

Docking, scoring and binding-affinity prediction in computer-aided drug discovery

I Development of a scoring function for quantifying binding affinities

II Developing a mQSAR model for the Pregnane X receptor

III Toxic potential evaluation for natural anti-trypanosomiasis compounds

Inauguraldissertation

ZUR

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Zhenquan Hu

aus Wuhan, China

Basel, 2015

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel

edoc.unibas.ch



Dieses Werk ist unter dem Vertrag „Creative Commons Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 3.0 Schweiz“ (CC BY-NC-ND 3.0 CH) lizenziert.

Die vollständige Lizenz kann unter

creativecommons.org/licenses/by-nc-nd/3.0/ch/

eingesehen werden.

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von:

Prof. Dr. Angelo Vedani

Prof. Dr. Matthias Hamburger

Basel, den 15. September 2015

Prof. Dr. Jörg Schibler
Dekan

Acknowledgements

First of all I would like to thank Prof. Dr. Angelo Vedani for giving me the great opportunity to do my doctoral dissertation in his group, and for his patience and long-term supports during my work and study in Basel.

I would also like to thank Prof. Beat Ernst for providing a challenging scientific environment to face it.

I further would like to thank Prof. Dr. Matthias Hamburger for accepting to be the co-referee of my thesis.

I would also like to thank Claudia Huber and Gabi Lichtenhahn for their administrative supports.

Many thanks to Dr. Martin Smiesko, Dr. Sameh Eid and Dr. Shuguang Yuan for their technique support during my PhD study. Thanks to all current and former members of the Institute of Molecular Pharmacy for sharing the experiences during the Monday seminars. Further more, I would like to thank the modellers Christoph, Oya and Joe, for the enjoyable working environment and their kind help and support.

Especially thank to Dr. Lijuan Pang and Dr. Fan Yang for their help and guiding during my early stay in Basel.

Additionally, I would like to thank the Association of Chinese Students and Scholars in Basel for providing me the opportunities to attend the fantastic events and to meet many new friends in Switzerland.

Finally, I would like to thank my parents, my parents in-law who always believe in me and support me. Many thanks to my dearest wife Dr. Yanlei Li for being with me through the times.

Abstract

Docking and scoring are widely used in nowadays drug discovery process. Scoring function is used as a fast method to estimate the docking results. In this thesis, a regional-defined genetic algorithm approach is developed to optimize the force-field based scoring function.

Human pregnane X receptor (PXR) is a nuclear receptor which is promiscuous in its affinity for ligands such as bile acid, steroid hormones, fat-soluble vitamins, prescription and herbal drugs, and environmental chemicals. In this thesis, the development and validation of in silico three-dimensional models for the pregnane X receptors is presented. These model aim at the screening of drug candidates for potential activity towards the PXR.

Potential side effects and toxicity of anti-trypanosomiasic active compounds were investigated using the *VirtualToxLab*. This technology identifies the binding mode of a small-molecule compound toward a series of 16 target proteins (nuclear receptors, cytochrome P450 enzymes, hERG, AhR) known or suspected to trigger adverse effects. The kinetic stability of the identified hits are evaluated by molecular dynamics simulations.

Contents

1 Introduction	5
1.1 Computer Simulation of Protein–Ligand interactions	5
1.1.1 Protein–Ligand Docking	6
1.1.2 Scoring functions	9
1.1.3 Quantitate Structure–Activity Relationships (QSAR)	12
1.1.4 <i>VirtualToxLab</i>	13
1.2 Pregnane X Receptor and Drug Metabolism	15
1.2.1 The Pregnane X receptor	15
1.2.2 Approaches in PXR modeling	16
1.2.3 mQSAR for PXR	17
1.3 Evaluation of the toxic potential of natural products with anti-trypanosomal activity	20
1.3.1 Human African trypanosomiasis	20
1.3.2 Natural products active for <i>T. b. gambiense</i>	21
1.3.3 In silico toxic potential profiling by the <i>VTL</i>	22
2 Methods	23
2.1 4D scoring function development	23
2.2 QSAR model for the PXR	27
2.2.1 Protein processing	27
2.2.2 Docking protocol	28
2.2.3 Molecular-Dynamics Simulations	30
2.2.4 Trajectory analysis	31
2.2.5 QSAR studies: Quasar & Raptor	32
2.3 Toxic potential estimation by means of the <i>VTL</i> and MD simulations	33
3 Results & Discussion	35

3.1	Scoring Function development	35
3.1.1	Flowchart of data processing	35
3.2	QSAR model development for the PXR	38
3.2.1	Binding-site analysis	38
3.2.2	MD simulation for the PXR with ligands in the PDB	39
3.2.3	Retrieval of the binding affinity data	46
3.2.4	Docking to the PXR	47
3.2.5	Multidimensional QSAR for the PXR	59
3.3	<i>In silico</i> evaluation of anti-trypanosome natural products	62
3.3.1	Compounds active towards the androgen receptor	64
3.3.2	The only compound active towards the Aryl hydrocarbon receptor	73
3.3.3	Compounds active towards the estrogen receptor β	75
3.3.4	Compounds active towards the glucocorticoid receptor	77
3.3.5	Compounds active towards the mineralocorticoid receptor	80
3.3.6	Compounds active towards the progesterone receptor	82
3.3.7	The only compound active toward the thyroid receptor β	88
3.3.8	Compounds active toward the Cytochrome P450 2D6	90
4	Conclusions	92
	Appendices	93
A	Chemical structure of the PXR ligands	94
A.1	Chemical structures	94
A.1.1	Estratrienes (E01–E08)	94
A.1.2	Androstans (A01–A18)	95
A.1.3	Pregnanes (P01–P23)	96
A.1.4	T091317 derivatives (T01–T12)	97
A.1.5	BMS-817399 derivatives (S01–S07)	97
A.1.6	Benzenesulfonamide derivatives (M01–M09)	98
A.1.7	2-Aryl indoles (N01–N08)	98
A.1.8	Substituted phenyl triazoles (G01–G16)	99
B	Anti-trypanosomal compounds	100

Chapter 1

Introduction

1.1 Computer Simulation of Protein–Ligand interactions

Protein–ligand interactions play a central role in living cells. From G protein–coupled signal transduction to nuclear–receptor inducing gene expression, from programming cell death to embryo stem-cell differentiation. In the era of ever expanding biological complexity, the interplay between structural, computational and chemical biology has brought major scientific advances to modern biomedical research. The complexity of interactions between a small-molecule ligand and its target protein, is often determined by the flexibility of the protein-binding site and by the structural rearrangements that occur upon binding. The forces that control protein behavior and their physical–chemical origins are inferred from equilibrium binding-kinetic measurements or are computed with molecular models. Calculated energies may be associated with physical and chemical interactions ruling protein function and behavior. Although *in silico* models are usually simplified and abstracted from the real world, computational techniques are continually evolving so as to broaden the range of feasible applications, and the accuracy of predictions and theoretical approaches are often supportive in guiding and interpreting experiments.¹

1.1.1 Protein–Ligand Docking

During the docking process, different poses are generated by the search algorithm, which ideally should sample the degrees of freedom of the protein–ligand complex not necessarily exhaustive but adequately enough as to include the true binding modes. These different poses are then evaluated and used to identify the true binding mode(s) for a given ligand, and to estimate its binding affinity. Hence, a scoring function should not only be able to ensure a distinction between different alternatives and ranking them accordingly, but also to represent the thermodynamics of interaction of the protein–ligand system accurately. Protein flexibility, water molecules and entropy are important factors that influence docking.

Ligand Flexibility

Ligand conformational sampling is an essential step to generate a multi-conformer dataset used in ligand sampling (e.g. *Glide*²). To reduce the complexity of conformational-space sampling for the ligand and also to collect the binding poses more efficiently, pharmacophore-based methods have been adopted to pre-align the ligand within the binding site (e.g. *Dolina*³), with the underlying algorithm incorporating geometric and chemical features based on known or computationally-identified template ligands.

Protein Flexibility

Protein flexibility, including side-chain reorientations and backbone motions, can significantly modulate the geometry and characteristics of the ligand binding site.⁴ Due to the high degrees of freedom, most of the strategies already implemented in protein–ligand docking programs accounting for side-chain flexibility only, with the inclusion of backbone flexibility would be computationally extremely demanding.⁵ Main strategies including some level of protein flexibility into protein–ligand docking are: (1) soft docking, allowing a certain overlap between receptor and ligand, limited to small scale rearrangements associated to side-chain plasticity, without the corresponding backbone adjustment. (e.g. *Glide*²); (2) side-chain flexibility, by either systematic exploring side-chain rotamers or using a stochastic-searching method. (e.g. *MedusaDock*,⁶ *Cheetah*^{7,8}); (3) ensemble docking, implicitly introducing flexibility into protein–ligand

docking involves the use of an ensemble of protein conformations (X-ray, NMR, molecular dynamics, Monte Carlo) as a target for docking instead of a single structure (e.g. *AutoDock*⁹); (4) induced-fit docking, consider both side-chain and backbone flexibility (e.g. *IFD*,¹⁰ *RosettaLigand*¹¹); (5) molecular-dynamics simulations, generating perturbed ensembles, which represent ligand-induced binding-site flexibility in a robust process.¹²

Water molecules

Water molecules on the interface of bimolecular complexes play an important role in protein–ligand interactions¹³ (hydrophobic interactions, residual mobility, bridged hydrogen bond, desolvation, dielectric properties). Active-site water molecules can be considered as an aspect of target flexibility during docking. Properly accounting for specific water molecules on the interface as well as for the general effect of solvation is a vital aspect of docking. Rossato *et al.* developed an algorithm (*AcquaAlta*⁸) to match predicted water positions (structural water molecules) with the Cambridge Structure Database; Zheng *et al.* introduced *wPMF*¹⁴ (a knowledge-based method) to predict the potential hydration sites of protein structure, Abel and Friesner *et al.* developed *WaterMap*¹⁵ to identify hydration sites in binding pockets and to evaluate the favorability of their displacement using an empirical formula based on the computed enthalpic and entropic contributions, Hu *et al.* developed *WATSite*¹⁶ to identify hydration sites using a MD trajectory, the thermodynamic profile of each hydration site is then estimated by computing the enthalpy and entropy of the water molecule throughout the simulation. Representation of bulk water is even more problematic than that of specific water molecules, particularly for polar or charged systems.¹⁷ Docking with flexible (on/off) water molecules (e.g. *MolDock*,¹⁸ *Cheetah*^{7,8}) led to an improvement in pose prediction.

Entropy

Entropic effects contribute substantially to the protein–ligand binding energy. They arise from a variety of aspects including the reduction of the translational and rotational degrees of freedom in the ligand, changes in the normal modes of protein vibration and of

the ligand during binding, from the arrangement of water layers around the two entities and even from protonation and deprotonation events. However, in most commonly used computational applications that deal with protein complexes, including free energy calculations, entropy is neglected altogether, or dramatically simplified such as counting the number of ligand rotatable bonds as a factor of ligand entropy. In Cheetah,^{7,8} the entropic contribution is estimated from the different conformational flexibility in the bound and unbound state.

Binding Modes—Exploration of the 4th dimension

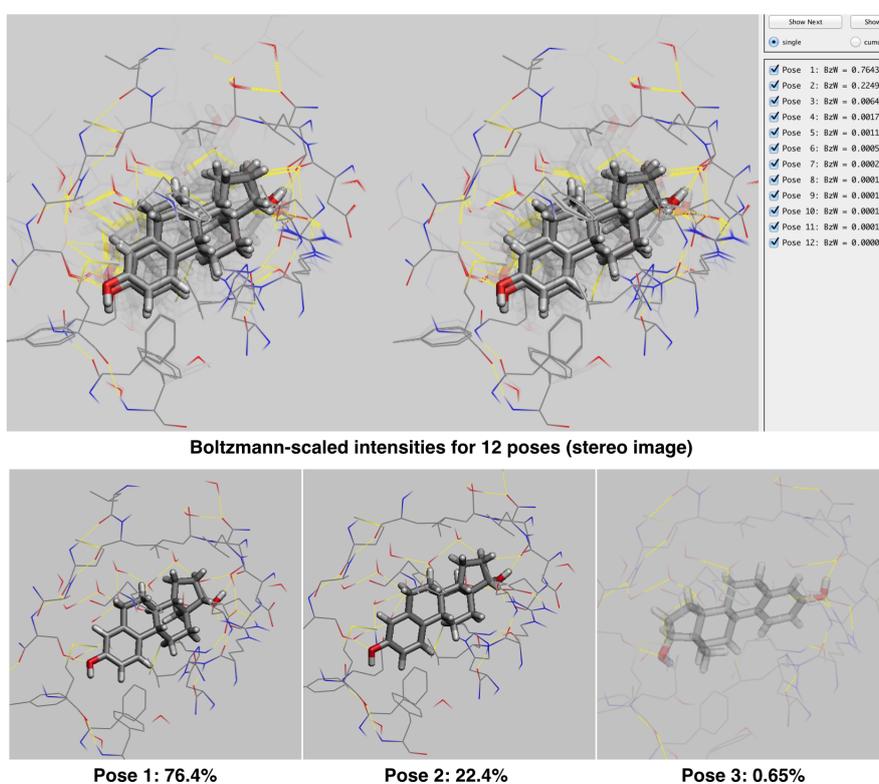


Figure 1.1: 4D-binding of 17 β -estradiol to the Pregnane X receptor

In general, a 3D structures of protein–ligand complex can provide a direct insight into the interactions between a ligand and its target protein. However, it is hard to discriminate the favorability of the binding poses. Biographics' 4D viewer allows inspecting all identified poses (potential binding modes) Boltzmann-weighted and at once.¹⁹ The poses are shown

with a scaled intensity corresponding to the individual contribution to the binding energy (cf. Boltzmann weight in the panel at right, figure 1.1). This method could provide a more meaningful interaction diagram for binding-pose identification.

1.1.2 Scoring functions

Various classes of scoring functions are used for pose prediction and ligand ranking, which can be classified as physics-based, knowledge-based, empirical-based and descriptor-based methods.²⁰

Physics-based scoring functions

Physics-based approaches the approximate binding free energy by combining non-bonded energy terms of molecular-mechanics force fields, solvation energy, with or without entropy.²¹ An example is MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area).^{22,23} In the MM-PBSA approach, a molecular-mechanics force field represents the solute (receptor, ligand, and receptor-ligand complex), and the PB equation represents the solvent molecules as a structureless dielectric continuum with ions distributed in a mean-field manner according to the Boltzmann distribution. The dielectric continuum treatment represents the solute as a low dielectric body, with a shape defined by the atomic coordinates, radii, molecular surface, and its 3D spatial charge distribution defined by the atomic coordinates and partial atomic charges. The non-polar solvation is divided into two terms: the repulsive (cavity) and attractive (dispersion) interactions, which corresponds to the creation of a cavity in water and the vdW interactions between the non-polar molecule and the water molecules, respectively. The polar solvation term describes the contribution to the free energy due to polarization of the solvent environment by the solute. The implicit water model of PB solver neglecting the degrees of freedom associated with each individual water and ion molecule. The main equations for the MM-PBSA method are shown below, where G denotes the Gibbs free energy:

$$\Delta G = G_{solvated}[Complex] - (G_{solvated}[Receptor] + G_{solvated}[Ligand]) \quad (1.1)$$

The free energies of each species are typically evaluated by the following relations, where T is the temperature, S is the entropy, E is the potential energy, evaluated as the terms in

molecular mechanics force field energy, used in place of the enthalpy:

$$G_{solvated} = G_{gas} + G_{solvation} - TS \quad (1.2a)$$

$$G_{gas} = E_{internal} + E_{elec} + E_{vdW} \quad (1.2b)$$

$$E_{internal} = E_{bond} + E_{angle} + E_{torsion} \quad (1.2c)$$

$$G_{solvation} = E_{solv,polar} + E_{solv,nonpolar} \quad (1.2d)$$

The polar solvation energy is obtained by solving the PB equation:

$$\nabla[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) - 4\pi\lambda(\mathbf{r})\sum_i z_i c_i \exp(-z_i\phi(\mathbf{r})/k_B T) \quad (1.3)$$

where $\epsilon(\mathbf{r})$ is the dielectric constant, $\phi(\mathbf{r})$ is the electrostatic potential, $\rho(\mathbf{r})$ is the solute charge, $\lambda(\mathbf{r})$ is the Stern layer-masking function, z_i is the charge of ion type i , c_i is the bulk number density of ion type i far from the solute, k_B is the Boltzmann constant, and T is the temperature; the summation is over all different ion types.

The adoption of the MM-PBSA methodology by the scientific community has resulted in an increased number of publications including MM-PBSA results, yet it has also seen an increase in the errors of the predicted binding affinities reported using this method with respect to the experimentally measured binding affinities. One reason is a net difference between a large desolvation penalty value and large electrostatic interaction energy could be very small, but it is that difference strongly affects the binding affinity, a very accurate determination is required to reduce the errors in the net difference, another reason is in part to the significant number of parameters in the PB equation that are not optimized by the end users. The dielectric-interface representation, the interior dielectric constant, handling of internal cavities, bridged waters are key parameters et to perform a valid MM-PBSA approach for protein-ligand interactions.²⁴ Advanced methods by incorporating QM/MM method within the binding pocket to optimize the geometry and partial charges²⁵ or by using a polarizable force field could give more insight into this aspect at higher computing expense.

Knowledge-based potentials

Knowledge-based scoring functions are exclusively built from statistical analyses of experimentally determined complex structures, based on the assumption that interatomic

distances occurring more often than some average value should represent favorable contacts, and vice versa.²⁶ The development from atom-pairwise correlation to group correlation (network motif, pharmacophore) and incorporation of extended physical-meaning potentials has optimized this process since its origin. Such as MotifScore,²⁷ which introduced tally motifs of protein–ligand interaction networks constructed from experimental structures, it captured frequently occurring protein–ligand interactions by network motifs instead of pairwise interactions. Another effort captured knowledge-vested pharmacophore to process a knowledge-guided scoring strategy. Neudert and Klebe developed DSX,²⁸ consisting of distance-dependent pair potentials, novel torsion angle potentials and solvent-accessible surface-dependent potentials and demonstrated good performance in pose prediction and ligand ranking. Zheng *et al.*¹⁴ presented orientation-dependent hydrogen potential. The preceding studies demonstrate the potential of capturing experimental data for improvement of scoring outcomes, compared with conventional functions. They also reveal the need for more consistent and extensive evaluation and comparison.

Empirical scoring functions

The underlying idea of empirical scoring functions is that the binding free energy of a non-covalent protein–ligand complex can be interpreted as a sum of localized, chemically intuitive interactions. Such energy decompositions can be a useful tool to understand binding phenomena. These average functional-group contributions can then be used to estimate a protein-independent binding energy for a compound that can be compared to experimental values. If the experimental value is approximately the same as or higher than the calculated value, one can infer a good fit between receptor and ligand and essentially all functional groups of the ligand are involved in protein interactions. If the experimental energy is significantly lower, one can infer that the compound can not fully form its potential interactions with the protein. Experimental binding affinities have also been analyzed on a per atom basis in quest of the maximal binding affinity of non-covalent ligands. Also referred to “empirical scoring functions”, on the other hand, are normally used to compute the fitness of protein–ligand binding by summing up the contributions of a number of individual terms, each representing an relevant energetic factor in protein–ligand binding.²⁹ The weights are assigned by regression technique by fitting

predicted and experimentally determined affinities to a given set of training complexes (Böhm and Stahl, 1999). For example, Glide Score:³⁰

$$GScore = 0.05 \cdot vdW + 0.15 \cdot Coul + Lipo + Hbond + Metal + Rewards + RotB + Site \quad (1.4)$$

GScore designed with an emphasis on recognizing the diversity in protein binding sites by rewarding or penalizing certain interaction patterns. Of particular interest is the classification of hydrogen bonds into neutral–neutral, neutral–charged, and charged–charged types and use of separate terms accounting for “hydrophobic enclosure” in addition to consideration of hydrophobic contacts between protein and ligand. The convenience of adding or removing individual terms also makes it possible to develop customized scoring functions for certain molecular systems to achieve better performance. On the other hand, adopting intuitive functional forms adds to the empirical nature of these methods. Empirical scoring functions include only common protein–ligand interaction patterns. Less common interaction patterns, despite being strong and specific such as cation- π interaction, are usually ignored because they are not significant in the regression analysis. Or, if a certain factor is not interpretable by human in a straight-forward manner, such as entropic factors, it is not likely to be included either. Thus, it is rather difficult, if not impossible, to establish a comprehensive and consistent description of all possible factors in protein–ligand binding within the framework of an empirical scoring function.

Machine-learning approaches

One of the postulated weaknesses of scoring resulting in poor affinity prediction, is the assignment of a common set of weights to the individual functional terms and the incorrect assumption that these weights are additive in their contribution to binding affinity. Recent literature has seen an increase of attempts to overcome these shortcomings by applying machine-learning and nonlinear-regression methods to the derivation of scoring functions.¹⁷ (e.g. eSimDock,³¹ SFCcore^{RF},³² ID-Score³³)

1.1.3 Quantitate Structure–Activity Relationships (QSAR)

QSAR attempt to describe and quantify the correlation between chemical structure and biological activity. The foundation of quantitative correlations between chemical structure

and biological effect is the entirely reasonable assumption that the differences in the physicochemical properties are responsible for the relative potency of the interactions of the drug with biological macromolecules. It is assumed in the first approximation that these contribute additively to the affinity of an active substance on its receptor. The concept of describing the biological activity of substances with mathematical models is derived from this approach.³⁴ Traditionally, the investigated substances interact with the same biological target should come from a chemically uniform series and display the same mode of action. With the advent of 3D-QSAR (CoMFA³⁵), a rational model representing the binding site could be generated by mapping physico-chemical properties onto a surface or a grid surrounding the ligand molecules, superimposing in 3D space (pharmacophore hypothesis), as such a model interacts with all ligands simultaneously, it represents but an averaged surrogate; a fundamental shortcoming as receptor–ligand adaptation (the specific alteration of protein conformations induced by the individual ligand) which leads to different physico-chemical fields experienced by the individual ligands, cannot be simulated with an averaged model. By introducing pseudo-receptor concept,³⁶ the Quasar methodology^{7,37} developed by Vedani *et al.* is one of the few QSAR approaches which accounts for ligand triggering induced-fit by specifically allowing for a topological adaptation of the receptor surrogate to the individual ligand molecules, solvation effects is also evaluated simultaneously.

1.1.4 *VirtualToxLab*

The *VirtualToxLab*³⁸ (VTL) is an *in silico* tool for predicting the toxic potential (endocrine and metabolic disruption, some aspects of carcinogenicity and cardiotoxicity) of drugs, chemicals and natural products. The toxic potential of a given compound is computed by simulation and quantification of the binding of a small molecule toward a series of 16 proteins known or suspected to trigger adverse effects. Those so-called “off targets” currently include the androgen, aryl hydrocarbon, estrogen α , estrogen β , glucocorticoid, liver X, mineralocorticoid, progesterone, thyroid α , thyroid β and peroxisome proliferator-activated receptor γ (PPAR γ), potassium voltage-gated channel (hERG) as well as the enzymes cytochrome CYP450 1A2, 2C9, 2D6 and 3A4.

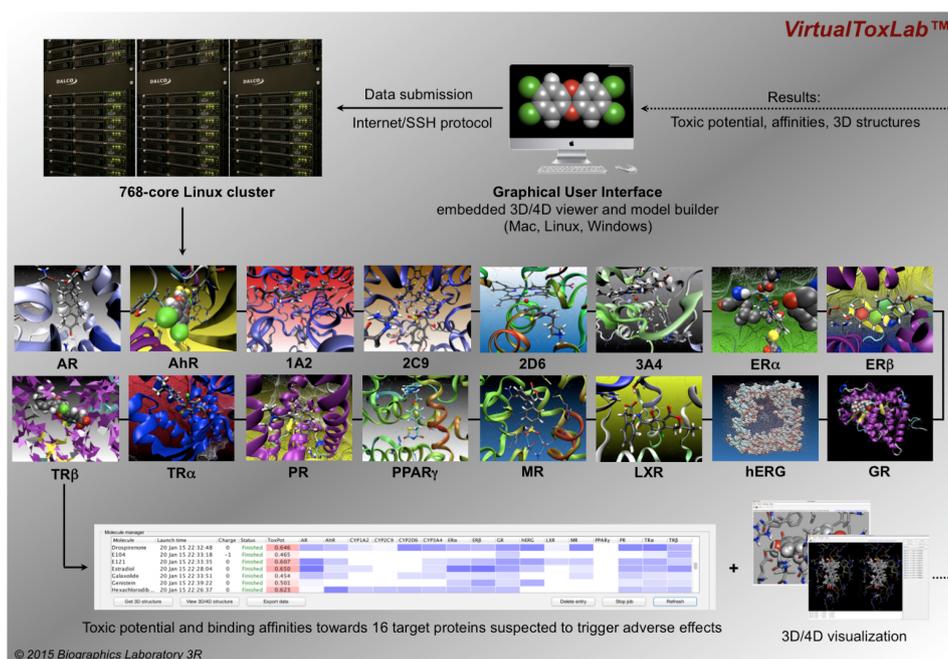


Figure 1.2: Pictorial flowchart of the VirtualToxLab¹⁹

The underlying technology to quantify the binding affinity (mixed-model QSAR), however, meets its limitations when attempting to estimate the binding affinity of compounds significantly different from those present in the training set as individual target proteins are typically validated using only a few chemical classes (for which affinity data, preferably determined at a single laboratory, are available). The identification of the binding mode itself does not depend on any training data as the compounds are docked to the three-dimensional structure of the target protein.

The binding affinity of a small molecule toward a macromolecular target may be estimated by generating and quantifying its 4D ensemble both in aqueous solution and at a target protein and computing the associated changes in the associated energies therefrom. Here, the term “4D” refers to the consideration of all energetically feasible poses (potential binding modes, i.e. different positions, orientations and conformations) of the small molecule, weighted e.g. by a Boltzmann function.

1.2 Pregnane X Receptor and Drug Metabolism

1.2.1 The Pregnane X receptor

The pregnane X receptor (PXR) is an orphan nuclear receptor expressed in mammalian liver, intestine and brain capillaries. It plays a key role in the regulation of both drug metabolism and efflux by inducing a network of genes, including those that encode cytochrome P450 enzymes (CYPs, particularly CYP3A4)³⁹ and the multi-drug resistance gene ABCB1, which encodes the P-glycoprotein.⁴⁰ Together, these drug-detoxification proteins are responsible for the elimination of more than 50% of all drugs.⁴¹

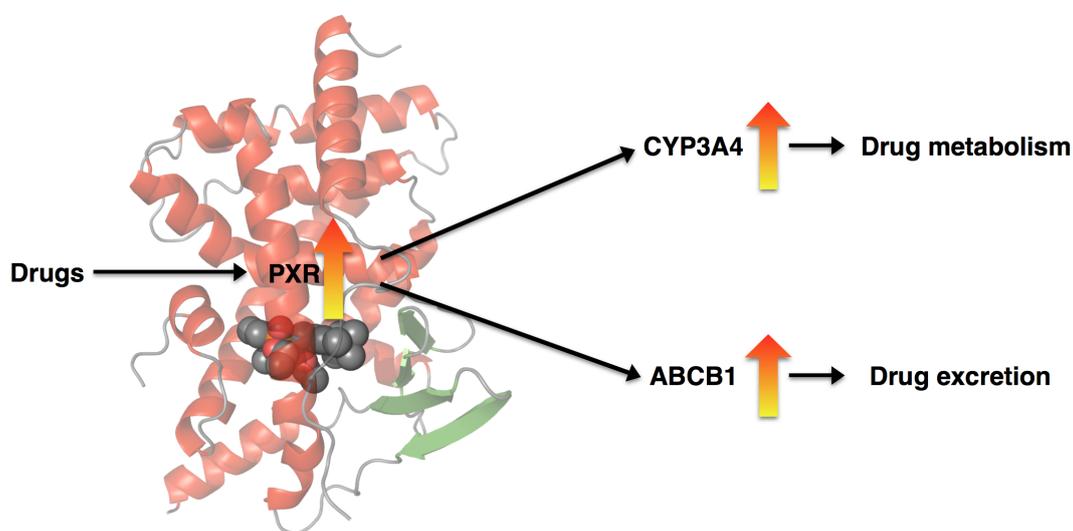


Figure 1.3: Drugs that bind and activate PXR, coordinately induce cytochrome P450 isoform 3A4 (CYP-3A4)-mediated drug metabolism and ABCB-1-P-glycoprotein-mediated drug efflux.

The PXR ligand-binding domain (LBD) is highly flexible and largely hydrophobic with five polar residues capable of both donating and accepting hydrogen bonds. The ligand binding cavity of PXR is lined with 28 amino acids, with an essentially elliptical shape.⁴² Because most proteins use shape to dictate specificity, the promiscuity of PXR appears to be greatly assisted by its relatively shapeless binding pocket. Drug candidates that display agonistic activity toward PXR might cause severe drug–drug interactions and should, consequently, be removed from the drug-discovery pipeline. PXR has evolved to detect structurally diverse compounds, resulting in promiscuous protein–ligand interactions.

Due to the high flexibility of the PXR ligand-binding domain, there were only limited successes with structure-based modeling approaches to predict PXR activators or non-activators. Therefore, new theoretical approaches that can effectively characterize the alterable interaction modes of PXR with its diverse ligands would be in need for solve this problem.

1.2.2 Approaches in PXR modeling

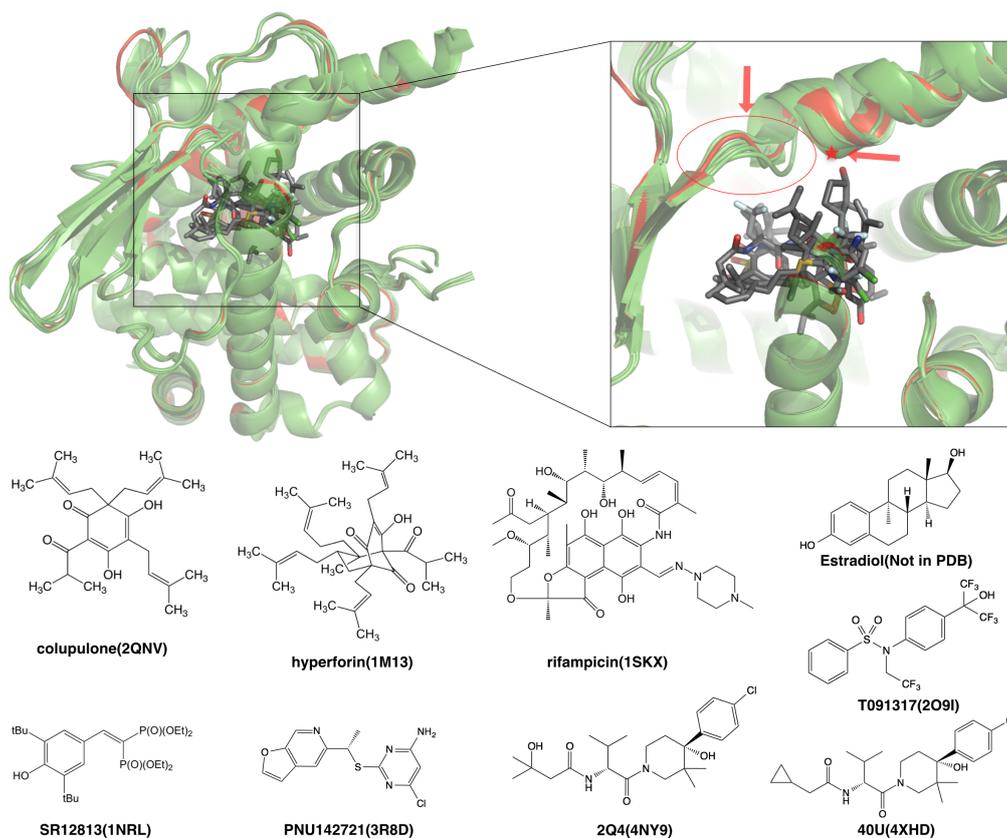


Figure 1.4: Top left: the superposed crystal structures of the PXR-LBD, the *apo* structure is colored in red(PDB ID 1ILG), the complex structures are colored in green (PDB ID: 1M13, 1NRL, 2O9I, 2QNV, 3R8D, 4NY9, 4XHD). The ligand represented as licorice. Top right: the binding pocket of PXR-LBD with the promiscuity hydrophobic binding site (red lines) and hydrogen-bond acceptor site (red star) highlighted. Bottom: Chemical structures solved by X-ray binding towards the PXR (PXR/estradiol was obtained from the author, not available through the PDB).

PXR has evolved to detect structurally diverse compounds (Figure 1.4), it shows conformation-change tolerance towards ligand binding. Theoretical approaches that can effectively characterize the alterable interaction modes of PXR with its diverse ligands are needed for the analysis and investigation of molecular recognition. A study using computational solvent mapping⁴³ suggested that Phe288, Trp299 and Tyr306 play an important role in forming hydrophobic interactions with ligands, and Gln285 is likewise essential in forming hydrogen-bond interactions with ligands. A comprehensive study using 3D, 4D and 5D-QSAR methods⁴⁴ suggested that only a 5D method⁷ could display some extend of success for predicting external test set of steroid compounds. Chen et. al. used multiple binding mode-based quantitative structure-activity relationship (MBMB-QSAR) method that characterizes the non-bonded interaction profile of human PXR with its ligands in multiple binding modes. From the predicted models it has been suggested that the hydrophobic forces and electrostatic interactions play an important role in hPXR–ligand binding, while the steric factor contributes moderately to the binding.⁴⁵ Handa et. al. used MD simulations to sample the active protein conformation for docking, combined with MM-GB/SA for binding free-energy calculation and found good correlations,⁴⁶ which further confirmed the importance of receptor flexibility.

1.2.3 mQSAR for PXR

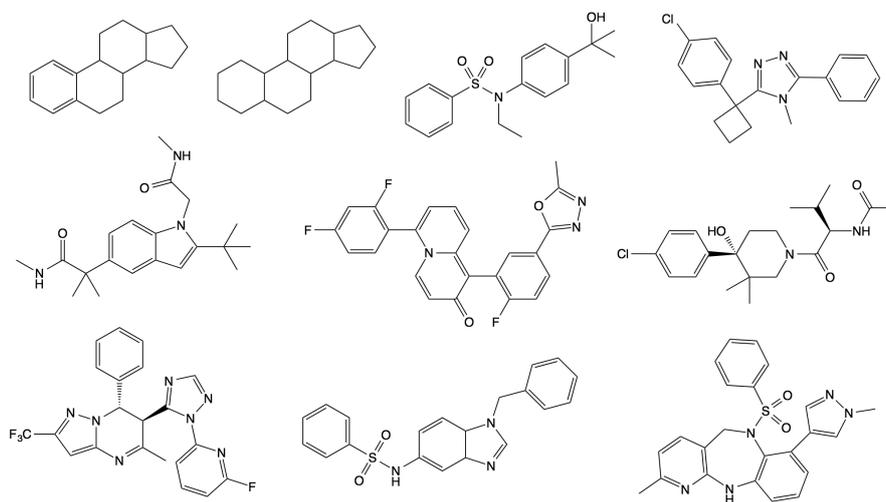


Figure 1.5: Ligand scaffolds used for the PXR QSAR study

In this work, affinities (EC_{50}) for PXR binding molecules were selected from a series of congeneric compounds and classified by chemical type.⁴⁷⁻⁵⁵ The binding modes for the high affinity representatives in each class are generated by manual docking with *Yeti* as well as automated docking with *Cheetah* and *Schrödinger IFD* and visually inspected and identified; the others are docked by both pharmacophore-based pre-aligned docking and Monte-Carlo metropolis sampling (software *Alignator/ Dolina*,³ *Cheetah*^{7,8}) based on the identified templates. The underlying algorithm particularly allows for two aspects of ligand-protein binding which would seem to be of utmost importance: 1. Simulation of induced fit, i.e. allowing the protein to adapt its shape to the different orientations and conformations of the small molecule during the search procedure and 2. Quantification of solvent effects (ligand desolvation, solvent stripping). The minimization algorithm is driven by a directional force field, which has been tailored for simulating hydrogen bonds and metal-ligand interactions, including ligand polarization terms, and to allow for dynamic solvation (switch off the water molecules occupying the ligand position *in situ*) of the binding pocket as well as for the evaluation of hydrogen-bond and hydrophobic saturation.^{7,37} The poses were scored and weighted by means of the 4D viewer.

For the binding poses of the most high-affinity ligands in each class, the kinetic stability and alternative binding modes are challenged by MD simulations. The binding modes of the seed compounds were generated based on these confirmed templates. Next, all energetically feasible binding modes were extracted and used as input for the mQSAR software *Quasar* and *Raptor*.⁵⁶ (Figure 3.24)

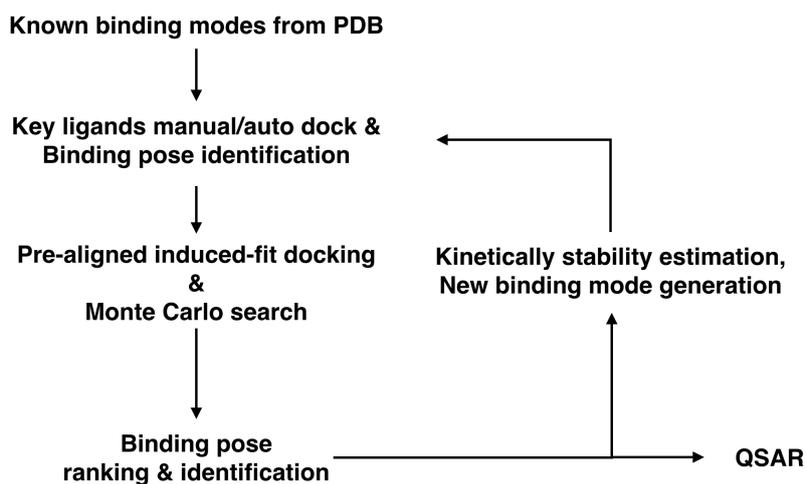


Figure 1.6: PXR QSAR process

1.3 Evaluation of the toxic potential of natural products with anti-trypanosomal activity

1.3.1 Human African trypanosomiasis

Human African Trypanosomiasis (HAT, or “sleeping sickness”) is a fatal disease caused by two distinct subspecies of *T. brucei* (*T. b. gambiense* and *T. b. rhodesiense*) transmitted by the bite of the Tsetse fly (*Glossina* spp., family Glossinidae) in Africa.⁵⁷ The clinical manifestations of this disease depend on the stage of infection and also on the subspecies of the parasite. After the bite of an infected fly, the parasite multiplies in the lymph and blood, causing headaches, fever, malaise, weakness, weight loss, arthralgia, and eventually vomiting and skin lesions. In the latter stages, the parasite crosses the blood-brain barrier, migrates to the CNS and the cerebral spinal fluid, and causes severe neurological and psychiatric disorders, leading to death. According to the WHO, around 60 million people are at risk to contract the sleeping sickness, and *T. b. gambiense* accounts for more than 90% of the disease.

