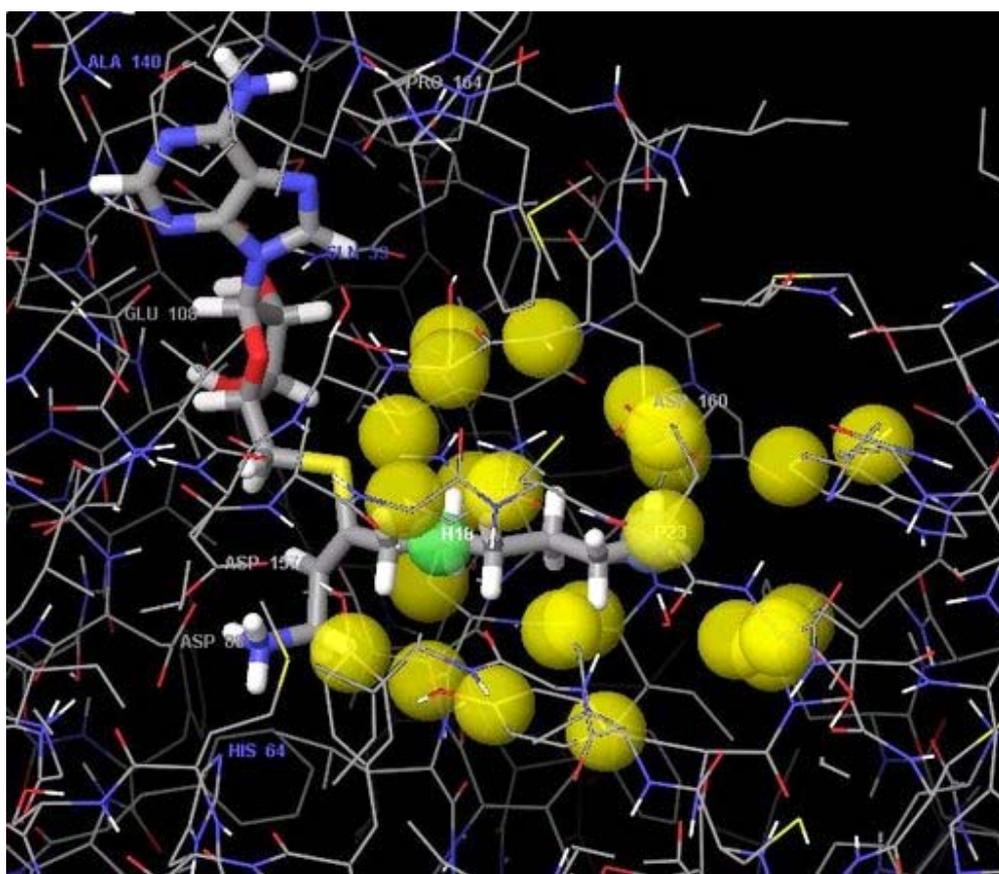


Figure 10 "Excluded volumes", displayed as yellow spheres, framing the pharmacophore features in *srmhyp2*.

The pharmacophore searches were performed as described in "Experimental procedure", resulting in 5489 hits for *srmhyp1* and 1866 hits for *srmhyp2*.

Grids for the Glide docking were calculated as described in "Material and methods". The size of the grid box was set by studying the binding pocket and the region potentially accessible for ligand binding. In the examination, a graphic representation of the molecular surface of the binding pocket was used (see "Material and methods" for a closer description). The examination resulted in the generation of grids with the dimensions 8x12x14 Å. No restraints were added to the grids. To test the reliability of the generated grids, a Glide SP dock was performed with 983 randomly selected "decoy-ligands" and five known binders: adoDATO, MTA (5'-methylthioadenosine, the original ligand for the target), APE (5-amino-1-pentene, a pentene with a nitrogen on the end), 4MCHA (*trans*-4-methylcyclohexylamine, a cyclohexane with a nitrogen in the 1' position and a methyl group in the 4' position) and CHA (cyclohexylamine, a cyclohexene with a nitrogen in the 1' position). For adoDATO and MTA, structural information of the binding to PfSRM was available. These two ligands were docked correctly. The other three ligands – known binders but with no information of their orientation in the binding stage - were docked in a likely position, but given a lower score than the ones taken from the crystal structures. The lower score is probably due to the small size of these three ligands.

To limit the result from the pharmacophore search and to eliminate false positives, a two-step screening approach was performed as follows:

The hits from the first pharmacophore searches were extended by LigPrep to all possible combinations of ionization states, tautomers and stereo-isomers. The resulting structures were docked to the target structure using Glide SP, storing 9 poses per structure. All resulting poses were once again filtered using *srmhyp1* and *srmhyp2* respectively. In this screen, the compounds were kept rigid. Due to a limitation of the software the docked poses were however allowed to translate and rotate as rigid bodies. The procedure is described more in detail in "Experimental procedure" in the sections "Glide SP dock" and "Second pharmacophore search". The intention with the Glide SP-docking and the following pharmacophore filtering step was to identify compounds that were predicted by Glide to bind to the target, without constraints, in a manner consistent with the pharmacophore hypothesis in Phase used to select them in the first place.

The procedure resulted in 1472 hit compounds from the search with the pharmacophore hypothesis *srmhyp1* and 394 hit compounds from the screen with pharmacophore hypothesis *srmhyp2*.

Glide XP-dock^b was performed with the compounds not being sorted out by the procedure described above. For the XP-docking, grids with the same dimensions as for the previous Glide SP-dock were used. The same compounds were ranked by MASC-scoring. For a description of the Glide XP-docking and MASC-scoring, see "Experimental procedure".

The best-scoring ligand poses were visually examined, to find ligand-target interactions considered probable to result in ligand binding. The visual analysis was quite allowing, meaning that good-scoring compounds with no obvious bad binding properties such as sterical clashes, mismatched charges or hydrophilic ligand parts in hydrophobic target regions or vice versa, were selected.

169 compounds were selected from the Glide XP-dock, and 26 compounds from the MASC-scoring.

The selected ligands were applied to a second Glide XP-dock, using the same grids as for the first XP-dock. This second XP-dock was the final computational step, and the visual analysis of this docking result was intended to result in the final selection of 30 compounds probable to bind to the target.

The best-scoring docking compounds were carefully studied. Hydrophilic and hydrophobic interactions and clashes, steric clashes and matches and/or miss-matches of charges were considered. All the docking poses were compared to the ligand-target interactions in the crystal structure (*fig 11*). Some examples of ligand docking poses estimated as good in comparison with the crystal structure are given in (*fig 12 - fig15*). The course of action in the visual analysis is described and explained in detail in "Material and methods" and "Experimental procedure".

The chemical properties of the compounds were also considered, rejecting compounds with no potential to act as an in-vivo inhibitor due to for example instability. However this criterion wasn't thoroughly applied, since the aim of this work was only to find potential binders to the target, with no consideration of the chemical properties of the ligands. Yet, some compounds obviously unsuitable as potential drugs were rejected. Some of the compounds were identical to other compounds found in the hit-list. In these cases, just one of the compounds was selected.

^b See "Material and methods" for a description of Glide docking, and a more detailed review of the Glide XP scoring function.

Finally, 30 compounds hypothesized to interact with the target and inhibit target function were selected. The names of the compounds as well as the 2-dimensional structures are found in (fig 16).

