

# A computational study of the substrate conversion and selective inhibition of aldosterone synthase

**Citation for published version (APA):**

Roumen, L. (2008). *A computational study of the substrate conversion and selective inhibition of aldosterone synthase*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Biomedical Engineering]. Technische Universiteit Eindhoven. <https://doi.org/10.6100/IR637195>

**DOI:**

[10.6100/IR637195](https://doi.org/10.6100/IR637195)

**Document status and date:**

Published: 01/01/2008

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

# A Computational Study of the Substrate Conversion and Selective Inhibition of Aldosterone Synthase

The cover illustrates my eclectic interpretation of this research project.

A catalogue record is available from the Eindhoven University of Technology Library.

ISBN: 978-90-386-1365-9

Copyright ©2008 by L. Roumen

All rights reserved. No part of this book may be reproduced, stored in a database or retrieval system, or published, in any form or in any way, electronically, mechanically, by print, photoprint, microfilm or any other means without prior written permission of the author.

Cover design: K. Pieterse, L. Roumen

Printed by PrintPartners Ipskamp, Enschede, The Netherlands

This research was financially supported by the Dutch Technology Foundation STW, applied science division of NOW and the Technology Program of the Ministry of Economic Affairs. Grant Number MFA 6504.

# A Computational Study of the Substrate Conversion and Selective Inhibition of Aldosterone Synthase

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de  
Technische Universiteit Eindhoven, op gezag van de  
Rector Magnificus, prof.dr.ir. C.J. van Duijn, voor een  
commissie aangewezen door het College voor  
Promoties in het openbaar te verdedigen  
op woensdag 1 oktober 2008 om 16.00 uur

door

Luc Roumen

geboren te Geldrop

Dit proefschrift is goedgekeurd door de promotor:

prof.dr. P.A.J. Hilbers

Copromotor:

dr. J.J.R.M. Hermans

## Table of Contents

Chapter 1 Aldosterone Synthase inhibitors, a new treatment option for heart failure?.....	1
1.1 Heart Failure.....	2
1.2 The Renin Angiotensin Aldosterone System.....	2
1.3 Heart failure treatment.....	3
1.4 Aldosterone Biosynthesis.....	4
1.5 Starting Structures.....	5
1.6 Aim and Scope of this Thesis.....	7
Chapter 2 Homology Modelling.....	11
2.1 Introduction.....	12
2.2 Cytochrome P450 enzymes.....	12
2.2.1 Nomenclature and generic function.....	12
2.2.2 Structural Architecture.....	13
2.3 Homology Modelling.....	17
2.4 Homology Models for CYP11B1 and CYP11B2.....	19
2.4.1 Modelling Criteria.....	20
2.4.2 Multiple Sequence Alignment.....	21
2.4.3 Template Selection.....	24
2.4.4 Template Construction.....	26
2.4.5 Additional Modelling Criteria: Point Mutants.....	27
2.5 Hypothesis: Steric Aspects Play an Important Role in Substrate Conversion.....	28
2.6 Construction of CYP11B Models.....	29
2.7 Results and Discussion.....	30
2.7.1 Model quality assessment.....	31
2.7.2 CYP11B1 and CYP11B2 model active site differences.....	36
2.7.3 Protein - Substrate interactions.....	36
2.7.3.1 DOC.....	39
2.7.3.2 18OH-DOC.....	39
2.7.3.3 B and 18OH-B.....	40
2.7.4 Triple-Mutant Influence.....	40
2.7.5 Proposed Synthesis Mechanism of Aldosterone.....	40
2.8 Conclusions.....	41
Chapter 3 Molecular Dynamics.....	47
3.1 Molecular Dynamics.....	48
3.2 Force fields.....	48
3.3 Molecular Dynamics of hCYP11B1 and hCYP11B2.....	49
3.4 Molecular Dynamics Settings.....	50
3.5 Results and Discussion.....	51
3.5.1 Protein Structure Stability.....	51
3.5.2 Protein-Ligand Interactions.....	53
3.6 Conclusions.....	58

Chapter 4 Molecular Docking .....	61
4.1 Molecular Docking .....	62
4.1.1 GOLD .....	63
4.1.2 Scoring Functions .....	63
4.2 Prediction of Binding Affinity.....	64
4.3 Molecular Docking in the CYP11B Models.....	65
4.4 Docking Settings.....	66
4.5 Statistical Evaluation of Docking Results .....	67
4.6 Results and Discussion .....	68
4.6.1 Substrate Docking.....	68
4.6.2 Docking of the Four Known Inhibitors .....	69
4.6.3 Docking of Fadrazole Analogues .....	70
4.6.3.1 Comparison of scoring functions and protein models.....	71
4.6.3.2 Prediction models for non-Rb-substituted fadrazole analogues .....	73
4.6.3.3 Prediction of novel inhibitors.....	75
4.7 Conclusions .....	80
Chapter 5 Structure Activity Relationships .....	83
5.1 Quantitative Structure Activity Relationships.....	84
5.1.1 QSAR Challenges .....	85
5.1.2 QSAR Model Construction and Cross-validation .....	86
5.1.3 Receiver-Operating-Curves .....	87
5.2 Problems Encountered By Performing QSAR on the Fadrazole Dataset .....	88
5.2.1 Initial Trends .....	89
5.2.2 QSAR Results: Why did we opt for Decision Tree Analysis? .....	90
5.3 Decision Tree Analysis .....	92
5.3.1 Fadrazole analogue main set.....	93
5.3.2 Substituent descriptors .....	95
5.3.3 Results and Discussion.....	96
5.3.3.1 Prediction of CYP11B1 and CYP11B2 potency for novel compounds .....	97
5.4 Conclusions .....	102
Chapter 6 Quantum Mechanics Calculations .....	105
6.1 Introduction .....	106
6.2 Quantum Mechanics.....	107
6.2.1 Quantum Mechanics Calculation Methods .....	107
6.2.2 Basis Sets .....	108
6.3 Catalytic cycle of cytochrome P450 enzymes .....	109
6.3.1 Oxygen Rebound in detail.....	110
6.3.2 Uncoupling .....	113
6.4 Mechanistic Knowledge.....	114
6.4.1 Steroid Conformation Generation .....	116
6.4.2 Results and Discussion: Mechanistic Knowledge .....	118
6.4.2.1 DOC Conformations .....	118
6.4.2.2 B Conformations .....	121
6.4.2.3 18OH-DOC Conformations.....	121
6.4.2.4 18OH-B Conformations .....	121
6.4.2.5 18diOH-B Conformations .....	122

6.4.3 Fukui index analysis.....	124
6.4.4 Fukui index analysis of sequential paths .....	124
6.4.4.1 Path 1 .....	125
6.4.4.2 Path 2 .....	126
6.4.4.3 Path 3 .....	126
6.4.4.4 Path 4 .....	126
6.4.4.5 Path 5 .....	126
6.4.4.6 Path 6 .....	127
6.4.5 Structural investigation of the hydroxylation paths.....	127
6.4.6 18OH-DOC as an in vitro inhibitor for CYP11B1 and CYP11B2.....	127
6.4.7 Study Limitations: Mechanistic Knowledge.....	128
6.5 Transition state analogues of the 18-hydroxycorticosterone conversion .....	129
6.5.1 Reduction of Heme Complexity .....	129
6.5.2 Reduction of Substrate Complexity.....	130
6.6 Results and Discussion: Transition state analogues.....	131
6.7 Conclusions: Transition State Analogues and Mechanistic Knowledge.....	134
Chapter 7 Concluding Remarks .....	141
Appendix A: Known inhibitors of the CYP11B family.....	147
Appendix B: Chemical Structures of Fadrazole Analogues.....	157
Appendix C.1.1: CYP11B1 Docking Results using GoldScore.....	161
Appendix C.1.2: CYP11B1 Docking Results using ChemScore.....	163
Appendix C.2.1: CYP11B2 Docking Results using GoldScore.....	165
Appendix C.2.2: CYP11B2 Docking Results using ChemScore.....	167
Appendix D: Decision Tree Analysis Results .....	169
Summary.....	171
Samenvatting .....	173
Dankwoord .....	175
Publications.....	177
Curriculum Vitae.....	179





## Chapter 1 Aldosterone Synthase inhibitors, a new treatment option for heart failure?

The work described in this PhD thesis was conducted in the setting of a multi-disciplinary STW project (MFA 6504, later renumbered as 06504). The title of the project is "Aldosterone Synthase inhibitors, a new treatment option for heart failure?". In order to develop specific inhibitors for aldosterone synthase and screen for their therapeutic feasibility, a research strategy has been planned in which different disciplines are used in an integrated way, combining molecular modelling with pharmacological experiments (Figure 1–1).

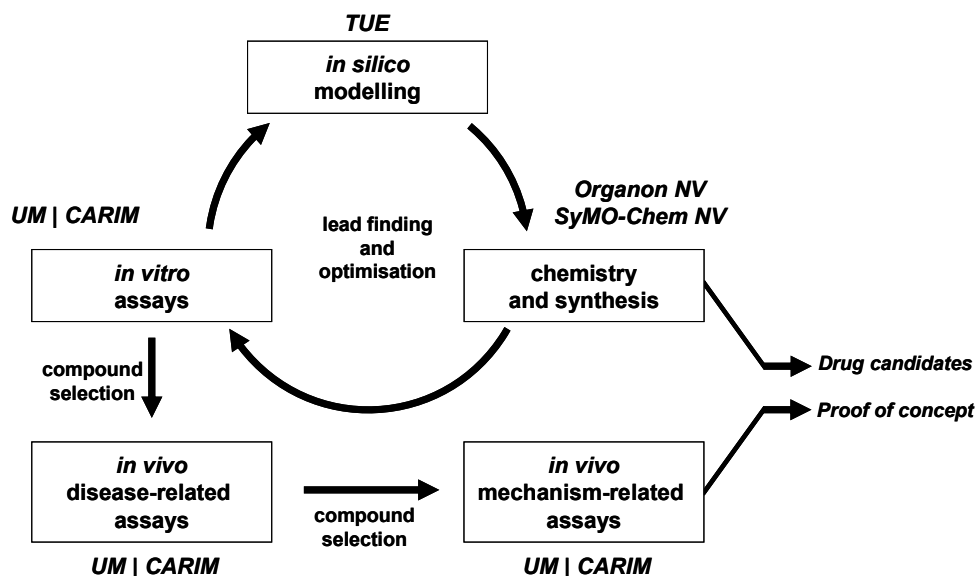


Figure 1–1 STW Project Setup. UM/CARIM: Maastricht University, Cardiovascular Research Institute Maastricht. TUE, Eindhoven University of Technology

The lead finding and lead optimisation strategy applied in the project encompasses the use of *in vitro* assays, *in silico* modelling and compound synthesis, resulting in the identification of drug candidates. Molecular modelling serves to elucidate protein-ligand interactions and subsequent prioritisation of novel compound synthesis. Finally, the most promising drug candidates are relayed to *in vivo* assays to determine the therapeutic application as aldosterone synthase inhibitors.

The participating project members are stationed in Maastricht (*in vitro* and *in vivo* measurements), Oss (chemistry and synthesis) and Eindhoven (molecular modelling, chemistry and synthesis). Because of our collaboration we have aptly chosen our compound acronym to be "Moeras", which means "Maastricht Oss and Eindhoven Reduce Aldosterone Synthesis".

### 1.1 Heart Failure

The term heart failure or sometimes congestive heart failure is commonly conceived to be the cessation of heartbeat (asystole) or the cessation of normal heart function that is followed by a collapse in blood flow and sudden death (cardiac arrest). These are, however, unfortunate misconceptions. Heart failure is neither of those afflictions; rather, it is a condition that can result from any functional or structural cardiac disorder that impairs the ability of the heart to provide the required blood flow to the body. This means that heart failure is a process where the heart is severely remodelled into a weakened myocardial structure. Remodelling evokes abnormalities in cardiac structure, rhythm, function, or conduction, whereas these abnormalities may also be the trigger for (further) remodelling.

The major causes of heart remodelling eventually resulting in chronic heart failure are myocardial infarction and hypertension. Since progression of heart remodelling is paralleled by modified functioning of other organs important in maintaining cardiovascular homeostasis, it is often portrayed as a vicious circle. Indeed, heart failure and damage to other organs such as kidneys or blood vessels often occur in parallel [1,2,3].

### 1.2 The Renin Angiotensin Aldosterone System

Currently, many drugs that are used for the treatment of heart failure target the Renin Angiotensin Aldosterone System (RAAS) (Figure 1–2). This physiological system is responsible for the regulation of electrolyte homeostasis and blood pressure. It is activated when the arterial pressure is decreased, the renal blood flow is decreased or in case of reduced plasma sodium chloride levels or enhanced sympathetic nervous activity. Once activated, the kidneys secrete the enzyme renin to the bloodstream. Renin cleaves the plasma protein angiotensinogen into a 10 amino acid peptide angiotensin I (DRVYIHPFHL), which is subsequently cleaved by angiotensin-converting enzyme (ACE) or chymase into the 8 amino acid peptide angiotensin II (DRVYIHPF). When angiotensin II is formed it binds to the angiotensin II type 1 receptor (AT1) and the angiotensin II type 2 receptor (AT2). The response of activation of both receptor types is generally opposite. While the response of the AT1 receptor is related to an increase of vasoconstriction, norepinephrine release, heart contractility, water retention and conservation, ventricular hypertrophy, myocardial fibrosis and aldosterone synthesis, the response of the AT2 receptor is related to vasodilatation, decreased norepinephrine levels and decreased myocardial fibrosis [4,5]. The opposing receptor activities seem to be important for the tuning of cardiac functioning. Finally, the last member of the cascade, the steroid aldosterone, binds to the mineralocorticoid receptor and amplifies some of the actions of angiotensin II. It can induce sodium and water retention, potassium excretion, and it can also increase sympathetic activation. The different ways of increasing blood volume, vascular resistance and increasing heart rate serve to maintain arterial pressure [6].

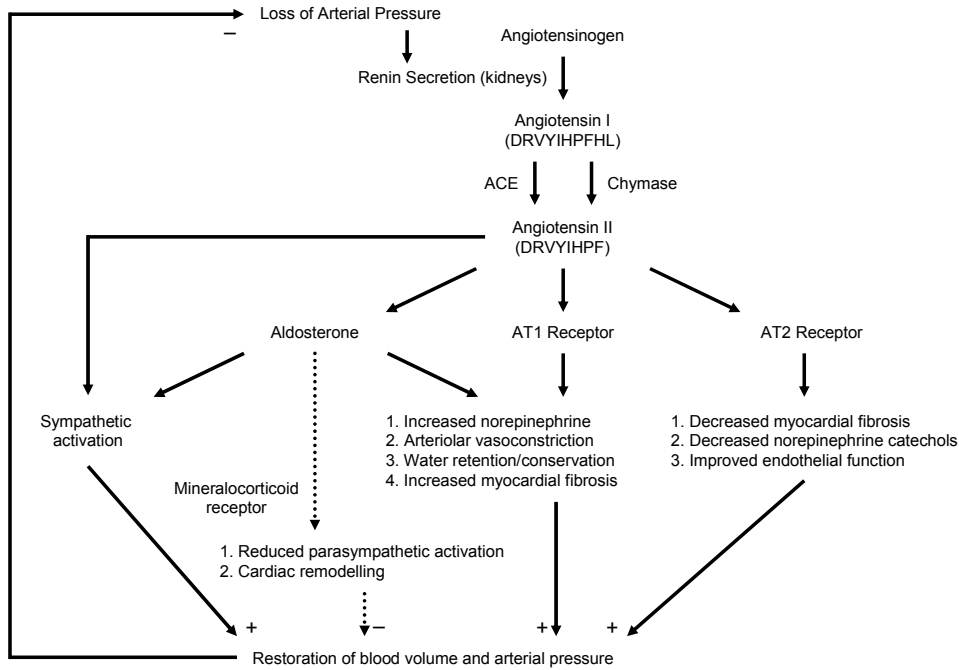


Figure 1–2 The Renin Angiotensin Aldosterone System (adapted from references 4, 16 and 17).

### 1.3 Heart failure treatment

Problems arise when the activation of angiotensin II is too large and the AT1 receptor activation takes the upper hand. To block the excessive actions of angiotensin II, the first medicines were designed to stop its synthesis with ACE blockers (CONSENSUS [7], SOLVD [8]). A great success has been achieved by treating patients with these ACE blockers, yet due to sustained activity of the ACE independent formation of angiotensin II, the effectiveness of the drug decreases in time. Indeed, 60% to 80% of the formation of angiotensin II seems to be independent from ACE and a close to full blockade of the RAAS requires both the inhibition of ACE and chymase [9,10,11]. The success and the apparent limitations of ACE blockers led to the use of angiotensin receptor blockers (ARBs) that selectively inhibit the AT1 receptors to shift the balance towards the AT2 receptor pathway. Unexpectedly, several studies have indicated that the effect of ARBs on the treatment of heart failure is equal to that of ACE inhibitors (ELITE I [12], ELITE II [13], RESOLVD [14]). Recently, various studies on the pathophysiology of heart failure have revealed that aldosterone plays an important role in the formation of myocardial hypertrophy, reactive myocardial fibrosis, vascular remodelling and electrolyte imbalance [15,16]. Furthermore, it has become apparent that increased levels of aldosterone can block myocardial norepinephrine uptake and reduce baroreceptor discharge, contributing to the development of arrhythmias [17,18,19,20]. It has been shown that blocking the action of aldosterone using mineralocorticoid receptor antagonists greatly reduces mortality and hospitalisation numbers

in patients with severe heart failure as well as patients that have suffered from a myocardial infarction even in situations that the RAAS is blocked (RALES [21], EPHEsus [22,23]). Nevertheless, there are still disadvantages for using mineralocorticoid receptor antagonists. Several side-effects exist and patients may possess interindividual variations in their response to mineralocorticoid receptor antagonists regarding pharmacodynamics and pharmacokinetics [21,24]. Furthermore, mineralocorticoid receptor antagonists induce compensatory aldosterone synthesis of which long term effects are unknown [25,26]. Finally, aldosterone is known to exert not only genomic but also rapid non-genomic effects that may play a role in the pathophysiology of heart failure, but that are not necessarily mediated by the mineralocorticoid receptor, and hence may not be blocked by mineralocorticoid receptor antagonists.

In our project we have investigated an alternative way to reduce aldosterone action, namely by prevention of aldosterone formation by intervening with the final steps of its biosynthesis [27]. The challenge in this approach lies in obtaining selectivity. The biosynthesis of aldosterone involves two very highly homologous cytochrome P450 enzymes that also possess overlapping substrate and product selectivities.

#### 1.4 Aldosterone Biosynthesis

The last steps in the biosynthesis of aldosterone are mediated by the mitochondrial cytochrome P450 11B family (Figure 1–3). The members of this protein family contain a heme prosthetic group in the core of the active site with which they catalyse (subsequent) oxidation reactions on C<sub>11</sub>, C<sub>18</sub> and C<sub>19</sub> on the  $\beta$ -side of the steroid skeleton. The enumeration of the steroid skeleton is also shown in Figure 1–3. In bovine [28], pig [29] and frog [30], aldosterone synthesis is performed by only one enzyme, CYP11B, but in man [31] and mouse [32] the synthesis involves two isoforms, CYP11B1 (cortisol synthase) and CYP11B2 (aldosterone synthase). Rat possesses four isoforms of which CYP11B1 is responsible for the bulk of corticosterone production (the main glucocorticoid in rodents as they lack 17-hydroxylase activity and hence do not produce cortisol), and CYP11B2 is responsible for aldosterone synthesis. CYP11B3 is only expressed in neonatal rat and carries the same activity as CYP11B2, and CYP11B4 encodes a pseudo gene [33].

The substrate specificity of the different CYP11B isoforms is nearly identical, but there are particular differences. In both human and rat, only the CYP11B2 isoform can perform the final oxidation of C<sub>18</sub> to produce the aldehyde aldosterone [31,33]. For the CYP11B1 isoform, the hydroxylation of C<sub>19</sub> has been reported for rat [33], and the CYP11B1 isoform in general is known to play an important role in the biosynthesis of glucocorticoids (Figure 1–3). It is not surprising that the C<sub>19</sub> can be oxidised by rat CYP11B1 because structurally, it is in close proximity to C<sub>11</sub> and C<sub>18</sub>. Other carbon atoms in close proximity to C<sub>11</sub> that might be oxidised on the  $\beta$ -side of the steroid skeleton are C<sub>1</sub>, C<sub>8</sub> and C<sub>12</sub>, but thus far no oxidation of these atoms by the CYP11B family has been reported. We have utilised this stereo- and regio-selective substrate hydroxylation to characterise the structural differences between CYP11B1 and CYP11B2 that is required for the design of selective inhibitors (as will be elaborated in Chapter 2). In addition, we have investigated the substrate conversion mechanism in Chapter 6.

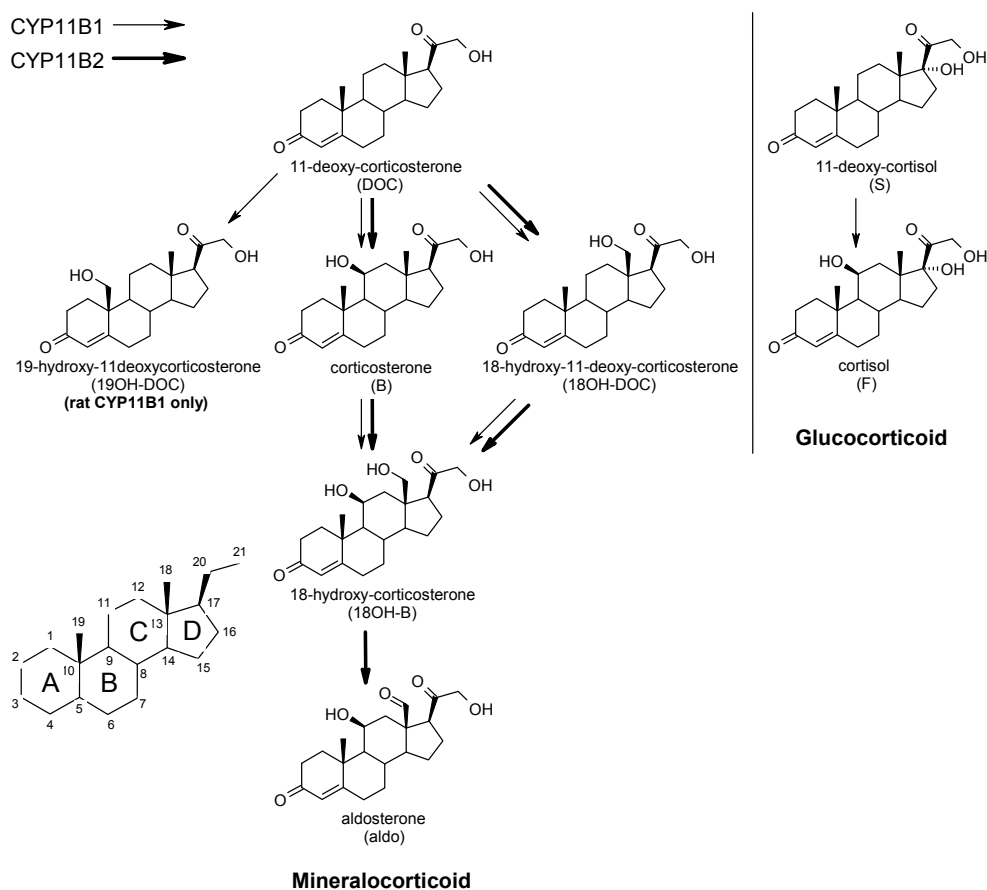


Figure 1–3 Biosynthesis of mineralocorticoids and glucocorticoid by the CYP11B family. Indicated with arrows are the possible substrate conversions performed by human CYP11B1 (thin) and CYP11B2 (thick) [31]. Rat CYP11B1 can CYP11B2 possess the same activities as the human isoforms, except that rat CYP11B1 can also oxidise 11-deoxycorticosterone on C<sub>19</sub> [33]. In rat, the glucocorticoids are not present. Instead, the primary glucocorticoid is corticosterone.

### 1.5 Starting Structures

At the beginning of the project, a preliminary literature search was performed to identify several chemical substances that possess inhibitory action on either CYP11B1 or CYP11B2. These known inhibitors of the cytochrome P450 11B family have provided an insight of the general structural features required for the design of novel inhibitors (Appendix A). The structures can be divided into three generic categories; (1) steroidal compounds that share substructure features of the CYP11B substrates, (2) heterocyclic compounds that were designed for the inhibition of other cytochrome P450 enzymes, and (3) other drugs.

Steroidal inhibitors could provide a good structural basis for CYP inhibition, although they have the tendency to cause undesirable side-effects. Steroids display biological effects by

binding to nuclear receptors that are involved in numerous metabolic, developmental and homeostatic processes, and designing a steroid selective to only CYP11B2 is expected to be a difficult task. Finally, because of their lipophilic nature, steroidal substances will easily penetrate into tissues and under certain conditions may accumulate, making the theoretical possibility for side effects even greater.

Several of the known inhibitors have been designed for the inhibition of cytochrome P450 enzymes related to CYP11B1 and CYP11B2, in particular aromatase. Aromatase, or CYP19, oxidises its substrates on the C<sub>19</sub> position of the steroid and converts the A-ring into an aromatic benzene ring. Since rat CYP11B1 can oxidise 11-deoxycorticosterone on the C<sub>19</sub> position, and both these hydroxylation sites are in very close proximity to one another, it can be anticipated that the CYP11B enzymes share some of the active site features with aromatase. The structural characteristic of the known inhibitors that define them as CYP inhibitors is the presence of a heterocyclic nitrogen atom (Appendix A). In particular, the accessibility of its electron lone pair allows the compound to form a strong complex with the heme iron atom which is a common feature of many non-steroidal CYP inhibitors [34,35].

Based on these considerations and the knowledge that fadrazole (Figure 1–4) decreases *in vivo* corticosteroid levels (in particular aldosterone levels) [36,37], we have chosen fadrazole as our lead structure. Fadrazole is a chiral compound. It is potent aromatase inhibitor that also possesses inhibition for CYP11B1 and CYP11B2 in the nanomolar range [38,39]. In an initial modelling study, we constructed homology models for CYP11B1 and CYP11B2 that predicted a stereoselectivity for fadrazole. After enantiomer separation and *in vitro* testing, we found that just as predicted, the *R*-enantiomer possesses CYP11B2 selectivity and the *S*-enantiomer possesses CYP11B1 selectivity. During the progression of the project, many modifications on the structure have been evaluated *in silico* and subsequently synthesised and tested, resulting in the lead structure 2 (Figure 1–4).

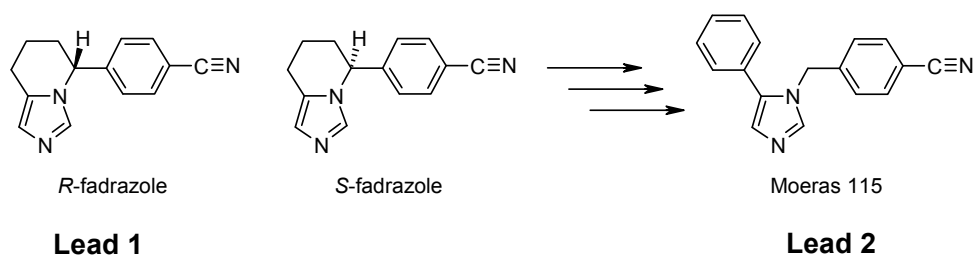


Figure 1–4 Lead Structures used throughout the progression of the project

## 1.6 Aim and Scope of this Thesis

Computer-aided molecular design (CAMD) is one of the essential tools for drug discovery. It refers to a collection of *in silico* molecular modelling techniques that are used to study molecular structures and properties of drug candidates as well as drug targets for the discovery and design of new drugs. The molecular modelling techniques allow rapid investigation and detailed information gathering of intramolecular and intermolecular interactions of molecular systems as well as the identification of the three dimensional characteristics of drug functional groups that are important for activity. The application of CAMD at the early stages of drug development can be vital for the creation of new drugs. Guided by insights obtained from the structural assessment of either target structures or drug lead structures, large numbers of compounds can be sampled with user-defined selection criteria to provide a rationale for compound chemistry and drug optimisation. In conjunction with compound synthesis and *in vitro* potency measurements, the *in silico* selection criteria can be refined for further modifications and optimisations of the lead structure.

In this thesis we have applied some commonly used *in silico* tools to provide a solid base on which to predict ligand potency and to steer compound synthesis. In this chapter we have focused on the rationale of our approach and the introduction of our drug target, the cytochrome P450 enzymes belonging to the 11B family. In the following three chapters we have utilised tools that are commonly applied to protein-based drug design, a field of expertise that requires knowledge of the three-dimensional properties of the target protein.

**Chapter 2** introduces the enzymatic activity and structural characteristics of the cytochrome P450 family. Using these properties we have constructed three dimensional models for CYP11B1 and CYP11B2, detailing the most conserved structural properties of cytochrome P450 enzymes. Additional models were constructed for the rat isoforms as well as a mutant protein for which the substrate conversion activity has been determined. In this step we discuss homology modelling difficulties and model quality assessment.

In **Chapter 3** we have conducted a molecular dynamics study on the structural integrity of the protein models. During this stage, the interaction of four of the known CYP11B inhibitors was evaluated in the protein active sites. The results have been used to determine the most important interactions of both substrates and ligands in the active sites.

Molecular docking is a method to quickly evaluate the protein-ligand interactions for multiple ligands. In **Chapter 4** we have used molecular docking to obtain information on the binding mode of our lead structure and derived analogues inside the active site of CYP11B1 and CYP11B2. Subsequently, we expanded the docking study with the different protein states that were sampled during the molecular dynamics study. By incorporating these active site conformations in our analysis we examined the influence of active site changes on the performance of the docking program. In addition, the *in silico* predicted potencies have been correlated with *in vitro* measured potencies for the application of virtual screening of new analogues.

After these protein-based approaches for medicine design, we evaluated a ligand-based approach called "decision tree analysis" in **Chapter 5**. For this method, only the physico-



chemical properties of different substructure components are considered. Subsequently, Boolean decisions based on the values of the physico-chemical properties describe the required or prohibited features of a potent inhibitor.

The topic of **Chapter 6** has a different basis than inhibitor design, although the underlying idea behind the work in this chapter still was the design of an alternative novel type of CYP11B2 inhibitors. In parallel to the previous studies, we have performed a mechanistic study on the conversion of steroids by cytochrome P450 enzymes. Using quantum mechanics calculations we have attempted to rationalise the precise steps taken by CYP11B2 during the production of aldosterone, as well as rationalise the required active site interactions that determine the regio-selective conversion profile of the CYP11B isoforms.

The thesis ends with concluding remarks in **Chapter 7**. In this chapter, the general findings and conclusions derived by the different molecular modelling methods are compared and summarised, and recommendations for future research are proposed.

