



Molecular modeling study of the testosterone metabolizing enzyme UDP-glucuronosyltransferase 2B17

Ingmar Trane^{a,b}, Georg Sager^a, Erik Sveberg Dietrichs^{a,c}, Aina Westrheim Ravna^{a,*}

^a Experimental & Clinical Pharmacology, Department of Medical Biology, Faculty of Health Sciences, University of Tromsø– The Arctic University of Norway, 9037 Tromsø, Norway

^b Research Group in Pharmacology, Department of Pharmacy, Faculty of Health Sciences, University of Tromsø– The Arctic University of Norway, 9037 Tromsø, Norway

^c Division of Diagnostic Services, Department of Clinical Pharmacology, University Hospital of North Norway, Tromsø, Norway

ARTICLE INFO

Keywords:

Molecular modelling
Testosterone
UDP2B17

ABSTRACT

The dominant sex hormone testosterone is mainly metabolized by liver enzymes belonging to the uridine-diphospho (UDP) glucuronosyltransferase (UGT) family. These enzymes are the main phase II enzymes, and they have an important role in the detoxification of endogenous and exogenous compounds in humans. The aim of the present study was to improve the understanding of the binding properties of UGT2B17. A homology modelling procedure was used to generate models of the UGT2B17 enzyme based on templates with known crystal structures. Molecular docking of inhibitors was performed to gain further insights in the interactions between ligand and binding site, and to determine which of the models had the best accuracy. ROC curves were made to evaluate the ability of the models to differentiate between binders (inhibitors) and non-binders (decoys). When comparing the four models, which were based on four different crystal structures, the model based on the 4AMG crystal structure was the most accurate in distinguishing between true binders and non-binders. Investigating pharmacological UGT2B17 inhibition may provide novel treatment for patients with low testosterone levels. Such treatment may elevate endogenous testosterone levels and provide a more predictable increase in serum concentrations rather than un-physiological elevation of serum levels through direct treatment with testosterone, and this could be favorable both for giving a predictable treatment regime with reduced chances of serious adverse effects. The present study may serve as a tool in the search for novel drugs aiming for increasing testosterone levels.

1. Introduction

Testosterone is the dominant male sex hormone, and it plays a key role in the male pubertal development of testes and prostate, as well as promoting masculine characteristics such as increased muscle and bone mass, height, and the growth of body hair. In males, testosterone is produced in testicular Leydig cells, while in females, testosterone is produced in the ovaries, giving serum concentrations of between 5 and 10% of male levels. Through adolescence testosterone helps maintain the libido, sperm production, muscle and bone mass, and male hair pattern. Male hypogonadism is a clinical condition characterized by low serum testosterone levels in combination with a variety of symptoms and signs such as reduced libido and vitality, reduced muscle mass, increased fat mass and depression.³⁰

Testosterone is mainly metabolized by liver enzymes belonging to

the uridine-diphospho (UDP) glucuronosyltransferase (UGT) family. UGTs are the main phase II enzymes, and they have an important role in the detoxification of endogenous and exogenous compounds in humans.¹⁹ UGTs catalyze the transfer of a glucuronyl group to a lipophilic substrate following the phase I reaction, forming a more water soluble and thus more rapidly excreted compound. The glucuronyl group is transferred from the uridine-diphosphoglucuronic acid (UDPGA) co-substrate.¹⁴

There are 22 known UGTs from families UGT1-3 and UGT8 (Meech et al. Physiological Reviews, 2019, 99, 1153–1222), and these are divided into families UGT1-3. Androgens can be metabolized by three different isoforms of the UGT2 family, UGT2B7, UGT2B15 and UGT2B17, with the latter being the most efficient. UGT2B17 also has the ability to conjugate dihydrotestosterone (DHT).^{19,25,11}

UGTs are composed of two functional domains, a variable N-terminal

* Corresponding author.

E-mail address: Aina.W.Ravna@uit.no (A. Westrheim Ravna).

<https://doi.org/10.1016/j.bmc.2021.116060>

Received 10 August 2020; Received in revised form 25 January 2021; Accepted 1 February 2021

Available online 20 February 2021

0968-0896/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

(NT) domain (residues 1–265) and a highly conserved C-terminal (CT) domain (residues 266–530), with a catalytic site in the cleft between the two domains. The NT domain contains a signal peptide that mediates the integration into the ER-lumen, the aglycone binding site, and a membrane interacting region. The CT domain contains most of the UDPGA cofactor binding site and a transmembrane helix near the carboxy-terminus with a cytosolic tail. The enzyme is predicted to form dimers in endoplasmic reticulum membranes, this may have an effect on function and receptor ligand specificity.^{19,25,20}

UGT inhibition is important to investigate when developing novel potential drugs. Inhibition of drug metabolizing enzymes is a major mechanism in drug-drug interactions, and a number of cases of drug-drug interactions via inhibition of UDP-glucuronosyltransferases (UGTs) have been reported (<https://dmd.aspetjournals.org/content/43/6/812>). Among approximately 500 known inhibitors are atazanavir, quinidine, diclofenac (<https://www.cyprotex.com/admek/in-vitro-metabolism/ugt-inhibition#:~:text=The%20following%20positive%20control%20inhibitors%20are%20used%20in,3%20UGT1A4%2C%20UGT1A6%2C%20UGT1A9%20and%20UGT2B7%20inhibitor%3A%20>) lapatinib, pazopanib, regorafenib and sorafenib (<https://www.sciencedirect.com/science/article/abs/pii/S0006295219303065>).

The aim of the present study was to improve the understanding of the binding properties of UGT2B17, thereby making it possible to develop selective inhibitors of the enzyme. Inhibitors of UGT2B17 could help maintain normal testosterone levels in patients with clinical conditions that reduce testosterone production. The models presented in this study may be used in drug development studies in order to treat male hypogonadism and other conditions with low testosterone, and coordinates are available as supplementary material.

A homology modelling procedure was used to generate models of the UGT2B17 enzyme based on templates with known crystal structures. Molecular docking of inhibitors on the models was performed to gain further insights in the interactions between ligand and binding site, and to determine which of the models had the best accuracy.

2. Materials and methods

2.1. Software and databases

The Molsoft Internal Coordinates Mechanics (ICM) (Version 3.8.7) program,² which gives a general modelling and structure prediction framework for many different tasks of structural biology and rational drug design, was used to build homology models of the enzymes and docking of ligands and decoys in this study. The ICM method has been extensively validated in bioinformatics and drug discovery projects.^{13,7,16} The Protein Data Bank (PDB),⁵ which is a data bank of experimentally determined 3D structures of biological macromolecules, was used to retrieve templates. The Universal Protein Resource Knowledgebase (UniProtKB),⁴ which is a comprehensive resource for protein sequence and functional information with detailed annotations, was used to find amino acid sequences for target and template proteins. The Basic Logical Alignment Search Tool (Blast),³ which is a search tool that finds regions of similarity between biological sequences, from the National Center for Biotechnology Information sequence database, was used to find potential templates with sequence homology of known 3D structures. The Structural Analysis and Verification Server (SAVES) v5.0 (<https://servicesn.mbi.ucla.edu/SAVES/>), which is a part of web services provided by the Molecular Biology Institute at the University of California, Los Angeles, was used to analyze and validate protein structures. ChEMBL,¹⁰ which is a database of bioactive compounds with drug-like properties, was used to search for compounds, targets and assays. DecoyFinder,⁸ which is a graphical tool designed to aid molecular docking programs by providing challenging decoys for a given group of active ligands, was used to retrieve decoys with similar physicochemical properties assumed to be inactive for the UGT2B17 enzyme. The software acquires the decoys directly from the ZINC compounds database.⁸

Table 1

Templates chosen for homology modelling.

| PDB ID | Name | Sequence identity | Resolution | Deposition author |
|--------|---------------------------------|-------------------|------------|-------------------|
| 3WAD | Glycosyltransferase VinC | 20% | 2.00 Å | ²¹ |
| 4AMG | Glycosyltransferase SnogD | 22% | 2.59 Å | ⁹ |
| 4M83 | Glycosyltransferase OleD | 21% | 1.70 Å | ²⁸ |
| 2O6L | UDP-glucuronosyltransferase 2B7 | 82% | 1.80 Å | ²⁰ |

3. Homology modelling

3.1. Template identification

The amino acid sequence of human UGT2B17 was retrieved from the UniProtKB database,⁴ with accession number O75795. Close homologues of the human UGT2B17 were found using the Blast search tool³ for a sequence similarity search with the target sequence as query. A standard protein–protein Blast was performed on the 530 residues of the human UGT2B17 amino acid sequence, resulting in a list of potential templates with available crystal structures.

Most of the potential templates were UGTs with a sequence identity of about 20% to UGT2B17, but one partial structure of an UGT had a high sequence identity to UGT2B17. The partial structure (PDB id: 2O6L)²⁰ consisted of the CT domain of the closely related enzyme human UDP-glucuronosyltransferase 2B7 (UGT2B7), and had a sequence identity of 82% with the query sequence. The CT domain included most of the residues that make up the binding site of the UDPGA cofactor. Consequently, this made the crystal structure useful as a template for a model of the CT domain of target, but also as a part of a multi template model where two templates are combined in the modelling process. The crystal structure of UGT2B7 was crystallized as a dimer, with chains designated A and B. Since chain B lacked some residues located close to the binding site, chain A was chosen for modelling.

Because of the high sequence identity of UGT2B7 with the target, combined with the low sequence identity in most of the other potential templates, a multiple template modelling procedure was the best option for an acceptable full length model of both domains. The rationale for building a model with both domains was based on studies indicating interactions between the co-factor UDPGA and residues in the NT domain.^{25,20} In addition, a partial model of the CT domain of the enzyme was built based on UGT2B7 alone. This model could lack some residues of importance to UDPGA binding, but would have higher sequence identity.

To model both domains of the target protein there was a need for templates with acceptable sequence identity to the NT, in addition to the partial structure of UGT2B7. A delta-Blast algorithm of the first 284 residues from the NT domain of target was performed, resulting in a long list of potential templates. UGTs were marked for next iteration, and followed by a psi-Blast algorithm. This resulted in a new list of 48 potential templates. The templates obtained from the Blast search tool were shortlisted and investigated further based on: (1) conserved UDPGA binding site, (2) sequence identity, (3) query cover, (4) resolution, and (5) expectation-value. All the templates considered had most of the UDPGA binding site conserved, but ultimately this region, which was focus of the present study, was modelled from the UGT2B7 template. The query cover of all templates were all over 80%, except from the partial structure with 31% query cover. The resolution of the chosen templates ranged from 1.7 Å to 2.59 Å. The expectation value (e-value) is a parameter describing the number of different alignments expected to occur by chance in a database search, the lower the e-value, the better the alignment. Proteins with an e-value above 0.0001 were excluded. Thus, the following crystal structures were chosen as templates for homology modelling, PDB id: 3WAD,²¹ 4AMG,⁹ 4M83²⁸ and 2O6L,²⁰ as shown in Table 1. The sequence identities between the target and