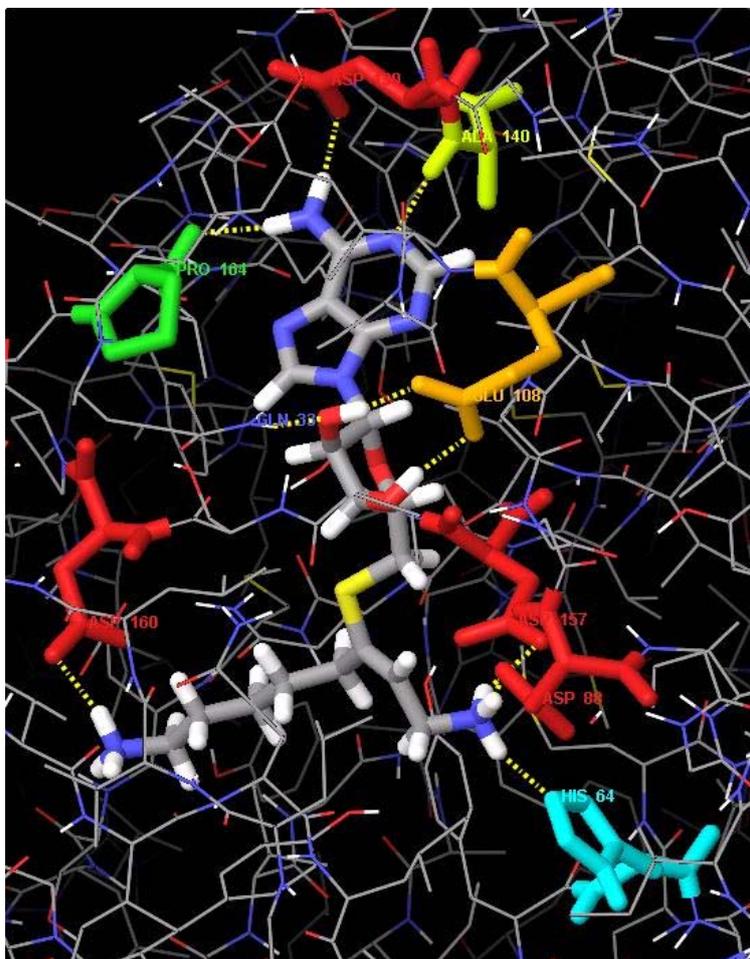
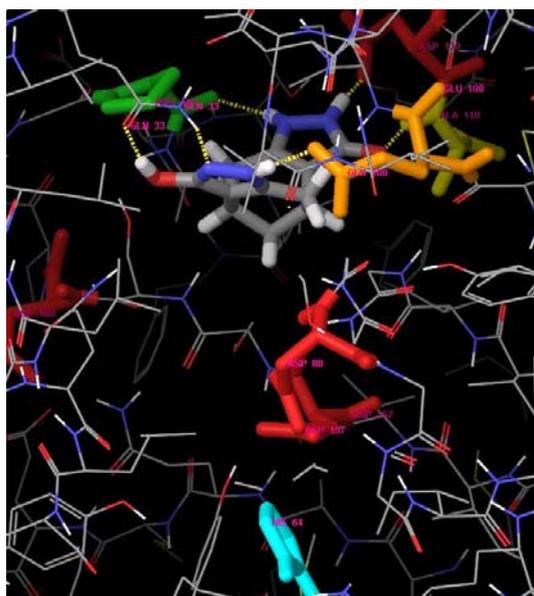
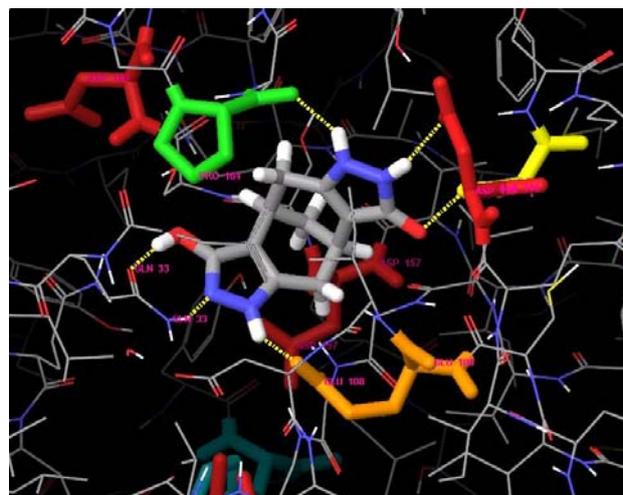


Figure 11
The binding of adoDATO in the crystal structure. The aa-residues interacting in ligand-target hydrogen bonds are drawn in tubes and coloured red (asp), green (pro), cyan (his), yellow (ala) and orange (glu).



(a)



(b)

Figure 12. Two different views of the docking of compound 4858005. The ligand is placed as a lid in the adenine-site. Hydrogen bonds stabilize the ligand binding all around the molecule.

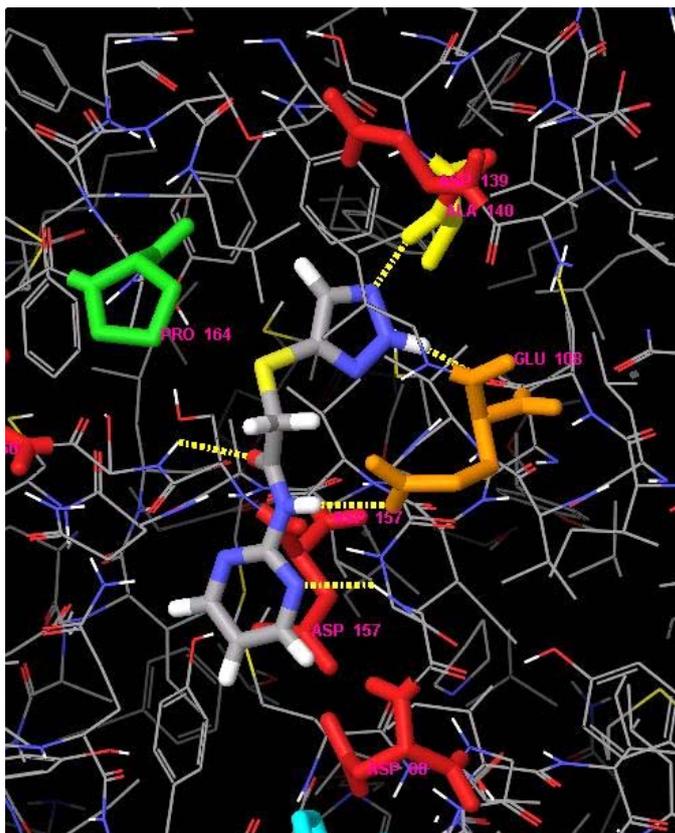


Figure 13

Binding of compound 5063695. The ligand stretches from adenine-site down to active site. Several hydrogen bonds stabilize the ligand binding throughout the molecule, both hydrogen bonds found in the crystal structure binding of adoDATO as well as interaction with aa-residues not participating in hydrogen bonding in the crystal structure.