## Literature

- 1 J.J.V. McMurray, M.A. Pfeffer, "Heart failure", *Lancet*, 2005, 365, 9474, 1877-1889
- 2 T.J. Wang, D. Levy, E.J. Benjamin, R.S. Vasan, "The epidemiology of "asymptomatic" left ventricular systolic dysfunction: implications for screening", *Ann Intern Med*, 2003, 138, 907-916
- 3 K. Hogg, K. Swedberg, J. McMurray, "Heart failure with preserved left ventricular systolic function", *J Am Coll Cardiol*, 2004, 43, 3, 317-327
- 4 M.J.Eisenberg, L.C. Gioia, "Angiotensin II receptor blockers in congestive heart failure", *Cardiol Rev*, 2006, 14, 1, 26-34
- 5 T.L. Goodfriend, M.E. Elliott, K.J. Catt, "Angiotensin receptors and their antagonists", *N Engl J Med*, 1996, 334, 25, 1649-1654
- 6 A.C. Guyton, J.E. Hall, "Textbook of Medical Physiology", *W.B. Saunders Company*, 1996, 9th edition, ISBN: 0-7216-5944-6
- 7 Consensus Trial Study Group, "Effect of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS)", *N Engl J Med*, 1987, 316, 23, 1429-1435
- 8 Solvd Investigators, "Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure", *N Engl J Med*, 1991, 325, 5, 293-302
- 9 H. Urata, B. Healy, R.W. Stewart, F.M. Bumpus, A. Husain, "Angiotensin II-forming pathways in normal and failing human hearts", *Circulation Research*, 1990, 66, 4, 883-890
- 10 H. Urata, A. Kinoshita, K.S. Misono, F.M. Bumpus, A. Husain, "Identification of a highly specific chymase as the major angiotensin II-forming enzyme in the human heart", *The journal of biological chemistry*, 1990, 265, 36, 22348-22357
- 11 A.A. Voors, Y.M. Pinto, H. Buikema, H. Urata, M. Oosterga, G. Rooks, J.G. Grandjean, D. Ganten, W.H. van Gilst, "Dual pathway for angiotensin II formation in human internal mammary arteries", *British journal of Pharmacology*, 1998, 125, 5, 1028-1032
- 12 B. Pitt, R. Segal, F.A. Martinez, G. Meurers, A.J. Cowley, I. Thomas, P.C. Deedwania, D.E. Ney, D.B. Snavey, P.I. Chang, "Randomised trial of losartan versus captopril patients over 65 with heart failure (Evaluation of Losartan in the Elderly Study, ELITE)", *Lancet*, 1997, 349, 9054, 747-752
- 13 R.S. McKelvie, S. Yusuf, D. Pericak, A. Avezum, R.J. Burns, J. Probstfield, R.T. Tsuyuki, M. White, J. Rouleau, R. Latini, A. Maggioni, J. Young, J. Pogue, "Comparison of candesartan, enalapril, and their combination in congestive heart failure: randomised evaluation of strategies for left ventricular dysfunction (RESOLVD) pilot study", *Circulation*, 1999, 100, 10, 1056-1064
- 14 B. Pitt, P.A. Poole-Wilson, R. Segal, F.A. Martinez, K. Dickstein, A.J. Camm, M.A. Konstam, G. Riegger, G.H. Klinger, J. Neaton, D. Sharma, B. Thiyagarajan, "Effect of Losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial - the Losartan Heart Failure Survival Study ELITE II", *Lancet*, 2000, 355, 1582-1587
- 15 C. Delcayre, J.S. Silvestre, A. Garnier, A. Oubenaissa, S. Cailmail, E. Tatara, B. Swynghedauw, V. Robert, "Cardiac aldosterone production and ventricular remodeling", *Kidney Int*, 2000, 57, 4, 1346-1351
- 16 A.D. Struthers, "Aldosterone: cardiovascular assault", *Am Heart J*, 2002, 144, 5, S2-S7
- 17 A. D. Struthers, "Why does spironolactone improve mortality over and above an ACE inhibitor in chronic heart failure?", *Br J Clin Pharmacol*, 1999, 47, 5, 479-482
- 18 K.T. Weber, "Extracellular matrix remodeling in heart failure: a role for de novo angiotensin II generation", *Circulation*, 1997, 96, 11, 4065-4082
- 19 C.S. Barr, C.C. Lang, J. Hanson, M. Arnott, N. Kennedy, A.D. Struthers, "Effects of adding spironolactone to an ACE inhibitor in chronic congestive heart failure secondary to coronary artery disease", *Am J Cardiol*, 1995, 76, 17, 1259-1265
- 20 K.M. Yee, A. D. Struthers, "Aldosterone blunts the baroreflex response in man", *Clin Sci*, 1998, 95, 6, 687-692
- 21 B. Pitt, F. Zannad, W.J. Remme, R. Cody, A. Castaigne, A. Perez, J. Palensky, J. Wittes, "The effect of spironolactone on morbidity and mortality in patients with severe heart failure", *N Engl J Med*, 1999, 341, 10, 709-717
- 22 B. Pitt, G. Williams, W.J. Remme, F. Martinez, J. Lopez-Sendon, F. Zannad, J. Neaton, B. Roniker, S. Hurley, D. Burns, R. Bittman, J. Kleiman, "The EPHEsus trial: eplerenone in patients with heart

- failure due to systolic dysfunction complicating acute myocardial infarction", *Cardiovasc Drugs Ther*, 2001, 15, 1, 79-87
- 23 B. Pitt, W. Remme, F. Zannad, J. Neaton, F. Martinez, B. Roniker, R. Bittman, S. Hurley, J. Kleiman, M. Gattlin, "Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction", *N Engl J Med*, 2003, 348, 14, 1309-1321
  - 24 D.A. Sica, "Pharmacokinetics and pharmacodynamics of mineralocorticoid blocking agents and their effects on potassium homeostasis", *Heart Fail Rev*, 2005, 10, 1, 23-29
  - 25 H. Krum, H. Nolly, D Workman, W. He, B. Roniker, S. Krause, K. Fakouhi, "Efficacy of eplerenone added to renin-angiotensin blockade in hypertensive patients", *Hypertension*, 2002, 40, 2, 117-123
  - 26 M.F. Rousseau, O. Gurné, D. Duprez, W. van Mieghem, A. Robert, S. Ahn, L. Galanti, J.-M. Ketelslegers, "Beneficial neurohormonal profile of spironolactone in severe congestive heart failure", *J Am Coll Cardiol*, 2002, 40, 9, 1596-1601
  - 27 R.W. Hartmann, U. Muller, P.B. Ehmer, "Discovery of selective CYP11B2 (aldosterone synthase) inhibitors for the therapy of congestive heart failure and myocardial fibrosis", *Eur J Med Chem*, 2003, 38, 4, 363-366
  - 28 A. Wada, T. Ohnishi, Y. Nonaka, M. Okamoto, T. Yamano, "Synthesis of aldosterone by a reconstituted system of cytochrome P-45011 $\beta$  from bovine adrenocortical mitochondria", *J Biochem (Tokyo)*, 1985, 98, 1, 245-256
  - 29 K. Yanagibashi, M. Haniu, J.E. Shively, W.H. Shen, P. Hall, "The synthesis of aldosterone by the adrenal cortex. Two zones (fasciculata and glomerulosa) possess one enzyme for 11 beta-, 18-hydroxylation, and aldehyde synthesis", *J Biol Chem*, 1986, 261, 8, 3556-3562
  - 30 Y. Nonaka, H. Takemori, S.K. Halder, T. Sun, M. Ohta, O. Hatano, A. Takakusu, M. Okamoto, "Frog Cytochrome P-450 (11 $\beta$ , aldo), a single enzyme involved in the final steps of glucocorticoid and mineralocorticoid biosynthesis", *Eur J Biochem*, 1995, 229, 1, 249-256
  - 31 A. Fisher, E.C. Friel, R. Bernhardt, C. Gomez-Sanchez, C. Connell, J.M.C. Fraser, E. Davies, "Effects of 18-hydroxylated steroids on corticosteroid production by human aldosterone synthase and 11 $\beta$ -hydroxylase", *J Clin Endocrinol Metab*, 2001, 86, 9, 4326-4329
  - 32 L.K. Domalik, D.D. Chaplin, M.S. Kirkman, R.C. Wu, W. Liu, T.A. Howard, M.F. Seldin, K.L. Parker, "Different isozymes of mouse 11 beta-hydroxylase produce mineralocorticoids and glucocorticoids", *Mol Endocrinol*, 1991, 5, 12, 1851-1861
  - 33 Y. Nonaka, M. Okamoto, "Functional expression of the cDNAs encoding rat 11 $\beta$ -hydroxylase [cytochrome P450(11 $\beta$ )] and aldosterone synthase [cytochrome P450(11 $\beta$ , aldo)]", *Eur J Biochem*, 1991, 202, 3, 897-902
  - 34 M. Murray, "Mechanisms of the inhibition of cytochrome P 450-mediated drug oxidation by therapeutic agents", *Drug Metab Rev*, 1987, 18, 1, 55-81
  - 35 T.L. Poulos, A.J. Howard, "Crystal structures of metyrapone- and phenylimidazole-inhibited complexes of cytochrome P-450cam", *Biochemistry*, 1987, 26, 25, 8165-8174
  - 36 R.W. Brueggemeier, J.C. Hackett, E.S. Diaz-Cruz, "Aromatase inhibitors in the treatment of breast cancer", *Endocrine Reviews*, 2005, 26, 3, 311-345
  - 37 P. Furet, C. Batzl, A. Bhatnagar, E. Francotty, G. Rihs, M. Lang, "Aromatase inhibitors: synthesis, biological activity, and binding mode ofazole-type compounds", *J. Med. Chem.*, 1993, 36, 10, 1393-1400
  - 38 S.W.J. Lamberts, H.A. Bruining, H. Marzouk, J. Zuiderwijk, P. Uitterlinden, J.J. Blijd, W.H.L. Hackeng, F.H. de Jong, "The new aromatase inhibitor CGS-16949A suppresses aldosterone and cortisol production by human adrenal cells in vitro", *Journal of Clinical Endocrinology and Metabolism*, 1989, 69, 4, 896-901
  - 39 L.E. Demers, J.C. Melby, T.E. Wilson, A. Lipton, H.A. Harvey, R.J. Santen, "The effects of CGS 16949A, an aromatase inhibitor on adrenal mineralocorticoid biosynthesis", *Journal of Clinical Endocrinology and Metabolism*, 1990, 70, 4, 1162-1166

## Chapter 2 Homology Modelling

The construction of a three-dimensional model of a protein structure provides crucial insights to protein-ligand interactions involved in ligand binding, ligand stabilisation and substrate conversion. Here, the construction of homology models is presented for the human and rat CYP11B isoforms, as well as an important CYP11B2 triple mutant that mainly possesses CYP11B1 activity. Using the knowledge of substrate hydroxylation sites, the general binding modes of the steroids in the CYP11B homology models are derived. As a next step, we explain the difference in substrate binding caused by amino acid differences in the active site as well as amino acids located in regions lining the active site. Finally, we propose the CYP11B2 specific ligand-binding characteristics of 18-hydroxycorticosterone that are required for aldosterone synthesis.

Part of this chapter is described in:

L. Roumen, M.P.A. Sanders, K. Pieterse, P.A.J. Hilbers, R. Plate, E. Custers, M. de Gooyer, J.F.M. Smits, I. Beugels, J. Emmen, H.C.J. Ottenheijm, D. Leysen, J.J.R. Hermans, "**Construction of 3D models of the CYP11B family as a tool to predict ligand binding characteristics**", *J Comput-Aided Mol Des*, 2007, 21, 8, 455-471

## 2.1 Introduction

When performing protein-based drug design, it is important to know the details of the target regarding its structure, function and regulation. The knowledge of these protein features can be used for the elucidation of protein-ligand interactions and the design of novel drugs. Therefore, we first introduce the functional and structural protein features of cytochrome P450 enzymes. Because no three-dimensional structure of CYP11B1 or CYP11B2 is available, the structural insights on cytochrome P450 proteins have been used to develop homology models for the members of the CYP11B family. In addition, knowledge on the catalytic activity of cytochrome P450 enzymes has been used to derive the binding mode of the CYP11B1 and CYP11B2 substrates in their protein active sites, as well as to quantify the regio-specific substrate hydroxylations performed by the different family members.

## 2.2 Cytochrome P450 enzymes

Cytochrome P450 is a large super family of proteins that contain a heme prosthetic group in the active site. The cytochrome P450 enzymes catalyse many types of reactions and are called mixed-function oxidases or mono-oxygenases, because they incorporate one atom of molecular oxygen into the substrate and one oxygen atom into water. This oxidation is performed by the heme-oxygen complex, often aided by several active site residues such as a catalytic threonine. Cytochrome P450 enzymes differ from di-oxygenases that incorporate both oxygen atoms into the substrate [1]. The enzymes are involved in numerous processes such as the biosynthesis and metabolism of sterols, bile acids and steroids, and the metabolism of endogenous fatty acids, drugs and other xenobiotic compounds [2].

### 2.2.1 Nomenclature and generic function

The name cytochrome is a combination of both *cytos* (cell) and *chromos* (colour) derived from Greek. The name P450 originates from the carbon-monoxide - ferrous-heme complex which produces a spectral peak pigment at 450 nm. The nomenclature of cytochrome P450 enzymes is based on the evolution of protein sequences. Genes that encode cytochrome P450 enzymes, and the enzymes themselves, are abbreviated with the root name CYP. This root name is followed by an Arabic numeral indicating the gene family, a capital letter indicating the subfamily, and a second Arabic numeral that identifies the individual gene [3,4,5,6,7,8]. Two cytochrome P450 enzymes belong to the same family if their sequence identity is higher than 40%, and they belong to the same subfamily if that sequence identity is above 55%. An example is CYP1A1, which belongs to family 1, subfamily A, and is the first isoform of the subfamily. Similarly, aldosterone synthase (CYP11B2) belongs to family 11, subfamily B and is the second isoform.

The functional classification of cytochrome P450 enzyme containing mono-oxygenase systems yields two main classes; the bacterial and mitochondrial class I and the microsomal class II. These classes are based on the functionality of the reduction system. Most cytochrome P450 enzymes require a protein partner to deliver one or more electrons to reduce the iron during the enzymatic activity. In general, class I enzymes interact with an

iron-sulphur reductase, so called ferredoxin reductases, and class II enzymes interact with a flavoprotein reductase. Unfortunately, the classification into two classes fails to describe the whole diversity of cytochrome P450 enzyme systems [1] and there are several exceptions to the classifications where cytochrome P450 enzymes belonging to one class interact with reductases belonging to the other class [9,10,11].

Cytochrome P450 enzymes are often involved in a complex of several proteins that perform the transfer of electrons to the substrate-bound heme [1]. These systems involve several redox-domains provided by proteins in various combinations. The main domains in the complex are (1) a flavin adenine dinucleotide (FAD) domain belonging to a reductase or flavoprotein, (2) an intermediate domain that can either be a flavin mononucleotide (FMN), ferredoxin or cytochrome b5 domain, and (3) the heme prosthetic group of the cytochrome P450 enzyme. In case of self-sufficient cytochrome P450 enzymes, the FAD domain and the intermediate domain are a part of the cytochrome P450 enzyme.

The catalytic cycle of a cytochrome P450 begins with the binding of a substrate to the cytochrome P450 active site. Next, the two proteins carrying the FAD and intermediate domains, respectively, bind to the cytochrome P450. The FAD domain uses NAD(P)H to supply electrons to the intermediate domain, which transfers these electrons to the heme of the cytochrome P450. In total, two electrons are supplied to the heme through the electron transfer and two protons are supplied to the heme by the solvent, before the catalytic state of the heme-oxygen complex has been completed. The heme-oxygen complex then oxidises the substrate, after which it is released to the environment and the cycle can start again [1]. The catalytic mechanism of cytochrome P450 enzymes is explained in more detail in Chapter 6.

### 2.2.2 Structural Architecture

During substrate conversion, the stability of the protein-substrate complex is determined by the spatial orientation of the enzyme active site. The active site is in turn stabilised by the three dimensional architecture of the remaining the protein structure. All cytochrome P450 enzymes consist of 12 alpha-helices annotated from A to L, as well as 5 beta-sheets (Figure 2–1). Additionally, several short helices are present in various cytochrome P450 enzymes (annotated B', F', G', J' and K') whilst being absent in others. The exact function of these short helices is unknown, but may be related to the recognition, binding and stabilisation of the cytochrome P450 redox partners.

The structural core of all cytochrome P450 enzymes consists of the four-helix bundle composed of helices D, E, I and L, and the two helices J and K [12,13,14,15]. The heme prosthetic group is linked to the protein near the beginning of alpha-helix L and is stabilised on one side by the helix bundle, and on the other side by variant regions around helices B, B' and C. These regions do vary between cytochrome P450 enzymes because they make up a part of the active site that is involved in ligand binding and substrate specificity. The helices A, B and H are more distant from the active site and are involved in the binding of redox partners. Lastly, helices F and G vary between cytochrome P450 enzymes, are flexible and can slightly move to allow substrates to penetrate the active site [14,15].

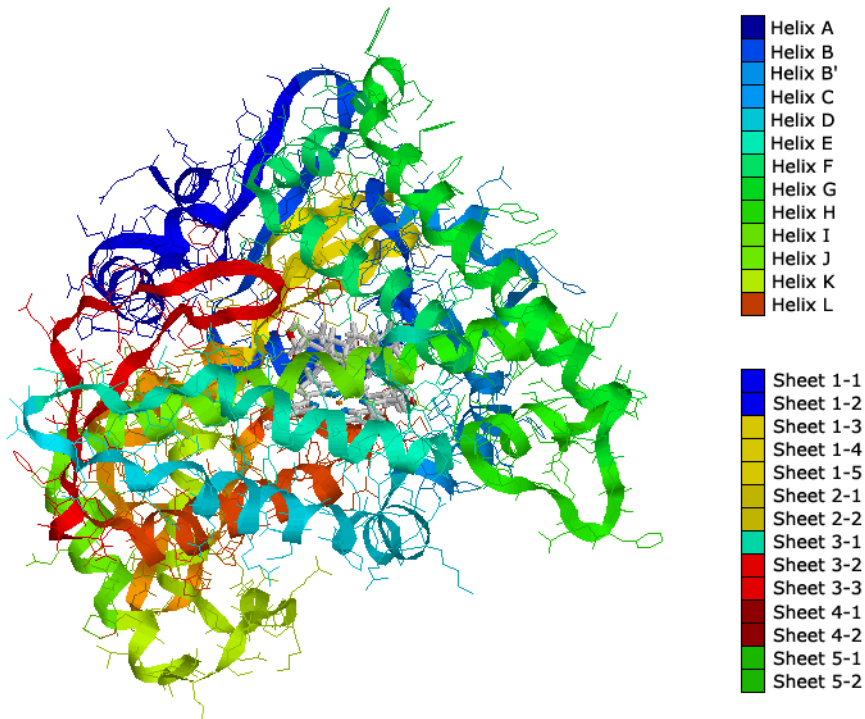


Figure 2–1 Generic structural fold of cytochrome P450 enzymes.

The structural features involved in ligand binding and substrate specificity are called substrate recognition sites (SRS). These SRS are located in helices B', F, G, I and their adjacent loop regions, in close proximity to the heme prosthetic group (Figure 2–2) [12]. Next to the SRS, all cytochrome P450 enzymes possess several characteristic sequence motifs in their amino acid sequence [13]. These are (1) a (W/H)(R/K)X(R/K)R motif in helix C that stabilises one of the propionic acid groups of the heme, (2) an XGXXTX motif in helix I that supplies a catalytic threonine for many cytochrome P450 enzymes, (3) an EXXR motif in helix K that interacts with the meander region that is important in cytochrome P450 - redox partner interactions [16], (4) a (W/F)XXPXX(F/Y)XPX(H/R)(W/F) motif after helix K' that comprises the meander region, and (5) an XXF(G/S)XGX(H/R)XCXGXX(L/F)AXXE motif that is found at the beginning of helix L and contains the cysteine that binds to the heme iron (Figure 2–2) [13].

```

SECOND: -----AAAAAAAAA---|b1-1|-----
CYP101: -NLAPLPPHVPEHLVDFDMYNPSNLSAGVQEAWAVLQESN---VPDLVWTRCNG-GHW
CYP102: -----TIKEMPQKTFGELKNLPLLNDKPVQALMKIADE-----LGEIFKFEAPGRVTR
CYP119: -----GKLPFGPTFPFIIGNILQIDA---KDISKSLTKFSECYGPVFTVYLGMKPTV
CYP2C5: -----GKLPPGPTFPFIIGNILQIDA---KDISKSLTKFSECYGPVFTVYLGMKPTV
CYP3A4: HSHGLFKKLGIPGPTPLPFLGNILSYHK---GFCMFDMECKKYGKVVGFYDQQPVLL
                                     |SRS1|
SECOND: |b1-2|BBBBBBBBB---|b1-5|-----B'B'B'B'-----
CYP101: IATRGQLIREAYEDYR--HFSSEC-PFIPREAGEAY-----DFIPTSMDP
CYP102: YLSSQRLIKEACDESFR---DKNL-----SQALKFVRDF-----AGDGLFTSWT
CYP119: QVFSYRYTKEVLNDFS--KFSSDLT-----GYHERLEDLRNGKIRFDIPTRYTMTLSDP
CYP2C5: VLHGVEAVKEALVDLGEFAGRGSV-----PILEKVS-----KGLGIAFSNA
CYP3A4: AITDPDMIKTVLVEKCVSVFTNRRP-----FGPVGFM-----KSA-ISIAED
                                |MOT1|
SECOND: CCC|CCCC|CCCCC-----DDDDDDDDDDDDDDDDDD---|b3-1|EEEEEE
CYP101: PE--QRQFRALANQVVGMP-VVD-KLENRIQELACSLIESLRPQ---GQ--CNFTEDYA
CYP102: HEKNWKAHNILLPSFSQQ-AMK-GYHAMMVDIAVQLVQKWERL---NADEHIEVPED-M
CYP119: PL--HDELRSMSADIFSPQ-KLQ-TLETFIRETTRSLDSIDP---RE--DDIVKLA
CYP2C5: KT--WKEMRFRSLMTRLNRFMGKRSIEDRIQEEARCLVEELRKT--NASP-CDOPTI-L
CYP3A4: EE--WKRLRSLLSPTFTSGKLKE--MVPIIAQYGDVLRNLRREAETGKP-VTLKDV-F
                                       |SRS2|
SECOND: EEEEEEEEE---FFFFFFFFFFFFFFFFFFFFFFFF'-----
CYP101: EPFPFIRIFMLLAGLP-----EEDIPHLYKLTQDMTRP-----
CYP102: TRLTLDITGLCGFNRYFNFSYRQPHFITSMVRALEAMNKLRANPD-----DP
CYP119: VPLPIIIVISKIL-GLPI-----EDKEKFEKESDLVAFRLG-----
CYP2C5: GCAPCNVICSVIFHNRFDY-----KDEEFKLMESLHENVELLGTWLVQVNNFPALLD
CYP3A4: GAYSMDVITSTSFGVNISSLNN-PQDPFVENTKLLRFDLDPFELSITVFPFLIPILE
                                |SRS3|
SECOND: ''-GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG-----HHHHHH|b5-1|b5-2|III
CYP101: ---DGSMFAEAKALYDYLPIIEQRKQPG---TDAISIVANGVNN--GRPITSDE
CYP102: AYDENKRFRFEDIKVMNDLVDKIADRKASGEQ--SDLLTHMLNGKDPETGEPLDDEN
CYP119: ---KPEIFELGKKYLELIGYVKDHLNSGT-----EVSVRVVNSN-----LSDIE
CYP2C5: YFPGIHKTLKNADYIKNFIMEKVKHQKLLDVNNPRDFIDCFIKMEQENNFLEFSLE
CYP3A4: VLNICVFPREVNTFLRKSVMKMKESRLEDQKH--RVDFLQMLDSQNS|SHKALSDE
                                |SRS4| |MOT2|
SECOND: II|IIIII|IIII|IIII|IIII|IIIIIIJJJJJJJJJJJJ-----'-----
CYP101: AKRMCGLLLVGGDLTVVNFLSFSEMEFLAKSPEHRQELIERP-----
CYP102: IRYQIITFLIAGHETTSGLLSFALYLVKNPHVLQKAAEEAARVLDVPV--PSYQVKQL
CYP119: KLGYIILLIAGNETTNLINSNSVIDFTRFN-LWQRIREE-----
CYP2C5: LVAVSDFLFGAGTETTSITLRYSLLLLLKHPEVAARVQEEIERVGRHSRPMQRDRSM
CYP3A4: LVAQSIIIFPAGYETTSVLSFI|MYELATHPDVQKQLQEEDI|AVLPNKAPPTYDTVLQM
                                |MOT3| |SRS5|
SECOND: KKKKKKKKKKK|b6-1|b1-4|b2-1b2-2|b1-3|-----
CYP101: ERIPAAEEELRFRFLV-A-DGRILTSDFYFHHG-VQLKKGQIILLPQLSGLDERENA-
CYP102: KYVGMVLNEALRLWPTAPA-FSLYAKEDTVLGGEYPLEKGDMLVIPQLHRDKTIWGD
CYP119: NLYLKAIEEALRYSPPVMR-TVRKTKERVKLGD-QTIEEGEVVVWIASANRDEEVFHF
CYP2C5: PYTDAVHFIQRFDLPTNLPHAVTRDVRFRN-YFIPKGTDIITSLTSLVHDEKAFP-
CYP3A4: EYLDMMVNETLRFPIAMR-LERVCKKDVING-MFIPKGVVVMIPSYALHRDPKYWT-
                                |MOT4| |MOT5|
SECOND: |meander|-----|-----LLLLLLLLLLLLLLLLLLLLL---|
CYP101: CPMHVDFSRQ-----KVSHTTFGHGSHLCLGQHARRREIIVTLKEWLTRIPDFS
CYP102: DVVEEFRPERFENPSAI---PQHAFKPFGNGQRACIGQQFALHEATLVLGMLLKHFDFE
CYP119: DGKGFIPDRN-----PNPHLSFGSGIHLCLGAPLARLEARIAIEEFSKRFRHIE
CYP2C5: NPKVFDGPHFLDESNGFKSD-YFMPPSAGKRMVCGEGLARMEFLFLTSLILQNFK-LQ
CYP3A4: EPEKELPFBPSKKNKNDIDPY-IYTFPGSGPRNCIGMRFALMNMKLALIRVLQNFNFS-FK
                                     |SRS6|
SECOND: |b3-3|---|b4-1||b6-2||b4-2b3-2|-----
CYP101: IAPGA---QIQH-KS-GIVSGVQ-ALPLVWPATTKAV
CYP102: DHTNY---ELDI-KE-TLTLKPE-GFVVKAKSKKIPL-
CYP119: IL-----DTEKVP-NEVLNGYK-RLVVRLLKSN-----
CYP2C5: SLVEPKDLDIRA-VVNGFVSVPP-SYQLCFPIIHH---
CYP3A4: PCKETQIPLKLS-LGGLLQPEKP-VVLKVESRDGT---

```

Figure 2–2 Location of cytochrome P450 sequence motifs (dotted squares) as well as substrate recognition sites (solid squares)



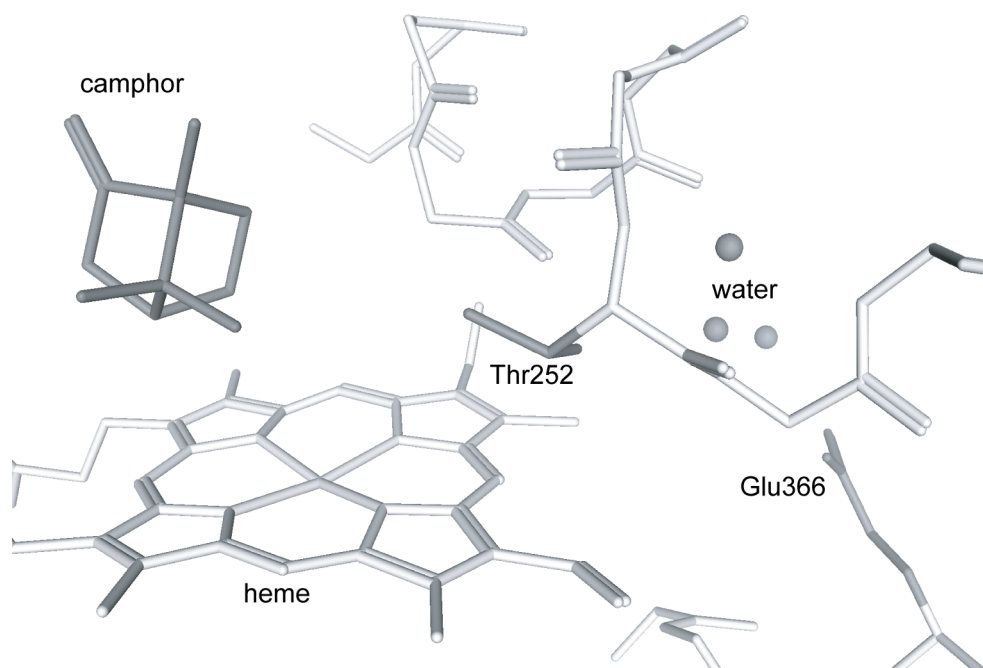


Figure 2–3 Putative water channel in cytochrome P450 enzymes that supplies protons from a glutamic acid residue through the centre of helix I, here CYP101, PDB-code 2CPP; Glu366. Indicated are also the catalytic threonine that plays an important role during the catalytic cycle of the cytochrome P450 enzyme (CYP101; Thr252), as well as the CYP101 substrate camphor.

A final important feature of the cytochrome P450 structure is a small cavity containing a glutamic acid residue. This cavity is connected to the active site and the catalytically important threonine residue through a putative water channel in helix I. The water channel may play a role in the delivery of protons during substrate conversion (to and from the glutamic acid), or may play a role in the decoupling of the cytochrome P450 reactive species if proton or electron delivery is impaired or desynchronised (Figure 2–3) [17].

The combination of these structural features determines the structural integrity of the cytochrome P450 fold and is important for the catalytic function. Thereby, all the helices are connected to each other to form a compact and highly stabilised structure. These features also link the heme prosthetic group tightly to the active site and place the catalytically important active site residues in the ligand binding site. A prerequisite for the modelling of CYP11B1 and CYP11B2 is that in the constructed models, these active site characteristics discussed above, are fulfilled.

### 2.3 Homology Modelling

Homology modelling or comparative modelling is widely applied to proteins for the purpose of understanding protein-ligand interactions. In particular, the inhibition of enzymes has been extensively investigated in this manner. The term homology modelling implies that a model is constructed for the protein of interest based on the features of another protein, called the template protein or template structure. Ideally, the best model accuracy is obtained when the three-dimensional architecture of the template protein closely resembles that of the target protein. Hence, modelling one cytochrome P450 requires another cytochrome P450 structure to act as a template.

In general, the atomic data of chemical structures can be acquired by performing X-ray crystallography, NMR spectroscopy and electron microscopy. Resolving atomic data for proteins is commonly performed by X-ray crystallography, since NMR spectroscopy is restricted to relatively small molecules and electron microscopy is restricted by resolution issues. For X-ray crystallography to succeed, a protein needs to be soluble or be made soluble. As a result, the three-dimensional structure of membrane-bound proteins and proteins with large hydrophobic regions are difficult to resolve. Many cytochrome P450 enzymes such as CYP11B1 and CYP11B2 are membrane-bound proteins, but fortunately, crystal structures have been derived for bacterial cytochrome P450 enzymes [18,19,20], as well as mammalian cytochrome P450 enzymes (often solubilised) [21,22,23,24,25,26]. These crystal structures can be found in the Brookhaven Protein Databank [27].

The resolved crystal structures not only reveal the overall packing of the protein, the structural stabilisation and flexibilities, but also detailed interactions between protein, ligand and water. Unfortunately, the three-dimensional structure has not been resolved for either CYP11B1 or CYP11B2. Because 3D insight on ligand binding plays an important role in drug design, homology modelling can be applied to construct a three-dimensional model for their architecture.

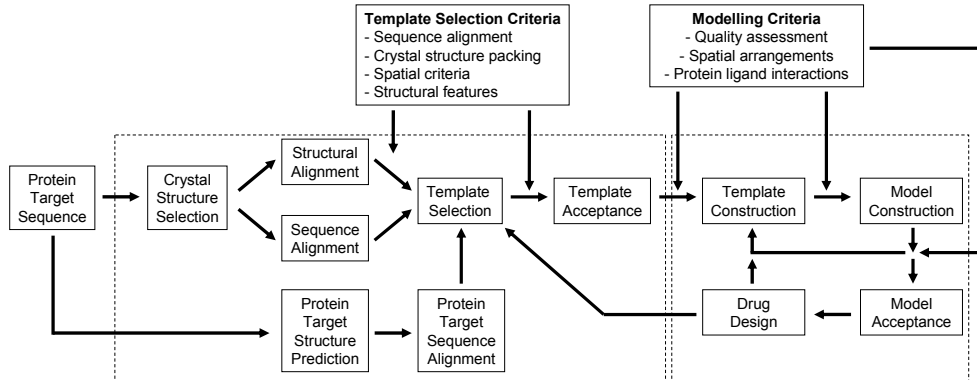


Figure 2–4 Flowchart of the homology modelling approach

The construction of a homology model is an iterative process involving sequence alignment, structural alignment, model building and structure assessment, before the final usage for drug design (Figure 2–4). For each modelling step, several criteria are tested to ensure the accuracy of model construction. The first step in homology modelling is the choice of the template structure. In this step, crystal structures of structurally related proteins are extracted from the protein databank [27]. The amino acid sequences of these proteins are aligned in accordance to their three-dimensional structure (topological alignment), and their three-dimensional structures are superimposed. Next, a secondary structure prediction is performed for the protein target, which is subsequently aligned to the multiple sequence alignment. By aligning these secondary structures to the three-dimensional overlay, the general side chain packing of the protein target can be assessed, and the best template structure can be chosen based on spatial arrangements of the various crystal structures. Several criteria play an important role in the acceptance of a template structure.