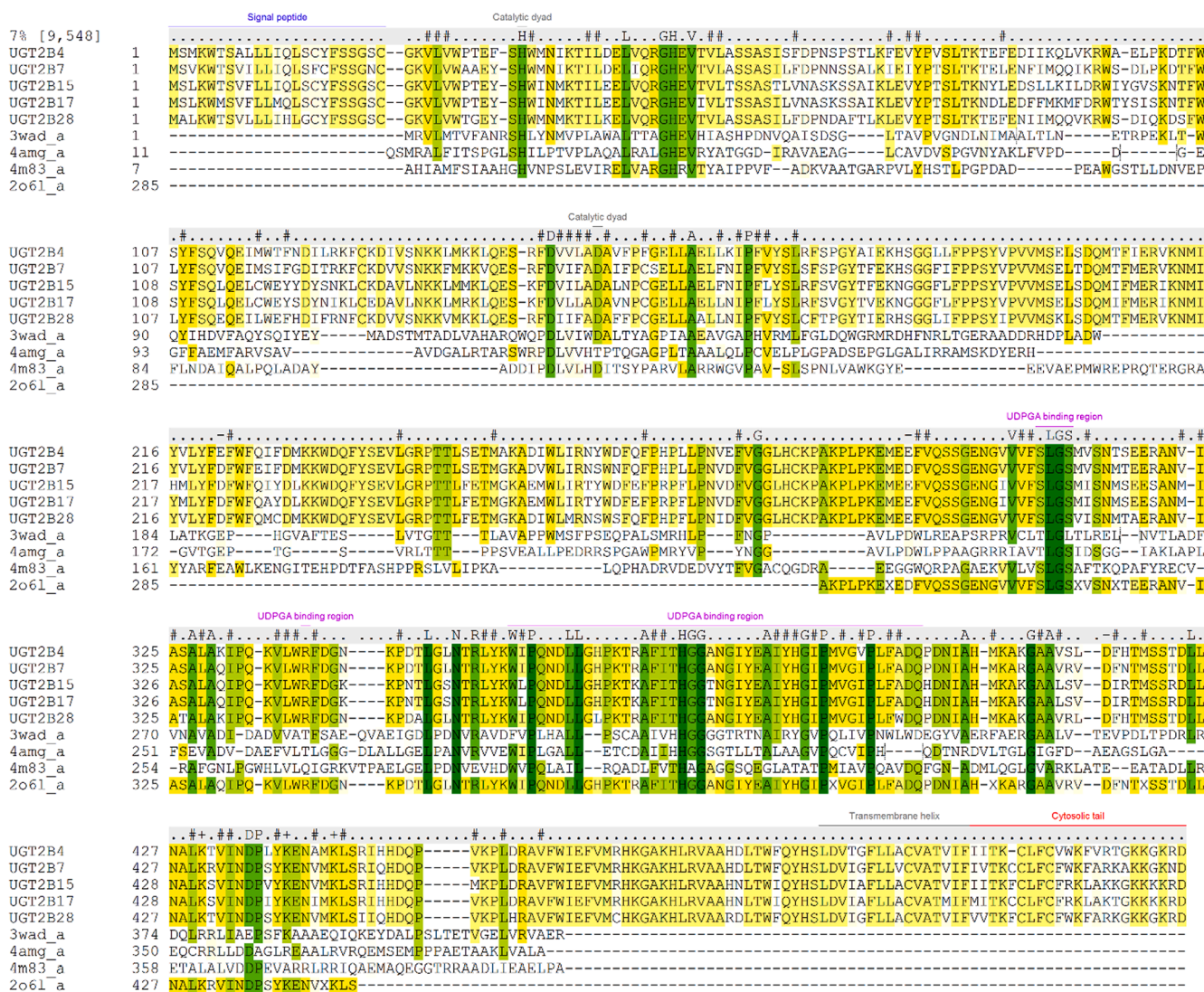


Figure 1. Multiple sequence alignment of several homologous UGTs and the chosen templates. Areas with dark green colour indicates fully conserved residues, yellow colour indicates semi conserved. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

template sequences are much lower than the requirement ($>$ or $=$ 30%) for a structure that can be used as a template for homology modeling, and the sequence similarities are 25.6% (3WAD), 24.8% (4AMG), 27.3% (4 M83) and 84.4% (2O6L). The values for 3WAD, 4AMD and 4 M83 are below what is regarded as acceptable in homology modeling, but since the binding site of interest was located in the CT-region where the 2O6L template was used, the sequence identity in the region of interest was acceptable.

3.2. Sequence alignment

The templates selected for homology modelling were aligned with the sequence of UGT2B17 using the alignment tool in the ICM software. The sequence of 2O6L²⁰ needed no adjustment because of the high sequence identity. The other templates had relatively low homology with target, and needed manual adjustment. By using a multiple sequence alignment of the templates combined with several other human UGTs, a basis for further adjustment was built, as shown in Figure 1.

Site-directed mutagenesis studies of human UGTs^{25,26,29,23} where used to guide the alignment process. Residues H35 and D152 act as a catalytic dyad in the catalytic reaction initiating the glucuronidation mechanism of the enzyme. Residues R49 and H51 have a role in function

and structural integrity required for optimal catalytic activity, but are not directly involved in substrate binding. The residue F90 forms aromatic ring stacking interactions with phenolic substrates. The residue S121 is required for the ability to conjugate C19 steroids at the 3 α -OH position, thereby being involved in steroid specificity. The residues S309 and R339, in

addition to many residues in the region 357–400 are involved in UDPGA binding, and forms the binding pocket.^{25,26,29,23} The adjusted sequence alignments shown in Figures 2-5 were used to build the models of UGT2B17.

3.3. Model building

One partial model of the CT domain, and three full length models of both domains of UGT2B17 were built based on four different crystal structures. All templates belong to the UGT family and were the most suitable candidates with regards to the criteria for template identification.

1. Bacterial glycosyltransferase VinC in complex with magnesium ion PDB id: 3WAD,²¹ resolution 2.0 Å, chain A, length 398 residues
2. Bacterial glycosyltransferase SnogD PDB id: 4AMG,⁹ resolution 2.59 Å, chain A, length 362 residues

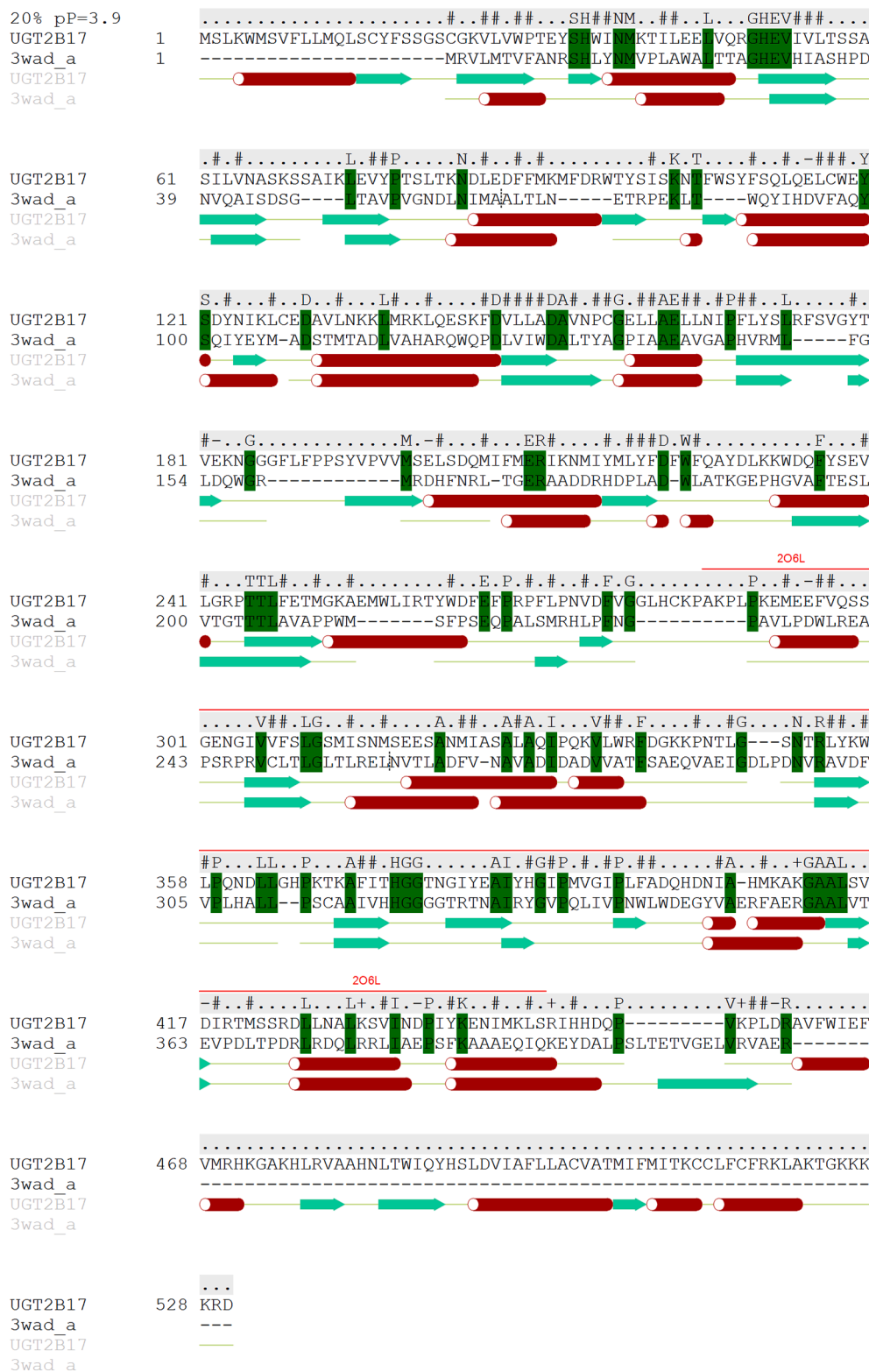


Figure 2. Sequence alignment for UGT2B17 and 3WAD used for homology modelling. Areas with green colour indicates conserved residues. Red annotation marks region for multiple templates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

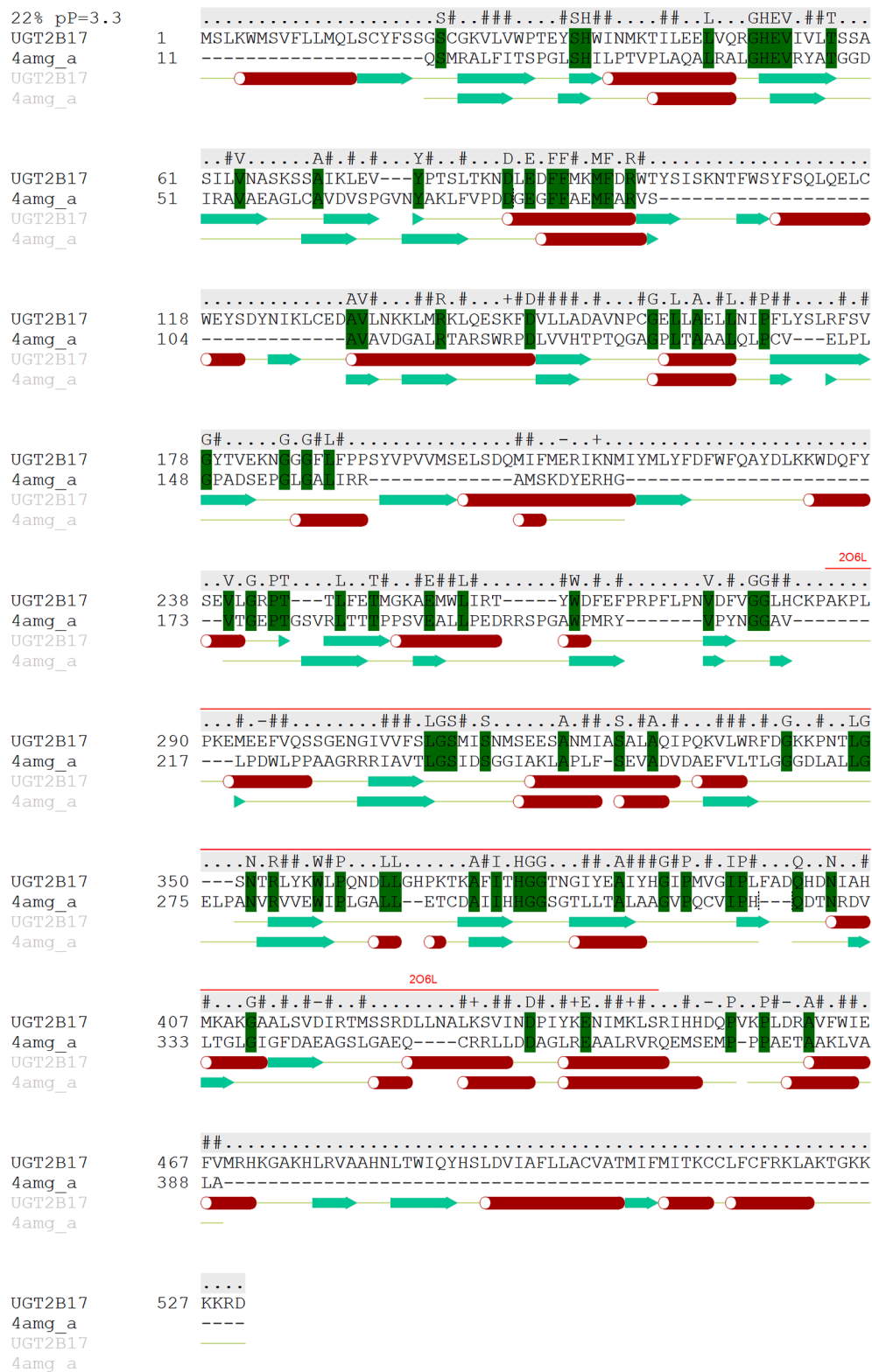


Figure 3. Sequence alignment for UGT2B17 and 4AMG used for homology modelling. Areas with green colour indicates conserved residues. Red annotation marks region for multiple templates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

