

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Ahinko, Mira; Niinivehmas, Sanna; Jokinen, Elmeri; Pentikäinen, Olli

**Title:** Suitability of MMGBSA for the selection of correct ligand binding modes from docking results

**Year:** 2019

**Version:** Accepted version (Final draft)

**Copyright:** © 2018 John Wiley & Sons A/S

**Rights:** In Copyright

**Rights url:** <http://rightsstatements.org/page/InC/1.0/?language=en>

**Please cite the original version:**

Ahinko, M., Niinivehmas, S., Jokinen, E., & Pentikäinen, O. (2019). Suitability of MMGBSA for the selection of correct ligand binding modes from docking results. *Chemical Biology and Drug Design*, 93(4), 522-538. <https://doi.org/10.1111/cbdd.13446>

Article type : Research Article

# Suitability of MMGBSA for the Selection of Correct Ligand Binding Modes from Docking Results

Mira Ahinko<sup>1‡</sup>, Sanna Niinivehmas<sup>1,2‡</sup>, Elmeri Jokinen<sup>2</sup>, Olli T. Pentikäinen<sup>1,2\*</sup>

<sup>1</sup> University of Jyväskylä, MedChem.fi, Department of Biological and Environmental Science & Nanoscience Center, University of Jyväskylä, P.O. Box 35, FI-40014 University of Jyväskylä, Finland

<sup>2</sup> University of Turku, MedChem.fi, Institute of Biomedicine, Integrative Physiology and Pharmacology, Kiinamyllynkatu 10, Medisiina C6, FI-20520 Turku, Finland

\* Corresponding author: e-mail: olli.pentikainen@utu.fi; ‡ Equal contribution

The estimation of the correct binding mode and affinity of a ligand into a target protein using computational methods is challenging. However, docking can introduce poses from which the correct binding mode could be identified using other methods. Here, we analyzed the reliability of binding energy estimation using the molecular mechanics-generalized Born surface area (MMGBSA) method without and with energy minimization to identify the likely ligand binding modes within docking results. MMGBSA workflow (1) outperformed docking in recognizing the correct binding modes of androgen receptor ligands and (2) improved the correlation coefficient of computational and experimental results of rescored docking poses to phosphodiesterase 4B. Combined with stability and atomic distance analysis, MMGBSA helped to (3) identify the binding modes and sites of metabolism of cytochrome P450 2A6 substrates. The standard deviation of estimated binding energy within one simulation was lowered by minimization in all three example cases. Minimization improved the identification of the correct binding modes of androgen receptor ligands. Although only three case studies are shown, the results are analogous and indicate that these behaviors could be generalized. Such identified binding modes could be further used, for example, with free energy perturbation methods to understand binding energetics more accurately.

## Introduction

Being able to accurately calculate differences in the binding free energies of protein–ligand and protein–protein interactions is essential in applications related to drug discovery [1,2]. A wide variety of methods have been developed to rank ligands based on their binding affinities. Molecular docking algorithms are the least computationally demanding but lack accuracy in the approximation of ligand binding affinity [3–9]. Methods based on molecular dynamics (MD) or Monte Carlo simulations consume more computational resources, with the aim of producing rigorous estimations of binding

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi:

10.1111/cbdd.13446

This article is protected by copyright. All rights reserved.

free energies. In principle, the most accurate predictions are generated by the alchemical free energy methods, such as free energy perturbation (FEP), which require intensive computational effort due to the sampling of unphysical intermediate states between the initial and final state of the system [10]. A less time-consuming MD-based option is offered by the molecular mechanics generalized Born or Poisson–Boltzmann surface area (MMGBSA or MMPBSA) methods that analyze only the free and bound states of the system [11].

The MMGBSA/MMPBSA methods have been widely used in the identification of active ligands and their affinities and binding modes [12]. With MMGBSA, protein–ligand and protein–protein interaction energies can be estimated relatively quickly when compared with alchemical free energy methods. Several studies have shown that MMGBSA generally outperforms the scoring functions of docking algorithms in terms of accuracy in ranking ligands by their binding affinity [13,14]. Reported correlations between experimental and MMGBSA-derived binding affinities range from excellent to poor, indicating that the performance of MMGBSA/MMPBSA methods is highly case-specific [15–18]. Because of this variance, the careful selection of essential parameters—such as the length of the MD simulation, dielectric constants, and whether to include or exclude the entropy term—has been advised, as such parameters can greatly influence the precision of the method [19]. Significant correlations between experimental and computational binding free energies have been obtained both by averaging the binding free energy over multiple protein–ligand complexes derived from MD simulation snapshots and by running an MMGBSA analysis on a single minimized snapshot [13,14,20,21]. Still, FEP has been shown to be generally more capable of identifying active ligands and their correct binding modes than MMGBSA, making it the preferable method for the final binding free energy estimation [22]. In general, MMGBSA method enables rapid estimation of binding affinity but only with the cost of crude approximations. Especially the lack of conformational entropy and free energy of solvent binding cause significant standard error of estimation, which again limits the applicability of the method in drug discovery [12].

The difficulty of adequately sampling the relevant ligand binding modes has traditionally been one of the major limitations of FEP and other alchemical free energy methods [23]. Especially in cases of multiple possible binding modes, this inability may cause a substantial decrease in the precision of the free energy calculation [24,25]. Generally, considering different ligand binding modes requires the utilization of multiple calculations, which significantly increases the computational cost of the protocol. This has constrained the use of alchemical free energy methods in many applications, such as virtual screening, where computational efficiency is one of the main priorities [26]. The problem has been addressed by utilizing methods, such as replica exchange with solute tempering (REST) and accelerated MD, that aim to enable consideration of multiple ligand binding modes during a single calculation by lowering the energetic barriers between them [27–33]. Another possible way to tackle the issue is to decrease the number of necessary calculations by identifying the biologically relevant binding modes with computationally less demanding methods prior to using the free energy methods. The reported good performance of MMGBSA indicates the usefulness of this method in filtering irrelevant ligand binding modes to facilitate the completion of the computationally more intensive alchemical free energy analysis.

Here, we have used MD simulations together with MMGBSA analysis to rescore docking results. Specifically, we have studied the stability of both formed complexes and binding energies using MD simulation snapshots without and with energy minimization. Three protein targets that we have studied also previously [15,34–37]—cytochrome P450 type 2A6 (CYP2A6), androgen receptor (AR), and phosphodiesterase 4B (PDE4B)—were utilized to evaluate the performance of MMGBSA in identifying ligand binding modes. CYP2A6 has a relatively small binding site for ligands, and thus, it is an important player in the oxidation of the smallest xenobiotics [38]. It is also an excellent model enzyme because its binding cavity is limited in size and has easily definable sites for polar interactions.

AR belongs to nuclear receptors and is a traditional target for the discovery of prostate cancer treatments [39]. PDE4B belongs to a large family of PDE-enzymes that degrade cAMP and cGMP by catalyzing the breakage of their phosphodiester bond. PDE4 has many implications in drug discovery, and it has been suggested that PDE4B inhibitors could act as therapeutic agents e.g. against depression and schizophrenia [40] and chronic obstructive pulmonary disease [41].

## Materials and methods

*Protein structures:* Protein structures were acquired from the Protein Data Bank (PDB; www.rcsb.org) [42,43]. The coumarin-bound structure was used for CYP2A6 (PDB-code: 1Z10 [44]). Twelve crystal structures for AR were chosen: 1GS4 [45], 1Z95 [46], 2AM9 [47], 2AMB [47], 2HVC [48], 2IHQ [49], 2OZ7 [50], 2PNU [51], 3B66 [52], 3RLL [53], 4OHA [54], 5V8Q [55]. In structure selection, the structural variability of the crystallized ligands and the resolution of the crystal structure were emphasized. For PDE4B we used crystal structure of PDE4B with bound inhibitor (PDB-code: 4W5E [56]) for two reasons: 1) it contains also the additional part to many other structures, which forms a lid into the inhibitor-binding site in PDE4B, and 2) due it is easier to compare the results to our previous results [37]. Hydrogens were added to protein structures using reduce 3.23 [57].

*Ligand preparation:* For CYP2A6, two known substrates, coumarin and 6-methylcoumarin, were selected. Both compounds are 7-hydroxylated in CYP2A6 oxidation [36,58]. For AR, ligands were separated from the chosen protein crystal structures. In the case of PDE4B, we used 152 previously published PDE4B inhibitors that have been measured with comparable methods [59–64]. Three-dimensional molecules for CYP2A6 and PDE4B were obtained by using LigPrep in Maestro 2017-4 (Schrödinger Inc, Portland, OR, USA), with following options: (1) OPLS2005 force field, (2) protonization at physiological pH 7.4±0.0, where ligands were ionized with Epik, (3) tautomers were generated, (4) specific chiralities were retained, (5) maximum number of tautomers and chiralities was set to 32/ligand. For tautomers and chiralities only one low energy conformation was kept. For AR, bonds and stereochemistry of ligands were checked to comply with that given in PDB. Hydrogens and partial charges were added using OPLS2005 force field in Maestro (Schrödinger Inc, Portland, OR, USA).

*Molecular docking:* Molecular docking for CYP2A6, AR, and PDE4B was performed with Plants v.1.2 [65] with chemplp scoring function to produce four binding modes for each ligand that differ significantly (RMSD = 2.5 Å (AR/PDE4B) or 2.0 Å (CYP2A6)). The docking area was set to the center of the binding cavity defined by the bound ligand in each protein structure. In the case of AR, the ligands separated from the protein crystal structures were cross-docked back to the different crystal structures. In all used three examples the ligand binding site does not contain large cavities to surrounding solvent, and accordingly, the docking area was relatively easy to define, and all of the docked ligand poses were located at the defined binding sites.

*Parametrization of ligands:* Parametrization of ligands for MD simulations were performed with two different methods: (1) the 3D structures of CYP2A6 substrates were optimized quantum mechanically with GAUSSIAN09 (Gaussian, Inc., Wallingford, CT) at the HF/6-31+G\* level using a polarizable continuum model. The RESP method [66] was used to calculate the atom-centered point charges from the electrostatic potentials; (2) AR ligands and PDE4B inhibitors were optimized and parametrized with Antechamber [67] method utilizing charge method AM1-bcc.