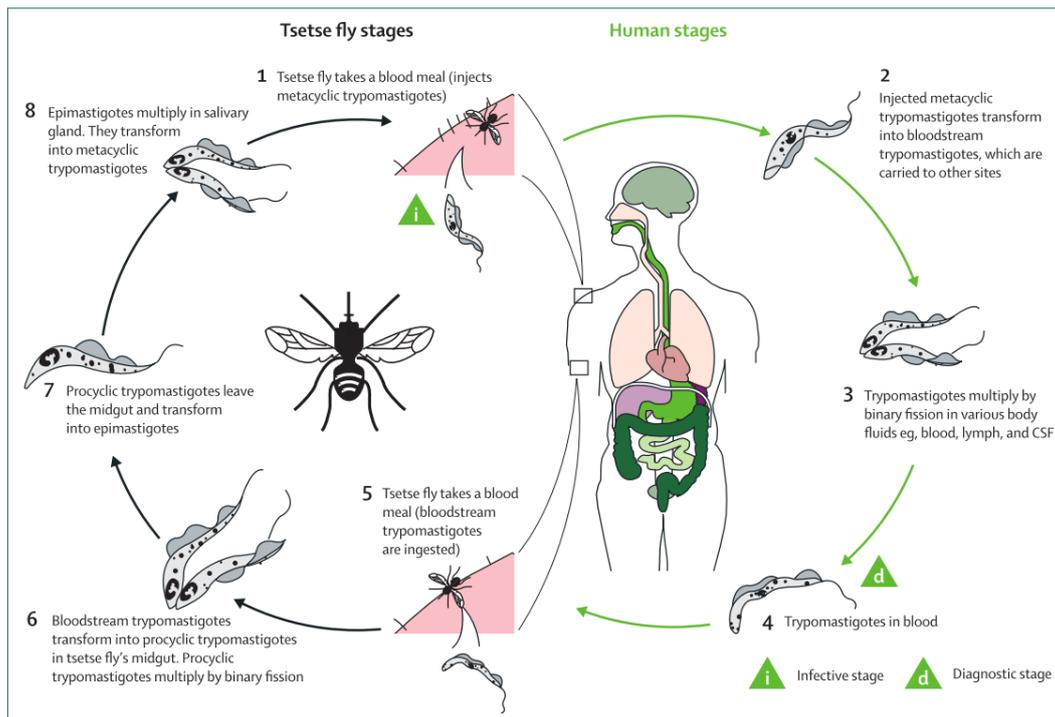


Figure 1.7: Lifecycle of the African trypanosome, image from CDC⁵⁸

According to the WHO, around 60 million people are at risk to contract the sleeping sickness, and *T. b. gambiense* accounts for more than 90% of the disease, which is one of the very few infectious diseases with a mortality rate of 100% if untreated. Currently four drugs are registered for the treatment of sleeping sickness. It usually managed with a combination of suramin (*T. b. rhodesiense*) and pentamidine (*T. b. gambiense*) for the first stage of the disease, prior to CNS involvement, being easier to administer. Eflornithine (*T. b. gambiense*) and melarsoprol (for both *T. b.*) are used in the latter stage of the disease and must cross the blood-brain barrier to reach the parasite, being very toxic and complicated to administer. Suramin, pentamidine, and melarsoprol have unknown mechanism of action, while eflornithine inhibits the enzyme ornithine decarboxylase, involved in the biosynthesis of polyamines required for cell growth. Although these drugs provide the cure of the infection in some cases, they show serious side effects, such as nausea, vomiting, fatigue, renal toxicity, neurological complications (suramin), hypoglycemia (pentamidine), fever, infections, hypertension, diarrhea, and neutropenia (eflornithine).

In the last decade, the screening of more than 700 new and existing nitro-heterocycles led the DNDi to find the fexinidazole with potential to treat advanced-stage sleeping sickness. It is the first new clinical drug candidate in 30 years and after preclinical studies entered as oral treatment in human phase I studies in September 2009.

1.3.2 Natural products active for *T. b. gambiense*

Considering the severe disadvantages of the existing drugs, there is a clear and pressing need for the development of safer and more effective drugs for the treatment of HAT. Many natural products have been reported to show antitrypanocidal activity, including flavonoids, xanthenes, lignans, terpenes, and alkaloids. Hamburger *et al.* recently identified several classes of natural products with antitrypanosomal activity.⁵⁹⁻⁶⁷ These compounds represent a diverse and challenging class of chemicals for *in silico* profiling against adverse effects.

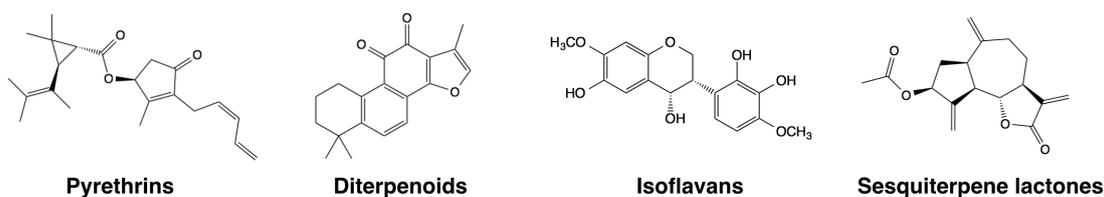


Figure 1.8: representative natural-product scaffolds yield HAT activity

1.3.3 In silico toxic potential profiling by the VTL

The toxic potential of these compounds was estimated employing the VirtualToxLab,³⁸ which simulates and quantifies the binding of a given compound against 16 target proteins known or suspected to trigger adverse effects. As the technology is based solely on thermodynamic considerations, the resulting protein–ligand complexes were subsequently challenged by MD simulations employing *AMBER*,⁶⁸ which allowed probing the kinetic stability of the complexes. If stable, a potential side-effect is likely to occur and the compound could be removed from evaluation pipeline in an early stage of drug development.

Chapter 2

Methods

2.1 4D scoring function development

The philosophy of our modeling efforts to quantify the changes in the free energy associated with small-molecule binding to protein targets is a “mixed-model approach” combined with multi-dimensional QSAR (mQSAR) based on a directional force field. In this approach, feasible binding modes of the small molecule at the protein are sampled using a Monte-Carlo search protocol at the three-dimensional structure of the macromolecular target and comprised into a 4D data set. As the errors associated with energy components extracted from such an entity are larger than the differences in free energies of ligand binding between two related molecules. An afterwards re-scoring process is needed to re-weight the energy components in order to fit binding energy to the experimental data. Such a re-weighted receptor environment should be feasible to generate a quasi-atomistic receptor model for binding affinity prediction.

For all binding modes identified by *Cheetah* (12 poses/ligand), the following quantities were calculated: protein–ligand interaction energy, ligand desolvation energy, ligand entropy and induced-fit energy ($E_{\text{ligand-receptor}}$, $E_{\text{ligand desolvation}}$, $T\Delta S$ and $E_{\text{induced fit}}$ respectively). Therefrom the binding energy was calculated according to the equation:

$$\begin{aligned} E_{\text{binding}} &= E_{\text{ligand-receptor}} - E_{\text{ligand desolvation}} - E_{\text{ligand strain}} - T\Delta S - E_{\text{induced fit}} \\ E_{\text{ligand-receptor}} &= E_{\text{electrostatic}} + E_{\text{van der Waals}} + E_{\text{hydrogen bonding}} + E_{\text{polarization}} \end{aligned} \quad (2.1)$$

The ligand–protein interaction energy was calculated by the *Yeti* force field:

$$\begin{aligned}
E_{\text{ligand-receptor}} = & \sum_{\text{bonds}} K_r(r-r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta-\theta_{eq})^2 + \sum_{\text{torsions}} \frac{V_n}{2}[1+\cos(n\phi-\gamma)] + \\
& \sum_{\text{nb pairs}} \frac{q_i \cdot q_j}{4\pi\epsilon_0 D(r)r_{ij}} + \sum_{\text{nb pairs}} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6}\right) + \\
& \sum_{\text{H-bonds}} \left(\frac{C}{r_{ij}^{12}} - \frac{D}{r_{ij}^{10}}\right) \cdot \cos^2(\theta_{\text{Don-H...Acc}}) \cdot \cos^n(\omega_{\text{H...Acc-LP}}) + \\
& \sum_{\text{metal pair}} \frac{q_i^{CT} \cdot q_j^{CT}}{4\pi\epsilon_0 D(r)r_{ij}} + \sum_{\text{metal pair}} \left(\frac{E}{r_{ij}^{12}} - \frac{F}{r_{ij}^{10}}\right) + \sum_{\text{atoms}} -\frac{1}{2}\alpha_i[\vec{E}_i^0 \cdot \vec{E}_i]
\end{aligned} \tag{2.2}$$

For the 12 docking poses to each ligand, *Cheetah* records the per-residue interaction energy (with both protein and explicit solvent) for each pose by electrostatic, van der Waals, hydrogen-bond and polarization energy contributions as E_{ele} , E_{vdw} , E_{hbd} and E_{pol} . In order to optimize the parameters, current common trait in the free-energy functions comprised a linear combination of terms/descriptors adopted from other scoring functions or force field, with linear weighting coefficients derived by fitting to a training set of complexes with known structure and affinity, such as *gCOMBINE* and *NeoScore*. However, these protocols do not take alternative poses (multiple conformations) into consideration for the ligands, which potentially oversimplified the binding process. In the meanwhile, the interaction energy is mainly contributed by residues within the binding pocket. Generally, the energy contribution from the protein residues could be decomposed into van der Waals interaction, hydrogen bond, electrostatic interaction and polarization energies. These fractions are equally weighted among residues and combined to yield the protein–ligand binding energy, nevertheless, the local environments within the binding pocket are not identical between amino acids. For example, a cooperative hydrophobic effect contributes stronger than isolated ones, the less frequent side-chain rotamers have a reduced contribution than more frequent rotamers. Weighting factors should not be identical among different residues. In order to optimize the weighting factors for each residue, coefficients of each residue-related component were optimized by a genetic algorithm, the re-weighted sum combined with E_{int} was regarded as

$E_{ligand-receptor}$, the contribution of $E_{ligand-receptor}$, ΔG_{solv} and $T\Delta S$ were optimized by multiple-linear regression. Furthermore, the QSAR model is a quasi-atomic surrogate concentered in the binding site, the binding energy components are distance-dependent, energy contribution by residues far from the binding site are neglected, so only contributions of residues within the binding pocket are considered for the binding.

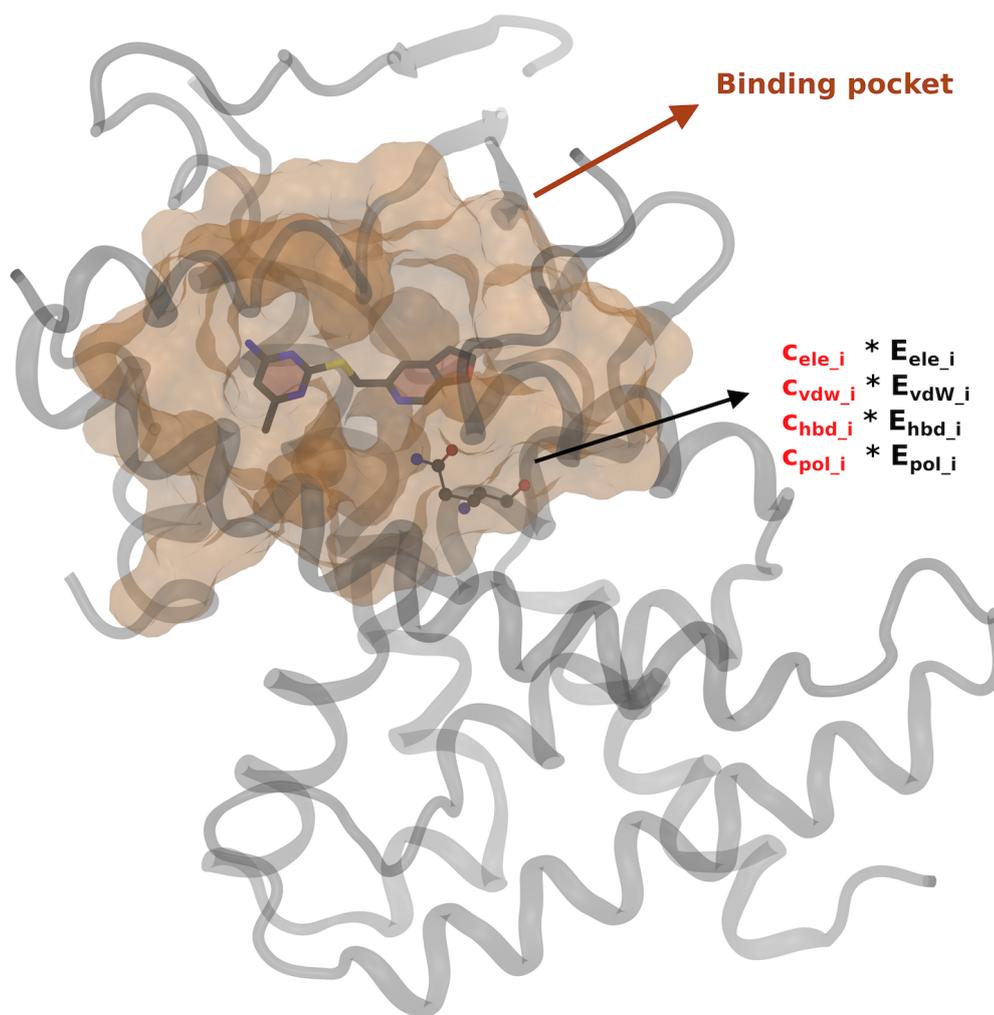


Figure 2.1: Region-defined interaction energy calculation

Genetic algorithms (GAs) are iterative optimization techniques inspired by the natural evolutionary process associated with passing genetic material from parents to their offspring. The basic idea is to randomly generate an initial population, whose members (**genes**) are candidate solutions to the problem (here: a combination of coefficients for all

energy components within the binding site), and evolve that population (**crossover** and **mutation**) under appropriate selection pressure — a fitness function, to obtain a better solution. The process starts by representing the search domain by chromosomes that can be mutated and altered.

The least-squares fitting (used as the fitness function), a mathematical procedure for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets, the coefficient β (4×1 matrix) for $E_{lig-rec}$, ΔG_{solv} , $T\Delta S$ and the slope, is generated by:

$$\beta = (X^T \cdot X)^{-1} \cdot X^T \cdot E_{exp}^T \quad (2.3)$$

X is a $n \times 3$ matrix, n stands for the number of ligands, 3 stands for the 3 energy components: $E_{lig-rec}$, ΔG_{solv} and $T\Delta S$. Therein, the r^2 was represented by:

$$r^2 = \Sigma(E_{cal_i} - \bar{E}_{exp})^2 / \Sigma(E_{exp_i} - \bar{E}_{exp})^2 \quad (2.4)$$

Each gene was components by random numbers between 0.001 and 2.0. For each optimization step, 1,000 genes were randomly generated and the r^2 was calculate and sorted. The top 50 genes were then selected for crossover. If the crossover operation yield a better result, then the new gene was inherited, otherwise the original gene was kept for the next selection round. Crossover and mutation were iteratively performed. If the r^2 doesn't yield better result for 10 iterations, the highest-deviation ligand was referred to as an outlier and removed.

2.2 QSAR model for the PXR

In order to predict the ligand binding affinity toward the PXR, a combination of methods were used to generate the stable ligand binding poses, calculating the binding energies, as well as fitting the results into a QSAR model. The modeling flowchart is illustrated in Figure 2.2:

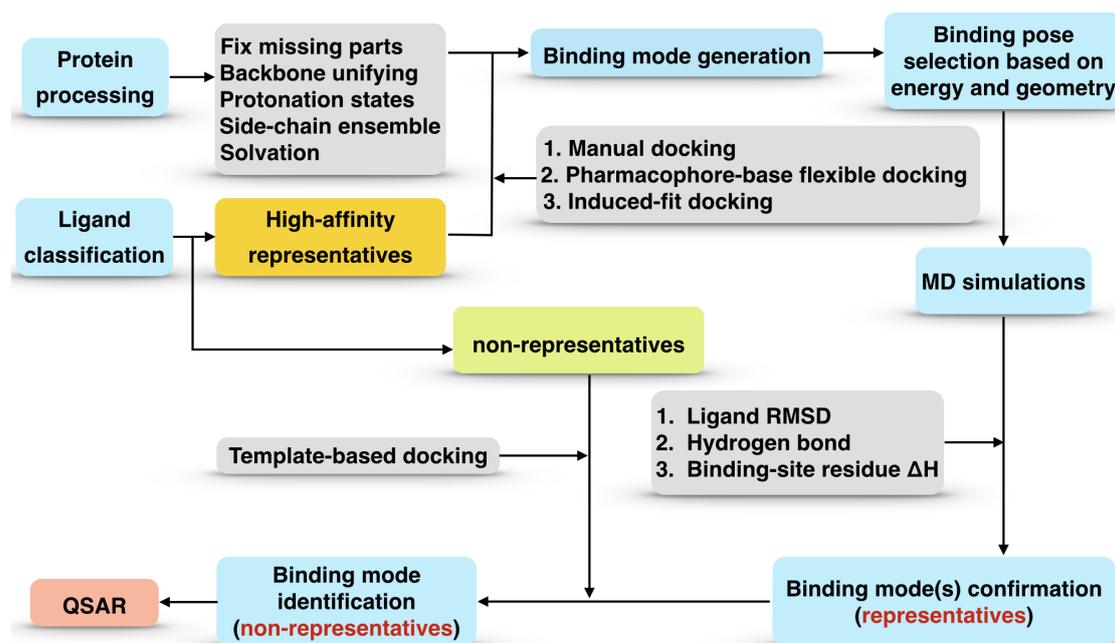


Figure 2.2: The PXR modeling flowchart

2.2.1 Protein processing

PXR can bind various kinds of ligands. In order to consider the pre-organization or ligand induced-fit effect within the binding pocket, multiple crystal structures were used to enrich the side-chain rotamer library within the binding pocket explicitly. The three-dimensional complex structures of the human pregnane X receptor ligand binding domain (hPXR-LBD) were obtained from the Protein Data Bank, with one additional complex (PXR/17 β -estradiol,⁶⁹ not yet deposited with the PDB) was kindly provided by Prof. Redinbo (University of North Carolina), here referred to as PXE2. The PDB structures employed in this project are listed in Table 2.1:

PDB ID	Ligand ID	res(Å)	Description
1M13	HYF	2.15	hPXR-LBD with hyperforin
1NRL	SRL	2.00	hPXR-LBD in complex with SRC-1 and SR12813
2O9I	444	2.80	hPXR-LBD in complex with SRC-1 and T091317
3R8D	PNU	2.80	hPXR-LBD with PNU-142721
4NY9	2Q4	2.80	hPXR-LBD with ligand 2Q4
4XHD	40U	2.40	hPXR-LBD with ligand 40U
PXE2	EST	2.65	hPXR-LBD with estradiol

Table 2.1: PDB structures used for PXR docking, SRC-1: steroid receptor coactivator-1

In general, all systems were pre-processed using *Maestro's Protein Preparation Wizard*: missing side chains were completed by *Prime*; the bond orders were automatically assigned and hydrogens were added accordingly; alternative locations were removed by comparing the positions with the other PDB structures; the ligand's ionization and tautomeric states were generated by *Epik*; the hydrogen-bond network were optimized on the basis of neighboring partners by flip the alternative position of His, Gln and Asn as the X-ray diffraction data for protein could not distinguish the atom type, the hydroxyl group directions were also optimized accordingly; Finally two restraint minimization step were performed to minimize the structure and remove strain: first by fix all the heavy atoms, then by converging heavy atoms to an RMSD of 0.15 Å. All the crystal waters were included in the process.

After the automatic optimization procedure, the system were further optimized by *Yeti*, which features a directional hydrogen-bond force field to optimize the hydrogen bonds (equation 2.2), a solvation protocol to solvate the hydrogen bond donor/acceptor site, as well as various refinement protocols to optimize the side chain orientations individually.

2.2.2 Docking protocol

The ligands were grouped by chemical classes and for each class the most active compound was manually docked into the binding site and minimized with *Yeti*. The minimization of the complex was performed over two steps: first, only the protein around

the ligand (within a distance of 8 Å) was minimized, then the ligand and the protein within the same zone were optimized. An automated pharmacophore-based, flexible-docking procedure was performed with the packages (*MacroModel*, *AMSOL*, *Aquarius*, *Alignator/Dolina*, *Cheetah*, *BzScore*) and analyzed with *VTLViewer4D*. The automated docking protocol is illustrated in figure 2.3.

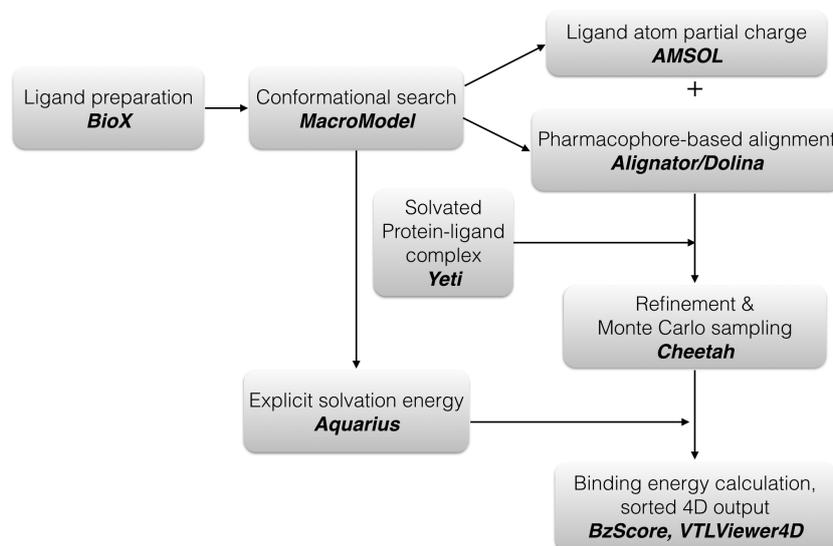


Figure 2.3: The pharmacophore-based, flexible-docking protocol

First the ligand was prepared and minimized with *BioX*, the solvated template protein-ligand complex was generated first by orient solute the binding pocket (no ligand) with *Yeti*, then added the template ligand back into the solvated protein. The ligand conformation was sampled by *MacroModel* and the according solvation energy was calculated by *Aquarius*, the ligand atom partial charges (CM1) were generated by *AMSOL*. The charged-ligand conformations were then aligning to the template's pharmacophores in the binding pocket by *Dolina* and refined as well as sampled by *Cheetah*, including dynamic solvation; at last the generated binding poses were sorted by the Boltzmann-weighted binding energy and the generated 4D binding poses were employed further.

In order to consider the main chain flexibility, as well as using an alternative docking procedure, the induced-fit docking (*IFD*) protocol was also performed to generate putative binding modes. To perform *IFD* designed by Schrödinger, all the water molecules were

removed from the system. The extended sampling protocol was used which could generate up to 80 poses using automatic docking settings, the *IFD* docking poses were ranked by *IFDscore*, which accounts for both the protein–ligand interaction energy (Coulomb term reweighed) and the system energy.

$$IFDscore = Prime_energy + 9.057 \cdot GlideScore + 1.428 \cdot Glide_Ecoul \quad (2.5)$$

The more negative the *IFDscore*, the more favorable the binding. The top 15% docking poses were visually analyzed by checking: 1. atom clash; 2. ligand intra-action; 3. hydrophobic/hydrophilic saturation; 4. hydrogen bond with the key residues (Ser247, Asn285, His324, His407, Arg410); 5. π - π stacking with the aromatic hydrophobic pore (Phe288, Trp299, Tyr306). The confirmed poses were selected for MD simulations.

2.2.3 Molecular-Dynamics Simulations

To analyze the interactions between PXR and its ligands from a dynamic point of view, molecular-dynamics simulations of protein–ligand complexes were performed. For ligands no crystal structures are available in the PDB, the binding modes obtained by manual docking as well as selected posed from automatic docking were both taken into consideration. MD simulations and in particular the analysis of the trajectories allowed for a more dynamic characterization of the protein–ligand interactions responsible for ligand binding and to gain insight into the binding mode(s). The MD simulation flowchart:

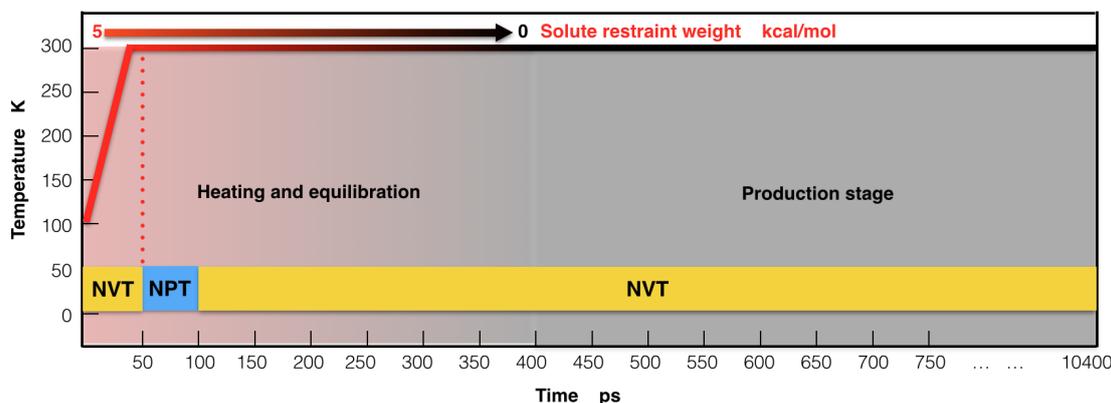


Figure 2.4: The MD simulation equilibration and production flowchart

All MD simulations were performed using the *AMBER12* package and corresponding

all-atom potential function *ff99SB-ILDN*. The *GAFF* force field was used to model the ligands, and the atomic partial charges were determined with *Gaussian09* at the RHF/6-31G* level and fitted by *RESP*. A 2 fs time step and a 9 Å cutoff value were used for truncation of non-bonded interactions. Simulation structures were solvated in an orthorhombic TIP3P water box with periodic boundary conditions at a minimum distance of 10 Å in each dimension from the solute. Counter-ions (Na⁺ or Cl⁻) were added to neutralize the systems. The particle-mesh Ewald (PME) summation method was adopted to treat long-range electrostatic interactions. The SHAKE algorithm was used to constrain all covalent bonds involving hydrogen atoms. Energy minimization was performed in two stages, with each stage employing 250 steps of steepest decent followed by 750 steps of conjugate gradient method using position restraints for the solute, the harmonic restraint weight was set at 25 kcal/mol and 5 kcal/mol for the first and the subsequent minimization step, respectively; following minimization, 40 ps of MD simulation was performed to heat the system from 100K to 300K followed with 10 ps equilibration in a NVT ensemble, the harmonic restraint weight was set at 5 kcal/mol; then the system was switched to NPT ensemble and equilibrated for 50 ps; after the initial equilibration, the system was changed back to NVT and equilibrated for another 6 stages (50 ps each) by gradually reducing the restraint weight of solute from 5 kcal/mol to zero. The production phase was performed at constant temperature (300K) and constant volume for at least 5.0 ns.

2.2.4 Trajectory analysis

For the MD trajectories, the ligand movement was recorded as the most important criteria for binding stability. A stable binding mode should have smaller movement compared with unstable ones during the simulations. Besides the geometric evaluation for protein–ligand interactions, conserved binding free energy is another criteria for binding stability. Hydrogen bond distances and angles, binding free energy contributions of residues in the binding pocket were also recorded for binding-stability check, Based on the results of trajectory analysis. The confirmed “stable” binding poses were used for template-based docking of the remaining compounds in each class.

2.2.5 QSAR studies: Quasar & Raptor

The software *Quasar* and *Raptor* were used to generate receptor models for the PXR. The ligand alignment obtained from the MD-confirmed docking and template-based docking results was used as input to build the QSAR model. The data set was split into a training set, used to build the model, and test set, used to evaluate it, in such a way that a maximal diversity of the training set with respect to binding affinity and chemical properties was obtained. In order to achieve this goal, the compounds were grouped according to their chemical class (i.e. sharing the same scaffold) and were ranked by affinity. For each group, the most and the least active compound was assigned to the training set. From the compounds remaining in the pool, compounds with different scaffolds and functionalities were selected to be part of the training set in order to achieve maximal chemical diversity. For the QSAR simulations, a 4:1 ratio for the training and test set.

In order to develop a model, the ligand alignment along with were added the experimental binding affinities, the solvation energies, the entropy values and the internal strain as calculated by *MacroModel*, were employed as input for the binding affinity.

2.3 Toxic potential estimation by means of the VTL and MD simulations

In the *VirtualToxLab*, the toxic potential of a compound is estimated by simulating and quantifying its interactions towards a series of macromolecular targets at the molecular level using automated flexible docking combined with 4D Boltzmann scoring.³⁸ The technical flowchart of *VirtualToxLab*TM is presented in Figure 2.5. For the anti-trypanosomal compounds identified by Hamburger *et al.* and their metabolites, the binding affinity towards the 16 targets were estimated by the automatic procedure.

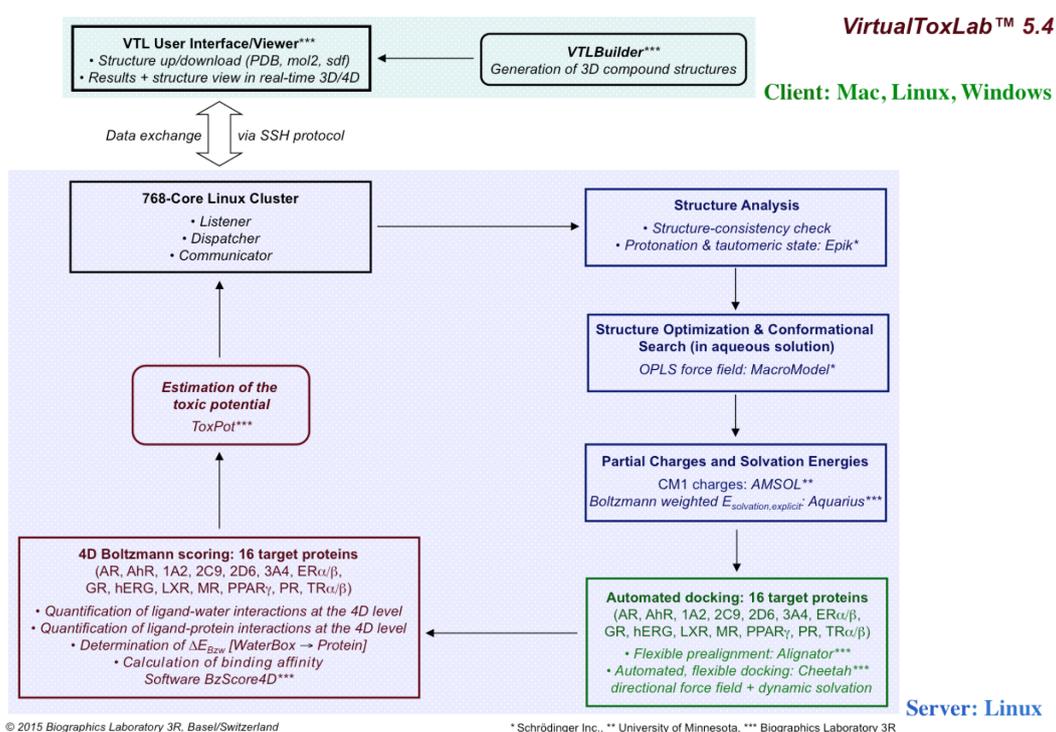


Figure 2.5: The technical flowchart of *VirtualToxLab*¹⁹

In order to provide a reliable *in silico* affinity estimation, the identified high-affinity docking poses generated by VTL were visually checked. For each identified high-affinity pose, the kinetic stability was challenged by MD simulations with the *AMBER12* package (with a minimum of 5 ns production stage).

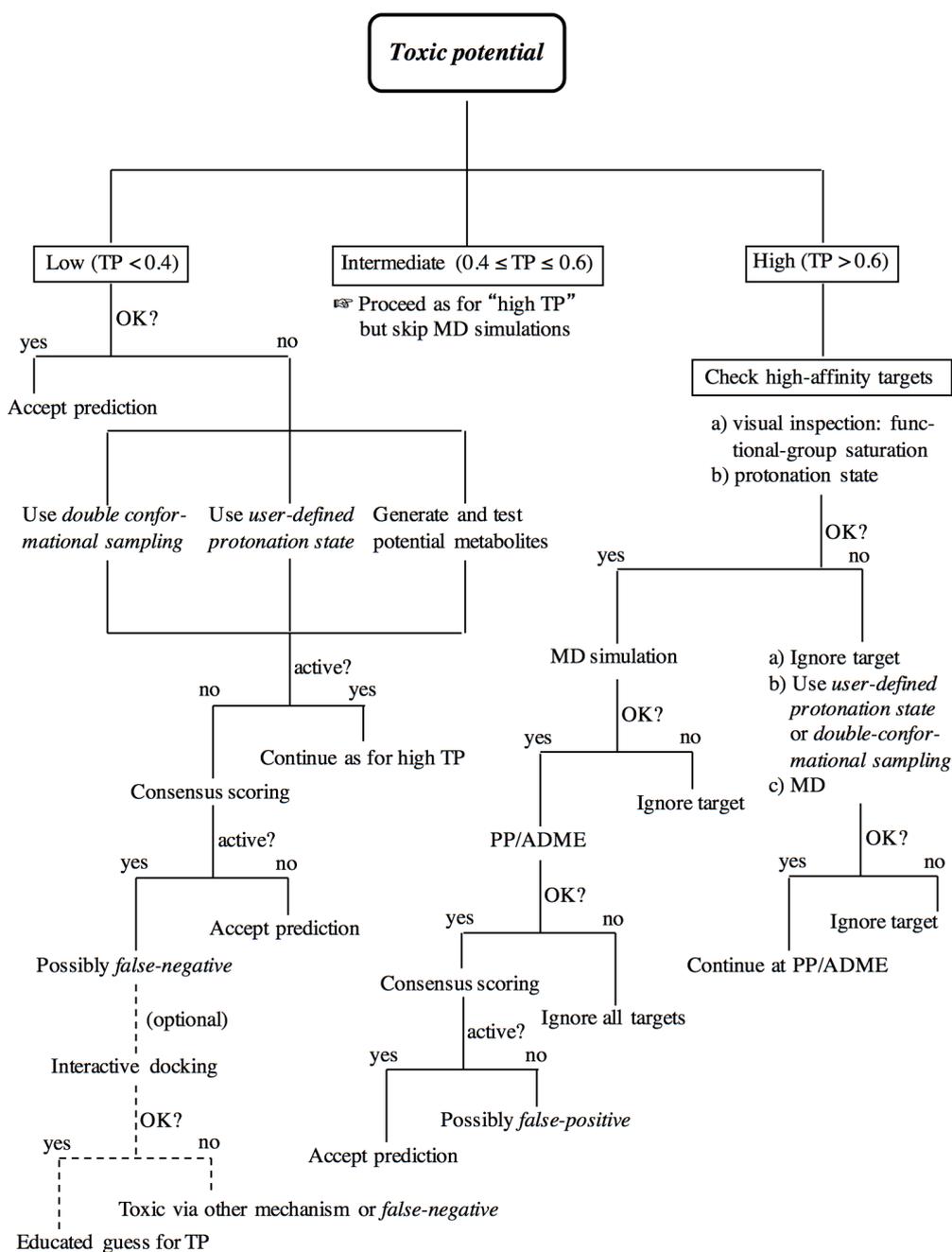


Figure 2.6: VirtualToxLab: decision tree for the verification of a prediction (binding mode, affinity, toxic potential). TP: toxic potential; MD: molecular-dynamics simulation, PP: physico-chemical properties, ADME: adsorption, distribution, metabolism and elimination properties, consensus scoring (CS) using other *in silico* approaches.¹⁹

Chapter 3

Results & Discussion

3.1 Scoring Function development

3.1.1 Flowchart of data processing

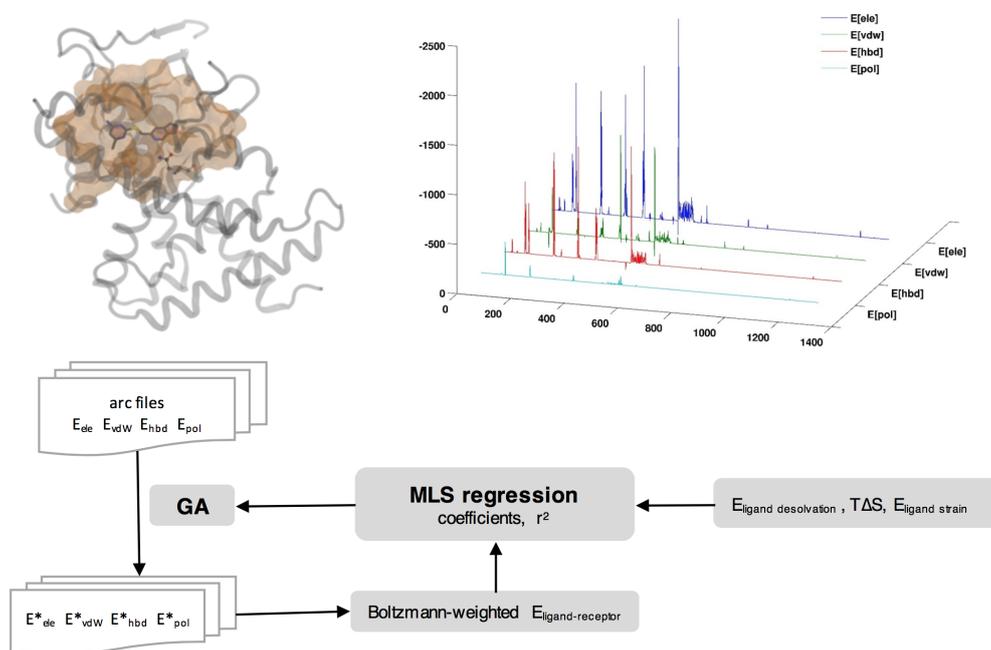


Figure 3.1: Left: per-residue binding energy for the whole dataset. Right: binding pocket.

The r^2 value obtained from multiple-linear regression was used as the fitness-function in the optimization algorithm, which aims at minimizing the least-square deviation between calculated and experimental binding affinities. The results are shown in fig. 3.2. The androgen ($r^2=0.82$), estrogen α ($r^2=0.79$) and liver-X receptor ($r^2=0.68$) yield the highest agreement among the 16 target proteins. For the cytochrome P450 enzymes, CYP3A4 ($r^2=0.58$) and CYP1A2 yield moderate $r^2=0.58$), CYP2D6 and CYP2C9 are poorly fitted ($r^2=0.02, 0.19$). The latter may be due to the fact that the experimental K_{r2} value spans only two orders of magnitude which is not truly discriminative. The aryl hydrocarbon receptor, estrogen receptor β , mineralocorticoid receptor, progesterone receptor, the hERG ion channel and PPAR γ all yield moderate results ($r^2=0.49, 0.35, 0.49, 0.57$ and 0.37), while the glucocorticoid and thyroid receptor α/β ($r^2=0.32, 0.20, 0.23$) did not fit well at this point.

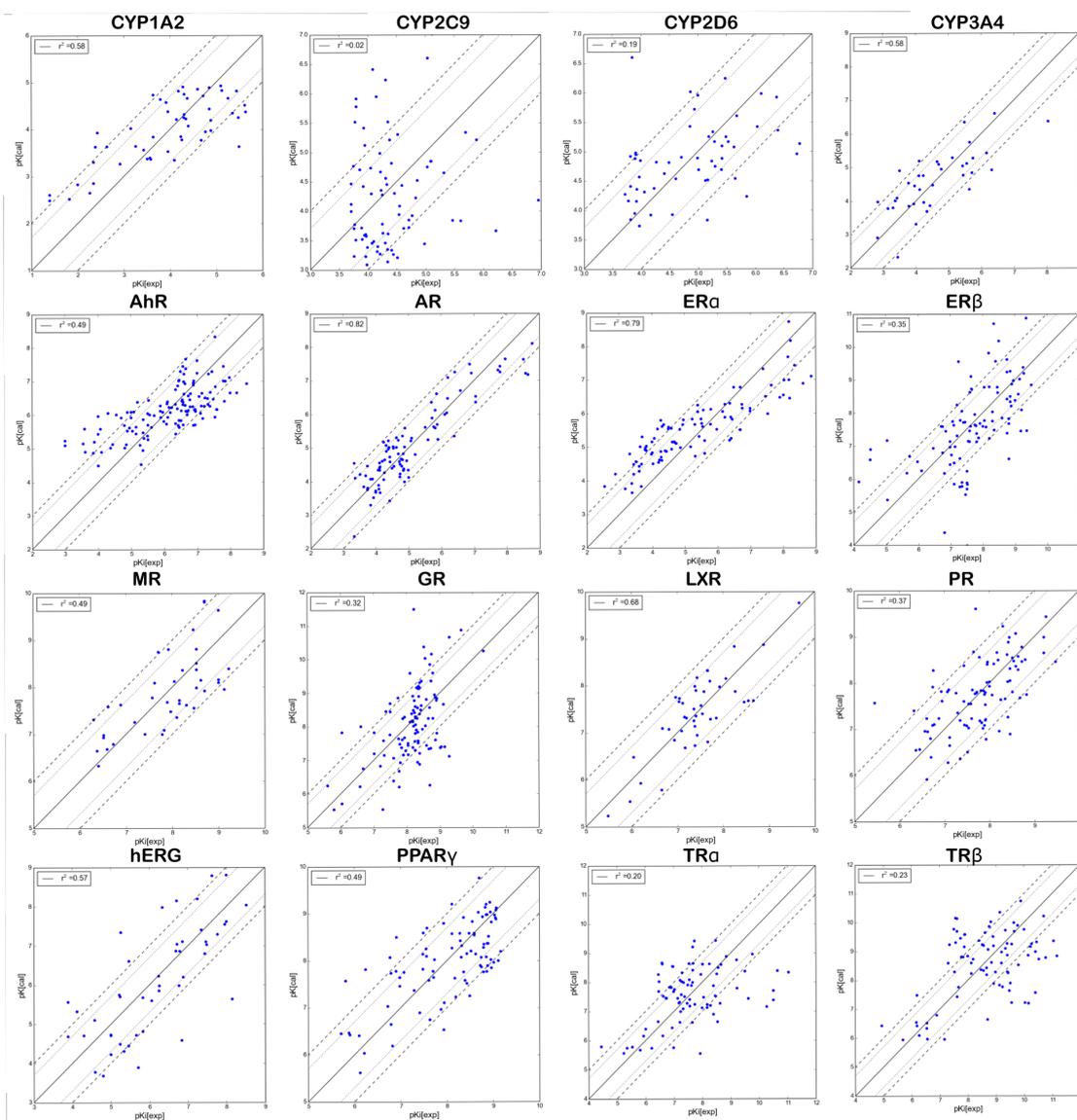


Figure 3.2: Fitting results for the 16 target proteins. Experimental values are shown on the horizontal axis, calculated values on the vertical axis. Dashed lines are drawn at a factor of 5.0 and 10.0 off the experimental value.