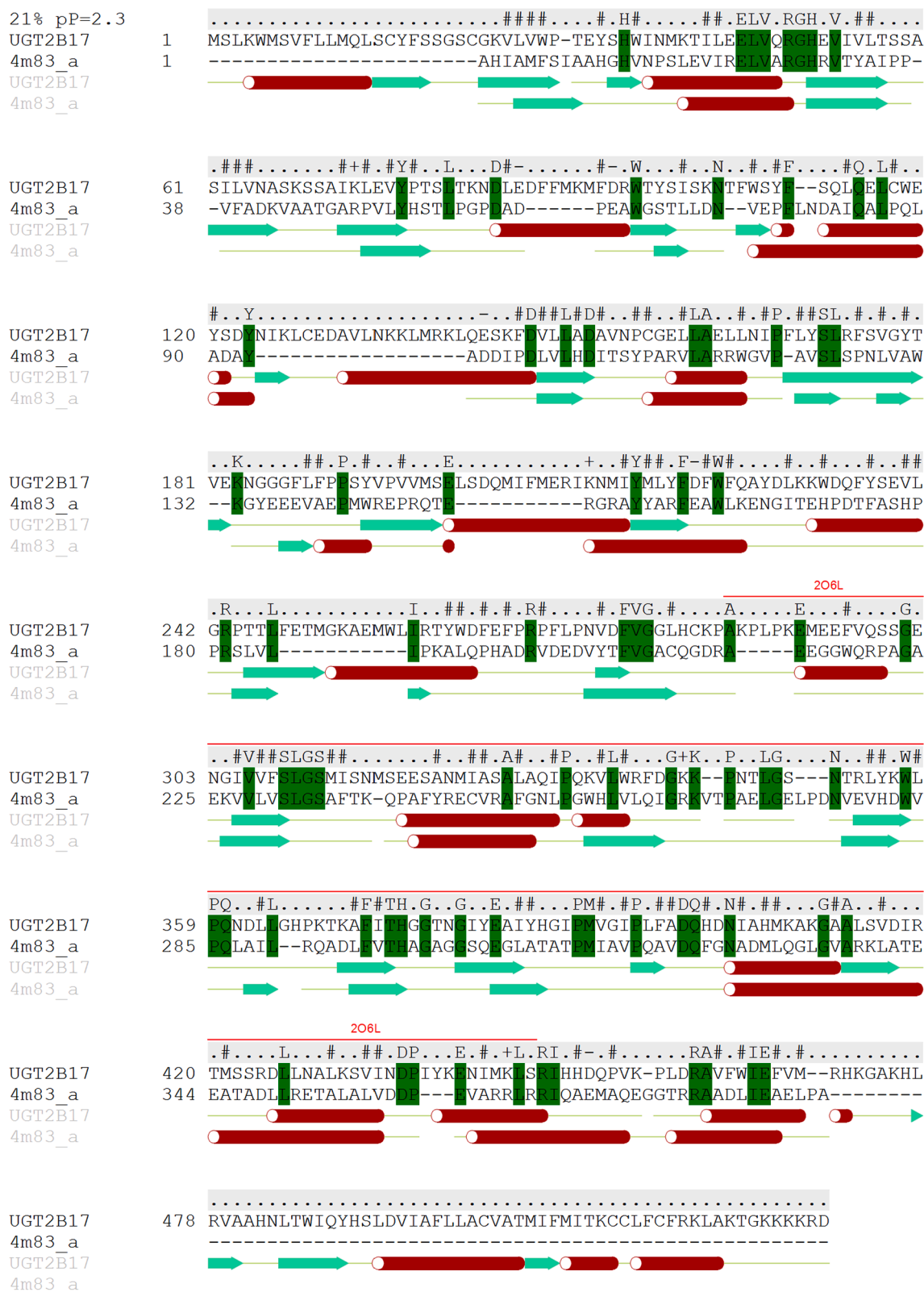


Figure 4. Sequence alignment for UGT2B17 and 4 M83 used for homology modelling. Areas with green colour indicates conserved residues. Red annotation marks region for multiple templates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

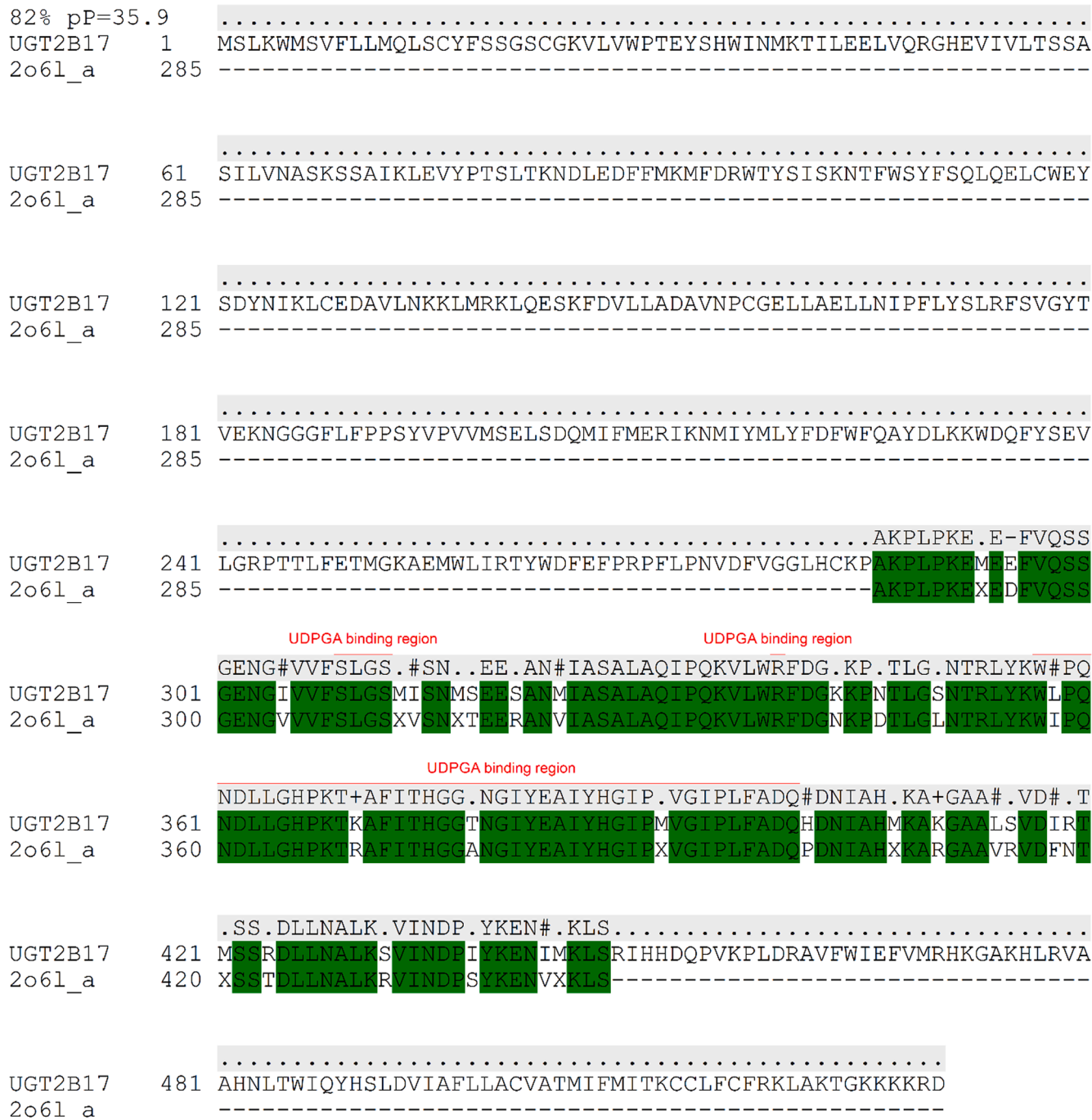











Figure 5. Sequence alignment for UGT2B17 and 206L used for homology modelling. Areas with green colour indicates conserved residues. Red annotation indicates the UDPGA binding region. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
UGT2B17 inhibitors.

| ID | Structure | Chemical composition | CHEMBL ID | PubMed ID | IC50 (μ M) | Ki (μ M) |
|----|---|----------------------|--------------|------------|-----------------|---------------|
| 1 |  | C29H34N7O | CHEMBL941 | 96,642,944 | 0.8 | 0.4 |
| 2 |  | C30H18O10 | CHEMBL63354 | 29,470,958 | 2.1 | 2.1 |
| 3 |  | C15H16O2 | CHEMBL418971 | 23,948,605 | | 19.9 |
| 4 |  | C21H3O | CHEMBL376840 | 17,474,732 | | 21.8 |
| 5 |  | C9H12N2O12P2 | CHEMBL130266 | 17,998,297 | | 100.0 |
| 6 |  | C19H16O4 | CHEMBL1464 | 25,393,417 | | 166.8 |
| 7 |  | C21H22O4 | CHEMBL139702 | 26,875,642 | 32.0 | |
| 8 |  | C31H33N3O6S | CHEMBL603 | 25,834,030 | 50.0 | |
| 9 |  | C14H11Cl2NO2 | CHEMBL139 | 19,643,121 | 65.0 | |
| 10 |  | C13H17O2 | CHEMBL521 | 19,643,121 | 1340.0 | |

3. Bacterial glycosyltransferase Oled in complex with Erythromycin A and UDP PDB id: 4M83,²⁸ resolution 1.7 Å, chain A, length 393 residues

4. Human UDP glucuronosyltransferase 2B7 PDB id: 2O6L,²⁰ resolution 1.8 Å, chain A, length 162 residues

The partial model and three complete initial models were made using the Homology macro in ICM. The Multi-Template Model Editor macro was used to improve the quality of the three GT based models by adding 2O6L²⁰ as a second template for their CT domain. Since the target sequence contained a signal peptide, a transmembrane region and a cytosolic tail not present in the GT templates, the excessive carboxy- and amino terminus tails generated by the modelling procedure were trimmed of the models to avoid them interacting with the secondary structures.

3.4. Model refinement

The Refine Model macro² of the ICM software was used to refine the

built models, a full refinement and optimization of backbone, sidechains and loops were carried out. This refinement macro included (1) Monte Carlo fast simulations for sampling of the conformational space of side chains, (2) repeated annealing of the backbone with tethers, and (3) a second run of Monte Carlo fast simulations on the side chains. Each repetition of Monte-Carlo fast samples the conformational space of the molecule with the ICM global optimization procedure, which consists of a random move followed by a local energy minimization, and then a complete energy calculation. Based on the energy and temperature, the repetition is either accepted or rejected.