One criterion for template selection is that the sequence alignment should display a high similarity with the template for important three-dimensional structures. This ensures that the homology model consists of structural features comparable to the template. An example is the correct modelling of hydrophobicity of a beta sheet that is on one side interacting with solvent and on the other side interacting with hydrophobic regions of the protein. An amino acid mismatch may lead to an inverse modelling of the hydrophobicity distribution.

A next criterion is to decrease the amount of amino acid insertions or deletions in important structural arrangements, especially alpha helices and beta sheets. Insertion or deletion of an amino acid in a helix may lead to the loss of secondary structure integrity and the unfolding of the helix. Important stabilising interactions between amino acids are not assumed and the model detail is inaccurate. Insertions and deletions in loop regions are less important for the structures accuracy, unless they are in contact with the protein active site. In such a case, careful investigation of the spatial arrangement of the loop region must be performed. One of the methods to model these regions accurately is to use information on protein-ligand interactions.

A last criterion is that the overall packing of the amino acid residues is tight, as gaps in the three-dimensional structure will collapse during model refinement. When these gaps are hydrophilic, they can be filled with water to prevent the collapse, however, if there is no apparent entrance for the water molecules, the three-dimensional structure prediction requires further optimisation.

When the template is chosen, the amino acids of the target protein are superimposed on the template structure, followed by a structural optimisation of the amino acid residues. As such, the generic structural integrity of the template structure is inherited by the models. Next, the constructed homology model is subjected to several optimisation steps to decrease amino acid clashes and improve hypothetical protein-ligand contact points. Resulting from protein model acceptance and insights obtained during drug design, protein models are often re-evaluated and improved.

The continuous optimisation and verification of the sequence alignment and the spatial arrangements of the protein model are the most crucial steps during homology modelling. This is especially true when the sequence similarity of the template and target is low,

because when the sequence alignment contains several misalignments, the homology model will possess inaccuracies that may lead to the misinterpretation of protein structural stability and protein-ligand interactions. In certain cases where amino acids are inserted in secondary structures, the resulting homology model may be more plausible if those amino acids are neglected from the final model, since secondary structure disruptions are very difficult to control.

In general, the regions in and around the active site are used for the elucidation of protein-ligand interactions. Therefore, the modelling detail of these regions must be of the highest accuracy. Regions outside the active site do play a role in the general structural stability and can influence protein function, however, they often require the least attention in modelling.

#### **2.4 Homology Models for CYP11B1 and CYP11B2**

Before the first mammalian cytochrome P450 structures became available, modelling attempts were classically performed on bacterial cytochrome P450 enzymes, in particular CYP101 [28,29,30]. These models possessed a sequence identity with their template lower than 25%, because no realistic alternatives were available. The introduction of the class II bacterial CYP102 (exception from class I) allowed for modelling the functional properties of eukaryotic class II cytochrome P450 enzymes [31]. Although these early homology models contained a low sequence identity with their template structures and are intuitively suboptimal, it has been shown that they can describe key features of protein-ligand interactions [30,31]. For example, features observed for inhibitor binding in aromatase models have provided important insights for the development of drugs [30,32]. Currently, models often still feature bacterial cytochrome P450 enzymes as template [33,34], but methods involving the use of multiple crystal structures for model construction may prove to be the future trend [35,36]. A model based on the structure of several known enzymes would be more accurate since every additional segment will improve similarity or spatial coordination of protein regions. However, it can be understood that structural flaws are expected at locations where the different template structures are joined together. Moreover, if these regions are within the active site, they also need to be thoroughly refined.

Modelling work on CYP11B1 and CYP11B2 has already been performed by Belkina et al [37] and Ulmschneider et al [38]. The models of Belkina et al consider the potential spatial arrangement of the amino acids in the active site and hypothesise the hydrogen-bonding network involved with heme stabilisation. Furthermore, the effects of several amino acid mutations have been detailed. Ulmschneider et al focus on describing protein-inhibitor interactions and structure activity relations of their developed inhibitors. Both models are thoroughly characterised for those specific purposes, however, the goal of our modelling work encompasses not only the prediction of novel CYP11B inhibitors, but also the investigation of the regio-selectivity of the natural ligands within the enzyme active sites, and to detail potential protein-ligand interactions.

We intend to construct homology models for both the human and rat isoforms of the CYP11B family. Subsequently, these models will be used to rationalise the substrate regio-specificity of the human and rat isoforms. The homology models are validated using molecular dynamics in Chapter 3 and molecular docking in Chapter 4.

### 2.4.1 Modelling Criteria

A template structure must be chosen for the construction of the homology models. To ensure modelling accuracy, we have defined several criteria that must be satisfied by the template structure. We have already discussed that the template structure for modelling the three-dimensional architecture of the CYP11B isoforms is ideally another cytochrome P450 enzyme. Since the structure of many cytochrome P450 enzymes has been elucidated with X-ray crystallography, this requirement will not pose a problem. However, the CYP11B isoforms are mitochondrial class I cytochromes that interact with the ferredoxin-like reduction partners. In order to increase the modelling accuracy of regions involved with protein-protein binding, the template structure should be built using class I cytochrome P450 enzymes.

The template structure must also possess a high sequence identity with the CYP11B amino acid sequences in order to maximise the accuracy of sequence alignment. Not only does a high sequence identity provide clues on how to align the secondary structures, but it will also provide insights on the packing of highly conserved three-dimensional structures. Next to the alignment of the secondary structures, the three-dimensional organisation of these features is also of importance. The structural core of the cytochrome P450 enzyme structure is the four-helix bundle composed of helices D, E, I and L, and the two helices J and K [12,13,14,15]. To maintain the accuracy of this structural basis, these helices should be properly aligned with the template structure. In addition, the crystallographic accuracy of these regions must be high for the template, such that the modelling is not performed on unreliable crystal structures.

Next, ligand-binding characteristics of the template structure should closely resemble that of the CYP11B family. The CYP11B family members possess a highly hydrophobic core to supplement the steroid rings and are most likely surrounded by hydrophilic contact points to stabilise the C3-ketone and C<sub>20,21</sub>-hydroxyacetyl groups. When the template structure possesses similar ligand-binding characteristics as the CYP11B family, the amino acid packing can effectively be mimicked by the homology amino acid replacement, i.e. the orientation of the active site side chains is easily compared and refined for the CYP11B models. However, if the template structure is more hydrophilic than the active site of CYP11B, it will be more difficult to reconstruct the correct residue packing within the active site. Furthermore, the protein-ligand binding interactions for the template structure need to be properly defined or the protein-ligand interactions for the CYP11B isoforms may end up based on incorrect structural insights of the template structure.

Lastly, the spatial positioning of template active site regions is vital. If for instance the template active site allows space for a small amino acid and the replacement in the CYP11B isoforms is large, the homology model will become strained and may deteriorate during model refinement. As such, the amino acid spatial arrangements of the template structure must be comparable to that of the CYP11B amino acids, or they need to be thoroughly refined. Important spatial arrangements are the abovementioned structural core of the cytochrome P450 protein fold, but also the substrate recognition sites (SRS), cytochrome P450 motifs, heme stabilisation and the putative helix I water channel [12,13,16,17].

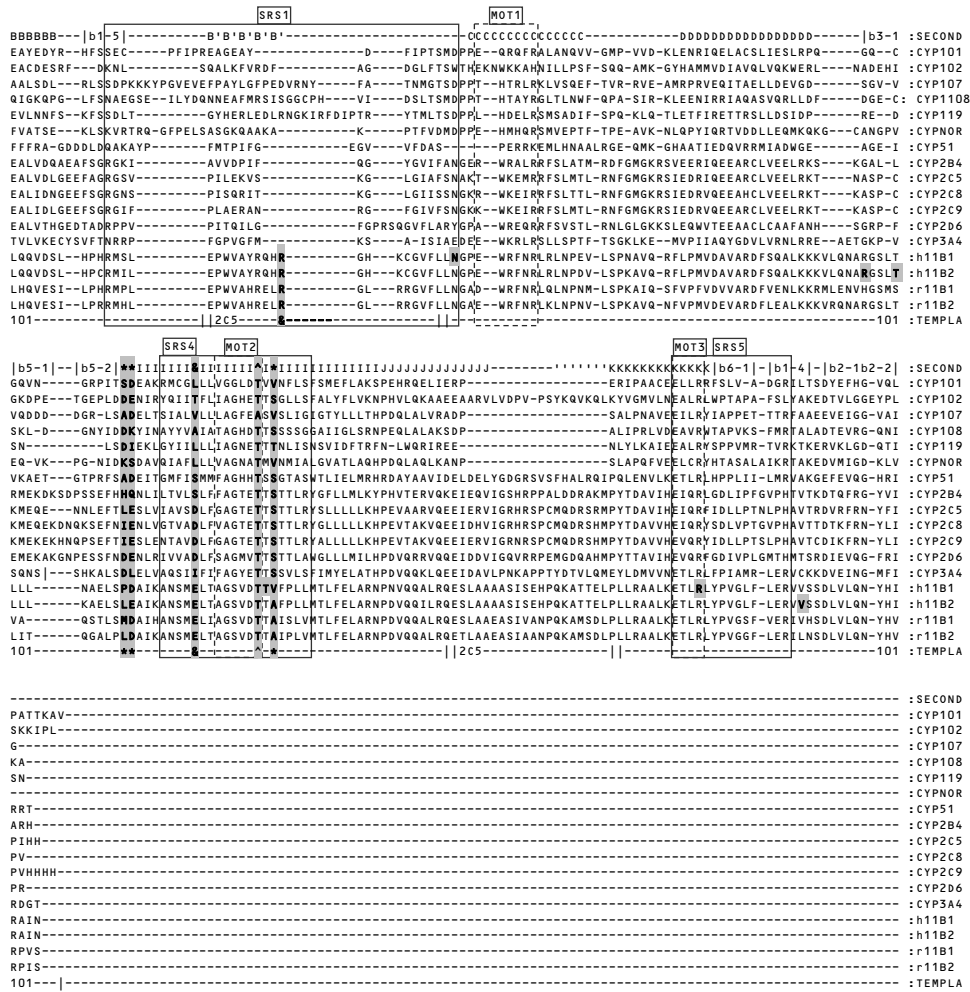
### 2.4.2 Multiple Sequence Alignment

For the creation of a multiple sequence alignment, the amino acid sequences of the CYP11B family have been taken from Swissprot [39] (human CYP11B1 accession P15538, human CYP11B2 accession P19099, rat CYP11B1 accession P15393, rat CYP11B2 accession P30099). From here on, the human isoforms will be denoted as hCYP11B1 and hCYP11B2, whereas the rat isoforms will be denoted as rCYP11B1 and rCYP11B2. The secondary structures for the proteins have been determined using the secondary structure program JPred [40]. The JPred program uses several secondary prediction methods to devise a consensus result for the location of alpha-helices, beta-sheets and random coils for the protein sequence. The prediction results have been used for an initial alignment of the secondary structures of the CYP11B isoforms to the secondary structures of cytochrome P450 enzymes for which a crystal structure is available. These crystal structures have been extracted from the Brookhaven Protein Databank and contain a unique four character code (subscript, Table 2-1) [27]. After the initial alignment, the multiple sequence alignment has been optimised using the MOE-Align sub-routine of the MOE modelling and visualisation package (Figure 2-5) [41].

CYP11B1 and CYP11B2 portray a high degree of homology possessing a pair wise sequence identity percentage as high as 94% for human and 83% for rat (highlighted, Table 2-1). This emphasises the difficulty of modelling the difference between the two isoenzymes, and the challenge of reaching the level of modelling accuracy that is required. The overall pair wise sequence identity of the CYP11B enzymes with cytochrome P450 enzymes for which a crystal structure has been elucidated is found to be less than 20%. Although this may seem like a limitation, early homology modelling attempts of cytochrome P450 structures were carried out for template and target possessing equally low sequence identities [28,29], and has proved to be successful for the investigation of inhibitor binding in aromatase models [30,32].

For each of the crystal structures, the active site residues have been investigated and compared with those of the CYP11B alignment. Taking into account those residues that contact the ligands in the crystal structures, a sequence identity measure of the protein active site is obtained (hydrophobic cut-off distance 4.5 Å). This method of active site residue comparison yields sequence identities of up to 32%. The low degree of homology indicates that none of the reviewed cytochromes clearly resembles the CYP11B isoforms. Hence, based on the sequence alignment score, no sole template can be selected as a representative for the CYP11B family. Therefore, the modelling of the CYP11B isoforms has been performed by creating a hybrid template that possesses important structural features extracted from multiple crystal structures.





Important amino acids are labelled with an asterisk (\*) and are indicated with a greyed background. These are: R123 in helix B', E310, the catalytic T318 and the triple-mutant L301P, E302D, A320V in helix I, and the conserved E459 in helix L. Individually greyed amino acids indicate known point mutants for CYP11B1 and CYP11B2.



Table 2–1 Generic pair wise sequence identity (in percentages) between the human and rat CYP11B isoforms and the cytochrome P450 enzymes for which a three dimensional structure has been elucidated.

Chains	101	102	107	108	119	55	51	2B4	2C5	2C8	2C9	2D6	3A4	h11B1	h11B2	r11B1	r11B2	PDB-CODE
CYP101	-	17.3	30.2	37.7	39.6	38.5	23.5	21.2	23.1	30.8	23.1	11.5	19.6	26.4	24.5	24.5	26.4	2CPP
CYP102	16.3	-	24.5	28.3	39.6	19.2	29.4	26.9	34.6	30.8	32.7	14.8	47.1	26.4	26.4	24.5	24.5	1BU7
CYP107	20.0	12.3	-	28.3	50.9	36.5	21.6	19.2	21.2	21.2	25.0	14.3	21.6	20.8	18.9	20.8	18.9	1JIN
CYP108	23.2	15.8	22.1	-	35.8	28.8	29.4	25.0	26.9	28.8	25.0	17.9	29.4	26.4	26.4	28.3	28.3	1CPT
CYP119	18.8	16.0	25.6	20.8	-	34.6	27.5	28.8	28.8	32.7	28.8	25.0	29.4	26.4	26.4	28.3	26.4	1F4U
CYP55	21.2	11.2	28.5	24.7	24.3	-	19.6	17.3	19.2	21.2	23.1	18.5	19.6	18.9	17.0	17.0	17.0	1ROM
CYP51	12.3	18.0	19.1	17.2	14.7	16.0	-	23.1	28.8	25.0	25.0	11.1	35.3	20.8	20.8	17.0	20.8	1EA1
CYP2B4	14.6	16.7	16.4	14.0	15.0	16.0	16.7	-	57.7	65.4	59.6	33.3	33.3	32.1	32.1	32.1	30.2	1SUO
CYP2C5	16.8	17.8	16.6	14.8	16.9	15.0	14.9	51.0	-	69.2	78.8	44.4	33.3	32.1	32.1	28.3	32.1	1NR6
CYP2C8	15.8	17.6	16.4	15.5	15.8	15.8	13.4	53.8	73.6	-	69.2	44.4	31.4	30.2	30.2	30.2	30.2	1PQ2
CYP2C9	15.6	18.0	17.4	15.0	16.6	15.3	14.3	51.0	77.3	78.4	-	40.7	33.3	32.1	32.1	28.3	30.2	1OG2
CYP2D6	13.1	16.9	14.9	14.3	14.7	16.0	16.7	39.6	40.0	40.6	38.5	-	14.8	25.0	25.0	25.0	21.4	2F9Q
CYP3A4	14.1	22.0	19.6	14.5	17.2	14.8	16.5	22.8	22.0	22.9	21.9	17.9	-	32.1	32.1	30.2	30.2	1TQN
h11B1	16.0	16.3	14.6	15.5	16.1	12.0	14.9	17.4	17.6	15.8	17.3	14.9	17.1	-	98.1	81.1	86.8	-
h11B2	15.3	16.9	13.6	14.8	15.8	12.0	14.5	17.4	17.6	16.2	17.7	15.5	17.5	93.6	-	83.0	88.7	-
r11B1	12.3	15.4	13.2	13.8	15.3	11.0	13.4	17.4	16.7	16.6	17.7	14.4	17.3	63.6	-	88.7	-	-
r11B2	15.3	15.8	14.4	14.5	16.6	13.0	14.7	15.9	16.2	15.8	17.1	13.5	16.8	68.2	68.8	82.6	-	-

The bottom triangle indicates the pair wise sequence identity of the whole protein, the top triangle indicates the pair wise sequence identity of the residues within 5.0 Å from the surface of the active site cavity of CYP101.

Species information: *Pseudomonas-Putida* 2CPP, *Bacillus Megaterium* 1BU7, *Saccharopolyspora-Erythraea* 1JIN, *Pseudomonas-SP*, 1CPT, *Archaeon Sulfolobus Solfataricus* 1F4U, *Fusarium-Oxysporum* 1ROM, *Mycobacterium Tuberculosis* 1EA1, *Oryctolagus Cuniculus* 1SUO and 1NR6, *Homo Sapiens*, 1PQ2, 1OG2, 1W0E and 2F9Q

### 2.4.3 Template Selection

Because of the low sequence identity, we have chosen to create a hybrid template for human CYP11B2 using MOE-Homology [41]. Using this method we have supplied the optimal structural information of multiple template structures and combined them into a hybrid template. Next, the amino acids of the template structure can be replaced for those of the cytochrome P450 11B isoforms.

Our criteria for selecting the template structure are (1) similarity in the cytochrome P450 reduction system, (2), high sequence identity of the structural core and confidence in the precision of the available crystal structures, (3) similarity in ligand characteristics, and (4) the spatial positioning of active site regions.

In accordance to the first criterion, the CYP11B family members belong to the class I cytochrome P450 enzymes that obtain electrons from ferredoxin reductases (paragraph 2.2.1 ) [42]. Therefore, using a bacterial cytochrome for the modelling of mitochondrial P450 enzymes is intuitively appropriate, rather than selecting one of the mammalian hepatic cytochrome P450 enzymes. In addition, the usage of bacterial proteins has been successfully applied to other mitochondrial P450 enzymes in the past [36,43].

As discussed above, none of the crystal structures possess a high sequence identity with the CYP11B isoforms, hence no distinction can be made based on that criterion. Another important consideration related to the crystal structure amino acid sequences is the requirement of a high confidence in the binding orientation of the ligand in the active site as well as the structural integrity of the protein fold. Investigation of the crystal structure of CYP2D6 (2F9Q) raises doubt on the appropriateness for its use as a template structure. The crystal structure is resolved at low resolution (3.0 Å), at which only few water molecules can be distinguished and the structure resolution only portrays the relative positioning of amino

acids in the structure, rather than detailed orientations and rotations of the amino acid side chains (for a high resolution below 2.0 Å). In addition, the crystal structure is resolved in a tetramer complex where numerous amino acids participate in an extensive interaction network between the four crystal structure monomers. Therefore, the structure was excluded from our modelling work. Conversely, the structure of CYP2C9 is highly accurate, however, it does not provide sufficient information about the molecular basis of regio-selective substrate hydroxylation. A modelling study on the substrate-binding of CYP2C9 showed that a conformational change of the active site is required to allow the substrate hydroxylation sites to contact the heme [44]. Because it is uncertain whether the active site fold around the CYP2C9 substrate is realistic, the structure was also excluded from our modelling work.

A third criterion for template selection involves the similarity of ligands characteristics. The natural ligands of the CYP11B family are steroids which possess a hydrophobic core as well as hydrophilic extremities. Several of the resolved class II mammalian cytochromes (CYP2C5, CYP2D6 and CYP3A4) can also convert steroids. In CYP2C5 and CYP2D6, steroids are oxidised on the  $\beta$ -side of the steroid skeleton at carbon atoms close to C<sub>11</sub> and C<sub>18</sub> (Figure 1–3 [45]), and their crystal structures may possess the necessary interaction features for model construction. CYP3A4, another hepatic cytochrome P450 enzyme, can also oxidise steroids, but its oxidation sites are on the  $\alpha$ -side of the steroid skeleton or on the different steroid rings. This implies that the steroids bind with a rotated or reversed orientation in the active site [46] and since steroids are slightly bent, the active site conformation of CYP3A4 possesses unfavourable spatial features. The substrates of bacterial CYP101 (cyclic alkanes) match the hydrophobic core of the steroid structure. In addition, the interaction of the alkanes in the active site is via a hydrogen bonding stabilisation on one end of the active site cavity (helix B'). These structural features are functionally similar to those of the natural ligands of CYP11B that only possess an additional hydrogen bonding interaction on the opposite region of the active site. The substrates for CYP2D6 are characterised by a basic nitrogen [47], and for CYP2C9 the substrates are mostly weakly acidic [48]. These properties are not comparable to those of the corticosteroids and it is anticipated that the amino acid packing of these active site residues is unsuitable for the modelling of the CYP11B isoforms. Hence, due to the difference in ligand properties, the structures of the human hepatic CYP2D6, CYP2C9 and CYP3A4 were found to be less suitable as template structure.

Lastly, as active site conformations play an important role in substrate binding, the spatial arrangement of the template structure determines the general architecture of the protein target. The structural core of CYP101 was found most suitable for the modelling of CYP11B isoforms because of an important active site feature. When comparing its active site cavity to that of the mammalian cytochrome P450 enzymes, we found that the steroidal ligands do not fit into the active site cavity of the mammalian P450s with C<sub>11</sub> and C<sub>18</sub> oriented to the heme, unless additional space is introduced near helix K. Because the most important interactions between protein and ligand take place near the heme, the regions lining the active site must be modelled with the highest accuracy. From the multiple sequence alignment with the CYP11B family one can see that beta-sheet 6-1 following helix K possesses a 1 amino acid insertion in the CYP101 structure and a 1 amino acid deletion in

the mammalian structures (SRS5, Figure 2–5). Although it is possible to model this site through a deletion in the mammalian structures, the resulting active site cavity becomes strained. Since the structure of CYP101 is elongated, modelling a 1 amino acid insertion and relaxing the fold creates a better spatial definition of this active site region.

Resulting from the various considerations, the crystal structures of cytochrome P450 class I CYP101 (based on spatial arrangements, ligand characteristics, pdb code 2CPP) and cytochrome P450 class II CYP2C5 (based on sequence identity, ligand characteristics, pdb code 1NR6) have been found representative for the construction of the hybrid model template. In comparison, the already published modelling attempts on cytochrome P450 11 family members have included mainly the usage of cytochrome P450 class II enzymes. These are CYP102 for the CYP11B1 and CYP11B2 modelling by Belkina et al [37] as well as for CYP11A1 modelling by Usanov et al [33], and CYP2C9 for the CYP11B1 and CYP11B2 modelling by Ulmschneider et al [38]. By performing hybrid modelling, we anticipate that we can construct the active site in detail by extracting the optimal parts from both templates.

#### 2.4.4 Template Construction

Taking into consideration that its reduction system and substrate are similar to that of the CYP11B family, and that the spatial features of CYP101 are most suitable, CYP101 is found to possess the best structural core for the modelling of the CYP11B isoforms. The structural core of the cytochrome P450 enzyme structure is the four-helix bundle composed of helices D, E, I and L, and the two helices J and K [12,13,14,15]. These regions have been modelled with the CYP101 structure to maintain the spatial accuracy of the structural basis. The remaining variant regions of the cytochrome P450 enzyme fold (helices B', F, G, J' and the meander region) have been modelled with CYP2C5, because several of these regions are SRS and because CYP2C5 possesses a higher sequence similarity to the CYP11B family in these regions. As mentioned above, CYP2C5 possesses substrate and active site properties comparable to the CYP11B family that allows for a proper modelling basis of these regions. In addition, CYP2C5 possesses slightly better spatial alignment with CYP101 in the connecting sequences than the other mammalian crystal structures, reducing the structural flaws for the transition between the two template structures.

The conformations of helix B' and the helices F and G have been emphasised in our modelling work, because this three-helical complex creates the entrance of the active site cavity and closes the active site pocket like a lid. The helix B' is predicted to be 3 turns long in the CYP11B family introducing an apparent insertion of 3 residues compared to helix B' in other cytochrome P450 structures (SRS1, Figure 2–5). The structure of CYP2C5 has been used to model this region by extending its helix from 2 turn length to a 3 turn length using an alpha-helical template extension. The template has been finalised by omitting the first 50 N-terminal residues corresponding to the membrane binding region of the human cytochrome 11B family, because no complementary sequence is present in the crystal structures of either CYP101 or CYP2C5. The final template can be reviewed in the multiple sequence alignment (lower line, Figure 2–5).

#### 2.4.5 Additional Modelling Criteria: Point Mutants

Several amino acid point mutants for CYP11B1 and CYP11B2 are known to change the expression of protein activity or determine the specificity of substrate binding and conversion. By taking into consideration some of these mutants the accuracy and confidence in the constructed homology models can be increased. In general, the effects of the naturally occurring mutants have not been determined in detail. However, by mapping the CYP11B1 and CYP11B2 amino acid sequences on the template structure we can determine the location of the mutants and obtain a general impression of their role. Subsequently, we can derive whether their effect can relate to active site changes or causes impairment in redox partner interaction.

For CYP11B1, several mutants have been identified in patients that have developed adrenal hyperplasia type 4, a recessive disease due to a defective synthesis of cortisol. These mutants are P42S, N133H, T318M, T319M, R374Q and R448H [49,50]. For CYP11B2, mutants have been reported in patients suffering from corticosterone methyl oxidase (CMO) type 1 or type 2 deficiencies, that cause a decreased synthesis of aldosterone, featuring the production of 18-hydroxycorticosterone to be low or normal (type 1), or increased (type 2). The mutant related to CMO-1 is L461P [51] and the mutants related to CMO-2 are R181W, T185I, E198D, V386A, T498A [52,53,54,55]. The sequential location of these mutants can be found annotated with an asterisk in Figure 2–5.

After fitting these mutants to the model template and investigating their location, we have found that most of the mutants are located on the protein surface and are likely involved in cytochrome P450-redox partner interactions, such as the mutants in helices D, E, K and L. A decreased or changed interaction between the protein-protein interaction may lead to a mismatch in the electron transfer chain and a decreased capacity for substrate conversion [56]. The two mutants for CYP11B1 (T318M, T319M) occur in a substrate recognition site, namely the catalytic centre of helix I (SRS4). Effects on decreased or changed protein action are expected but have not been assessed as it is out of the scope of this thesis. In analogy with mutants for other cytochrome P450 enzymes, their effect will be related to the uncoupling of the cytochrome P450 iron-oxygen species due to loss of substrate- or iron-oxygen-stabilisation [57,58,59,60].

In addition to the naturally occurring mutants, Bernhardt et al have conducted point mutation experiments on CYP11B1 and CYP11B2 changing the catalytic activity of the enzymes [61,62]. One particular experimental result showed that mutation of three residues outside the active site (L301P, E302D, A320V, helix I) is sufficient to convert the catalytic activity of CYP11B2 into that of CYP11B1. This experimental result has been used by Belkina and Ulmschneider to rationalise their models [37,38] and will also be used to rationalise the CYP11B1 and CYP11B2 differences of our models. This CYP11B2 triple mutant has led us to devise a hypothesis to explain the regio-selective hydroxylation of steroids by CYP11B1 and CYP11B2 (next paragraph), which is to be tested by homology modelling.

## 2.5 Hypothesis: Steric Aspects Play an Important Role in Substrate Conversion

After mapping of the CYP11B1 and CYP11B2 amino acid sequences on the model template, the triple-mutant introduced by Bernhardt et al [62] is found in close proximity of the substrate active site, yet does not directly contact the substrate. Instead, their effect must be indirectly mediated by amino acids lining the active site. This has led us to postulate that remote steric aspects can play an important role in substrate binding and substrate conversion [63]. Extending this hypothesis, we postulate that the difference in substrate conversion induced by the triple mutant is caused by a difference in the relative positioning of the substrate above the heme in the active site. It may indicate that there is a relation between substrate selectivity and the substrate hydroxylation distance, the distance between the heme iron and the substrate carbon, in the sense that the binding mode of the natural substrate dictates which carbon atom is oxidised first, with conversion taking place on the carbon atom which is in closest proximity to the iron-oxygen complex [63].

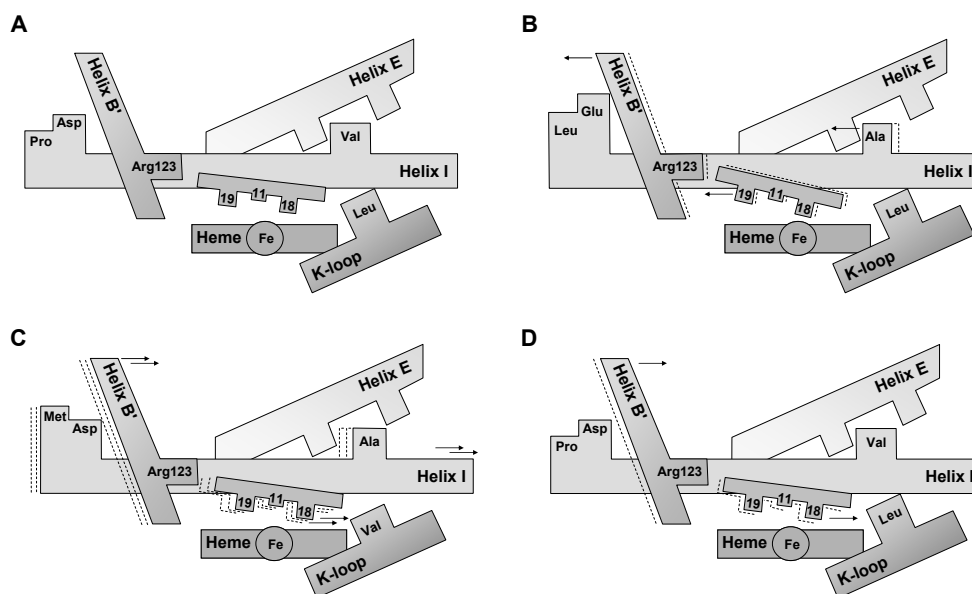


Figure 2-6 Schematic representation of the putative origin of substrate selectivity within (A) hCYP11B1, (B) hCYP11B2 and rCYP11B2, (C) rCYP11B1 and (D) hCYP11B2-TM. Minor steric differences in amino acids lining the active site can play an important role in the positioning of the substrate above the iron atom, resulting in a different substrate regio-specificity of the enzymes.

Using our hypothesis, an initial explanation for the substrate regio-selectivity of the cytochrome 11B isoforms (both human and rat) can be derived. For hCYP11B1, the hypothesis implies that C<sub>11</sub> and C<sub>18</sub> are to be in close proximity to the catalytic iron atom, with C<sub>11</sub> in closest proximity to the iron atom (Figure 2–6a). rCYP11B1 possesses a similar binding mode, yet is expected to also present C<sub>19</sub> to the iron atom in a position allowing oxidation (Figure 2–6c). Explaining the preference for C<sub>18</sub>-hydroxylation, hCYP11B2 and rCYP11B2 would bind with C<sub>18</sub> in a much closer proximity to the iron than expected for hCYP11B1 and rCYP11B1 (Figure 2–6b).

Extending this to the effects of the triple-mutant in helix I, the general interactions of the substrate in the active site of the different isoforms is schematically displayed in Figure 2–6d. The main putative stabilisation of the substrate in the active site cavity involves the hydrogen bonding interaction of its C<sub>3</sub>-carbonyl group with R123 in helix B'. The valine residue in helix I resides in a hydrophobic pocket and the leucine residue on the right-hand-side connects the steroidal hydroxyacetyl side chain in the pocket. A change of residue size in these positions would allow the substrate to shift spatially in the pocket. The two mutants at the beginning of helix I stabilise the spatial arrangement of helix B' in the active site. If in case of rCYP11B2 and hCYP11B2 the mutants can interact more strongly with helix B', it can be pulled further to the left-hand-side of the active site, and since the substrate is strongly bound to R123 inside the helix, it will move as well.