3.2 QSAR model development for the PXR

3.2.1 Binding-site analysis

All available X-ray crystal structures of the PXR share a similar backbone conformation. For the structures with a bound small molecule and/or a co-factor, the main-chain C_{α} RMSD values are lower than 0.4 Å compared with the highest-resolution structure (1NRL, resolution 2.0 Å). Consequently, employing a single backbone conformation should be acceptable for the docking studies. On the other hand, the loop defined by Cys207–Val211 displays major deviations in position and further induces a drift in its adjacent α_2 helix, inside of which the backbone of Asp205 acts as hydrogen bond acceptor to 17 β -estradiol (the bound ligand in the PXE2 structure). This change impedes docking 17 β -estradiol to the PXR when using the protein coordinates of 1NRL. Alternatively, by adapting the α_2 helix from PXE2 to 1NRL (fig. 3.3A) and followed with energy minimization, the optimized structure could adopt similar binding modes for 17 β -estradiol as well as most of the other PXR ligands available in crystal structures by interactive docking with *Yeti*^X (fig. 3.4).

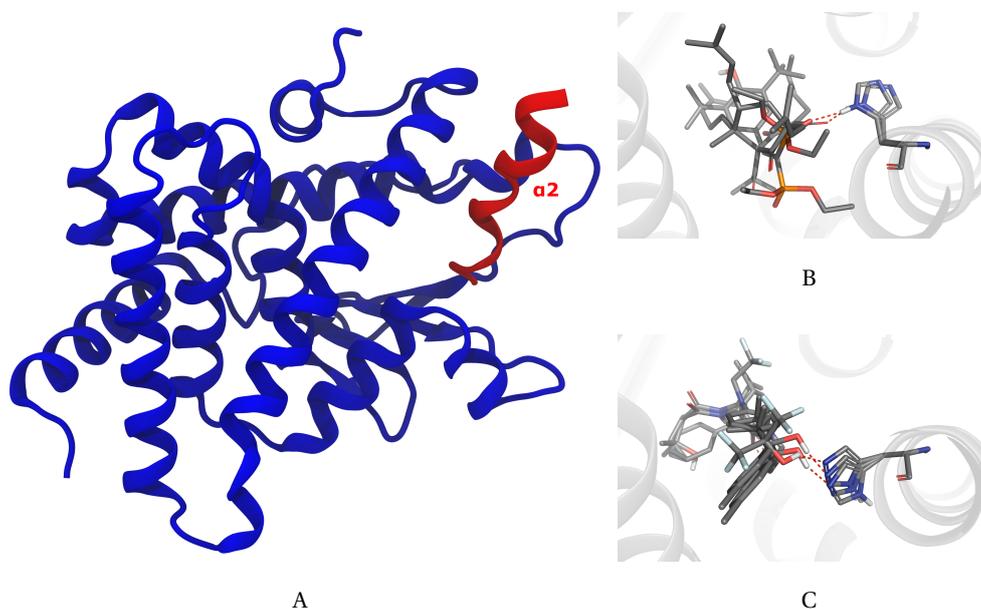


Figure 3.3: A: Hybrid structure of the PXR backbone. The 1NRL portion show in blue, the PXE2 part in red. B and C: His407 act as hydrogen-bond donor (B) and acceptor (C).

After optimizing the hydrogen-bond network by *Maestro's Protein Preparation Wizard* and the software *Yeti^X*: His168, His242, His359, His386 and His418 were protonated at the N_δ; His327 and His336 were protonated at the N_ε according to the potential hydrogen bonds with the local environment. In the binding pocket, His407 can act as both H-bond donor (N_δ: 1M13, 1NRL) and H-bond acceptor (N_ε: 2O9I, 4NYL, 4XHD), therefore only the N_δ was protonated (figs. 3.3B and 3.3C).

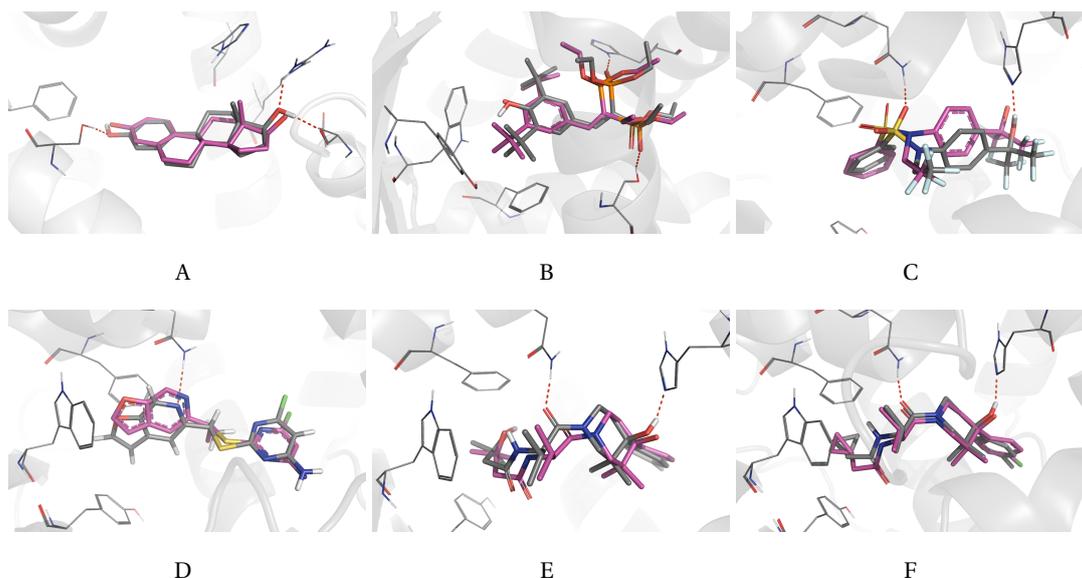


Figure 3.4: Superimposed ligands of the crystal structure (carbon atoms colored in pink) and of the docked poses (carbon atoms colored in grey) with optimized structure. A: 17 β -estradiol, B: SRL12813, C: T091317, D: PNU142721, E: 2Q4 (N-(2R)-1-[(4S)-4-(4-chlorophenyl)-4-hydroxy-3,3-dimethylpiperidin-1-yl]-3-methyl-1-oxobutan-2-yl-3-hydroxy-3-methylbutanamide), F: 40U (N-(2R)-1-[(4S)-4-(4-chlorophenyl)-4-hydroxy-3,3-dimethylpiperidin-1-yl]-3-methyl-1-oxobutan-2-yl-2-cyclopropylacetamide).

3.2.2 MD simulation for the PXR with ligands in the PDB

To analyze the interactions between the PXR and its ligands available in the PDB from a dynamic point of view, and to profile the key residues within the binding pocket, MD simulations of protein–ligand complexes followed with trajectory analysis were performed. A total of four ligands were chosen to comprise representative subset of the studied PXR complexes: **17 β -estradiol**, **T091317**, **PNU142721** and **2Q4** (SRL12813 is not

selected as there is no similar ligands with reported binding affinity to the PXR, **40U** is not selected because it is quite similar to 2Q4). As starting point for the MD studies, for each ligand the binding mode obtained by manual docking with *Yeti*^X was considered. Details of the simulation protocols are reported in the section 2.2.3, in addition, water molecules which are important in the binding pocket were restrained during the equilibration stage.

17 β -estradiol

In the complex of **17 β -estradiol** with the PXR (fig. 3.4A), the hydroxyl group in D-ring (fig. 3.5) engages two hydrogen bonds with the backbone-carbonyl group of Asp205 and the side-chain of Arg410 respectively, the oxygen atom on the hydroxyl group of A-ring is involved in a hydrogen bond with Ser247. Upon challenging the binding by 50.0 ns MD simulation: the ligand RMSD remained within 1.5 Å throughout the simulation (fig. 3.6A), the two hydrogen bonds of the ligand with Asp205 and Arg410 are partly retained, the hydrogen bond between the oxygen atom on A-ring's hydroxyl group and Ser247 was lost. Instead, the hydrogen atom on the same hydroxyl group engaged a new hydrogen bond with Ser247, and Ser247 was further stabilized by a hydrogen bond with Met243. The MM-PBSA results indicate that the three hydrogen-bonding residues are the main contributors for the binding free energy (fig. 3.6C). These findings suggest that the docking pose is kinetically stable and Asp205, Arg410 and Ser247 are the key residues for the binding.

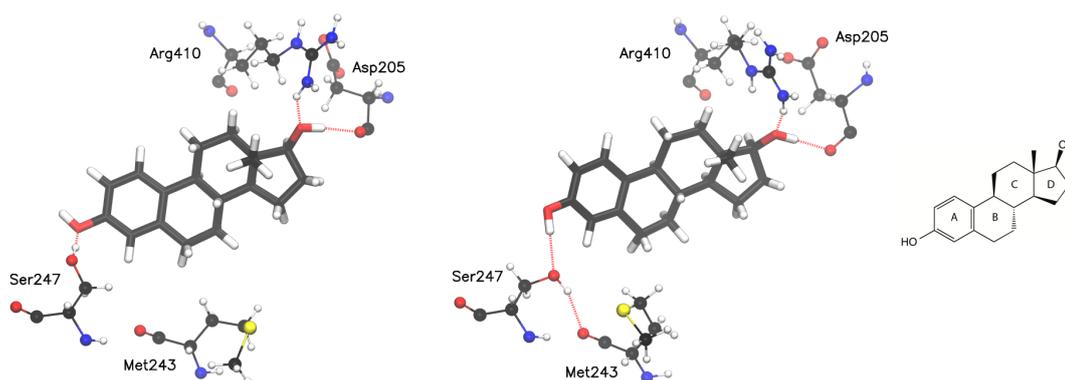


Figure 3.5: Details of 17 β -estradiol binding to the PXR. Left: docking pose; right: MD snapshot.

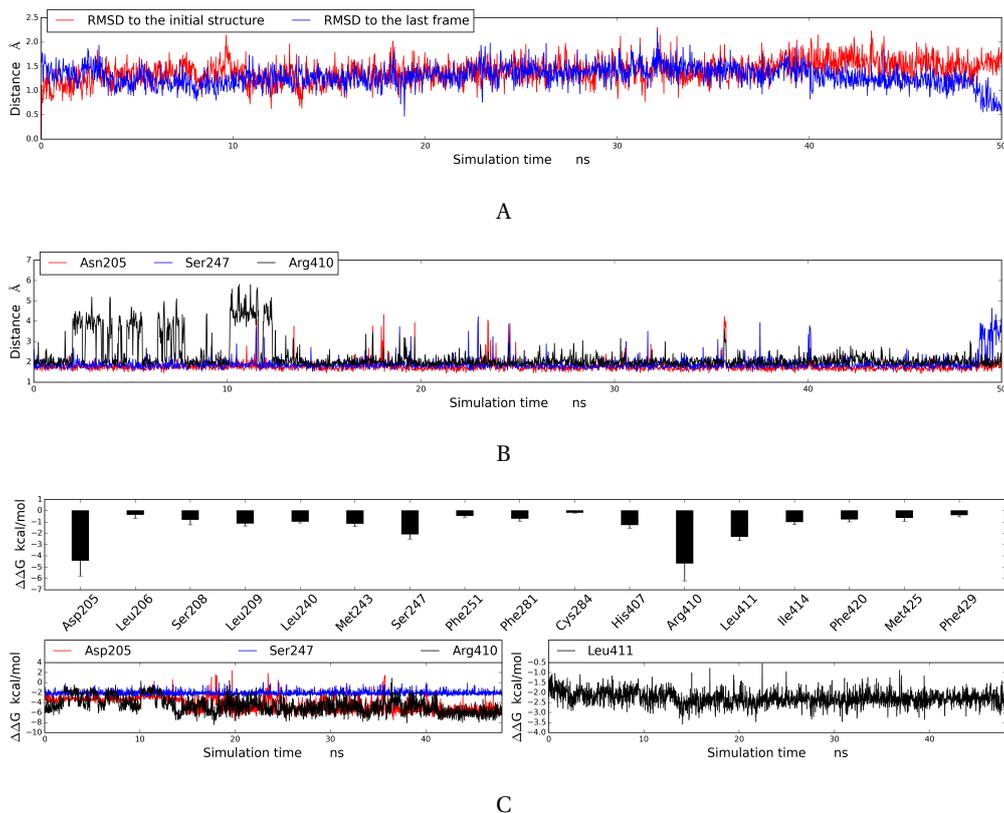


Figure 3.6: A: Ligand RMSD 17 β -estradiol to the PXR. B: Key-residue hydrogen bond distances for 17 β -estradiol with the PXR. C: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues of the PXR binding with 17 β -estradiol.

T091317

In the complex of **T091317** with the PXR, the ligand's hydroxyl group is stabilized by engaging a hydrogen bond with His407, the sulfonyl group is stabilized by a hydrogen bond with Gln285 (which is further stabilized by His327), the sulfonyl-benzyl group is accommodated by the hydrophobic pocket formed by Phe288–Trp299–Tyr306. Upon challenging the binding by a 50.0 ns MD simulation, the two hydrogen bonds are retained in 85% of the time. The ligand RMSD is within 2.0 Å. The results suggest that the docking pose is kinetically stable, Gln285 and His407 turned out to be key residues for the binding, the Phe288–Trp299–Tyr306 pocket is important for hydrophobic interactions (especially for π - π stacking).

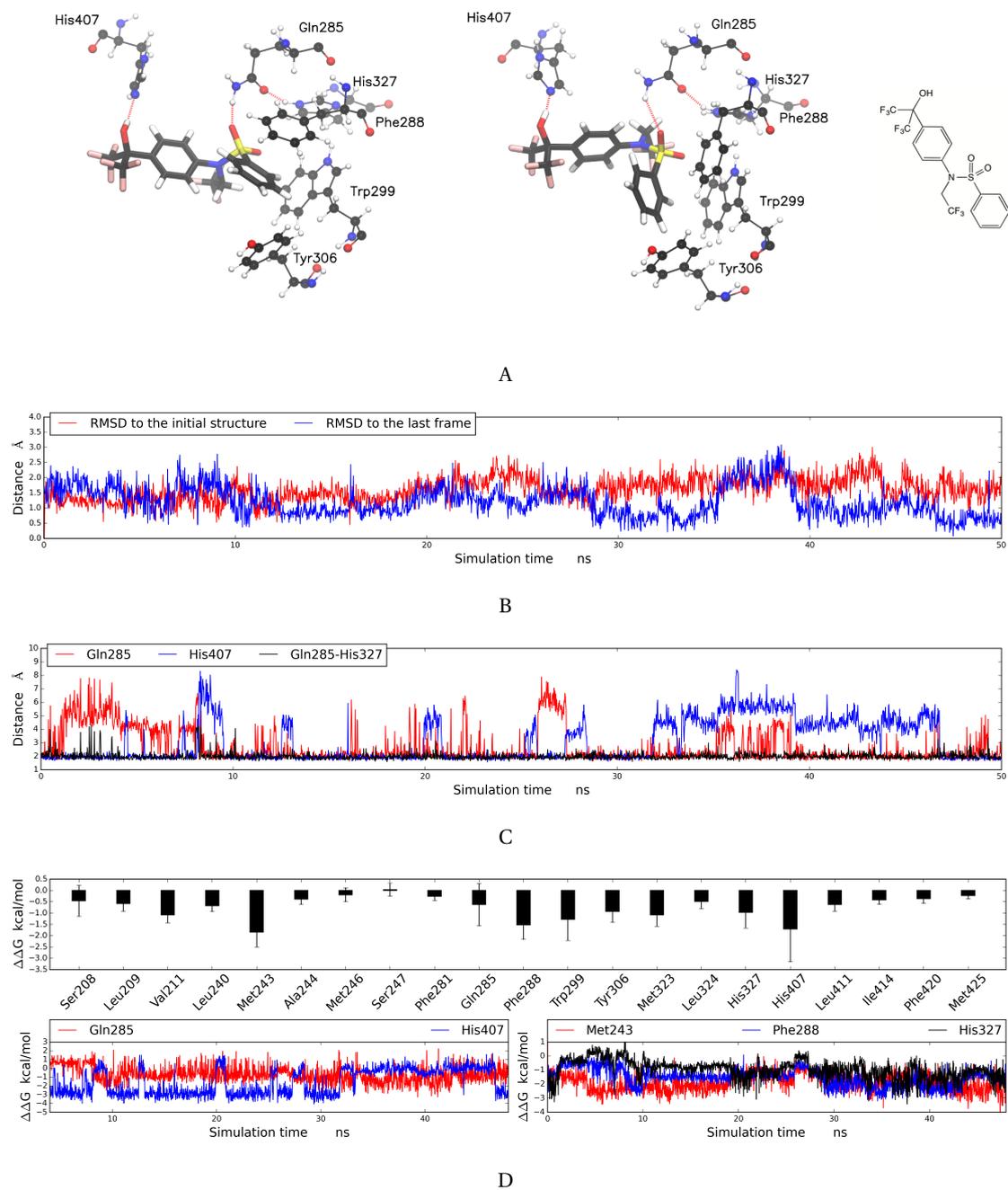


Figure 3.7: A: Details of T091317 binding to the PXR. Left: docking pose; right: the last MD frame. B: Ligand RMSD. C: Key-residue hydrogen bond distances. D: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues.

PNU142721

In the complex of **PNU142721** with the PXR (fig. 3.7A): the furo[2,3-*c*]pyridine part features hydrophobic interactions with the pocket Phe288–Trp299–Tyr306, the nitrogen atom on the furo[2,3-*c*]pyridine part engages a hydrogen bond with Gln285. In contrast, there is no specific interactions with the pyrimidine part, which contains higher b-factors in the PDB and refers to weaker stabilization by the protein. After challenging the binding by 50.0 ns MD simulation: the RMSD values of the furo[2,3-*c*]pyridine part retains within 1.5 Å throughout the simulation (fig. 3.8B), the pyrimidin part rotated its direction and engaged a second hydrogen bond with the backbone of Ser208. The hydrogen-bond distance between the ligand and Gln285 is retained (fig. 3.9A), which is similar to **T091317**. The MM-PBSA results suggest that Gln285, Met243, Phe288 and Tyr306 are the key contributors for the binding free energy (fig. 3.9B).

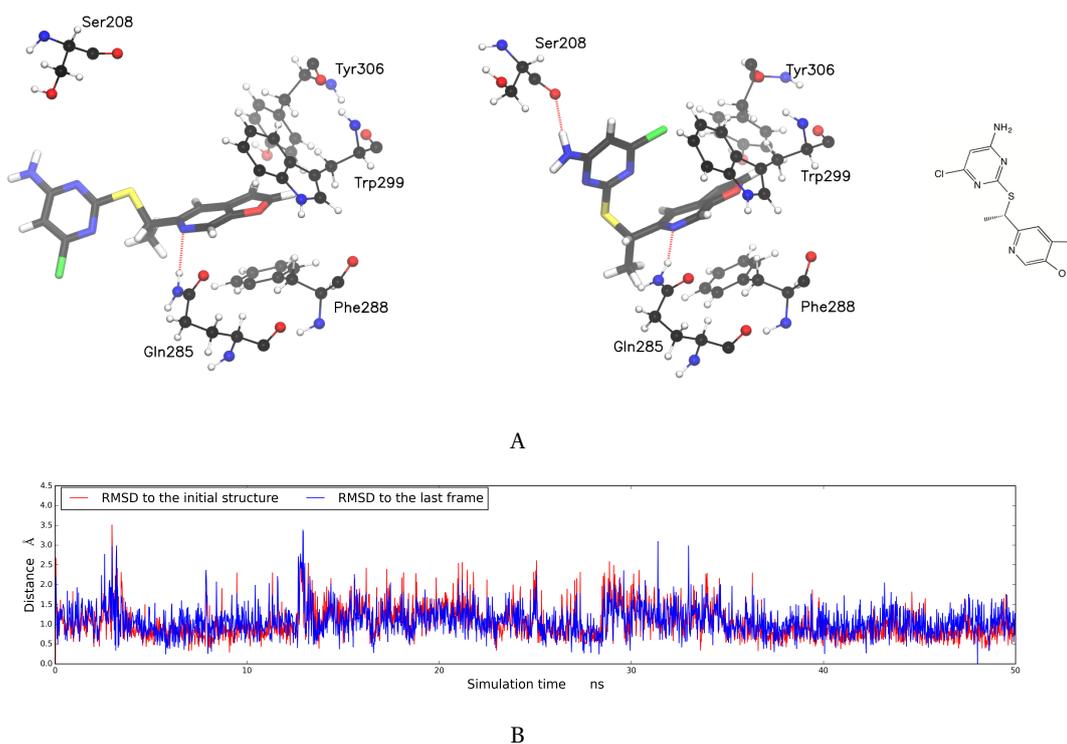


Figure 3.8: A: details of PNU142721 binding to the PXR. Left: docking pose; right: MD snapshot. B: ligand RMSD of PNU142721 to the PXR.

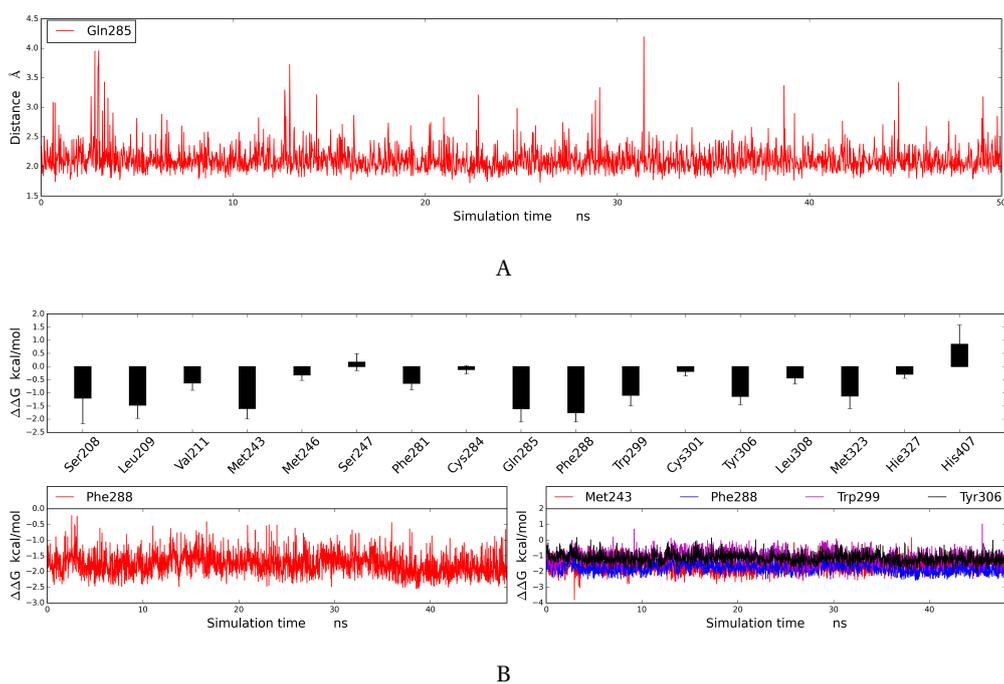


Figure 3.9: A: key-residue hydrogen bond distances. B: binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues.

2Q4

In the complex of **2Q4** with the PXR, the hydroxyl group on the 6-membered ring engages a hydrogen bond with His407, the carbonyl group linking to the 6-membered ring is involved in another hydrogen bond with Gln285. After challenging the binding by 50.0 ns MD simulation: the ligand RMSD is within 2.0 Å throughout the MD simulation, the two hydrogen-bond distances are retained and His407 is the residue contributes most to the binding free-energy. The docking pose is kinetically stable.

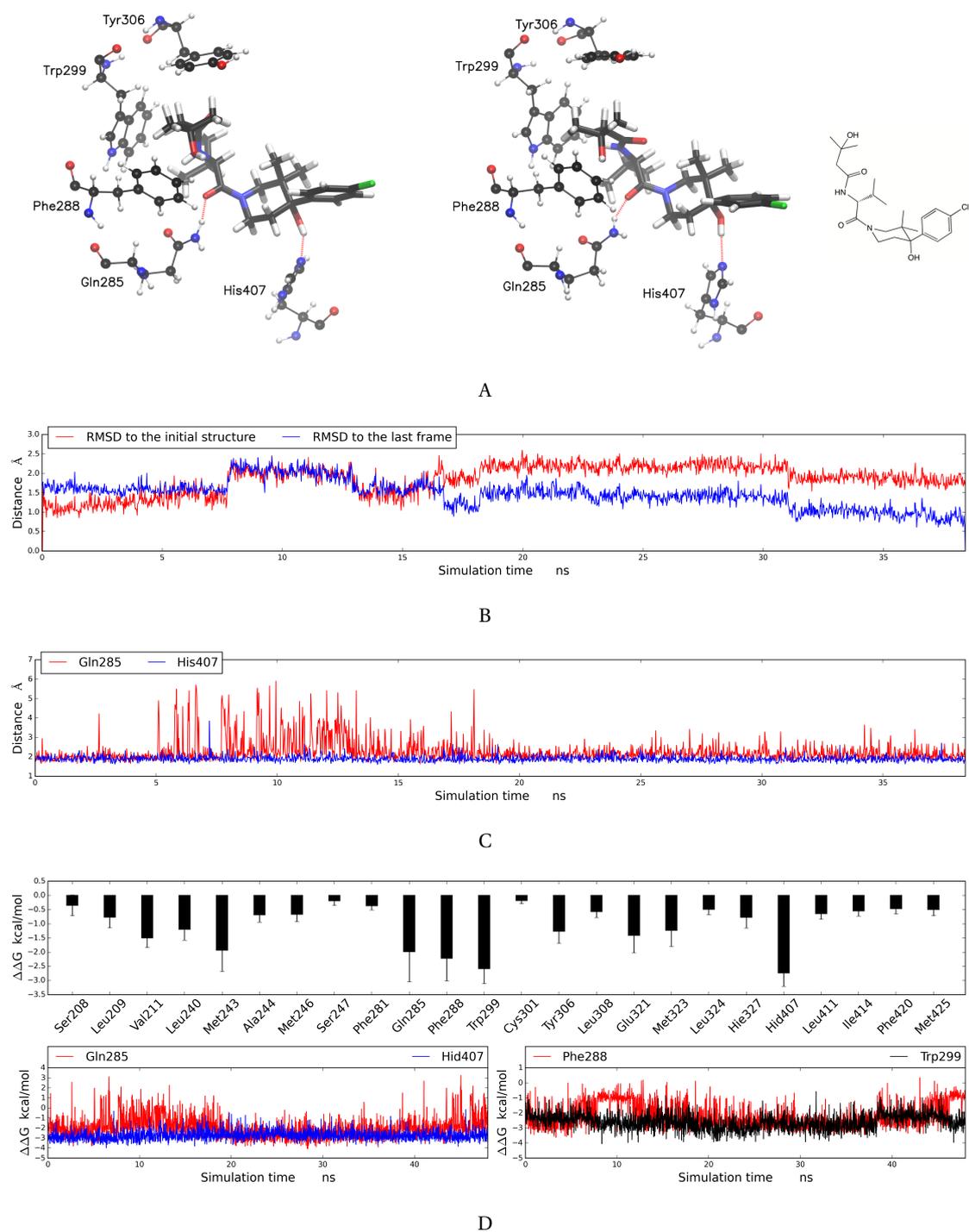


Figure 3.10: A: Ligand RMSD of 2Q4 to the PXR. Left: docking pose; right: MD snapshot. B: Key-residue hydrogen bond distances for 2Q4 with the PXR. C: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues of the PXR binding with 2Q4.

3.2.3 Retrieval of the binding affinity data

The pharmacological data for the 101 PXR-binding compounds was obtained from multiple sources. The experimental EC₅₀ values range from 0.71 nM to 72.4 μM. The majority of affinities, however, lies within two orders of magnitude (10⁻⁶–10⁻⁴M, fig. 3.11). EC₅₀ values were converted in K_i values for the binding-energy calculation.

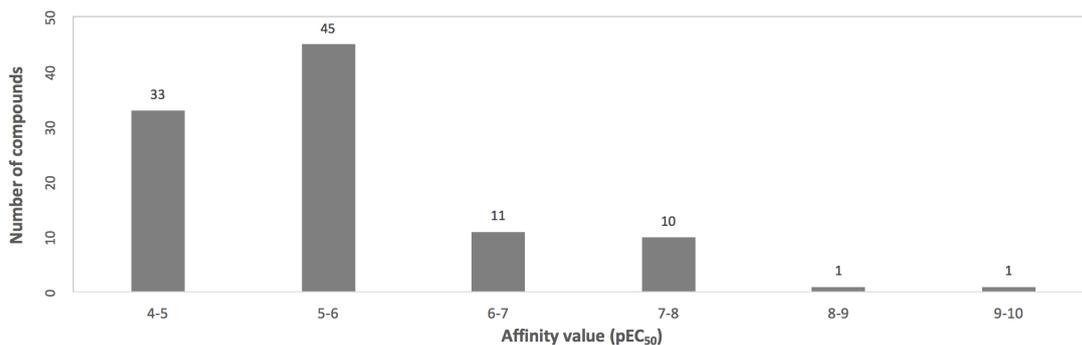


Figure 3.11: Distribution of pEC₅₀ values: 77% of the affinities cluster within two logarithmic units, while the whole data set spans six orders of magnitude (0.71 nM–72.4 μM)

The ligands comprise seven different chemical classes (fig. 3.12), three of which (E, T, S) have crystal structures deposited with the PDB.

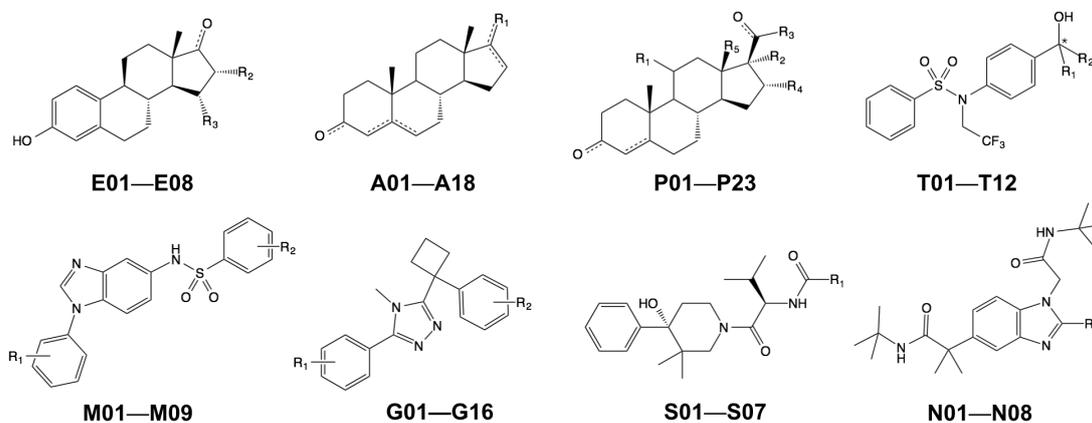


Figure 3.12: Eight classes of compounds used in the QSAR study. All 101 structures are given in Appendix A.

3.2.4 Docking to the PXR

In order to obtain realistic binding mode(s) towards the PXR, a combination of protocols was adopted for the docking process. First, the ligand binding modes from known structures (in PDB) and their respective stable state after the MD simulation were used as templates for derivative compounds to be docked interactively with *Yeti^X*. Then, the remaining ligands were assigned to one class each and several automated docking protocols (*AutoDock/Cheetah*, Schrödinger's QPLD and *IFD*) were used to sample and rank the potential binding modes, the ligand conformation compiled from the first step on was used as boundary criteria for the selection of potential binding mode(s), relevant hydrogen bonds and hydrophobic interactions with the Phe288–Trp299–Tyr306 pocket were considered as key factors to select the most potentially correct binding mode(s). Finally, the identified stable binding mode(s) of the representative compounds were used as templates for their respective derivative ligands and interactively docked with *Yeti^X*.

Crystal structure-based docking

For compounds which contain derivative structures bound to the PXR in the PDB, interactive template-based docking was performed, wherein the binding modes from both the crystal structure and the MD simulations were employed as templates. Protein structures with different bound ligands were superposed to the hybrid protein and its side-chain residues within the binding pocket were interactively adjusted to adapt to the very ligand in crystal structures or MD frames. For each ligand, the geometry was optimized by *MacroModel* and atomic partial charges were calculated with *AMSOL*. The ligands in each class were automatically superposed to their templates and interactively docked to their respective optimized form of the hybrid protein (*Symposar* and *Yeti^X*).

E02–E11 share a common ring scaffold with **E01** (17 β -estradiol). The crystal structure and MD/MM-PBSA results of the PXR-**E01** complex suggest that key hydrogen-bonding residues (Ser247, Arg410 and Asp205) are the main contributors for the binding (figs. 3.5 and 3.13A). Although the ring scaffold was slightly rotated, its position based on the anchored hydroxyl group (by Asp205 and Arg410) in D-ring, the hydrogen bond between Ser247 and hydroxyl group in A-ring were retained during the simulation. Superposition of the docking results are shown in fig. 3.13B.

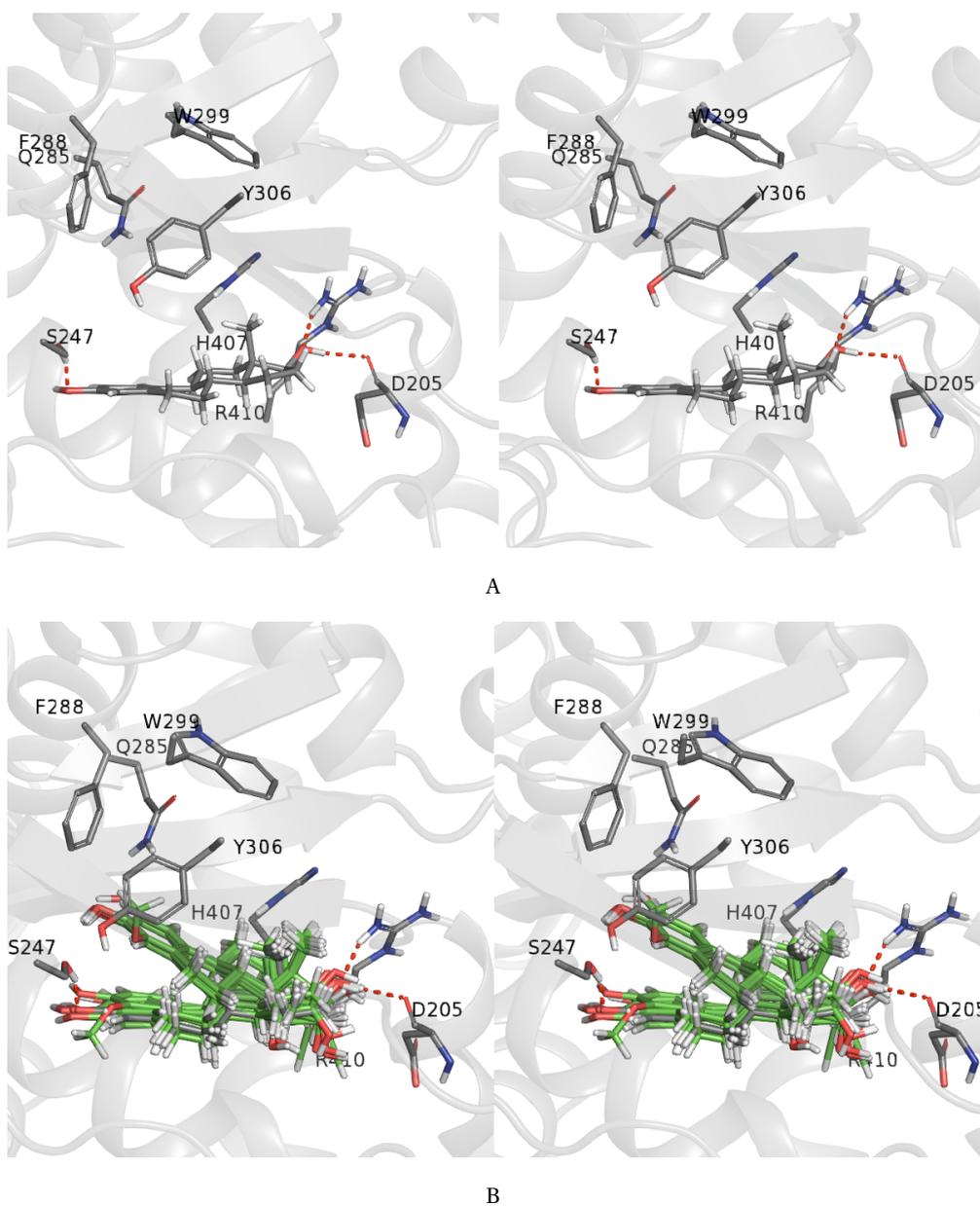


Figure 3.13: A: binding mode of **E01** (17β-estradiol) to the PXR in stereo view. The ligand has been completed with hydrogens. The protein is represented as cartoon, key amino-acid residues and the ligand as sticks. B: the Docking results for **E01–E08**. The carbon atoms of the ligands are colored in green.

T02–T12 are derivative compounds of **T01** (T091317). The template structure engages two hydrogen bonds with Gln285 and His407, the sulfonyl-benzyl group is stabilized through

the hydrophobic pocket formed by Phe288-Trp299-Tyr306. These interactions are retained for all derivative compounds as shown in fig. 3.14.