3.5. Model validation

The SAVES metaserver (<https://servicesn.mbi.ucla.edu/SAVES/>) was used to analyze and validate the built models. Of the different programs available in the metaserver, ProCheck¹⁷ and WhatCheck¹² were chosen for the validation. ProCheck investigates the stereo chemical quality of a protein structure by analyzing the overall and residue-by-residue geometry, and the result of the analysis is represented by a

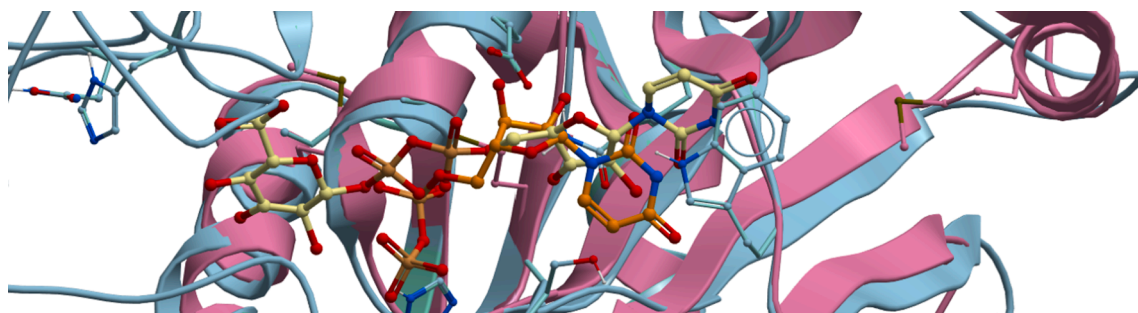


Figure 6. Superimposition of UDPGA (yellow) binding site with 2O6L (pink) and 4 M83 (light blue) with UDP (orange). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

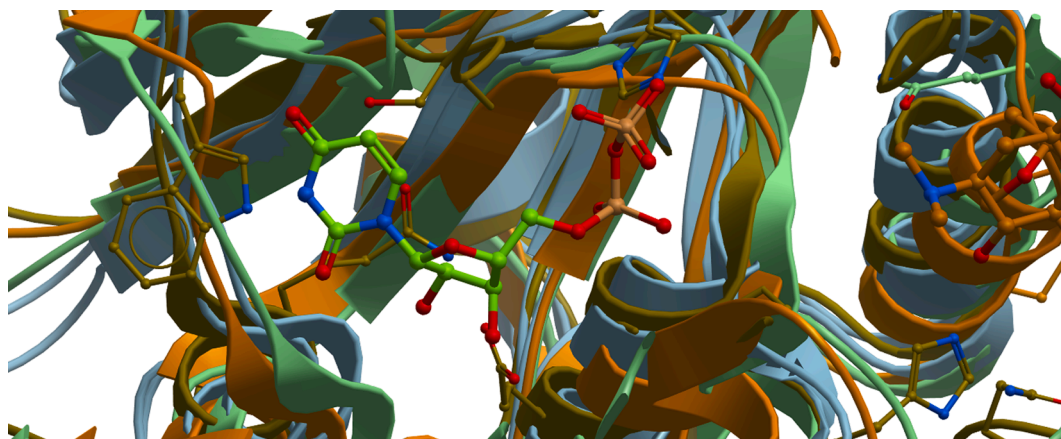


Figure 7. Superimposition of all four models. Color coding: 4 M83 (pink), 2O6L (yellow), 3WAD (orange), AMG (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Ramachandran plot. WhatCheck did extensive checking of many stereochemical parameters of the residues in the models.

In order to identify structural differences between the models and their templates, the root-mean-square-deviation (RMSD) was also calculated by ICM for the CT domain and for the binding pocket. RMSD is a measure of the degree of similarity of two protein 3D structures, and it calculates the average distance between equivalent backbone C α atoms by superimposing the models on their templates.^{18,16}

4. Molecular docking

Molecular docking is a method used to predict protein ligand interactions within a targeted binding site, and score their potential complementarity. Exploring these interactions are important for our understanding of how the protein function, and for the development of new drug molecules. Studies on the co-factor binding site of UGT2B17 may provide insights on the formation of the ligand-protein complex, and the intermolecular forces deciding specificity and affinity of a ligand. The formation of a ligand-protein complex may lead to structural changes in both ligand and protein. Retrieving information about protein ligand interactions can assist in designing new inhibitors with a good fit in the binding pocket.²²

4.1. Inhibitors and decoys

To validate the models ability to differentiate between inhibitors and decoys, a set of known inhibitors of UGT2B17 were retrieved from PubChem and ChEMBL, examining studies where the UDPGA binding site in UGT2B17, or UGTs in general had been investigated. 17 inhibitors with varying ability to inhibit UGT2B17 were identified, of these were 10 selected based on known IC₅₀ or K_i values, as shown in Table 2.

Ideally, the docking would be performed with experimentally determined decoys for target, but none were available at time of this study. To acquire decoys, the known inhibitors were entered as templates into the Decoyfinder software.⁸ The DecoyFinder software finds molecules which have similar number of rotational bonds, HBA, HBD, logP value, and molecular weight, but are chemically different from the active ligands used as input. The software generated a set of 145 decoy substances with similar physiochemical properties as the inhibitors. The decoys were inserted into a chemical table with the inhibitors, giving a dataset of 155 substances ready for docking.

4.2. Ligand and model preparation

Inhibitors and decoys were converted from 2D structures to 3D conformations, and formal charges were assigned to physiological

conditions (pH = 7.0) The models had their hydrogens optimized, any missing side chains were hidden, and the residues Histidine, Proline, Asparagine, Glutamine and Cysteine were also optimized.

4.3. Identification of ligand binding pocket

The UGT2B17 enzyme have an aglycone binding site, and a co-factor binding site, the latter being the focus of this study. Comparison of the CT domain crystal structure of human UGT2B7 to other GT family enzymes revealed that UDPGA binds to the same site as the co-factor in these enzymes. Bacterial enzymes are part of the GT family and use UDP-glucose as co-factor substrate, while humans use UDP-glucuronic acid. One of the chosen templates, namely the structure of GT OleD (PDB id: 4M83²⁸) was crystalized in complex with UDP, this indicated the putative binding site. Interestingly, human UGTs can utilise multiple UDP-sugars with diverse glycone moieties, however, UDP-glucuronic acid remains the primary co-factor.

Since UDP lacked the glucuronic acid moiety of UDPGA, the binding site determined from UDP by ICM would have been too short and missed several residues of importance. This was solved by an initial docking of UDPGA in the pocket of Model_2O6L indicated by the superimposed UDP, giving an excellent pose and a good score for the docked co-factor. This UDPGA pose was later superimposed on all models, and residues in the models in a 5 Å vicinity to the superimposed ligand were selected, thereby defining the binding pocket to be used in the main docking procedure. In the procedure of using UDPGA to define the binding pocket, the binding pocket of 2O6L and 4 M83 has been compared to confirm the validity of the superimposition of UDPGA with UDP (Figure 6). The cofactor binding site of the CT domain are shown in Figure 7.

4.4. Docking of inhibitors and decoys

Docking of known inhibitors of UGT2B17 and decoys into the putative binding pocket of the models was carried out to investigate the accuracy of the models. A semi-flexible docking approach was used in this study. This keeps the ligands fully flexible, and the homology models are represented as rigid structures. Protein structure backbone and sidechains of enzymes are considered flexible in nature, with an approach using a rigid binding pocket in the docking, this flexibility is not taken into account.

The binding pocket used in the docking is visualized as an energy grid, with pre-calculated energy maps representing ligand binding interactions such as van der Waals, hydrogen-bonding, electrostatics, and hydrophobic interactions. The box defining the energy grid maps was set to include the entire binding pocket, and to exclude neighboring cavities

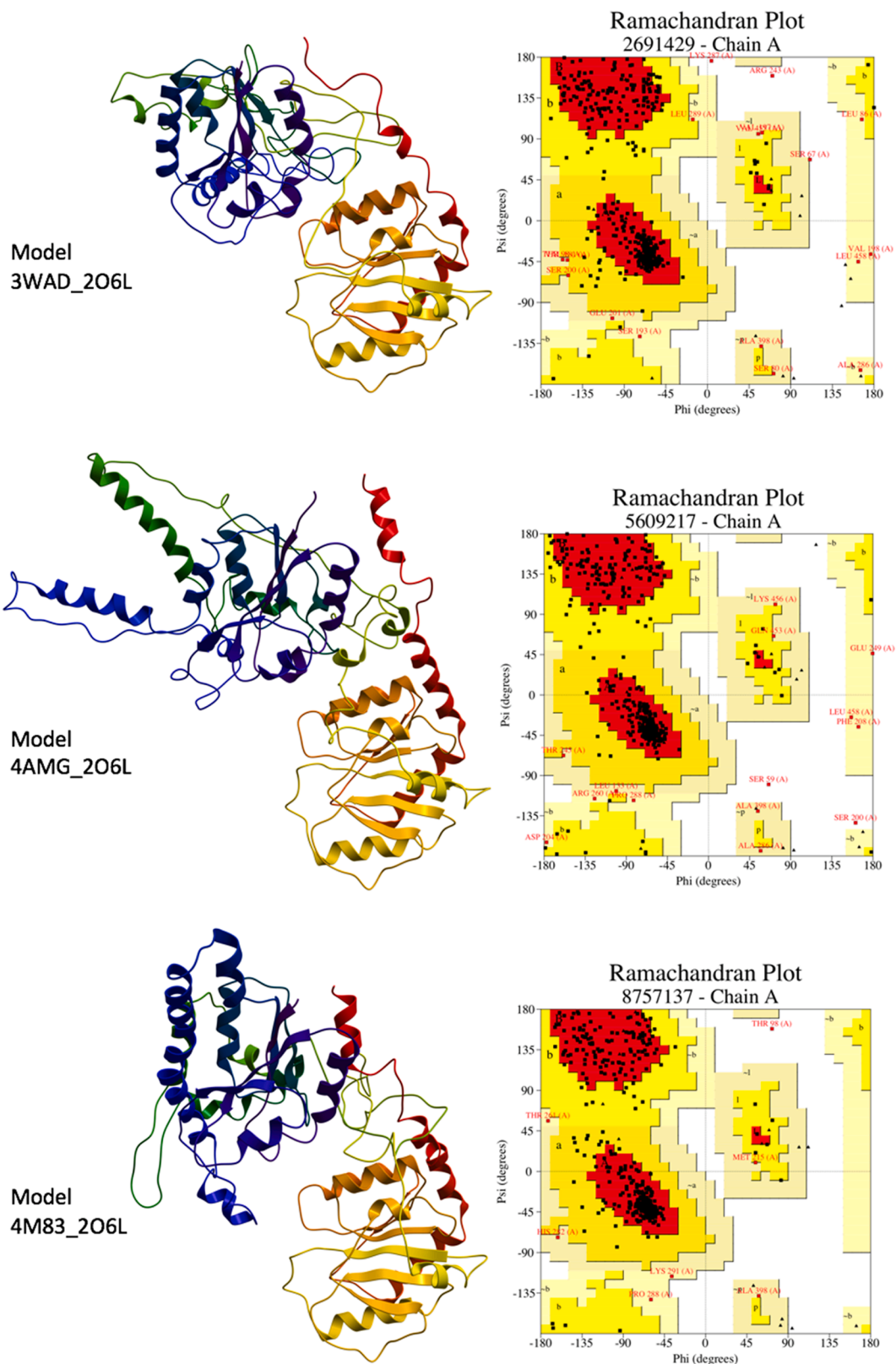


Figure 8. Homology models and their corresponding Ramachandran Plot. The models are visualized as ribbons, with the protein chain colour scheme of a rainbow, from blue at the amino-terminus to red at the carboxy-terminus. The Ramachandran Plot was generated by ProCheck, showing residues in the most favoured regions (red), additionally allowed regions (yellow), generously allowed regions (beige) and disallowed regions (white). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