## 2.6 Construction of CYP11B Models

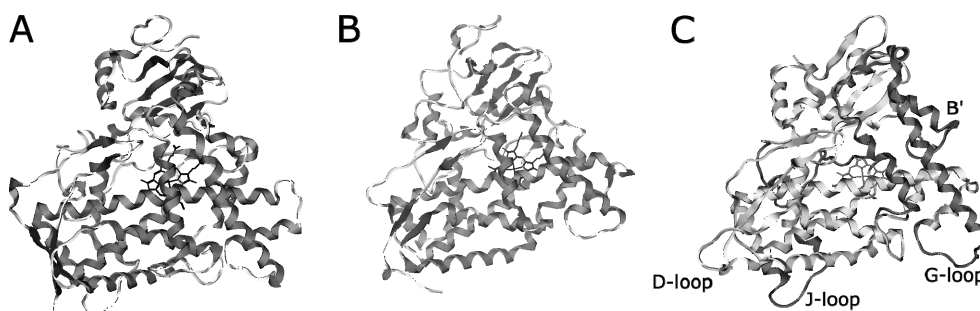
To test the hypothesis, a homology model has been constructed for each of the human and rat CYP11B isoforms based on the CYP101/CYP2C5 template structure. The initial model that has been constructed is hCYP11B2. Subsequently, models for hCYP11B1, rCYP11B1, and rCYP11B2 have been derived by amino acid replacement in the model of hCYP11B2, followed by a structural relaxation of the protein architecture using MOE [41]. In addition, we have investigated an hCYP11B2-triple mutant model (hCYP11B2-TM) by amino acid replacement in the model of hCYP11B2, to investigate the influence of the L301P, E302D, A320V triple-mutant. The three dimensional architecture of this structure should be similar to the active site of hCYP11B1.

The hCYP11B2 model has been constructed using MOE-Homology [41], followed by manual adjustments of amino acid side chains to compensate for large steric hindrances. Subsequently, the model has been refined by energy minimisation using a MOE succession method of steepest descent (for a quick minimum search), followed by conjugate gradient and truncated Newton (for a refined minimum optimisation) until the minimisation method had reached an RMS gradient of 0.1 kcal/(mol Å). This model refinement is an important step to reduce amino acid clashes and relax the overall protein structure. The minimisation has been performed using the Charmm22 force field [64] which is one of the best force fields optimised for modelling protein structures. In addition, all backbone atoms have been tethered with a force constant of 100 kcal/(mol Å<sup>2</sup>) to prevent large movements and unfolding of loop regions.

The four steroids 11-deoxycorticosterone (DOC), 18-hydroxy-11-deoxycorticosterone (18OH-DOC) corticosterone (B), and 18-hydroxycorticosterone (18OH-B) have been fit in the active sites of all CYP11B models matching all three CYP11B hydroxylation sites (C11, C18, C19) to the oxygen atom occupying the 6th ligating position of the heme iron. First the steroids have been docked into the active site using automated docking with GOLD v3.2 [65] (paragraph 4.6.1 ) to get an impression of the relative positioning of the C11, C18 and C19 above the heme. Next, docking of alternate orientations of the steroids has been investigated to test which ligand binding mode is optimal. A distance of 2.5 Å has been chosen as a starting distance between the oxygen and carbon atoms, which is a near-optimal distance for hydrogen atom abstraction during hydroxylation [66]. A threshold of 5.0 Å has been estimated to be the representative maximal distance for hydroxylation [66]. Subsequently, the models have been equilibrated with MOE without tethering the backbone atoms, allowing the model active sites to diverge.

## 2.7 Results and Discussion

The hCYP11B2 model constructed from the template based on CYP101 and CYP2C5 can be seen in Figure 2–5. The MOE optimisation and minimisation protocol carried out on the model has optimised the flexible loop regions on the outside of the protein, especially the regions following alpha helices D and G that carry amino acid insertions. In addition, we have carefully optimised helix B', which was extended by one helix turn during template construction.



**Figure 2–7** Crystal structures for a) CYP101 and b) CYP2C5. c) hCYP11B2 model after template refinement and model optimisation. The dark coloured helices B', F, F', G and G' and the J-loop are modelled based on CYP2C5.

### 2.7.1 Model quality assessment

The quality of homology models is of great importance for the application in drug design. The accuracy of the model has a direct connection to the elucidation of structure-function relationships. Therefore, to obtain confidence in the constructed models, several quality assessment tools have been used to check their validity.

One of the theorems commonly applied to the quality assessment of a homology model is the Ramachandran Plot. This plot is an indicative measure for the correctness of the residue torsion angles. The plot graphically displays torsion angle  $\phi$  ( $C_{n-1}-N_n-C_{\alpha,n}-C_n$ ) versus torsion angle  $\psi$  ( $N_n-C_{\alpha,n}-C_n-N_{n+1}$ ) for each residue of the protein (with  $n$  as the amino acid numbering). The alpha-helical character of a protein backbone is located roughly in the region where  $-60^\circ < \phi < -30^\circ$  and  $-120^\circ < \psi < -30^\circ$ , and the beta-sheet character is located roughly in the region where  $-180^\circ < \phi < -60^\circ$  and  $90^\circ < \psi < 180^\circ$  [67]. In Table 2–2, the results of the Ramachandran Plot are summarised for the CYP11B models after they have been equilibrated with the ligand 18-hydroxycorticosterone.

For both hCYP11B1 and hCYP11B2 around 95% of the residues are positioned in the favoured and core regions of the Ramachandran Plot. The distribution of amino acids in the Ramachandran plot for crystal structures is generally similar, indicating that our hybrid models are of comparable quality. Due to the high quantity of alpha-helices and beta-sheets, the majority of residues are positioned in the expected regions. The residues which are situated in disallowed and unfavoured regions of the plot are located in loop regions outside the active site. In total, 9 residues in the hCYP11B1 model are situated in the disallowed regions and 15 residues in the unfavoured regions. For the hCYP11B2 model, 10 residues are situated in disallowed regions and 14 residues in the unfavoured regions. The causes for these disparities are several insertions or deletions introduced in the models for which the structural minimisation was not sufficiently adequate to correct the backbone dihedrals. In particular, these regions are a relatively large insertion between alpha-helix D and beta-sheet 3-1, and an insertion between helix G and H.

Table 2–2 Validation results for the homology models of CYP11B1 and CYP11B2, part I.

Structure	Ramachandran Distribution (Core Regions) (%)	Ramachandran Distribution (favourable regions) (%)	Errat Score (quality factor) (%)
Template (PDB, resolution)			
CYP101 (2CPP, 1.63Å)	92.1	100.0	96.0
CYP2C5 (1NR6, 2.10Å)	87.8	99.2	93.6
Models			
hCYP11B1	78.8	94.7	84.1
hCYP11B2	78.7	94.7	87.5
hCYP11B2-TripMut	80.6	96.5	81.1
rCYP11B1	79.7	96.5	80.2
rCYP11B2	82.4	96.5	80.1

Errat outliers are positioned in the loop regions before helix F where the two template structures are connected, and at the end of helix D, where a large insertion was introduced.



Additionally, the amino acid environment of the models was evaluated with Errat [68] and Verify3D [69]. Errat determines the non-bonded interactions for each amino acid and compares the results to statistical data derived from 96 reliable protein structures from the Brookhaven Protein Databank. A reliability score is assigned to each amino acid, and the results are smoothed using a 9 amino acid sliding window. As a result, the Errat score for the first and last amino acids in the sequence has no meaning. The default Errat score uses a confidence rejection limit of 95% and 99% to determine which protein regions possess unfavourable packing (Figure 2–8). Errat also determines an overall quality factor, for which a score above 95% is expected for crystal structures resolved at a resolution of 2.5 Å (Table 2–2). By using Errat, the reliability of the CYP11B model folding and packing can be investigated and compared to the template structures to determine structural disparities. Verify3D determines the environment for each amino acid and calculates the probability to encounter the amino acid in the environment based on statistical data derived from 16 reliable protein structures (and several related protein structures) from the Brookhaven Protein Databank. A score is assigned for each amino acid and the results are smoothed using a 21 amino acid sliding window. A Verify3D total score is supplied to determine the overall matching, however, the individual scoring portrays the outliers.

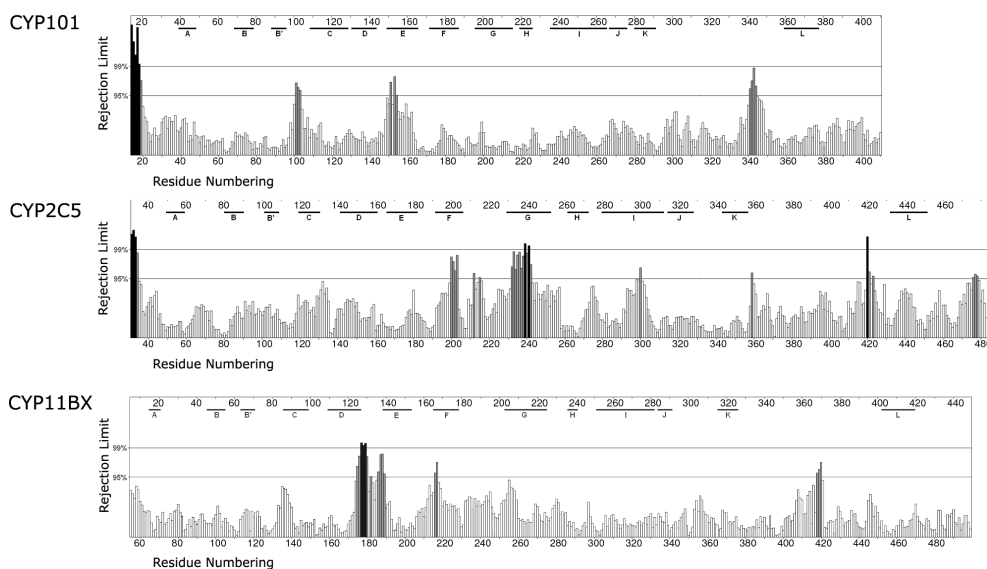


Figure 2–8 Errat results for the template structures of CYP2C5 (1NR6) and CYP101 (2CPP) as well as for the CYP11B2 model. Crossing of the rejection limit for the CYP11B2 model are observed for the region between helices D and E, as well as at the start of helix F. Other high values coincide with those observed for either CYP2C5 or CYP101.

By comparing the results to the data gathered from the Ramachandran plot, it can be seen that the models all score equally well but are less accurate than the template structures (Table 2–2). Using the Errat score per amino acid, the structural deviations have been located in the loop regions at the end of helix E where the two template structures connect and at the end of helix D where a large insertion was introduced (Figure 2–8). Outliers in helices F and G of CYP2C5 are expected since these helices interact with the membrane and several point mutants have been introduced to make this region soluble. Because CYP2C5 has been used as a template, the CYP11B models also possess structural deviations, however, they are found to be less severe.

Investigation of the amino acid environment with Verify3D resulted in similar conclusions as found with Errat and the Ramachandran data. The low scoring of Verify3D can be attributed to a bad folding of the regions around the helices D and F, and additional potential errors were located at the end of helix G, where again the two templates have been connected to each other (Figure 2–9). The low Verify3D score at the beginning of helix L is caused by the linking of the heme group, which is not included in the Verify3D scoring function.

Finally, the stereochemistry and the amino acid conformation of the models was analysed with MOE [70]. From Table 2–3 it can be concluded that the amino acid conformation of the backbone (omega, phi and psi characteristics), the side chains (chirality and alkyl-gauche characteristics) are comparable to the general conformation as found in crystal structures (reference values). No flaws were detected by these measurements for the conformation of the amino acids in the protein models, indicating that the overall conformation of the protein is of good quality.

The occurrence of Ramachandran errors and problems with modelling external loop regions seems an inevitable circumstance in homology modelling [71,72] and will probably not pose a problem for modelling protein-ligand interactions, since these regions do not interact with the active site. It is also anticipated that connecting several template structures to another introduces structural flaws at those amino acids. During the modelling, the connection of the template regions has been performed in regions far from the active site centre, as to not introduce structural flaws in the protein active site. All the above-mentioned errors are expected to be alleviated during further investigation of protein-ligand interactions using molecular dynamics simulations, and the current models have been selected as appropriate starting points for further analyses.

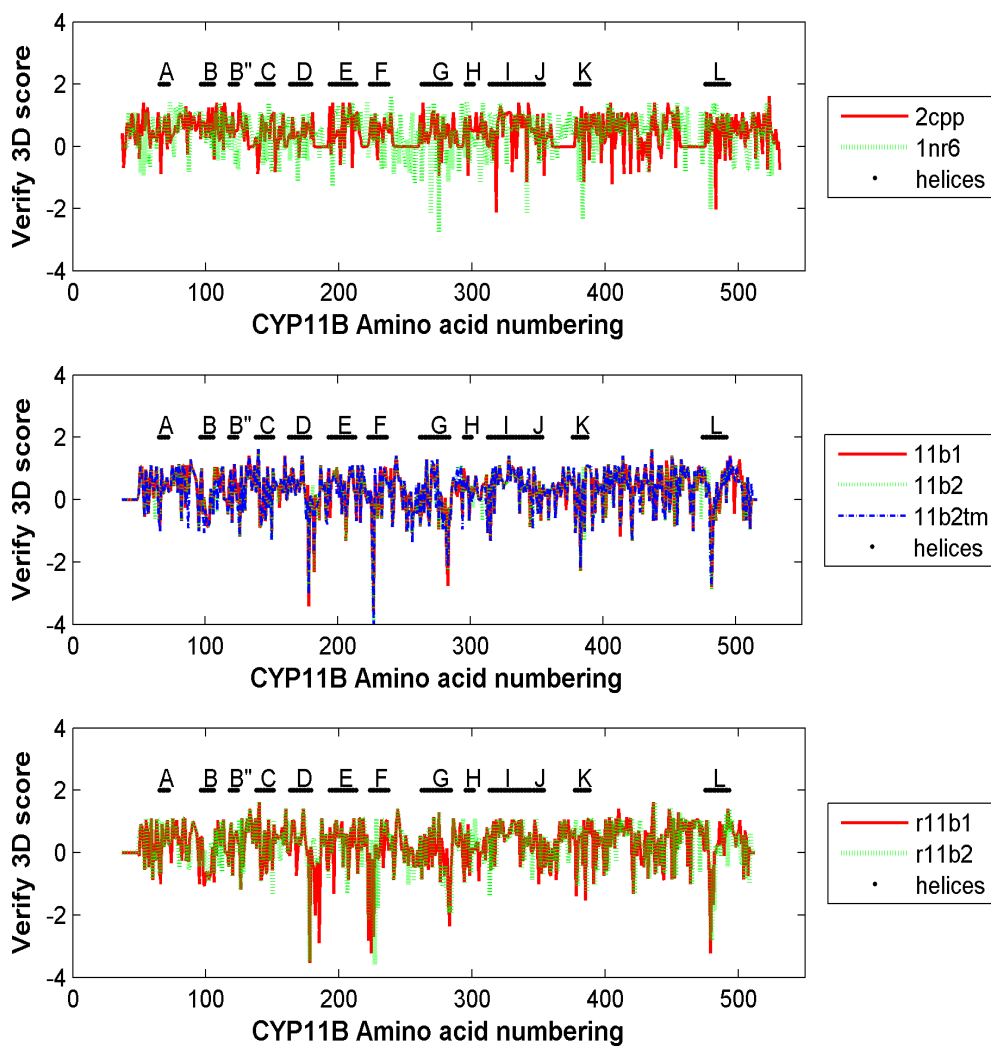


Figure 2-9 Verify3D results for the template structures of CYP2C5 (1NR6) and CYP101 (2CPP) as well as for the different CYP11B models. Unreliability for the CYP11B models is observed in the regions between helix D and E, the start of helix F and at the end of helix G.

Table 2–3 Validation results for the homology models of CYP11B1 and CYP11B2, part II.

MOE-protein report	CYP101 (2CPP)		CYP2C5 (1NR6)		hCYP11B1 (model)		hCYP11B2 (model)		Reference Values	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Trans-Omega dihedral	176.6	2.7	178.9	0.9	172.0	7.7	172.0	7.7	180.0	5.8
Helix Phi	-65.2	11.9	-67.4	15.5	-60.7	19.7	-61.1	19.8	-65.3	11.9
Helix Psi	-41.2	16.5	-37.6	17.9	-42.5	25.1	-41.9	25.2	-39.4	11.3
Proline Phi	-65.8	11.3	-61.9	9.5	-65.3	19.6	-67.6	20.5	-65.4	11.2
Ca chirality	32.8	3.6	34.3	1.7	30.8	11.1	30.8	10.8	33.8	4.2
Chi1-gauche minus	-63.0	17.4	-63.3	14.8	-62.3	21.7	-62.6	22.1	-66.7	15.0
Chi1-gauche plus	55.4	20.7	56.3	16.4	51.5	26.7	53.0	28.3	64.1	15.7
Chi1-trans	185.3	13.3	184.2	12.9	186.6	21.9	186.8	20.3	183.6	16.8
Dihedral outliers	0		4		15		17			
Bond Angle outliers	0		0		3		4			
Bond Length outliers	0		0		0		1			

Results were generated with the MOE module: Protein Eval. The thresholds were chosen as 5 for the Z-score and 70 for van der Waals contacts. Reference values were published in a statistical survey of the high-resolution data in the Protein Data Bank [70]

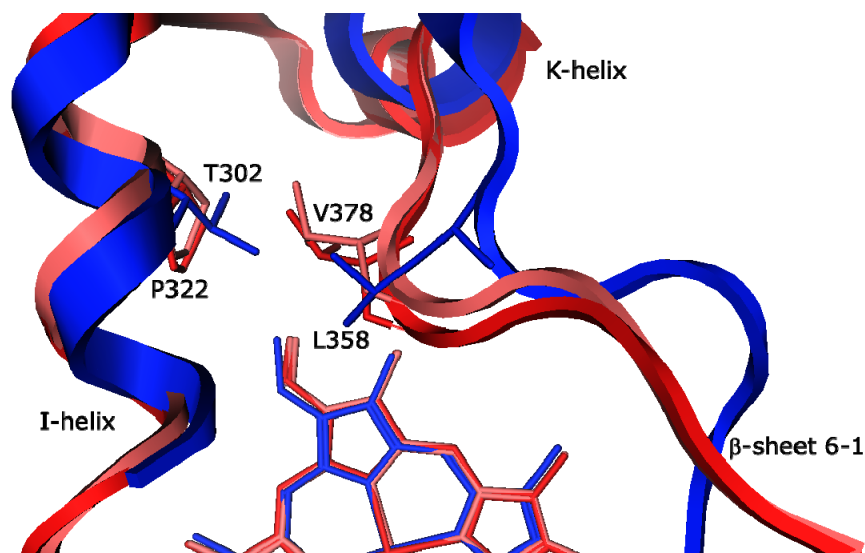


Figure 2–10 Active site difference between the hCYP11B1 (red) and hCYP11B2 (pink) models. Indicated are P322 and V378 for hCYP11B1 and hCYP11B2, as well as the corresponding amino acid residues of CYP2C5 (blue), T302 and L358 respectively. As is clear from the picture, the connection between helix I and the loop following helix K is much closer for the hCYP11B models.

### 2.7.2 CYP11B1 and CYP11B2 model active site differences

There are two striking differences between the active sites of the CYP11B1 and CYP11B2 models. Firstly, if both species are compared, the active site cavity near beta-sheet 6-1 is smaller in both CYP11B2 models compared to the CYP11B1 models. The rCYP11B1 active site near this region is the largest. The cavity that creates the differences between the CYP11B models is formed by the loop region between helix K and beta-sheet 6-1 (Figure 2–10, Figure 2–11). Comparing this cavity in both the hCYP11B1 and hCYP11B2 models, we find that in hCYP11B1 this cavity has a 9–14% larger volume, dependent on amino acid flexibility as well as the probe radius chosen for calculation of the volume (differing from 1.0 to 1.5 Å). This difference in size is caused by the folding of L407 which induces a 1.0 Å outward shift of the loop region. A comparison of the models with the crystal structures of the CYP2 family reveals that the loop region in the CYP11B models is in closer proximity to helix I (by 1.5 Å). This contact is defined by two amino acids having relatively smaller side chains (P322 and V378 in CYP11B) than observed in the CYP2C family (generally threonine or valine and leucine or isoleucine, T302 and L358 in CYP2C5, Figure 2–10). Changes in helix I such as the A320V between hCYP11B1 and hCYP11B2 are likely to have a direct influence on the folding of this region.

Secondly, helix B' is shifted outward by 1.1 Å in both CYP11B2 models (the backbone RMSD of helix B' of hCYP11B1 and hCYP11B2 is 2.13 Å). The change in spatial positioning of this helix is probably caused by L301P and E302D in helix I, but because this region is on the surface of the protein, the exact cause is less clear. In the hCYP11B2 model, E302 contacts helix B' at K127, which forms the counter charge for the C-terminal end of this helix. Hasemann et al have posed that movement of helices B', F and G is involved in the opening of the active site cavity [16]. In conclusion, changing the stabilising environment of helix B' has a likely effect on the structural stability of the helix and may result in a loss of activity.

### 2.7.3 Protein - Substrate interactions

After quality assessment of the constructed protein models encompassing the various ligands, and investigation of the protein model differences, the different steroid binding modes have been investigated for each of the models. All docking results from GOLD favoured the  $\beta$ -side of the steroid oriented to the heme with the C<sub>3</sub>-carbonyl group pointing towards helix B' (Figure 2–11). For details on the docking settings and docking results, see paragraphs 4.4 and 4.6.1. Visual inspection revealed that in case of reverse orientation of the steroid, unfavourable clashes would occur for the C<sub>3</sub>-carbonyl group in the opposite side of the pocket.

All ligands showed two very distinct interactions in the modelled active site cavities. Firstly, the ligands possess a steric fit for the C<sub>20</sub>-carbonyl group and the C<sub>21</sub>-hydroxyl group in a small cavity created between helix K and beta-sheet 6-1 (Figure 2–11). Inside this cavity, the C<sub>21</sub>-hydroxyl group possesses two hydrogen bonding interactions with the protein backbones of G379 and F381. The presence of these amino acids in the active site cavity coincides with the models of Belkina et al., but for those models, no interactions between protein and ligand were discussed [37].

Secondly, the ligands possess an interaction between the C<sub>3</sub>-carbonyl group and active site residue R123 in helix B'. R123 is stabilised by E310 in helix I, which has further stabilising interactions with the protein backbone (Figure 2–11). E310 coincides with an aspartic acid in the CYP2 family which from visual inspection of the crystal structures of CYP2C5 (1NR6) and CYP2C9 (1OG5) seems to play a specific stabilising role for helix B'. Therefore, it is expected that E310 has the same stabilising role. The difference in chain length between an aspartic acid and a glutamic acid also determines the flexibility of helix B'. It can move 1.5 Å further out of the active site cavity in all CYP11B models, which creates more room and allows the steroid to fit parallel to the heme. The ligands also possess many hydrophobic interactions in this region, particularly with F130 (Figure 2–11). Due to the close interactions with the A and B rings of the steroid skeleton, this amino acid might play an important role in substrate stabilisation. The presence of F130 in our models coincides with the models of Ulmschneider et al. [38] where it is seemingly involved in ring stacking with their inhibitors. After docking the steroids into the active site, the protein–ligand complexes were subjected to minimisation with the Charmm22 force field. After minimisation, the distance between the heme iron atom and the different substrate hydroxylation sites (C<sub>11</sub>, C<sub>18</sub> and C<sub>19</sub>) was measured (Table 2–4).

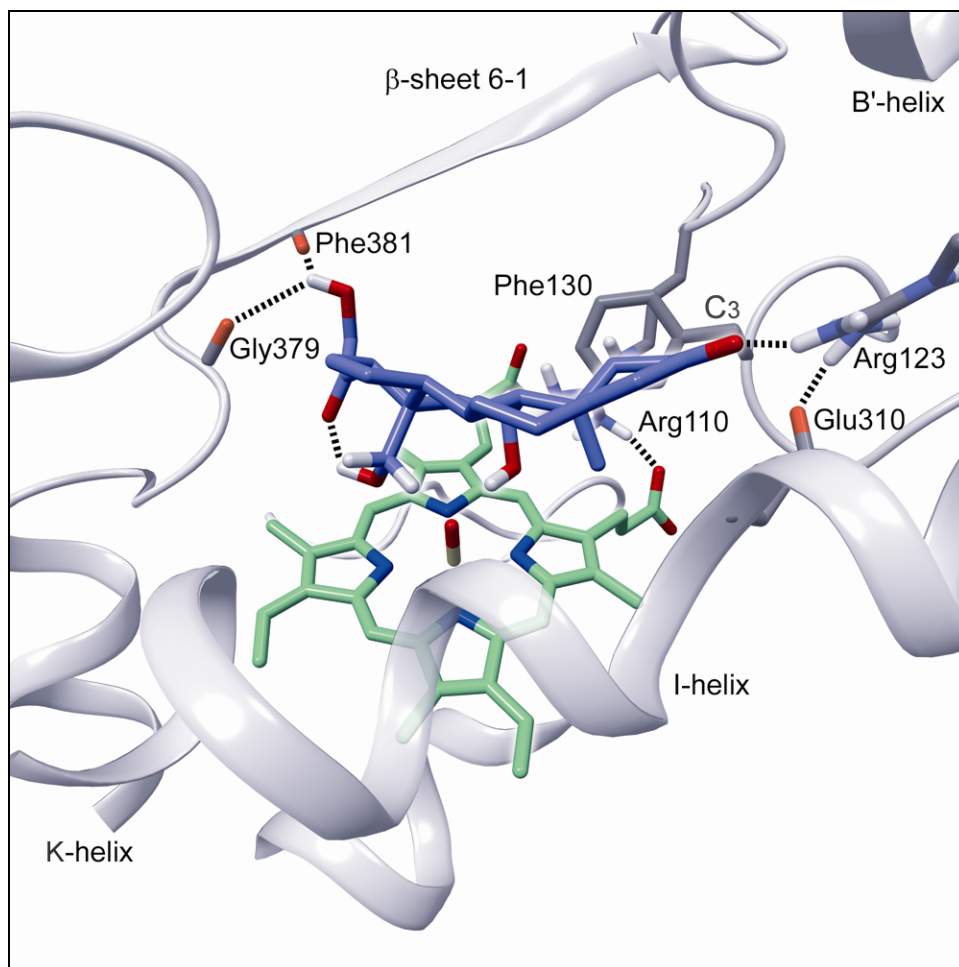


Figure 2–11 Hypothetical binding of 18-hydroxycorticosterone (18OH-B) inside the active site of CYP11B2 for the synthesis of aldosterone. The heme prosthetic group contains a bound oxygen atom needed for catalytic function. The CYP11B2 model features R123 in helix B' that is stabilised by E310 in helix I. 18OH-B is stabilised by several hydrogen bonds in the active site: one internal hydrogen bond between the C<sub>18</sub>-hydroxyl group and the C<sub>20</sub>-carbonyl group, two hydrogen bonds between the C<sub>21</sub>-hydroxyl group and the backbone carbonyl groups of G379 and F381, and finally a hydrogen bond between the C<sub>3</sub>-carbonyl and R123.

Table 2–4 Hydroxylation distance table (iron atom-carbon atom) after minimisation with MOE (distances are in Angstrom)

Ligand	hCYP11B1			hCYP11B2-TripMut			hCYP11B2			rCYP11B1			rCYP11B2		
	C11	C18	C19	C11	C18	C19	C11	C18	C19	C11	C18	C19	C11	C18	C19
DOC	4.30	4.56	5.48	4.37	4.65	5.32	4.72	4.30	5.61	4.30	4.75	4.83	4.70	4.24	5.54
18OH-DOC	4.31	4.60	5.19 b	4.31	4.51	5.21 b	4.33	4.30	5.42 a	4.30	4.68	5.17 b	4.32	4.31	5.39 a
B	5.43	4.39	5.28	5.37	4.40	5.22	5.39	4.06	5.46	5.33	4.49	4.94	5.28	4.21	5.20
18OH-B	5.38	4.62	5.26 d	5.42	4.64	5.29 c	4.86	4.21	5.50 a	5.47	4.62	5.28 d	5.29	4.35	5.29 a

The ligands: DOC (11-deoxycorticosterone), B (corticosterone), 18OH-DOC (18-hydroxycorticosterone) and 18OH-B (18-hydroxycorticosterone).

- a) Ligand C18-hydroxyl group forms a hydrogen bond with the C20-ketone group of the ligand
- b) Ligand C18-hydroxyl group forms a hydrogen bond with the iron-oxygen active species of the protein
- c) Ligand C18-hydroxyl group forms a hydrogen bond with the C11-hydroxyl group of the ligand
- d) Ligand C11-hydroxyl group forms a hydrogen bond with the C18-hydroxyl group of the ligand

### 2.7.3.1 DOC

For DOC, hCYP11B1 and rCYP11B1 preferentially catalyse C<sub>11</sub>-hydroxylation, whereas hCYP11B2 and rCYP11B2 preferentially catalyse C<sub>18</sub>-hydroxylation. For both CYP11B2 models, the shifted alpha-helix B' causes DOC to present its C<sub>18</sub> closest to the iron as it is repositioned by the strong interaction of the C<sub>3</sub>-carbonyl group with R123 in alpha-helix B'. Additionally, the larger active site of rCYP11B1 around beta-sheet 6-1 allows DOC to fit further into the niche presenting C<sub>19</sub> into hydroxylation range (distance less than 5.0 Å, Table 2–4).

One argument against a preferred C<sub>18</sub>-hydroxylation of DOC by CYP11B2 is that *in vitro* measurements indicate higher levels of the C<sub>11</sub>-hydroxylated product B [73,74]. Therefore, it is questionable what the true regio-selectivity of CYP11B2 really is, and whether it is in fact a C<sub>11</sub>,C<sub>18</sub>-hydroxylase. However, in the *in vitro* cells, both C<sub>11</sub>- and C<sub>18</sub>-hydroxylated products of DOC can be promptly consumed as a substrate for the production of 18OH-B and subsequently aldosterone [73,74]. This apparent discrepancy between observed and predicted regio-selectivity of DOC hydroxylation by CYP11B2 may possibly indicate that other factors than hydroxylation distances are involved in the formation of the actual products, such as active site influences and charge distributions.

### 2.7.3.2 18OH-DOC

For 18OH-DOC, the C<sub>18</sub>-hydroxyl group forms a hydrogen bond with the iron–oxygen limiting hydroxylation to C<sub>11</sub> which is in closest proximity to the iron. In the CYP11B2 models both C<sub>11</sub>- and C<sub>18</sub> are in approximately equal distance to the iron and an internal hydrogen bond is formed by the substrate between the C<sub>18</sub>-hydroxyl group and the C<sub>20</sub>-carbonyl group. Although both C<sub>11</sub>- and C<sub>18</sub>-hydroxylation are shown to be possible in such a complex, oxidation on the unsubstituted C<sub>11</sub> is likely to form a more stable reaction intermediate, since the C<sub>18</sub>,<sub>18</sub>-dihydroxylated product has not been determined in *in vitro* experiments [73,74].



### 2.7.3.3 B and 18OH-B

In all models, the two C<sub>11</sub>-hydroxylated ligands B and 18OH-B only portray C<sub>18</sub> in close proximity to the iron, as the C<sub>11</sub>-hydroxyl group blocks access of C<sub>11</sub> to the heme iron. The positioning of the C<sub>18</sub>-hydroxyl group of 18OH-B in the active site cavity appears to determine conversion into aldosterone. In both CYP11B2 models, the natural substrate is shifted above the heme, which creates a slightly larger active site cavity near beta-sheet 6-1 for the C<sub>18</sub>-hydroxyl group to rotate in. This difference in size allows an internal hydrogen bond between the C<sub>18</sub>-hydroxyl group and the C<sub>20</sub>-carbonyl group for both 18OH-DOC and 18OH-B. In rCYP11B1 and hCYP11B1, the C<sub>18</sub>-hydroxyl group forms a hydrogen bond with the heme oxygen atom which blocks C<sub>18</sub> for hydroxylation. Interestingly, in the hCYP11B2-TripMut, the C<sub>18</sub>-hydroxyl group forms an internal hydrogen bond with the C<sub>11</sub>-hydroxyl group, and thereby also blocks the C<sub>18</sub> for hydroxylation.

### 2.7.4 Triple-Mutant Influence

Introducing the triple mutant investigated by Bernhardt et al. [62] by means of the hCYP11B2-TripMut model showed that the hydroxylation distance pattern of the substrates shifts as expected from hCYP11B2 to that of hCYP11B1 (Table 2–4). Both hCYP11B1 and hCYP11B2-TM display similar active site cavities near beta-sheet 6-1 (backbone RMSD 1.43 Å) with only slight deviations in the hydroxylation distances for the ligands. This confirms the behaviour of the triple mutant in enzymatic activity found by the study of Bernhardt et al.