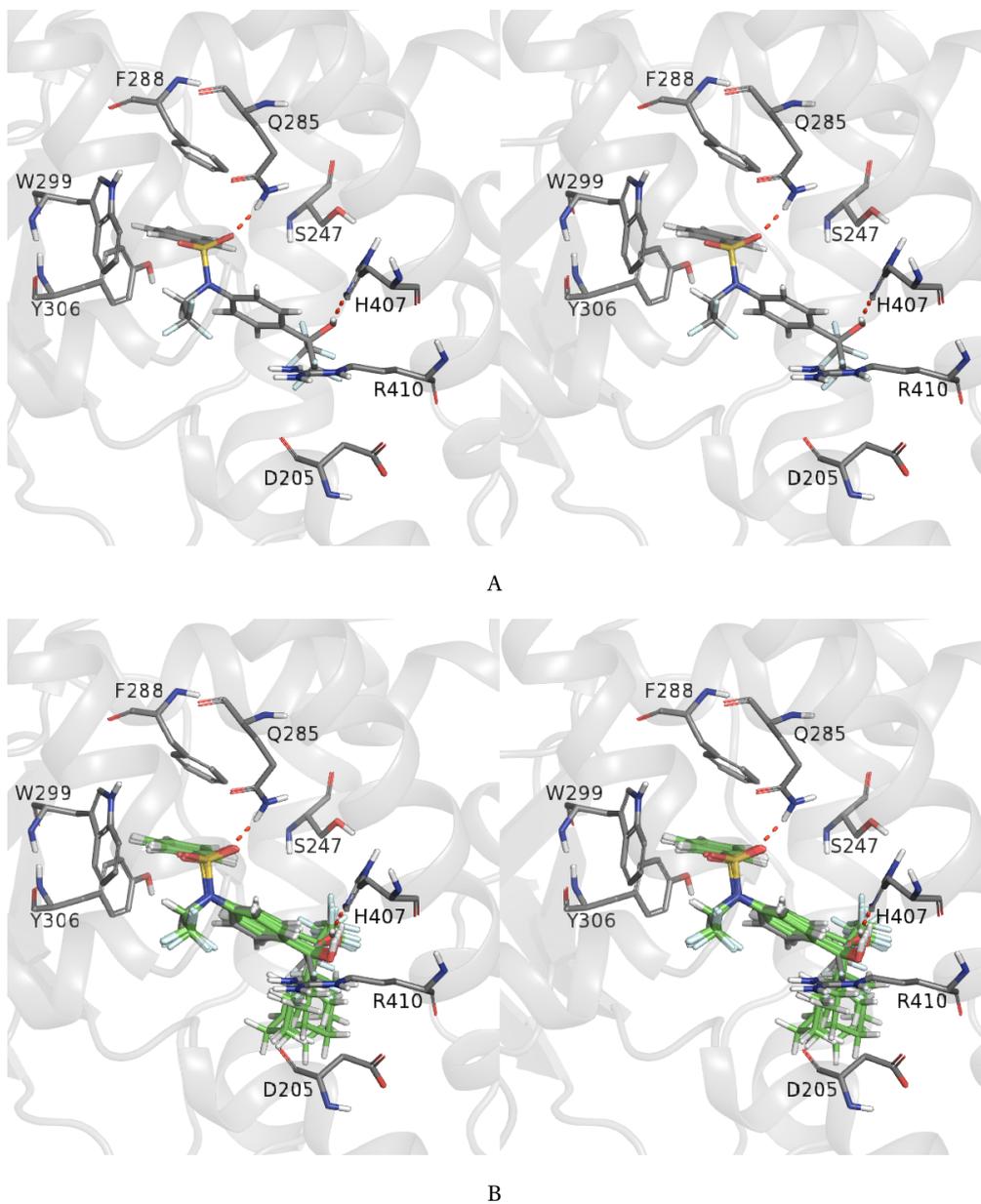


Figure 3.14: A: binding mode of **T01** (T091317) to the PXR in stereo view. The ligand has been completed with hydrogens. The protein is represented as cartoon, key amino-acid residues and the ligand as sticks. B: the Docking results for **T01–T12**. The carbon atoms of the ligands are colored in green.

Similarly to **T091317**, **2Q4** also features hydrogen-bond interactions with Gln285 and His407, as well as hydrophobic interactions towards the Phe288-Trp299- Tyr306 pocket. The superposed docking results of **S02–S25** are shown in fig. 3.15.

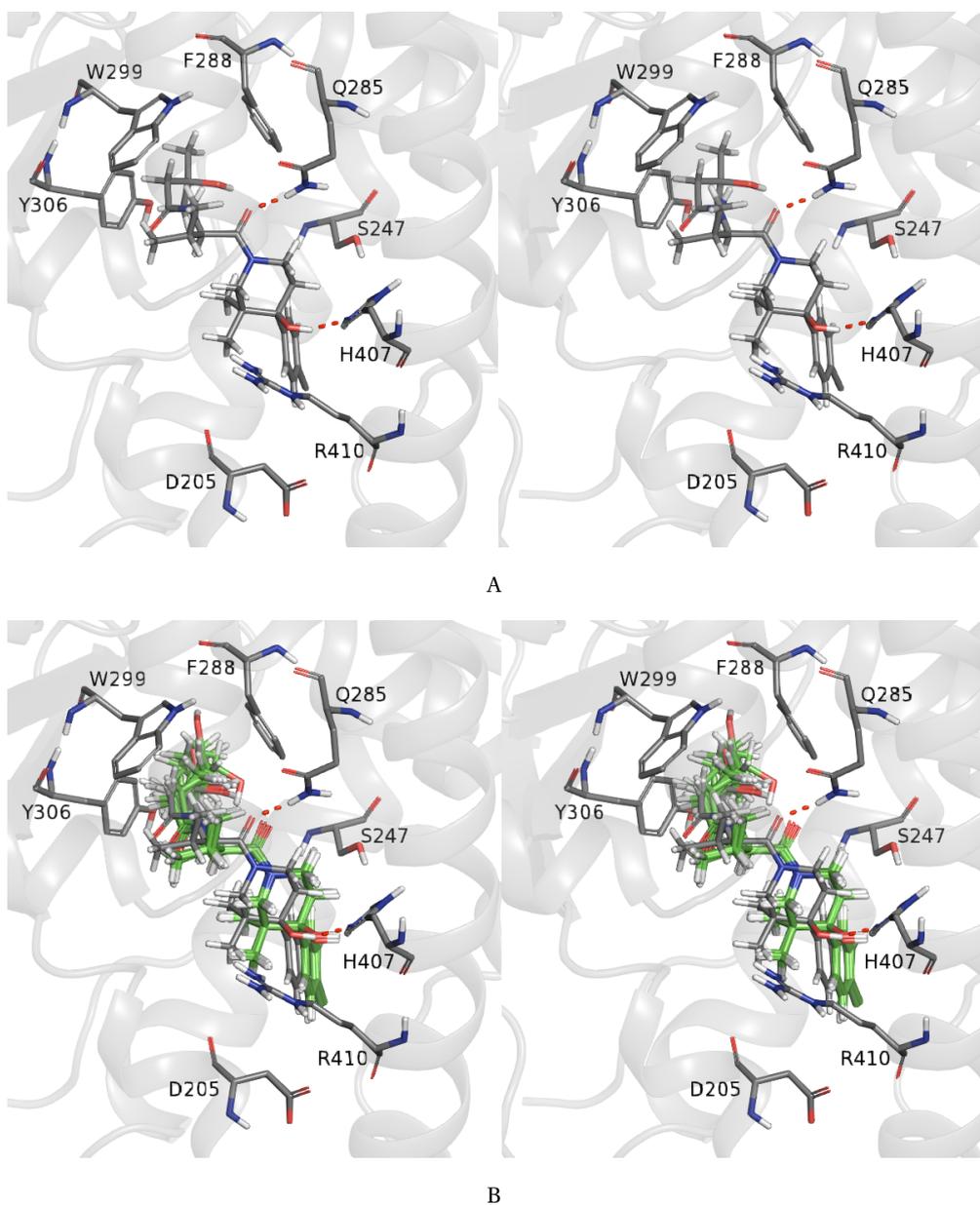


Figure 3.15: A: binding mode of **S01** (**2Q4**) to the PXR in stereo view. The ligand has been completed with hydrogens. The protein is represented as cartoon, key amino-acid residues and the ligand as sticks. B: the Docking results for **S01–S25**. The carbon atoms of the ligands are colored in green.

Ensemble docking and MD simulation

For compounds lacking a similar ligand in the PDB bound to the PXR, automated docking approaches (*AutoDock/Cheetah*, Schrödinger's QPLD and *IFD*) were performed to sample and rank the potential binding modes for the representative ligands in each class (**A01–A18**, **P01–P23**, **M01–M07**, **B01–B09**). Then the ligand conformations compiled from the first step on were used as boundary criteria for the selection of potential binding mode(s), hydrogen bonds with crucial residues (Gln285, His407, Ser247, Asp205 and Arg410) and hydrophobic interactions with the Phe288–Trp299–Tyr306 pocket were considered as key factors to select the most realistic binding mode(s). The kinetic stability for the selected binding modes was challenged by MD simulations. Finally, the identified stable binding mode(s) of the representative compounds were used as templates for their respect derivate ligands and interactively docked with *Yeti^X*.

A: Case Study 1 — Docking of 5 β -androstan-3 α -ol-17-one (steroid)

For the steroid compounds, the top-ranked poses as generated by *Schrödinger's IFD* mainly occupy the hydrophobic pocket defined by Phe288–Trp299–Tyr306. However, 17 β -estradiol does not occupy this pocket (PXE2), it features hydrogen bonds with Asp205, Ser247 and Arg410. Such interactions were not observed in any of the *IFD* poses. These results suggest that although the *IFD* protocol trends to yield high-ranking order for hydrophobic interactions, it is not truly suitable for generating and ranking poses in which the hydrogen-bond interactions play a central role for the binding, such as steroids to the PXR. On the other hand, the automatic docking protocol implemented in *Cheetah*, which employs a directional force field, combined with a template-based alignment protocol, showed a top-ranking order for the pose similar to the crystal structure. For example, 5 β -androstan-3 α -ol-17-one, which displays a appreciable binding affinity toward the PXR ($EC_{50}=20.0\text{nM}$), the top-ranked binding pose shows good agreement (RMS=) compared with its template 17 β -estradiol. The androstan is stabilized by generating hydrogen bond with Asp205 and Arg410. The kinetic stability of the binding pose was challenged by means of a 10.0 ns MD simulation. The ligand remained in its original position and orientation, and the hydrogen bonds with Gln285 and Asp205 were retained. As the hydrophilic part in some of the ligands are missing, top-ranked docking poses with both programs were accepted for template-based docking in order to consider both hydrophobic interactions and hydrogen bonds.

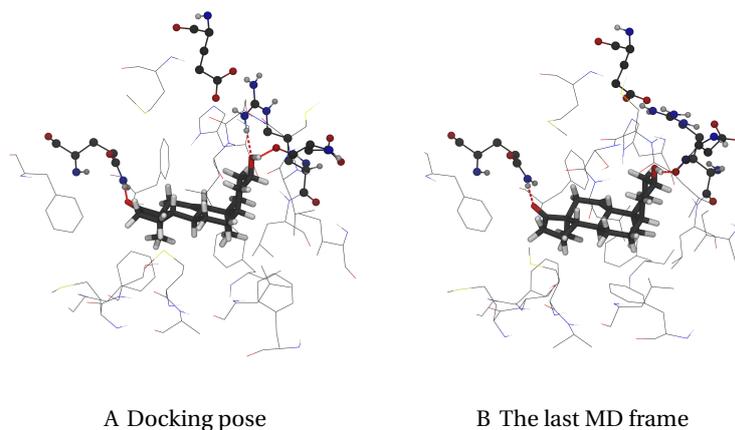


Figure 3.16: Ligand binding of 5 β -androstan-3 α -ol-17-one to the PXR.

A01–A04 contain a L-shaped ring conformation (ring A and ring B are linked in cis-conformation) and two types of binding modes were identified from the docking results (fig. 3.17). In the first type (fig. 3.17A), His407 and Arg410 could act as hydrogen-bond donor for the ligands; In the second type (fig. 3.17B), the 3-hydroxyl group is stabilized by engaging hydrogen bonds with Asp205 and Arg410, the 17-carbonyl group engaged a hydrogen bond with Gln285.

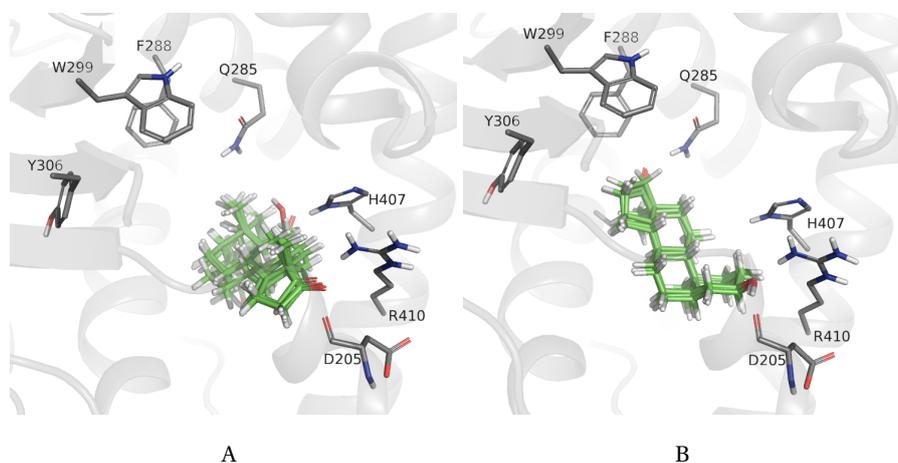


Figure 3.17: Docking results for **A01–A04**.

A05–A18 contain a flat 6-6-6-5 ring conformation (ring A and ring B are in

trans-conformation) and mainly contain four types of binding modes depending on their respect hydrophilic functional groups in each compound (fig. 3.18). The first and second types (fig. 3.18 A and B) are similar to the two binding modes of 17 β -estradiol, which features hydrogen bonds towards Ser247, Asp205 and Arg410. The third and fourth types contain only one hydrogen bond towards Ser247 and His410 respectively, unlike type A and B, they also engage hydrophobic interactions towards the Phe288- Trp299-Tyr306 pocket.

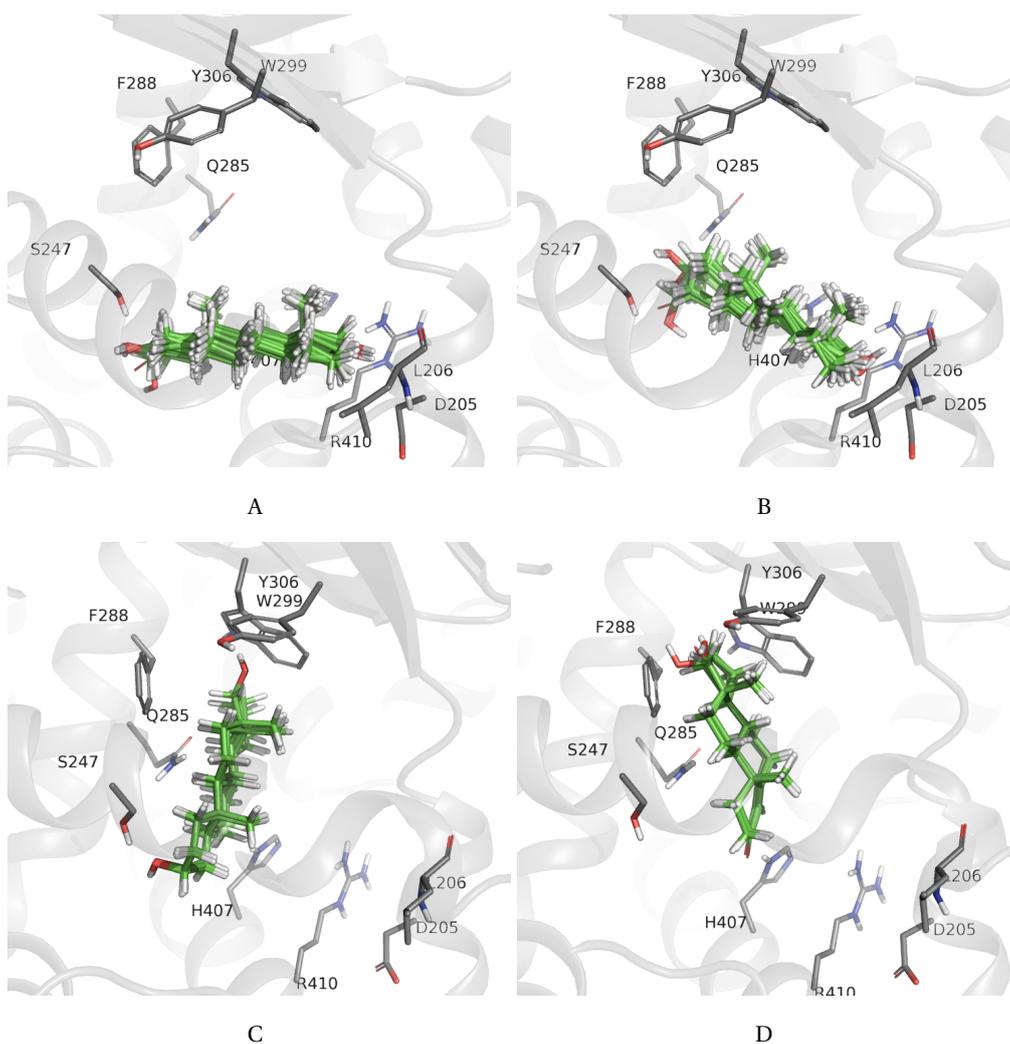


Figure 3.18: Docking results for A05–A18.

Ring A and ring B in **P01–P08** are in cis conformation, two types of binding modes are considered. Type A (fig. 3.19A) features hydrophobic interactions towards the Phe288-Trp299-Tyr305 pocket and a hydrogen bond with His407; type B (fig. 3.19B) mainly engages hydrogen bonds towards Gln285, Asp205 and Arg410. For **P09–P23**, ring A and ring B are in trans-conformation, the binding modes are familiar with type A ligands (fig. 3.19C and D), except that type D is less convergent than type B.

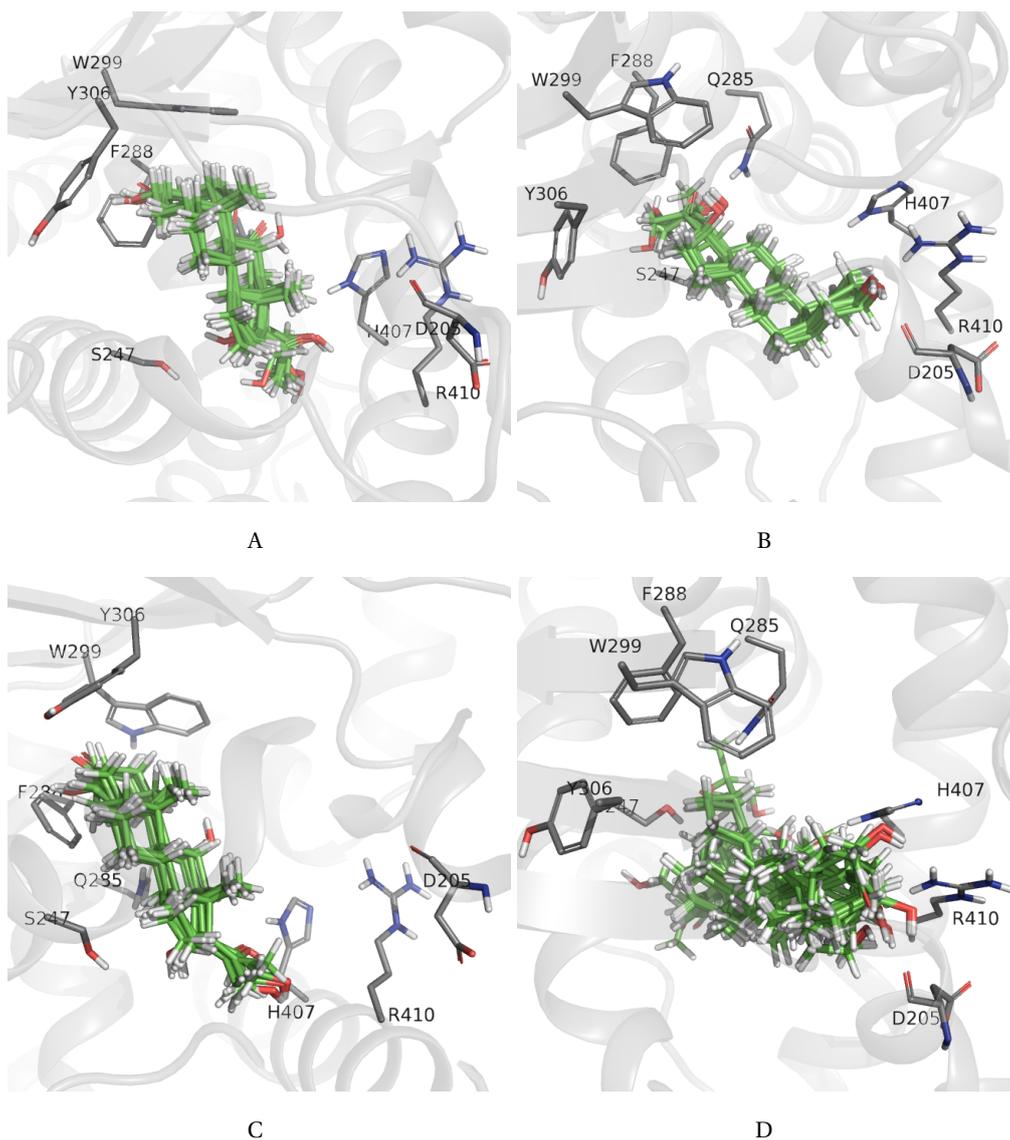


Figure 3.19: Docking results for **P01–P23**.

B: Case Study 2 — Docking for M01

M01 is a high-affinity PXR agonist ($EC_{50} = 0.7$ nM) which contains two aromatic ring, the reported putative binding mode was generated by mimicking the binding mode of T091317 (PDB: 2O9I).⁴⁹

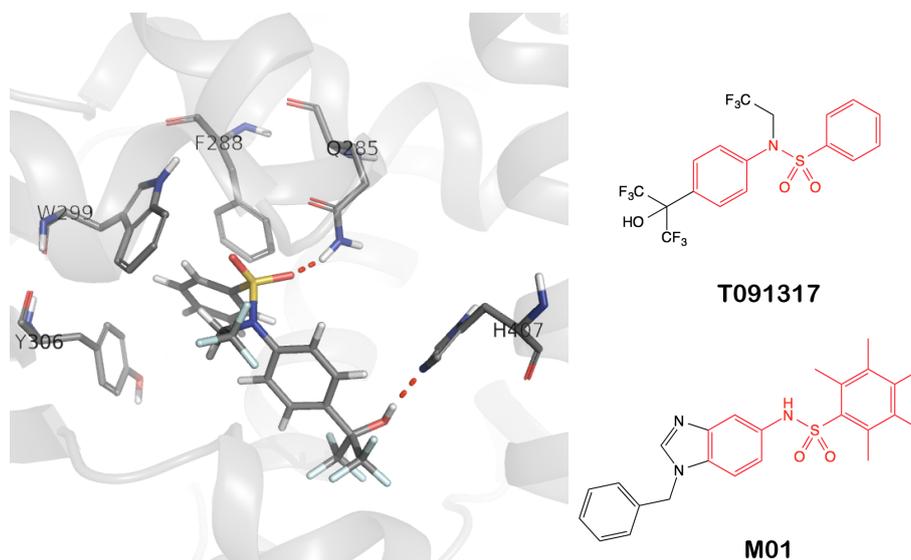


Figure 3.20: A: ligand binding site of T091317, B: T091317 and **M01**. **highlighted** are the similar fragments.

In the crystal structure, the benzyl-sulfonamide group was found binding to the hydrophobic pocket Phe288–Trp299–Tyr306, forming face-face π - π stacking with Phe288 and T-shaped π - π stacking with Trp299. The sulfonyl group was further stabilized by engaging a hydrogen bond with Gln285. However, the five methyl substituted aromatic ring in **M26** prevents a similar pose due to steric hindrance. Docking both by *Cheetah* and *IFD* were performed to generate putative binding modes. Top-ranked poses in both protocol were interactively inspected, four poses identified from *Cheetah* and three from *IFD* were selected and the kinetic stability were challenge by MD simulations with the *AMBER* package.

Ligand RMSD values were recorded for the 10 ns production-phase for each of the selected poses (fig. 3.21). The *Cheetah* docking poses were all above 2 Å, which suggest that the thermodynamic identified conformation is kinetically not stable. On the other hand, two

poses generated by *IFD* yielded low RMSDs (around 1 Å), which indicate a higher kinetic stability. For these two trajectories, a more detailed analysis was performed: the start/end state conformations, key hydrogen-bond distances and per-residue binding energy were recorded.

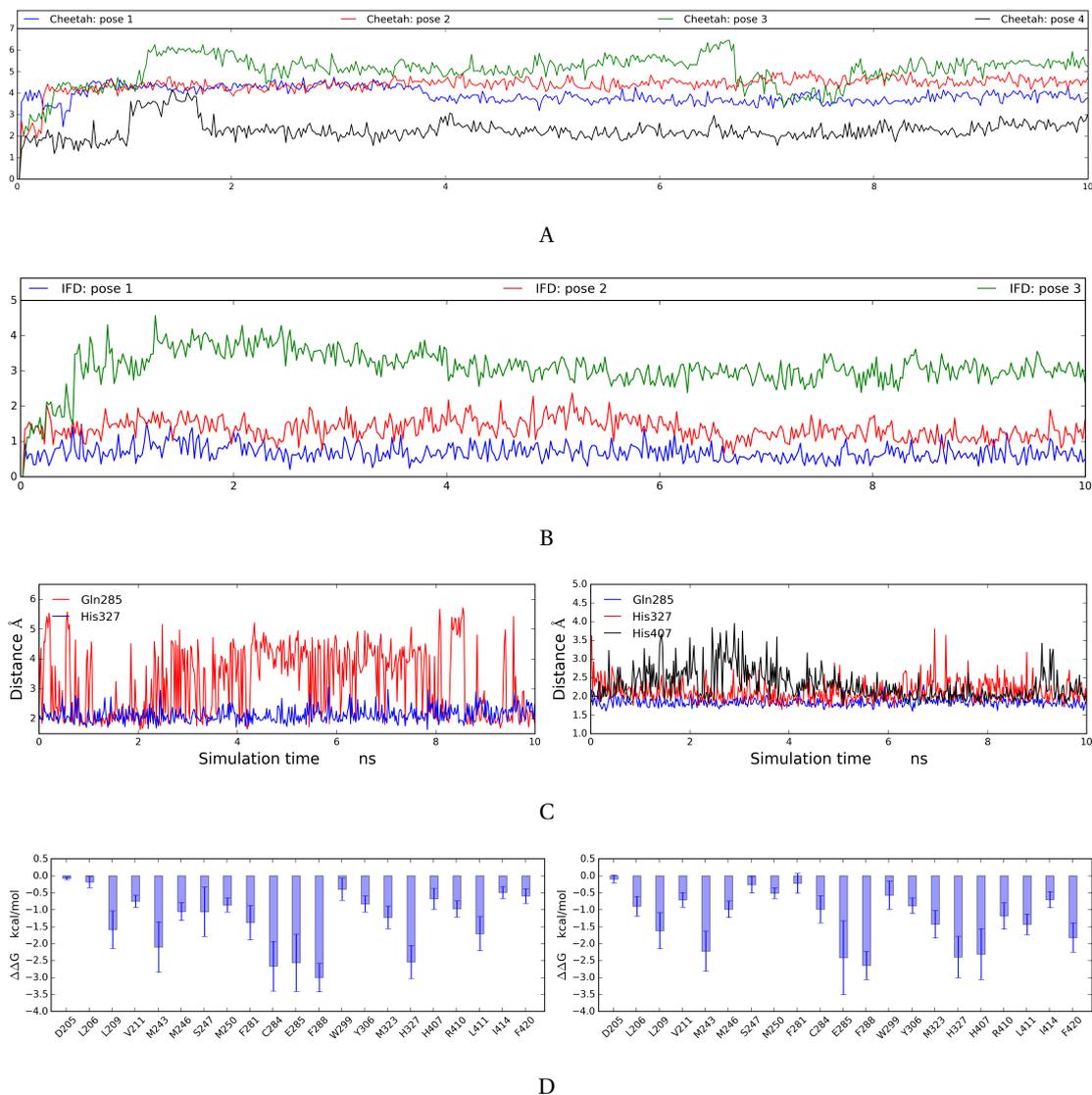


Figure 3.21: A and B: Ligand RMSD of MD simulation for PXR-M26 binding poses generated by *Cheetah* (A) and *IFD* (B); C: H-bond distance of pose 1 (left) and pose 2 (right); D: Binding energy decomposition by MM-PBSA of *IFD* pose 1 (left) and pose 2 (right).

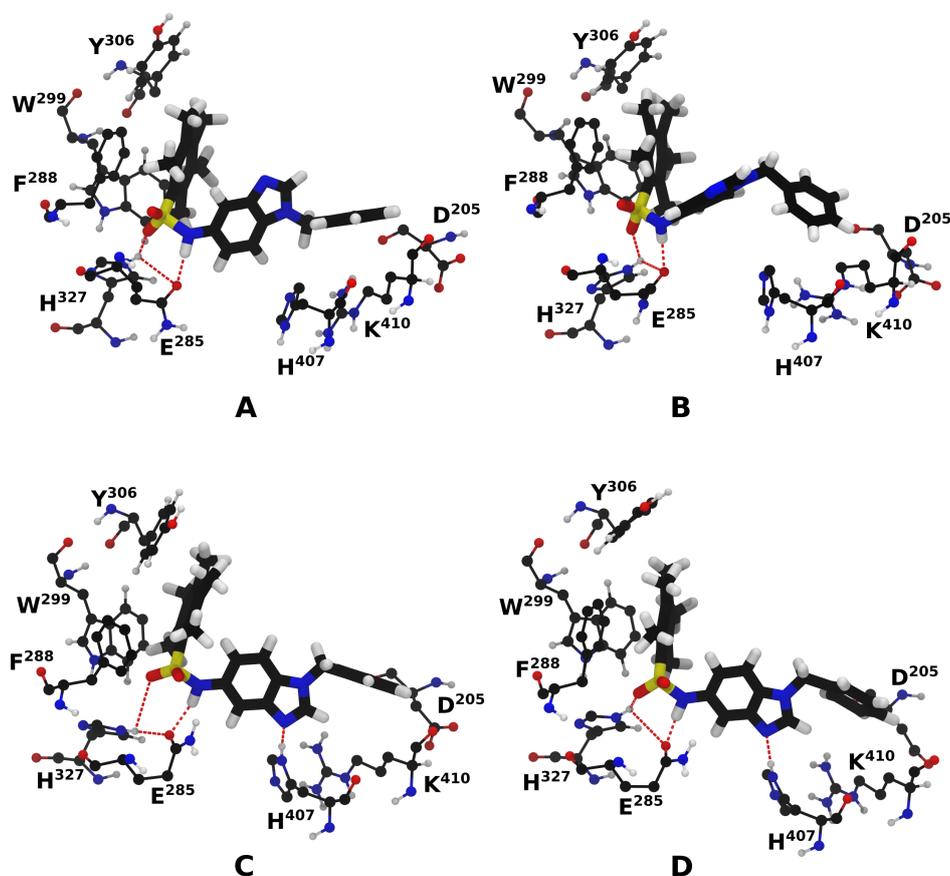


Figure 3.22: Binding modes before and after MD. (A) IFD pose 1 before MD, (B) IFD pose 1 after MD, (C) IFD pose 2 before MD, (D) IFD pose 2 after MD.

Pose 1 and 2 share a common binding pattern toward the F288–W299–Y306 pocket, as well as the hydrogen bonds to His327 and Gln285, this interaction pattern is conserved during the 10 ns MD simulations. The benzimidazole ring in pose 2 is further stabilized by forming hydrogen bond with His407, which locked the flexibility of the benzimidazole ring as in pose 1 it is more flexible than pose 2. The *IFD* protocol performed better results for docking ligands with aromatic/hydrophobic moieties. The empirical scoring function is trained for pattern recognition, which has benefits for the recognition of π - π stacking. Binding modes for the derivative compounds of **M01** (**M02**–**M09**) were docked to the PXR based on the stable *IFD* poses by interactive docking with *Yeti*.

For **S01–S07**, Gln285 act as hydrogen-bond donor for the binding, The hydrophobic pocket Phe288–Trp299–Tyr306 could either engage hydrophobic interactions with the aromatic ring or cyclobutane of the ligands, identified binding poses are shown in fig. 3.23. Finally, top-ranked poses generated by *Cheetah* are also included as augment for the QSAR study.

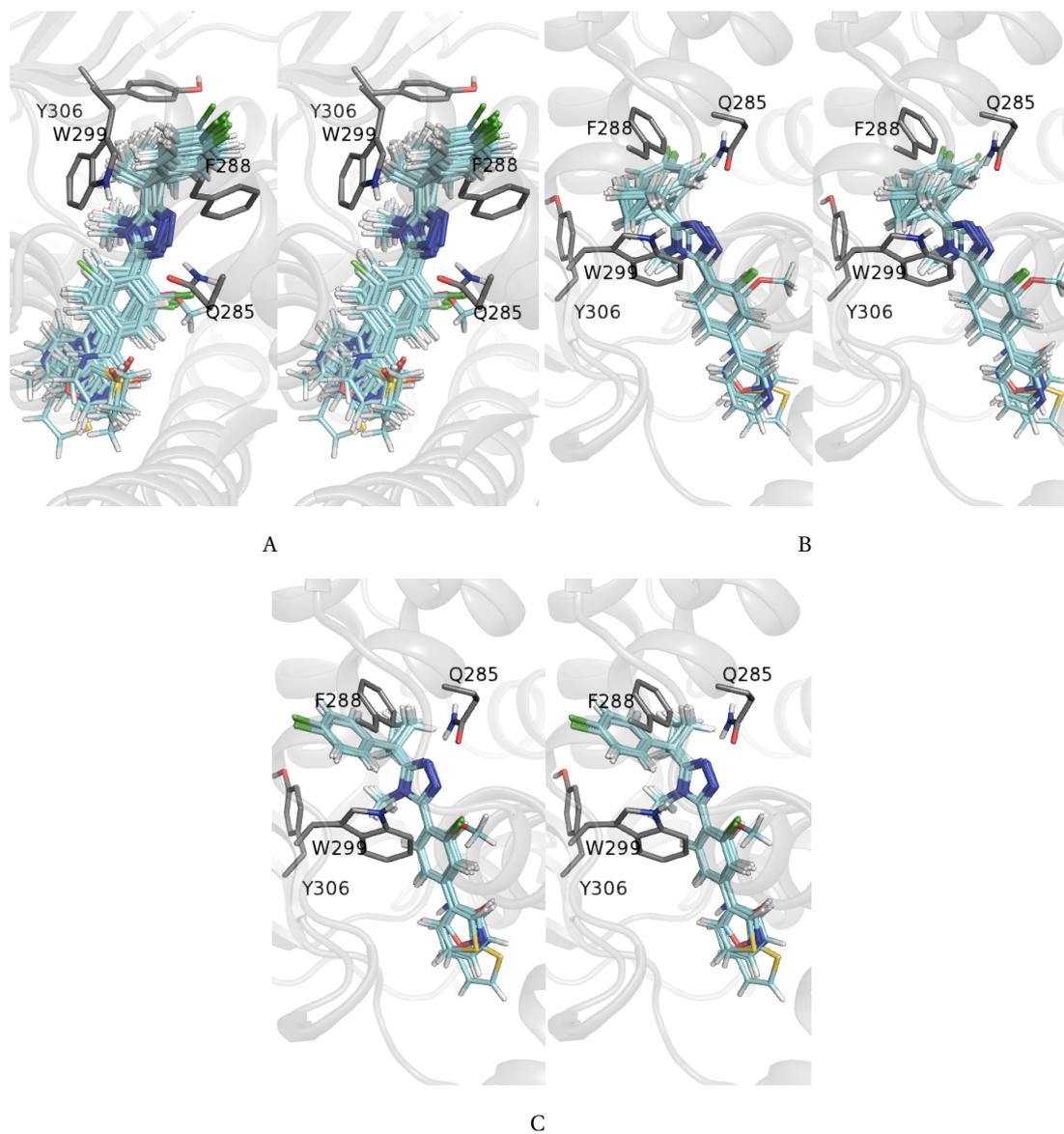


Figure 3.23: Stereo view of three kinds of binding modes for **S01–S07** to the PXR.

3.2.5 Multidimensional QSAR for the PXR

The ligand data from the docking poses were combined as input for the quantitative structure-activity relationship software *Quasar*, including ligand-specific information: its free energy of ligand binding (ΔG_{exp} , converted from EC_{50} data), ligand-desolvation energy, loss of entropy ($T\Delta S$) upon ligand binding as well as the increase of ligand-internal energy (strain) when binding from an aqueous environment to a hydrophobic receptor typically hydrophobic in nature.

In the *Quasar* simulation, the model family of the PXR (fig. 3.24 and table 3.1) converged at cross-validated r^2 of 0.812 for the 73 training compounds leaving one third of the set out, and yielded a predictive r^2 of 0.854 for the 24 test ligands. The average deviation (rms) between experimental and calculated affinities is of a factor of 2.0 for the training and 1.3 for the test set. The maximal deviation in the prediction of binding affinities for a compound is of a factor of 22.3 for training set and 4.7 for the test set, respectively.

When compared to the crystal structure of the PXR complexed with the two ligands, the receptor surrogate generated by *Quasar* properly reproduces properties observed for the amino-acid residues in the binding pocket: two hydrogen bond donors (big|yellow) are located close to the position occupied by Ser247 and Arg410, and two hydrogen bond acceptor (big|blue) are located close to the position occupied by Asn205 and Gln285. Moreover, hydrophobic properties (gray and brown) populate great part of the surface, correctly reflecting the hydrophobic character of the binding pocket.

To challenge the model, a second software (*Raptor*) — using the same ligand alignment and selection — was applied to yield an r^2 of 0.870 and a predictive r^2 of 0.646. When compared with *Quasar*, the *Raptor* simulation would only seem to yield a modest predictive power. Considering the limited range of experimental activity (77% of compounds cluster within two orders of magnitude), the compound's chemical diversity and the different literature source for the affinities, the *Raptor* model can be considered acceptable in terms of quality.

A representation of the receptor surrogate with bound the compound **17 β -estradiol** and **T091317** is depicted in fig. 3.24.

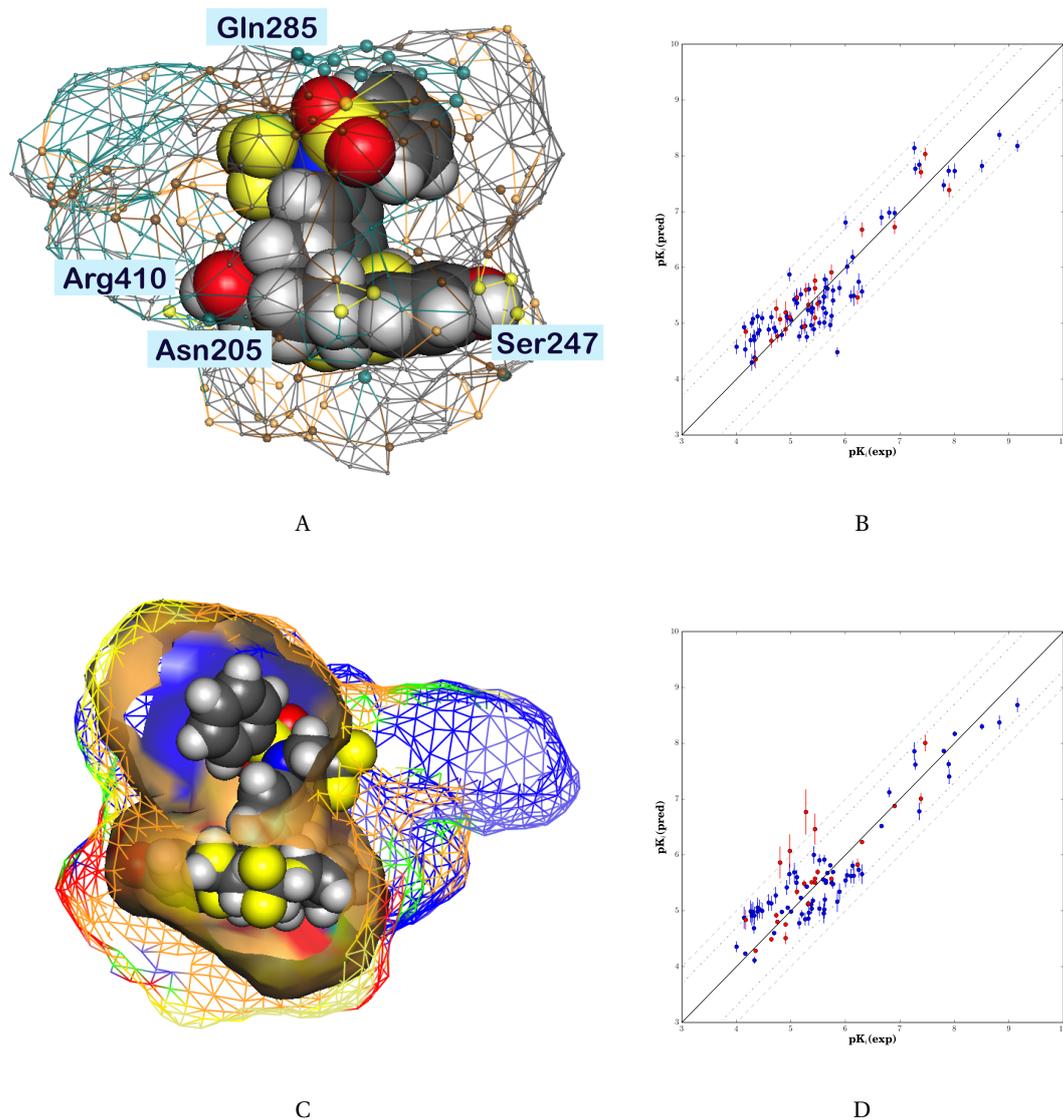


Figure 3.24: A: Representation of the *Quasar* model of PXR surrogate (*Quasar*) with bound compound **17 β -estradiol** and **T091317**(space-filling). The mapped quasi-atomistic properties are sized|colored as follows: big|blue (H-bond donor), big|yellow (H-bond acceptor), middle|saddle brown (hydrophobic, positively charged), middle|chocolate brown (hydrophobic, negatively charged), tiny|grey (hydrophobic, neutral), tiny|blue (solvent water). B: Comparison of experimental and predicted binding affinities of the training set (blue circles) and test set (red triangles) for the PXR by *Quasar*. C: *Raptor* model of the PXR with bound **17 β -estradiol** and **T091317**. D: Comparison of experimental and predicted binding affinities of the training set (blue circles) and test set (red triangles) for the PXR by *Raptor*.