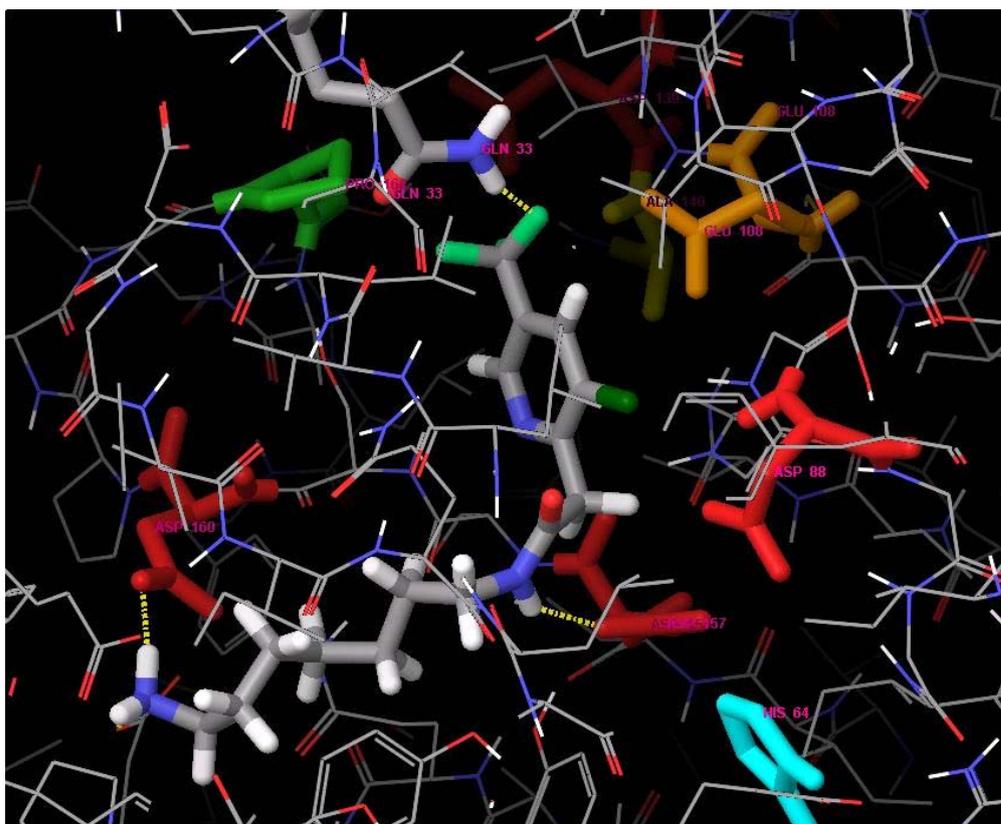


Figure 14
Compound 6630557 occupies the sugar-site, the active site and the substrate pocket in the docked position shown in the picture. Hydrogen bonds present in the crystal structure as well as hydrophobic patterns are fulfilled.

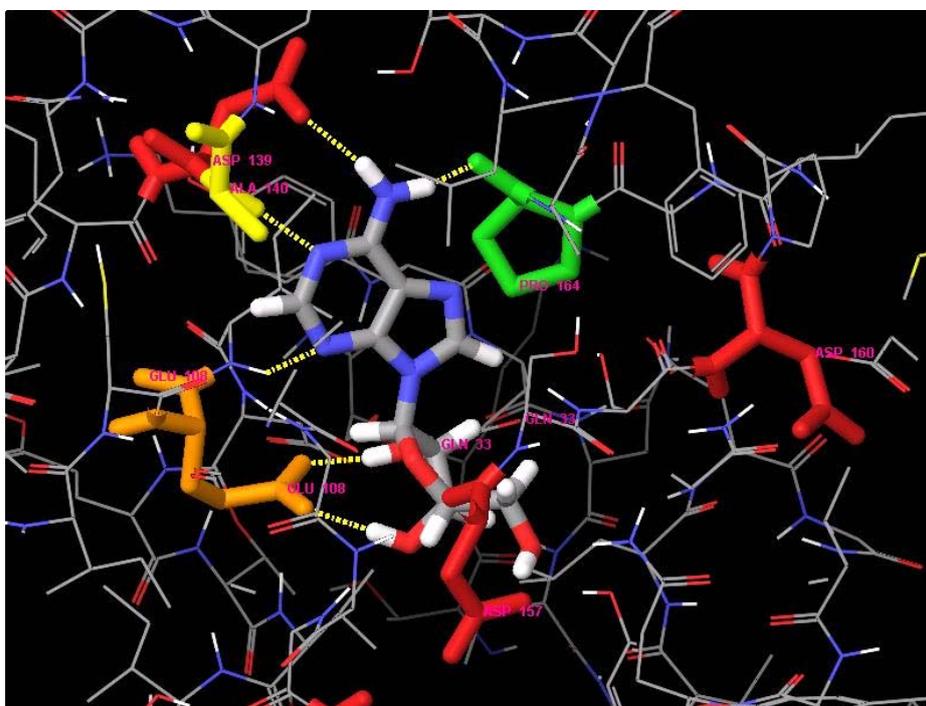


Figure 15
Compound 6704121 is adenosine, i.e. the adenine- and sugar part of adoDATO. The compound is docked to the adenine- and sugar site, and hydrogen-bonds in the same way as adoDATO.

Compound	Chemistry	Compound	Chemistry	Compound	Chemistry	Compound	Chemistry
6687756		5256636		5816377		5805272	
6630557		5051850		7045547		6047283	
6232475		5752042		4857087		5397340	
6232435		6235670		4895899		7072536/ 5031951	
4997359		5063695/ 6257137		6258946		5886166	
5267831		5467248		5797946		4939835	
5900218		4858005		5999229		6704121/ 6889981	
		7241768		6823081			

Figure 16 The selected compounds hypothesized to interact with Pf-SRM.

4.3.2 *Mt-DXR*

The crystal structure 2JCZ was examined to find features probable to be crucial for ligand-target interaction. The 3D-structure was aligned with other DXR-structures to find conserved features in the ligand-target interaction. The crystal structures compared to 2JCZ, and the species of origin, were: 1Q0Q (*Escherichia coli*), 1JVS (*Escherichia coli*) and 2C82 (*Mycobacterium tuberculosis*).

The binding pocket was divided into five different regions according to the identified binding properties of each region. The regions were named as follows: nucleoside site, diphosphate site, sugar site, nicotinamide site, fom-mn region and fom-p region (*fig 17*).

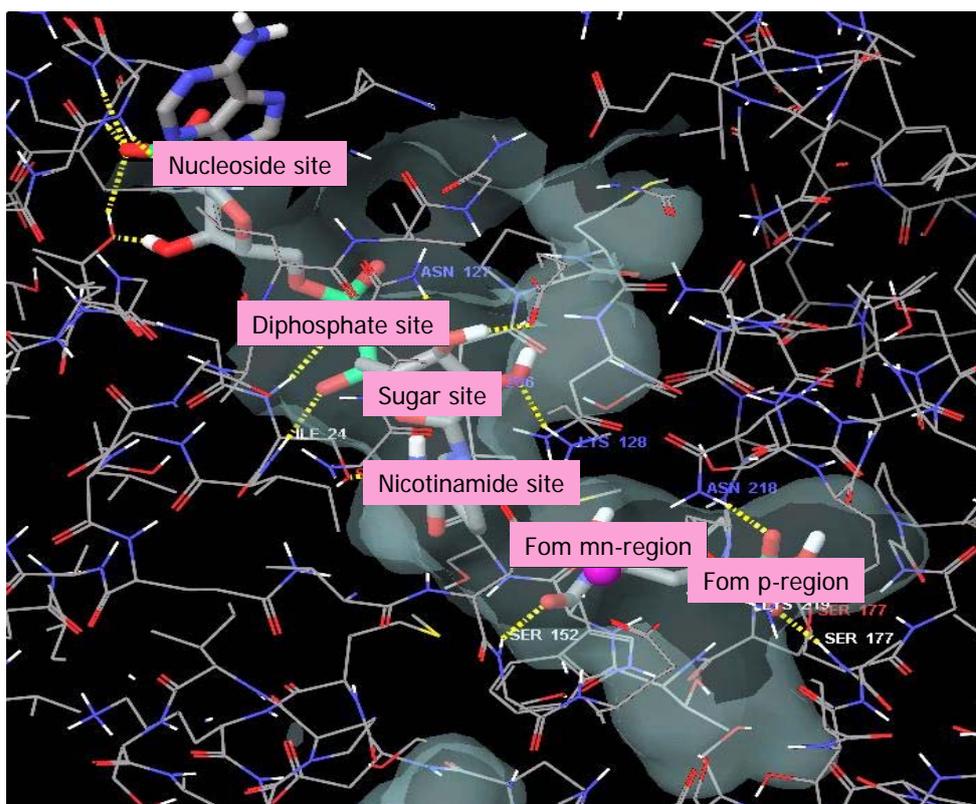


Figure 17

The different parts of the binding pocket in DXR identified according to the ligand-target binding features. The ligand (FOM) and co-ligand (NADPH) in the crystal structure are drawn in tubes. The given names of the different parts of the binding pocket are shown in pink boxes. The surface of the protein, meaning the border between solvent-exposed and non-solvent exposed areas, is shown as a grey shadow.