To speculate further, Bernhardt et al have shown that the A320V mutant alone or the L301P/E302D mutant alone, is not enough to significantly change aldosterone synthesis [62]. Only mutation in both regions results in an almost complete loss of the aldosterone synthesis capabilities of hCYP11B2. It is likely that the subtle changes on both sides of the active site go hand-in-hand to (nearly) completely convert the activity of hCYP11B2 into that of hCYP11B1.

### 2.7.5 Proposed Synthesis Mechanism of Aldosterone

Comparing the hydroxylation distances and the internal hydrogen bonds observed for the protein-substrate interactions, a generic consensus can be derived for the conversion of 18OH-B. In the rCYP11B1, hCYP11B1 and hCYP11B2-TripMut models, the C<sub>18</sub> has been blocked from further oxidation due to hydrogen bonds formed between the C<sub>18</sub>-hydroxyl group and an acceptor moiety in the catalytic centre of the proteins. However, for the rCYP11B2 and hCYP11B2 models, an internal hydrogen bond between this C<sub>18</sub>-hydroxyl group and the C<sub>20</sub>-carbonyl group presents another hydrogen atom to the iron-oxygen species, ready for hydrogen abstraction.

Therefore, we propose that the immediate folding of the active site around the substrate C<sub>18</sub>-hydroxyl group provides the key difference between the two isoforms leading to the production of aldosterone by only CYP11B2 and not CYP11B1. The formation of a stable internal hydrogen bond for the C<sub>18</sub>-hydroxyl group of the ligand may be essential to stabilise its C<sub>18</sub> for further oxidation. Subsequently, a second oxygen atom can be inserted between

C<sub>18</sub> and one of its hydrogen atoms, forming a C<sub>18</sub>-gem-diol intermediate which forms aldosterone by elimination of water.

The possible formation of a C<sub>18</sub>-gem-diol has been proposed before, yet the exact mechanism of oxidation is still unknown [63,75,76]. This proposed method of substrate stabilisation may explain the regio-selectivity of the two CYP11B isoforms and can be attributed to the subtle changes observed on two sides of the active site cavity; the cavity near helix K and the positioning of helix B'.

## 2.8 Conclusions

We have constructed homology models of the two isoforms of the human CYP11B family, as well as promising models for the rat isoforms. Several amino acid residues and active site regions have been identified to play a role in the stabilisation of the active site pocket. On one side the active site is enclosed by helix B' that is stabilised on the inside by E310 in helix I and on the outside by P310 and D302 in hCYP11B1, and L310 and E302 in hCYP11B2. The difference in the residues on the outside of helix B' allows it to slightly shift inward or outward and determines part of the difference between the active site cavity of hCYP11B1 and hCYP11B2. The other side of the active site is enclosed by a loop region following helix K. This helix is interacting tightly with helix I through a hydrophobic amino acid complex involving V378 in the loop structure and P322 in helix I. The main difference between hCYP11B1 and hCYP11B2 in the amino acids lining this region is V320 in hCYP11B1 and A320 in hCYP11B2.

The most important interactions for the natural substrates in the active site cavity connect to both of these active site enclosing regions. The substrates are stabilised by a hydrogen bonding interaction between the substrate C<sub>3</sub>-carbonyl group with R123 in helix B', a hydrogen bonding interactions of the C<sub>21</sub>-hydroxyl with the backbone residues of E379 and F381 in the loop region following helix K, and the presence of F130 in helix B'.

Since changes in the amino acids directly lining both sides of the active site cavity are able to change the oxidation preference of the enzymes [62], we have investigated this by substrate minimisation study in the active site. To accomplish this, we have defined the substrate specificity as the differences in hydroxylation distances between the active site and the C<sub>11</sub>, C<sub>18</sub> and C<sub>19</sub> of the steroid skeleton. Since both hCYP11B1 and hCYP11B2 possess similar active sites, the steric fit of the different steroidal ligands seems to be one of the strongest determinants for the substrate specificity.

From our calculations, we have found that within the active site of the hCYP11B2 model, the endogenous ligand 18-hydroxycorticosterone forms a stabilising internal hydrogen bond. This is not observed for the hCYP11B1 model, or the hCYP11B2-TripMut model that possesses hCYP11B1 activity. The absence of the stabilisation of the C<sub>18</sub>-hydroxyl group in the hCYP11B1 model might rationalise why 18-hydroxycorticosterone is solely a substrate for CYP11B2 to yield aldosterone.

## Literature

- 1 K.N. Degtyarenko, T.A. Kulikova, "Evolution of bioinorganic motifs in P450-containing systems", *Biochemical Society Transactions*, 2001, 29, 2, 139-146
- 2 D.W. Nebert, D.W. Russell, "Clinical importance of the cytochromes P450", *Lancet*, 2002, 360, 9340, 1155-1162
- 3 D.W. Nebert, M. Adesnik, M.J. Coon, R.W. Estabrook, F.J. Gonzalez, F.P. Guengerich, I.C. Gunsalus, E.F. Johnson, B. Kemper, W. Levin, I.R. Phillips, R.Sato, M.R. Waterman, "The P450 gene superfamily: recommended nomenclature", *DNA*, 1987, 6, 1, 1-11
- 4 D.W. Nebert, D.R. Nelson, M. Adesnik, M.J. Coon, R.W. Estabrook, F.J. Gonzalez, F.P. Guengerich, I.C. Gunsalus, E.F. Johnson, B. Kemper, W. Levin, I.R. Phillips, R. Sato, M.R. Waterman, "The P450 superfamily: updated listing of all genes and recommended nomenclature for the chromosomal loci", *DNA*, 1989, 8, 1, 1-13
- 5 D.W. Nebert, D.R. Nelson, M.J. Coon, R.W. Estabrook, R. Feyereisen, Y. Fujii-Kuriyama, F.J. Gonzalez, F.P. Guengerich, I.C. Gunsalus, E.F. Johnson, E.F. Loper, R. Sato, M.R. Waterman, D.J. Waxman, "The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature", *DNA Cell Biol.*, 1991, 10, 1, 1-14
- 6 D.R. Nelson, T. Kamataki, D.J. Wasman, F.P. Guengerich, R.W. Estabrook, R. Feyereisen, F.J. Gonzalez, M.J. Coon, I.C. Gunsalus, O. Gotoh, K. Okuda, D.W. Nebert, "The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature", *DNA Cell Biol.*, 1993, 12, 1, 1-51
- 7 D.R. Nelson, L. Koymans, T. Kamataki, J.J. Stegeman, R. Feyereisen, D.J. Waxman, M.R. Waterman, O. Gotoh, M.J. Coon, R.W. Estabrook, I.C. Gunsalus, D.W. Nebert, "P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature", *Pharmacogenetics*, 1996, 6, 1, 1-42
- 8 D.R. Nelson, D.C. Zeldin, S.M. Hoffman, L.J. Maltais, H.M. Wain, D.W. Nebert, "Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants", *Pharmacogenetics*, 2004, 14, 1, 1-18
- 9 P.S. Stayton, T.L. Poulos, S.G. Sligar, "Putidaredoxin competitively inhibits cytochrome b5-cytochrome P-450cam association: a proposed molecular model for a cytochrome P-450cam electron-transfer complex", *Biochemistry*, 1989, 28, 20, 8201-8205
- 10 A. bridges, L. Gruenke, Y.T. Chang, I.A. Vakser, G. Loew, L. Waskell, "Identification of the binding site on cytochrome P450 2B4 for cytochrome b5 and cytochrome P450 reductase", *J. Biol. Chem.*, 1998, 273, 27, 17036-17049
- 11 H. Shimada, S. Nagano, H. Hori, Y. Ishimura, "Putidaredoxin-cytochrome P450cam interaction", *J. Inorg. Biochem.*, 2001, 83, 4, 255-260
- 12 O. Gotoh, "Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analysis of amino acid and coding nucleotide sequences", *J. Biol. Chem.*, 1992, 267, 1, 83-90
- 13 D.F. Lewis, E. Watson, B.G. Lake, "Evolution of the cytochrome P450 superfamily: sequence alignments and pharmacogenetics", *Mutat. Res.*, 1998, 410, 3, 245-270
- 14 J.A. Peterson, S.E. Graham, "A close family resemblance: the importance of structure in understanding cytochromes P450", *Structure*, 1998, 6, 9, 1079-1085
- 15 J.A. Peterson, S.E. Graham, "How similar are P450s and what can their differences teach us", *Arch. Biochem. Biophys.*, 1999, 369, 1, 24-29
- 16 C.A. Hasemann, R.G. Kurumbail, S.S. Boddupalli, J.A. Peterson, J. Deisenhofer, "Structure and function of cytochromes P450: a comparative analysis of 3 crystal-structures", *Structure*, 1995, 3, 1, 41-62
- 17 T.L. Poulos, R. Raag, "Cytochrome P450cam: crystallography, oxygen activation, and electron transfer", *FASEB J.*, 1992, 6, 2, 674-679
- 18 T.L. Poulos, B.C. Finzel, A.J. Howard, "High-resolution crystal structure of cytochrome P450cam", *J. Mol. Biol.*, 1987, 195, 3, 687-700
- 19 I.F. Sevrionkova, H. Li, H. Zhang, J.A. Peterson, T.L. Poulos, "Structure of a cytochrome P450-redox partner electron-transfer complex", *Proc. Natl. Acad. Sci. U.S.A.*, 1999, 96, 5, 1863-1868

- 20 J.K. Yano, L.S. Koo, D.J. Schuller, L. Hi, P.R. Ortizz de Montellano, T.L. Poulos, "Crystal structure of a thermophilic cytochrome P450 from the archaeon *sulfolobus solfataricus*", *J. Biol. Chem.*, 2000, 275, 40, 31086-31092
- 21 J. Cosme, E.F. Johnson, "Engineering microsomal cytochrome P450 2C5 to be a soluble, monomeric enzyme: mutations that alter aggregation, phospholipid dependence of catalysis, and membrane binding", *J. Biol. Chem.*, 2000, 275, 4, 2545-2553
- 22 P.A. Williams, J. Cosme, V. Sridhar, E.F. Johnson, D.E. McRee, "Mammalian microsomal cytochrome 450 monooxygenase: structural adaptations for membrane binding and functional diversity", *Mol. Cell.*, 2000, 5, 1, 121-131
- 23 M.R. Wester, E.F. Johnson, C. Marques-Soares, S. Dijols, P.M. Dansette, D. Mansuy, C.D. Stout, "Structure of mammalian cytochrome P450 2C5 complexed with doclofenac at 2.1 Å resolution: evidence for an induced fit model of substrate binding", *Biochemistry*, 2003, 42, 31, 9335-9345
- 24 P.A. Williams, J. Cosme, A. Ward, H.C. Angove, D. Matak Vinkovic, H. Jhoti, "Crystal structure of human cytochrome P450 2C9 with bound warfarin", *Nature*, 2003, 424, 6947, 464-468
- 25 P. Rowland, F.E. Blaney, M.G. Smyth, J.J. Jones, V.R. Leydon, A.K. Oxbrow, C.J. Lewis, M.G. Tennant, S. Modi, D.S. Eggleston, R.J. Chenery, A.M. Bridges, "Crystal structure of human cytochrome P450 2D6", *J. Biol. Chem.*, 2006, 281, 11, 7614-7622
- 26 J.K. Yano, M.R. Wester, G.A. Schoch, K.J. Griffin, C.D. Stout, E.F. Johnson, "The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution", *J. Biol. Chem.*, 2004, 279, 37, 38091-38094
- 27 F.C. Bernstein, T.F. Koetzle, G.J. Williams, E.F. Meyer Jr, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, "The protein data bank. a computer-based archival file for macromolecules", *Eur. J. Biochem.*, 1977, 80, 2, 319-324
- 28 Y.-T. Chang, O.B. Stiffelman, I.A. Vakser, G.H. Loew, A. Bridges, L. Waskell, "Construction of a 3D model of cytochrome P450 2B4", *Protein. Eng.*, 1997, 10, 2, 119-129
- 29 L.M.H. Koymans, N.P.E. Vermeulen, A. Baarslag, G.M. Donne-Op den Kelder, "A preliminary 3D model for cytochrome P450 2D6 constructed by homology model building", *J. Comput.-Aided. Mol. Des.*, 1993, 7, 3, 281-289
- 30 A. Cavalli, G. Greco, E. Novellino, M. Recanatini, "Linking CoMFA and protein homology models of enzyme-inhibitor interactions: an application to non-steroidal aromatase inhibitors", *Bioorg. Med. Chem.*, 2000, 8, 12, 2771-2780
- 31 J.J. Lozano, E. Lopez-de-Brinas, N.B. Centeno, R. Guigo, F. Sanz, "Three-dimensional modelling of human cytochrome P450 1A2 and its interaction with caffeine and MeIQ", *J. Comput.-Aided. Mol. Des.*, 1997, 11, 4, 395-408
- 32 M. Recanatini, A. Cavalli, P. Valenti, "Nonsteroidal aromatase inhibitors: recent advances", *Med. Res. Rev.*, 2002, 22, 3, 282-304
- 33 S.A. Usanov, S.E. Graham, G.I. Lepesheva, T.N. Azeva, N.V. Strushkevich, A.A. Gilep, R.W. Estabrook, J.A. Peterson, "Probing the interaction of bovine cytochrome P450<sub>scc</sub> (CYP11A1) with adrenodoxin: evaluating site-directed mutations by molecular modeling", *Biochemistry*, 2002, 41, 26, 8310-8320
- 34 N. Mast, S.E. Graham, U. Andersson, I. Bjorkhem, C. Hill, J. Peterson, I.A. Pikuleva, "Cholesterol binding to cytochrome P450 7A1, a key enzyme in bile acid biosynthesis", *Biochemistry*, 2005, 44, 9, 3259-3271
- 35 N. Mast, D. Murtazina, H. Liu, S.E. Graham, I. Bjorkhem, J.R. Halpert, J. Peterson, I.A. Pikuleva, "Distinct binding of cholesterol and 5β-cholestane-3α,7α,12α-triol to cytochrome P450 27A1: evidence from modeling and site-directed mutagenesis studies", *Biochemistry*, 2006, 45, 14, 4396-4404
- 36 M.S. Gomaa, C. Simons, A. Brancale, "Homology model of 1α,25-dihydroxyvitamin D3 24-hydroxylase cytochrome P450 24A1 (CYP24A): active site architecture and ligand binding", *J. Steroid. Biochem. Mol. Biol.*, 2007, 104, 1-2, 53-60
- 37 N.V. Belkina, M. Lisurek, A.S. Ivanov, R. Bernhardt, "Modelling of three-dimensional structures of cytochromes P450 11B1 and 11B2", *J. Inorg. Biochem.*, 2001, 87, 4, 197-207
- 38 S. Ulmschneider, U. Muller-Vieira, M. Mitrenga, R.W. Hartmann, "Synthesis and evaluation of imidazolylmethylentetrahydronaphthalenes and imidazolylmethyleneindanes: potent inhibitors of aldosterone synthase", *J. Med. Chem.*, 2005, 48, 6, 1796-1805

- 39 A. Bairoch, B. Boeckmann, "The SWISS-PROT protein sequence data bank", *Nucleic. Acids. Res.*, 1991, 19, S2247-2249
- 40 J.A. Cuff, M.E. Clamp, A.S. Siddiqui, M. Finlay, J.G. Barton, "JPred: a consensus secondary structure prediction server", *Bioinformatics*, 1998, 14, 10, 892-893
- 41 MOE (The Molecular Operating Environment) Version 2007.08, Chemical Computing Group Inc., Sherbrooke Street West, Suite 910, Montreal Canada H3A 2R7. <http://chemcomp.com>
- 42 A.V. Grinberg, F. Hanneman, B. Schiffer, J. Muller, U. Heinemann, R. Bernhardt, "Adrenodoxin: structure, stability, and electron transfer properties", *Proteins*, 2000, 40, 4, 590-612
- 43 S. Vijayakumar, J.C. Salerno, "Molecular modeling of the 3-D structure of cytochrome P-450sc", *Biochim. Biophys. Acta.*, 1992, 1160, 3, 281-286
- 44 A. Seifert, S. Tatzel, R.D. Schmidt, J. Pleiss, "Multiple molecular dynamics simulations of human P450 monooxygenase CYP2C9: the molecular basis of substrate binding and regioselectivity toward warfarin", *Proteins*, 2006, 64, 1, 147-155
- 45 T. Hiroi, W. Kishimoto, T. Chow, S. Imaoka, T. Igarashi, Y. Funae, "Progesterone oxidation by cytochrome P450 2D isoforms in the brain", *Endocrinology*, 2001, 142, 9, 3901-3908
- 46 T.L. Domanski, J. Liu, G.R. Harlow, J.P.R. Halpert, "Analysis of four residues within substrate recognition site 4 of human cytochrome P450 3A4: role in steroid hydroxylase activity and alpha-naphthoflavone stimulation", *Arch. Biochem. Biophys.*, 1998, 350, 2, 223-232
- 47 S. Ekins, M.J. de Groot, J.P. Jones, "Pharmacophore and three-dimensional quantitative structure activity relationship methods for modelling cytochrome P450 active sites", *Drug Metab. Dispos.*, 2001, 29, 7, 936-944
- 48 B.E. Daikh, J.M. Lasker, J.L. Raucy, D.R. Koop, "Regio- and stereoselective epoxidation of arachidonic acid by human cytochromes P450 2C8 and 2C9", *J. Pharmacol. Exp. Ther.*, 1994, 271, 3, 1427-1433
- 49 P.C. White, J. Dupont, M.I. New, E. Leiberman, Z. Hochberg, A. Rosler, "A mutation in CYP11B1 (Arg-448-His) associated with steroid 11 beta-hydroxylase deficiency in Jews of Moroccan origin", *J. Clin. Invest.*, 1991, 87, 5, 1664-1667
- 50 K. Joehrer, S. Geley, E.M. Strasser-Wozak, R. Azziz, H.A. Wollmann, K. Schmitt, R. Kofler, P.C. White, "CYP11B1 mutations causing non-classic adrenal hyperplasia due to 11 beta-hydroxylase deficiency", *Hum. Mol. Genet.*, 1997, 6, 11, 1829-1834
- 51 S. Nomoto, G. Massa, F. Mitani, Y. Ishimura, K. Miyahara, K. Toda, I. Nagano, T. Yamashiro, S. Ogoshi, J.-I. Fukata, S. Onishi, K. Hashimoto, Y. Doi, H. Imura, Y. Shizuta, "CMO I deficiency caused by a point mutant in exon 8 of the human CYP11B2 gene encoding steroid 18-hydroxylase (P450C18)", *Biochemical and Biophysical Research Communications*, 1997, 234, 2, 382-385
- 52 L. Pascoe, K.M. Curnow, L. Slutsker, A. Rosler, P.C. White, "Mutations in the human CYP11B2 (aldosterone synthase) gene causing corticosterone methyl oxidase 2 deficiency", *Proc. Natl. Acad. Sci. USA*, 1992, 89, 11, 4996-5000
- 53 M. Peter, K. Bunger, J. Solyom, G. Sippell, "Mutation Thr-185 Ile is associated with corticosterone methyl oxidase deficiency type II", *Eur. J. Pediatr.*, 1998, 157, 5, 378-381
- 54 S. Portrat-Doyen, J. Tourniaire, O. Richard, P. Mulatero, B. Aupetit-Faisant, K.M. Curnow, L. Pascoe, Y. Morel, "Isolated aldosterone synthase deficiency caused by simultaneous E198D and V386A mutations in the CYP11B2 gene", *J. Clin. Endocrinol. Metab.*, 1998, 83, 11, 4156-4161
- 55 F.M. Dunlop, P.A. Crock, J. Montalto, J.W. Funder, K.M. Curnow, "A compound heterozygote case of type 2 aldosterone synthase deficiency", *J. Clin. Endocrinol. Metab.*, 2003, 88, 6, 2518-2526
- 56 A.W. Munro, H.M. Girvan, K.J. McLean, "Variations on a (t)heme-novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily", *Nat. Prod. Rep.*, 2007, 24, 3, 585-609
- 57 R. Raag, T.L. Poulos, "Crystal structures of cytochrome P-450cam complexed with camphane, thiocamphor, and adamantine: factors controlling P-450 substrate hydroxylation", *Biochemistry*, 1991, 30, 10, 2674-2684
- 58 T.L. Poulos, R. Raag, "Cytochrome P450cam: crystallography, oxygen activation, and electron transfer", *the FASEB journal*, 1992, 6, 2, 674-679
- 59 S.A. Martinis, W.M. Atkins, P.S. Stayton, S.G. Sligar, "A conserved residue of cytochrome P-450 is involved in heme-oxygen stability and activation", *J. Am. Chem. Soc.*, 1989, 111, 26, 9252-9253

- 60 M. Imai, H. Shimada, Y. Watanabe, Y. Matsushima-Hibiya, R. Makino, H. Koga, T. Horiuchi, Y. Ishimura, "Uncoupling of the cytochrome P-450cam monooxygenase reaction by a single mutation, threonine-252 to alanine or valine: possible role of the hydroxy amino acid in oxygen activation", *Proc. Natl. Acad. Sci. U.S.A.*, 1989, 86, 20, 7823-7827
- 61 S. Bechtel, N. Belkina, R. Bernhardt, "The effect of amino-acid substitutions I112P, D147E, and K152N in CYP11B2 on the catalytic activities of the enzyme", *Eur. J. Biochem.*, 2002, 269, 4, 1118-1127
- 62 B. Bottner, H. Schrauber, R. Bernhardt, "Engineering a mineralocorticoid- to a glucocorticoid-synthesizing cytochrome P450", *J. Biol. Chem.*, 1996, 271, 14, 8028-8033
- 63 L. Roumen, M.P.A. Sanders, K. Pieterse, P.A.J. Hilbers, R. Plate, E. Custers, M. de Gooyer, J.F.M. Smits, I. Beugels, J. Emmen, H.C.J. Ottenheijm, D. Leysen, J.J.R. Hermans, "Construction of 3D models of the CYP11B family as a tool to predict ligand binding characteristics", *J. Comput.-Aided Mol. Des.*, 2007, 21, 8, 455-471
- 64 A.D. MacKerell Jr, D. Bashford, M. Bellott, R.L. Dunbrack Jr, J.D. Evanseck, M.J. Field, S. Fisher, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins", *J. Phys. Chem. B*, 1998, 102, 18, 3586-3616
- 65 M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, "Improved protein-ligand docking using GOLD", *Proteins*, 2003, 52, 4, 609-623
- 66 J.G. Hackett, R.W. Brueggemeier, C.M. Hadad, "The final catalytic step of cytochrome P450 aromatase: a density functional theory study", *J. Am. Chem. Soc.*, 2005, 127, 14, 5224-5237
- 67 G.N. Ramachandran, C. Ramakrishnan, "Conformations of polypeptides and proteins", *Adv. Prot. Chem.*, 1968, 23, 283-437
- 68 C. Colovos, T.O. Yeates, "Verification of protein structures: patterns of nonbonded atomic interactions", *Protein Sci*, 1993, 2, 9, 1511-1519
- 69 J.U. Bowie, R. Luthy, D. Eisenberg, "A method to identify protein sequences that fold into a known three-dimensional structure", *Science*, 1991, 253, 5016, 164-170
- 70 R.A. Laskowski, D.S. Moss, J.M. Thornton, "Main-chain bond lengths and bond angles in protein structures", *J. mol. Biol.*, 1993, 231, 4, 1049-1067
- 71 B. Rupp, S. Raub, C. Marian, H.D. Holtje, "Molecular design of two sterol 14 $\alpha$ -demethylase homology models and their interactions with theazole antifungals ketoconazole and bifonazole", *J. Comput.-Aided. Mol. Des.*, 2005, 19, 3, 149-163
- 72 T. Tanaka, T. Okuda, Y. Yamamoto, "Characterization of the CYP3A4 active site by homology modelling", *Chem. Pharm. Bull. (Tokyo)*, 2004, 52, 7, 830-835
- 73 A. Fisher, E.C. Friel, R. Bernhardt, C. Gomez-Sanchez, C. Connell, J.M.C. Fraser, E. Davies, "Effects of 18-hydroxylated steroids on corticosteroid production by human aldosterone synthase and 11 $\beta$ -hydroxylase", *J Clin Endocrinol Metab*, 2001, 86, 9, 4326-4329
- 74 Y. Nonaka, M. Okamoto, "Functional expression of the cDNAs encoding rat 11 $\beta$ -hydroxylase [cytochrome P450(11 $\beta$ )] and aldosterone synthase [cytochrome P450(11 $\beta$ , ald)]", *Eur J Biochem*, 1991, 202, 3, 897-902
- 75 S. Takemori, S. Kominami, "The role of cytochromes P-450 in adrenal steroidogenesis", *Trends In Biochemical Sciences*, 1984, 9, 9, 393-396
- 76 J.O. Johnston, C.L. Wright, G.W. Holbert, "Enzyme-activated inhibitors of steroidal hydroxylases", *J. Steroid Biochem. Molec. Biol.*, 1995, 52, 1, 17-34



## Chapter 3 Molecular Dynamics

In the previous chapter, we have discussed the construction of homology models for the CYP11B isoforms. Analysis of protein-substrate interactions yielded an explanation for the regio-specific hydroxylation of steroids and a hypothesis for the active site requirements for aldosterone synthesis by CYP11B2. This promising result has given us enough confidence in the models for the design of CYP11B2 inhibitors. However, the evaluation of protein-ligand interactions using a static protein model does not always imply that the best interaction of a ligand in the active site is taken into account. The conformation of amino acids in the active site or just outside the active site may have a large influence on the binding of a ligand. Here, we employ molecular dynamics simulations to sample multiple conformations of the protein active site to determine whether the constructed models possess sufficient structural integrity, i.e., whether the regions outside the active site are modelled accurately. In addition, the ligand binding characteristics of four known CYP11B inhibitors are detailed (*R*-fadrazole, *S*-fadrazole, *R*-etomidate and metyrapone). The reason for choosing these inhibitors is explained in paragraph 3.3 .

Part of this chapter is described in:

L. Roumen, M.P.A. Sanders, K. Pieterse, P.A.J. Hilbers, R. Plate, E. Custers, M. de Gooyer, J.F.M. Smits, I. Beugels, J. Emmen, H.C.J. Ottenheijm, D. Leysen, J.J.R. Hermans, "**Construction of 3D models of the CYP11B family as a tool to predict ligand binding characteristics**", *J Comput-Aided Mol Des*, 2007, 21, 8, 455-471



### 3.1 Molecular Dynamics

Molecular dynamics is the science of simulating the motions of a system of particles. It provides insight on the fluctuation behaviour of the particle positions. The molecular dynamics method was first introduced by Alder to study the interactions of hard spheres [1,2]. Later, the first molecular dynamics simulations on proteins have been performed by McCammon et al on the dynamic behaviour of bovine pancreatic trypsin inhibitor (BPTI) [3]. Nowadays, the application of molecular dynamics simulations addresses the behaviour of solvated proteins, protein folding, assessment of protein-ligand interactions, and even X-ray crystallography and NMR structure determination. Furthermore, novel modelling methods have emerged, involving combinations with either Monte Carlo or quantum mechanics calculations.

### 3.2 Force fields

Molecular dynamics simulations require a definition of a potential function that describes how particles interact. This potential function is usually referred to as the force field of the system. The force field is a combination of the functional form and the corresponding parameter sets that are used during the simulation to describe the potential energy of the system. Force fields include a standard set of parameters describing the bonded and non-bonded interactions of the model system. The bonded terms of the potential function are defined by a combination of a bond potential, an angle potential and a torsional potential. Additionally, the torsional contributions can be grouped into proper dihedrals (such as the description for the cis, trans and gauche conformations of alkyl chains), and improper dihedrals to enforce planarity of conjugated systems and aromatic rings. Many non-bonded interactions between particles can be described in the force field. However, for the investigation of protein structures, force fields are commonly limited to electrostatic contributions (Coulomb's Law) and van der Waals contributions (Lennard Jones potential).

As a result, several types of non-bonded interactions between atoms or structures are simplified. Important interactions such as hydrogen bonds can be modelled using an additional explicit potential function, but are commonly modelled implicitly using the electrostatics contributions. Similarly, the stacking of aromatic ring systems is implicit to the electrostatic and van der Waals contributions. Another limitation of force fields is that atomic charges do not fluctuate during molecular dynamics. Hence, atoms cannot polarise their charge as a result of an (un)favourable atom environment. Furthermore, the bonding character of atoms is included in the molecule definition and bonds are not allowed to be broken or formed. Because of these limitations, molecular dynamics simulations are less suitable to apply in the study of chemical reactions.

Using molecular dynamics in this thesis implies the proper force field choice for the protein simulations. Several force field have been empirically derived for the modelling protein structures such as AMBER [4], CHARMM [5], OPLS [6], MMFF94 [7] and GROMOS [8]. A typical parameter set includes the values for the mass and van der Waals radius of the individual atoms. Additionally, equilibrium values are included for the bond length, angles

and torsion angles with a spring constant corresponding to the different potentials (bond, angle or torsional). The parameter set contains multiple entries for each element, since they behave differently when they are for instance sp<sup>2</sup> or sp<sup>3</sup> hybridised. Next to the parameter set, a topology set is used comprising the definition of atomic assemblies such as molecules, individual amino acids, ions and water. In the topology set, each atom in the molecular assembly is defined by an atom type (corresponding to the expected elements behaviour), atomic partial charge and atom connectivity (bonds, angles and torsion angles). Often topology sets comprise so-called internal coordinates that are used to reconstruct the atomic coordinates of atoms, for which no coordinates have been provided by the user.

### 3.3 Molecular Dynamics of hCYP11B1 and hCYP11B2

In this thesis, molecular dynamics simulations have been performed to gain insight into the stability of the protein model packing, the dynamic behaviour of the protein models and to evaluate protein-ligand interactions. To investigate the stability of the protein model packing, we have performed a molecular dynamics simulation using 18-hydroxycorticosterone. For the validation of the constructed models of hCYP11B1 and hCYP11B2, we have attempted to correlate the in vitro activity of four known inhibitors to in silico data. The inhibitors that have been chosen for this study are metyrapone [9], R-etomidate [10], R-fadrazole and S-fadrazole [11] (Figure 3–1). These inhibitors were chosen for the following reasons. Metyrapone is a known inhibitor of CYP11B1 and is clinically used in the diagnosis of Cushing Syndrome [9,12]. R-etomidate is clinically used as anaesthetic, but it is known to be a highly potent suppressor of the CYP11B family [10,13]. Racemic fadrazole was designed for the selective inhibition of aromatase (CYP19), a cytochrome P450 enzyme closely related to the CYP11B family as it oxidises steroids on C<sub>19</sub>. Next to aromatase inhibition, racemic fadrazole also shows considerable inhibition of members of the CYP11B family [11,14].

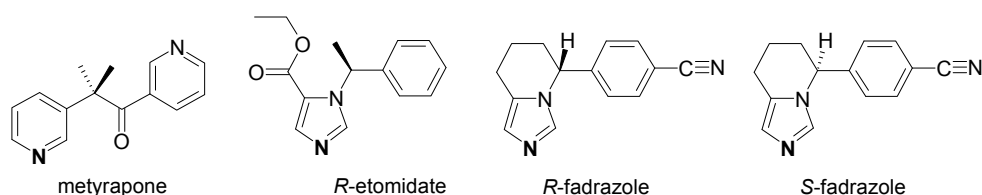


Figure 3–1 Chemical structure of the four known inhibitors. The aromatic nitrogen atoms interacting with the heme iron are indicated bold.

### 3.4 Molecular Dynamics Settings

Molecular dynamics simulations were performed to investigate the stability of the enzyme models of hCYP11B1 and hCYP11B2. The molecular dynamics package used in these simulations is NAMD [15] in combination with the CHARMM force field that has been best parameterised for TIP3P explicit solvent simulations [16] (TIP3P being the current standard model system for modelling water). The simulation settings include the usage of periodic boundary conditions. The interaction between individual atoms is calculated using the standard CHARMM cut-off of 13.0 Å. A simulation box needs to be defined that is large enough to prevent interaction of atoms on opposite sides of the protein structure. The maximum width of the hCYP11B1 and hCYP11B2 protein models is 65.4 Å and 64.2 Å, respectively. Hence, the protein models of hCYP11B1 and hCYP11B2 were solvated in an 80 Å × 80 Å × 80 Å equilibrated TIP3P water box, by placing the largest width in on the diagonal of the water box. Subsequently, excess water was removed using a cut-off value of 2.5 Å to prevent water from overlapping the van der Waals radius of the protein atoms.