*Preparations for MD:* Tleap in Antechamber 17 [68] was used to: (1) generate force field parameters for protein and combine protein-ligand parametrizations; (2) add hydrogen atoms; and (3) solvate the system with a rectangular box of transferable intermolecular potential three-point (TIP3P) water molecules [69] extending 13 Å in every dimensions around the solute. The ff14SB force field [70] was

used for the protein structures. For CYP2A6, parameters suitable for heme with a six-coordinate iron were used [71] (<https://personalpages.manchester.ac.uk/staff/Richard.Bryce/amber/index.html>).

*MD simulations and MMGBSA:* MD simulations were performed with NAMD 2.9 [72] in four steps: (1) 15,000 steps of energy minimization where Ca-atoms were restrained with 5kcal/mol; (2) 15,000 steps of energy minimization without restraints; (3) 180,000 steps of MD where Ca-atoms were restrained with 5kcal/mol; (4) 1,200,000 steps (2.4 ns) of MD without restraints. The energy minimization of snapshot structures from the MD simulation was performed for 15,000 steps without constraints. Otherwise all parameters were kept similar as in our previous study [17]. All energy minimizations were performed using the default conjugate gradient energy minimization method. Superpositioning of protein-ligand complexes was performed with Vertaa in Bodil [73]. All MD trajectory analysis were made with Cpptraj [74]. RMSD calculations in relation to crystal structures were conducted with rmsd.py script in Schrödinger release 2017-4 (Schrödinger Inc, Portland, OR, USA). MMGBSA analyses were performed with MMPBSA.py script [75] implemented in Antechamber. For generalized Born calculations, three different igb-models, namely igb1, igb2, and igb5, were employed. Shortly, the main difference between these three igb-models is that the original description of solute boundaries in term of effective radii (igb1) have been rescaled in igb2 and igb5 to mimic primarily the computationally more demanding Poisson-Boltzmann model.

## Results and discussion

The general aim of this research was to study the ability of MMGBSA to identify the biologically relevant binding modes of small molecules in various target proteins. First, a key challenge in drug discovery is understanding the metabolism, and especially the consequent metabolites for the developed molecules. Accordingly, we studied the possible binding modes of coumarin and 6-methylcoumarin into CYP2A6 to determine whether their site of metabolism (SOM) can be predicted using MMGBSA. Second, the identification of biologically relevant binding modes is an essential quality in computational methods. As such, we used AR to study the binding modes that are verifiable from the protein crystal structures. Last, we tested the effect of ligand binding pose selection using MMGBSA for the correlation coefficient between experimental and computational data with PDE4B. Short (2.4 ns) MD simulations with relatively fast MMGBSA-analysis were employed, although, longer simulations could expose to larger conformational changes and the usage of implicit solvation model has significant drawback in case the solvent molecules form explicit hydrogen-bond network in order to stabilize the ligand binding. From these two possible problems, the short simulation time was compensated with the usage of multiple starting poses from docking. Another issue is the rather limited number of study cases. However, it has been previously shown that a small number of study cases can demonstrate the usability of the method, although more cases are needed to draw broad conclusions. For example, it has been shown that ensemble docking for only two [76] or three study cases, one of which was AR,[77] was enough to indicate the applicability of the method. In addition to that, with molecular docking and MD, promising results have been obtained in studying the binding of AR ligands in detail to predict agonism and antagonism [78], and MD combined with MMGBSA showed success in studying interaction mechanism ARs with only one specific compound [79].

### *CYP2A6: Identification of the SOM*

The recognition of the correct SOM(s) is crucial in drug discovery because it can act as the first step in identifying the formed metabolites of new molecules and thus understanding their possible toxicity effects in living organisms. Because MMGBSA has been used somewhat successfully in the prediction of binding affinities, our aim was to study whether the most likely binding mode for CYP2A6 substrates can be recognized solely by using MMGBSA. All calculations were directly performed using MD simulation snapshots and energy minimized snapshots.



Coumarin is the most used reference compound with CYP2A6 because the catalysis of coumarin to fluorescent 7-hydroxylcoumarin is specific to the 2A6 form in the array of human CYP enzymes. The crystal structure of CYP2A6 in complex with coumarin [44] explains, at the structural level, the likely mechanism of catalysis. The carbonyl oxygen atom of coumarin accepts a hydrogen bond from Asn297 (Figure 1A: orange dotted line), and position 7 (C7) locates towards the heme iron (Figure 1A). Furthermore, there is a small electron density in the vicinity of the iron and C7 of coumarin, where oxygen (O, O2, OH, etc.) could be placed for the catalysis to occur (Figure 1A).

*MD simulations with coumarin:* The MD simulations were started from four coumarin docking poses, of which the pose 2 is closest to the crystal structure binding mode (Figure 1B). Twenty snapshots with their minimized correspondents from each MD trajectory were used in the binding energy estimation with MMGBSA. A comparison of the numerical MMGBSA values from the simulation snapshots without and with minimization (Table 1) reveals that all four binding poses, with all three igb-models, give a clearly lower energy value with the minimized snapshots (igb1:  $-3.7 \pm 0.4$ ; igb2:  $-4.7 \pm 0.4$ ; igb5:  $-5.8 \pm 0.3$ ) and that the standard deviations of the values are also clearly lower (igb1:  $-0.7 \pm 0.2$ ; igb2:  $-0.7 \pm 0.1$ ; igb5:  $0.7 \pm 0.3$ ). The difference is highest with igb5, which seems to be very sensitive to hydrogen bond distances and angles. In the binding poses without minimization, the large variations in the igb5 energies are highly in accordance with the variation of the distance and angle of the interaction of Asn297 with the carbonyl oxygen atom of coumarin.

Table 1. Docking scores and binding energy estimations (MMGBSA; kcal/mol) with different generalized Born models (igb1, 2, and 5) for four coumarin and 6-methylcoumarin binding poses (#1-4) within CYP2A6.

Molecule	Pose	Docking	Without minimization			RMSD1	RMSD20	Minimized			RMSD1	RMSD20
			igb1	igb2	igb5			igb1	igb2	igb5		
coumarin	1	-71.0	-20.8±1.6	-20.5±1.7	-22.9±1.9	2.31	2.67	-24.9±0.9	-25.4±1.0	-28.7±1.4	2.23	2.52
	2	-70.4	-21.7±1.6	-21.4±1.8	-23.9±2.1	<b>0.62</b>	0.98	<b>-25.6±0.8</b>	-26.3±0.9	-30.0±1.1	<b>0.59</b>	1.11
	3	-70.4	-21.3±3.6	-21.0±4.0	-23.3±4.1	4.30	<b>0.71</b>	-24.9±3.1	-25.6±3.5	-29.4±3.3	4.36	<b>1.00</b>
	4	-67.4	<b>-22.4±1.9</b>	<b>-22.3±1.9</b>	<b>-25.2±2.1</b>	2.63	2.71	-25.5±1.1	<b>-26.4±1.3</b>	<b>-30.5±1.6</b>	2.46	2.63
6-methyl-coumarin	1	-77.8	<b>-25.8±1.7</b>	<b>-25.9±1.9</b>	<b>-28.4±2.2</b>	-	-	<b>-30.0±1.0</b>	<b>-31.2±1.2</b>	<b>-34.9±1.5</b>	-	-
	2	-73.9	-19.4±1.6	-18.4±1.6	-20.9±1.9	-	-	-22.7±1.3	-22.5±1.4	-26.2±2.0	-	-
	3	-70.1	-24.4±1.8	-23.9±1.8	-26.1±1.8	-	-	-28.9±1.8	-29.5±1.9	-32.9±2.0	-	-
	4	-68.6	-24.4±1.3	-23.8±1.5	-26.1±1.8	-	-	-29.2±1.8	-29.8±2.0	-33.6±2.2	-	-

Underline: Original docking pose that resembles crystal structure binding mode (coumarin) or is in agreement with other experimental results (6-methylcoumarin); bold: The best binding energy and the lowest RMSD value of the molecule's binding poses within the MMGBSA method; ±: Standard deviation in MD snapshots. RMSD1=RMSD value to the crystal structure from the beginning of the MD simulation (the first frame). RMSD20=RMSD value to the crystal structure from the end of the MD simulation (the last frame).

The MMGBSA binding energies rank the coumarin pose 4 highest both without with minimization in igb2 and igb5, while docking ranked the initial pose the lowest (Table 1). The MMGBSA-based order of the poses stayed the same between snapshots without/with minimization in igb2 and igb5, and they changed only slightly in the igb1 energies. Only the igb1 energy from minimized snapshots rank the crystal structure-like pose 2 the highest. The comparison of energy values between the binding poses shows that in the snapshots without minimization, the difference between the most favorable and the worst pose is much higher than in the minimized snapshots (igb1: 0.8; igb2: 0.8; igb5: 0.5). In addition, without minimization, the differences between the best and the second-best poses are approximately three to ten times those of the corresponding minimized snapshots (without/with minimization: igb1: 0.74/0.08; igb2: 0.90/0.09; igb5: 1.27/0.51). However, as the standard deviations are also clearly higher for the coumarin binding poses without minimization, the MMGBSA values for all poses are within the error limits of each other. The gain in binding energy from minimization is similar in all coumarin poses with each igb model (e.g. igb1 pose 1: -4.1; pose 2: -3.9; pose 3; -3.6; pose 4: -3.1). This results from, firstly, the above-mentioned observation that the improvement of hydrogen bonding interactions is clearly a significant factor for lowering the MMGBSA value in minimization, and secondly, that all binding poses of coumarin eventually formed a hydrogen bond to CYP2A6 during the MD simulations (see below) that could be improved in the minimization step. As the MMGBSA values of the coumarin binding poses are within error limits of each other, the binding energy estimations alone cannot be used to identify the most likely binding mode for coumarin.