Due to the absence of structural information of target binding properties, potential ligand-target interactions of importance in ligand binding were examined by docking known binder to the binding pocket in the presence of NADPH, using Glide SP dock. Most of the known binders used in the dockings were taken from the article "AFMoC Enhances Predictivity of 3D QSAR: A Case Study with DOXP-reductoisomerase"²⁸. The ligands used from this article were FOM, FR900098, KL562, LD_1_1, LD_2_6, LD_2_7, LD_2_8, LIHP_88, LIHP_83, LIHP_84, LTB65, LTB66, TK54, UK213, UK513, UK613, UK621. In addition, the two ligands IMB and CBQ were fetched from the article "Crystallographic Structures of Two Bisphosphonate:1-Deoxyxylulose-5-Phosphate Reductoisomerase Complexes"²⁹. The original substrate, DXP, was included as well. The 2D-structures of the ligands are shown in (fig 18). The 3D-structure of the ligands were built in Maestro and expanded by LigPrep to all possible combinations of ionization states, tautomers and stereo-isomers. The resulting ligand structures were docked to the target structure using Glide SP using grids with the dimensions 14x14x14 Å, without restraints.

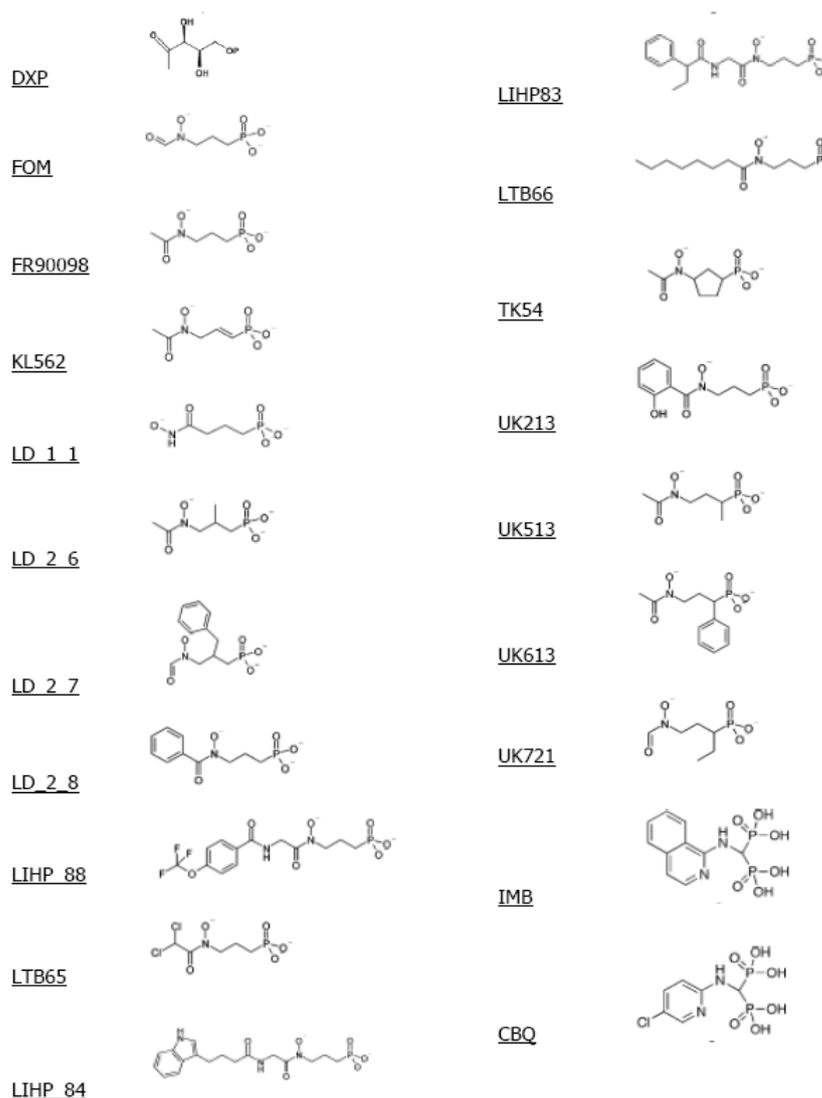


Figure 18
The 2D-structure of known binders used for docking to the target in the examination of *Mt*-DXR ligand-target interactions

Most of the ligands were docked in an expected manner, and the interactions shown were used in the creation of a pharmacophore hypothesis. The ligand-target interactions resulting from the dockings were carefully studied. A graphical visualization of hydrophobic and hydrophilic surfaces (see "Visual analysis" in "Material and methods") was used for examining the probability of the ligand positions. In the hypothesis, features in the mn-fom region, nicotinamide site and sugar site were considered as most important for ligand binding. Another factor found to be of great importance was the manganese ion between the mn-fom and nicotinamide site, which was left in the target structure in the further calculations. The created pharmacophore hypothesis, *dxrhyp*, contained four features: two hydrogen bond acceptors in the mn-fom place, and one hydrophobic feature each in the nicotinamide site and sugar site (*fig 19*). For a more detailed description of the creation of *dxrhyp*, see "Experimental procedure".

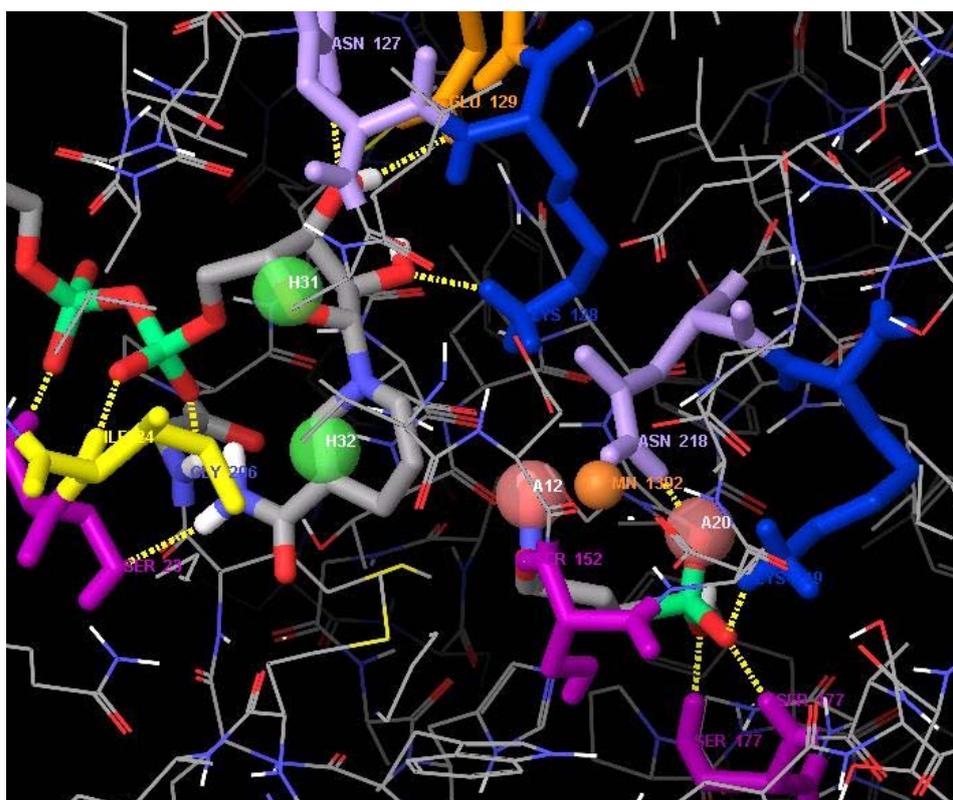


Figure 19
The pharmacophore hypothesis *dxrhyp*. The hypothesis contains two hydrophobic features where NADPH is to be found in the crystal structure (green spheres labeled H) and two hydrogen bond acceptor features in the FOM-region (red spheres labelled A). NADPH and FOM are drawn in tubes, and aa-residues interacting with the ligands are drawn in tubes and coloured according to the aa-residue properties: lys (blue), asn (light purple), ser (magenta), ile (yellow) and glu (orange)