Counter-ions were added to ensure an overall net charge of zero. The protein-ligand complexes were slightly equilibrated for 25 ps at a temperature of 100 K and were subsequently simulated for 1ns at a temperature of 310 K in an NPT ensemble. This particular ensemble implies a constant temperature, pressure and amount of particles. To calculate the electrostatic interactions we chose the implementation of NAMDs Particle Mesh Ewald [17].

The ligands have been docked flexibly into the active site cavity using GOLD (paragraph 4.6.2 ). Using the program Maximal Speed Molecular Surface (MSMS [18]), the volume of the active site and of the different ligands has been calculated (Table 3–1). As a result from the tight docking pose of 18-hydroxycorticosterone, no space was available to add water molecules to the active site cavity. To accommodate the filling of the active site cavity volume by the slightly smaller ligands, three and four explicit water molecules were positioned in the active sites of hCYP11B1 and hCYP11B2, respectively. The positions of these water molecules were optimised during the equilibration step. The protein-ligand interactions found in the docking study were monitored during the simulation to gain an impression of the stability of the ligand in the active site.

Table 3–1 Molecular volume for the ligands and the active site of the proteins

Molecule	hCYP11B1	hCYP11B2	18OH-B	Etomidate	Fadrazole	Metyrapone	Water
Volume (Å <sup>3</sup> )	360.0	334.0	315.6	269.7	241.9	247.7	15.5

### 3.5 Results and Discussion

#### 3.5.1 Protein Structure Stability

To investigate the structural integrity of the protein fold, the relative root mean square deviation (RMSD) was calculated over all heavy atoms (Figure 3–2). The first 25 ps of equilibration, the structure relaxes to the water environment and optimises to a local potential energy minimum. After equilibration, the increase in thermal motion allows for a diverse sampling of the protein structure. During the first 500 ps, the RMSD increased and the protein still adapted towards its optimal conformation. After this point in time hardly any change in the three dimensional structures of the proteins was observed other than rotations of loop regions far from the active site.

The largest fluctuations of the protein were found in the flexible regions with peak values located in the structures around helix A, helix D and the beginning of helix E (Figure 3–3). The reason for these fluctuations is a large insertion of 7 amino acids introduced by the modelling procedure. To accommodate the insertion of amino acids, we had elongated helix D by one turn before it connects to the following beta-sheet. Inside the water box, this region is found to protrude into the water without any stabilising protein interactions and unfolds due to interaction with water.

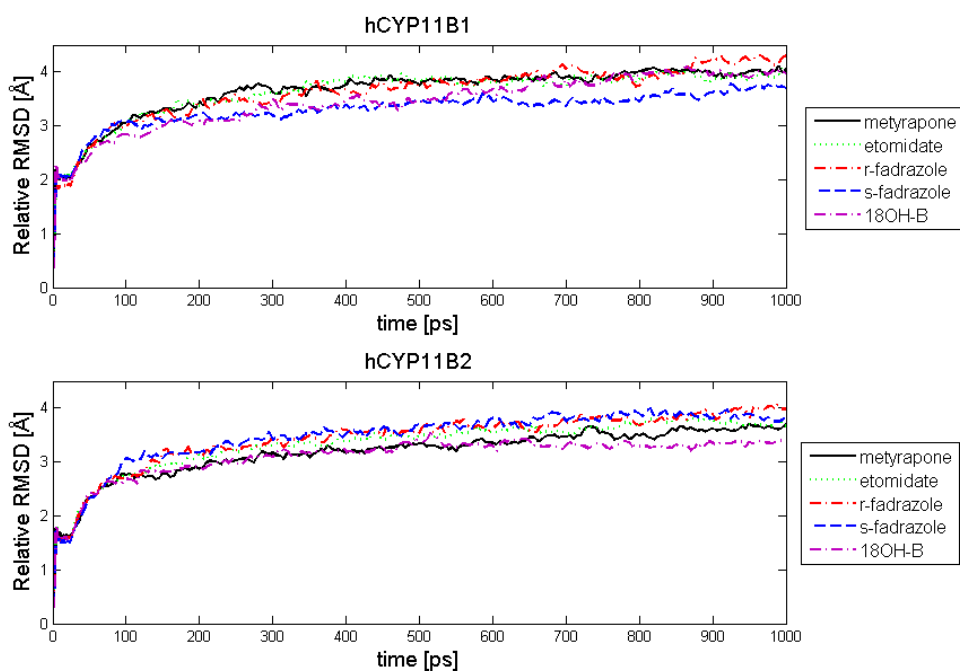


Figure 3–2 Relative RMSD for the heavy atoms versus time. All combinations for the protein-ligand complexes behave in a similar fashion

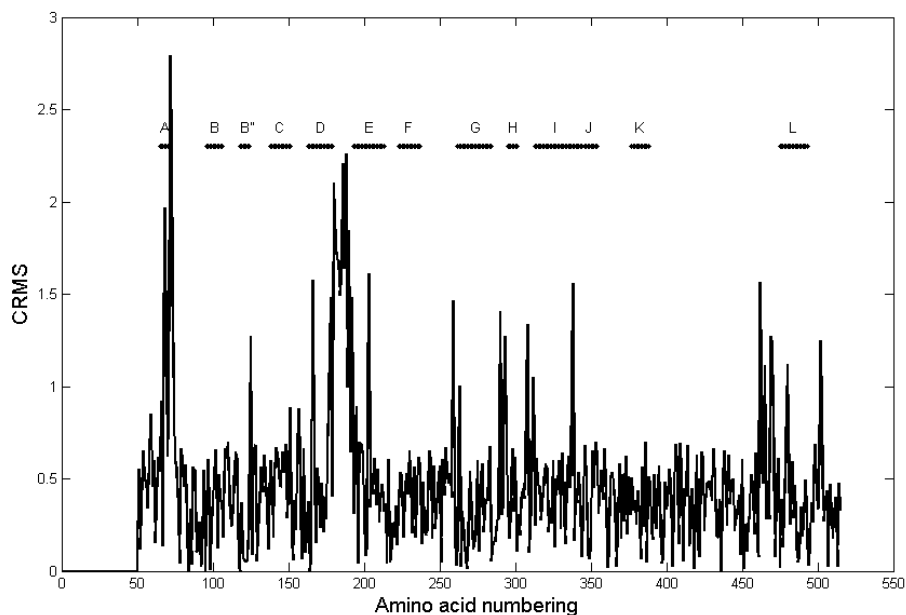


Figure 3–3 Average RMSD for the amino acid backbone for the hCYP11B2 model with 18OH-B. Large deviations are found for helix A and the loop region between helices D and E. Slight movement in helix I is observed due to interaction with the water molecules in the water channel.

In all the simulations we observed an opening of the active site and the continuous flow of water molecules in and out of the active site cavity. Several water molecules retained key positions. The water molecules that make up the channel towards the conserved E459 cannot leave their cavity and remain stable during the entire simulation (Figure 3–4a and b). Although they interact with helix I and challenge the integrity of the helix structure, the helix remained stable and only slight deviations were observed in its RMS profile (Figure 3–3). Interactions made by another water molecule to stabilise a heme propionate group, is also preserved during the simulation (Figure 3–4c and d). The other propionate group also possesses the same stabilising interactions with W137 and R141 in the (W/H)(R/K)X(R/K)R motif of helix C.

The stability of helix B' is ensured by the interaction of R123 with E310 in helix I. Stabilising interactions between the outside of helix B' and the acidic mutant in helix I (D302 in hCYP11B1 and E302 in hCYP11B2) can be observed during the simulations, yet are frequently lost due to the interference of water. Since these are on the outside of the protein, it is uncertain whether these interactions are realistic and of relevance for the protein catalytic and ligand-binding properties.

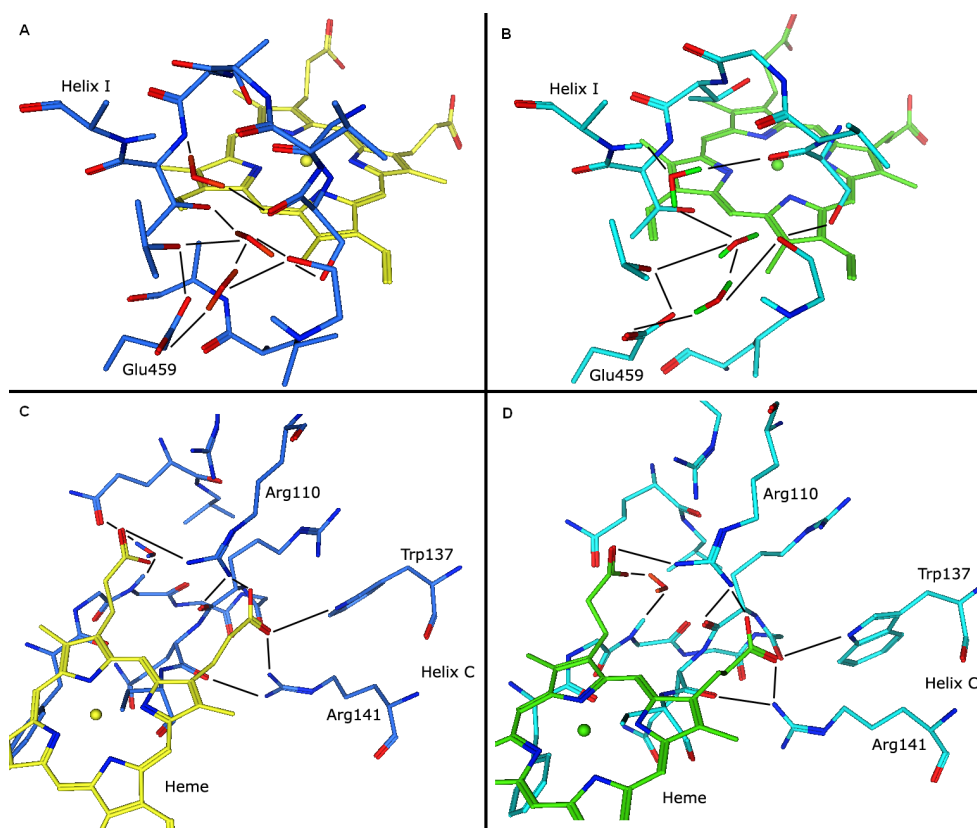


Figure 3-4 Conserved interactions around the heme prosthetic group for hCYP11B2 with R-fadrazole. Hydrogen bonding network for the water channel from E459 to helix I; A) at the beginning of the simulation and B) at the end of the simulation. Hydrogen bonding network featuring conserved interactions for a water molecule that mediates interactions for heme propionate group stabilisation for C) the beginning of the simulation and D) the end of the simulation.

### 3.5.2 Protein-Ligand Interactions

Nearly all of the ligands maintained the same interactions as observed in the docking study, indicating that the interaction predicted by docking was stable. The exceptions to this observation are R-fadrazole in the hCYP11B1 model and S-fadrazole in the hCYP11B2 model. Both these protein-ligand interactions were disrupted when their imidazole group lost its interaction with the heme group in exchange for a water molecule. This happened several times during the simulation until finally the ligand lost all interactions with the heme and moved to the top of the active site (Figure 3-5a-f). Repeated simulations with slightly different starting conformations of the ligand were unable to change these scenarios. Apparently, these ligands can not be stabilised in their respective protein models.

The steroid 18-hydroxycorticosterone (18OH-B) contains a ring stacking between its steroid A-ring and F130 (Figure 3-5b, e). In addition to this interaction, 18OH-B contains two

hydrogen bonding interactions in the active site for both hCYP11B1 and hCYP11B2. These are between R123 and the C<sub>3</sub>-ketone group and between the backbone of F379 and the C<sub>21</sub>-hydroxyl group (Figure 3–5c, f). All interactions of the steroid were maintained during the simulation.

The stable interactions of the inhibitors metyrapone and R-etomidate are displayed in Figure 3–6. Etomidate possesses a ring stacking in the active site with F130 that is located in helix B'. The substituents at the chiral centre of etomidate bestow it certain rigidity. This causes short distortions in the ring-ring interaction during the simulation for several time steps. However, in general the interaction is maintained (>90% of the simulations, Figure 3–5b, e). Next to the ring-ring interaction, R-etomidate also possesses a hydrogen bond with the catalytic threonine, T318. This interaction is found in both the hCYP11B1 and hCYP11B2 models, which may explain why R-etomidate possesses a similar inhibitory action for both protein isoforms. The interaction is also strongly maintained during the simulation (Figure 3–5c, f).

Metyrapone contains a ring-stacking with F130 with its pyridine ring as well. Metyrapone also mediates a hydrogen bond via water with the R110 in helix B' that stabilises the heme in the active site. These interactions were maintained throughout the simulation (Figure 3–5b, c, e, f).

The poses of R-fadrazole in the hCYP11B2 active site and S-fadrazole in the hCYP11B1 active site are compared in Figure 3–7. In both models the position of helix B' is stabilised by the interaction of R123 in helix B' with E310 in helix I. There is a strong difference in the active site cavity of the hCYP11B1 and hCYP11B2 models. The helix B' is moved further out of the active site in the hCYP11B2 model, and helix I is slightly twisted because of this interaction. There is also a difference in the protein-ligand interactions observed in both models. In the hCYP11B2 model, R-fadrazole possesses a polar interaction with R123 and a horizontal ring stacking with F130, whereas in the hCYP11B1 model, S-fadrazole possesses the same interactions, yet the ring stacking with F130 is vertical. These interactions with the protein are the same interaction points as observed for 18-hydroxycorticosterone (Figure 3–5a-f).

Because the complex was stable over the last 500 ps, we sampled the non-bonded interactions between the protein and the ligand for this time period (Table 3–2). These interactions contribute to the binding free energy, although the solvation effects of free ligand still need to be subtracted [19]. These energies follow the same general trend as the in vitro screening results, which emphasise the strong interactions for R-etomidate in both hCYP11B models and the strong interactions of S-fadrazole in hCYP11B1 and R-fadrazole in hCYP11B2 respectively.

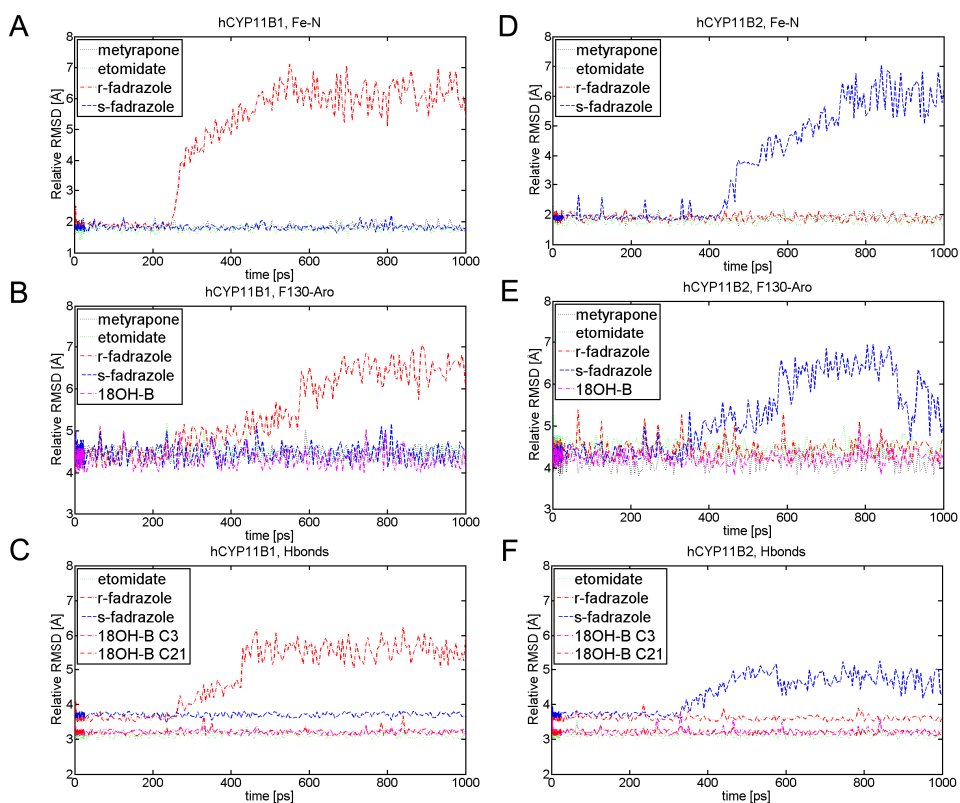


Figure 3–5 Non-bonded interactions observed during the molecular dynamics simulation. Fe-N indicates the polar interaction between the heme iron and the aromatic nitrogen for the ligands (1.8 Å - 2.5 Å). F130-Aro indicates the ring-ring interaction between F130 and the phenyl ring of the ligands (3.8 Å - 5.0 Å). The interaction with metyrapone is by the pyridine ring, and the interaction with 18OH-B is by the steroid A-ring. Hydrogen bonding interactions are between T318 and the ester group of etomidate (3.2 Å - 3.8 Å), R123 and the para-cyano group of fadrazole (3.8 Å - 4.4 Å, measured from the cyano nitrogen atom to the guanidinium carbon atom), R123 and the C<sub>3</sub>-ketone group of 18OH-B, and F379 backbone and the C<sub>21</sub>-hydroxyl group of 18OH-B.



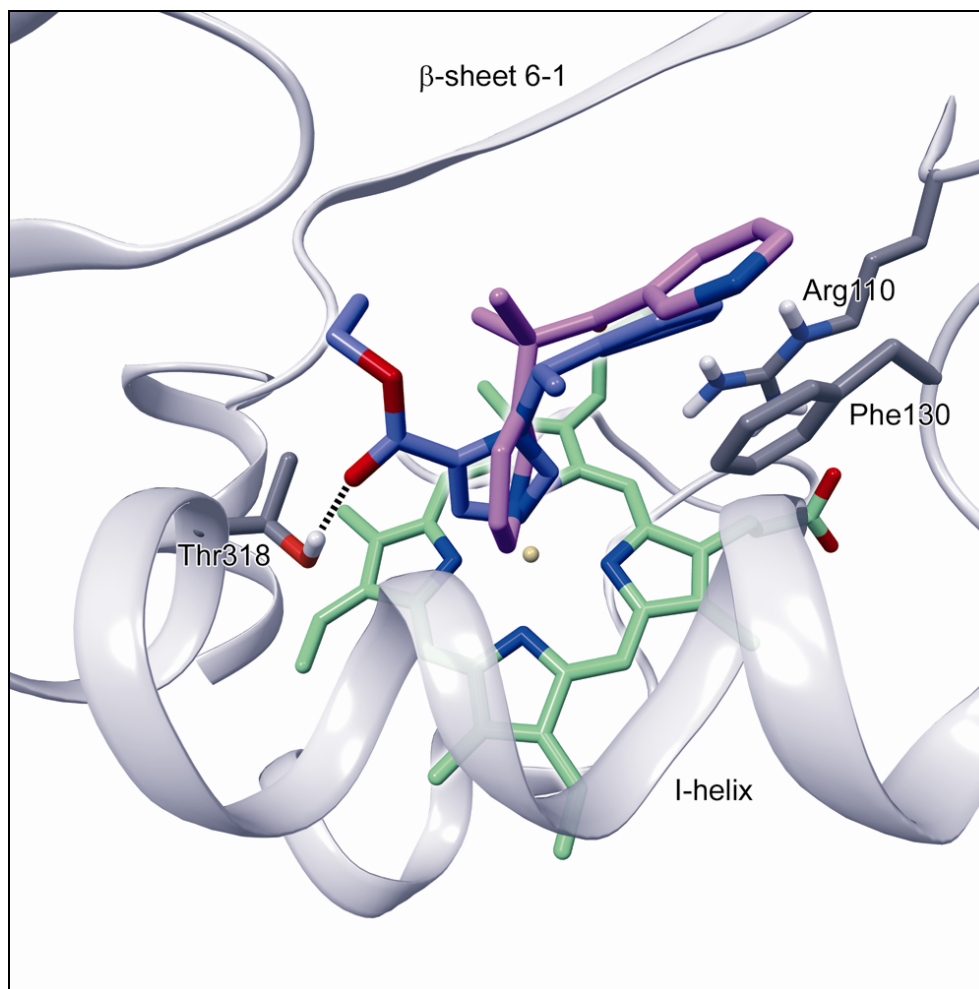


Figure 3–6 hCYP11B2 active site model with metyrapone (purple) and R-etomidate (blue) in the active site. Indicated are the stabilising interaction between R110 and the heme group, as well as F130 that accommodates a ring stacking interaction with the ligands. The ester group of R-etomidate possesses a hydrogen bond with the catalytic T318, an interaction also found in the hCYP11B1 model.

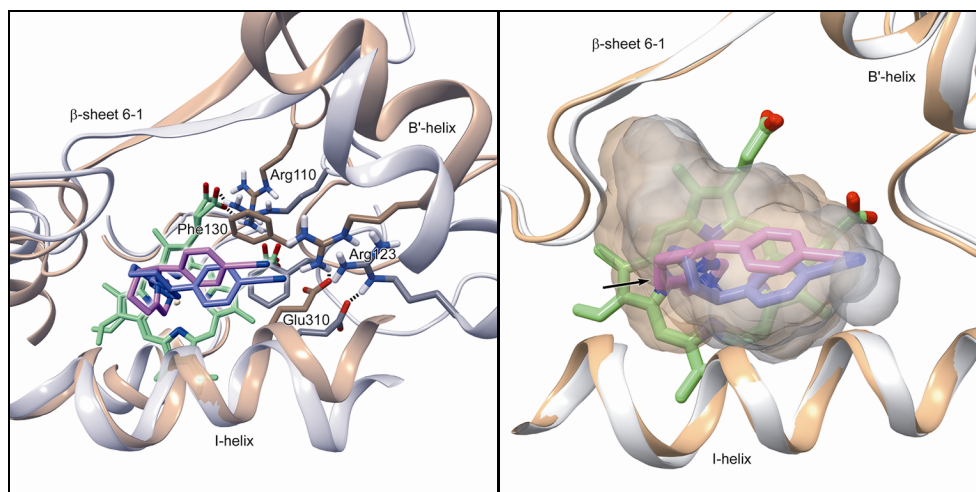


Figure 3–7 (Left) hCYP11B1 model (orange) containing S-fadrazole (purple) in the active site, and the hCYP11B2 model (white) containing R-fadrazole (blue) in the active site. Helix B' is stabilised through the interaction of R123 in this helix with E310 in helix I. The helix B' is shifted in the hCYP11B2 model positioning R123 further away from the active site cavity. Both R-fadrazole and S-fadrazole interact with R123 in helix B' and possess a ring stacking interaction with F130 in helix B'. The ring stacking is horizontal for R-fadrazole and vertical for S-fadrazole.

Figure 3–8 (Right) Active site volume difference between hCYP11B1 (white) and hCYP11B2 (orange). Shown for both hCYP11B1 and hCYP11B2 are the active site volume and the backbone trace. R-fadrazole and S-fadrazole are indicated with blue and purple respectively. It is clear to see that hCYP11B1 contains a larger active site between helix I and beta-sheet 6-1. This cavity allows S-fadrazole to fit in the cavity, but not R-fadrazole (black arrow). On the other side of the active site near Helix B' and R123, hCYP11B2 contains the larger cavity, which might rationalise the better fit of R-fadrazole in the cavity.

Table 3–2 Correlation of molecular dynamics results to in vitro inhibitor potency

	IC50 (nM)		Unon-bonded (kcal/mol)	
	mean	s.d.	mean	s.d.
hCYP11B1				
Metyrapone	46.4	10.4	-48.4	3.7
R-Etomidate	0.5	0.2	-56.0	2.4
R-Fadrazole	118.6	8.9	-38.4	2.5
S-Fadrazole	39.5	4.4	-56.3	3.4
hCYP11B2				
Metyrapone	207.8	4.5	-36.2	7.5
R-Etomidate	1.7	0.9	-54.4	2.9
R-Fadrazole	6.0	1.9	-55.9	3.3
S-Fadrazole	171.2	51.7	-44.3	1.8

### 3.6 Conclusions

We have conducted a molecular dynamics study to investigate the structural integrity of the hCYP11B1 and hCYP11B2 models. In general, the structural core of the protein models did not change during the simulation, as portrayed by the relative RMSD as well as the average RMSD per residue. The important interactions that stabilise the heme in the active site as well as a water channel were also conserved during the simulation.

Simulation deviations were observed for two regions. The first region is helix A which does not strongly interact with the other parts of the protein structure. The second is the loop region between helices D and E for which we had introduced a 7 amino acid insertion in the modelling study of Chapter 2. The accuracy of this loop region was already questioned and defined as low in the structure validation study in paragraph 2.7.1. The molecular dynamics simulation seems to confirm this. The conformation of these active site regions should be improved by further model optimisation.

The ligands that have been used in the molecular dynamics study are 18-hydroxycorticosterone (18OH-B), metyrapone, R-etomidate, R-fadrazole and S-fadrazole. 18OH-B forms a hydrogen bond between its C<sub>21</sub>-hydroxyl group and F379 and between its C<sub>3</sub>-ketone group and R123. It also forms a stable ring-stacking in the active site with F130. During the simulations, 18OH-B maintained its hydrogen bonding interactions on both extremities of the ligand as well as the ring-ring interaction.

All other ligands bind in a similar fashion to the protein active sites as 18OH-B. They all occupy the same space as 18OH-B and possess a ring-ring interaction with F130. Both fadrazole enantiomers also form a hydrogen bonding interaction with R123. The inhibitors are not large enough to interact with the active site cavity where the steroid mediates an interaction with F379. There are only small differences in the binding mode of Metyrapone and R-etomidate in the hCYP11B1 and hCYP11B2 models, suggesting that these ligands are not selective inhibitors.

The binding mode of the fadrazole enantiomers is different for both hCYP11B1 and hCYP11B2; R-fadrazole preferring the shape of the hCYP11B2 active site cavity and S-fadrazole preferring the shape of the hCYP11B1 active site cavity. During the simulation of the enantiomers in the opposite protein models, their interactions were lost, indicating that the simulations corroborate the stereoselectivity of the enantiomers. Moreover, calculation of the protein-ligand interactions and comparison to the *in vitro* potency of the inhibitors supports the general trend in binding strength.

## Literature

- 1 B.J. Alder, T.E. Wainwright, "Phase transition for a hard sphere system", *J. Chem. Phys.*, 1957, 27, 5, 1208-1209
- 2 B.J. Alder, T.E. Wainwright, "Studies in molecular dynamics. I. general method", *J. Chem. Phys.*, 1959, 31, 2, 459-466
- 3 J.A. McCammon, B.R. Gelin, M. Karplus, "Dynamics of folded proteins", *Nature*, 1977, 267, 5612, 585-590
- 4 Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations", *J. Comp. Chem.*, 2003, 24, 16, 1999-2012
- 5 A.D. MacKerell Jr, D. Bashford, M. Bellott, R.L. Dunbrack Jr, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins", *J. Phys. Chem. B.*, 1998, 102, 18, 3586-3616
- 6 W.L. Jorgensen, J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin", *J. Am. Chem. Soc.*, 1988, 110, 6, 1657-1666
- 7 T.A. Halgren, "Merck molecular force field. I. basis, form, scope, parameterization, and performance of MMFF94", *J. Comp. Chem.*, 1996, 17, 5-6, 490-519
- 8 C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, "A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6", *J. Comp. Chem.*, 2004, 25, 13, 1656-1676
- 9 M. Schonshofer, B. Schefzig, W. Oelkers, "Evidence of adrenal 18-hydroxylase inhibition by metyrapone in man", *Horm. Metab. Res.*, 1979, 11, 4, 306-308
- 10 H.G. Dorr, U. Kuhnle, H. Holthausen, F. Bidlingmaier, D. Knorr, "Etomidate: a selective adrenocortical 11 beta-hydroxylase inhibitor", *Klin. Wochenschr.*, 1984, 62, 21, 1011-1013
- 11 L.M. Demers, J.C. Melby, T.E. Wilson, A. Lipton, H.A. Harvey, R.J. Santen, "The effects of CGS 16949A, an aromatase inhibitor on adrenal mineralocorticoid biosynthesis", *J. Clin. Endocrinol. Metab.*, 1990, 70, 4, 1162-1166
- 12 P.C. Avgerinos, J.A. Yanovski, E.H. Oldfield, L.K. Nieman, G.B. Cutler Jr, "The metyrapone and dexamethasone suppression tests for the differential diagnosis of the adrenocorticotropic-dependent Cushing Syndrome: a comparison", *Ann. Intern. Med.*, 1994, 121, 5, 318-327
- 13 I.W. Fellows, M.D. Bastow, A.J. Byrne, S.P. Allison, "Adrenocortical suppression in multiply injured patients: a complication of etomidate treatment", *Br. Med. J.*, 1983, 287, 6408, 1835-1837
- 14 K. Schieweck, A.S. Bhatnagar, A. Matter, "CGS 16949A, a new nonsteroidal aromatase inhibitor: effects on hormone-dependent and -independent tumors in vivo", *Cancer. Res.*, 1988, 48, 4, 834-838
- 15 J.C. Philips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, "Scalable molecular dynamics with NAMD", *J. Comp. Chem.*, 2005, 26, 16, 1781-1802
- 16 W.L. jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, "Comparison of simple potential functions for simulating liquid water", *J. Chem. Phys.*, 1983, 79, 2, 926-935
- 17 P.F. Batcho, D.A. Case, R. Schlick, "Optimized particle-mesh ewald/multiple-time step integration for molecular dynamics simulations", *J. Chem. Phys.*, 2001, 115, 9, 4003-4018
- 18 M.F. Sanner, A.J. Olson, J.C. Spehner, "Reduced surface: an efficient way to compute molecular surfaces", *Biopolymers*, 1996, 38, 3, 305-320
- 19 Y.Y. Sham, Z.T. Chu, H. Tao, A. Warshel, "Examining methods for calculations of binding free energies: LRA, LIE, PDL-LRA, and PDL/S-LRA calculations of ligand binding to an hiv protease", *Proteins. Struct. Funct. Genet.*, 2000, 39, 4, 393-407



## Chapter 4 Molecular Docking

In the previous chapter we have investigated the dynamic behaviour of the homology models as well as elucidated the ligand-binding characteristics of four known inhibitors. The structural stability of the protein-ligand interactions has confirmed the structural stability of the amino acid packing involved with ligand binding. In a next step, application of the homology models for drug design can be tested by performing molecular docking of novel CYP11B inhibitors. In this chapter, we present the results for molecular docking in the homology models of human CYP11B1 and CYP11B2, as well as a comparison of the predicted docking affinities to the *in vitro* measured ligand affinities. By incorporating the protein states sampled by the molecular dynamics simulations, rather than just the static homology model, we have ensured that the dynamic behaviour of the protein active site has been taken into consideration. This not only allows the elucidation of important protein-ligand interactions, but also approximates the effects of ligand-induced changes of the overall protein conformation.

Part of this chapter is described in:

L. Roumen, M.P.A. Sanders, K. Pieterse, P.A.J. Hilbers, R. Plate, E. Custers, M. de Gooyer, J.F.M. Smits, I. Beugels, J. Emmen, H.C.J. Ottenheijm, D. Leysen, J.J.R. Hermans, "**Construction of 3D models of the CYP11B family as a tool to predict ligand binding characteristics**", *J Comput-Aided Mol Des*, 2007, 21, 8, 455-471

#### 4.1 Molecular Docking

Molecular docking is a method that predicts the preferred binding mode of a ligand and a target molecule, often a protein. It is used for rapid in silico screening of protein-ligand interactions and is also used for the prediction of ligand binding affinities. The preferred binding mode of a ligand in the protein active site is determined by the docking protocol that samples different ligand (and often protein residue) conformations to optimise the protein-ligand interactions.

A conventional method to sample protein-ligand interactions is to generate so-called interaction grids for the different types of contributions that are used to determine the total interaction energy of the protein-ligand complex. Many docking programs such as Dock [1], AutoDock [2], ICM [3] and Glide [4,5] pre-generate these matching grids to perform a fast sampling of the optimal binding modes. The grids also possess an alternate function as they can be manipulated to construct attraction and repulsion wells to force a compound to match or deny a particular site. The matching of a ligand to the protein (or interaction grid) can be performed by using incremental ligand reconstruction (Dock, FlexX [6]), Monte Carlo search (ICM, Glide) or a genetic algorithm search (AutoDock, GOLD [7]). The obtained poses are subsequently ranked by a so-called scoring function that evaluates the protein-ligand interactions and determines the best interacting binding mode for the ligand.