Examining the coumarin orientation in simulations from the four different starting poses, it is notable that from each, the substrate rotates in such way that the carbonyl oxygen accepts a hydrogen bond from Asn297 (Table 2), which is the only polar residue in the vicinity of the coumarin binding site. Furthermore, from two starting poses (poses 2 and 3), C7 is or becomes the closest to the heme iron (Table 2; average distance without/with minimization 3.6/3.4 Å) and the whole ligand pose is or becomes similar to the crystal structure based on both visual and RMSD analysis (Table 1). Simultaneous to the C7 flip towards the heme in pose 3, the energy of the complex drops notably as the coumarin forms the hydrogen bond with Asn297 (without/with minimization: igb1: -7.6/-6.7; igb2: -8.5/-7.4; igb5: -8.6/-6.9). From the other two starting poses (poses 1 and 4), position 5 (C5) is or becomes closest to the iron, but the distance of the C5 is farther than that of the C7 in poses 2 and 3 (Table 2; average distance without/with minimization 3.9/3.8 Å). When considering the typical distances of the groups closest to the iron in the crystal structures of CYP2A6 (e.g., coumarin: 3.2 Å; methoxsalen (PDB-code:1Z11 [44]): 3.3 Å), the C7 of the minimized binding modes, 2 and 3 are at a similar optimal distance to typical values, while the C5 of poses 1 and 4 are slightly farther compared to typical values. The variations in the distances are also clearly lower when C7 is the closest to iron (without minimization: 0.38 vs. 0.59; minimized: 0.23 vs. 0.45). In both the simulations without and with minimization, conformation 2 is the most stable based on the RMSD from both the ligand and the heme, but it is within the error limits of binding modes 1 and 4 (Table 2).

**Table 2.** The stability, closest carbon atom to heme iron, and Asn297 side chain nitrogen distance (Å) to the carbonyl group of coumarin and 6-methylcoumarin binding poses (#1-4) with CYP2A6 in simulations without and with minimization.

Molecule	Pose	Without minimization					Minimized				
		RMSD <sub>L</sub>	RMSD <sub>LH</sub>	CX	d	d (N-O)	RMSD <sub>L</sub>	RMSD <sub>LH</sub>	CX	d	d (N-O)
coumarin	1	0.16±0.03	1.19±0.20	C5	4.0±0.8	3.0±0.4	0.02±0.01	1.03±0.17	C5	3.8±0.7	2.9±0.0
	2	0.20±0.04	<b>1.04±0.18</b>	C7	3.7±0.4	2.9±0.2	0.04±0.02	<b>0.96±0.17</b>	C7	3.4±0.3	2.8±0.0
	3	0.14±0.04	2.34±0.40	C7	4.0±1.2	4.5±2.8	0.04±0.01	2.29±0.34	C7	4.0±1.3	4.5±1.0
	4	0.14±0.03	1.17±0.15	C5	3.8±0.4	3.0±0.3	0.04±0.02	1.07±0.16	C5	3.7±0.3	2.9±0.0
6-methyl-coumarin	1	0.60±0.17	1.20±0.07	C7 6M	3.9±0.3 3.9±0.4	2.9±0.1	0.60±0.20	1.1±0.05	C7 6M	3.8±0.2 3.9±0.2	2.9±0.0
	2	0.53±0.21	1.70±0.19	C4	3.4±0.7	10.8±0.9	0.51±0.24	1.77±0.16	C4	3.5±0.7	10.3±0.8
	3	0.53±0.26	2.04±0.10	C4	6.2±0.7	3.2±1.6	0.56±0.24	1.83±0.07	C4	5.7±0.5	3.1±1.1
	4	0.59±0.19	<b>1.05±0.13</b>	C5	4.4±0.4	3.0±0.2	0.43±0.27	<b>0.94±0.18</b>	C5	4.2±0.3	2.9±0.0

Underline: Original docking pose that resembles crystal structure binding mode (coumarin) or is in agreement with other experimental results (6-methylcoumarin); RMSD: Root mean square deviation of ligand (L) or ligand and heme (LH) in relation to the first snapshot; bold: the lowest RMSD of ligand and heme; ±: Standard deviation in MD snapshots; CX: The closest substrate carbon to the heme iron on average; d: The distance of the closest carbon atom to the heme iron on average; d (N-O): The distance of Asn297 side chain nitrogen to coumarin carbonyl; 6M: 6-methyl.

While the binding energy or the stability alone do not strictly differentiate the preference of coumarin hydroxylation for C7 over C5, the distance and variation of the closest substrate carbon atom to the heme iron does suggest that C7 is more susceptible for oxidation by CYP2A6, which is in line with experimental data [58]. Altogether, the combination of MMGBSA binding energy estimation, stability calculation, and considering the distance (together with its variations) of the coumarin's atomic positions to the heme iron form an *in silico* protocol that gives detailed insight into the binding mode of coumarin and provides an *in silico* workflow for the prediction of the SOM, in this case the hydroxylation of position 7, of coumarin.

*MD simulations with 6-methylcoumarin:* The docking of 6-methylcoumarin into CYP2A6 yielded four starting poses for MD simulations (Figure 1C), where the closest ligand positions to the heme iron are as follows: (1) position 6-methyl and position 7; (2) positions 4 and 6; (3) positions 3 and 4; and (4) position 6-methyl and 5. In the MD simulations, poses 1 and 4 were clearly the most stable ones, based on the RMSD of the ligand and the heme group (Table 2), and in pose 1, the C7 and 6-methyl groups are in close proximity to the heme iron (Table 2). There was still enough space to accommodate the catalytic oxygen needed for oxidation in pose 1, as seen with coumarin in the crystal structure (Figure 1A). In addition, the distance of C7 from the iron remained low enough for catalysis, although not as low as it was for C7 in the coumarin simulations (Table 2). The estimations of binding energy, done using MMGBSA, indicated that the pose 1 was energetically preferred (Table 1). As opposed to the case of coumarin where MMGBSA recognized the native binding mode better than docking,



MMGBSA and docking ranked the same 6-methylcoumarin pose 1 highest (Table 1). However, the MMGBSA binding energy estimations for poses 3 and 4 were always within the error limits of pose 1, while pose 2 was clearly worse, as it was the only pose that did not rotate in the catalytic site to accept a hydrogen bond from Asn297 during the simulation. Pose 2 also gained the least in binding energy in minimization due to the lack of a hydrogen bond with the enzyme (e.g. igb1 pose 1: -4.2; pose 2: -3.3; pose 3: -4.5; pose 4: -4.8). In the case of 6-methylcoumarin, there is no difference between the MD snapshots without and with minimization for the selection of the preferred binding mode. The ligand-heme RMSD- and the MMGBSA-based order of the poses stayed generally the same from both without and with minimization. Altogether, the stability, the distance of the closest carbon to the heme iron, and the binding energy estimations suggested that the primary SOM in 6-methylcoumarin would be C7 in pose 1 (Tables 1 and 2). This result, where C7 is the preferred SOM, correlates with the experimental results, which show that the primary metabolite is hydroxylated at C7 [36].

*Identifying the correct binding mode and SOM:* Although the simulation times were short, from a catalytic point of view it is crucial to understand how well the substrate is stabilized into favorable reaction coordinates. As a simple measurement, the RMSD of docked substrate and substrate plus the heme was calculated from the simulation trajectories (Table 2). In the case of two rigid substrates, the substrate-only RMSD values originate mainly from slight variations in bond distances and methyl group rotation in 6-methylcoumarin, and accordingly, it is highly beneficial to calculate the RMSD together with the heme. Adding the heme to the RMSD calculation brings forward the bigger shifts in the substrate's orientation in relation to the catalytic site. With coumarin, binding poses 1, 2 and 4 were again within error of the measurement of this RMSD value, although pose 2, in which the C7 is towards the heme, has a slightly lower RMSD. With 6-methylcoumarin, the most stabilized substrate and heme positions are for poses 1 and 4. The minimization step did not bring added value to either MMGBSA or RMSD analysis, but it stabilized the substrate positions that were towards the heme iron a little closer to the distances seen in crystal structures. Using both the ligand and heme-based RMSD and the distance of the substrate atoms to the heme iron to analyze the stabilization of the substrate to the favorable reaction coordinates is suggested to be an essential addition to MMGBSA binding energy estimations for finding the active binding modes (i.e., those binding modes in which the substrate is truly oxidized) and SOM(s) for CYP substrates.

When considering both the predicted binding energy and the stabilization of the reaction coordinates (Tables 1 and 2), it would be logical to suggest that for coumarin, the most favored binding mode is either pose 2 or 4, and for 6-methylcoumarin, it is pose 1. The results suggest that it is possible that coumarin has two favorable SOMs instead of just one, while 6-methylcoumarin has one clearly favorable site for CYP2A6 catalyzed oxidation. Altogether, this is in line with typical phenomena among xenobiotic-metabolizing CYP enzymes. A single CYP form can quite often oxidize the same substrate from multiple sites, as can be found in databases of SOMs for substrates of different CYP enzymes [80,81]. The malleability of the catalytic sites allows the substrate to orient into several binding modes in the site, resulting in the ability of the enzyme to oxidize various positions of the small molecule. Even in the relatively small and rigid catalytic site of CYP2A6, the variations in the shape of the catalytic site can play a major role in determining the success of structure-based regioselectivity prediction of the enzyme's substrates [82].

The flexibility of xenobiotic-metabolizing CYP enzymes and the multitude of possible binding modes for substrates create a challenge for finding the active binding mode(s) of CYP substrates. Furthermore, finding the active binding mode from multiple possible equally stable and energetically favorable ones is complicated by the fact that the substrate may also have an inhibiting binding mode. For instance, pilocarpine that is a competitive inhibitor of CYP2A6 [83,84] has been crystallized in the inhibiting binding mode, where its lone-pair nitrogen is coordinated with the heme iron [85]. However, it is also oxidized by CYP2A6 at its position 3 [86], which, in the crystal structure, is further

away from the usual reaction coordinates that lie near the heme iron. This is where, for instance, the SOM C7 of coumarin is placed in the crystal structure complex with CYP2A6 (Figure 1A) [44]. The inhibiting effect can occur when a non-reactive group faces the heme (i.e., the substrate is stabilized into an orientation where oxidation cannot occur efficiently), or when the CYP mediated catalysis results in a reactive metabolite that can irreversibly inhibit the enzyme. Finding both the active and inhibitive compounds and their binding modes in different CYP forms using *in silico* methods can be a helpful step in the risk assessment of new molecules [87]. Altogether, projecting the simulations here to a larger context, the workflow and observations can provide a plausible starting point into following studies for gaining a more detailed view of the defining interactions between small molecules and CYP enzymes.