Table 3.1: Summary of the *Quasar* and *Raptor* simulations for the 73 training and 24 test compounds

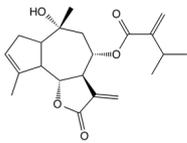
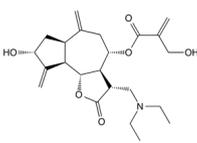
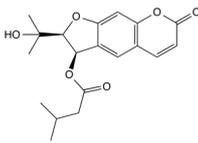
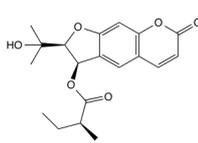
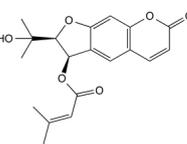
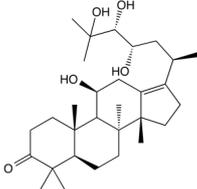
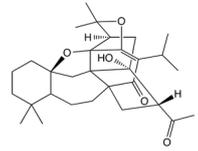
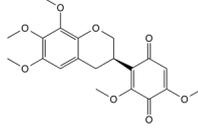
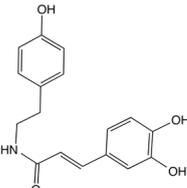
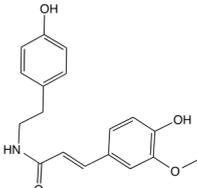
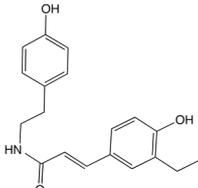
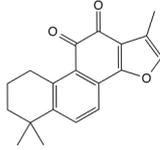
Simulation	r^2	q^2	rms. training	max. training	p^2	rms. test	max. test
<i>Quasar</i>	0.812	0.815	2.0	22.3	0.854	1.3	4.7
<i>Raptor</i>	0.870	n/a	0.9	3.9	0.646	3.8	16.8

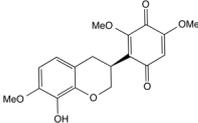
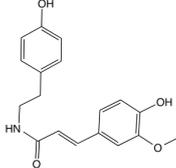
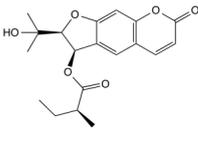
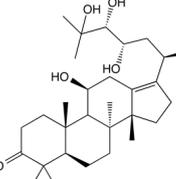
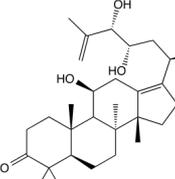
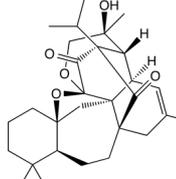
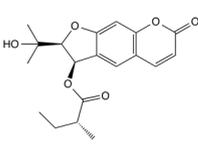
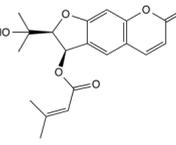
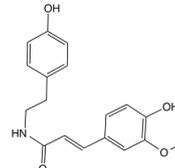
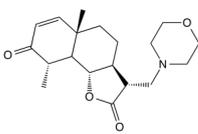
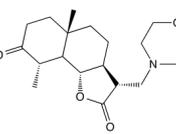
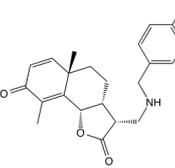
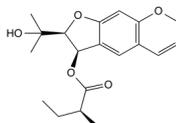
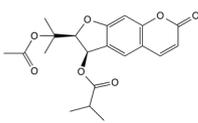
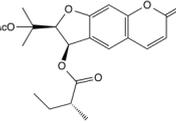
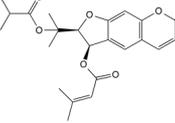
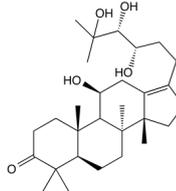
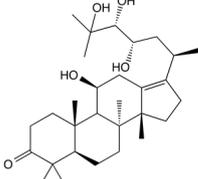
r^2 : correlation coefficient, q^2 : cross-validated r^2 , p^2 : predictive r^2 ; the *rms* and maximal deviation from the experimental binding affinity is given as a factor (off) in K_i .

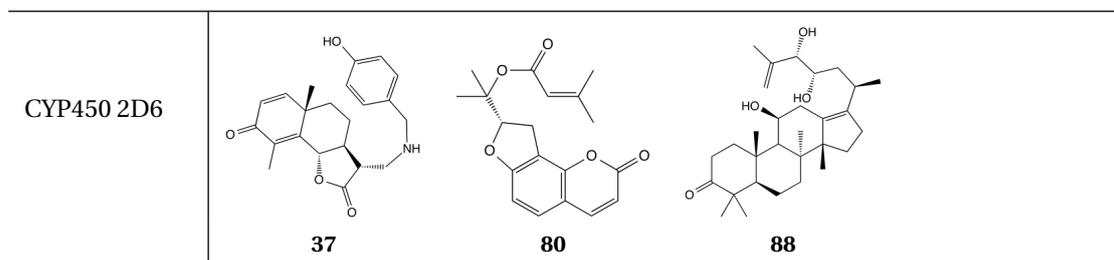
3.3 *In silico* evaluation of anti-trypanosome natural products

The binding of natural products, potentially active against the Human African Trypanosomiasis (HAT), towards 16 off-targets was simulated and quantified by means of the *VirtualToxLab*.³⁸ For each of the binding affinities computed below 100 nM, a subsequent 5.0–10.0 ns MD simulation was performed to probe the kinetic stability of the protein–ligand complex. Table 3.2 shows the selected compounds with their respective targets. In the following section, the underlying binding modes and affinities are discussed in detail.

Table 3.2: Active compounds towards selected targets as identified by the *VirtualToxLab*

Target	Compound(s)				
Androgen receptor	 16	 27	 70	 (S)-71	
	 72	 85	 91	 93	
	 97	 98	 100		
	Aryl hydrocarbon receptor	 54			

<p>Estrogen receptor β</p>	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>92</p> </div> <div style="text-align: center;">  <p>98</p> </div> </div>
<p>Glucocorticoid receptor</p>	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>(S)-71</p> </div> <div style="text-align: center;">  <p>85</p> </div> <div style="text-align: center;">  <p>88</p> </div> <div style="text-align: center;">  <p>89</p> </div> </div>
<p>Mineralocorti- coid receptor</p>	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>(R)-71</p> </div> <div style="text-align: center;">  <p>72</p> </div> <div style="text-align: center;">  <p>98</p> </div> </div>
<p>Progesterone receptor</p>	<div style="display: flex; flex-direction: column; align-items: center;"> <div style="display: flex; justify-content: space-around; width: 100%;"> <div style="text-align: center;">  <p>33</p> </div> <div style="text-align: center;">  <p>34</p> </div> <div style="text-align: center;">  <p>39</p> </div> <div style="text-align: center;">  <p>(S)-71</p> </div> </div> <div style="display: flex; justify-content: space-around; width: 100%; margin-top: 20px;"> <div style="text-align: center;">  <p>74</p> </div> <div style="text-align: center;">  <p>(R)-76</p> </div> <div style="text-align: center;">  <p>77</p> </div> <div style="text-align: center;">  <p>85</p> </div> </div> </div>
<p>Thyroid receptor β</p>	<div style="text-align: center;">  <p>85</p> </div>



3.3.1 Compounds active towards the androgen receptor

Eleven compounds were identified as active towards the AR by the *VirtualToxLab*, of which two are sesquiterpene lactones **16** and **27** (with computed IC_{50} values of 31.4 nM and 19.3 nM respectively); three are psoralen derivatives **70** (88.1 nM), (*S*)-**71** (35.4 nM) and **72** (25.7 nM); and three are paprazine derivatives **97** (22.9 nM), **98** (6.49 nM) and **100** (92.0 nM).

For **16** and **27**: the docking poses contributing most to the binding affinity are quite different (fig. 3.25). Although they both engage in hydrogen bonds with Asn705 and Thr877, **16** donates its hydroxyl group from the 7-membered ring while **27** provides it from its allyl-alcohol side chain—which is more flexible. In addition, **27** is further stabilized by Met745 and Arg752. After challenging the docking (input structure: thermodynamic lowest-energy pose) by means of 5 ns MD simulation each, the ligand movement was analyzed by recording the RMSD of the heavy atoms on its rings. The position and orientation of **16** became stable after 4 ns production run when its RMSD was fluctuated around 0.5 Å, which indicating that the docking pose of **16** with the AR is unfavorable due to ligand flexibility. In contrast hereto, the RMSD of **27** remained around 0.5 Å for the whole production phase. The moderate movement during the equilibration stage is probably due to receptor adaption (induced fit including the main chain) and interactions with the surrounding water molecules.

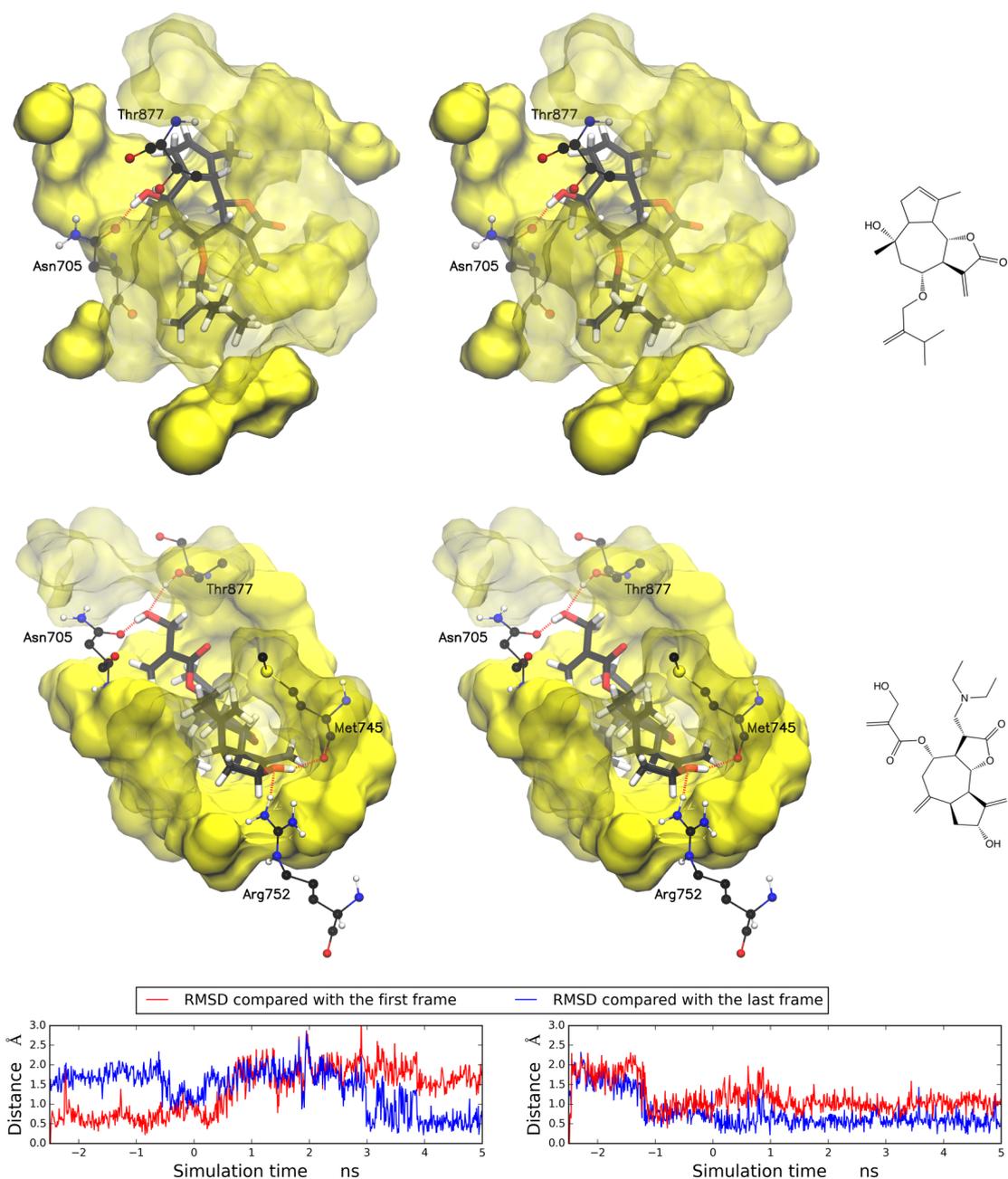
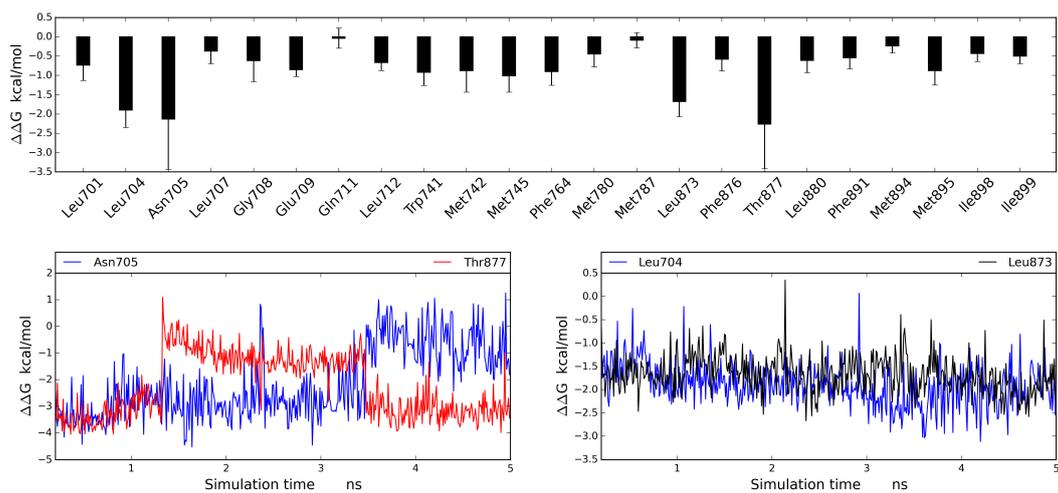
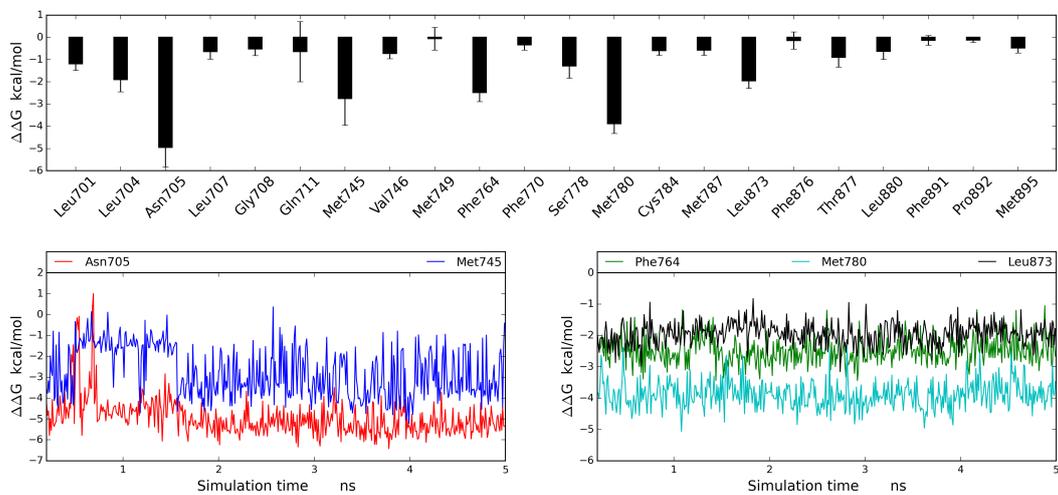


Figure 3.25: Top and middle: Details of the binding of **16** (top) and **27** (middle) to the AR. The ligands are represented as licorice, key amino-acid residues as balls and sticks, hydrophobic residues within the binding pocket as surface colored in yellow. Carbon, nitrogen, oxygen and hydrogen atoms are colored in black, blue, red and white respectively. Bottom: Ligand RMSD during the MD simulation for **16** (left) and **27** (right), negative values in the x-axis refer to the equilibration stage.

The binding energies were obtained by the MM-PBSA protocol (fig. 3.26). **16** lost its hydrogen bond with Asn705 during the simulation while the interaction with Thr877 was retained, the ligand is further stabilized mainly by the hydrophobic residues Leu704 and Leu873. **27** lost its hydrogen bonds with Arg752 and Thr877, in contrast, the hydrogen bonds with Asn705 or Met745 were retained, the ligand is further stabilized mainly by Phe764, Met780 and Leu873.



A



B

Figure 3.26: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues with respect of **16** (a) and **27** (b) binding to the AR.

For the psoralen derivatives—**70**, (*S*)-**71** and **72**: the docking poses contributing most to the binding affinity are quite similar (fig. 3.27), which indicates that the interactions are conservative in nature. Asn705 and Thr877 generated bridging hydrogen bonds to all three ligands; **70** and **72** formed a third hydrogen bond with Gln711.

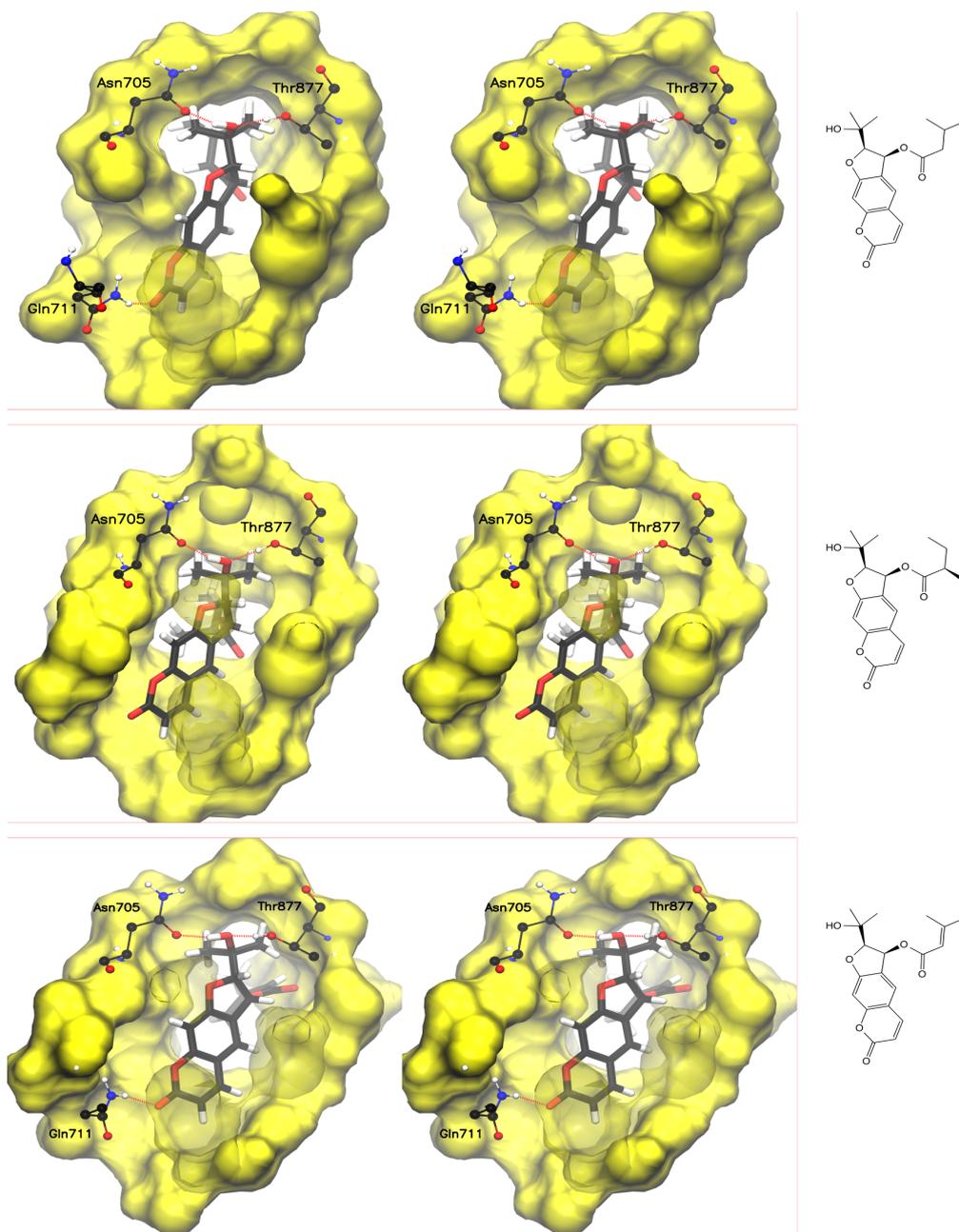


Figure 3.27: Details of the binding of **70** (top), (*S*)-**71** (middle) and **72** (bottom) to the AR.

The ligand RMSDs (of the ring heavy atoms) and the hydrogen-bond distances are shown in fig. 3.28. The RMSDs remained below 1.5 Å within 8 ns of the production stage compared with the initial frame for all the three ligands. The hydrogen bond between (S)-**71** and Gln711 became stable only after 5 ns, all the other hydrogen bonds remained stable throughout the simulation (Gln711 could flip the symmetrically related HN21 and HN22 during the MD simulation; consequently, the shorter of the two was recorded), except for **70**, which lost the interaction with Thr877 after 2 ns of the production stage.

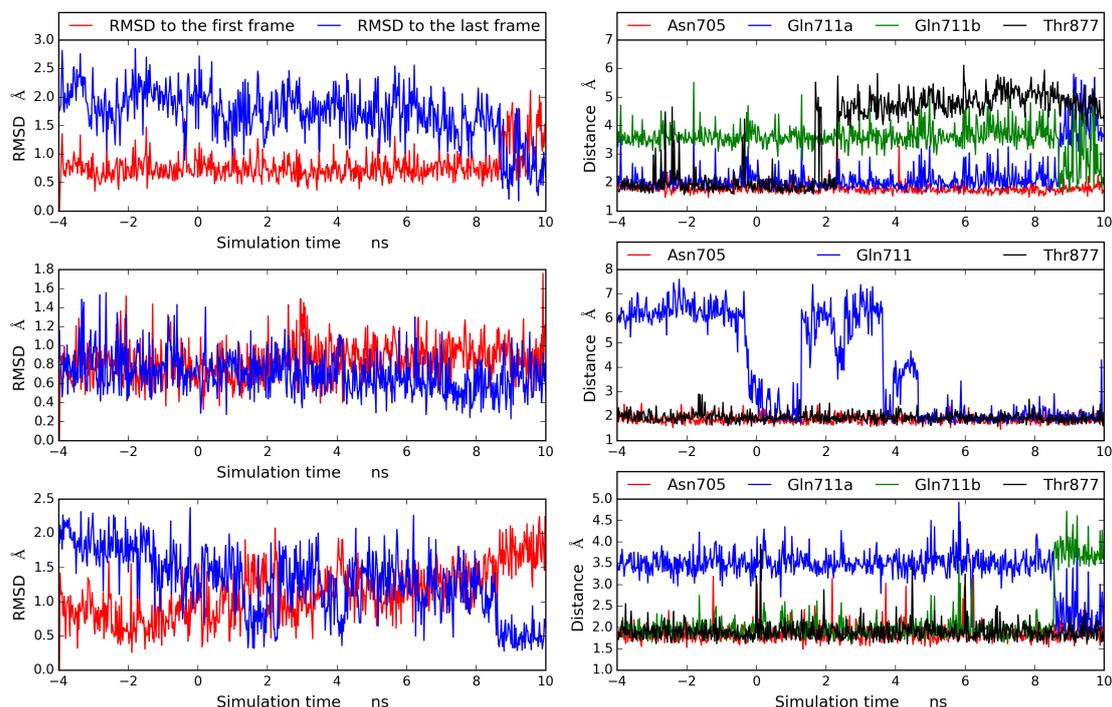


Figure 3.28: The ligand core RMSDs (aromatic ring atoms) are shown on the left panel, the hydrogen-bond distances are shown on the right. The complexes of **70**, (S)-**71** and **72** with the AR are arranged at the top, middle and bottom, respectively.

MM-PBSA energy (fig. 3.29): the free-energy contribution of Asn705 towards (S)-**71** and **72** are retained; the free-energy contribution of Gln711 towards (S)-**71** and **73** are retained throughout the MD simulation, while for **72** it is retained after 5ns production, which is in agreement with the trends of hydrogen-bond distance; the free-energy contribution of Thr877 towards all three ligands are retained throughout the simulation. In addition, **70** is

further stabilized mainly by Leu704, Arg752 and Leu873 within the binding pocket; (S)-71 is further stabilized mainly by Leu704 and Met745; 72 is further stabilized mainly by Leu704, Arg780 and Leu873.

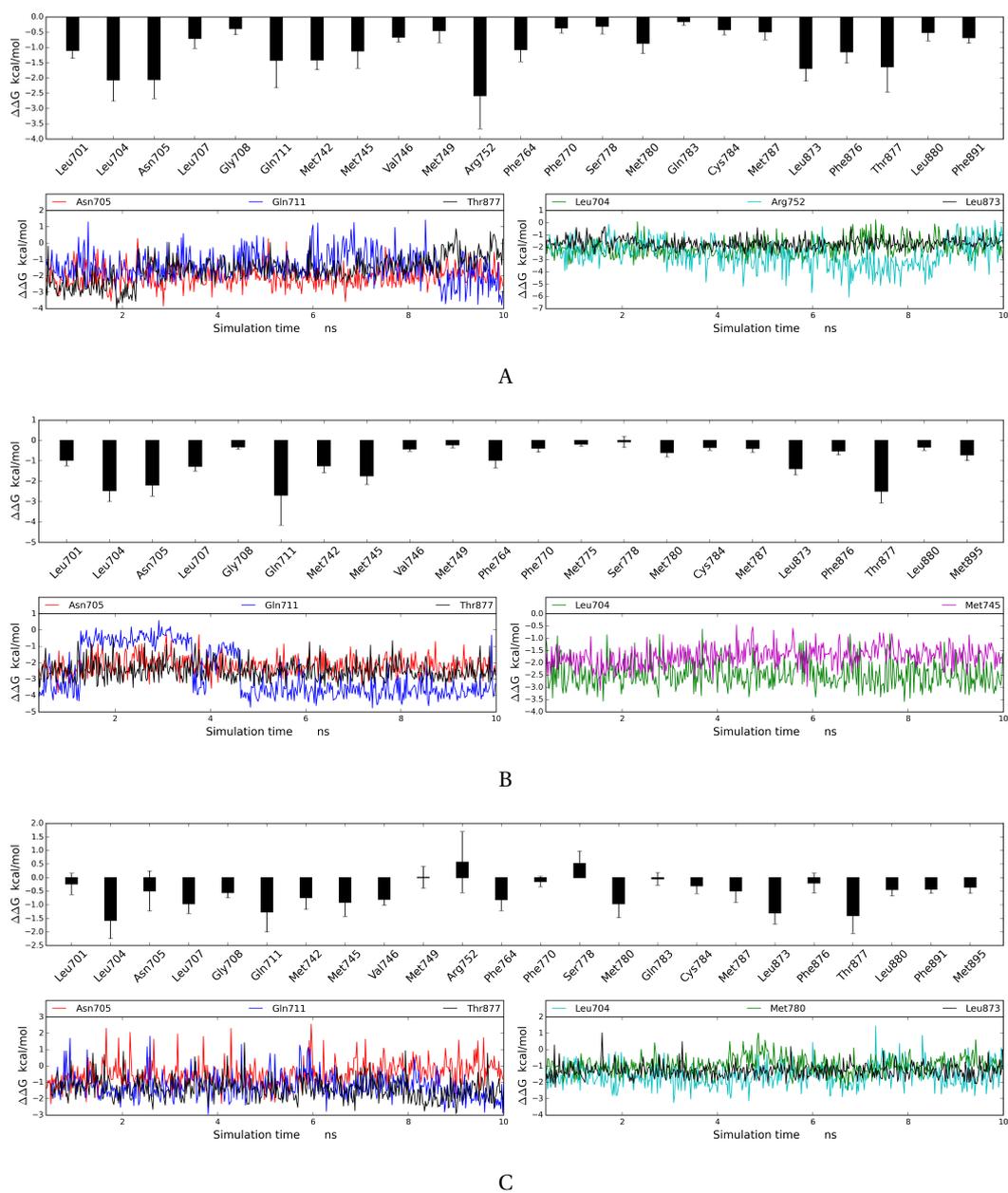


Figure 3.29: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues with respect of 70 (a), (S)-71 (b) and 72 (c) binding to the AR.

For the paprazine compounds—**97**, **98** and **100**: the thermodynamic poses contributing most to their respective affinity are displayed in fig. 3.30. They all feature hydrogen bonds with Asn705 and Thr877; in addition, for the complex of the AR and **100**, the amide group is stabilized by Gln711.

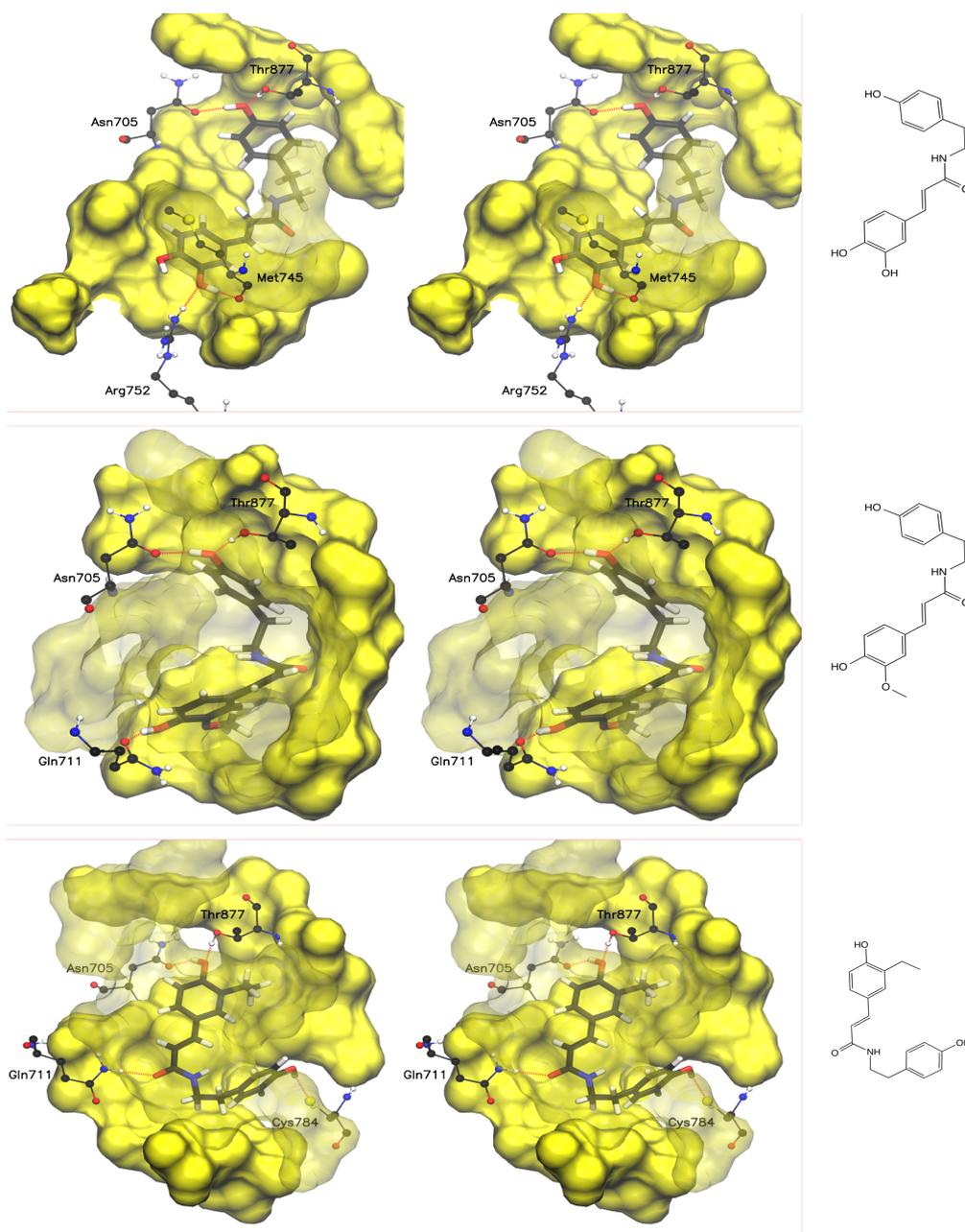


Figure 3.30: Details of the binding of **97** (top), **98** (middle) and **100** (bottom) to the AR.

The ligand RMSDs for **97** and **98** increased above 3.0 Å compared with their docking poses, which means they would not be kinetically stable; for **100** the ligand RMSD remains within 1.5 Å compared with both the thermodynamic lowest-energy pose and the last MD frame (fig. 3.31A). The hydrogen-bond distances of **100** with Asn705, Gln711 and Thr877 are within 2.0 Å during most of the production stage, which implies that these three residues continuously stabilize the ligand (fig. 3.31B). MM-PBSA calculation (fig. 3.31C) also suggest the free-energy contributions of Asn705, Gln711 and Thr877 to **100** remains stable during the whole simulation.

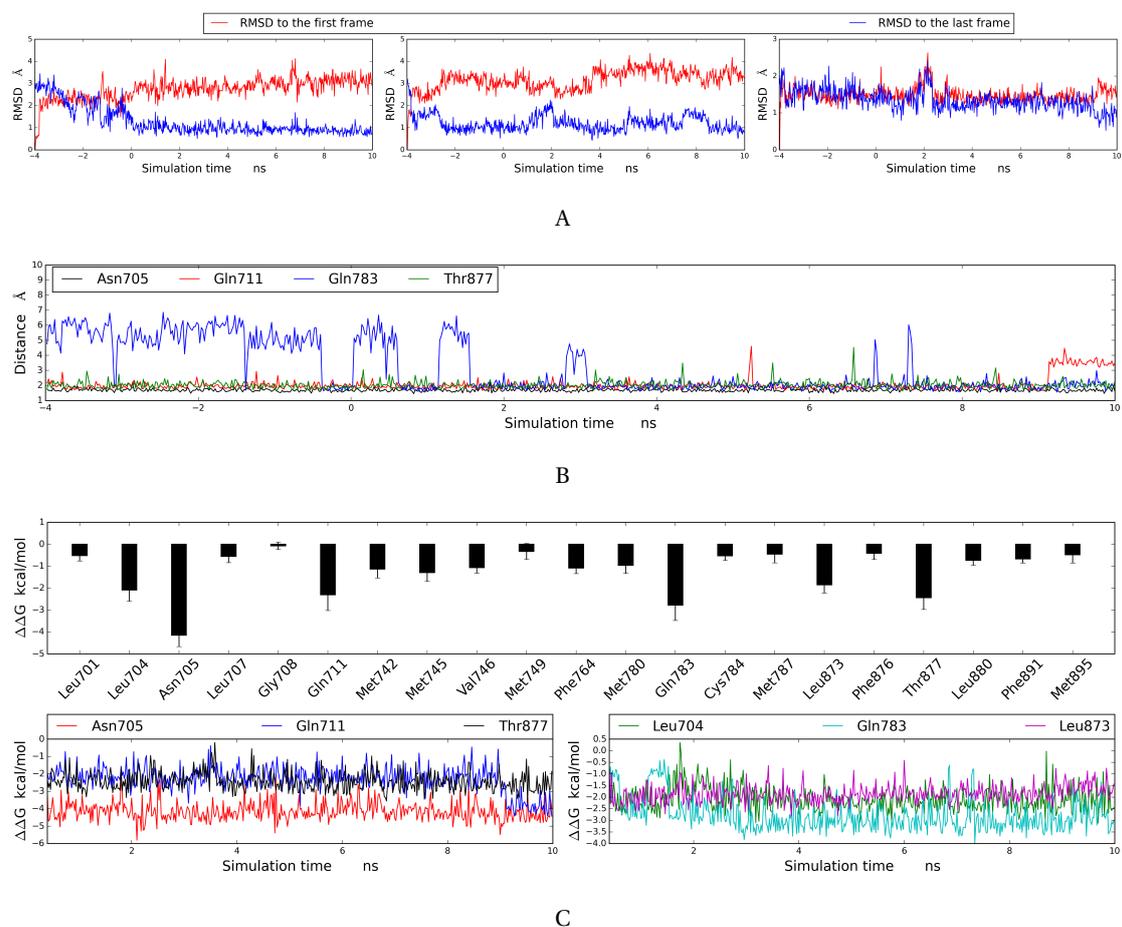
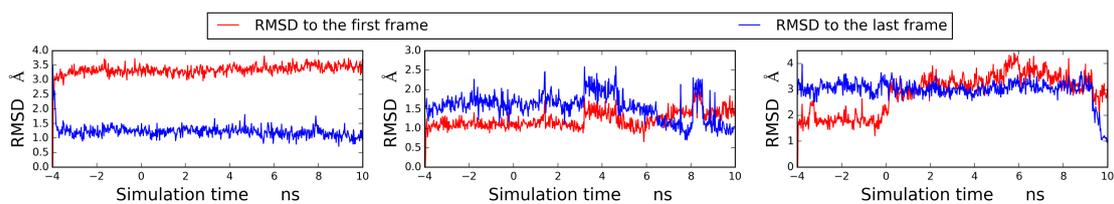
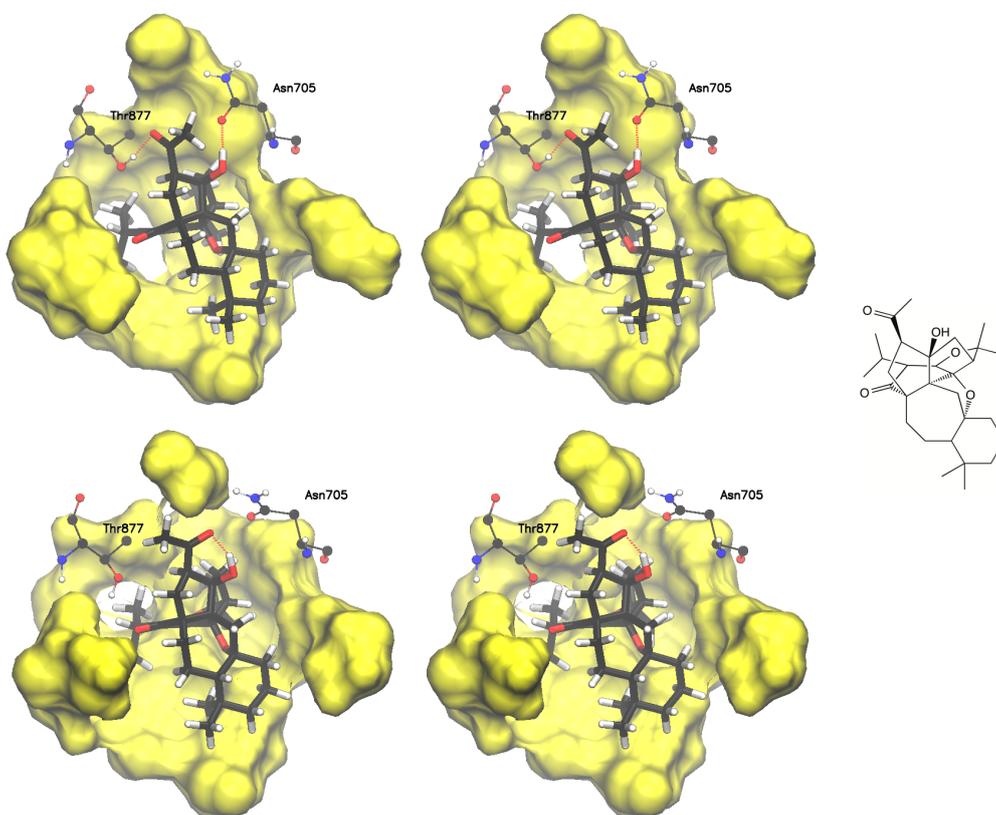


Figure 3.31: a: Ligand RMSDs of the AR with **97** (left), **98** (middle) and **100** (right). b: key-residue hydrogen-bond distances for the complex of AR and **100**. c: Binding-free energy contributions within the binding pocket and time-resolved details for key amino-acid residues of the AR binding with **100**.