In the pharmacophore hypothesis filtering, the radius of the hydrophobic features was increased from the default value of 1.7 Å to 2.5 Å. The two hydrogen bond acceptor features were set as required and the two pharmacophore features as optional. A hit had to fulfill three of the four pharmacophore features i.e. two hydrogen bond acceptors and one hydrophobic feature. Excluded volumes were added to the pharmacophore hypothesis as spatial restraints (*fig 20*)

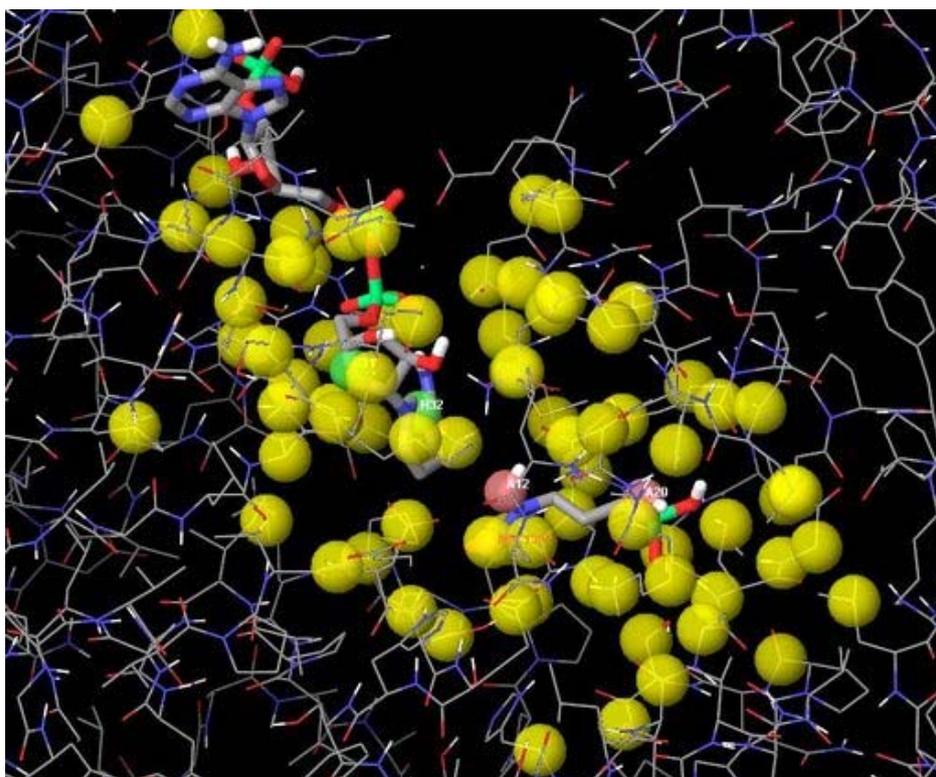


Figure 20
 "Excluded volumes", displayed as yellow spheres, were added to the pharmacophore hypothesis *dxrhyp*, acting as spatial restraints in the search for ligands suiting the hypothesis.

The pharmacophore hypothesis filtering was performed as described in "Experimental procedure" and resulted in 67 068 hits.

Grids for the Glide docking were calculated as described in "Material and methods". The size of the grid box was set by studying the binding pocket and the region potentially accessible for ligand binding. In the examination, a graphic representation of the molecular surface in the binding pocket was used (see "Material and methods" for a description of what a molecular surface is and how it can be used). The examination resulted in the generation of grids with 14x14x14 Å, and set the length of docked ligands to maximum 25 Å.

The same two-step procedure as described for *Pf*-SRM was used to limit the result from the pharmacophore search and get rid of false positives:

The hits from the first pharmacophore searches were expanded by LigPrep to all possible combinations of ionization states, tautomers and stereo-isomers. The resulting structures were docked to the target structure using Glide SP, storing 9 poses per structure. All resulting poses were once again filtered *dxrhyp*, with the compounds kept rigid but being allowed to translate and rotate as rigid bodies. The procedure is described more in detail in "Experimental procedure" in the sections "Glide SP dock" and "Second pharmacophore search".

The procedure yielded 2144 compounds. These were expanded by LigPrep to all possible combinations of ionization states, tautomers and stereo-isomers and docked to the target using Glide XP and grids of the same dimensions as the former Glide dock. The same expanded ligands were ranked by MASC-scoring. The theory and methodology for Glide XP scoring as well as MASC-scoring is found in "Material and methods" and "Experimental procedure".

The best-scoring docking positions were carefully studied. Hydrophilic and hydrophobic interactions and clashes, steric clashes and matches and/or mis-matches of charges were considered. All the docking positions were compared to the ligand-target interactions in the crystal structure (*fig 21*). Some examples of ligand docking poses estimated as good in comparison with the crystal structure are given in (*fig 22 – fig 24*). The procedure in the visual analysis is described and explained in detail in “Material and methods” and “Experimental procedure”.

Although the chemical properties of the compounds were not intended to be part of the selection-criterion, some ligands were rejected due to their chemical properties. For example compounds with high probability to hydrolyse rapidly were not accepted.

The final visual analysis resulted in 26 compounds hypothesized to interact with the target and inhibit the target function. The names of the compounds as well as the 2-dimensional structures are given in (*fig 25*).