*AR: Identification of binding modes*

*Docking:* For virtual screening purposes, it is important that the docking protocol can utilize any protein structure to produce biologically relevant and reliable binding poses for variable compounds [6,7]. In general, molecular docking identifies “correct” (i.e., crystal structure-like) binding mode for AR ligands very well. In cross-docking, with ten ligands out of twelve, the best ranked docking pose resembles the crystal structure ligand binding mode and is thus considered correct. Also, with the two “failed” docking results (i.e., for which crystal structure-like pose is not the best ranked pose), the correct binding mode is among the four saved poses (Table 3). One of the failed docking structures, the 3RLL ligand, was cross-docked into the 3B66 protein structure. In the 3B66 protein, Met745 turns towards the binding pocket, and thus, the binding pocket is not favorable for 3RLL ligand binding (Figure 2A). In addition, the original 3RLL protein has a Thr877Ala mutation in the binding site; however, the absence of this mutation should not affect the docking result much because it does not restrict the binding area. In self-docking, the 3RLL ligand repeats the crystal structure binding pose. The 5V8Q ligand was docked into the 2AMB protein. In the 2AMB protein, the position of Thr877 is altered in the binding area, but this does not affect the usable binding area. However, the docking scores for different poses of the 5V8Q ligand are relatively similar (-81.5 – -77.6; -78.9 for the crystal structure-like pose; Table 3). In self-docking, the crystal structure-like pose of the 5V8Q ligand is identified as the best pose. It is notable that, in self-docking, the scale of the docking scores is very different than that of the scores in cross-docking (-118.5 vs. -102.4 for 3RLL and -90.6 vs. -78.9 for 5V8Q). All in all, molecular docking seems to be a reasonable way to relatively quickly identify biologically relevant binding modes, at least in cases where there is some prior knowledge about the binding site. However, here, only ligands that are known to bind to AR were cross-docked into varying protein conformations.

[Table 3. Docking scores, binding energy estimations (MMGBSA; kcal/mol) with different generalized Born models (igb1, 2, and 5) and RMSD values for twelve cross-docked AR ligands.

Molecule	Pose	Docking	Without minimization				Minimized					
			igb1	igb2	igb5	RMSD1	RMSD20	igb1	igb2	igb5	RMSD1	RMSD20
1GS4	1	-87.4	<b>-47.7± 2.1</b>	<b>-46.1± 2.1</b>	<b>-52.1± 2.3</b>	<b>0.63</b>	<b>0.63</b>	-53.8± 1.5	-54.4± 1.9	<b>-62.6± 2.1</b>	<b>0.83</b>	<b>0.83</b>
	2	-78.9	-45.3± 2.5	-43.2± 2.8	-46.7± 3.2	7.39	7.42	-51.7± 1.8	-51.7± 2.3	-57.2± 2.7	7.40	7.39
	3	-76.5	-44.2± 2.2	-43.0± 2.8	-49.2± 3.8	3.03	3.28	-51.1± 2.8	-51.8± 3.7	-60.5± 5.0	3.02	3.34
	4	-63.3	-46.8± 2.6	-45.6± 3.0	-50.6± 3.8	7.12	7.07	<b>-54.6± 2.1</b>	<b>-55.0± 2.3</b>	-61.5± 3.0	7.07	7.11
1Z95	1	-79.7	<b>-53.8± 2.4</b>	<b>-50.7± 2.6</b>	<b>-55.1± 3.1</b>	<b>1.30</b>	<b>1.32</b>	<b>-60.3± 0.9</b>	<b>-58.7± 1.0</b>	<b>-64.5± 1.5</b>	<b>1.25</b>	<b>1.27</b>
	2	-78.0	-47.1± 2.1	-42.4± 2.3	-45.5± 2.5	2.56	3.00	-54.4± 1.5	-51.2± 1.9	-55.7± 2.4	2.79	3.05
	3	-72.7	-46.2± 2.5	-40.2± 2.6	-40.3± 3.0	7.34	7.96	-53.6± 2.7	-48.4± 2.6	-49.4± 3.2	7.42	7.86
	4	-72.4	-43.8± 2.8	-38.5± 2.9	-40.6± 3.0	7.93	7.82	-50.7± 1.2	-46.5± 1.0	-49.9± 0.9	8.10	7.95
2AM9	1	-91.6	<b>-45.4± 2.0</b>	<b>-48.5± 2.0</b>	<b>-58.0± 1.7</b>	<b>0.49</b>	<b>0.49</b>	<b>-51.9± 0.9</b>	<b>-57.1± 1.1</b>	<b>-68.3± 1.4</b>	<b>0.30</b>	<b>0.23</b>
	2	-83.1	-43.5± 2.9	-46.0± 2.9	-54.3± 2.9	3.02	3.19	-49.4± 2.0	-54.3± 1.8	-64.7± 1.6	2.99	3.10
	3	-73.8	-43.0± 2.7	-46.4± 3.1	-54.9± 3.4	6.70	6.80	-49.6± 2.4	-55.1± 3.0	-65.4± 4.0	6.71	6.71
	4	-68.3	-41.7± 2.9	-44.8± 3.1	-53.3± 3.1	6.64	6.71	-48.2± 2.6	-53.5± 2.6	-64.0± 2.6	6.63	6.74

				3.2	3.4				3.2	3.8		
<b>2AMB</b>	1	-107.6	<b>-52.6± 2.0</b>	<b>-54.7± 2.4</b>	<b>-63.9± 2.9</b>	<b>0.43</b>	<b>0.41</b>	<b>-60.1± 0.8</b>	<b>-64.7± 1.0</b>	<b>-76.5± 1.2</b>	<b>0.39</b>	<b>0.41</b>
	2	-77.2	-44.9± 2.6	-46.4± 3.1	-53.9± 3.4	6.82	6.87	-51.2± 1.0	-55.3± 1.3	-65.1± 1.6	6.79	6.84
	3	-74.4	-44.9± 2.2	-44.5± 2.2	-49.8± 2.2	3.24	3.60	-51.4± 2.0	-52.8± 2.2	-60.2± 2.4	3.31	3.65
	4	-68.4	-44.8± 2.6	-45.1± 2.8	-51.3± 3.2	3.40	3.46	-52.8± 1.5	-55.4± 1.9	-64.6± 2.7	3.42	3.54
<b>2HVC</b>	1	-97.4	<b>-50.4± 2.6</b>	<b>-46.1± 3.0</b>	<b>-49.4± 3.6</b>	<b>0.59</b>	<b>0.90</b>	<b>-58.4± 1.3</b>	<b>-55.0± 1.1</b>	<b>-58.7± 1.3</b>	<b>0.71</b>	<b>0.81</b>
	2	-89.4	-40.2± 2.0	-36.6± 2.4	-40.2± 2.9	3.28	3.27	-46.9± 1.5	-44.1± 1.7	-48.5± 1.8	3.16	3.33
	3	-88.6	-37.7± 1.9	-34.4± 2.1	-37.6± 2.5	6.77	7.06	-44.3± 0.9	-42.0± 1.1	-45.9± 1.5	6.83	6.87
	4	-84.7	-41.2± 1.7	-37.5± 1.7	-40.2± 1.9	6.33	6.36	-47.5± 1.2	-44.7± 1.7	-48.2± 2.3	6.25	6.33
<b>2IHQ</b>	1	-77.1	<b>-44.7± 1.8</b>	<b>-41.3± 1.7</b>	<b>-43.7± 1.7</b>	<b>0.43</b>	<b>0.96</b>	<b>-50.1± 1.0</b>	<b>-47.3± 1.0</b>	<b>-50.6± 1.0</b>	<b>0.38</b>	<b>0.64</b>
	2	-70.8	-39.2± 2.1	-36.1± 2.2	-39.0± 2.3	2.90	2.84	-44.5± 1.0	-42.4± 1.2	-46.7± 1.5	2.84	2.86
	3	-67.0	-42.5± 1.9	-38.3± 2.2	-40.9± 2.5	2.08	2.20	-47.8± 0.9	-44.5± 1.0	-48.4± 1.2	2.07	2.06
	4	-64.3	-39.0± 2.4	-36.0± 2.7	-38.9± 3.2	6.59	6.67	-45.6± 1.1	-43.2± 1.5	-46.9± 2.1	6.61	6.63
<b>2OZ7</b>	1	-62.3	<b>-59.7± 1.8</b>	<b>-58.6± 2.0</b>	<b>-61.5± 2.5</b>	<b>0.79</b>	<b>0.88</b>	<b>-68.0± 1.1</b>	<b>-68.1± 1.2</b>	<b>-71.6± 1.3</b>	<b>0.84</b>	<b>0.82</b>
	2	-60.0	-54.1± 2.6	-52.4± 4.1	-53.9± 9.1	7.37	7.45	-61.3± 1.5	-60.5± 5.1	-62.5± 5.1	7.42	7.40