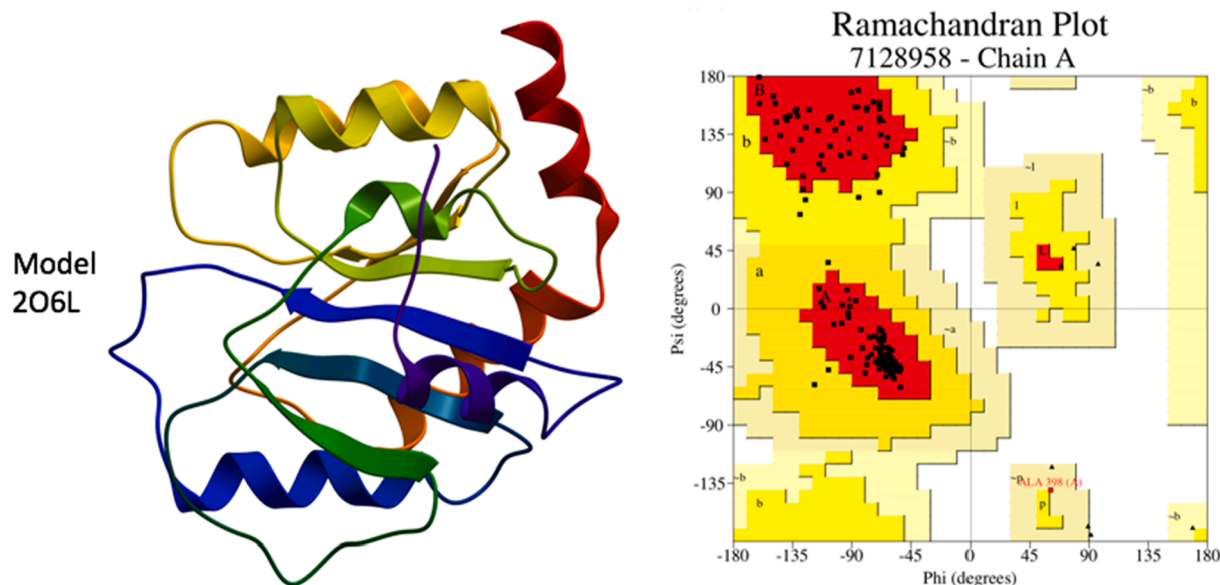


Figure 9. Homology model 2O6L and its corresponding Ramachandran Plot. The model is visualized as ribbon, with the protein chain colour scheme of a rainbow, from blue at the amino-terminus to red at the carboxy-terminus. The Ramachandran Plot was generated by ProCheck, showing residues in the most favoured regions (red), additionally allowed regions (yellow), generously allowed regions (beige) and disallowed regions (white). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Ramachandran plot statistics generated by ProCheck.

| Model | Most favoured regions | Additionally allowed regions | Generously allowed regions | Disallowed regions |
|-------|-----------------------|------------------------------|----------------------------|--------------------|
| 3WAD | 83.1% | 12.6% | 3.6% | 0.8% |
| 4AMG | 86.0% | 10.8% | 2.8% | 0.5% |
| 4 M83 | 87.5% | 11.0% | 1.3% | 0.3% |
| 2O6L | 94.2% | 5.0% | 0.7% | 0.0% |

which could disturb the docking. The ligand binding probe in the binding pocket was kept at default, as predicted by ICM using the Monte Carlo global optimization procedure.

The chemical table of inhibitors and decoys was docked into the binding pocket using the docking macro of ICM. Three parallel dockings runs were done on all four models. Once the docking was finished, a collection of the most energetically favorable poses of the ligands were

collected and could be displayed interactively inside the binding pocket.

4.5. Evaluation of docking

The docking was evaluated using ROC curves, giving insights to the overall predictability of the built models. The scores obtained by the docking process were analyzed using the inbuilt ROC-curve command in ICM. The positives (inhibitors) docked were labelled as 1, while negatives (decoys) were labelled as 0. The results were displayed as ROC curves, the AUC was calculated and interpreted.

5. Results

The final models, Model_3WAD, Model_4AMG and Model_4M83, and Model_2O6L (partial model), are shown in Figures 8 and 9. The partial model contained the co-factor binding site, and all full length models contained the typical structural characteristics of GTs and UGTs, an NT domain with the aglycone binding site, a conserved CT domain with the

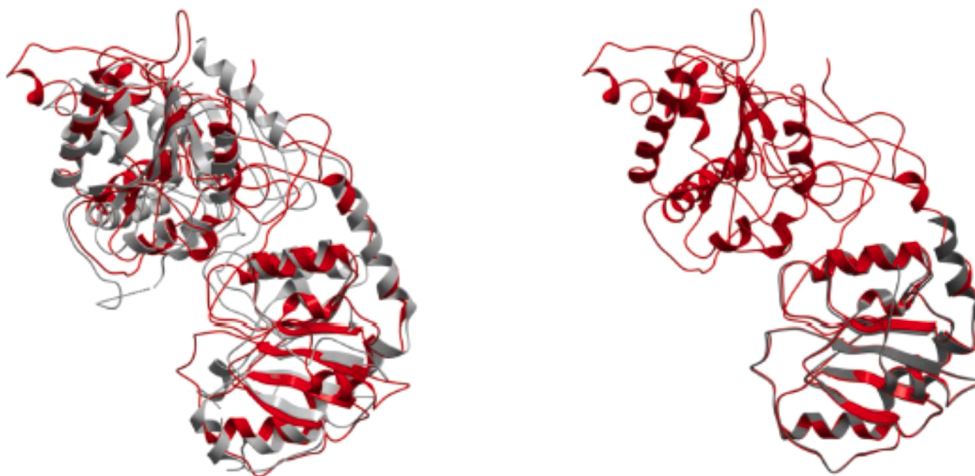


Figure 10. Model_3WAD superimposed on templates. Model shown in red, templates in grey. Left panel shows 3WAD template, and right panel shows 2O6L template. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

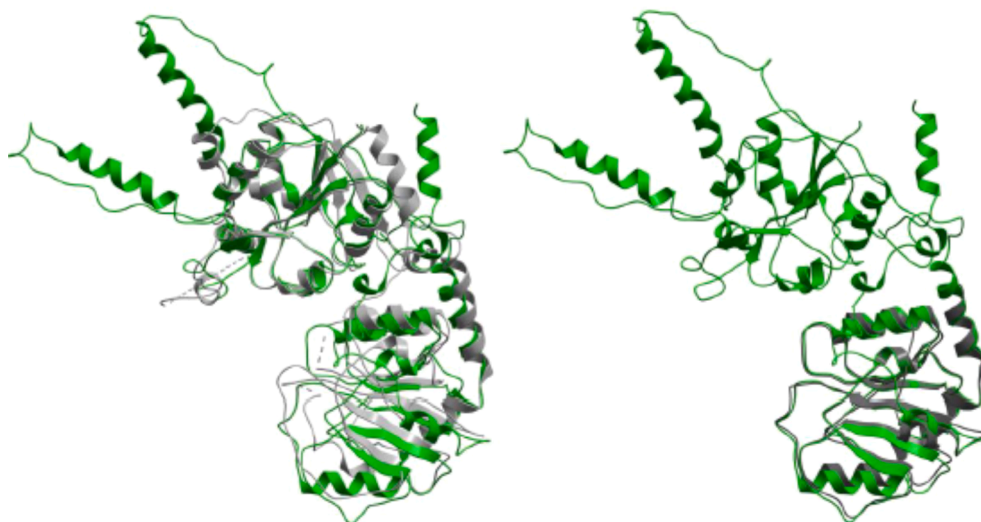


Figure 11. Model_4AMG superimposed on templates. Model shown in green, templates in grey. Left panel shows 4AMG template, and right panel shows 2O6L template. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

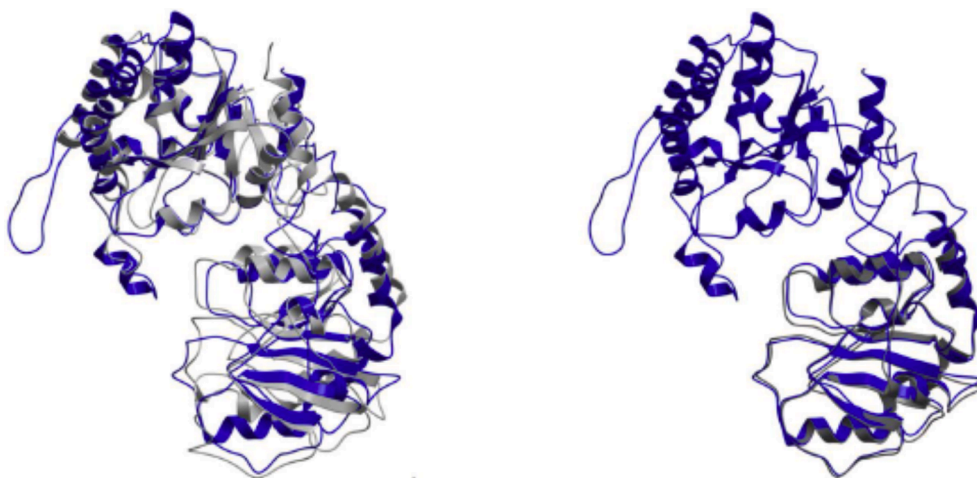


Figure 12. Model_4M83 superimposed on templates. Model shown in blue, templates in grey. Left panel shows 4M83 template, and right panel shows 2O6L template. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

co-factor binding site, and a catalytic cleft between them.

The Ramachandran plots of the models are displayed in Figures 8 and 9, and their ProCheck statistics are shown in Table 3. The Ramachandran plot visualizes the stereo chemical quality of the models, and their overall and residue-by-residue geometry.

The evaluation of models the WhatCheck tool, which did extensive checking of many stereochemical parameters of the residues in the models, showed that all models passed the overall summary rapport, confirming that the models were of satisfactory quality. In conclusion, Model_2O6L was of good quality, the others were of acceptable quality.^{17,12}

Figures 10-13 show the models superimposed on their UGT template for investigation of whether the CT domain of models resemble their template. The calculated RMSDs of the models' backbone C α (CT domain and the binding pocket) are shown in Table 4. The RMSD describes the degree of similarity between superimposed structures, and RMSD values are presented in Ångstrom (Å). Low RMSD values below 2 Å means the two structures are similar, while a value of 0 Å implicate that two structures are identical in conformation. The structures of the models were relatively conserved on the CT domain template, with RMSD values ranging from 0.157 Å to 1.759. The binding pockets were

more conserved, with RMSD values ranging from 0.149 Å to 0.840 Å.

Enzymes that are part of the GT family share a common co-factor binding site located in the CT domain of the protein. Several experimentally determined crystal structures have confirmed this, including the templates used in this study. Superimposing the structure of GT Oled in complex with UDP on the models indicated the putative binding site. An initial docking run of UDPGA in the binding pocket of Model_2O6L indicated by the superimposed UDP was performed. This gave the co-factor UDPGA an excellent pose and fit in the binding pocket, with a good docking score of -49, as shown in Figure 14.

A chemical table of 10 inhibitors and 145 decoys were docked in a semi-flexible mode into the putative binding site of UGT2B17, to evaluate the ability of the homology models to differentiate between them. The experimentally determined binding affinities of the inhibitors are shown in Table 2. The inhibitors were a diverse set of compounds, with different degree of inhibition of the target enzyme. The binding poses of the docked inhibitors were investigated, and the score values analyzed. Residues of importance for binding of UDPGA are shown as bold in Table 5. Most of the inhibitors were docked into the electronegative centrum of the pocket, as shown in Figure 15.

ROC curves were made to evaluate the ability of the models to



Figure 13. Model_2O6L superimposed on template. Model shown in yellow, template in grey. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Calculated root-meansquare-deviation for the models.

| Model | RMSD for the CT domain | RMSD for the binding pocket |
|-------|------------------------|-----------------------------|
| 3WAD | 1.264 Å | 0.445 Å |
| 4AMG | 1.759 Å | 0.840 Å |
| 4 M83 | 1.184 Å | 0.685 Å |
| 2O6L | 0.157 Å | 0.149 Å |

differentiate between binders (inhibitors) and non-binders (decoys). The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the model. A diagonal curve represents a model which is not able to discriminate between true positives and false positives. The calculated AUC for the ROC curves is a measure of the accuracy of the models, shown in percentage. Three parallel docking runs were performed for the results to be statistically viable. The ROC curves of all docking runs are shown in Figures 14-17, the calculated AUCs are shown in Table 6. No relationship was observed between docking scores of UGT2B17 inhibitors and their inhibitory activities when comparing docking scores and inhibitory activities for each of the compounds, but the ROC curves indicated that the models are able to discriminate between binders and non-binders. Fig 18.