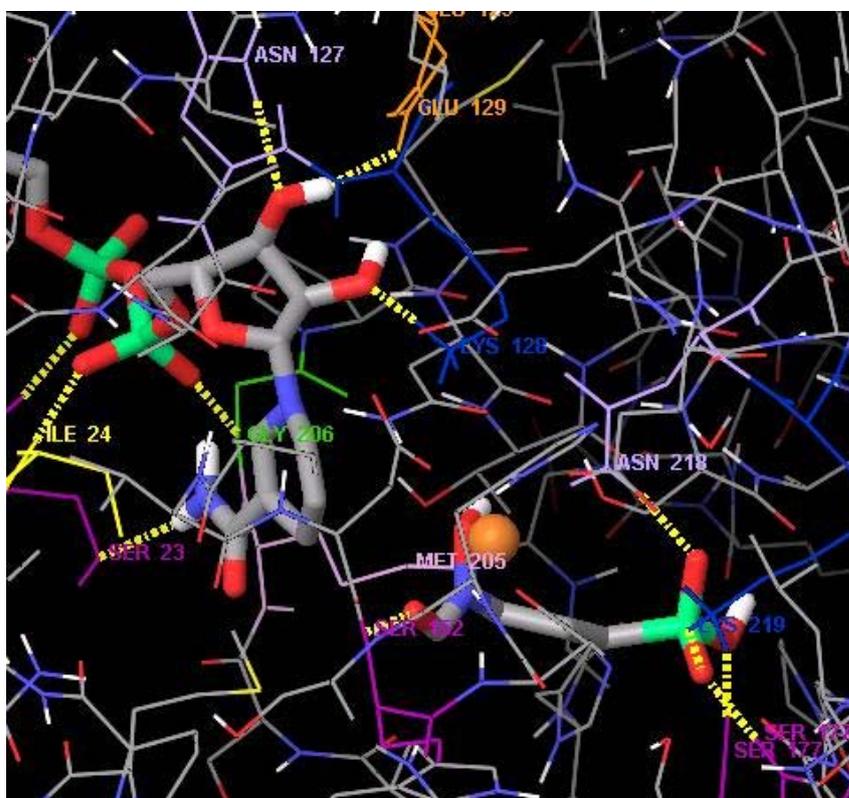


Figure 21

The ligand-target hydrogen bond interactions present in the crystal structure in the part of the binding pocket were *dxrhyp* features are located. The aa-residues interacting with FOM and NADPH are labelled and coloured according to the residue properties: asn (bright purple) ser (dark purple), glu (orange), lys (blue) and glu (green)

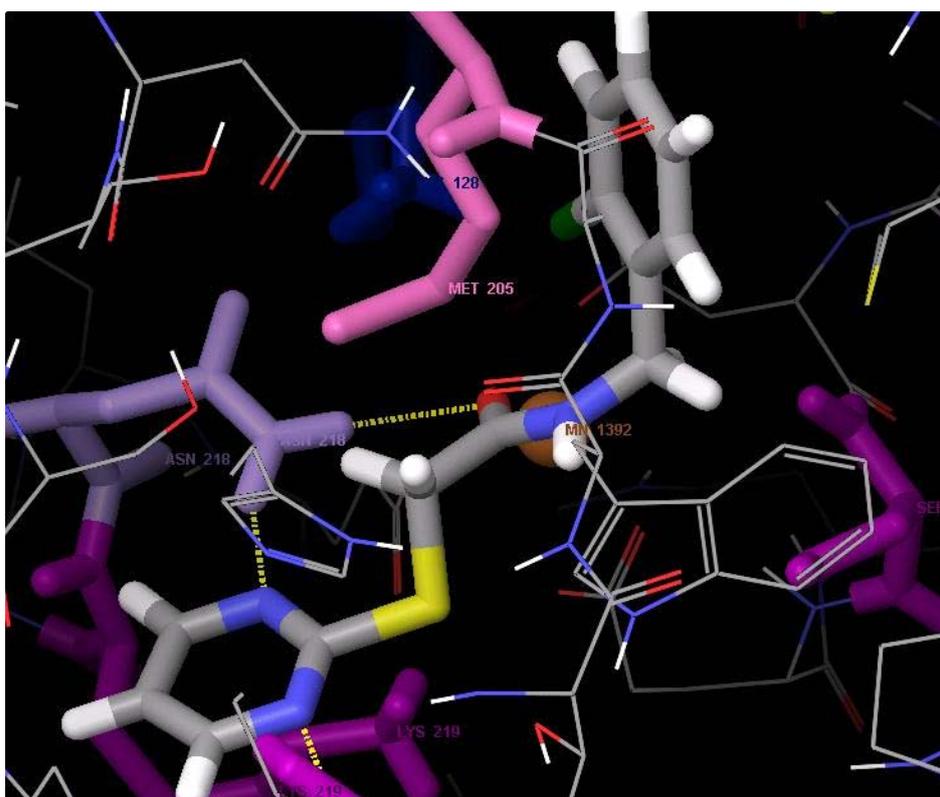


Figure 22
The docked position of compound 5536131 makes it possible to fulfill both hydrogen bonds and hydrophobic interactions found in the crystal structure. The lower part of the ligand interacts with aa-residues in the P-form part of the ligand pocket. Behind the amide bond in the ligand the manganese ion is visible, possible to interact with the oxygen in the compound. The upper part of the ligand occupies the Nicotinamide site in the binding pocket. The ligand is drawn in tubes, and aa-residues interacting with the ligand in the crystal structure are drawn in tubes and coloured differently according to the residue properties.

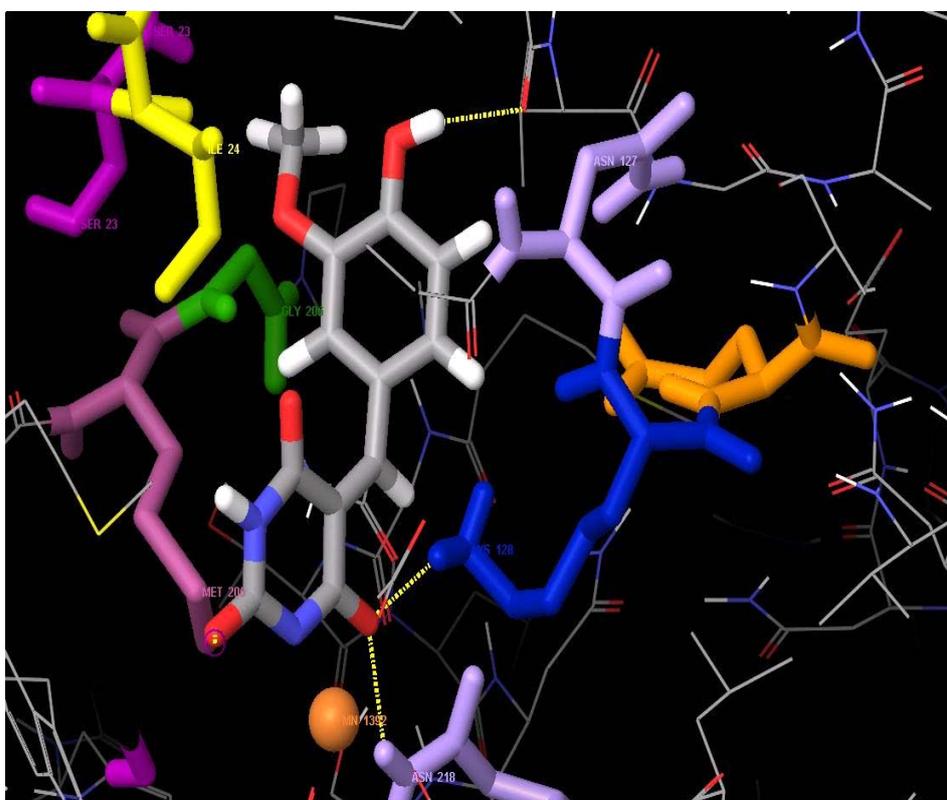


Figure 23
Compound 6307919 is mainly docked into the Nicotinamide- and sugar site. It is stabilized with hydrogen bonds in both ends of the molecule, and hydrophobic interactions found in the crystal structure are fulfilled as well. The ligand is able to interact with the manganese ion in the lower part of the molecule.