				2.7	3.0				1.4	1.5		
	3	-45.3	-55.4± 2.4	-55.1± 2.6	-59.0± 3.3	4.03	4.00	-61.7± 1.4	-63.1± 1.8	-68.3± 2.3	4.10	4.02
	4	-45.0	-50.3± 2.4	-47.2± 2.4	-46.2± 2.4	8.98	9.87	-55.8± 1.6	-53.2± 1.2	-52.2± 1.3	9.16	9.98
<b>2PNU</b>	1	-127.0	<b>-66.7± 2.7</b>	-70.4± 2.9	-83.1± 3.3	0.50	<b>0.51</b>	<b>-76.5± 1.3</b>	<b>-83.2± 1.5</b>	-98.1± 1.7	0.47	<b>0.35</b>
	2	-82.6	-66.5± 2.9	-69.9± 3.2	-83.1± 3.6	0.54	0.78	-76.2± 1.3	-82.8± 1.7	<b>-98.4± 2.2</b>	0.63	0.52
	3	-80.8	-58.2± 1.7	-59.3± 1.6	-68.5± 1.6	8.59	8.81	-65.5± 1.2	-68.5± 1.3	-79.5± 1.7	8.65	8.72
	4	-73.7	-66.6± 1.9	<b>-70.8± 2.1</b>	<b>-84.3± 2.8</b>	<b>0.43</b>	0.61	-75.4± 1.6	-82.7± 1.8	-98.2± 2.0	<b>0.46</b>	0.60
<b>3B66</b>	1	-75.6	<b>-54.5± 1.8</b>	<b>-54.6± 2.0</b>	<b>-63.1± 2.2</b>	<b>0.70</b>	<b>0.84</b>	<b>-63.1± 1.2</b>	<b>-65.9± 1.2</b>	<b>-77.3± 1.3</b>	<b>0.53</b>	<b>0.85</b>
	2	-70.7	-51.4± 4.5	-50.3± 4.8	-57.8± 5.1	7.17	6.59	-60.6± 2.8	-61.4± 2.6	-71.6± 2.9	7.60	6.59
	3	-68.2	-47.0± 2.5	-44.6± 2.3	-51.2± 2.5	2.46	2.53	-54.6± 1.3	-54.7± 1.2	-64.6± 1.7	2.42	2.35
	4	-67.7	-50.9± 2.6	-49.0± 2.8	-55.5± 3.2	2.29	2.40	-57.7± 1.8	-57.4± 2.3	-66.3± 3.3	2.26	2.38
<b>3RLL</b>	1	-67.4	-53.2± 3.1	-50.9± 3.3	-54.1± 3.6	6.97	7.07	-61.7± 1.9	-60.6± 1.8	-65.5± 2.1	7.00	6.97
	2	-57.2	-57.3± 3.0	-54.6± 3.0	<b>-58.0± 3.3</b>	<b>0.57</b>	<b>1.00</b>	-64.9± 1.4	<b>-63.5± 1.4</b>	<b>-68.3± 1.8</b>	<b>0.54</b>	<b>0.86</b>
	3	-57.2	-57.2± 1.9	-54.1± 1.9	-56.8± 2.1	3.03	3.13	-64.9± 1.7	-63.2± 2.0	-67.3± 2.4	3.05	3.10
	4	-54.8	<b>-58.7± 2.3</b>	<b>-54.8± 2.8</b>	-57.2± 2.8	7.25	7.49	<b>-65.9± 2.2</b>	-63.4± 2.4	-67.2± 2.8	7.25	7.49



				<b>2.4</b>	2.8				2.5	3.0		
<b>4OHA</b>	1	-81.0	<b>-39.6± 2.2</b>	<b>-37.5± 2.6</b>	<b>-39.2± 2.9</b>	<b>1.55</b>	<b>0.85</b>	<b>-46.5± 1.5</b>	<b>-45.6± 2.0</b>	<b>-48.6± 2.3</b>	<b>1.31</b>	<b>0.49</b>
	2	-77.3	-38.7± 1.9	-35.7± 2.1	-37.1± 2.6	2.90	2.67	-45.8± 1.1	-44.1± 1.6	-46.9± 2.1	2.91	2.62
	3	-63.9	-33.4± 2.6	-30.8± 2.7	-31.5± 2.9	6.77	7.05	-40.3± 1.8	-38.5± 1.7	-39.8± 2.1	6.77	6.96
	4	-63.8	-32.6± 2.4	-29.5± 2.6	-29.3± 2.9	6.28	6.31	-38.0± 1.8	-35.7± 2.0	-36.3± 2.2	6.31	6.29
<b>5V8Q</b>	1	-81.5	-39.5± 2.2	-36.9± 2.3	-39.6± 2.6	2.39	2.45	-46.7± 1.3	-45.4± 1.6	-49.4± 1.8	2.41	2.45
	2	-81.0	-38.9± 1.8	-36.6± 1.9	-39.7± 2.0	3.36	3.19	-45.2± 0.8	-43.9± 1.0	-48.0± 1.2	3.23	3.22
	3	-78.9	<b>-42.3± 2.8</b>	<b>-39.4± 2.9</b>	<b>-41.3± 3.0</b>	<b>0.71</b>	<b>0.80</b>	<b>-48.6± 1.2</b>	<b>-46.9± 1.1</b>	<b>-50.4± 0.9</b>	<b>0.53</b>	<b>0.89</b>
	4	-77.6	-39.8± 1.5	-37.6± 1.8	-39.2± 2.4	6.45	6.33	-44.5± 1.4	-43.1± 1.7	-45.5± 2.4	6.48	6.34

Underline: Docking pose that resembles crystal structure binding mode; bold: The best binding energy and the lowest RMSD value of the molecule's binding poses within the MMGBSA method; ±: standard deviation in MD snapshots.  
 RMSD1=RMSD value to the crystal structure from the beginning of the MD simulation (the first frame).  
 RMSD20=RMSD value to the crystal structure from the end of the MD simulation (the last frame).

*MD simulations and MMGBSA:* As mentioned, the general aim of this study was to examine the ability of MMGBSA to identify biologically relevant binding modes of small molecules. The AR ligands seemed to be rather stable during the simulations, and none of the docking poses were drastically changed, which was observed both in visual inspection and by calculated RMSD values (Table 3). This could be partly due to the relatively short MD simulations (2.4 ns). However, we expect that usage of several starting poses with such a short simulation time reflects better the situation than if only one starting pose would be employed. After the minimization of the MD snapshots, the energy values were lower and the standard errors were smaller in all cases (Table 3).

In coarse numerical comparison, MMGBSA without minimization using igb5 recognizes the correct binding mode of cross-docked 3RLL ligand (which failed in docking). After minimization, in addition to igb5, igb2 recognizes the correct binding mode (Figure 2B). This shows that small changes in the protein conformations are crucial for the identification of biologically relevant small molecule

binding modes, at least when identification is based on coarse numerical comparison. When MD simulations and MMGBSA originate from self-docked 3RLL ligand, MMGBSA calculations with and without minimization clearly recognize the correct binding mode as the best pose (data not shown). With 5V8Q (which failed in docking), MMGBSA recognizes the correct binding mode with and without minimization, using all three igbs.

With the 1GS4 ligand, after minimization using igb1 and igb2, pose No. 4 ranks over correct pose No. 1 in coarse numerical comparison (Table 3; Figure 3). However, the differences in the energy estimations are not significant and are clearly within the error ranges (Table 3). The 1GS4 protein is a double mutant (Leu701His, Thr877Ala; Figure 3A); as such, when the 1GS4 ligand is cross-docked into wild type protein structure (2AM9), the reverse binding mode also seems reasonable (Figure 3C). Unfortunately, there is no experimental data or crystal structure of the 1GS4 ligand with wild type protein, so the actual binding mode remains unclear. However, this shows that novel and potentially biologically relevant binding modes can be obtained via computational methods.

With the 2PNU ligand, only igb1 without minimization recognizes the correct binding mode in coarse numerical comparison. After minimization, igb2 also raises the correct binding mode into the first place. The 2PNU ligand is a steroidal ligand with a bulky chain (Figure 4A) directed towards helix 12. This creates a challenge when docking. Basically, the ligand has two more or less reasonable binding modes: crystal structure-like and reversed (Figure 4B). The formation of other poses is sterically hindered, leading to a situation in which three poses (out of four saved) resemble the crystal structure binding mode after MD simulation (Figure 4C). All these binding modes behave similarly, with or without minimization in MMGBSA (Table 3; 2PNU conformations 1, 2, and 4). Different docking poses can be produced by increasing the required RMSD value. However, novel poses created this way are unreasonable and, thus, easily separated from the biologically relevant binding mode, both with docking scores and with MMGBSA energy estimations (data not shown).

In all other cases, MMGBSA with and without minimization identifies the correct binding pose to be the best choice in coarse numerical comparison. The MMGBSA results are clearly over the error limits, with ligands from 1Z95, 2AMB, and 2HVC, with and without minimization, with all igbs; with ligands from 2IHQ and 2OZ7 with minimized igb1 and igb2; and with ligands from 3B66 with minimized igb2 and igb5. In other cases, differences in the biologically relevant binding modes are not over the error limit. For most of the cases, MD/MMGBSA does not bring added value because the method is computationally demanding, and correct binding poses can already be identified in molecular docking. Molecular docking, MD, and MMGBSA have all been previously used to study the binding of AR ligands. For example, molecular docking and MD were used to create antagonist AR structures, which led to a vast improvement in docking AR antagonists and separating them from decoy compounds, [78] and MD/MMGBSA has proven accurate enough to study the interaction mechanism between R-bicalutamide and wildtype/mutant ARs [79]. Also, in this study MD/MMGBSA facilitated the identification of some AR ligands and their relevant binding poses. MD helped to improve two mediocre docking poses into biologically relevant forms, which were identified as correct via MMGBSA (2PNU, pose No. 2, and pose No. 4). One igb model before minimization and two igb models after minimization identified the 3RLL ligand, which failed in docking. 5V8Q also failed in docking but was identified using all igb models, both with and without minimization.

#### *PDE4B*

Last, we tested the effect of ligand binding pose selection with MMGBSA to the correlation between experimental and computational data. For this, PDE4B was used as a test case, along with 152 published inhibitors whose inhibition data has been measured using a similar method, enabling a rough assumption about the homogeneity of the biological data. The selected data set is reasonable for

the identification of possible correlation between computed and experimental data, as the pIC<sub>50</sub> spreads seven log units (3.4–10.5).

*Effect of inhibitor binding pose.* All compounds were docked four times into the PDE4B structure, which yielded a total of 488 protein–ligand complexes. The RMSD difference between the acceptable poses was set to 2.5, ensuring that all the binding modes were truly different. All 488 complexes were subjected to short MD simulation with MMGBSA post-processing to determine whether MMGBSA can produce free energy binding estimations that correlate with experimentally obtained IC<sub>50</sub> values. Already, igb models have found that the top poses from docking produce R<sub>2</sub> values of 0.64 to 0.71 (Figure 5: blue). This shows that the data used is likely meaningful. When the lowest MMGBSA energy for each molecule is selected, the correlation with experimental results increases significantly, up to 0.77 (Figure 5: red; igb1, with and without energy minimization of snapshots). Interestingly, when the lowest binding energy is sought, the second ranked docking result is most often selected, by a slight margin; however, together, all four results seem highly useful as the ranked docking solutions are roughly selected in a ratio of 25:32:25:18, with only small variation between the different igb models.

*Effect of energy minimization of MD in simulation snapshots.* Not surprisingly, as with CYP2A6 and AR, the minimized snapshots of protein–ligand complexes yield systematically lower free energy of binding estimations, on average, from 10 kcal/mol (igb1 and igb2) to 13 kcal/mol (igb5) (Figure 5). Furthermore, the average error of the all measurements is lowered by 0.7 kcal/mol, regardless of igb model used. The minimization of the MD snapshots that were used in the MMGBSA calculations barely affects the correlation coefficient (Figure 5).