Model_3WAD had ROC curves closer to the diagonal than the other, a mean AUC value of 73.93, indicating that this model was acceptable at differentiating true binders from decoys. Model_4M83 had a mean AUC value of 81.93. Model_2O6L had a mean AUC value of 82.61. These

Table 5
Residues forming the co-factor binding site in the models. Residues most likely to have contact with ligands are shown as bold.

| Model | Residues forming the co-factor binding site |
|-------|--|
| 3WAD | E32, Y33, S34 , H35, I37, N38, V278, G279, G280, L281, S309 , G311, S312, M313, R339 , K356, W357 , L358, P359, Q360 , N361, L364, F372, T374 , H375 , G376, G377, T378, N379 , G380 , I381, E383 , F397, A398, D399 , Q400 , N403 |
| 4AMG | E32, Y33, S34 , V278, K284, S309 , G311, S312, M313, R339 , K356, W357 , L358, P359, Q360 , N361, L364, F372, T374 , H375 , G376, G377, T378, N379 , G380 , I381, E383 , F397, A398, D399 , Q400 , N403 |
| 4 M83 | D88, F90, M93, H282, K284, S309 , G311, S312, M313, R339 , K356, W357 , L358, P359, Q360 , N361, L364, T374 , H375 , G376, G377, T378, N379 , G380 , I381, Y382, E383 , F397, A398, D399 , Q400 , N403 |
| 2O6L | S309 , G311, S312, M313, R339 , K356, W357 , L358, P359, Q360 , N361, L364, F372, T374 , H375 , G376, G377, T378, N379 , G380 , I381, E383 , F397, A398, D399 , Q400 , N403 |

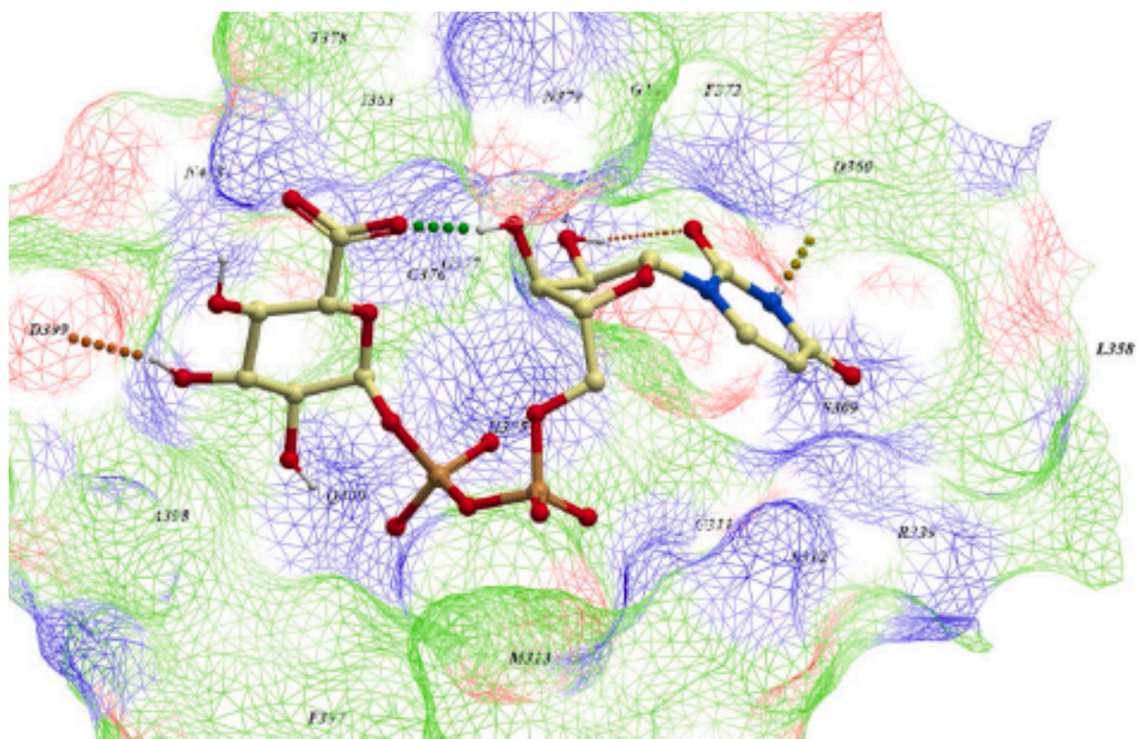


Figure 14. UDPGA docked into binding pocket of Model_2O6L. The surface of the pocket is coloured by the binding properties. Aromatic lipophilic shown as white. Aliphatic lipophilic shown as green, HBA potential shown as red, HBD potential shown as blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

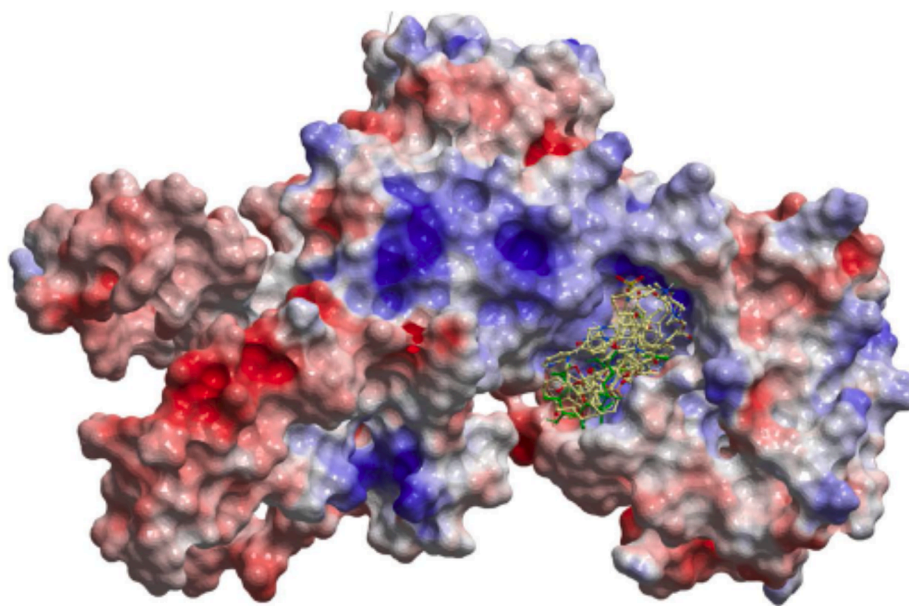


Figure 15. Model_4AMG shown as electrostatic potential. Areas coloured blue represent positive areas, red represents negative areas, and white represents neutral areas. The ten docked inhibitors shown in binding pocket. UDPGA added as reference, shown in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

models had higher AUCs, but had ROC curves crossing the diagonal line. Model_4AMG had ROC curves furthest away from the diagonal line curve and never crossing it. This model had the highest calculated AUC of 86.09, and mean AUC value was 84.68, which was the highest mean value of the models. In conclusion, Model_4AMG was the most accurate in distinguishing between true binders and non-binders. [Fig 19](#).

6. Discussion

The multiple sequence alignment with the templates used in the present study, combined with several other human UGTs, were about 130 residues shorter than the target. Furthermore, they had low sequence identity with target, leading to a difficult alignment process. The membrane interacting region in the NT domain added extra complexity, since the GT templates lacked this region. This resulted in several gaps in the alignments, and these gaps were shifted to the loop regions of the alignment if possible. A few site-directed mutagenesis studies were available mutagenesis,^{25,26,29,23} and these aided in the alignment process by highlighting regions of importance. The sequence identities for the adjusted alignments were 82% for UGT2B7, 20% for GT VinC, 22% for GT SnogD, and 21% for GT OleD, as shown in [Table 1](#).

The sequence identity between target and template strongly correlates with model accuracy, and three of the alignments had a low sequence identity. Because of the low sequence identity between GT templates and the target, a multi template model procedure was performed. After an initial construction of models based solely on their GT template, the partial structure of UGT2B7 was added as a second template for the CT domain, improving the quality of the models in this region. To be able to utilize homology models for VLS, a sequence identity above 60% would be preferred. By combining two templates, the overall sequence identity was raised to approximately 55% for the three full length models.

In Ramachandran plots, a good quality model is expected to have over 90% of the residues within the most favored regions. According to the Ramachandran plots, Model_2O6L was within the threshold of a good quality model with a percentage of 94.2%. The other models had 83.1% 86.0% and 87.5% respectively, which is just below the limit for good quality models. Of the residues making up the presumed co-factor binding pocket, none were in the disallowed regions, and only A398 was

in the generously allowed region.

The uncertainty in these models were most profound in the NT domains, due to gaps in the alignments and low sequence identity. This resulted in substantial structural differences, with some long secondary structures pointing out of the NT domain of the models. Model_4AMG had two long helices pointing outwards, while the other models had a few extra loop regions. The models superimposed on their corresponding GT templates are shown to the left in [Figures 10-13](#). Since the co-factor binding site built up by the CT domain, and some residues in the core of the NT domain, were the main areas of interest in this study, the uncertainty in the peripheral secondary structures of the NT domain were of less importance.

When the four models were superimposed, as shown in [Figure 20](#), the structural differences and similarities of the models became evident. The NT domain of the templates were highly variable, resulting in very different models. The CT domains were highly conserved in all the templates, and adding a second template for this region made the structural similarity even better.

Some of the residues from the site-directed mutagenesis studies were conserved in the models, as seen in the alignments. The two residues involved in the catalytic reaction of the enzyme (H35 and D152) were positioned close to the UDPGA binding site in the core of all the models, as needed to initiate the catalysis, as shown in [Figure 21](#). The residue F90 responsible for ring stacking interactions with the aglycone was also in close proximity in Model_4AMG and Model_4M83. These conserved residues at key positions increased the possibility that the built models were similar to the target in the core region of the enzymes, despite the relatively low sequence identity in the NT domain.

This UDPGA pose was superimposed on all models, and residues in a 5 Å vicinity to the superimposed UDPGA ligand were selected, thereby defining the binding pocket to be used in the docking process. Model_3WAD, Model_4AMG, Model_4M83 and Model_2O6L had 37, 32, 33 and 27 residues defining the binding pocket, respectively. Model_3WAD had the linker region between the two domains in close vicinity to the co-factor binding pocket, adding extra residues and narrowing the pocket. [Table 5](#) lists all residues forming the binding pocket of the models.