A scoring function can be based on a force field or empirically derived, both implying a variety of limitations in their usage. Force field-derived scoring functions often only take into account the electrostatic and van der Waals interactions, yet neglect important terms such as desolvation effects and entropic effects. They are also subject to the limitations of the force field parameters and how these parameters have been derived. The advantage of using force field-derived scoring functions is that they are applicable to other chemical series and do have a strong physical basis.

Empirically derived scoring functions have been derived from known protein-ligand complexes (crystal structures) with well-matched interactions and have been optimised to maintain these interactions. Their goal is to reproduce the experimental values of binding energy and importantly, their global minimum is directed at the binding mode observed in the crystal structure. As a result, the docking process may have little tolerance for small changes in binding orientation or conformational changes and subsequently the classification of sub-optimal docking poses can be incorrect. Certain empirical scoring functions take desolvation energies and metal interactions implicitly into account, which unfortunately only suits the ligands used for the derivations of the scoring functions themselves. Other empirical scoring functions possess explicit terms with weight factors derived from a certain compound series. Because all empirical scoring functions are based on a limited subset of protein-ligand complexes, they are generally self-consistent but not transferable to other chemical series. Thus, the prediction of binding affinity using empirical scoring functions remains questionable.

#### 4.1.1 GOLD

We have used GOLD v3.0 and v3.2 (Genetic Optimisation for Ligand Docking) for the flexible docking of ligands in the CYP11B models. GOLD uses a genetic algorithm for the docking of the compounds. The algorithm starts with a population of potential solutions for each ligand encoded as a so-called chromosome. Each of the chromosomes contains the flexible ligand conformation as well as two mappings between the protein and ligand atoms for the hydrogen bonding and van der Waals partners, respectively. For each ligand, the chromosomes are assigned a fitness score based on their predicted binding affinity. During each genetic algorithm iteration, the chromosomes can mutate to a different ligand conformation or two chromosomes may merge to obtain a different ligand conformation. The mutations and merging steps are biased towards the optimisation of a higher fitness score. The different binding modes that are obtained through docking are ranked using a scoring function. In general, the scoring function of a docking program consists of an electrostatic contribution and a van der Waals contribution to the protein-ligand interaction that is penalised by unfavourable clashes in protein-ligand overlap or ligand conformations. This is similar to the contributions that are often considered in molecular dynamics simulations (Chapter 3), however, the motility of the protein is often limited to torsional changes of the nearby protein amino acid residues.

#### 4.1.2 Scoring Functions

GOLD v3 comes with two scoring functions GoldScore [7,8,9,10] and ChemScore [11,12]. In v3.2, the docking function ASP (Astex Statistical Potential [13]) has been supplemented, however, due to project time limitations we have not used this scoring function in our analyses. The different scoring functions have been reported to be equally reliable, but given a certain problem, one or the other scoring function may result in a better prediction. In this thesis, the docking of compounds has been performed using the GoldScore, and many poses have been rescored using the ChemScore. Overall, the docking poses contributed by the GoldScore have provided more reliable results in our study.

The ChemScore scoring function has been derived empirically from a subset of 82 protein-ligand complexes for which the measured binding affinities were available. The ChemScore estimates the binding of free energy ( $\Delta G$ ) as a contribution of hydrogen bonds ( $\Delta G_{hb}$ ), metal-ligand interactions ( $\Delta G_{metal}$ ), lipophilic contact ( $\Delta G_{lipo}$ ), rotational entropy ( $\Delta G_{rot}$ ) as well as an intercept term assuming a minimal affinity ( $\Delta G_0$ ). In addition to these scores, an additional clash penalty ( $E_{clash}$ ) and internal torsional terms ( $E_{int}$ ) are added to filter out close protein-ligand contacts and poor internal conformations (Equation 4–1).

$$\text{Equation 4-1} \quad \text{ChemScore} = \Delta G_0 + \Delta G_{hb} + \Delta G_{metal} + \Delta G_{lipo} + \Delta G_{rot} + E_{clash} + E_{int}$$

The GoldScore fitness function is made up of a protein-ligand interaction component comprised of hydrogen bonding ( $S_{hb\_ext}$ ) and van der Waals energies ( $S_{vdw\_ext}$ ), and ligand intramolecular components comprised of hydrogen bonding ( $S_{hb\_int}$ ), van der Waals and



torsional strain energies (included in the van der Waals contributions,  $S_{vdw\_int}$ ) (Equation 4–2). The fitness score is taken as the negative sum of the component energy terms, so that larger fitness scores are better. There is an additional empirical weight factor ( $w$ ) of 1.375 introduced for the protein-ligand van der Waals score to encourage protein-ligand hydrophobic contact. Because the active site of our protein models is largely hydrophobic, this is probably the reason why the GoldScore performs better on average than the ChemScore.

$$\text{Equation 4-2} \quad \text{GOLD Fitness} = S_{hb\_ext} + w \times S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int}$$

#### 4.2 Prediction of Binding Affinity

Structure based drug design is commonly based on the prediction of the binding affinity of protein-ligand complexes. The correlation of *in silico* predictive data to biological data is rather difficult. The most conventional method is to correlate a calculated and an experimentally determined free energy of binding named  $\Delta G_{bind}$  and  $\Delta G_{exp,bind}$ , respectively. The calculation of the  $\Delta G_{bind}$  can be performed using several methods such as force fields used by molecular dynamics and (empirical) scoring functions used in molecular docking. Assuming thermodynamic equilibrium conditions for the protein-ligand complex formation, the  $\Delta G_{exp,bind}$  can be quantified by using the ligand binding constant  $K$  (Equation 4–3, Figure 4–1). Here, the inhibitor dissociation constant  $K_i$  can be compared to the substrate dissociation constant  $K_d$  or the negative association constant  $K_a$ .

$$\text{Equation 4-3} \quad \Delta G_{exp,bind} = -RT \ln K_a = RT \ln K_d$$

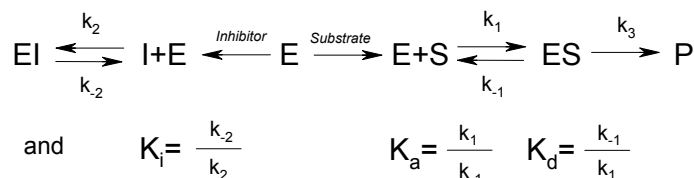


Figure 4–1 Kinetic scheme for competitive inhibition,  $E$  enzyme,  $S$  substrate,  $I$  inhibitor,  $P$  product.  $K_a$ , association constant,  $K_d$ , dissociation constant,  $K_i$ , inhibitor dissociation constant

In this thesis, the inhibitor potency has been measured by determining the IC<sub>50</sub> value which is the inhibitor concentration at which binding is reduced to 50%. If the measurement data has been obtained under the same experimental conditions, the IC<sub>50</sub> is proportional to the  $K_i$  and can be used to deduce binding free energy differences. We have transformed the IC<sub>50</sub> by taking the negative logarithm (Equation 4–4) as a representative for the  $\ln K_i$ . As such, a linear relationship between the biological data and the *in silico* predicted free energy is obtained (Equation 4–3). Because the potency for the best inhibitors is in the order of 1nM, the pIC<sub>50</sub> of our compounds reaches values of 9.

$$\text{Equation 4-4} \quad pIC50 = -\log(IC50)$$

Using the ChemScore for evaluating ligand binding poses allows the calculation of their free energy of binding ( $\Delta G_{bind}$ ). By correlating the ChemScore docking results to the pIC<sub>50</sub> values, it is possible to construct a predictive model for inhibitor potency. However, since the biological value used is the IC<sub>50</sub> rather than the  $K_i$ , and because the ChemScore only provides an estimation for the free energy, care must be taken with the interpretation of the docking results. The GoldScore is optimised for the evaluation of binding poses, but has not been optimised for the prediction of affinity. Nevertheless, the GoldScore does allow an accurate correlation between the ligand docking score and binding affinity [7]. Taking these considerations into account, the results of our docking study are only used to describe the general trends in inhibitor potency and determine those inhibitor substituents most important for inhibitor potency and selectivity.

### 4.3 Molecular Docking in the CYP11B Models

We have conducted three different molecular docking studies for the CYP11B protein models. In Chapter 2, we have investigated possible binding modes of steroids (Figure 4–2) in the active site of the constructed homology models. This docking study has been performed for the human and rat CYP11B isoforms, as well as the hCYP11B2-TripMut model. The resulting conformations have been used as input for the protein-substrate minimisation study discussed in paragraph 2.7.3 .

In Chapter 3, we have attempted to validate the constructed hCYP11B1 and hCYP11B2 models by correlating the *in vitro* activity of four known inhibitors (Figure 4–2) to *in silico* data. The *in silico* data that have already been discussed are the averaged interaction energies between the protein model and ligand structure during the molecular dynamics study (paragraph 3.5.2 ). It has been shown that the calculated interaction energy of the ligands correlates rather well with the *in vitro* activity profile (Table 3–2).

Also in Chapter 3, we have expanded the static structure of the initial protein models of hCYP11B1 and hCYP11B2 to include the dynamic behaviour of the proteins, i.e. we have included multiple protein states. These states are the conformations of the protein structures that have been sampled during the molecular dynamics simulations. In this chapter, the amount of inhibitors has also been increased to include fadrazole analogues that have been synthesised during the progression of the project (Figure 4–2).

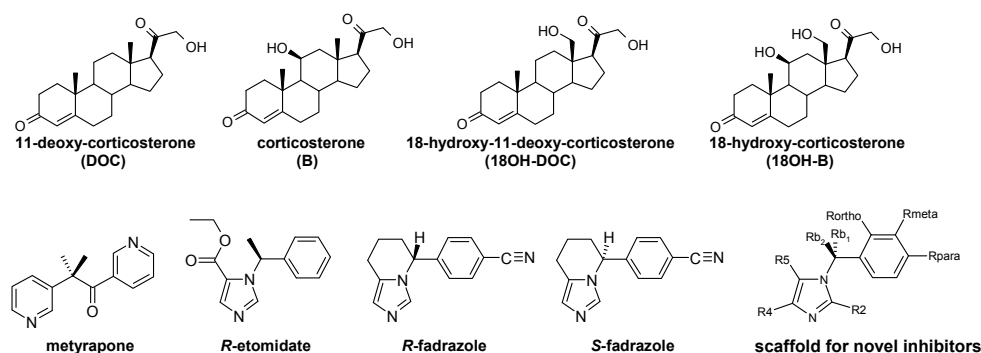


Figure 4–2 Structures of the four steroids, the four inhibitors and the general scaffold of the novel CYP11B inhibitors.

#### 4.4 Docking Settings

Substrate docking was performed using GOLD v3.0. The docking parameters for the genetic algorithm were taken from the *default /* settings in GOLD. The protein structures used for the docking of the substrates were the unequilibrated structures of hCYP11B1 and hCYP11B2. An important detail is that the docking was performed in the presence of an iron-bound oxygen atom. Each steroid was docked for 100 poses to obtain an initial binding mode for the steroid in the active site cavity. Their resulting conformations were afterwards checked to investigate alternate orientations of the steroids in the active site cavity, and to investigate the positioning of C<sub>11</sub>, C<sub>18</sub> and C<sub>19</sub>.

Similar to the docking of the steroids, the docking of the 4 known inhibitors was performed using GOLD v3.0. The docking parameters were taken from the *default /* settings, and extra cytochrome P450 parameters [8] were added to the scoring functions to simulate the correct interaction strength between the protein iron atom and different ligand acceptor groups. The protein structures used for this docking study were hCYP11B1 and hCYP11B2 after they were equilibrated using the ligand 18-hydroxycorticosterone (paragraph 2.7.3 ). This protein-ligand complex was chosen, because it defines the key differences between the catalytic activity of hCYP11B1 and hCYP11B2, being that 18-hydroxycorticosterone is only a substrate for hCYP11B2. Each ligand was docked 5 times for 100 poses each to ensure an appropriate sampling of ligand conformations in the protein active sites. The GOLD fitness score was obtained by using the standard GoldScore function, which was subsequently averaged for the top 10 ranking poses. To obtain an impression of the correlation between the docking score and in vitro data, the GoldScore measure of binding affinity was calculated by the method reported by Verdonk et al [7] per  $\Delta G_{\text{binding}} = -0.1075 * \text{GoldScore} - 2.2665$  ( $R^2 = 0.5529$ ,  $N=60$ ,  $\Delta G_{\text{binding}}$  expressed in kcal/mol). The average poses were additionally rescored using the ChemScore function. The best ranked poses were used for the molecular dynamics study in Chapter 3.

The docking of the novel inhibitors has been performed using GOLD v3.2. The same *default /* settings were chosen for the genetic algorithm, and the extra cytochrome P450 parameters have been added to the scoring function. The protein structure resulting from the ligand

equilibration phase has been used for docking. It is labelled as the homology model because it is still comparable to the starting structure (with an RMSD < 2.5 Å for all heavy atoms). The protein states sampled during the molecular dynamics simulation were extracted every 25 ps of the simulation and were subsequently used for docking. Each ligand has been docked in each of the protein states for 50 poses. The ligands that have been used for docking can be found in Appendix B, carrying the "prediction" label. The GOLD fitness score was used for the docking and ranking, after which the poses were rescored using the ChemScore function.

#### 4.5 Statistical Evaluation of Docking Results

A statistical analysis has been carried out for the docking in the different protein states. This analysis has been performed to obtain an indicative measure for the correlation between docking score and the potency measured by in vitro experiments. First, a student's t-test has been performed on the null-hypothesis that there is no correlation between the docking scores ( $y$ ) and the pIC50 values. In this case, a P-value close to 0 means that the null-hypothesis can be rejected, hence both series are correlated. Next, a least squares regression has been performed for a linear relationship as well as the analysis of variance (ANOVA) to produce an F-statistic for the significance of the regression fit ( $\hat{y}$ ). This least squares analysis involves the calculation of the total sum of squares (TSS, Equation 4–5), the explained sum of squares (ESS, Equation 4–6) and the residual sum of squares (RSS, Equation 4–7). The TSS represents the variance of the docking data, the ESS the variance of the fit that can be explained due to fitting, and the RSS the discrepancy between the docking data and the regression fit. The Pearson correlation coefficient of the fit can be determined using Equation 4–8 and the F-statistic of its significance using Equation 4–9 (with  $n$  as the degrees of freedom of the docking data, and  $p$  the degrees of freedom of the fit).

$$\text{Equation 4-5} \quad TSS = \sum_{i=1}^N (y_i - \bar{y})^2, \text{ with } \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

$$\text{Equation 4-6} \quad ESS = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$$

$$\text{Equation 4-7} \quad RSS = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

$$\text{Equation 4-8} \quad R_{\text{Pearson}}^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$\text{Equation 4-9} \quad F = \frac{RSS/p}{ESS/(n-p-1)}$$

## 4.6 Results and Discussion

### 4.6.1 Substrate Docking

The docking results of the steroids all portrayed the steroid skeleton in one and the same orientation in the active site parallel to the heme. This conformation contains the  $\beta$ -side of the steroid skeleton facing the heme, with all three hydroxylation sites C<sub>11</sub>, C<sub>18</sub> and C<sub>19</sub> in proximity to the iron oxygen species. In this conformation the C<sub>3</sub>-ketone group of the steroid interacts with R123 in helix B', and the C<sub>20,21</sub>-hydroxyacetyl group interacts with the loop region following helix K (Figure 2–11).

The GoldScore of 18-hydroxy-11-deoxycorticosterone and 18-hydroxycorticosterone are higher for both human and rat CYP11B2 than for human and rat CYP11B1. Although this indicates that their affinity is higher for the CYP11B2 models, no conclusions can be drawn because the protein-ligand interactions have not yet been optimised (pose 1, Table 4–1). The steroids were also docked with the  $\beta$ -side oriented to the heme and the positions of the C<sub>3</sub>-ketone and the C<sub>20,21</sub>-hydroxyacetyl groups exchanged. This resulted in a consistently low GoldScore, indicating that this is a less profitable conformation for the steroids (pose 2, Table 4–1). Orientations of the steroid with the  $\alpha$ -side facing the heme was not observed during the docking.

As such, the  $\beta$ -side orientation of the steroid with the C<sub>3</sub>-ketone interacting with R123 and the C<sub>20,21</sub>-hydroxyacetyl group interacting with the active site loop were used for the protein-substrate equilibration study in paragraph 2.7.3. Because the three hydroxylation points of the steroid are presented to the heme, and the steroids possess such pronounced interactions in the model active sites, we conclude that is the most likely binding mode for the steroids in the CYP11B proteins.

Table 4–1 Docking results for the different steroids. Pose 1 possesses C<sub>11</sub>, C<sub>18</sub> and C<sub>19</sub> close to the iron-oxygen, Pose 2 only C<sub>18</sub> and C<sub>19</sub>.

GoldScore	hCYP11B1		hCYP11B2		hCYP11B2		rCYP11B1		rCYP11B2	
	mean	s.d	mean	s.d	mean	s.d.	mean	s.d.	mean	s.d.
Pose 1 (n=10)										
DOC	41.16	1.00	42.56	0.81	41.42	0.42	42.05	1.21	41.36	0.77
B	42.40	3.14	43.40	1.21	41.68	1.05	43.28	1.19	42.46	0.66
18OH-DOC	41.32	2.27	41.88	1.12	45.95	1.14	41.72	0.77	44.53	
18OH-B	39.32	1.15	41.20	1.11	46.09	2.23	41.37	0.69	45.52	1.60
Pose 2 (n=10)										
DOC	35.68	0.44	33.82	0.99	32.02	1.11	30.37	0.68	35.38	1.04
B	35.02	1.25	35.17	0.94	35.43	2.10	31.22	1.40	35.66	2.07
18OH-DOC	36.12	1.17	34.82	1.34	33.44	1.43	34.39	2.15	34.18	2.68
18OH-B	34.03	1.34	35.77	1.60	34.81	2.68	34.21	2.24	36.46	1.15

#### 4.6.2 Docking of the Four Known Inhibitors

The four known inhibitors all bind in one dominant orientation in the hCYP11B1 and hCYP11B2 models. All inhibitors possess their aromatic nitrogen atom interacting with the heme iron atom. In addition to this interaction, they also portray a ring stacking interaction between the distant aromatic ring and the F130 residue located in helix B'. The docking score of the ligands can be found in Table 4–2.

As discussed in Chapter 3, Metyrapone possesses two pyridine rings that can both bind to the iron atom, however, the nitrogen indicated in Figure 3–1, is the one that predominantly binds the heme. The carbonyl moiety of metyrapone is not observed to interact with any active site residue, but points in the general direction of R110 that stabilises the heme propionate groups in the active site. The two methyl groups possess hydrophobic interactions in the active site cavity overlapping the cavity where the B-ring of the steroids binds. R-etomidate possesses a hydrogen bond to the catalytic threonine T318 in both the active sites of hCYP11B1 and hCYP11B2. Due to the presence of rotatable bonds, the ethyl group of R-etomidate protrudes either towards the top of the active site or towards the active site loop following helix K. The overall orientation of these conformations was very similar and they overlap the region in the cavity where the D-ring of the steroids binds. R-fadrazole and S-fadrazole possess an electrostatic interaction with R123 in helix B'. Steric aspects of the active site cavity around helix I and helix B' are important for the binding mode of the two enantiomers. In hCYP11B1, the active site cavity favours the S-enantiomer, whereas the R-enantiomer is favoured by hCYP11B2. The precise differences have been described in paragraph 3.5.2 , and the best ranking orientation of all ligands has been used as the input for the molecular docking study in Chapter 3.

Table 4–2 Trend for the predicted docking scores and in vitro inhibitor potency

	IC50 (nM)		GoldScore (Fitness)		ΔG GoldScore (kcal/mol)		ΔG ChemScore (kcal/mol)	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
hCYP11B1								
Metyrapone	46.4	10.4	57.33	0.86	-8.43	0.09	-8.73	0.15
R-Etomidate	0.5	0.2	66.21	0.66	-9.38	0.07	-9.25	0.12
R-Fadrazole	118.6	8.9	54.01	0.48	-8.07	0.05	-8.14	0.21
S-Fadrazole	39.5	4.4	56.67	0.61	-8.36	0.07	-8.77	0.05
hCYP11B2								
Metyrapone	207.8	4.5	49.99	0.92	-7.64	0.10	-7.95	0.22
R-Etomidate	1.7	0.9	65.21	0.51	-9.28	0.05	-9.21	0.10
R-Fadrazole	6.0	1.9	63.20	0.76	-9.06	0.08	-9.38	0.12
S-Fadrazole	171.2	51.7	53.81	0.73	-8.05	0.08	-8.12	0.24

#### 4.6.3 Docking of Fadrazole Analogues

Even though the usage of molecular docking as a predictive tool for inhibitor potency possesses several caveats (as discussed in paragraph 4.2 ), we have investigated the correlation between docking score and in vitro measurement data as a guide for the potency prediction of novel inhibitors. To investigate whether the homology models have performed accurately so far, the docking study has been expanded to include multiple protein active site conformations as sampled by the molecular dynamics study in Chapter 3. The main requirement for the docking pose of the inhibitors is the interaction between the imidazole nitrogen atom lone-pair with the heme iron atom. Several of the docked compounds were unable to obtain this interaction and were discarded from the results ("Prediction" label in Appendix B). These ligands are either substituted on the R2 and R4 positions of the imidazole, which obstructs their interaction with the heme in our models (n=7), possess carboxylate or ester groups, which are sometimes docked interacting with the heme iron (n=5), or fail for other reasons (too big or too rigid).

Performing the t-test revealed that the correlation between the docking results and the pIC50 values is significant for all protein models (P-value less than  $10^{-7}$ , Table 4-3). Next, the expected linear relationship between the docking score and the pIC50 values is fit on the data, extracting the slope (A), intercept (B), correlation coefficient ( $R^2$ ) and the F-ratio for the linear relationship (Table 4-3). All F-ratios are above their critical value for a 1% confidence level, hence the correlation fit is significant.

Table 4–3 Results for the T-test, ANOVA correlation and the F-test for the docking with the GoldScore and ChemScore in the various CYP11B1 and CYP11B2 models. The ligand indicates whether the protein state results from molecular dynamics using 18-hydroxycorticosterone or fadrazole (*R*-fadrazole for CYP11B2 and *S*-fadrazole for CYP11B1). Indicated are the P-value for the T-test and the F-ratio for the F-test, as well as the correlation coefficient (R2) for a linear equation, with slope (A) and intercept (B). The average model results are obtained by averaging the docking results for all protein states of the molecular dynamics simulation.

Score	Model	Ligand	Entire Dataset (n=106, Fcrit=2.41, α=0.01)					Reduced Dataset (n=87, Fcrit=2.46, α=0.01)				
			T-test	R2	F-value	A	B	T-test	R2	F-value	A	B
GoldScore 11B1	Homology	18OH-B	e-15	0.50	93	4.52	17.86	e-24	0.75	234	6.43	10.58
	Average	18OH-B	e-17	0.53	104	4.22	21.75	e-23	0.74	212	4.82	17.60
	Homology	Fadrazole	e-12	0.41	65	4.15	22.28	e-21	0.70	171	4.88	16.89
	Average	Fadrazole	e-13	0.44	76	3.92	23.35	e-21	0.71	178	4.44	19.89
GoldScore 11B2	Homology	18OH-B	e-13	0.45	74	4.52	20.80	e-26	0.78	260	5.21	16.74
	Average	18OH-B	e-12	0.41	63	3.33	26.05	e-22	0.72	191	4.06	19.85
	Homology	Fadrazole	e-14	0.45	77	3.77	22.90	e-20	0.69	162	4.27	18.98
	Average	Fadrazole	e-13	0.45	71	3.42	24.12	e-20	0.69	162	3.99	19.30
ChemScore 11B1	Homology	18OH-B	e-8	0.29	38	-3.49	-10.59	e-15	0.57	96	-4.01	-5.54
	Average	18OH-B	e-11	0.37	55	-3.36	-9.85	e-15	0.58	100	-2.63	-7.02
	Homology	Fadrazole	e-8	0.30	39	-3.27	-9.72	e-12	0.48	68	-3.54	-6.86
	Average	Fadrazole	e-10	0.35	49	-3.20	-11.17	e-15	0.58	100	-2.78	-10.33
ChemScore 11B2	Homology	18OH-B	e-11	0.39	58	-3.74	-5.32	e-14	0.56	93	-3.92	-3.37
	Average	18OH-B	e-9	0.30	40	-3.12	-9.88	e-15	0.58	102	-3.44	-6.20
	Homology	Fadrazole	e-11	0.37	54	-3.53	-6.58	e-13	0.53	82	-3.65	-4.96
	Average	Fadrazole	e-7	0.26	32	-2.67	-12.03	e-13	0.51	76	-2.78	-10.33

#### 4.6.3.1 Comparison of scoring functions and protein models

Overall, the pIC50 correlates better with the GoldScore than with the ChemScore. This can be derived from either the T-test or the F-test. The reason for the higher performance is that the hydrophobic contribution to the potency is very high and provides problems for accurate predictions made by the ChemScore. The cavity is hydrophobic around the regions where the steroid rings fit, with small hydrophilic regions around the extremes of the steroid, being the C<sub>3</sub>-ketone group and the C<sub>20,21</sub>-hydroxyacetyl group. The fadrazole analogues possess a hydrophobic benzyl-imidazole scaffold that binds this cavity and is responsible for the majority of its interactions. Of course the catalytic T318 in the hydrophobic core of the active site can supply a hydrophilic interaction to the ligand, such as observed for *R*-etomidate (paragraph 3.5.2), however, the binding mode of the imidazole obstructs all but the R5-substituents from interacting with T318. Hence, not many ligands possess a polar or hydrogen bonding partner.

Unlike the difference in the scoring function, the results of the docking for the homology models is comparable to that for the MD sampled protein states (Figure 4–3). This indicates that the docking of ligands in the homology models of CYP11B1 and CYP11B2 is not improved by taking the dynamic behaviour of the proteins into account. Furthermore, the relevance of which ligand is used for equilibration is negligible, although the highest



correlation and lowest man square error are obtained from protein states equilibrated with the ligand 18OH-B. Intuitively, the models equilibrated with fadrazole would allow better docking because the ligands investigated are all analogues of fadrazole, however, 18OH-B fills up the entire cavity, thus preventing the active site from collapsing inward and making more space for larger ligands to fit.

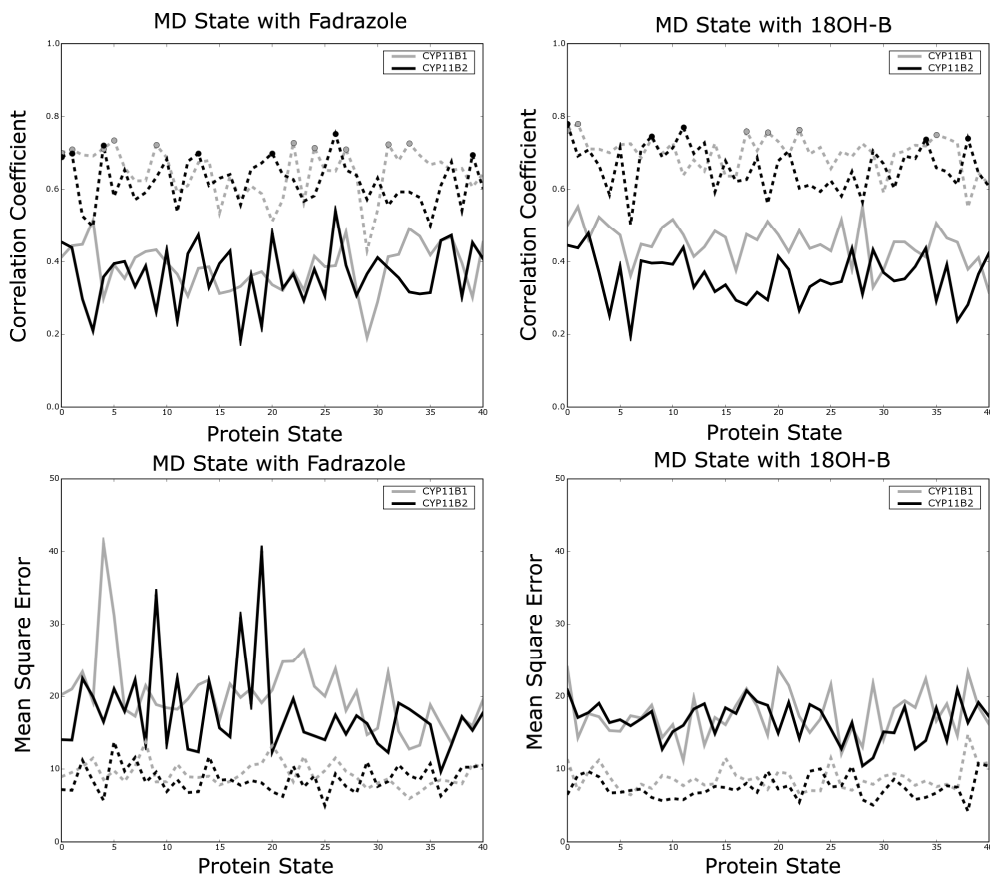


Figure 4–3 GoldScore docking results for the different protein conformations sampled by molecular dynamics using the ligands Fadrazole and 18-hydroxycorticosterone (18OH-B). Indicated are the correlation coefficient and the mean square error for the entire dataset (solid lines) and the pruned dataset (dashed lines). CYP11B1 is indicated in grey, whereas CYP11B2 is indicated in black. Protein states possessing a comparable quality as the homology models (frame 0) are indicated with a circle. The circles are coloured according to that of the respective proteins.

#### 4.6.3.2 Prediction models for non-Rb-substituted fadrazole analogues

Since the initial correlation between the docking score and the pIC50 was low, the dataset has been investigated towards the docking behaviour of ligands with similar substituent groups. An example of the results can be seen in Figure 4–4 for CYP11B2 docking results. All other results are depicted in Appendices C.1.1 to C.2.2. Ligands possessing a substituent that belongs to a certain group are coloured black in the subplot for that particular group. Many ligands contain more than one of the investigated substituent groups and are thus represented in more than one subplot. From the subplots it can be seen that a mismatch for the linear relationship is observed for the unsubstituted group (R5-H) as well as the substituent groups Rp-CN, R5-CCCR and Rb-X (with X as any substituent except for H).

Comparing these groups, the only one that contains almost all of the outliers is that of the Rb-substituted fadrazole analogues. These compounds are mispredicted because the Rb-substituent points into the active site entrance. During the simulation the active site entrance opens and the Rb-substituents of the docked ligands will protrude out of the active site, obtaining small interaction with the protein. The docking of most of these ligands in the homology model creates outliers (Appendices C.1.1 to C.2.2), indicating that perhaps the model accuracy of the active site entrance is not sufficient for the prediction of the potency of this group.