It is impossible to determine whether the selection of the binding pose for each molecule is correct in large data sets, especially when the structural data is not available. Accordingly, if possible, the methods used in such case should be verified with existing experimental data. In general, for example, docking methods are reasonable for the identification of possible binding modes for each molecule, but the scoring functions associated with them need to compromise between speed and accuracy. Thus, these scoring functions do not recognize the “correct” binding mode as their first option, and their ability to rank active ligands is quite often poor [3,8,35]. Therefore, the choice of the binding pose for each molecule or the identification of active molecules should be done using an alternative scoring function(s) [34,88,89]. In many cases, yet another rapid scoring may help with this; although this is a more time-consuming method, it might give more relevant data, as we show here using MMGBSA, which resulted in a relatively high correlation coefficient between the calculated and experimental data. Whether the selected poses truly represent the biologically relevant complex structure cannot be verified. It is notable that the correlation coefficient for this data set is relatively high and is significantly higher when compared to docking alone (Plants R<sub>2</sub>=0.66; Glide SP R<sub>2</sub>=0.03), or to the use of the best docking pose with Prime MMGBSA (from Plants poses R<sub>2</sub>=0.38; from Glide SP poses R<sub>2</sub>=0.25) or all poses from Plants (R<sub>2</sub>=0.53) [37]. As in this case, the MMGBSA results clearly show that highly active PDE4B inhibitors (pIC<sub>50</sub>>7.5) can be identified relatively reliably, suggesting that this combination of methods could be employed in the future studies where novel PDE4B inhibitors are searched.

## Conclusions

Here, we have used MMGBSA, with and without the energy minimization of trajectory snapshots, to study whether the correct binding conformation of a small molecule to its target protein can be identified using several docking conformations. To provide a wider perspective, we employed three different types of problems: (1) the identification of a substrate binding mode in CYP2A6

enzyme, (2) the identification of AR ligand binding modes, verified using crystal structures, and (3) the improvement of correlation coefficient between experimental and computational data using a set of PDE4B inhibitors. Not surprisingly, in all cases, the minimization of MD simulation snapshots lowered the free energy of binding estimations. This phenomenon was most profoundly seen with *igb5*, which suggests that *igb5* is the most sensitive for precise distances and angles in interactions. In the case of CYP2A6, it was possible to identify the likely binding mode(s) for the substrates by combining the estimation of free energy of binding to a heme and ligand-based RMSD and atomic distance calculations from MD simulation snapshots. In the case of AR, MMGBSA improved the identification of binding conformations in contrast to docking, and with minimized MD trajectory snapshots, it was possible to identify the correct binding mode within error limits. With PDE4B, the correlation between the experimental and computational results was significantly improved when the docking conformation was selected based on MMGBSA. Although energy minimization helped only in the case of AR, it was obvious that the minimization of MD snapshots tends to lower the errors in the free energy of binding estimation throughout the MD simulation, which often helps in the selection of the best binding pose. The choice of short MD simulations (2.4 ns) was based on previous studies that suggest that longer simulations do not necessarily improve the outcome but already minimization of protein-ligand complexes are quite often enough.[16,19] Obviously, if large conformational changes would be expected, then longer simulation time would be a more reasonable choice.

Acknowledgements: The project was funded by the Jenny and Antti Wihuri Foundation (MA), the Emil Aaltonen Foundation (MA) and the Finnish Cultural Foundation, Varsinais-Suomi Regional fund (EJ). CSC, The Finnish IT Centre for Science is acknowledged for their generous computational grants (O.T.P. projects *jyy2516*, *jyy2585*, and *jyy2586*).

**Figure 1.** CYP2A6 crystal structure and docked substrate binding modes. (A) Coumarin (orange carbon atoms) binding to CYP2A6 and unassigned electron density (purple) in the crystal structure of the complex (PDB-code: 1Z10 [44]). Binding poses of coumarin (B) and 6-methylcoumarin (C) after docking. Gray carbon atoms=crystal structure Asn297 and heme, blue carbon atoms=conf. No. 1, green carbon atoms=conf. No. 2, pink carbon atoms=conf. No. 3, brown carbon atoms=conf. No. 4. In (A) the numbering scheme for coumarin is shown.

**Figure 2.** Effect of protein conformation on identification of biologically relevant small molecule binding modes. A) Small changes in protein conformation affect recognition of crystal structure-like ligand poses in docking. For example: 3RLL ligand (turquoise) was docked into 3B66 protein structure (light blue) in which Met745 has turned towards binding pocket. This makes binding pocket unfavorable for 3RLL ligand binding. B) In MD Met745 turns away from the binding pocket (3B66 protein, light blue) and correct binding mode for 3RLL ligand (orange) is recognized with MMGBSA *igb2*. For comparison 3RLL crystal structure complex is shown in grey in both A and B.

**Figure 3.** Potential binding modes of 1GS4 ligand. A) Original crystal structure (PDB: 1GS4) is a double mutant which may affect binding mode. Original crystal structure ligand binding mode is shown with purple carbon atoms. Binding modes in wildtype protein for B) crystal structure-like pose No.1 (orange carbon atoms) and C) reversed pose No.4 (turquoise carbon atoms). Carbon atoms in amino acids are shown in gray in all panels.

**Figure 4.** Binding poses for 2PNU ligand after molecular docking and MD simulations. A) Original crystal structure binding mode of 2PNU ligand (purple carbon atoms) and binding mode with the lowest RMSD value (conf. No. 1, blue carbon atoms; RMSD from last MD frame=0.35). B) Binding modes after docking and C) after MD simulations for three correctly identified binding modes. Blue carbon atoms=conf. No. 1, green carbon atoms=conf. No. 2, brown carbon atoms=conf. No. 4.

**Figure 5.** Effect of choice of ligand pose and energy minimization to correlation coefficient for experimentally measured PDE4B inhibitors. Already the usage of MMGBSA values for the best docking conformation yields meaningful correlation with experimental results (blue) but are clearly better when the best MMGBSA value is used (red). In contrast, the minimization of MD trajectory snapshots does not improve the correlation (right panels) to those without minimization (left panels). The experimental pIC50 values are shown on X-axis and binding energy estimations from MMGBSA are on Y-axis.

#### Conflict of Interest

The authors declare that they have no conflict of interest.

#### References

1. Gilson, M.K.; Zhou, H.-X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42, doi:10.1146/annurev.biophys.36.040306.132550.
2. Jorgensen, W.L. Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009**, *42*, 724–733, doi:10.1021/ar800236t.
3. Warren, G.L.; Andrews, C.W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M.H.; Lindvall, M.; Nevins, N.; Semus, S.F.; Senger, S.; Tedesco, G.; Wall, I.D.; Woolven, J.M.; Peishoff, C.E.; Head, M.S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
4. Feliu, E.; Oliva, B. How different from random are docking predictions when ranked by scoring functions? *Proteins* **2010**, *78*, 3376–3385, doi:10.1002/prot.22844.
5. Plewczynski, D.; Lazniewski, M.; Augustyniak, R.; Ginalski, K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* **2011**, *32*, 742–755, doi:10.1002/jcc.21643.
6. Ramírez, D.; Caballero, J. Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules* **2018**, *23*.
7. Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964–12975, doi:10.1039/C6CP01555G.
8. Cross, J.B.; Thompson, D.C.; Rai, B.K.; Baber, J.C.; Fan, K.Y.; Hu, Y.; Humblet, C. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455–1474.
9. McGaughey, G.B.; Sheridan, R.P.; Bayly, C.I.; Culberson, J.C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.-F.; Cornell, W.D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47*, 1504–1519, doi:10.1021/ci700052x.
10. Shirts, M.R.; Mobley, D.L.; Chodera, J.D. Chapter 4 Alchemical Free Energy Calculations: Ready for Prime



Time? In; Spellmeyer, D.C., Wheeler, R., Eds.; Annual Reports in Computational Chemistry; Elsevier, 2007; Vol. 3, pp. 41–59.

11. Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham, T.E. 3rd Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.
12. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461.
13. Sun, H.; Li, Y.; Shen, M.; Tian, S.; Xu, L.; Pan, P.; Guan, Y.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22035–22045, doi:10.1039/c4cp03179b.
14. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **2011**, *32*, 866–877, doi:10.1002/jcc.21666.
15. Niinivehmas, S.P.; Virtanen, S.I.; Lehtonen, J. V; Postila, P.A.; Pentikäinen, O.T. Comparison of virtual high-throughput screening methods for the identification of phosphodiesterase-5 inhibitors. *J. Chem. Inf. Model.* **2011**, *51*, 1353–63, doi:10.1021/ci1004527.
16. Virtanen, S.I.; Niinivehmas, S.P.; Pentikäinen, O.T. Case-specific performance of MM-PBSA, MM-GBSA, and SIE in virtual screening. *J. Mol. Graph. Model.* **2015**, *62*, 303–318, doi:10.1016/j.jmgm.2015.10.012.
17. Ylilauri, M.; Pentikäinen, O.T. MMGBSA as a tool to understand the binding affinities of filamin-peptide interactions. *J. Chem. Inf. Model.* **2013**, *53*, doi:10.1021/ci4002475.
18. Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* **2014**, *16*, 16719–16729, doi:10.1039/c4cp01388c.
19. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69–82.
20. Karami, L.; Saboury, A.A.; Rezaee, E.; Tabatabai, S.A. Investigation of the binding mode of 1, 3, 4-oxadiazole derivatives as amide-based inhibitors for soluble epoxide hydrolase (sEH) by molecular docking and MM-GBSA. *Eur. Biophys. J.* **2017**, *46*, 445–459, doi:10.1007/s00249-016-1188-0.
21. Rastelli, G.; Del Rio, A.; Degliesposti, G.; Sgobba, M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2010**, *31*, 797–810, doi:10.1002/jcc.21372.
22. Kaus, J.W.; Harder, E.; Lin, T.; Abel, R.; McCammon, J.A.; Wang, L. How To Deal with Multiple Binding Poses in Alchemical Relative Protein–Ligand Binding Free Energy Calculations. *J. Chem. Theory Comput.*

2015, 11, 2670–2679.