The docking poses of **85** or **93** with the AR did not remain stable according to high ligand RMSDs (fig. 3.32A); for **91**, though the ligand RMSD was mainly below 1.5 Å, the hydrogen bonds generated by **91** with Asn705 and Thr877 were not retained during the simulation, instead the ligand formed an intra-molecular hydrogen bond, which does not contribute to the binding affinity (fig. 3.32B).



A



B

Figure 3.32: a: Ligand RMSD of **85** (left), **91** (center) and **93** (right) with the AR; b: detail of the docking mode of **91** (up) and the AR and the last MD frame (down).

3.3.2 The only compound active towards the Aryl hydrocarbon receptor

Tanshinone IIA (54) was identified as the only compound active towards the AhR, with a computed IC_{50} of 26 nM. The docking pose contributing most to the binding affinity (as identified by the *VirtualToxLab*: thermodynamic solution) and the corresponding structure toward the end of 5.0 ns MD simulation (kinetic solution) are shown in fig. 3.33.

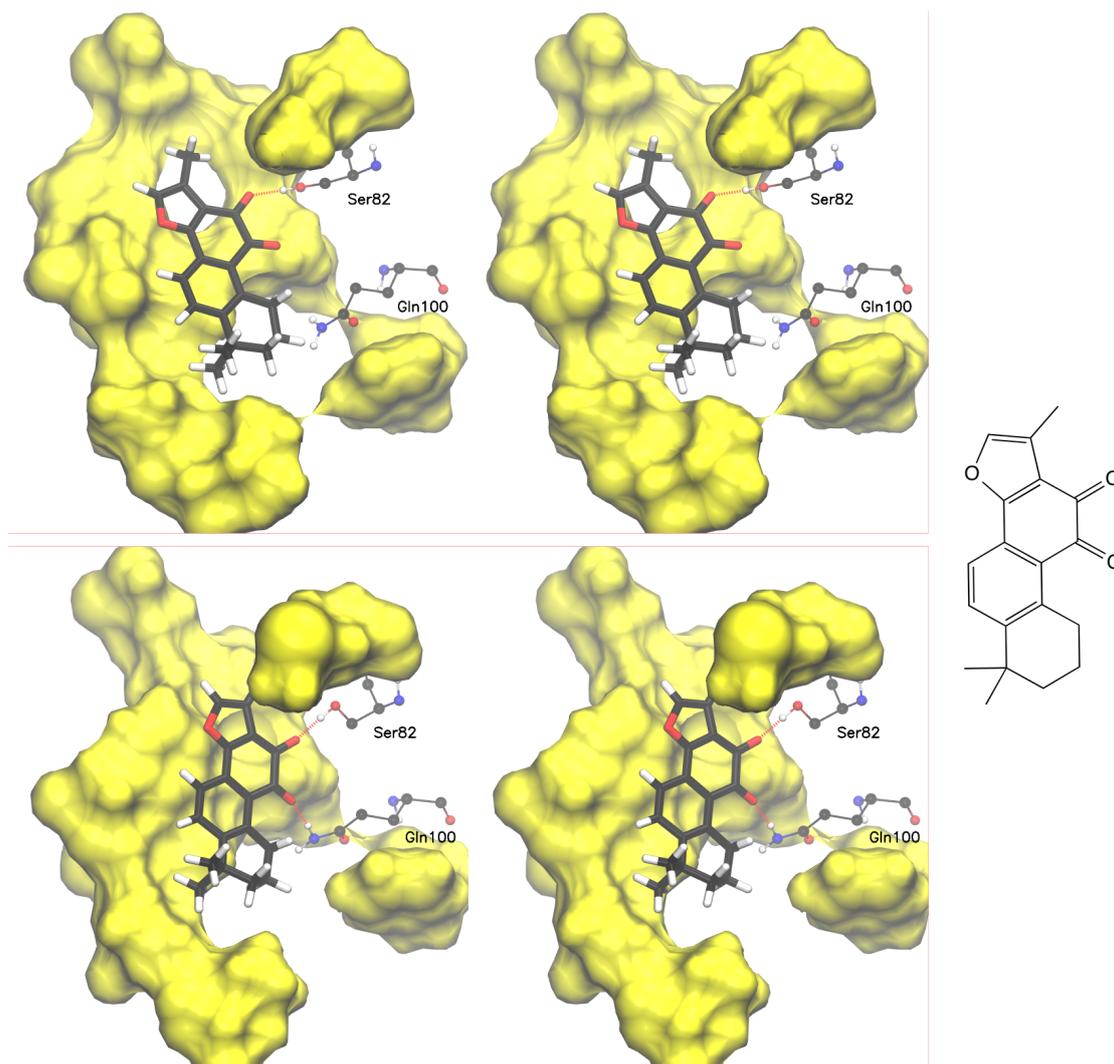


Figure 3.33: Details of the binding of tanshinone IIA to the AhR as identified by the *VirtualToxLab* (top) and the last MD frame (bottom).

In the binding pocket, tanshinone IIA is accommodated by the hydrophobic residues Phe4, Phe12, Pro14, Leu25, Leu32, Phe41, Ile42, Met57, Phe68, Leu70, Ala84 and Val98. The ligand is further stabilized by engaging a hydrogen bond with the side-chain hydroxyl group of Ser82, which was retained throughout the entire simulation period. An additional hydrogen bond with the side-chain amide group of Gln100 surfaced after the equilibration step and remained stable thereafter. The system turned out to be stable after 1.3 ns, where the RMSD was reduced to 1 Å compared with the last MD frame. During the whole 5.0 ns production phase, the ligand RMSD remained below 1.0 Å compared with the last frame, which suggests that the equilibrated pose was stable during the whole production stage (fig. 3.34A). The MM-PBSA results (fig. 3.34B) also confirmed the stable free-energy contributions of Ser82 and Gln100, and further suggest that Ile42 plays an important role for the binding.

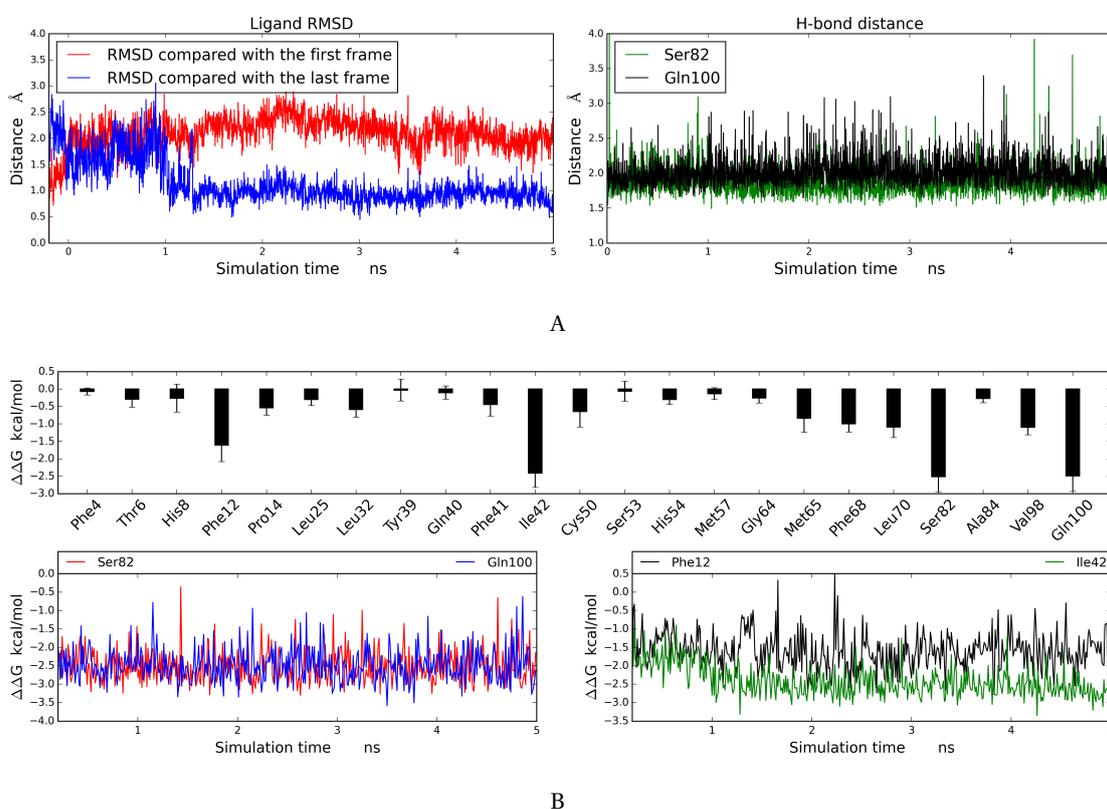


Figure 3.34: a: Time-resolved details of ligand RMSD and key-residue hydrogen-bond distances of **54** binding to the AhR. b: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues of the AhR binding with **54**.

3.3.3 Compounds active towards the estrogen receptor β

Abruquinone I (**92**) and moupinamide (**98**) were predicted to be active towards the ER β , with computed IC₅₀ of 78.2 nM and 24.8 nM respectively. After challenging the docking poses contributing most to the affinities (as identified by the *VirtualToxLab*) by means of 10 ns MD simulations, abruquinone I slightly changed its binding while moupinamide underwent a substantial movement as illustrated in fig. 3.35, indicating that the thermodynamically identified docking pose of moupinamide is not favorable.

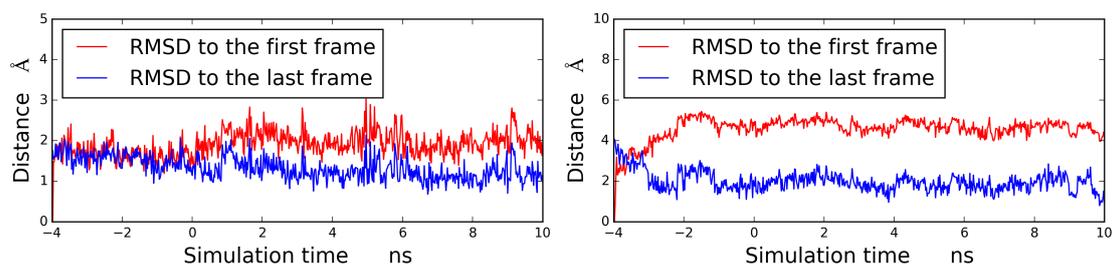


Figure 3.35: Ligand RMSD of **92** (left) and **98** (right) binding to the ER β .

For the docking pose of abruquinone I to the ER β , the ligand is stabilized by a hydrogen bond with Glu305. The negatively charged Glu305 forms a salt bridge with the positively charged Arg346. After the MD simulation, the ligand scaffold retains at its original shape. The hydrogen-bond stabilization of the ligand shifted to the backbone of Leu339 (fig. 3.37A). As depicted in fig. 3.36, the hydrogen bond generated between the ligand and the ER β was engaged in a conjugated switch between Glu305 and Leu339. The MM-PBSA result also suggest Leu339 retains free-energy contribution to the binding towards abruquinone I (fig. 3.37B).

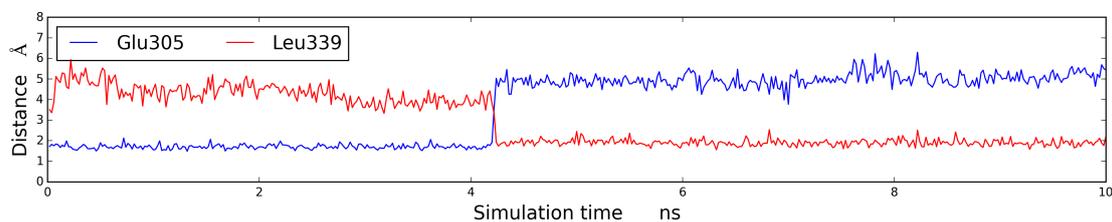


Figure 3.36: Hydrogen bond distances of abruquinone I binding to the ER β .

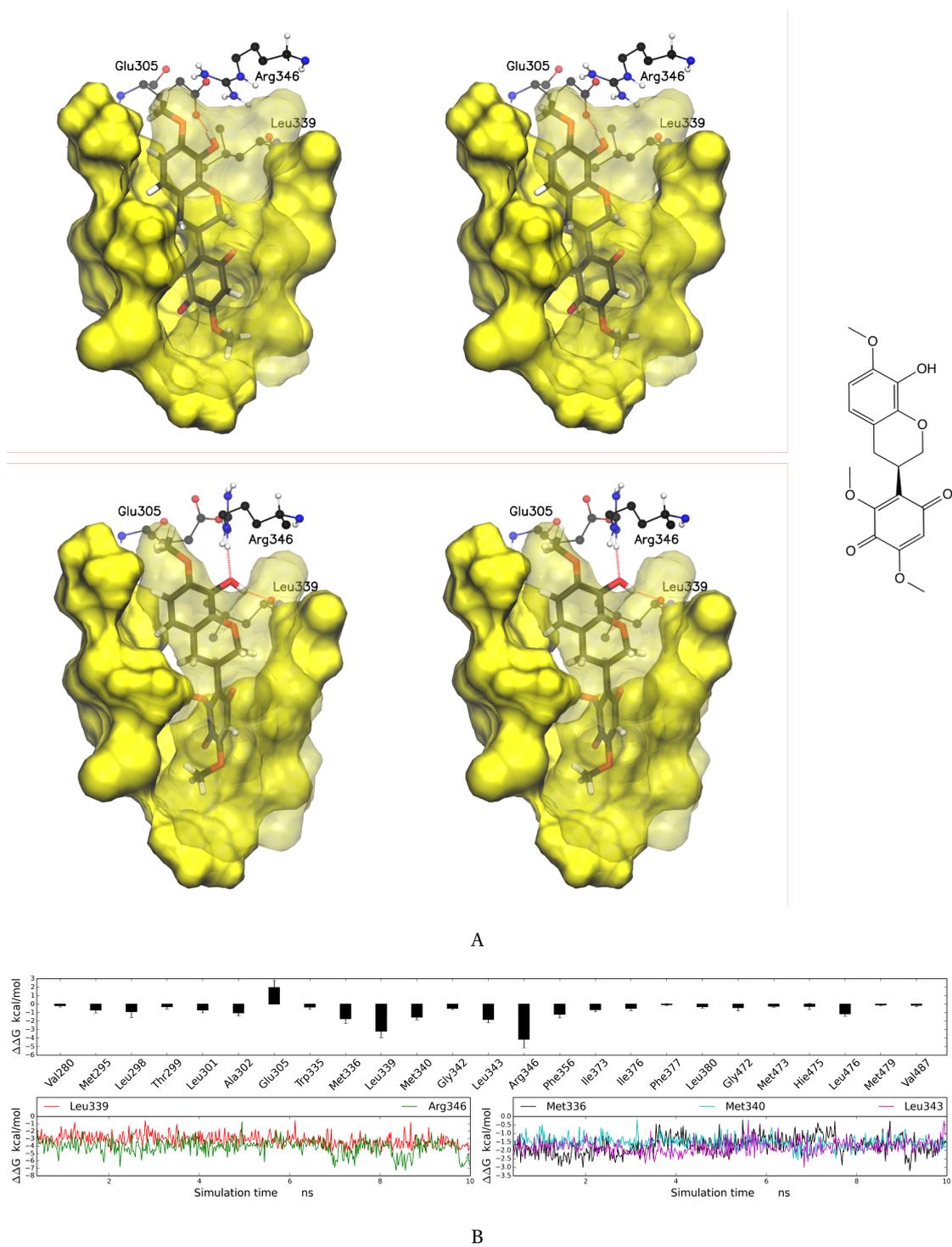


Figure 3.37: a: Details of the binding of **92** to the ER β as identified by the *VirtualToxLab* (top) and the last MD frame (bottom). b: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues of the ER β binding with **92**.

3.3.4 Compounds active towards the glucocorticoid receptor

S-71, **85**, **88** and **89** were predicted to be active towards the GR, with computed IC_{50} of 45.8 nM, 68.1 nM, 19.7 nM and 87.2 nM respectively. For *S-71*, **85** and **88**, the thermodynamic poses contributing most to their respective affinity are displayed in fig. 3.39. They all feature hydrogen bonds with Asn564 and Gln570, in addition, **85** and **88** are also stabilized by engaging a hydrogen bond with Gln642. After challenging the docking by 10 ns MD simulation, the RMSD of **89** increased to 4 Å compared with its starting pose, indicating a large movement, while the other ligands remain stable throughout the production phase and showed only a moderate movement during the equilibration stage (fig. 3.38A). The hydrogen-bond distances between Asn564 and all three ligands are retained during the simulation; only the hydrogen-bond distance between Gln570 and **88** is retained; for **88** and **89**, the ligand is further continuously stabilized by Cys638 and Gln642 respectively (fig. 3.38B).

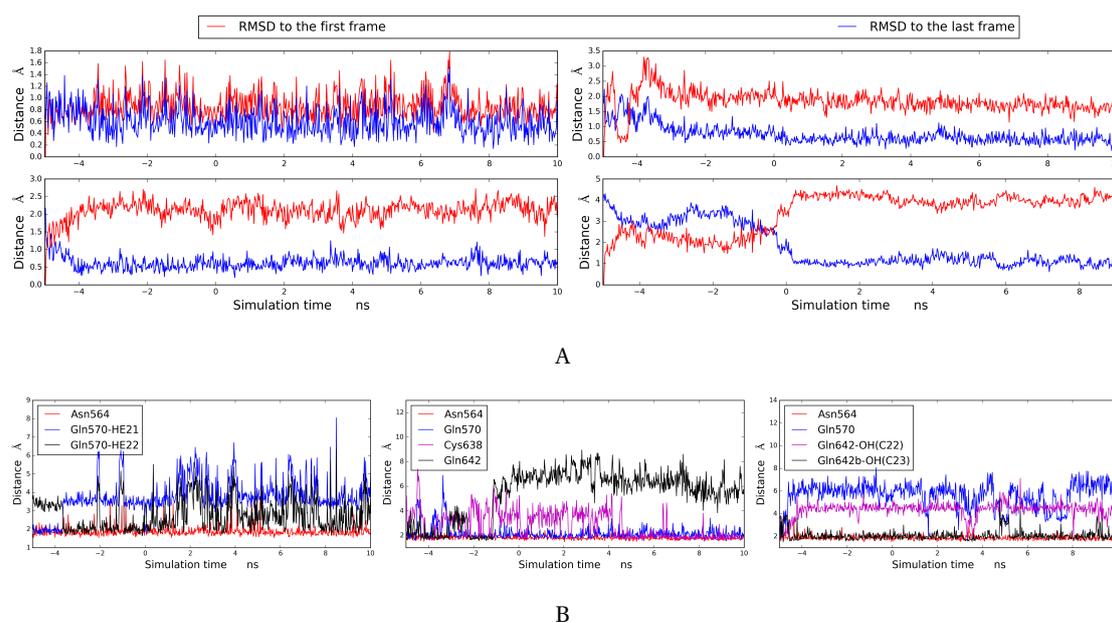


Figure 3.38: a: Ligand RMSD of *S-71* (top left), **85** (top right), **88** (bottom left) and **89** (bottom right) binding to the GR. b: Time-series key-residue hydrogen-bond distances for *S-71* (left), **85** (middle) and **88** (right) binding to the GR.

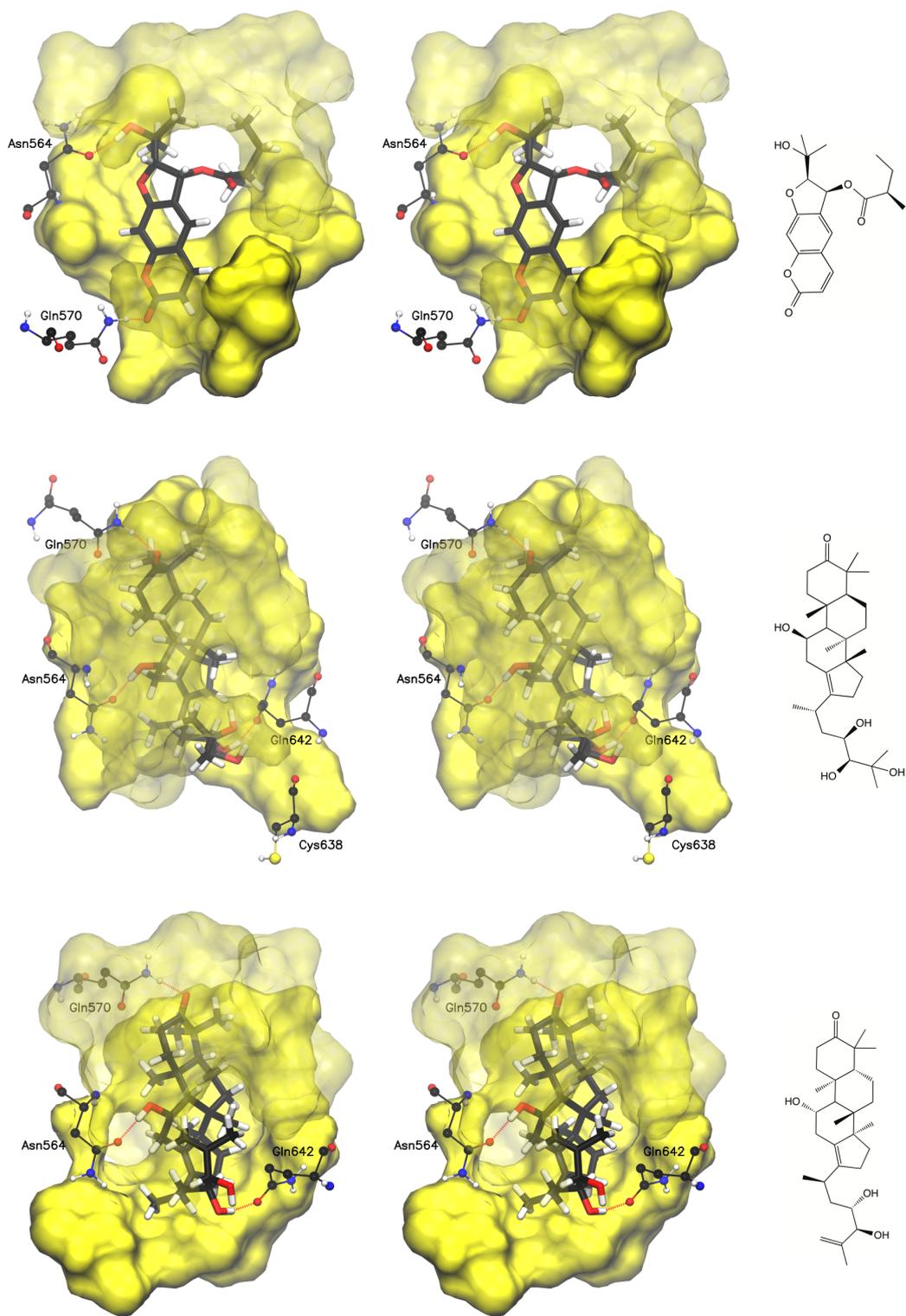


Figure 3.39: Details of the binding of *S*-71 (top), 85 (middle) and 88 (bottom) to the GR.

The MM-PBSA results further suggest the agreement with the time-series hydrogen-bond distance results (fig. 3.42): the free-energy contribution of Asn564 to all three ligands are retained; the contribution of Gln570 is retained for **71** and **85**; the contribution of Cys638 to **85** is retained after 5.0 ns and the contribution of Gln642 is retained for **88** throughout the MD simulation.

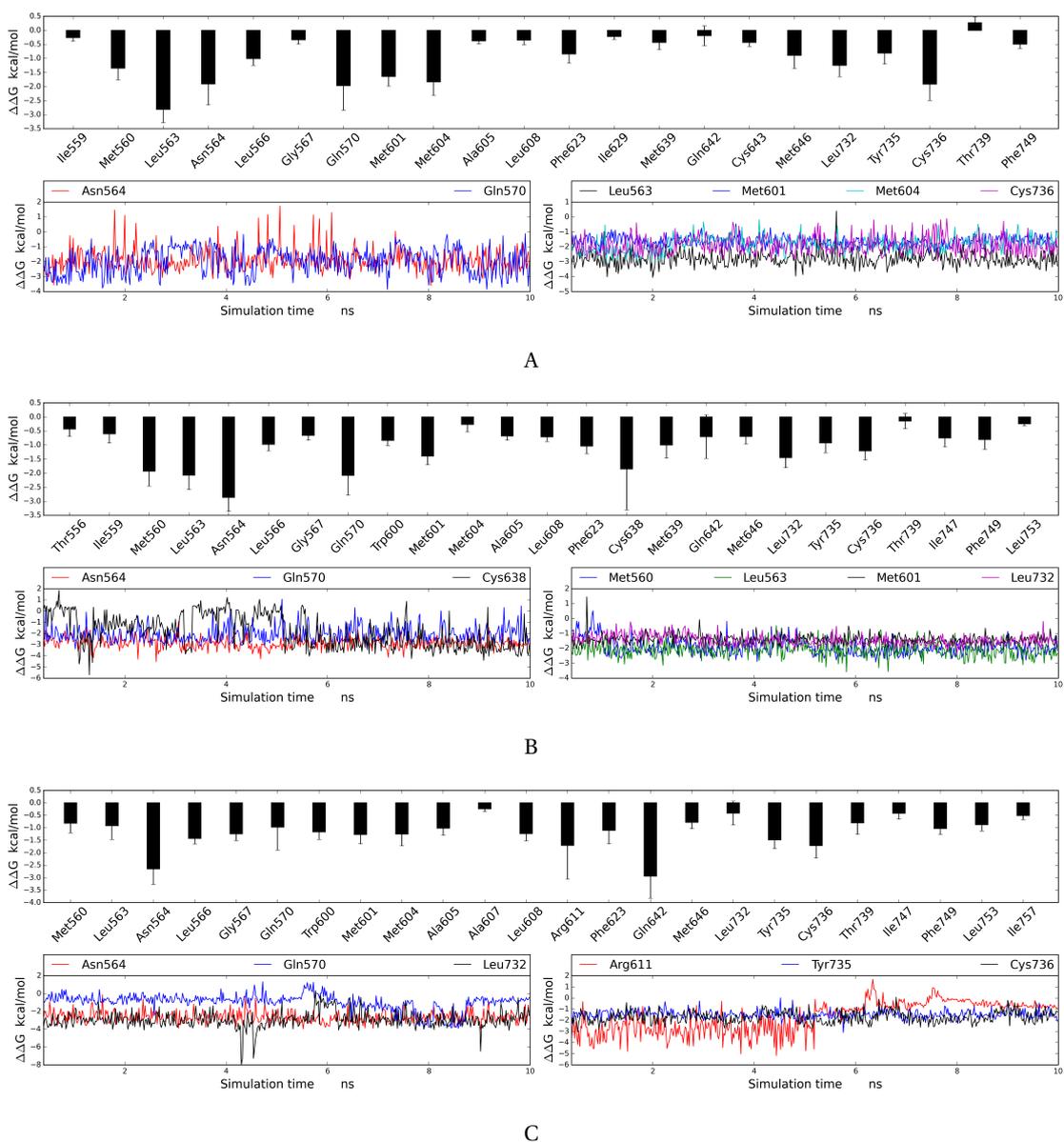


Figure 3.40: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues with respect of (*S*)-**71** (a), **85** (b) and **88** (c) binding to the GR.

3.3.5 Compounds active towards the mineralocorticoid receptor

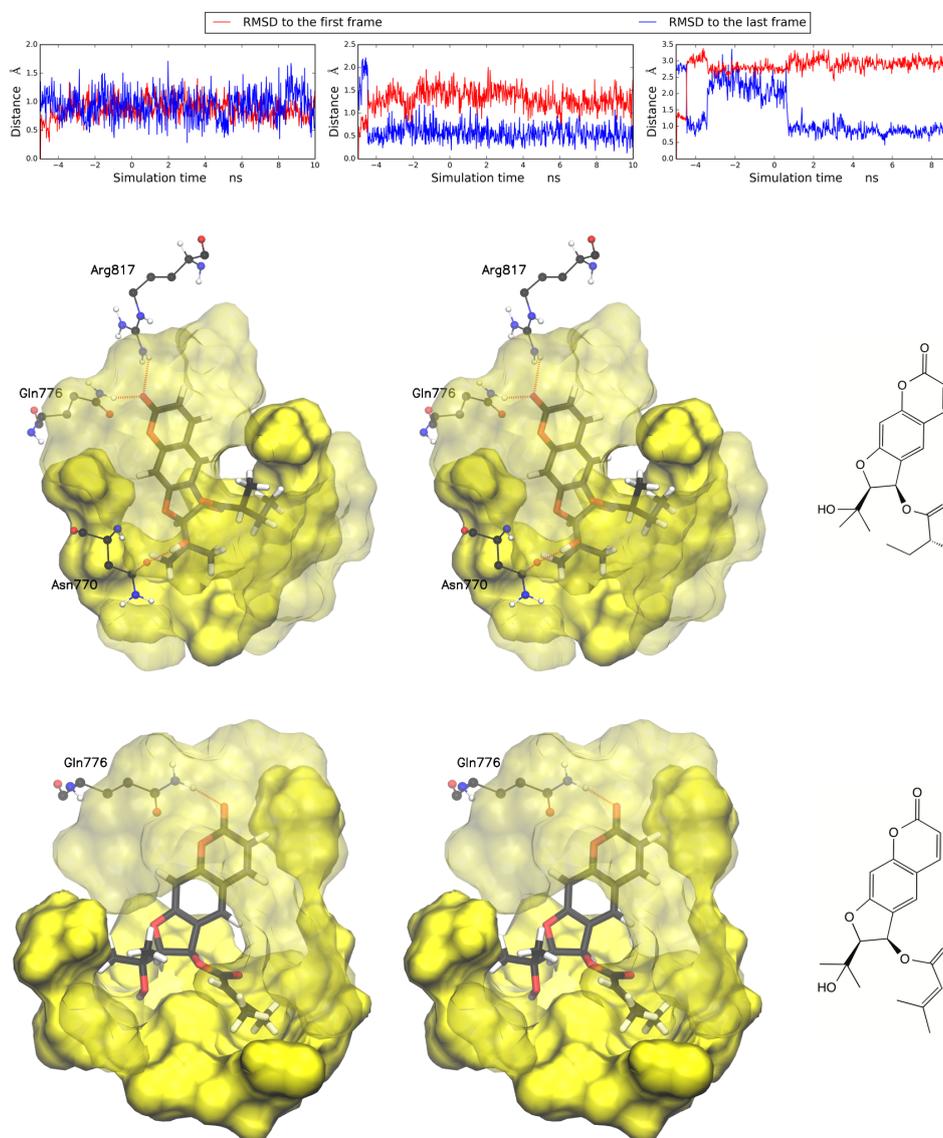


Figure 3.41: Top: Ligand RMSD of *R*-71, 72, and 98 with the MR; middle: details of the binding of 71S with the MR; bottom: details of the binding of 72 to the MR.

R-71, 72 and 98 are predicted to be active towards the MR, with computed IC_{50} of 73.4 nM, 84.8 nM and 15.0 nM respectively. After challenging the docking poses contributing most to the affinities (as identified by the *VirtualToxLab*) by means of 10 ns MD simulations, the

ligand RMSDs of *R*-71 and 72 remained below 1.5 Å, while the ligand RMSD of 98 increased to 3 Å during the equilibration stage, indicating that the conformations of *R*-71 and 72 are kinetically stable to the MR. The hydrogen bond distances between Gln776 with and the ligands are retained during the simulation (fig. 3.42). MM-PBSA results further suggest that the binding free energy of Gln776 is retained for both (*R*)-71 and 72.

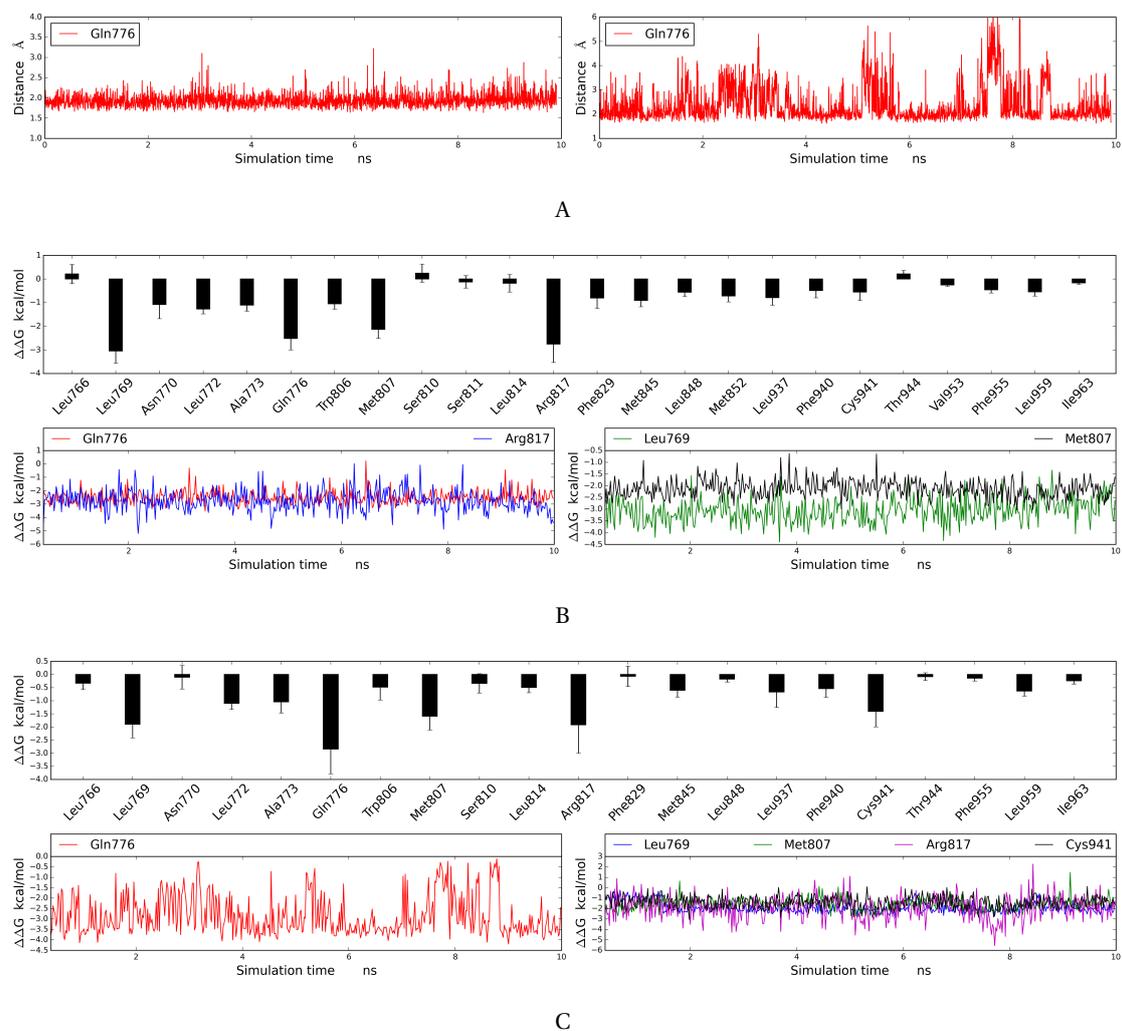


Figure 3.42: a: Hydrogen-bond distance of Gln776 with (*R*)-71 (left) and 72 (right). b and c: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues as respect of (*R*)-71 (b) and 72 (c) binding to the MR.

3.3.6 Compounds active towards the progesterone receptor

33, **34**, **39**, (*S*)-**71**, **74**, (*R*)-**76**, **77**, and **85** are predicted to be active towards the PR, with computed IC₅₀ values of 3.69 nM, 19.7 nM, 35.0 nM, 98.6 nM, 69.2 nM, 56.2 nM, 27.3 nM and 32.2 nM, respectively.

The docking poses of **33** and **34** contributing most to their respect binding affinities are quite similar (fig. 3.43). **33** features hydrogen bonds with both Arg766 and Thr894 while **34** engages only a hydrogen bond with Arg766 only.

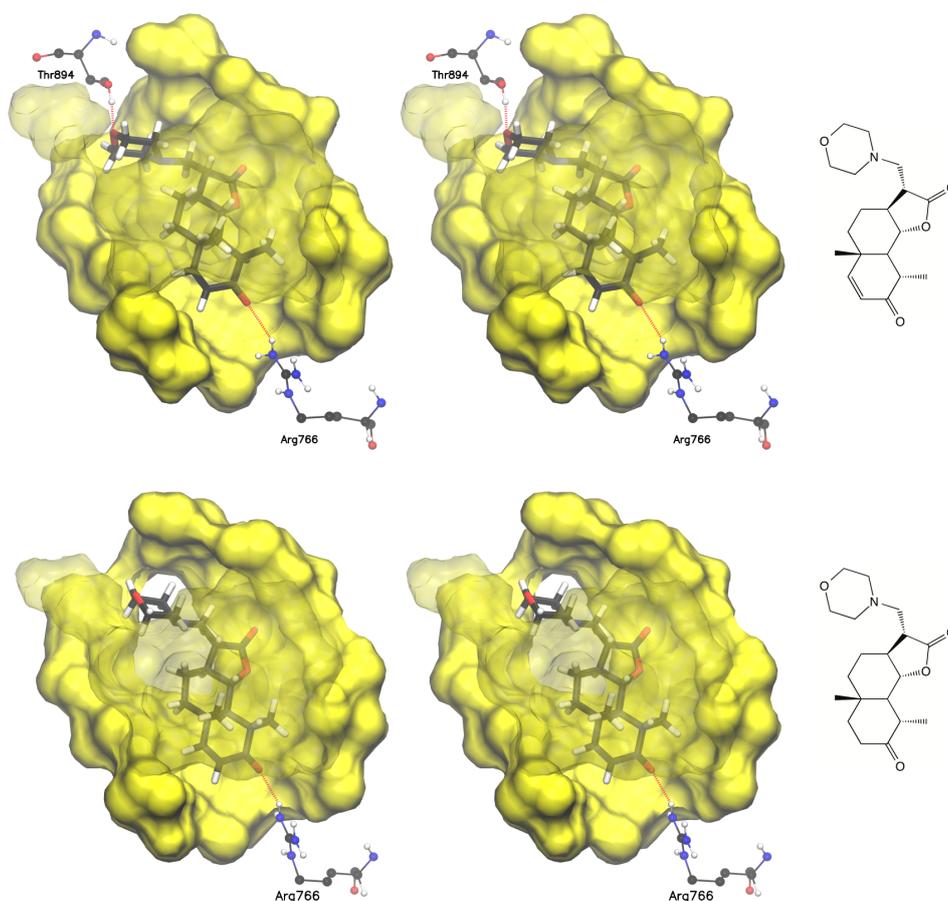


Figure 3.43: Details of the binding of **33** and **34** to the PR.

After challenging the binding by a 5.0 ns MD simulation: the ligand RMSDs are within 2.0 Å for both compounds (fig. 3.44A); the hydrogen-bond distances between Arg766 and the two ligands are retained at approximately 2.4 Å, while the hydrogen-bond between Thr894

and the ligands are not retained (fig. 3.44B). The MM-PBSA results (figs. 3.44C and 3.44D) further suggest that Arg766 provides stable free-energy contributions for **33** and **34** during the MD simulation.