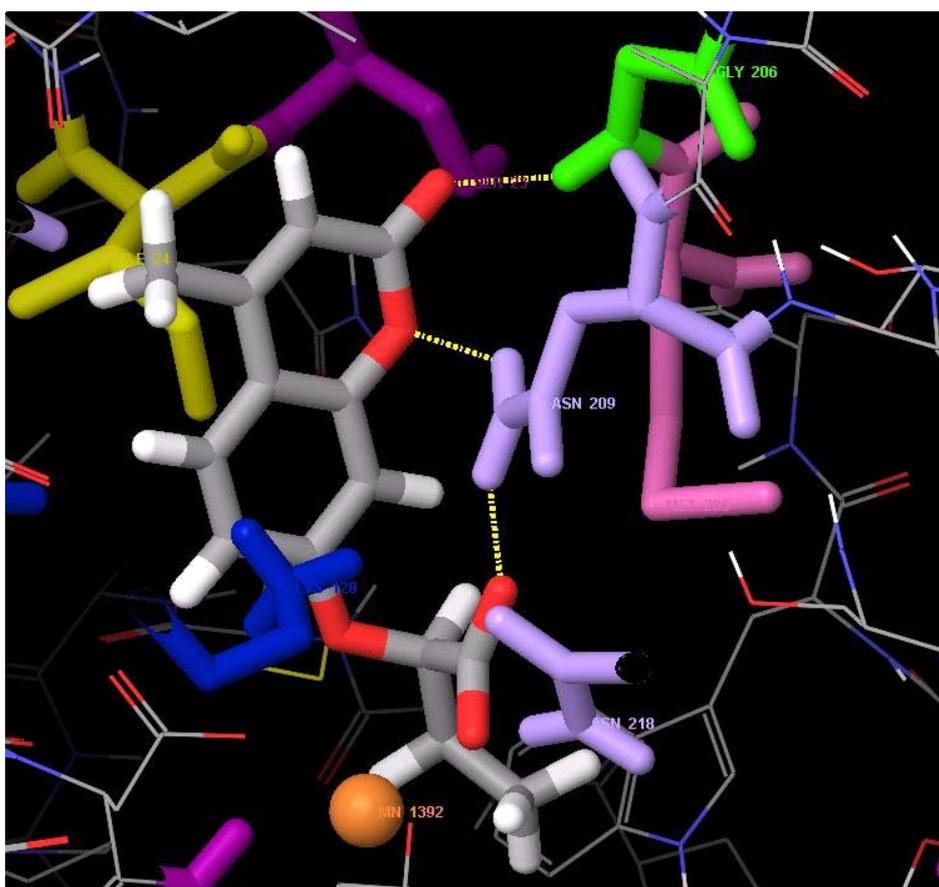


Figure 24
The sugar- and nikotinamide site is the place where to compound 6632003 is docked. Hydrogen bonds as well as hydrophobic interactions stabilize the position, and the ligand is able to interact with the manganese ion. A part of the ligand stretches down through the ion down to the Fom-mn region.

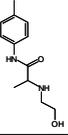
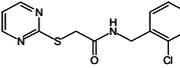
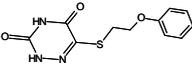
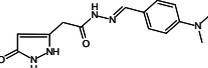
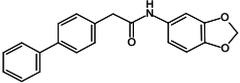
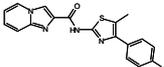
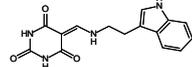
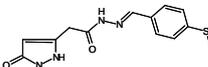
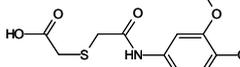
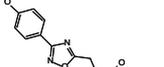
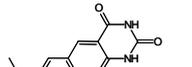
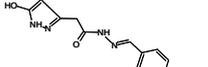
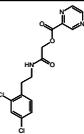
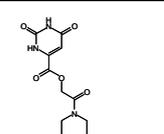
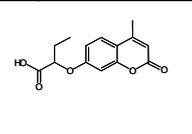
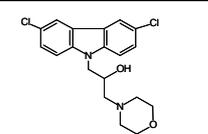
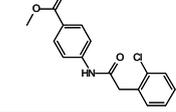
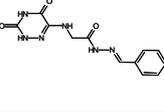
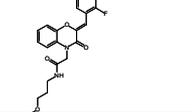
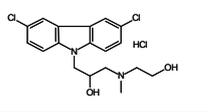
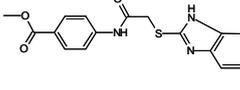
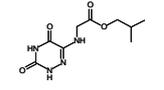
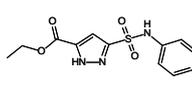
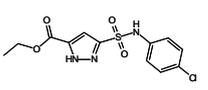
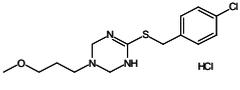
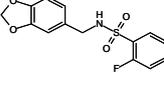
Chemistry	Compound	Chemistry	Compound	Chemistry	Compound	Chemistry
	5536131		6300173		6856666	
	5639015		6325967		6907932	
	6253845		6307919		7175844	
	5336919		6813010		6916459	
	5844050		7350638		4949841/224320	
	4899429		6073704		6073733	
	5685291					

Figure 25 The selected compounds hypothesized to interact with *Mt*-DXR.

5. Conclusion

The aim of this project was to identify about 30 potential inhibitors to each of the targets *Pf*-SRM (using the PDB structure file 217C) and *Mt*-DXR (using the PDB structure file 2JCZ) from a database containing 2,6 million compounds. The experimental approach was to combine pharmacophore filtering, docking and two scoring methods and evaluate the hits by visual analysis.

The project work has been successful, since 30 potential inhibitors to *Pf*-SRM and 26 potential inhibitors to *Mt*-DXR have been selected.

The approach with an initial pharmacophore filtering of the database made the computational approach quite swift. The combination of Phase pharmacophore filtering and Glide docking resulted in a specifically selective method, making it possible to narrow the compound bank of potential inhibitors from 2,6 million to a couple of thousand in three consecutive steps.

However, the approach has been impaired with a few practical complications, both because of bugs in the programs and non-automatized parts in the experimental procedure. A large amount

of compounds has successfully been screened, but the procedure could be even smoother by automatization using more scripts.

6. Discussion

The starting point in the search for potential inhibitors to the targets has been experimentally verified ligand-target interactions. The use of Phase pharmacophore filtering as an initial selecting algorithm made it possible to screen large amounts of compounds with experimentally verified binding features identified as important for ligand-target interaction. By combining Glide and Phase selecting algorithms, the approach has been more specifically selective than utilization of just one of the sorting algorithms.

All high-throughput docking programs use the rigid receptor approximation. This means that the target structure is regarded as a rigid body, with no conformational flexibility. This could potentially mean that compounds will be docked or scored incorrectly. However if small changes in the protein structure such as changing the torsion of an OH group or turning an aspartate residue towards the ligand would be allowed, it would lead to specific interactions being formed, resulting in a much better ligand score. During the visual inspection of docked poses, this was taken into account. Thus, if a specific interaction seemed to be able to form if the protein would be allowed to accommodate the ligand by small conformational changes, that specific interaction was assumed to exist. However, since not all docked compounds were visually inspected, the rigid body approximation probably has resulted in a number of false negatives.

Since the scores from the first Glide SP docking were not used, but Phase pharmacophore filtering of all docked poses regardless of score was used instead, the amount of false negatives is probably reduced compared to what it should be if the first Glide docking score was used as a ground for selection.

The decision to use just one target structure instead of an ensemble of different target structures for each target is a factor contributing to the amount of false negatives. An ensemble of different structures would screen for compounds potentially able to bind to different target states, while the used approach yields binders to one specific target structure. However, the project approach was to find some binders to the target and not all binders, meaning that the risk of false negatives was well known and taken into account. In an approach where the aim is to find as many potential binders as possible, the choice to use an ensemble of target structures would probably have been wiser.

Two different scoring methods were used in parallel before the visual analysis; XP Scoring and MASC scoring. It can be noticed that the MASC scoring was of less importance, since it to a large extent gave good scores to ligand poses that in the visual analysis seemed to bind poorly to the target structure.

As already mentioned in "Conclusion", the approach could be made smoother with less need of manual intervention by script-based automatization. However, the project work has described a method where combining two different screening applications makes it possible to specifically select compounds with certain features from a huge compound database.

In addition to the work described in this report, some laboratory work has been made. With assistance from the Structural Genomics Consortium –(SGC) Stockholm *Pf*-SRM has been expressed and purified with the intention to facilitate further experimental testing of the selected compounds.