23. Chodera, J.D.; Mobley, D.L.; Shirts, M.R.; Dixon, R.W.; Branson, K.; Pande, V.S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160, doi:10.1016/j.sbi.2011.01.011.
24. Mobley, D.L.; Graves, A.P.; Chodera, J.D.; McReynolds, A.C.; Shoichet, B.K.; Dill, K.A. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **2007**, *371*, 1118–1134, doi:10.1016/j.jmb.2007.06.002.
25. Boyce, S.E.; Mobley, D.L.; Rocklin, G.J.; Graves, A.P.; Dill, K.A.; Shoichet, B.K. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. *J. Mol. Biol.* **2009**, *394*, 747–763, doi:10.1016/j.jmb.2009.09.049.
26. Homeyer, N.; Stoll, F.; Hillisch, A.; Gohlke, H. Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. *J. Chem. Theory Comput.* **2014**, *10*, 3331–3344, doi:10.1021/ct5000296.
27. Liu, P.; Kim, B.; Friesner, R.A.; Berne, B.J. Replica exchange with solute tempering: a method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13749–13754, doi:10.1073/pnas.0506346102.
28. Wang, L.; Friesner, R.A.; Berne, B.J. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431–9438, doi:10.1021/jp204407d.
29. Moors, S.L.C.; Michielssens, S.; Ceulemans, A. Improved Replica Exchange Method for Native-State Protein Sampling. *J. Chem. Theory Comput.* **2011**, *7*, 231–237, doi:10.1021/ct100493v.
30. Wang, L.; Berne, B.J.; Friesner, R.A. On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1937–1942, doi:10.1073/pnas.1114017109.
31. Hamelberg, D.; Mongan, J.; McCammon, J.A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929, doi:10.1063/1.1755656.
32. de Oliveira, C.A.F.; Hamelberg, D.; McCammon, J.A. Coupling Accelerated Molecular Dynamics Methods with Thermodynamic Integration Simulations. *J. Chem. Theory Comput.* **2008**, *4*, 1516–1525, doi:10.1021/ct800160q.
33. Arrar, M.; de Oliveira, C.A.F.; Fajer, M.; Sinko, W.; McCammon, J.A. w-REXAMD: A Hamiltonian Replica Exchange Approach to Improve Free Energy Calculations for Systems with Kinetically Trapped Conformations. *J. Chem. Theory Comput.* **2013**, *9*, 18–23, doi:10.1021/ct300896h.
34. Kurkinen, S.T.; Niinivehmas, S.; Ahinko, M.; Lätti, S.; Pentikäinen, O.T.; Postila, P.A. Improving docking performance using negative image-based rescoring. *Front. Pharmacol.* **2018**, *9*, doi:10.3389/fphar.2018.00260.

35. Virtanen, S.I.; Pentikäinen, O.T. Efficient virtual screening using multiple protein conformations described as negative images of the ligand-binding site. *J. Chem. Inf. Model.* **2010**, *50*, 1005–1011, doi:10.1021/ci100121c.
36. Juvonen, R.O.; Kuusisto, M.; Fohrgrup, C.; Pitkänen, M.H.; Nevalainen, T.J.; Auriola, S.; Raunio, H.; Pasanen, M.; Pentikäinen, O.T. Inhibitory effects and oxidation of 6-methylcoumarin, 7-methylcoumarin and 7-formylcoumarin via human CYP2A6 and its mouse and pig orthologous enzymes. *Xenobiotica* **2016**, *46*, 14–24, doi:10.3109/00498254.2015.1048327.
37. Shubina, V.; Niinivehmas, S.; Pentikäinen, O.T. Reliability of virtual screening methods in prediction of PDE4B-inhibitor activity. *Curr. Drug Discov. Technol.* **2015**, *12*, 117–126.
38. Raunio, H.; Rahnasto-Rilla, M. CYP2A6: Genetics, structure, regulation, and function. *Drug Metabol. Drug Interact.* **2012**, *27*, 73–88.
39. Heinlein, C.A.; Chang, C. Androgen receptor in prostate cancer. *Endocr. Rev.* **2004**, *25*, 276–308, doi:10.1210/er.2002-0032.
40. Maxwell, C.R.; Kaness, S.J.; Abel, T.; Siegel, S.J. Phosphodiesterase inhibitors: A novel mechanism for receptor-independent antipsychotic medications. *Neuroscience* **2004**, *129*, 101–107, doi:10.1016/j.neuroscience.2004.07.038.
41. Brown, W.M. Treating COPD with PDE 4 inhibitors. *Int. J. COPD* **2007**, *2*, 517–533.
42. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242, doi:10.1093/nar/28.1.235.
43. Rose, P.W.; Prlic, A.; Altunkaya, A.; Bi, C.; Bradley, A.R.; Christie, C.H.; Costanzo, L. Di; Duarte, J.M.; Dutta, S.; Feng, Z.; Green, R.K.; Goodsell, D.S.; Hudson, B.; Kalro, T.; Lowe, R.; Peisach, E.; Randle, C.; Rose, A.S.; Shao, C.; Tao, Y.-P.; Valasatava, Y.; Voigt, M.; Westbrook, J.D.; Woo, J.; Yang, H.; Young, J.Y.; Zardecki, C.; Berman, H.M.; Burley, S.K. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **2017**, *45*, D271–D281, doi:10.1093/nar/gkw1000.
44. Yano, J.K.; Hsu, M.-H.; Griffin, K.J.; Stout, C.D.; Johnson, E.F. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nat. Struct. Mol. Biol.* **2005**, *12*, 822–823, doi:10.1038/nsmb971.
45. Matias, P.M.; Carrondo, M.A.; Coelho, R.; Thomaz, M.; Zhao, X.-Y.; Wegg, A.; Crusius, K.; Egner, U.; Donner, P. Structural basis for the glucocorticoid response in a mutant human androgen receptor (AR(ccr)) derived from an androgen-independent prostate cancer. *J. Med. Chem.* **2002**, *45*, 1439–1446.
46. Bohl, C.E.; Gao, W.; Miller, D.D.; Bell, C.E.; Dalton, J.T. Structural basis for antagonism and resistance of bicalutamide in prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6201–6206, doi:10.1073/pnas.0500381102.
47. Pereira de Jesus-Tran, K.; Côté, P.-L.; Cantin, L.; Blanchet, J.; Labrie, F.; Breton, R. Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals

- molecular determinants responsible for binding affinity. *Protein Sci.* 2006, 15, 987–999.
48. Wang, F.; Liu, X.; Li, H.; Liang, K.; Miner, J.N.; Hong, M.; Kallel, E.A.; van Oeveren, A.; Zhi, L.; Jiang, T. Structure of the ligand-binding domain (LBD) of human androgen receptor in complex with a selective modulator LGD2226. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 2006, 62, 1067–1071.
49. Sun, C.; Robl, J.A.; Wang, T.C.; Huang, Y.; Kuhns, J.E.; Lupisella, J.A.; Beehler, B.C.; Golla, R.; Sleph, P.G.; Seethala, R.; Fura, A.; Krystek, S.R.J.; An, Y.; Malley, M.F.; Sack, J.S.; Salvati, M.E.; Grover, G.J.; Ostrowski, J.; Hamann, L.G. Discovery of potent, orally-active, and muscle-selective androgen receptor modulators based on an N-aryl-hydroxybicyclohydantoin scaffold. *J. Med. Chem.* 2006, 49, 7596–7599, doi:10.1021/jm061101w.
50. Bohl, C.E.; Wu, Z.; Miller, D.D.; Bell, C.E.; Dalton, J.T. Crystal structure of the T877A human androgen receptor ligand-binding domain complexed to cyproterone acetate provides insight for ligand-induced conformational changes and structure-based drug design. *J. Biol. Chem.* 2007, 282, 13648–13655, doi:10.1074/jbc.M611711200.
51. Cantin, L.; Faucher, F.; Couture, J.-F.; de Jesus-Tran, K.P.; Legrand, P.; Ciobanu, L.C.; Frechette, Y.; Labrecque, R.; Singh, S.M.; Labrie, F.; Breton, R. Structural characterization of the human androgen receptor ligand-binding domain complexed with EM5744, a rationally designed steroidal ligand bearing a bulky chain directed toward helix 12. *J. Biol. Chem.* 2007, 282, 30910–30919, doi:10.1074/jbc.M705524200.
52. Bohl, C.E.; Wu, Z.; Chen, J.; Mohler, M.L.; Yang, J.; Hwang, D.J.; Mustafa, S.; Miller, D.D.; Bell, C.E.; Dalton, J.T. Effect of B-ring substitution pattern on binding mode of propionamide selective androgen receptor modulators. *Bioorg. Med. Chem. Lett.* 2008, 18, 5567–5570, doi:10.1016/j.bmcl.2008.09.002.
53. Duke, C.B.; Jones, A.; Bohl, C.E.; Dalton, J.T.; Miller, D.D. Unexpected binding orientation of bulky-B-ring anti-androgens and implications for future drug targets. *J. Med. Chem.* 2011, 54, 3973–3976, doi:10.1021/jm2000097.
54. Hsu, C.-L.; Liu, J.-S.; Wu, P.-L.; Guan, H.-H.; Chen, Y.-L.; Lin, A.-C.; Ting, H.-J.; Pang, S.-T.; Yeh, S.-D.; Ma, W.-L.; Chen, C.-J.; Wu, W.-G.; Chang, C. Identification of a new androgen receptor (AR) co-regulator BUD31 and related peptides to suppress wild-type and mutated AR-mediated prostate cancer growth via peptide screening and X-ray structure analysis. *Mol. Oncol.* 2014, 8, 1575–1587, doi:10.1016/j.molonc.2014.06.009.
55. Aikawa, K.; Asano, M.; Ono, K.; Habuka, N.; Yano, J.; Wilson, K.; Fujita, H.; Kandori, H.; Hara, T.; Morimoto, M.; Santou, T.; Yamaoka, M.; Nakayama, M.; Hasuoka, A. Synthesis and biological evaluation of novel selective androgen receptor modulators (SARMs) Part III: Discovery of 4-(5-oxopyrrolidine-1-yl)benzotrile derivative 2f as a clinical candidate. *Bioorg. Med. Chem.* 2017, 25, 3330–3349, doi:10.1016/j.bmc.2017.04.018.
56. Goto, T.; Shiina, A.; Murata, T.; Tomii, M.; Yamazaki, T.; Yoshida, K.I.; Yoshino, T.; Suzuki, O.; Sogawa, Y.; Mizukami, K.; Takagi, N.; Yoshitomi, T.; Etori, M.; Tsuchida, H.; Mikkaichi, T.; Nakao, N.; Takahashi, M.; Takahashi, H.; Sasaki, S. Identification of the 5,5-dioxo-7,8-dihydro-6H-thiopyrano[3,2-d]pyrimidine derivatives as highly selective PDE4B inhibitors. *Bioorganic Med. Chem. Lett.* 2014, 24, 893–899,

doi:10.1016/j.bmcl.2013.12.076.