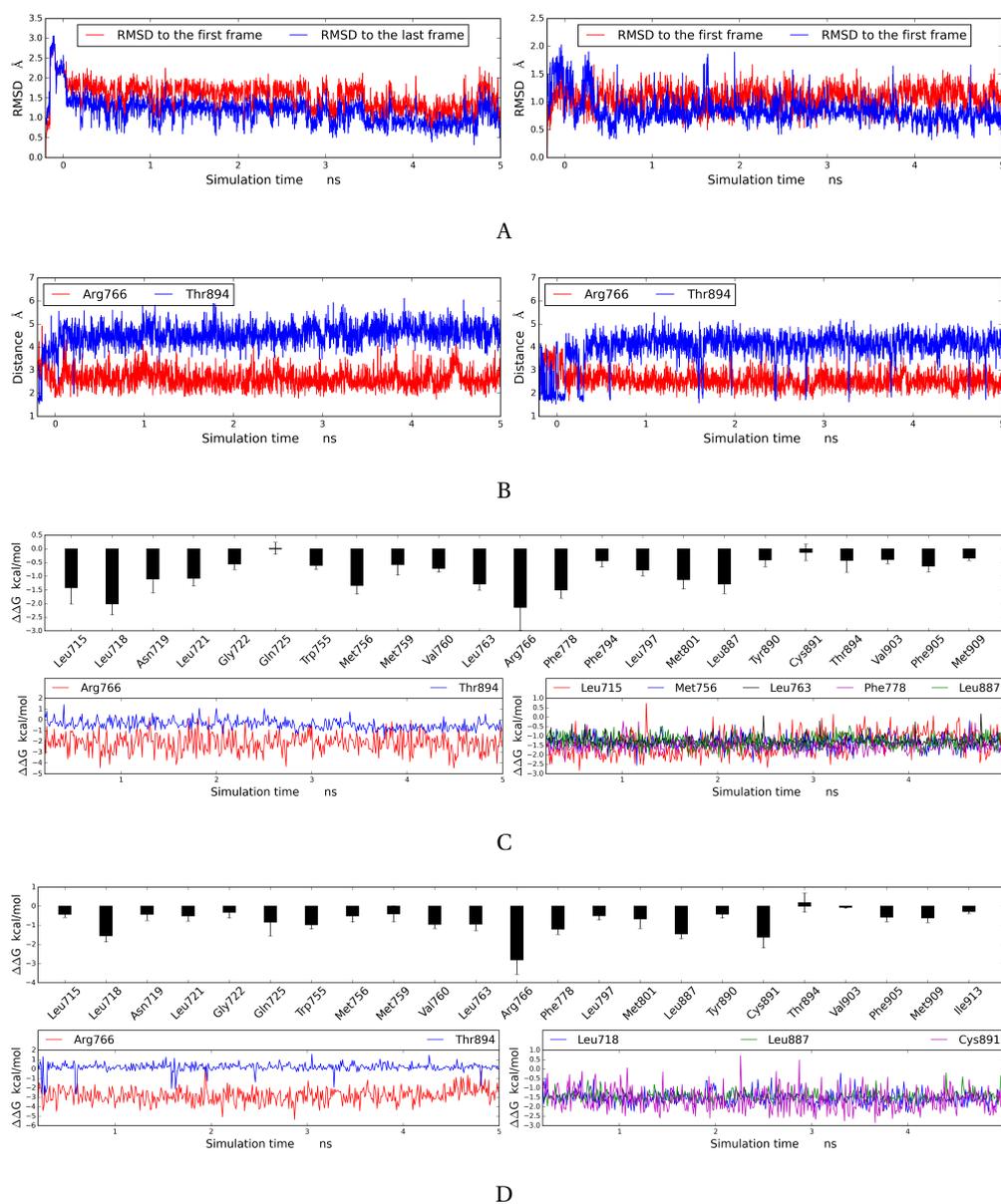


Figure 3.44: a: Ligand RMSD of **33** and **34** to the PR. b: key-residue hydrogen-bond distances of **33** and **34** to the PR. c and d: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues of the PR binding with **33** (c) and **34** (d).

39 is stabilized by engaging hydrogen bonds with Leu718, Gln725 and Cys891 while **85** is stabilized by Gln776, Arg817 and Met852, respectively (fig. 3.45).

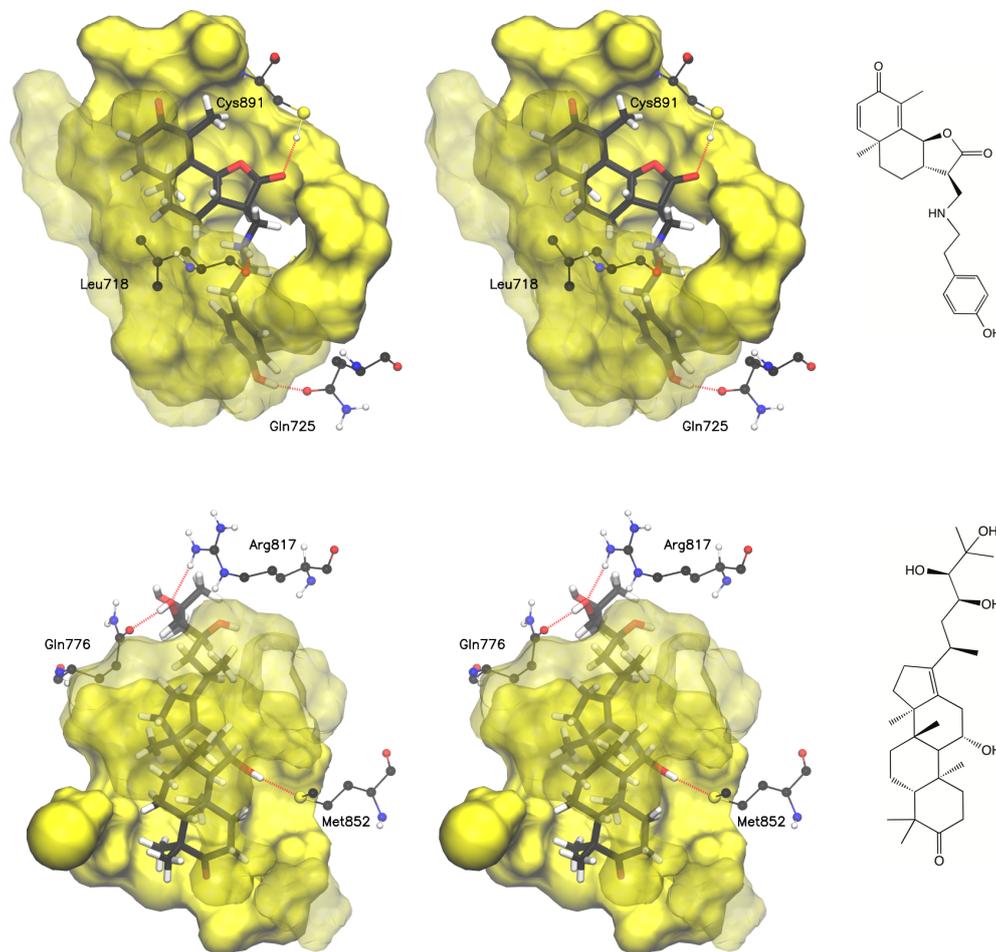
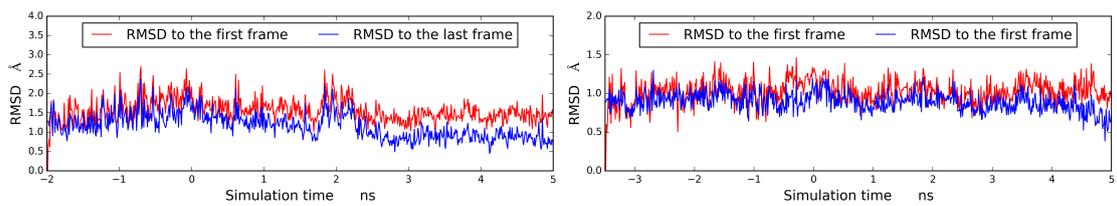
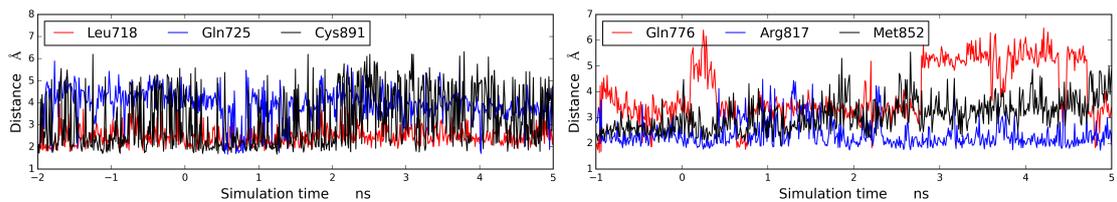


Figure 3.45: Details of the binding of **39** and **85** to the PR.

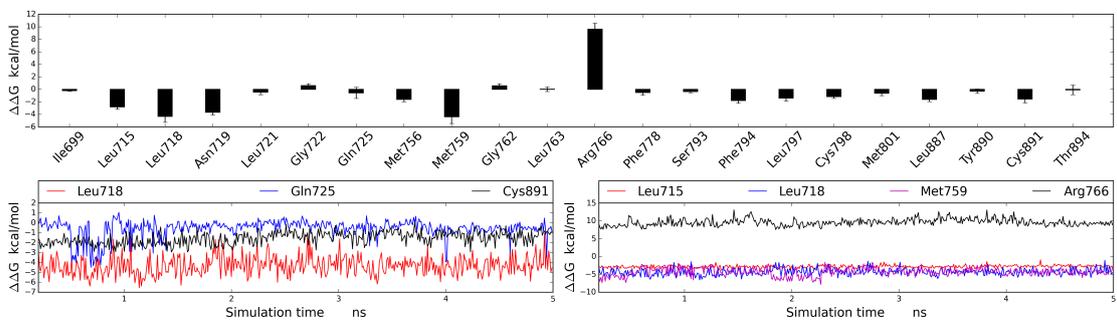
After challenging the binding by a 5.0 ns MD simulation: the ligand RMSDs are within 2.0 Å for both ligands (fig. 3.46A); the hydrogen-bond distance between Leu718 and **39** is retained at approximately 2.5 Å, the hydrogen-bond between Arg817 and **85** is kept at approximately 2.3 Å (fig. 3.46B), the other hydrogen bonds are not retained after MD simulation. MM-PBSA results (figs. 3.46C and 3.46D) further suggest that Arg766 features an unfavorable $\Delta\Delta G$ contribution for **33**, the $\Delta\Delta G$ fluctuation of Arg766 toward **85** is quite high, which suggests that these two poses are less likely kinetically stable.



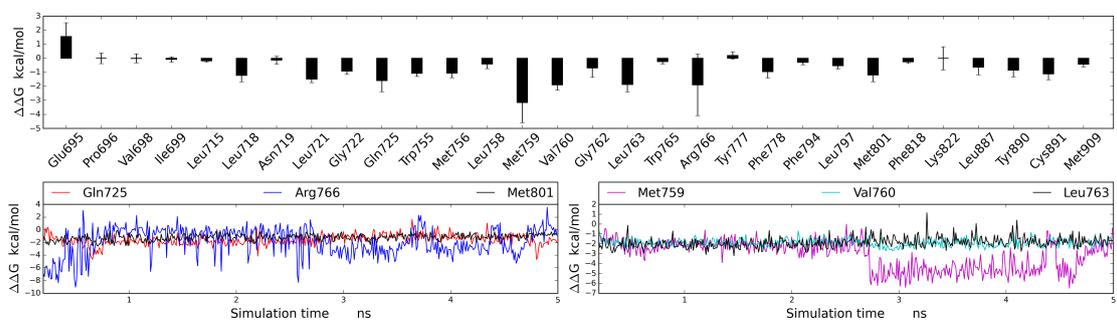
A



B



C



D

Figure 3.46: a: Ligand RMSD of **39** and **85** to the PR. b: key-residue hydrogen-bond distances of **39** and **85** to the PR. c and d: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues of the PR binding with **39** (c) and **85** (d).

For (*S*)-**71**, **74**, (*R*)-**76** and **77**, the docking poses contributing most to the binding affinity are quite similar, which indicates that the interactions are conservative in the nature. The compounds are all stabilized by engaging a hydrogen bond with Arg766. After challenging the binding by 5 ns of MD simulation: the ligand RMSDs are within 2.0 Å, and the hydrogen bond is retained for all the poses.

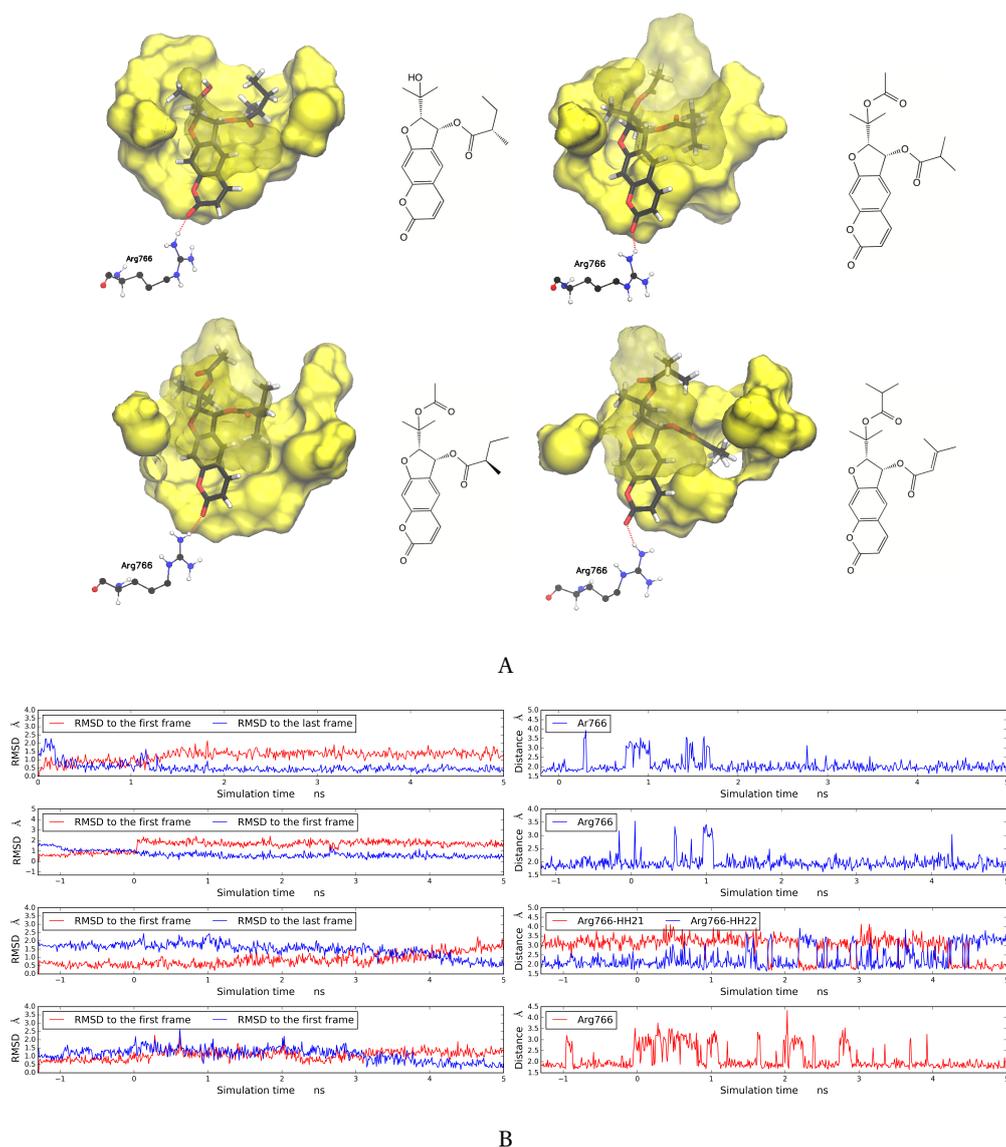


Figure 3.47: a: Details of the binding of (*S*)-**71** (top left), **74** (top right), (*R*)-**76** (bottom left) and **77** (bottom right) to the PR. b: Ligand RMSD (left) and key-residue hydrogen-bond distances (right) to the PR, from top to bottom: (*S*)-**71**, **74**, (*R*)-**76** and **77**.

The MM-PBSA results further suggest the free-energy contribution within the binding pocket is similar for all the four ligands. Arg766 substantially contributes to the major $\Delta\Delta G$ throughout the MD simulation for all ligands. Furthermore, Leu718 is another major contributor for $\Delta\Delta G$ mainly due to hydrophobic interactions.

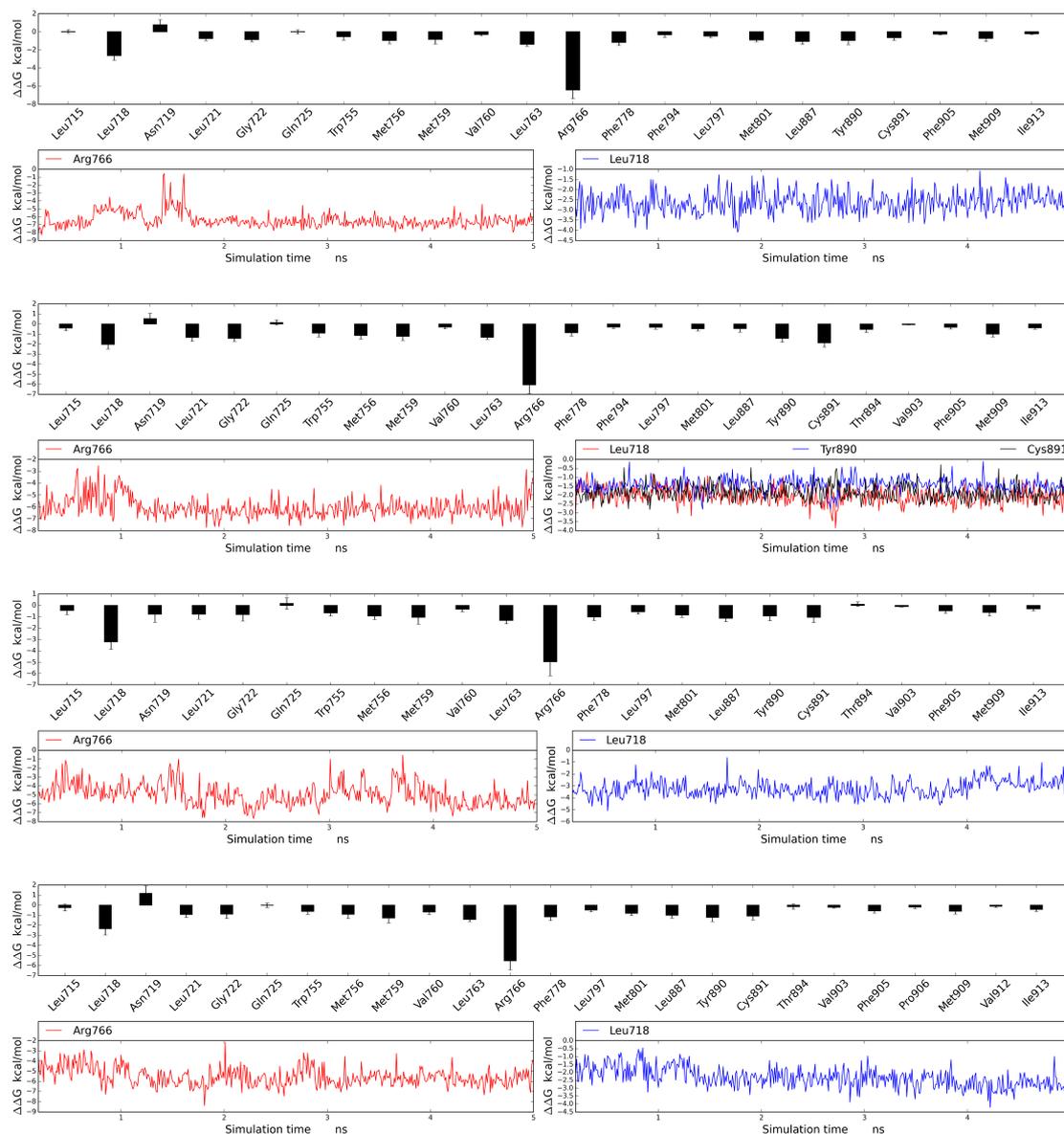


Figure 3.48: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues, from top to bottom: (*S*)-71, 74, (*R*)-76 and 77.

3.3.7 The only compound active toward the thyroid receptor β

From the *VirtualToxLab* screening, alisol A (**85**) was identified as active towards the TR β (computed IC₅₀ value = 33.5 nM). The docking pose contributing most to the binding affinity is shown in fig. 3.49. The hydrophobic skeleton (6-6-6-5 ring system) is stabilized by the hydrophobic residues lining in the binding pocket. Alisol A is further stabilized by engaging a hydrogen bond with Arg282.

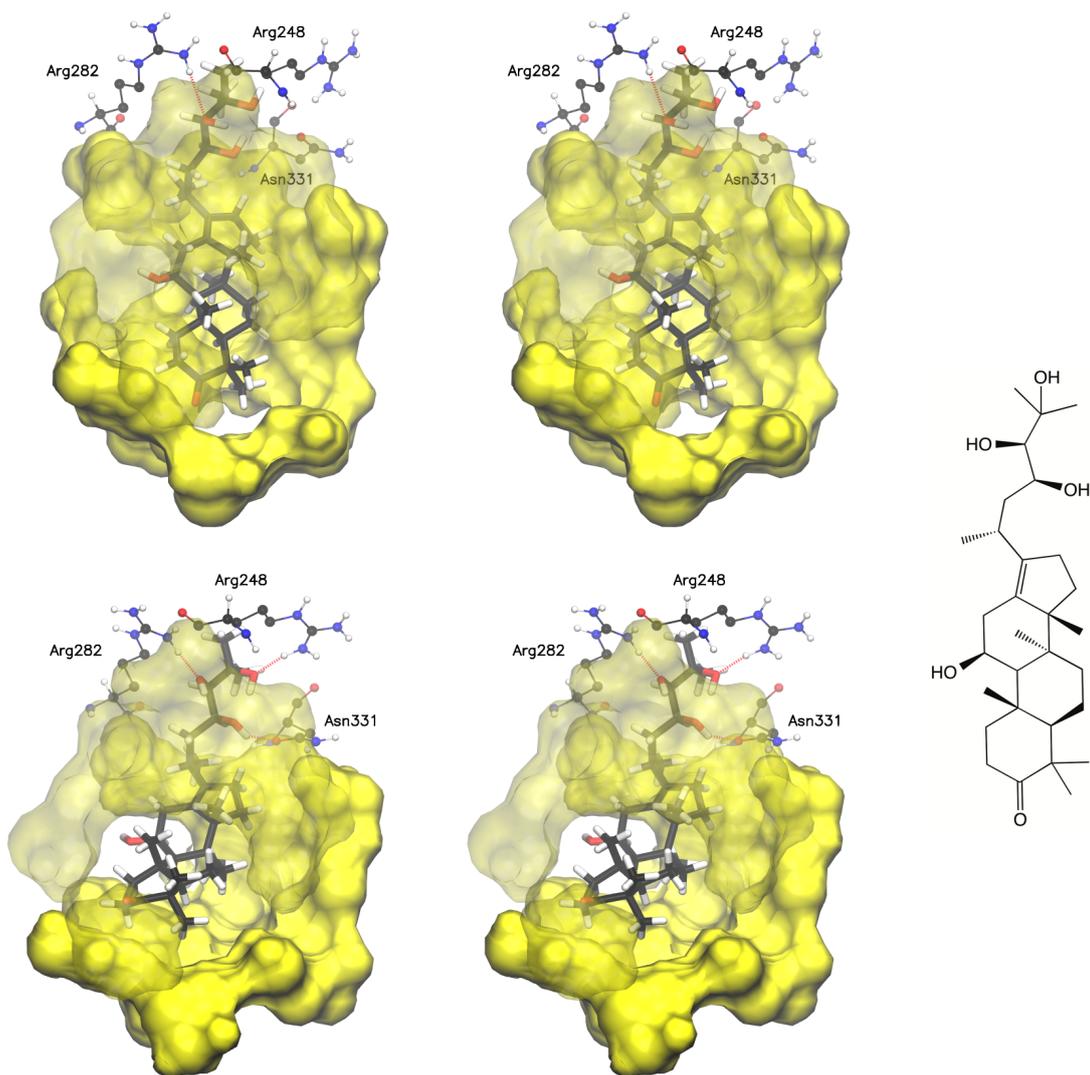


Figure 3.49: Details of the binding of **85** to the TR β by the *VirtualToxLab* (top) and the last MD frame (bottom).

After challenging the binding by means of 5.0 ns MD simulation, the ligand RMSD is within 2.0 Å compared with the first frame (fig. 3.50A). Aside from Arg282, the three adjacent hydroxyl group further engaged stable hydrogen bonds with Arg306 and Asn331, the bond distances are retained throughout the MD simulation (figs. 3.49 and 3.50B). The MM-PBSA results (fig. 3.50C): Arg282, Arg316 and Asn332 would seem to be the major contributors for the binding, as well as Met313 and His435 also yield favorable contributions mainly due to hydrophobic interactions.

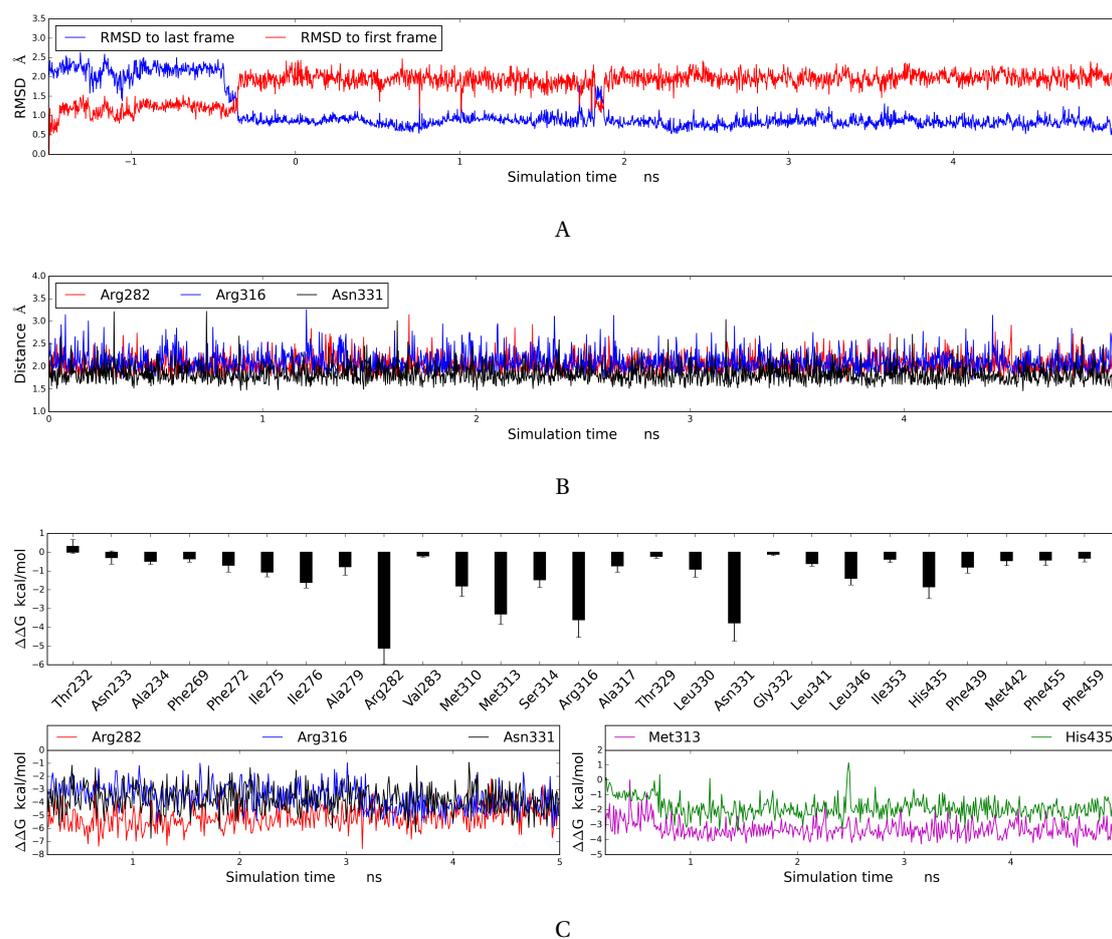


Figure 3.50: a: Ligand RMSD relative to the first and the last simulation frame. b: Hydrogen-bond distances of Arg282, Lys306 and Asn331 with alisol A. c: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues for **85** to the TR β .

3.3.8 Compounds active toward the Cytochrome P450 2D6

37, **80** and **88** were identified as active towards the CYP2D6, with computed IC_{50} values of 12.6 nM, 7.86 nM and 3.61 nM, respectively. The docking pose contributing most to the binding affinity is shown in fig. 3.51.

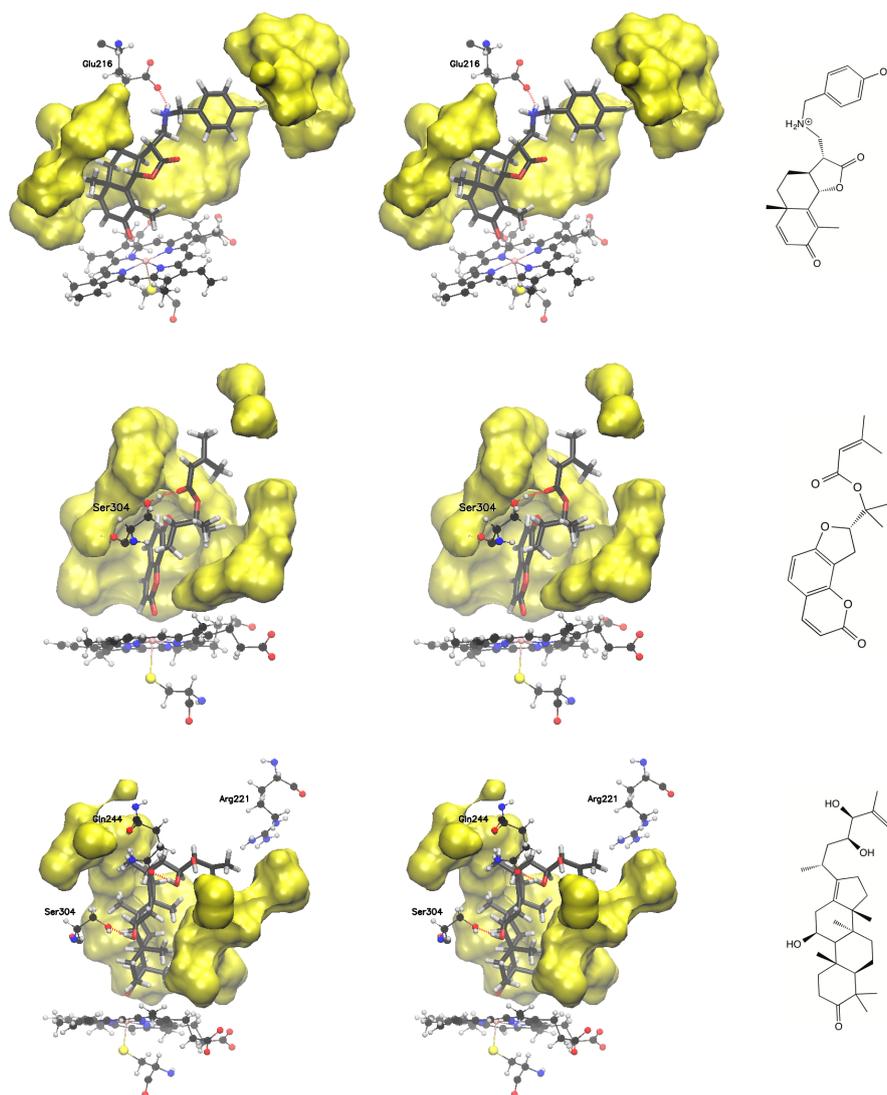
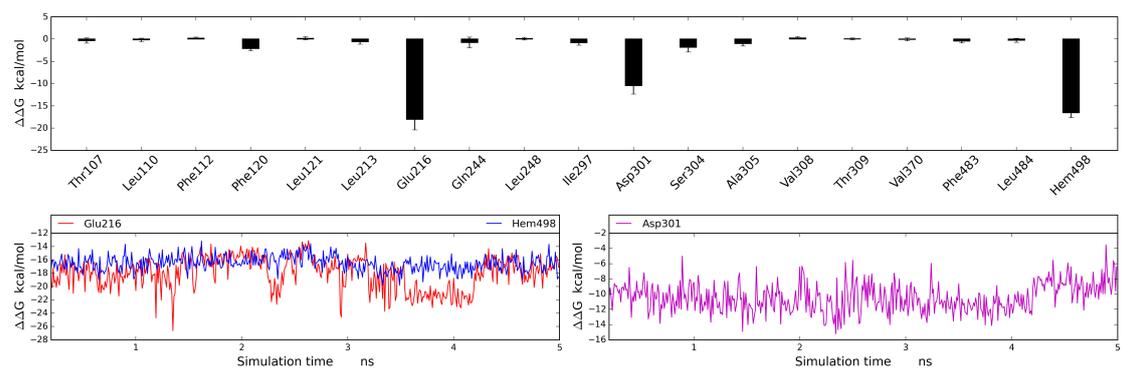
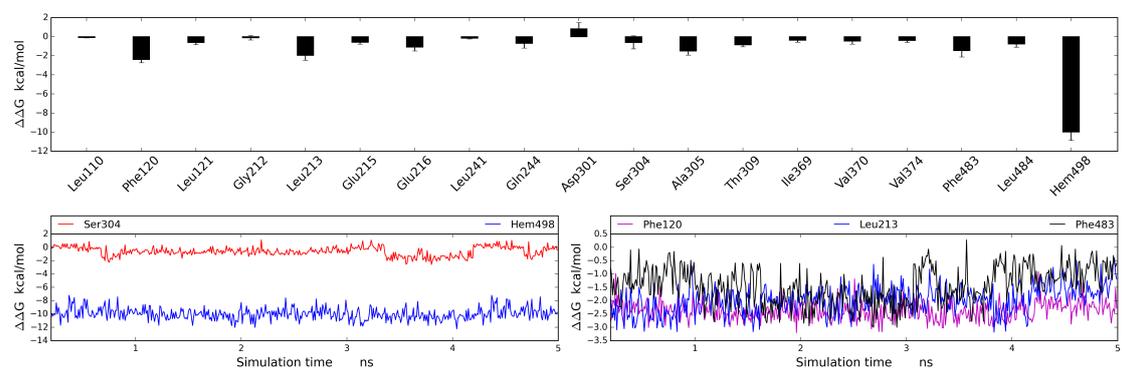


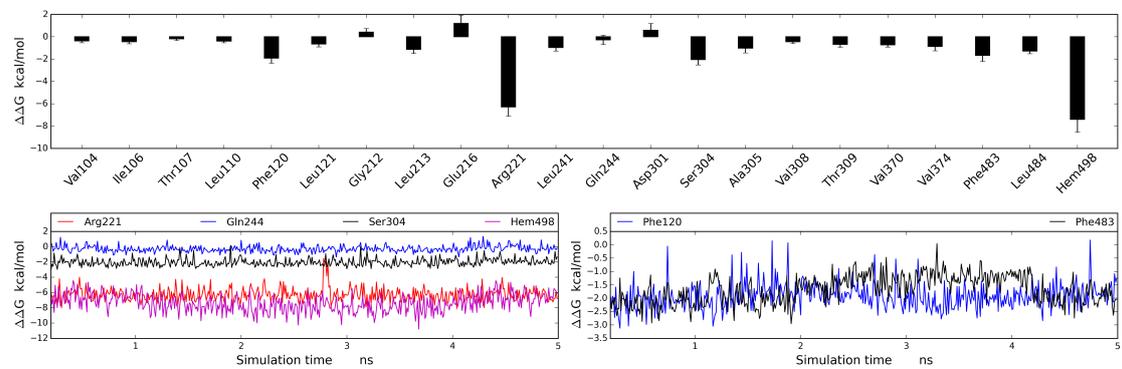
Figure 3.51: Details of the binding of **37**, **80** and **88** to the CYP2D6.



A



B



C

Figure 3.52: Binding-energy decomposition within the binding pocket and time-resolved details for key amino-acid residues for **37** (a), **80** (b) and **88** (c) to the CYP2D6.

Chapter 4

Conclusions

Employing a genetic algorithm, a 4D scoring function based on a total of 1,288 compounds binding to 16 different proteins (enzymes, receptors, ion channels) has been developed and compared against existing concepts. In the new function, the interaction ligand–protein energies are decomposed into electrostatic, van der Waals, hydrogen bonding and polarization components, which allows for a direct estimation of the associated binding affinity. This algorithm can be employed in consensus-scoring mode to existing concepts for the prediction of binding affinities based on three-dimensional ligand–protein structures. Current limitations of the scoring function include larger induced-fit movements and, if present, substantial entropic contribution of solvent released upon ligand binding.

Using the *Quasar* software, a QSAR model for the pregnane X receptor (PXR) was developed and validated. The binding modes of the individual compounds were either obtained from a crystal structure or, when absent, identified by interactive docking, followed by extensive molecular-dynamics simulations. These poses were subsequently employed as templates for the flexible docking of 101 compounds (comprising eight chemical classes) and the generation of the final model. It converged at a cross-validated r^2 of 0.812 (for 72 training compounds) and yielded a predictive r^2 of 0.854 (for 29 test compounds). Consensus scoring with the *Raptor* software yielded corresponding values of 0.870 and 0.646, respectively. This suggests that the models can be applied to predict the binding affinity of novel drug candidates towards the PXR, which will be extremely valuable in the early stages of the drug-discovery process. At the molecular level, residues

Asn205, His247, Gln285, His407 and Arg410 would seem to play a key role for stabilizing potential ligand molecules through hydrogen bonds. Hydrophobic stabilization of the ligand–protein complex, on the other hand, is mainly achieved through the residues Phe288, Trp299 and Tyr306. Limitations of the models involve the applicability domain, which does only include neutral species. The availability of high-quality experimental data turned out to be a limiting factor for the choice of a larger dataset. The application of interactive docking is not adequate for high-throughput screening, where automated procedures are necessary. Further improvement of the automated docking process could be achieved by implementation of knowledge-based rules. In summary, this model represents a basis for a future extension of the VirtualToxLab — an *in silico* tool for predicting side effects and toxicity of drugs, chemicals and natural compounds.

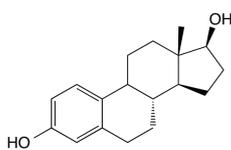
Potential side effects and toxicity of anti-trypanosomiasis active compounds were investigated using the *VirtualToxLab*. This technology identifies the binding mode of a small-molecule compound toward a series of 16 target proteins (nuclear receptors, cytochrome P450 enzymes, hERG, AhR) known or suspected to trigger adverse effects. As this technology provides thermodynamic information only, all relevant ligand–protein complexes were challenged by subsequent molecular-dynamics simulations. Sesquiterpene lactones showed a potential affinity toward the androgen, glucocorticoid and the mineralocorticoid receptor. Tanshinone and its derivatives displayed an affinity toward the aryl-hydrocarbon and the thyroid receptor β . Alisol A and its derivatives bind to the androgen and the glucocorticoid receptor. Smyrniolin and its derivatives showed a substantial affinity toward the glucocorticoid, mineralocorticoid and the progesterone receptor. Isoflavan showed a high affinity toward the thyroid receptor α . This protocol would seem to be a promising approach for probing the interactions of ligands with relevant target proteins.