The binding of some of the predicted inhibitors for Pf-SRM has subsequently been experimentally verified, as a follow-up to this project. The Pf-SRM part of this project work together with the experimental testing of the predicted binders was recently accepted for publication in Journal of Medical Chemistry³⁰.

-
- ¹ WHO. World Health Organization:
<http://www.who.int/mediacentre/factsheets/fs094/en/index.html> (10 Apr. 2008).
- ² WHO, World Health Organization:
<http://www.who.int/mediacentre/factsheets/fs104/en/index.html> (10 Apr. 2008).
- ³ Leach A R, Shoichet B K, Peishoff C E, "Prediction of protein-ligand interactions. Docking and scoring: successes and gaps", *J Med Chem.* **2006**, vol. 49, no. 20, pp.5851-5855.
- ⁴ Warren L G, Andrews C W, Campelli A-M, Clarke B, LaLonde J, Lambert M H, Lindvall M, Nevins N, Semus S F, Senger S, Tedesco G, Wall I D, Woolven J M, Peishoff C E, Head M S, "A critical Assessment of Docking Programs and Scoring Functions", *J. Med. Chem.* **2006**, vol. 49 no. 20 pp. 5912-5931.
- ⁵ Leach A R, Shoichet B K, Peishoff C E, "Docking and Scoring, Prediction of Protein – Ligand Interactions. Docking and Scoring: Successes and Gaps, " *J. Mol. Chem.* **2006**, vol. 49, no. 20, pp. 5851-5855.
- ⁶ Vigers G P, Rizzi J P "Multiple active site corrections for docking and virtual screening", *J. Med. Chem.* **2004** vol. 47, pp. 80-89.
- ⁷ Maestro, version 8.0, Schrödinger, LLC, New York, NY, 2007.
- ⁸ Dixon S L, Smondyrev A M, Knoll E H, Rao S N, Shaw D E, Friesner R A, "PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening. 1. Methodology and Preliminary Results", *J. Comput. Aided Mol. Des.* **2006** vol. 20, pp. 647-671.
- ⁹ Phase, version 2.5, Schrödinger, LLC, New York, NY, 2007.
- ¹⁰ LigPrep, version 2.1, Schrödinger, LLC, New York, NY, 2005.
- ¹¹ Glide, version 4.5, Schrödinger, LLC, New York, NY, 2007.
- ¹² Impact, version 4.5, Schrödinger, LLC, New York, NY, 2005.
- ¹³ Leach A, "Molecular Modelling: Principles and Applications (2nd Edition)", *Pearson Education Limited* **2001**.
- ¹⁴ Friesner R A, Banks J L, Murphy R B, Halgren T H, Klicic J J, Mainz D T, Matthew P, Repasky M P, Knoll E H, Shelley M, Perry J K, Shaw D E, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy", *J. Med. Chem.* **2004**, vol. 47 no. 7 pp. 1739-1749.
- ¹⁵ Vigers G P, Rizzi J P, "Multiple active site corrections for docking and virtual screening", *J. Med. Chem.* **2004**, vol. 47 pp. 80-89.
- ¹⁶ Vigers G P, Rizzi J P, "Multiple active site corrections for docking and virtual screening", *J. Med. Chem.* **2004**, vol. 47 pp. 80-89.
- ¹⁷ Jacobsson M, Karlén A, "Ligand Bias of Scoring Functions in Structure-Based Virtual Screening", *J. Chem. Inf. Model* **2006**, vol. 46 pp. 1334-1343.

-
- ¹⁸ Connolly M L, "Analytical molecular surface calculation", *J Appl Cryst* **1983**, vol. 16 no. 5 pp. 548–558.
- ¹⁹ Goodford P, "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", *J. Med. Chem.* **1985**, vol 28, pp. 849.
- ²⁰ Dufe V T, Qiu W, Müller I B, Hui R, Walter R D, Al-Karadaghi S, "Crystal structure of *Plasmodium falciparum* spermidine synthetase in complex with the substrate decarboxylated S-adenosylmethionine and the potent inhibitors 4MCHA and AdoDATO", *J. Mol. Biol.* **2007**, vol. 373 pp. 166-167.
- ²¹ Henriksson L M, Unge T, Jones T A, Mowbray S L, "Structures of Mycobacterium Tuberculosis 1-Deoxy-D-Xylulose 5-Phosphate Reductoisomerase Provide New Insights Into Catalysis", *J. Biol. Chem* **2007**, vol. 282, no. 27, pp. 19905–19916.
- ²² Wrenger C, Lüersen K, Krause T, Müller S, Walter R D, "The *Plasmodium falciparum* bifunctional ornithine decarboxylase, Sadenosyl-l-methionine decarboxylase, enables a well balanced polyamine synthesis without domain–domain interaction", *J Biol Chem* **2001**, vol. 276, no. 32, pp. 29651–29656.
- ²³ Kaiser A, Gottwald A, Wiersch C, Lindenthal B, Maier W, Seitz HM. "Effect of drugs inhibiting spermidine biosynthesis and metabolism on the in vitro development of *Plasmodium falciparum*", *Parasitol Res* **2001**, vol. 87 pp. 963–972.
- ²⁴ Dufe V T., Qiu W., Müller I B, Hui R, Walter R D, Al-Karadaghi S., "Crystal structure of *Plasmodium falciparum* Spermidine Synthase in Complex with the Substrate Decarboxylated S-adenosylmethionine and the Potent Inhibitors 4MCHA and AdoDATO", *J. Mol. Biol.* **2007**, vol. 373, pp. 167-177.
- ²⁵ Takahashi S, Kuzuyama T, Hiroyuki W, Seto H., "A 1-deoxy-xylulose 5-phosphate reductoisomerase catalysing the formation of 2-C-methyl-D-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis", *Proc. Natl. Acad. Sci.* **1998**, vol. 95, pp 9879-9884.
- ²⁶ Kuzuyama T, Takahashi S, Watanabe H, Seto H, "Direct formation of 2-c-methyl-D-Erythritol 4-phosphate from 1-deoxy-d-xylulose 5-phosphate by 1-deoxy-d-xylulose 5-phosphate reductoisomerase, a new enzyme in the non-mevalonate pathway to isopentenyl diphosphate", *Tetrahedron letters* **1998**, vol. 39, pp. 4509-4512.
- ²⁷ Henriksson L, Unge T, Carlsson J, Åqvist J, Mowbray S L, Jones T A, "Structures of Mycobacterium tuberculosis 1-Deoxy-D-xylulose-5-phosphate Reductoisomerase Provide New Insights into Catalysis", *J. Biol. Chem* **2007**, vol. 282, no. 27, pp. 19905–19916.
- ²⁸ Silber K, Heidler P, Kurz, T, Klebe G, "AFMoC Enhances Predictivity of 3D QSAR: A Case Study with DOXP-reductoisomerase", *J. Med. Chem.* **2005**, vol. 48, pp. 3547-3563.
- ²⁹ Yajima S, Hara K, Sanders J M, Yin F, Ohsawa K, Wiesner J, Jomaa H, Oldfield E, "Crystallographic Structures of Two Bisphosphonate:1-Deoxyxylulose-5-Phosphate Reductoisomerase Complexes", *J. Am. Chem. Soc.* **2004**, vol. 126, no. 35, pp. 10824 -10825.

³⁰ Jacobsson M, Gäredal M, Schultz J, Karlén A, "Identification of Plasmodium falciparum Spermidine Synthase Active Site Binders through Structure-Based Virtual Screening", *J. Med. Chem.* **2008**, vol. 51, no. 9, pp. 2777 – 2786.