57. Word, J.M.; Lovell, S.C.; Richardson, J.S.; Richardson, D.C. Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-chain Amide Orientation. *J. Mol. Biol.* **1999**, *285*, 1735–1747, doi:10.1006/jmbi.1998.2401.
58. Pelkonen, O.; Rautio, A.; Raunio, H.; Pasanen, M. CYP2A6: A human coumarin 7-hydroxylase. *Toxicology* **2000**, *144*, 139–147.
59. Kuang, R.; Shue, H.J.; Xiao, L.; Blythin, D.J.; Shih, N.Y.; Chen, X.; Gu, D.; Schwerdt, J.; Lin, L.; Ting, P.C.; Cao, J.; Aslanian, R.; Piwinski, J.J.; Prelusky, D.; Wu, P.; Zhang, J.; Zhang, X.; Celly, C.S.; Billah, M.; Wang, P. Discovery of oxazole-based PDE4 inhibitors with picomolar potency. *Bioorganic Med. Chem. Lett.* **2012**, *22*, 2594–2597, doi:10.1016/j.bmcl.2012.01.115.
60. Kuang, R.; Shue, H.J.; Blythin, D.J.; Shih, N.Y.; Gu, D.; Chen, X.; Schwerdt, J.; Lin, L.; Ting, P.C.; Zhu, X.; Aslanian, R.; Piwinski, J.J.; Xiao, L.; Prelusky, D.; Wu, P.; Zhang, J.; Zhang, X.; Celly, C.S.; Minnicozzi, M.; Billah, M.; Wang, P. Discovery of a highly potent series of oxazole-based phosphodiesterase 4 inhibitors. *Bioorganic Med. Chem. Lett.* **2007**, *17*, 5150–5154, doi:10.1016/j.bmcl.2007.06.092.
61. Irvine, M.W.; Patrick, G.L.; Kewney, J.; Hastings, S.F.; MacKenzie, S.J. Rhodanine derivatives as novel inhibitors of PDE4. *Bioorganic Med. Chem. Lett.* **2008**, *18*, 2032–2037, doi:10.1016/j.bmcl.2008.01.117.
62. Allcock, R.W.; Blakli, H.; Jiang, Z.; Johnston, K.A.; Morgan, K.M.; Rosair, G.M.; Iwase, K.; Kohno, Y.; Adams, D.R. Phosphodiesterase inhibitors. Part 1: Synthesis and structure-activity relationships of pyrazolopyridine-pyridazinone PDE inhibitors developed from ibudilast. *Bioorganic Med. Chem. Lett.* **2011**, *21*, 3307–3312, doi:10.1016/j.bmcl.2011.04.021.
63. Ochiai, K.; Takita, S.; Kojima, A.; Eiraku, T.; Iwase, K.; Kishi, T.; Ohinata, A.; Yageta, Y.; Yasue, T.; Adams, D.R.; Kohno, Y. Phosphodiesterase inhibitors. Part 5: Hybrid PDE3/4 inhibitors as dual bronchorelaxant/anti-inflammatory agents for inhaled administration. *Bioorganic Med. Chem. Lett.* **2013**, *23*, 375–381, doi:10.1016/j.bmcl.2012.08.121.
64. Ochiai, K.; Takita, S.; Kojima, A.; Eiraku, T.; Ando, N.; Iwase, K.; Kishi, T.; Ohinata, A.; Yageta, Y.; Yasue, T.; Adams, D.R.; Kohno, Y. Phosphodiesterase inhibitors. Part 4: Design, synthesis and structure-activity relationships of dual PDE3/4-inhibitory fused bicyclic heteroaromatic-4,4-dimethylpyrazolones. *Bioorganic Med. Chem. Lett.* **2012**, *22*, 5833–5838, doi:10.1016/j.bmcl.2012.07.088.
65. Korb, O.; Stütze, T.; Exner, T.E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49*, 84–96, doi:10.1021/ci800298z.
66. Bayly, C.I.; Cieplak, P.; Cornell, W.; Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280, doi:10.1021/j100142a004.
67. Wang, J.M.; Wang, W.; Kollman, P.A. Antechamber: An Accessory Software Package for Molecular Mechanical Calculations. *Abstr. Pap. Am. Chem. S.* **2001**, *222*, U403–U403.



68. Wang, J.; Wang, W.; Kollman, P.A.; Case, D.A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.
69. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935, doi:10.1063/1.445869.
70. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713, doi:10.1021/acs.jctc.5b00255.
71. Giammona, D.A. Force field modifications for united atom heme plus flexible water, University of California, 1984.
72. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802, doi:10.1002/jcc.20289.
73. Lehtonen, J. V.; Still, D.-J.; Rantanen, V.-V.; Ekholm, J.; Björklund, D.; Iftikhar, Z.; Huhtala, M.; Repo, S.; Jussila, A.; Jaakkola, J.; Pentikäinen, O.; Nyrönen, T.; Salminen, T.; Gyllenberg, M.; Johnson, M.S. BODIL: a molecular modeling environment for structure-function analysis and drug design. *J. Comput. Aided. Mol. Des.* **2004**, *18*, 401–419.
74. Roe, D.R.; Cheatham, T.E. 3rd PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095, doi:10.1021/ct400341p.
75. Miller, B.R.; McGee, T.D.; Swails, J.M.; Homeyer, N.; Gohlke, H.; Roitberg, A.E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321, doi:10.1021/ct300418h.
76. Motta, S.; Bonati, L. Modeling Binding with Large Conformational Changes: Key Points in Ensemble-Docking Approaches. *J. Chem. Inf. Model.* **2017**, *57*, 1563–1578, doi:10.1021/acs.jcim.7b00125.
77. Swift, R. V.; Jusoh, S.A.; Offutt, T.L.; Li, E.S.; Amaro, R.E. Knowledge-Based Methods to Train and Optimize Virtual Screening Ensembles. *J. Chem. Inf. Model.* **2016**, *56*, 830–842, doi:10.1021/acs.jcim.5b00684.
78. Wahl, J.; Smieško, M. Endocrine Disruption at the Androgen Receptor: Employing Molecular Dynamics and Docking for Improved Virtual Screening and Toxicity Prediction. *Int. J. Mol. Sci.* **2018**, *19*.
79. Liu, H.; Han, R.; Li, J.; Liu, H.; Zheng, L. Molecular mechanism of R-bicalutamide switching from androgen receptor antagonist to agonist induced by amino acid mutations using molecular dynamics simulations and free energy calculation. *J. Comput. Aided. Mol. Des.* **2016**, *30*, 1189–1200, doi:10.1007/s10822-016-9992-2.
80. Sheridan, R.P.; Korzekwa, K.R.; Torres, R. a.; Walker, M.J. Empirical regioselectivity models for human cytochromes P450 3A4, 2D6, and 2C9. *J. Med. Chem.* **2007**, *50*, 3173–3184, doi:10.1021/jm0613471.
81. Zaretski, J.; Rydberg, P.; Bergeron, C.; Bennett, K.P.; Olsen, L.; Breneman, C.M. RS-predictor models

augmented with SMARTCyp reactivities: Robust metabolic regioselectivity predictions for nine CYP isozymes. *J. Chem. Inf. Model.* **2012**, *52*, 1637–1659, doi:10.1021/ci300009z.

82. Sheng, Y.; Chen, Y.; Wang, L.; Liu, G.; Li, W.; Tang, Y. Effects of protein flexibility on the site of metabolism prediction for CYP2A6 substrates. *J. Mol. Graph. Model.* **2014**, *54*, 90–99, doi:10.1016/j.jmglm.2014.09.005.
83. Kimonen, T.; Juvonen, R.O.; Alhava, E.; Pasanen, M. The inhibition of CYP enzymes in mouse and human liver by pilocarpine. *Br. J. Pharmacol.* **1995**, *114*, 832–836, doi:10.1111/j.1476-5381.1995.tb13279.x.
84. Kinonen, T.; Pasanen, M.; Gynther, J.; Poso, A.; Jarvinen, T.; Alhava, E.; Juvonen, R.O. Competitive inhibition of coumarin 7-hydroxylation by pilocarpine and its interaction with mouse CYP2A5 and human CYP2A6. *Br. J. Pharmacol.* **1995**, *116*, 2625–2630.
85. DeVore, N.M.; Meneely, K.M.; Bart, A.G.; Stephens, E.S.; Battaile, K.P.; Scott, E.E. Structural comparison of cytochromes P450 2A6, 2A13, and 2E1 with pilocarpine. *FEBS J.* **2012**, *279*, 1621–1631, doi:10.1111/j.1742-4658.2011.08412.x.
86. Endo, T.; Ban, M.; Hirata, K.; Yamamoto, A.; Hara, Y.; Momose, Y. Involvement of CYP2A6 in the formation of a novel metabolite, 3-hydroxypilocarpine, from pilocarpine in human liver microsomes. *Drug Metab. Dispos.* **2007**, *35*, 476–483, doi:10.1124/dmd.106.013425.
87. Raunio, H.; Kuusisto, M.; Juvonen, R.O.; Pentikänen, O.T. Modeling of interactions between xenobiotics and cytochrome P450 (CYP) enzymes. *Front. Pharmacology* **2015**, *6*, 1–14, doi:10.3389/fphar.2015.00123.
88. Clark, R.D.; Strizhev, A.; Leonard, J.M.; Blake, J.F.; Matthew, J.B. Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model.* **2002**, *20*, 281–295, doi:10.1016/S1093-3263(01)00125-5.
89. Charifson, P.S.; Corkery, J.J.; Murcko, M.A.; Walters, W.P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109, doi:10.1021/jm990352k.