Appendix A

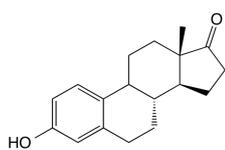
Chemical structure of the PXR ligands

A.1 Chemical structures

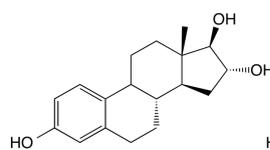
A.1.1 Estratrienes (E01–E08)



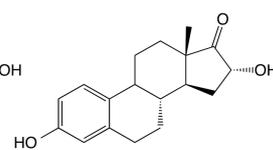
E01



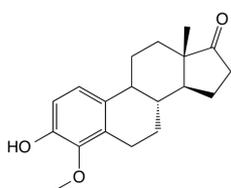
E02



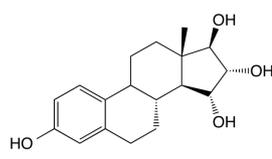
E03



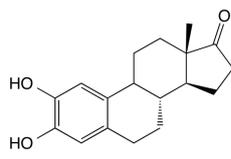
E04



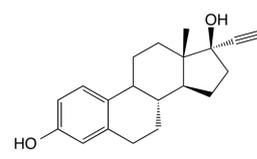
E05



E06

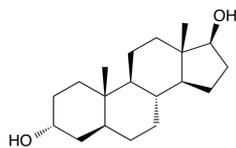


E07

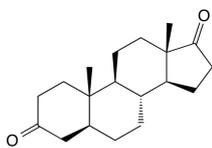


E08

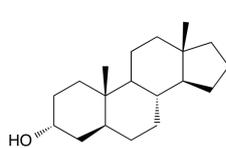
A.1.2 Androstans (A01–A18)



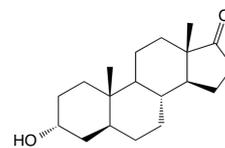
A01



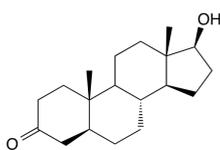
A02



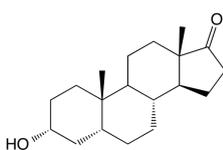
A03



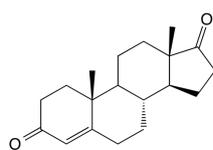
A04



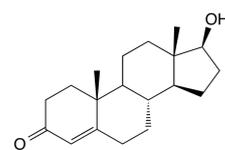
A05



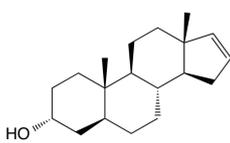
A06



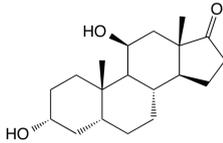
A07



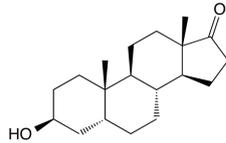
A08



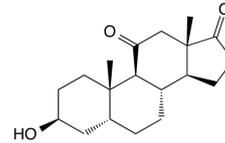
A09



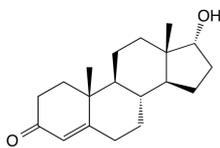
A10



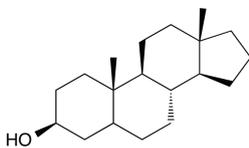
A11



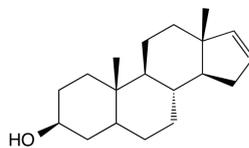
A12



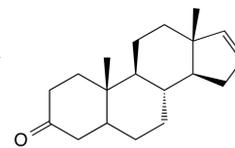
A13



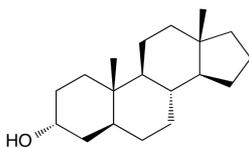
A14



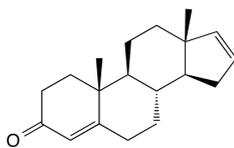
A15



A16

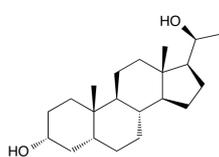


A17

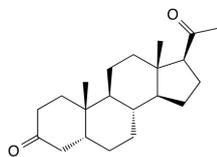


A18

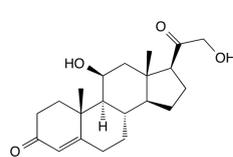
A.1.3 Pregnanes (P01–P23)



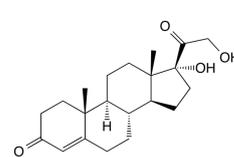
P01



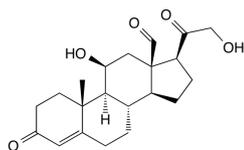
P02



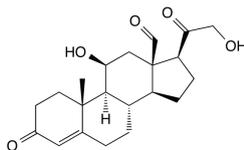
P03



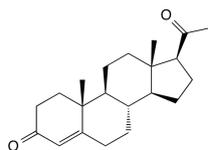
P04



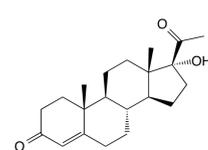
P05



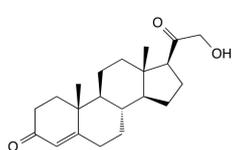
P06



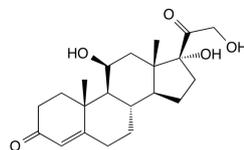
P07



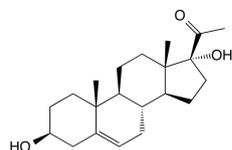
P08



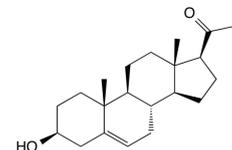
P09



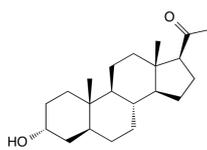
P10



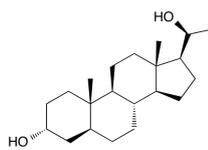
P11



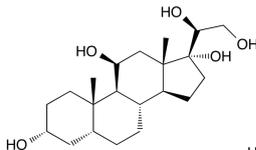
P12



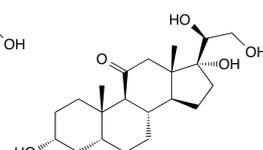
P13



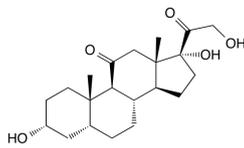
P14



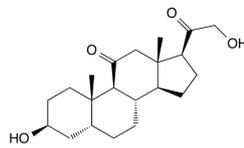
P15



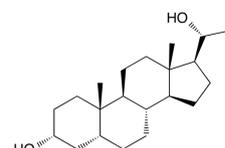
P16



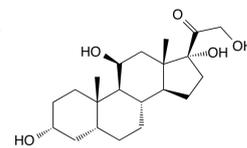
P17



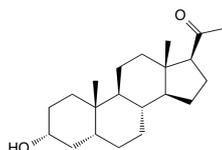
P18



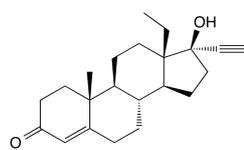
P19



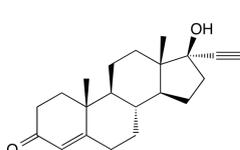
P20



P21

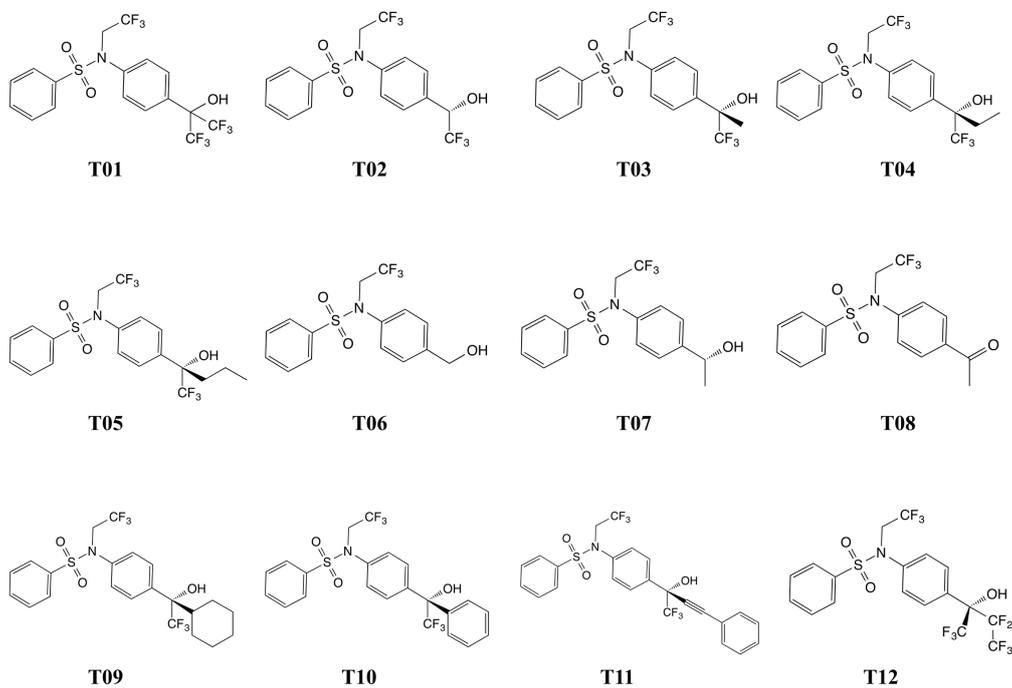


P22

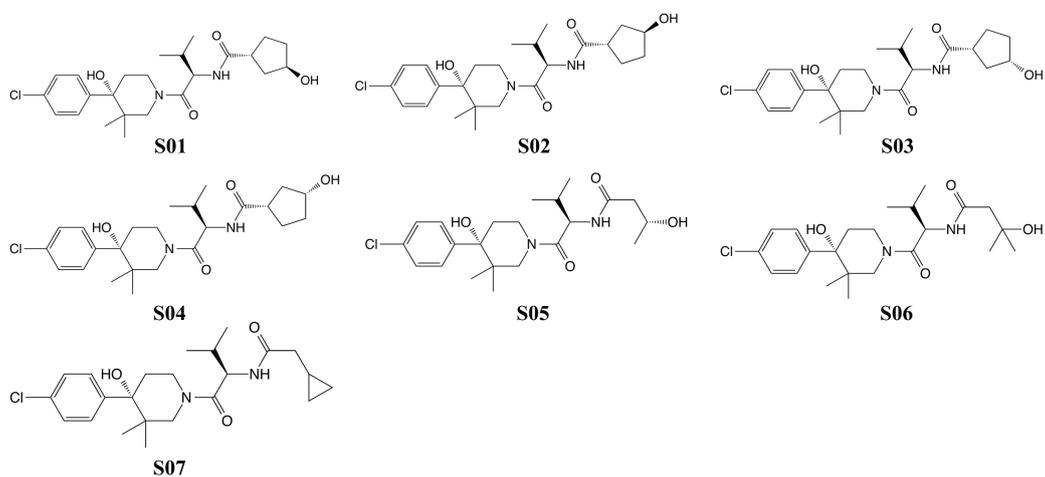


P23

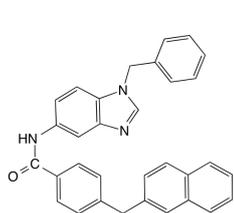
A.1.4 T091317 derivatives (T01–T12)



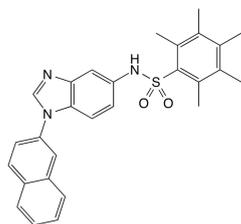
A.1.5 BMS-817399 derivatives (S01–S07)



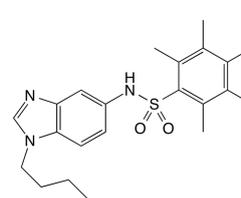
A.1.6 Benzenesulfonamide derivatives (M01–M09)



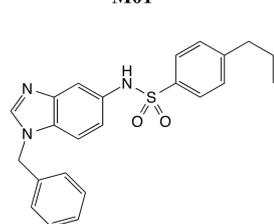
M01



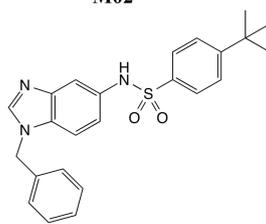
M02



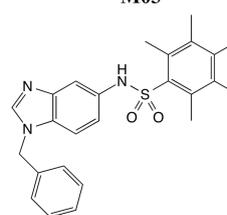
M03



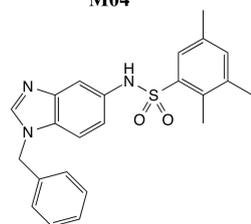
M04



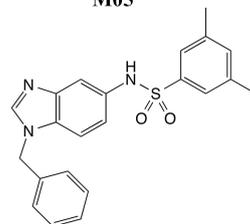
M05



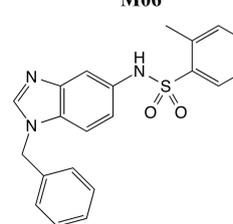
M06



M07

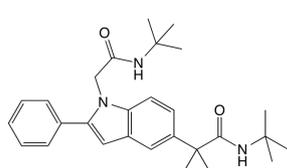


M08

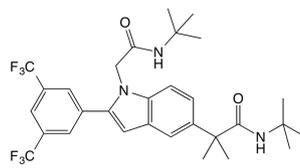


M09

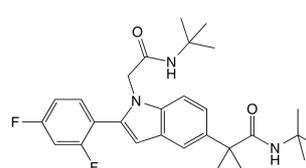
A.1.7 2-Aryl indoles (N01–N08)



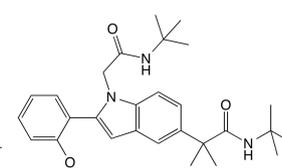
N01



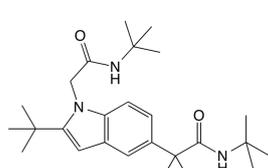
N02



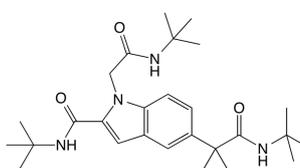
N03



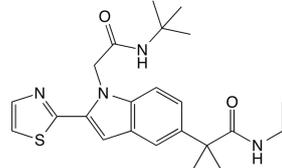
N04



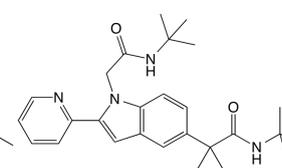
N01



N02

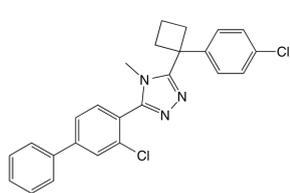


N03

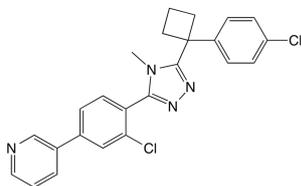


N04

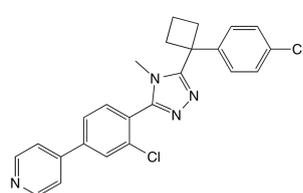
A.1.8 Substituted phenyl triazoles (G01–G16)



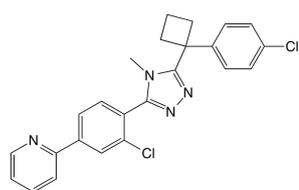
G01



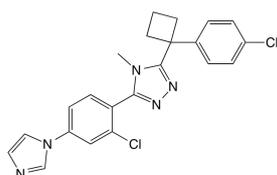
G02



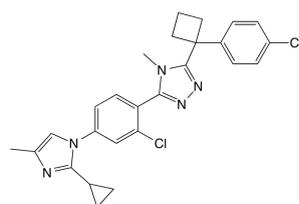
G03



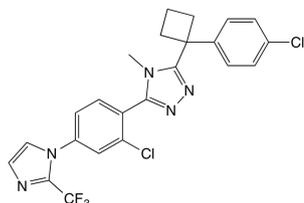
G04



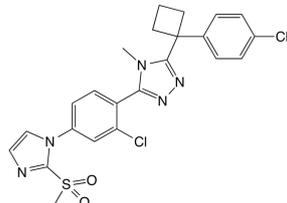
G05



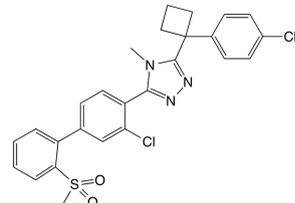
G06



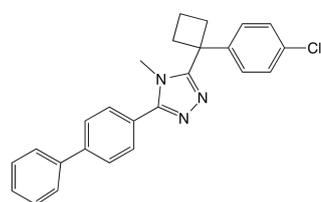
G07



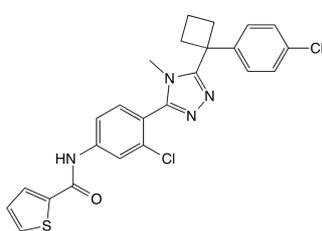
G08



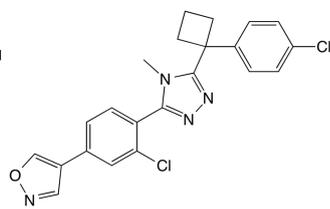
G09



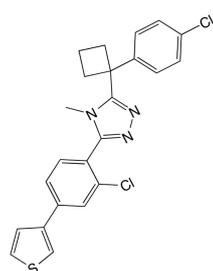
G10



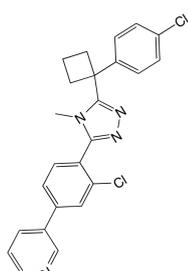
G11



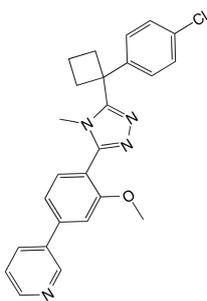
G12



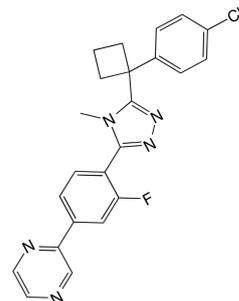
G13



G14



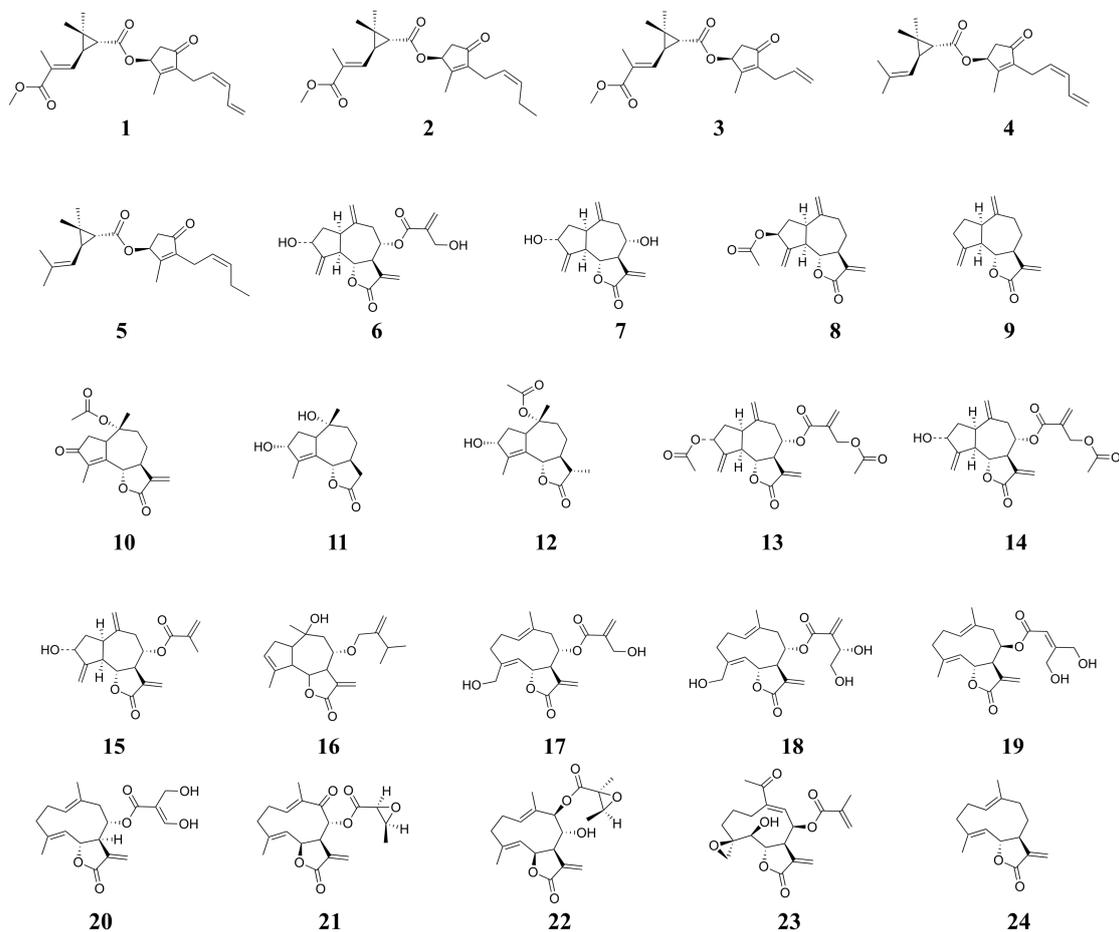
G15

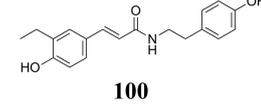
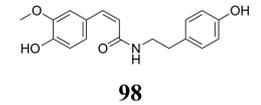
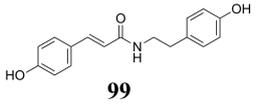
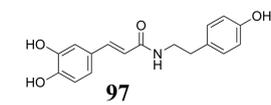
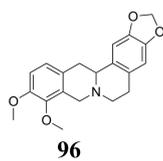
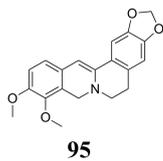
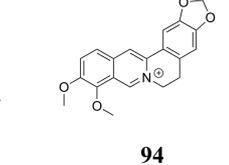
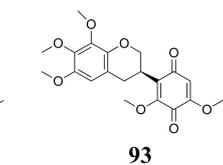
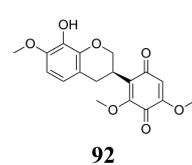
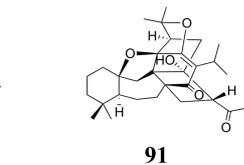
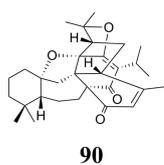
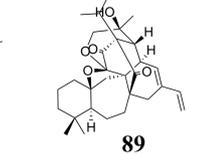
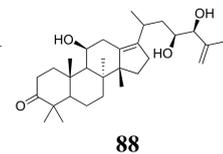
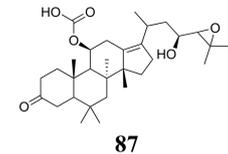
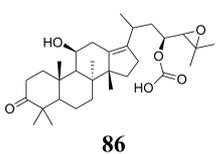
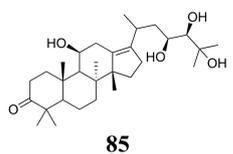
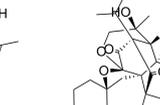
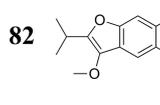
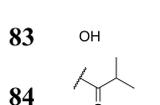
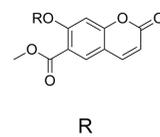
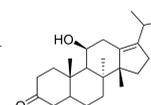
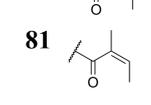
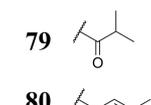
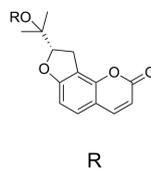
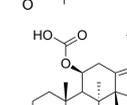
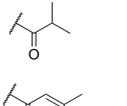
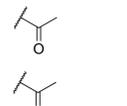
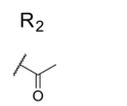
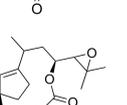
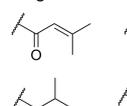
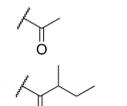
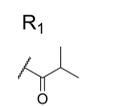
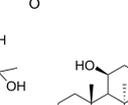
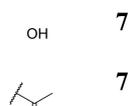
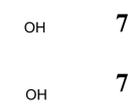
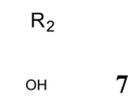
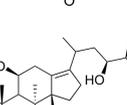
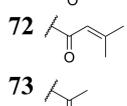
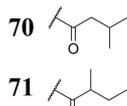
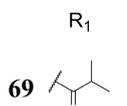
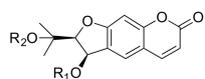
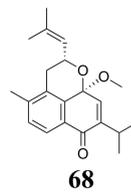
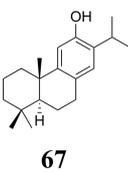
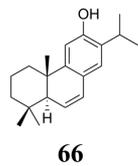
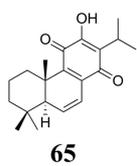
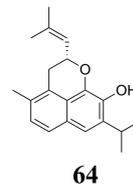
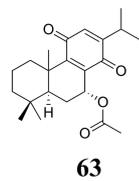
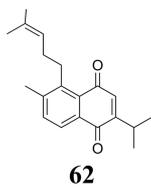
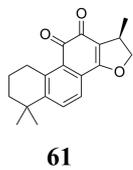
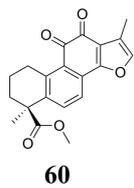


G16

Appendix B

Anti-trypansomal compounds





Bibliography

- [1] Hassan, S. A. *Protein–Ligand interactions, methods and applications*, volume 305 of *Methods In Molecular Biology*. Humana Press, 2005.
- [2] Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. a., Sanschagrin, P. C., and Mainz, D. T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.*, 49:6177–6196, 2006.
- [3] Smieško, M. DOLINA - Docking based on a local induced-fit algorithm: Application toward small-molecule binding to nuclear receptors. *J. Chem. Info. Model.*, 53:1415–1423, 2013.
- [4] Sousa, S. E., Ribeiro, a. J. M., Coimbra, J. T. S., Neves, R. P. P., Martins, S. a., Moorthy, N. S. H. N., Fernandes, P. a., and Ramos, M. J. Protein-ligand docking in the new millennium—a retrospective of 10 years in the field. *Curr. Med. Chem.*, 20:2296–314, 2013.
- [5] Sherman, W., Day, T., Jacobson, M. P., Friesner, R. a., and Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.*, 49:534–553, 2006.
- [6] Ding, E, Yin, S., and Dokholyan, N. V. Rapid flexible docking using a stochastic rotamer library of ligands. *J. Chem. Inf. Model.*, 50:1623–1632, 2010.
- [7] Vedani, A. and Dobler, M. 5D-QSAR: the key for simulating induced fit? *J. Med. Chem.*, 45:2139–2149, 2002.
- [8] Rossato, G., Ernst, B., Vedani, A., and Smieško, M. AcquaAlta: A directional approach to the solvation of ligand-protein complexes. *J. Chem. Info. Model.*, 51:1867–1881, 2011.

- [9] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selected receptor flexibility. *J. Comput. Chem.*, 30:2875–2971, 2009.
- [10] *Induced Fit Docking protocol 2015-1, Glide version 6.4, Prime version 3.7*. Schrödinger, LLC, New York, 2015.
- [11] Baron, R., editor. *Computational drug discovery and design*, volume 819 of *Methods in Molecular Biology*. Humana Press, 2012.
- [12] Chaudhuri, R., Carrillo, O., Laughton, C. A., and Orozco, M. Application of drug-perturbed essential dynamics/molecular dynamics (ED/MD) to virtual screening and rational drug design. *J. Chem. Theory Comput.*, 8:2204–2214, 2012.
- [13] Klebe, G. Applying thermodynamic profiling in lead finding and optimization. *Nat. Rev. Drug Discovery*, 14:95–110, 2015.
- [14] Zheng, M., Li, Y., Xiong, B., Jiang, H., and Shen, J. Water PMF for predicting the properties of water molecules in protein binding site. *J. Comput. Chem.*, 34:583–592, 2013.
- [15] Abel, R., Young, T., Farid, R., Berne, B. J., and Friesner, R. a. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.*, 130:2817–2831, 2008.
- [16] Hu, B. and Lill, M. a. WATsite: hydration site prediction program with PyMOL interface. *J. Comput. Chem.*, 35:1255–60, 2014.
- [17] Yuriev, E. and Ramsland, P. a. Latest developments in molecular docking: 2010-2011 in review. *J. Mol. Recognit.*, 26:215–239, 2013.
- [18] Thomsen, R. and Christensen, M. H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem.*, 49(11):3315–3321, 2006.
- [19] <http://www.biograf.ch>.
- [20] Liu, J. and Wang, R. Classification of Current Scoring Functions. *J. Chem. Info. Model.*, 55:475–482, 2015.

- [21] Kitchen, D. B., Decornez, H., Furr, J. R., and Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discovery*, 3:935–949, 2004.
- [22] Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.*, 72:1047–69, 1997.
- [23] Zou, X., Sun, Y., and Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using the generalized-born model. *J. Am. Chem. Soc.*, 121:8033–8043, 1999.
- [24] Rocchia, W. and Spagnuolo, M. *Computational Electrostatics for Biomolecular Systems*. Springer, 2014.
- [25] Wichapong, K., Rohe, A., Platzer, C., Slynko, I., Erdmann, F., Schmidt, M., and Sippl, W. Application of docking and QM/MM-GBSA rescoring to screen for novel Myt1 kinase inhibitors. *J. Chem. Info. Model.*, 54:881–93, 2014.
- [26] Böhm, H.-J. and Stahl, M. *Reviews in computational chemistry*, volume 18. Wiley, 2002.
- [27] Xie, Z.-R. and Hwang, M.-J. An interaction-motif-based scoring function for protein-ligand docking. *BMC Bioinformatics*, 11:298, 2010.
- [28] Neudert, G. and Klebe, G. DSX: A Knowledge-based scoring function for the assessment of protein–ligand complexes. *J. Chem. Info. Model.*, 51:2731–2745, 2011.
- [29] Böhm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput.-Aid. Mol. Des.*, 8:243–256, 1994.
- [30] *Glide, version 6.6*. Schrödinger, LLC, New York, 2015.
- [31] Brylinski, M. Nonlinear scoring functions for similarity-based ligand docking and binding affinity prediction. *J. Chem. Info. Model.*, 53:3097–112, 2013.
- [32] Zilian, D. and Sotri, C. a. A random forest–based scoring function for improved affinity prediction of protein–ligand complexes. *J. Chem. Info. Model.*, 53:1923–1933, 2013.

- [33] Li, G.-B., Yang, L.-l., Wang, W.-J., Li, L.-L., and Yang, S.-Y. ID-Score: A new empirical scoring function based on a comprehensive set of descriptors related to protein–ligand interactions. *J. Chem. Info. Model.*, 53:592–600, 2013.
- [34] Klebe, G. *Drug Design: Methodology, concepts, and mode-of-action*. Springer, 2013.
- [35] Cramer, R. D., Patterson, D. E., and Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.*, 110:5959–5967, 1988.
- [36] Tanrikulu, Y. and Schneider, G. Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening. *Nat. Rev. Drug Discovery*, 7:667–77, 2008.
- [37] Vedani, A., Dobler, M., and Lill, M. A. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J. Med. Chem.*, 48:3700–3703, 2005.
- [38] Vedani, A., Dobler, M., Hu, Z., and Smieško, M. OpenVirtualToxLab—A platform for generating and exchanging in silico toxicity data. *Toxicol. Lett.*, 232:519–532, 2015.
- [39] Willson, T. M. and Kliewer, S. A. PXR, CAR and drug metabolism. *Nat. Rev. Drug Discovery*, 1:259–266, 2002.
- [40] Löscher, W. and Potschka, H. Drug resistance in brain diseases and the role of drug efflux transporters. *Nat. Rev. Neuroscience*, 6:591–602, 2005.
- [41] Ekins, S. and Schuetz, E. The PXR crystal structure: The end of the beginning. *Trends Pharmacol. Sci.*, 23:49–50, 2002.
- [42] Watkins, R. E., Wisely, G. B., Moore, L. B., Collins, J. L., Lambert, M. H., Williams, S. P., Willson, T. M., Kliewer, S. A., and Redinbo, M. R. The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science*, 292:2329–33, 2001.
- [43] Ngan, C.-H., Beglov, D., Rudnitskaya, A. N., Kozakov, D., Waxman, D. J., and Vajda, S. The structural basis of pregnane X receptor binding promiscuity. *Biochemistry*, 48:11572–81, 2009.

- [44] Ekins, S., Kortagere, S., Iyer, M., Reschly, E. J., Lill, M. a., Redinbo, M. R., and Krasowski, M. D. Challenges predicting ligand-receptor interactions of promiscuous proteins: the nuclear receptor PXR. *PLoS Comput. Biol.*, 5:e1000594, 2009.
- [45] Chen, S., He, N., Chen, W., Sun, F., Li, L., Deng, R., and Hu, Y. Molecular insights into the promiscuous interaction of human pregnane X receptor (hPXR) with diverse environmental chemicals and drug compounds. *Chemosphere*, 96:138–45, Feb. 2014.
- [46] Handa, K., Nakagome, I., Yamaotsu, N., Gouda, H., and Hirono, S. Three-dimensional quantitative structure-activity relationship analysis for human pregnane X receptor for the prediction of CYP3A4 induction in human hepatocytes: Structure-based comparative molecular field analysis. *J. Pharm. Sci.*, 104:223–232, 2015.
- [47] Ekins, S., Reschly, E. J., Hagey, L. R., and Krasowski, M. D. Evolution of pharmacologic specificity in the pregnane X receptor. *BMC Evol. Biol.*, 8:103, 2008.
- [48] Xue, Y., Chao, E., Zuercher, W. J., Willson, T. M., Collins, J. L., and Redinbo, M. R. Crystal structure of the PXR-T1317 complex provides a scaffold to examine the potential for receptor antagonism. *Bioorg. Med. Chem.*, 15:2156–2166, 2007.
- [49] Benod, C., Subra, G., Nahoum, V., Mallavialle, A., Guichou, J. F., Milhau, J., Roblés, S., Bourguet, W., Pascussi, J. M., Balaguer, P., and Chavanieu, A. N-1H-Benzimidazol-5-ylbenzenesulfonamide derivatives as potent hPXR agonists. *Bioorg. Med. Chem.*, 16:3537–49, 2008.
- [50] Zhu, Y., Olson, S. H., Hermanowski-Vosatka, A., Mundt, S., Shah, K., Springer, M., Thieringer, R., Wright, S., Xiao, J., Zokian, H., and Balkovec, J. M. 4-Methyl-5-phenyl triazoles as selective inhibitors of 11 β -hydroxysteroid dehydrogenase type I. *Bio. Med. Chem. Lett.*, 18:3405–3411, 2008.
- [51] Sun, W., Maletic, M., Mundt, S. S., Shah, K., Zokian, H., Lyons, K., Waddell, S. T., and Balkovec, J. Substituted phenyl triazoles as selective inhibitors of 11 β -Hydroxysteroid Dehydrogenase Type 1. *Bio. Med. Chem. Lett.*, 21:2141–5, 2011.
- [52] Tyagarajan, S., Chakravarty, P. K., Park, M., Zhou, B., Herrington, J. B., Ratliff, K., Bugianesi, R. M., Williams, B., Haedo, R. J., Swensen, A. M., Warren, V. a., Smith, M., Garcia, M., Kaczorowski, G. J., McManus, O. B., Lyons, K. a., Li, X., Madeira, M., Karanam, B., Green, M., Forrest, M. J., Abbadie, C., McGowan, E., Mistry, S., Jochnowitz,

- N., and Duffy, J. L. A potent and selective indole N-type calcium channel (Ca_v2.2) blocker for the treatment of pain. *Bio. Med. Chem. Lett.*, 21:869–873, 2011.
- [53] Tynebor, R. M., Chen, M. H., Natarajan, S. R., O'Neill, E. A., Thompson, J. E., Fitzgerald, C. E., O'Keefe, S. J., and Doherty, J. B. Synthesis and biological activity of 2H-quinolizin-2-one based p38 α MAP kinase inhibitors. *Bio. Med. Chem. Lett.*, 20:2765–2769, 2010.
- [54] Chobanian, H. R., Guo, Y., Liu, P., Chioda, M., Lanza, T. J., Chang, L., Kelly, T. M., Kan, Y., Palyha, O., Guan, X. M., Marsh, D. J., Metzger, J. M., Gorski, J. N., Raustad, K., Wang, S. P., Strack, A. M., Miller, R., Pang, J., Madeira, M., Lyons, K., Dragovic, J., Reitman, M. L., Nargund, R. P., and Lin, L. S. Discovery of MK-7725, a potent, selective bombesin receptor subtype-3 agonist for the treatment of obesity. *ACS Med. Chem. Lett.*, 3:252–256, 2012.
- [55] Finlay, H. J., Jiang, J., Caringal, Y., Kover, A., Conder, M. L., Xing, D., Levesque, P., Harper, T., Hsueh, M. M., Atwal, K., Blanar, M., Wexler, R., and Lloyd, J. Triazolo and imidazo dihydropyrazolopyrimidine potassium channel antagonists. *Bio. Med. Chem. Lett.*, 23:1743–1747, 2013.
- [56] Lill, M. A., Vedani, A., and Dobler, M. Raptor: Combining dual-shell representation, induced-fit simulation, and hydrophobicity scoring in receptor modeling: Application toward the simulation of structurally diverse ligand sets. *J. Med. Chem.*, 47:6174–6186, 2004.
- [57] Rahman, A.-u., editor. *Studies in Natural Products Chemistry*, volume 44. Elsevier, 2015.
- [58] <http://www.cdc.gov/parasites/sleepingsickness/biology.html>.
- [59] Hata, Y., Zimmermann, S., Quitschau, M., Kaiser, M., Hamburger, M., and Adams, M. Antiplasmodial and antitrypanosomal activity of pyrethrins and pyrethroids. *J. Agric. Food Chem.*, 59:9172–9176, 2011.
- [60] Moridi Farimani, M., Bahadori, M. B., Taheri, S., Ebrahimi, S. N., Zimmermann, S., Brun, R., Amin, G., and Hamburger, M. Triterpenoids with rare carbon skeletons from *salvia hydrangea*: Antiprotozoal activity and absolute configurations. *J. Nat. Prod.*, 74:2200–2205, 2011.

- [61] Julianti, T., Hata, Y., Zimmermann, S., Kaiser, M., Hamburger, M., and Adams, M. Antitrypanosomal sesquiterpene lactones from *Saussurea costus*. *Fitoterapia*, 82:955–959, 2011.
- [62] Ślusarczyk, S., Zimmermann, S., Kaiser, M., Matkowski, A., Hamburger, M., and Adams, M. Antiplasmodial and antitrypanosomal activity of tanshinone-type diterpenoids from *Salvia miltiorrhiza*. *Planta Med.*, 77:1594–1596, 2011.
- [63] Mokoka, T. A., Xolani, P. K., Zimmermann, S., Hata, Y., Adams, M., Kaiser, M., Moodley, N., Maharaj, V., Koorbanally, N. A., Hamburger, M., Brun, R., and Fouche, G. Antiprotozoal screening of 60 south African plants, and the identification of the antitrypanosomal germacranolides schkuhrin I and II. *Planta Med.*, 79:1380–1384, 2013.
- [64] Hata, Y., Ebrahimi, S. N., Mieri, M. D., Zimmermann, S., Mokoka, T., Naidoo, D., Fouche, G., Maharaj, V., Kaiser, M., Brun, R., Potterat, O., and Hamburger, M. Antitrypanosomal isoflavan quinones from *Abrus precatorius*. *Fitoterapia*, 93:81–87, 2014.
- [65] Zimmermann, S., Fouché, G., De Mieri, M., Yoshimoto, Y., Usuki, T., Nthambeleni, R., Parkinson, C. J., van der Westhuyzen, C., Kaiser, M., Hamburger, M., and Adams, M. Structure–activity relationship study of sesquiterpene lactones and their semi-synthetic amino derivatives as potential antitrypanosomal products. *Molecules*, 19:3523–3538, 2014.
- [66] Ayyari, M., Salehi, P., Ebrahimi, S. N., Zimmermann, S., Portmann, L., Krauth-Siegel, R. L., Kaiser, M., Brun, R., Rezadoost, H., Rezazadeh, S., and Hamburger, M. Antitrypanosomal isothiocyanate and thiocarbamate glycosides from *moringa peregrina*. *Planta Med.*, 80:86–89, 2014.
- [67] Usuki, T., Sato, M., Hara, S., Yoshimoto, Y., Kondo, R., Zimmermann, S., Kaiser, M., Brun, R., Hamburger, M., and Adams, M. Antitrypanosomal structure–activity–relationship study of synthetic cynaropicrin derivatives. *Bioorg. Med. Chem. Lett.*, 24:794–798, 2014.
- [68] Case, D., Babin, V., Berryman, J. T., Betz, R., Cai, Q., Cerutti, D. S., Cheatham, III, T. E., Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Luo, R.,

Madej, B., Merz, K. M., Paesani, F., Roe, D. R., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., X., W., and A., K. P. *AMBER 12*. University of California, San Francisco., 2012.

- [69] Xue, Y., Moore, L. B., Orans, J., Peng, L., Bencharit, S., Kliewer, S. a., and Redinbo, M. R. Crystal structure of the pregnane X receptor-estradiol complex provides insights into endobiotic recognition. *Mol. Endocrinol.*, 21:1028–1038, 2007.