Comparing several crystal structures of GTs in complex with ligand (including GT OleD) with crystal structures without a ligand, have

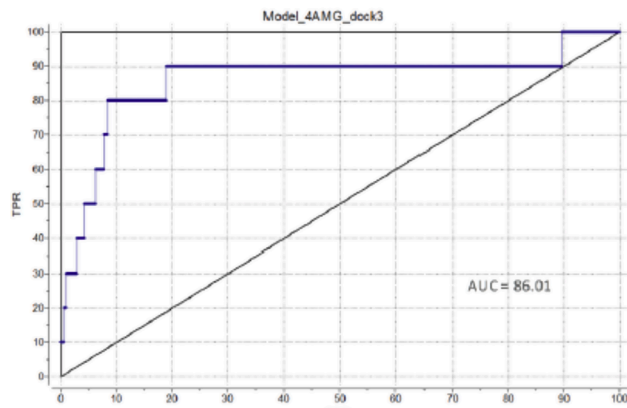
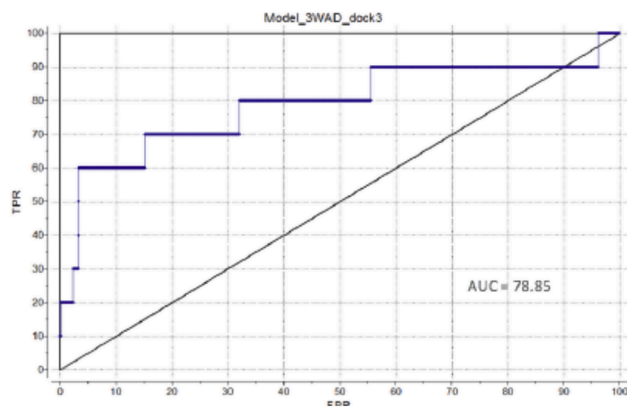
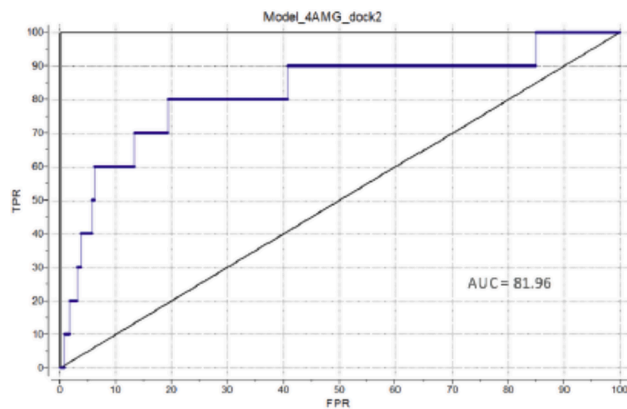
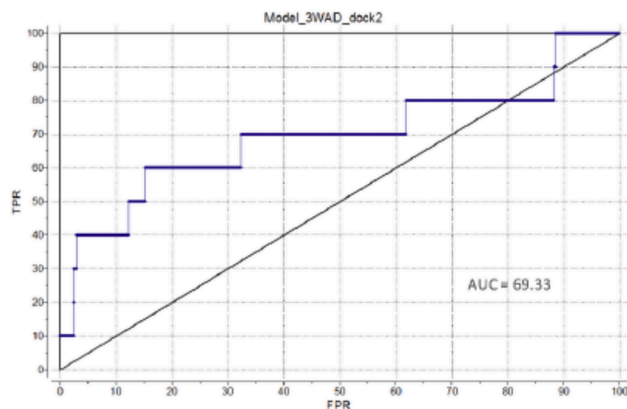
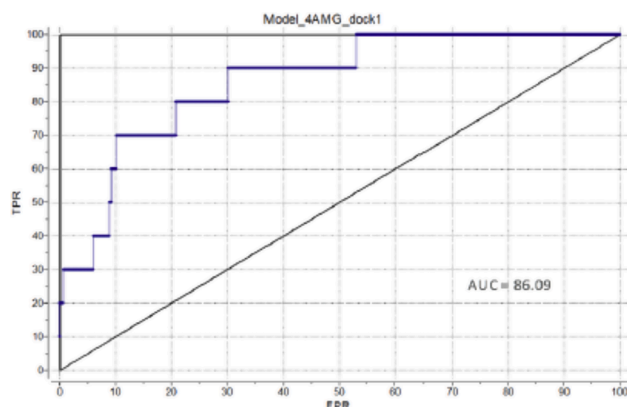
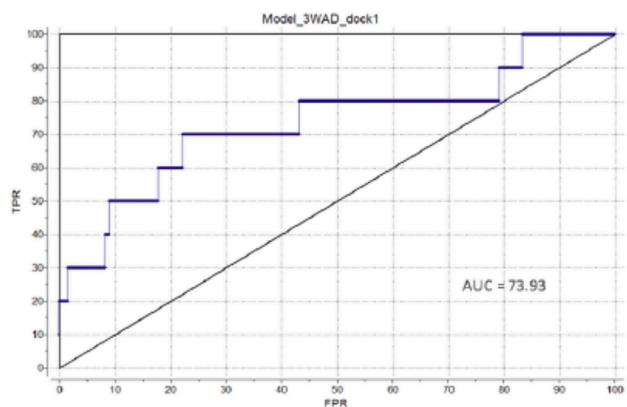


Figure 16. ROC curves for 3 parallel dockings on Model_3WAD. True positive rate on y-axis, false positive rate on x-axis.

shown a conformational change of W357, moving the residue closer to the ligand. The conformational change, presumably initiated by co-factor binding, could make aromatic ring stacking interactions possible between the aromatic ring of W357 and the uracil of UDPGA. This conformational change makes the residue important for ligand binding, despite its initial peripheral placement in the binding pocket^{24,20}.

Figure 22 shows Compound 2 and its putative binding interactions with the binding pocket. Hydrophobic interactions were observed with Y33, L63, H375, G377, T378, and N379. One of the aromatic rings of Compound 2 could participate in stacking interactions with Y33. Hydrogen bonds were observed to S34, S312, D399 and Q400, the distance being 2.9 Å, 2.4 Å, 3.2 Å and 2.5 Å respectively. There was also a possibility that hydrogen bonds could also be formed with residues Q360, T378 and N379.

The crystal structure of UGT2B7 shows the presence of water

Figure 17. ROC curves for 3 parallel dockings on Model_4AMG. True positive rate on y-axis, false positive rate on x-axis.

Table 6
Calculated AUC for all docking runs.

| Model | Docking parallel | AUC | mean AUC |
|-------|------------------|-------|----------|
| 3WAD | 1 | 73.61 | 73.93% |
| | 2 | 69.33 | |
| | 3 | 78.85 | |
| 4AMG | 1 | 86.09 | 84.68% |
| | 2 | 81.96 | |
| | 3 | 86.01 | |
| 4 M83 | 1 | 78.34 | 81.93% |
| | 2 | 84.59 | |
| | 3 | 82.88 | |
| 2O6L | 1 | 83.05 | 82.61% |
| | 2 | 81.59 | |
| | 3 | 83.29 | |

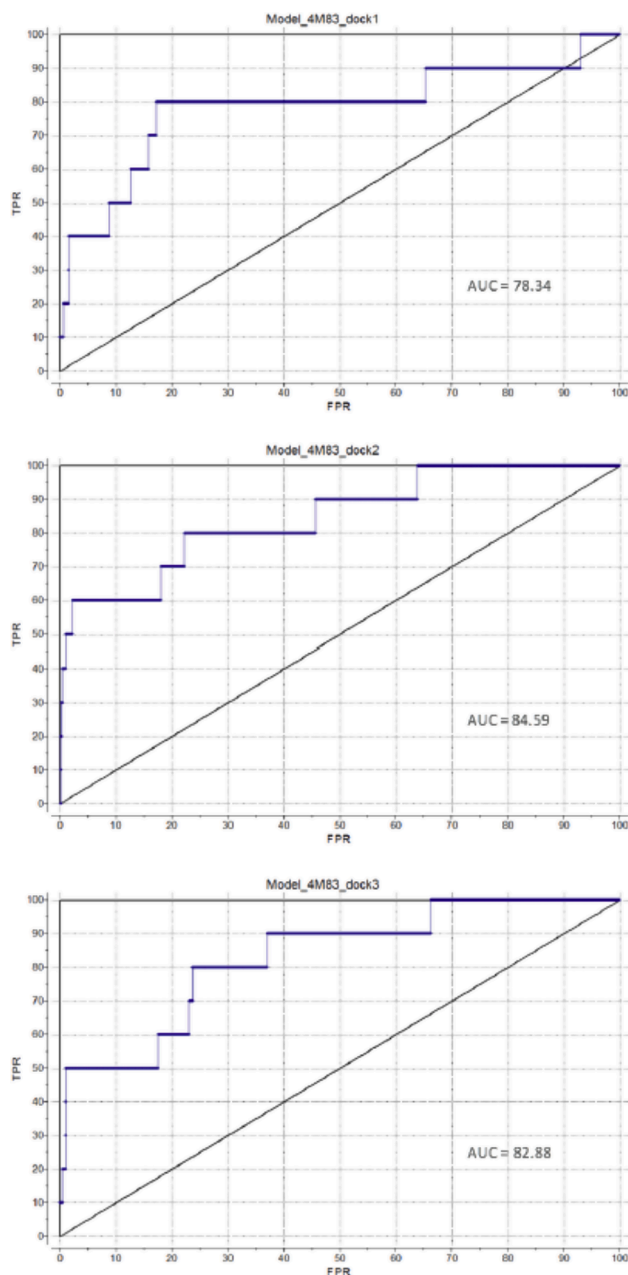


Figure 18. ROC curves for 3 parallel dockings on Model_4M83. True positive rate on y-axis, false positive rate on x-axis.

molecules in the binding pocket, suggesting that some of these could be involved in hydrogen bond interactions between ligands and the protein. The presence of water in the pocket was not accounted for in the docking process, but when superimposing the water molecules of the UGT2B7 template on Model_4AMG, water mediated interactions were possible with R339 and T374.

The ligand binding interactions described are not as static as the figures shows. In reality, both the enzyme and the ligand have natural structural flexibility and motion, making it easier to interact with residues in the binding pocket. The presumed conformational change upon ligand binding may also affect the binding pocket interactions. These figures are snapshots of how the ligand–protein complex could appear in reality.

Site directed mutagenesis studies were available for other UGTs.^{25,26,29,23} These confirmed that residues S34, S309, R339, W357, Q360, E373, T374, H375, N379, G380, E383, D399 and Q400,

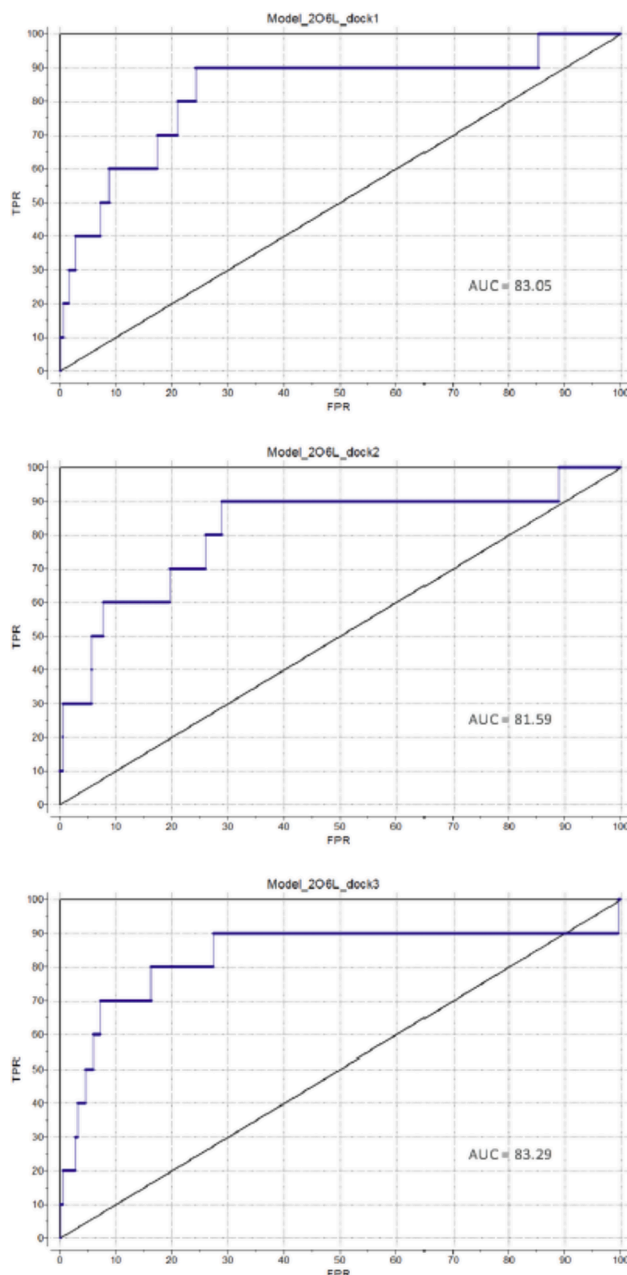


Figure 19. ROC curves for 3 parallel dockings on Model_2O6L. True positive rate on y-axis, false positive rate on x-axis.

positioned in the binding pocket, were involved in co-factor binding. Several of these residues were involved in binding interactions with compound 2, confirming the putative binding site in models. This study has shown that residues Y33, L63, S312, G377, and T378 may also be involved in ligand binding interactions. These residues along with the other NT domain residues listed in Table 5 could be interesting to study further in future site-directed mutagenesis studies. If experimental studies through crystal structures or site-directed mutagenesis studies could confirm these residues as important, it would be possible to conclude that the models were partially correct, and that the proposed residues were involved in ligand binding.^{25,20}

Model_4AMG was, according to the ROC curves, better at discriminating between binders from decoys than Model_2O6L, despite the lower sequence identity and RMSD of binding pocket. This indicates that the NT domain residues of Model_4AMG could be important for ligand binding.

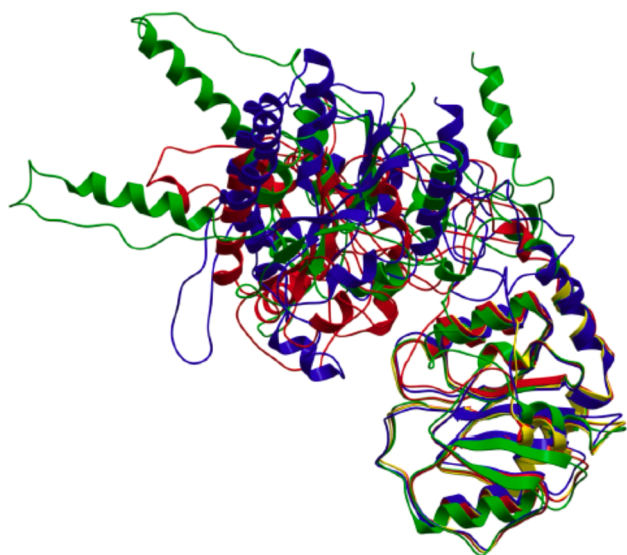


Figure 20. The four homology models superimposed. Model_3WAD shown as red, Model_4AMG shown as green, Model_4M83 shown as blue, and Model_2O6L shown as yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A poor value from the calculated AUC can indicate that a model is unable to distinguish between decoys and inhibitors. However, this does not necessarily imply that the model is inaccurate. In general, the ROC curve depends heavily on the choice of decoys. Decoys for the ROC curves were generated by the Decoyfinder software,⁸ since none were available from experimental data. These decoys are compounds with similar physiochemical properties and MW, presumed to be inactive. Without experimentally determined decoys, some of the theoretical decoys could actually be true binders, and may have generated false negatives, affecting the TPR in the curves.

In this study, homology models were constructed as working tools to aid in the design of experimental studies related to UGT2B17. Molecular docking of known ligands into the putative binding sites of UGT2B17

models was carried out to explore its overall predictability and aid in designing rational inhibitors. Molecular docking is a widespread approach for determining protein–ligand interactions, and it is used to predict potential ligand binding site of a target protein. Molecular studies on the binding sites of UGT2B17 may provide insightful information about key interactions and characteristics that may aid in designing new drugs.

Pharmacological UGT2B17 inhibition could provide novel treatment for patients with low testosterone levels. Use of currently available pharmacological testosterone replacement may have unwanted side effects. In a study of men aged 65 years or older, increased coronary artery plaque volume was found after 1 year of treatment.⁶ Inability to provide patients with a stable elevation of serum testosterone levels, could be related to adverse effects. Swerdloff et al. found that transdermal testosterone gel application gives a highly variable and unpredictable increase in serum concentrations.²⁷ Rather than un-physiological elevation of serum levels through direct treatment with testosterone,

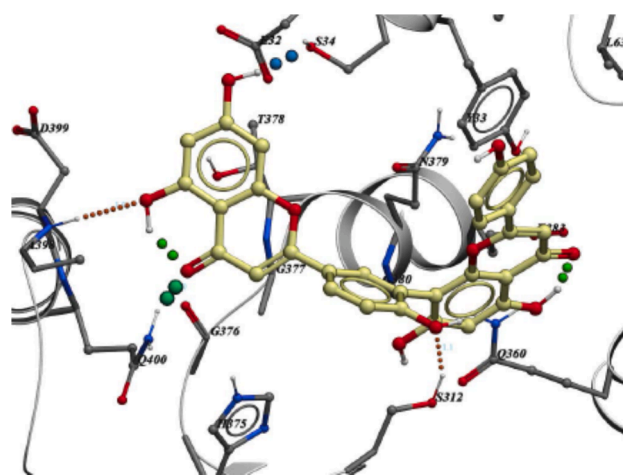


Figure 22. Presumed binding interactions between compound 2 and the binding pocket.

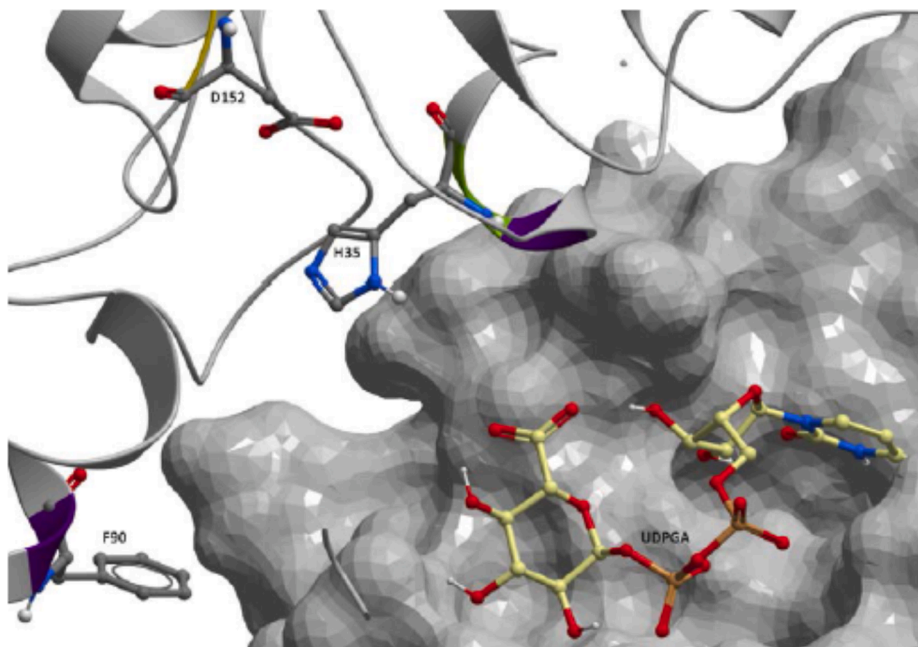


Figure 21. Residues H35 and D152 forming the catalytic dyad, and residue F90 crucial for interactions with the aglycone, all in close proximity to UDPGA situated in the binding pocket of Model_4AMG. NT domain shown as ribbon, CT domain shown as mesh.

UGT2B17 inhibition may elevate endogenous testosterone levels and provide a more predictable increase in serum concentrations. This could be favorable both for giving a predictable treatment regime with a stable testosterone substitution and reduce chance for serious adverse effects.

Coordinates are available upon request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [2] Abagyan R, Totrov M, Kuznetsov DN. ICM - a new method for protein modeling and design. Applications to docking and structure prediction from the distorted native conformation. *J Comp Chem*. 1994;15:488–506.
- [3] Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–3402.
- [4] Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res*. 1999;27(1):49–54.
- [5] Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235–242.
- [6] Budoff MJ, Ellenberg SS, Lewis CE, et al. Testosterone Treatment and Coronary Artery Plaque Volume in Older Men With Low Testosterone. *JAMA*. 2017;317(7):708–716.
- [7] Cavasotto CN, Orry AJ, Murgolo NJ, et al. Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J Med Chem*. 2008;51(3):581–588.
- [8] Cereto-Massague A, Guasch L, Valls C, Mulero M, Pujadas G, Garcia-Vallve S. DecoyFinder: an easy-to-use python GUI application for building target-specific decoy sets. *Bioinformatics*. 2012;28(12):1661–1662.
- [9] Claesson M, Siitonen V, Dobritzsch D, Metsa-Ketela M, Schneider G. Crystal structure of the glycosyltransferase SnogD from the biosynthetic pathway of nogalamycin in *Streptomyces nogalater*. *FEBS J*. 2012;279(17):3251–3263.
- [10] Gaulton, A., L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani and J. P. Overington (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 40(Database issue): D1100–1107.
- [11] Gauthier-Landry L, Belanger A, Barbier O. Multiple roles for UDP-glucuronosyltransferase (UGT)2B15 and UGT2B17 enzymes in androgen metabolism and prostate cancer evolution. *J Steroid Biochem Mol Biol*. 2015;145:187–192.
- [12] Hoofst RW, Vriend G, Sander C, Abola EE. Errors in protein structures. *Nature*. 1996;381(6580):272.
- [13] Katritch V, Byrd CM, Tseitin V, et al. Discovery of small molecule inhibitors of ubiquitin-like poxvirus proteinase 17L using homology modeling and covalent docking approaches. *J Comput Aided Mol Des*. 2007;21(10–11):549–558.
- [14] King CD, Rios GR, Green MD, Tephly TR. UDP-glucuronosyltransferases. *Curr Drug Metab*. 2000;1(2):143–161.
- [16] Kufareva I, Rueda M, Katritch V, Stevens RC, Abagyan R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure*. 2011;19(8):1108–1126.
- [17] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*. 1993;26:283–291.
- [18] Maiorov VN, Crippen GM. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J Mol Biol*. 1994;235(2):625–634.
- [19] Meech R, Mackenzie PI. Structure and function of uridine diphosphate glucuronosyltransferases. *Clin Exp Pharmacol Physiol*. 1997;24(12):907–915.
- [20] Miley MJ, Zielinska AK, Keenan JE, Bratton SM, Radominska-Pandya A, Redinbo MR. Crystal structure of the cofactor-binding domain of the human phase II drug-metabolism enzyme UDP-glucuronosyltransferase 2B7. *J Mol Biol*. 2007;369(2):498–511.
- [21] Nango E, Minami A, Kumasaka T, Eguchi T. Crystallization and preliminary X-ray analysis of vicenismaminyltransferase VinC. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2008;64(Pt 6):558–560.
- [22] Orry AJW, A. R. (2012). *Homology Modeling : Methods and Protocols*. Humana Press.
- [23] Patana AS, Kurkela M, Finel M, Goldman A. Mutation analysis in UGT1A9 suggests a relationship between substrate and catalytic residues in UDP-glucuronosyltransferases. *Protein Eng Des Sel*. 2008;21(9):537–543.
- [24] Qasba PK, Ramakrishnan B, Boeggeman E. Substrate-induced conformational changes in glycosyltransferases. *Trends Biochem Sci*. 2005;30(1):53–62.
- [25] Radominska-Pandya A, Czernik PJ, Little JM, Battaglia E, Mackenzie PI. Structural and functional studies of UDP-glucuronosyltransferases. *Drug Metab Rev*. 1999;31(4):817–899.
- [26] Senay C, Ouzzine M, Battaglia E, et al. Arginine 52 and histidine 54 located in a conserved amino-terminal hydrophobic region (LX2-R52-G-H54-X3-V-L) are important amino acids for the functional and structural integrity of the human liver UDP-glucuronosyltransferase UGT1*6. *Mol Pharmacol*. 1997;51(3):406–413.
- [27] Swerdloff RS, Pak Y, Wang C, et al. Serum Testosterone (T) Level Variability in T Gel-Treated Older Hypogonadal Men: Treatment Monitoring Implications. *J Clin Endocrinol Metab*. 2015;100(9):3280–3287.
- [28] Wang, F., Helmich, K.E., Xu, W., Singh, S., Olmos Jr., J.L., Martinez iii, E., Bingman, C.A., Thorson, J.S., Phillips Jr., G.N. (2013). Ensemble refinement of protein crystal structure (2IYF) of macrolide glycosyltransferases OleD complexed with UDP and Erythromycin A. *Enzyme Discovery for Natural Product Biosynthesis (NatPro)*.
- [29] Xiong Y, Bernard D, Bratton S, et al. Phenylalanine 90 and 93 are localized within the phenol binding site of human UDP-glucuronosyltransferase 1A10 as determined by photoaffinity labeling, mass spectrometry, and site-directed mutagenesis. *Biochemistry*. 2006;45(7):2322–2332.
- [30] Yasaman Aghazadeh BRZ, Papadopoulos Vassilios. Chapter Seven - Pharmacological Regulation of the Cholesterol Transport Machinery in Steroidogenic Cells of the Testis. *Vitam Horm*. 2015;98:189–227.