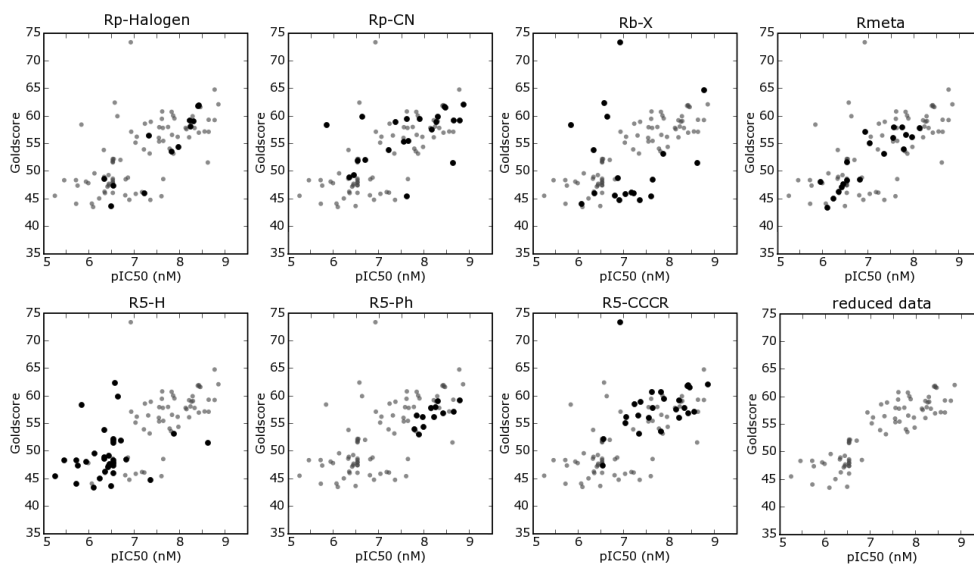


Figure 4–4 Comparison of compound docking results in CYP11B2 models to their pIC50 values. Docking has been performed in protein states resulting from molecular dynamics simulations with the ligand 18-hydroxycorticosterone. Each subplot also shows the docking results for specific substituent groups in black. The last subplot shows the reduced dataset

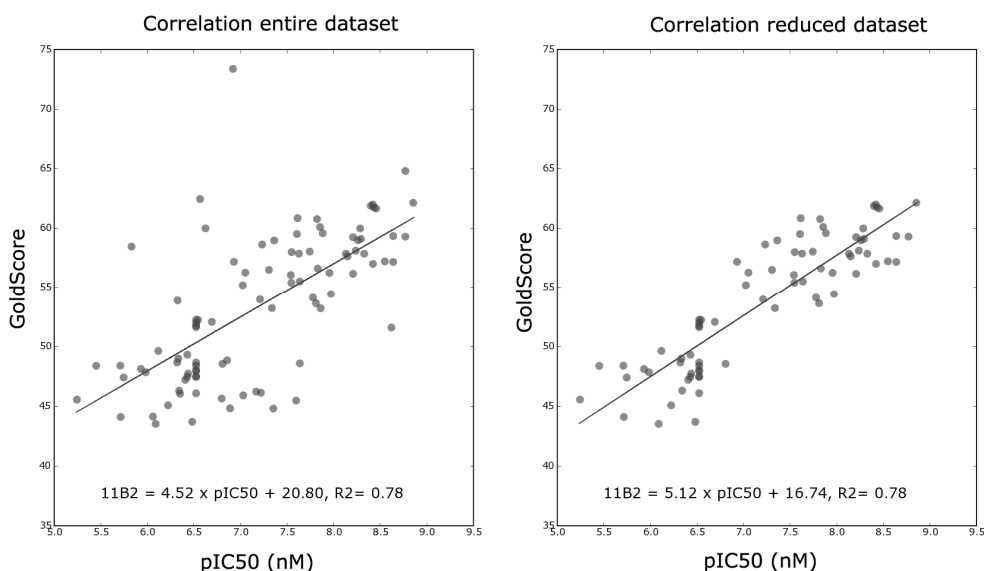


Figure 4–5 Correlation between the GoldScore and the pIC50 for all data points. The left picture indicates the entire dataset, whereas the right picture the reduced dataset that does not contain Rb-substituted compounds. The fit for a linear relationship is also indicated as well as its formula.

By removing the Rb-substituted compounds from the dataset, a more accurate prediction model can be constructed for novel fadrazole analogues that are not Rb-substituted. A similar T-test and F-test analysis has been carried out for this reduced dataset as depicted in Table 4–3, and the correlation and mean square errors have been determined as depicted in Figure 4–3. The docking result for the CYP11B2 homology model equilibrated using the ligand 18OH-B can be seen in Figure 4–5. After dataset reduction, the slope of the correlation has increased and the intercept has decreased. These changes are higher for both the CYP11B1 and CYP11B2 homology models equilibrated with 18OH-B compared to the changes for the other models. They also possess a relatively high correlation, although their mean square error remains comparable to the other protein states (Figure 4–3). Because both homology models either equilibrated with fadrazole or 18OH-B possess among the highest of correlations, they can be used as an accurate tool for the potency prediction of novel CYP11B inhibitors. In addition, protein states extracted from the molecular dynamics simulation with comparable accuracy can be used as validation.

#### 4.6.3.3 Prediction of novel inhibitors

A selection of novel CYP11B inhibitors (Figure 4–6) has been docked into the different protein conformations, followed by the prediction of their in vitro potency based on the regression models. These compounds can also be found in Appendix B, with the Prediction label. For the compounds that do not contain an Rb-substituent, the pruned regression models are used, whereas for the other compounds the regression models based on the entire dataset are used.

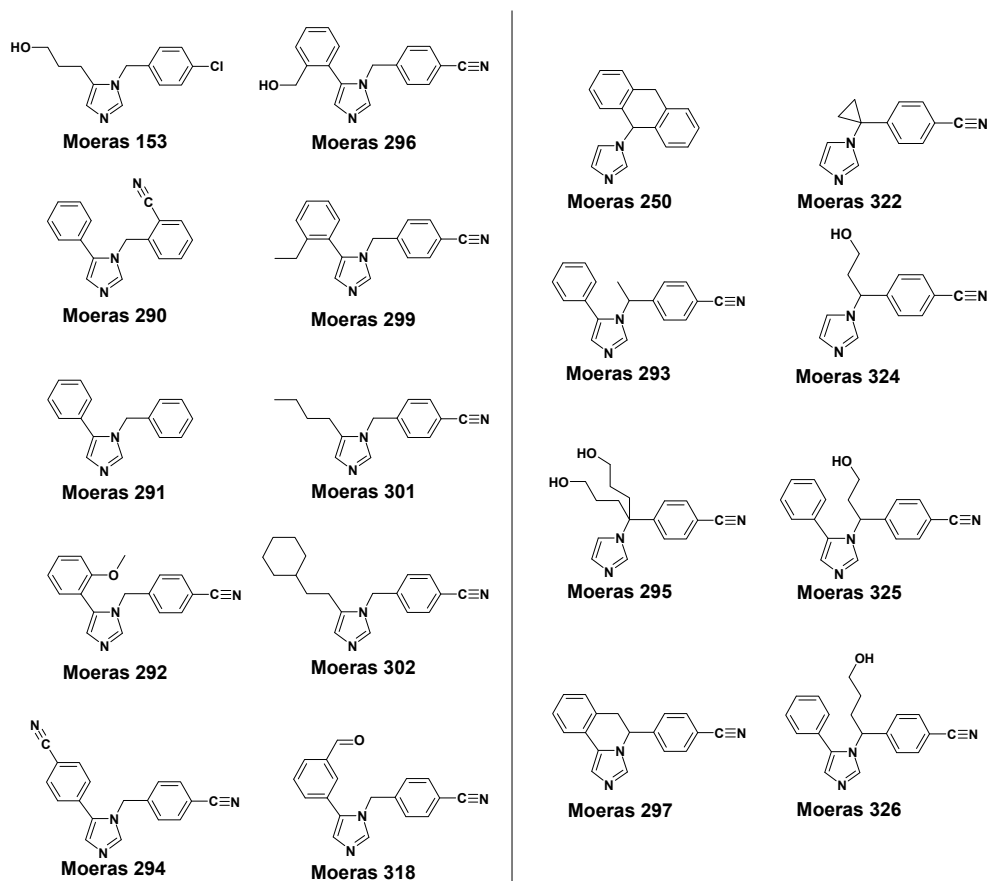


Figure 4–6 Novel CYP11B inhibitors that have not been used to derive the regression models, but are used to predict their binding affinity. Compounds that do not possess an Rb-substituent are shown on the left; compounds possessing an Rb-substituent are shown on the right.

The CYP11B1 and CYP11B2 regression models are able to predict inhibitor potency for two-third of the compounds (Figure 4–7). The largest outliers, Moeras 294 and 302 contain an elongated chain as R5-substituent which provides a large steric contribution to the docking score. This contribution may be incorrect and the cavity size may in reality be smaller, decreasing the affinity. The other deviations can be attributed to the variance with the regression models. For the Rb-substituted compounds, the regression models also behave rather well considering their low correlation coefficient. Moeras 250, 295 and 297 are predicted too potent, which is probably related to their overall size which fills up the entire active site cavity.

The pIC50 is systematically underpredicted by the homology model equilibrated with 18OH-B. This is unexpected since this model is able to distinguish the individual data points better than the other regression models. The explanation is the presence of a systematic error in the regression models for compounds substituted with the R5-phenyl group (pIC50 underestimation is  $0.78 \pm 0.10$  for CYP11B1 and  $0.81 \pm 0.14$  for CYP11B2,  $n=19$ ). The cause of this discrepancy is a sub-optimal orientation of F487 in the active site, as seen in Figure 4–8. For the homology model, the phenylalanine ring is twisted to optimise its interaction with the steroid, however, the orientation of the R5-phenyl ring results in an obstruction (light model). If the phenylalanine ring is turned, such as observed for the molecular dynamics CYP11B2 protein state 8 ( $R^2_{\text{reduced}}$  is 0.74), the R5-phenyl ligands will create a ring-stacking, increasing their interaction and thus docking score (dark model).

Because the error is systematic, the prediction by the 18OH-B homology model may be optimised by performing baseline correction. The average deviation of the R5-phenyl compounds is calculated and the prediction results are adjusted accordingly (Figure 4–9). The adjustment term is only unfavourable for Moeras 294 in the CYP11B2 model, however, as mentioned, the pIC50 prediction of this compound may be debatable.

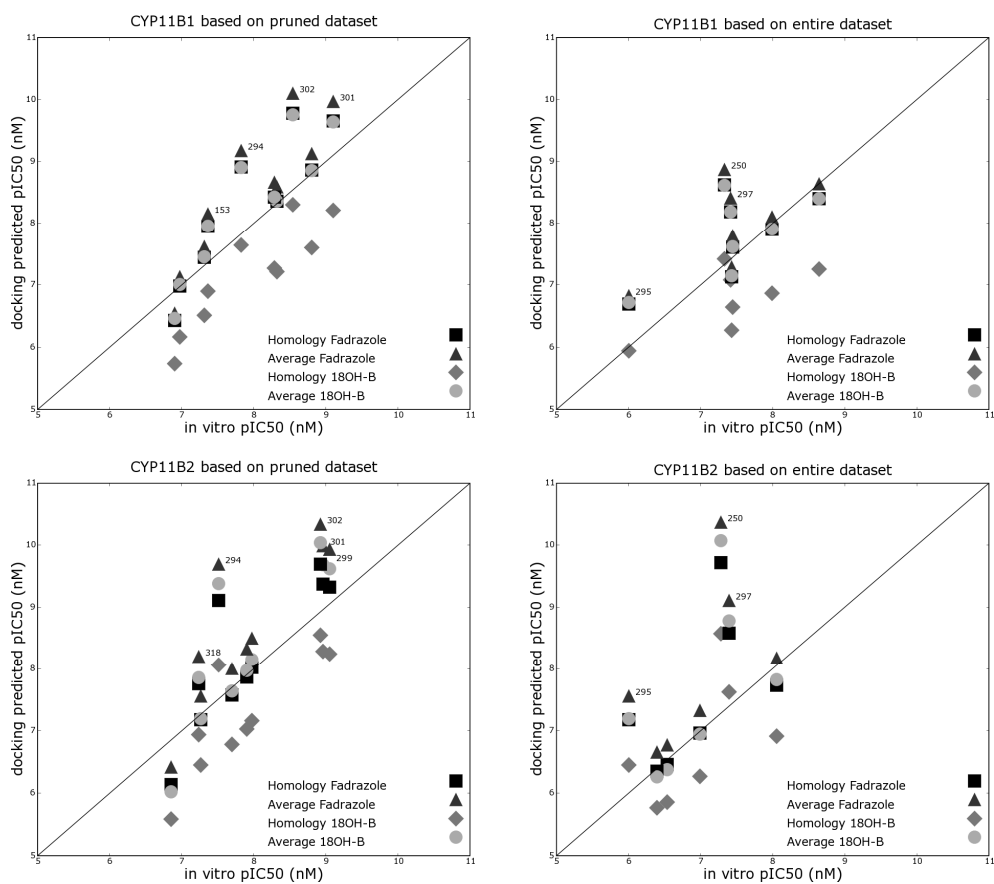


Figure 4–7 Predicted CYP11B1 and CYP11B2 pIC50 values by docking using the correlation equations of the pruned dataset for non-Rb-substituted compounds and the correlation equations of the entire dataset for Rb-substituted compounds. Linear performance is indicated by the black line. Outliers are indicated with their Moeras numbering. The legend symbols are shown on the bottom right of the pictures.

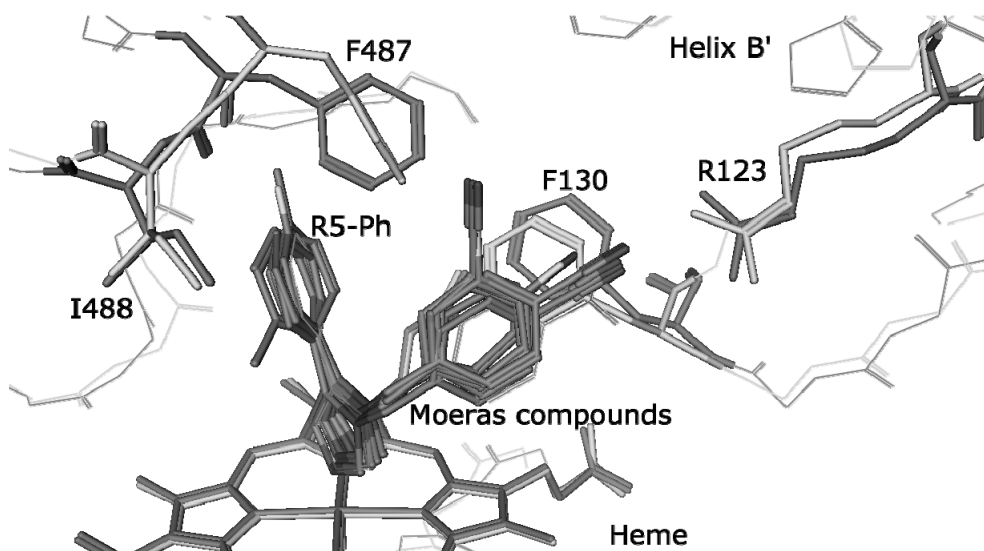


Figure 4–8 CYP11B2 homology model equilibrated with 18-hydroxycorticostere (dark grey) compared to one of its molecular dynamics states (light grey, state 8,  $R^2_{\text{reduced}}=0.74$ )

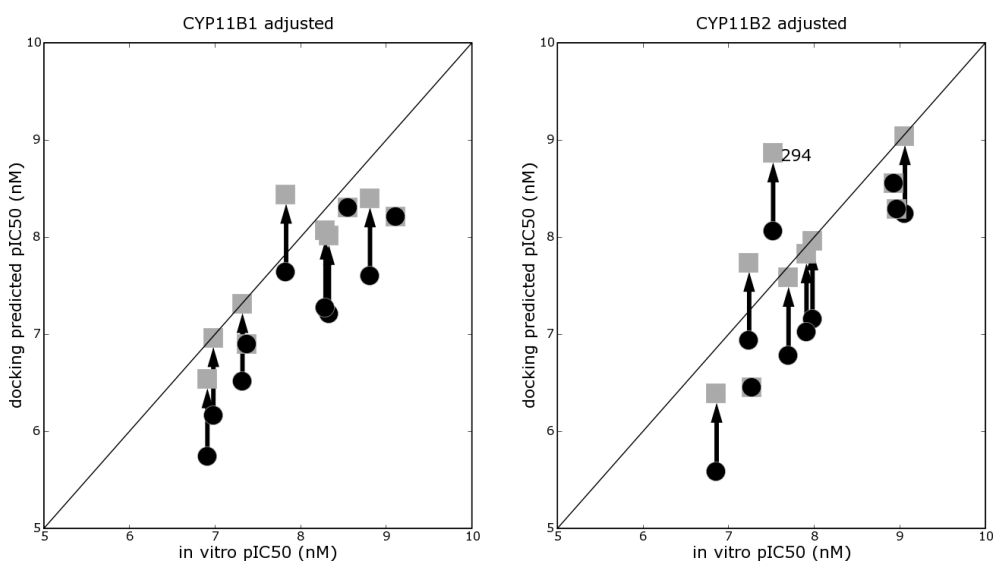


Figure 4–9 Adjusted prediction for the regression using the homology model with the ligand 18OH-B. The blue dots indicate the original regression result. The arrows and red squares indicate the adjustment of the predicted pIC<sub>50</sub> values.

An alternative method is to investigate the docking of R5-substituted compounds in the protein states extracted from molecular dynamics. The investigation of the CYP11B2 protein state 8 equilibrated with 18OH-B shows that the baseline prediction of the R5-substituted compounds is better than that observed for the homology model (Figure 4–10). The downside of this protein state is that the baseline for the R5-alkyl substituent (R5-CCCR) is differing from the correlation fit. Analysis of other protein states that possess a high correlation after dataset reduction showed that CYP11B1 protein state 2 and 19 also possesses a rotated F487 and consequently allow prediction for R5-phenyl substituted compounds. None of the other molecular dynamics states clearly possess a proper baseline for R5-substituted compounds.

Prediction of the potency of the novel CYP11B inhibitors using the molecular dynamics protein states portrays a very good correlation for all compounds (Figure 4–11). As such, different protein states can be used for the prediction of substituted benzyl-imidazoles, although each of the protein states will possess a different predictive value dependent on the substituent position and type.

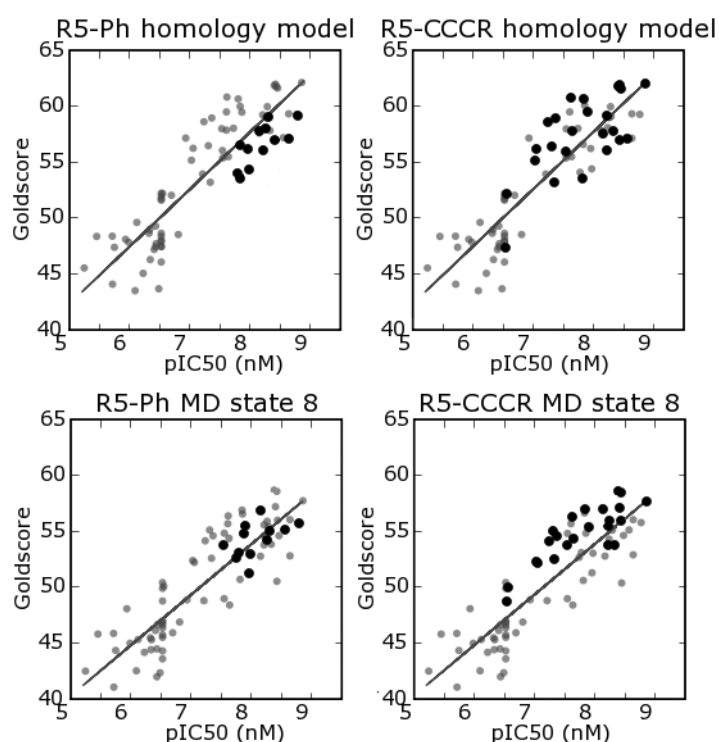


Figure 4–10 Docking results for the reduced datasets in the CYP11B2 homology model and the molecular dynamics protein state 8, both equilibrated with 18-hydroxycorticosterone. Indicated are the correlation of the R5-phenyl substituted compounds as well as the R5-alkyl substituted compounds. The R5-phenyl and R5-alkyl substituted compounds are indicated with black dots.



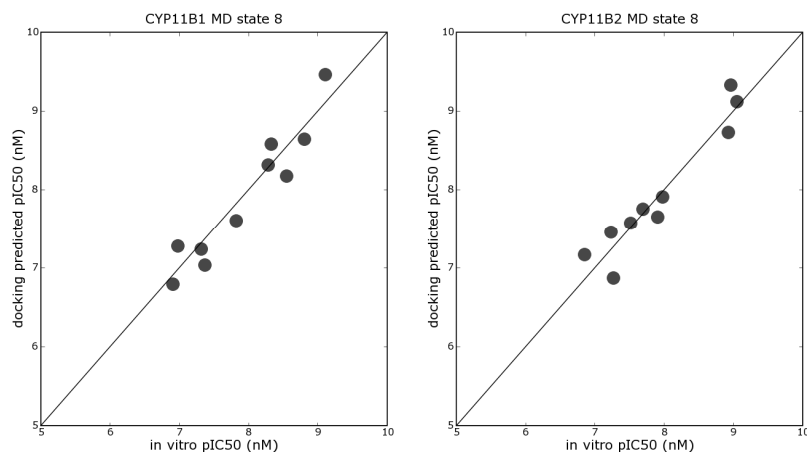


Figure 4–11 Predicted CYP11B1 and CYP11B2 pIC50 values by docking using the correlation equations of the pruned dataset for non-Rb-substituted compounds for CYP11B1 protein state 2 and CYP11B2 protein state 8. Optimal performance is indicated by the black line.

#### 4.7 Conclusions

The docking of the steroids resulted in one dominant binding mode featuring an interaction between the steroid C<sub>3</sub>-ketone group with R123 in helix B' and the C<sub>21</sub>-hydroxyl group interacting with the backbone of G379 and F381. The presence of F130 is also important to stabilise the steroid structure in the active site. This orientation of the ligand provides all three hydroxylation sites in close proximity to the heme iron, strengthening the confidence in the docking pose. Alternative orientations of the ligand are allowed, however they result in a much lower docking score and possess less or no hydroxylation sites close to the heme.

The known non-steroidal CYP11B inhibitors metyrapone, R-etomidate, R-fadrazole and S-fadrazole were postulated to be localised in a comparable manner as the endogenous substrates. All inhibitors bind to the heme with their aromatic nitrogen atom and interact with F130 forming a stabilising ring-ring interactions. The two scoring functions GoldScore and ChemScore both corroborate the measured activity data from in vitro experiments, supporting the validity of the binding modes predicted by the docking study. An important observation is that the docking rationalises the enantioselectivity of fadrazole, with the R-enantiomer being most potent on CYP11B2 and the S-enantiomer being most potent on CYP11B1.

Next, the docking study was expanded to include multiple protein active site conformations as sampled by the molecular dynamics study. The performance of a large-scale docking study of various fadrazole analogues showed that the correlation of the docking results with the measured activity data is low ( $R^2_{\text{Goldscore}} \sim 0.5$  and  $R^2_{\text{Chemscore}} \sim 0.3$ ). Investigation of compounds with similar substituent groups revealed that the outliers are mainly caused by Rb-substituted compounds. By neglecting the docking data of these compounds, a

regression can be obtained for the prediction of non-Rb-substituted compounds. The correlation coefficient of these regression models is good ( $R^2 \sim 0.7$ ) and significant.

Comparing the docking results in the molecular dynamics sampled protein states reveals that the results are comparable to the docking obtained in the homology models. As such, using the dynamic behaviour of the constructed models for docking might be an appropriate tool for the prediction of protein-ligand interactions and the determination of inhibitor potency.

The flexibility of F487 in the active site cavity plays an important role in the docking of R5-substituted compounds. The homology models of CYP11B1 and CYP11B2 equilibrated with 18OH-B, possess the highest correlation between docking score and *in vitro* measured potencies, however, the R5-phenyl substituted compounds are systematically underpredicted. This is caused by a suboptimal orientation of F487 which obstructs the phenyl group in the active site cavity. Investigation of different protein states sampled by molecular dynamics has led to the elucidation of three models that possess the alternate orientation of F487. Docking of the R5-phenyl substituted compounds results in a ring-ring interaction between the phenyl groups of the amino acid residue and ligand. Unfortunately, these models lose the ability to predict other R5--substituents correctly, as their interaction with F487 has also changed. Thus, the rotation of F487 plays an important role for the stabilisation of different R5-substituents, which implies that for the molecular docking of R5-substituted compounds, multiple active site conformations are of importance and need to be investigated.

Molecular dynamics of the protein active sites has resulted in an unfolding of the active site entrance, causing the prediction of Rb-substituted compounds to become inaccurate for all investigated protein states. The prediction of Rb-substituted compounds remains difficult and requires further model optimisation, starting at the level of homology model construction. In particular, detail should be spent in elucidating the correct conformation of the helices F and G that close the active site lid. Currently the conformation of these helices is similar to that of the template CYP2C5 and their position is relatively the same for all other cytochrome P450 crystal structures. As such, no immediate information can be derived for the remodelling of these helices and, therefore, we have not performed this exercise.

## Literature

- 1 D.T. Moustakas, P.T. Lang, S. Pegg, E. Pettersen, I.D. Kuntz, N. Brooijmans, R.C. Rizzo, "Development and validation of a modular, extensible docking program: DOCK 5", *J. Comput.-Aided Mol. Des.*, 2006, 20, 10-11, 601-619
- 2 G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function", *J. Comp. Chem.*, 1999, 19, 14, 1639-1662
- 3 R. Abagyan, M. Totrov, D. Kuznetsov, "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, 5, 488-506
- 4 R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, D.E. Shaw, M. Shelley, J.K. Perry, P. Francis, P.S. Shenkin, "Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy", *J. Med. Chem.*, 2004, 47, 7, 1739-1749
- 5 T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, J.L. Banks, "Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening", *J. Med. Chem.*, 2004, 47, 7, 1750-1759
- 6 M. Rarey, B. Kramer, T. Lengauer, G. Klebe, "A fast flexible docking method using an incremental construction algorithm", *J. Mol. Biol.*, 1996, 261, 3, 470-489
- 7 M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, "Improved protein-ligand docking using GOLD", *Proteins: Structure, Function, and Genetics*, 2003, 52, 4, 609-623
- 8 B. Kirton, C.W. Murray, M.L. Verdonk, R.D. Taylor, "Prediction of binding modes for ligands in the cytochromes P450 and other heme-containing proteins", *Proteins: Structure Function and Bioinformatics*, 2005, 58, 4, 836-844
- 9 M.L. Verdonk, G. Chessari, J.C. Cole, M.J. Hartshorn, C.W. Murray, J.W.M. Nissink, R.D. Taylor, R. Taylor, "Modeling water molecules in protein-ligand docking using GOLD", *J. Med. Chem.*, 2005, 48, 20, 6504-6515
- 10 M.J. Hartshorn, M.L. Verdonk, G. Chessari, S.C. Brewerton, W.T.M. Mooij, P.N. Mortenson, C.W. Murray, "Diverse, high-quality test set for the validation of protein-ligand docking performance", *J. Med. chem.*, 2007, 50, 4, 726-741
- 11 M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, "Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes", *J. Comput-Aided Mol. Des.*, 1997, 11, 5, 425-445
- 12 C.A. Baxter, C.W. Murray, D.E. Clark, D.R. Westhead, M.D. Eldridge, "Flexible docking using tabu search and an empirical estimate of binding affinity", *Proteins: Structure, Function, and Genetics*, 1998, 33, 3, 367-382
- 13 W.T.M. Mooij, M.L. Verdonk, "General and targeted statistical potentials for protein-ligand interactions", *Proteins: Structure, Function, and Bioinformatics*, 2005, 61, 2, 272-287

## Chapter 5 Structure Activity Relationships

The previous chapters have discussed protein-ligand interactions with focus on the protein characteristics that contribute to the ligand binding affinity. We have shown that certain amino acid conformations are important for the correct elucidation of protein-ligand interactions. However, it is possible that the dynamic sampling has not encompassed all possible protein states responsible for ligand binding. Because the protein homology models are static structures and active ligand binding protein states can be missed by molecular dynamics sampling, a ligand-based approach can lead to the elucidation of new structural insights. In this chapter we introduce quantitative structure activity relationships, which are used to determine whether the potency of ligands can be related to the presence of specific ligand substructure features. In particular, we have employed the methodology of decision tree analysis for the potency prediction of novel CYP11B inhibitors.

## 5.1 Quantitative Structure Activity Relationships

In this thesis, we have derived structure-activity relationships for the prediction of newly synthesised CYP11B2 inhibitors as well as the prediction of their selectivity. A structure-activity relationship (SAR) analysis is the basis for understanding the effect of the structural features of both the ligands as well as the protein active site, even if the latter initially is a black box. The analysis involves the investigation of ligand properties or on occasion protein properties for the elucidation of the optimal protein-ligand interactions. The ligand is regarded as a combination of substructures that each carry their own physico-chemical properties that allow the substructures to interact with local regions within the protein active site. The substructures are referred to as pharmacophores if the substructures prove to be essential for protein-ligand interactions. The evaluation of the physico-chemical properties and their effect on ligand binding allows the identification of those molecular entities that are most determining for inhibitor potency. Therefore, structure-activity relationships aid in the design of more potent inhibitors by combining several of the best substituents or substituent properties.

When the physico-chemical properties of structures are correlated to biological activity of the ligands, the analysis is called a quantitative structure-activity relationship (QSAR). Originally, SARs were descriptive relating the biological activity to only a single molecular property (also called a descriptor). This exercise is the simplest form of QSAR revolving around univariate linear regression models. Methods currently involve multiple regression analysis, so-called 2D-QSAR [1]. Here, the biological activity ( $X$ ) is said to be a linear combination of  $N$  different substituent properties ( $P_i$ ) that all receive their own weight constant to the equation ( $w_i$ ). The weight constants are determined by fitting the variations in the physico-chemical properties to the biological data (Equation 5-1).

Equation 5-1 
$$X = \sum_{i=1}^N w_i P_i$$

Early physico-chemical properties involved with 2D-QSAR methods are measures of molecular size and shape, hydrophobicity, hydrogen-bonding capabilities and electronic contributions and surface properties [2,3,4,5,6,7,8,9]. Since then, a multitude of topological descriptors have been introduced, such as the polar surface area [10], various connectivity, shape and size descriptors [11,12,13,14,15] and even atom-type descriptors [16,17].

A novel type of investigation is 3D-QSAR, which involves the correlation of the biological activity to actual three-dimensional data. The most well-known methods are Comparative Molecular Field Analysis (CoMFA [18,19,20]), Comparative Molecular Similarity Index Analysis (CoMSIA [21,22]) and Adaptation of Fields for Molecular Comparison (AFMoC) [23,24,25]. The CoMFA method constructs a three-dimensional overlay of the compounds in the dataset and determines steric, electrostatic and hydrogen-bonding fields for the molecules and attempts to correlate the field energy terms with the biological data. As such, the most important three dimensional features that all active compounds have in common are derived. The CoMSIA method is very similar to CoMFA, however, the fields used in the approach are based on similarity indices of the molecules for the regression with the

biological data. AFMoC is the direct opposite from CoMFA in that it is not the ligands that are used to construct the molecular fields, but the protein environment. The benefits of these 3D-QSAR methods are that they reflect the bioactive conformation of molecules and provide the very important spatial insight on protein-ligand interactions.

Next to these QSAR methods, other predictive analytic tools are used in drug design projects for the abstraction of structure-activity relationships. The commonly used tools are genetic algorithms [26], artificial neural networks such as self-organising maps [27,28], support-vector machines [29] and decision tree analysis [30]. Genetic algorithms start with a genetic representation of the training data as well as a fitness function for the evaluation of data-correlation. Following evolutionary optimisation techniques of inheritance, mutation, selection and crossover, the best description can be found between biological data and molecular descriptors. Neural networks use a non-linear modelling technique of interconnecting nodes (neurons) to learn the data from the training in an attempt for the resulted mapping to represent the underlying relationship with the biological data. Self-organising maps is a method of neural networks where the data is set up to a structure of interconnected neurons which can be used to provide a clustered visualisation of molecular descriptors important for biological activity. Support vector machines is a different type of machine learning that maps multi-dimensional data and defines hyperplanes to separate, cluster and classify the data towards a description of its relationship with biological data. Finally, decision tree analysis is a multivariate analysis similar to support vector machines as it attempts to uncover the most determinant descriptors by separating the data on a 1D scale following distribution into user-defined classifications based on the biological activity. Each of the QSAR methods possesses its own benefits. Either they are able to determine complex relationships between the data and descriptors (genetic algorithms, neural networks, self-organising maps, 3D-QSAR), provide easily interpretable quantitative results (2D-QSAR, support vector machines, decision tree analysis) or provide a visual aid to interpreting structure activity relationships (self-organising maps, 3D-QSAR, decision tree analysis).

### 5.1.1 QSAR Challenges

There are several difficulties in performing or interpreting QSAR analyses. Firstly, if the analysis is performed using a relatively homogenous class of compounds, it is difficult to expand the results to other compound classes. The success of the (2D-) QSAR model thus depends on a broad choice of compound variance as input or requires the knowledge of the limitations of the constructed model. Secondly, if the compounds possess multiple substructures that are important to the activity, modifications on either of the substructures may be dependent on the characteristics of the other substructure, hence will not relate the same effects on the biological activity. In other words, a substituent change on one side of the ligand may change the structural requirements for the other substructure. Thirdly, the presence of multiple binding sites in the target makes it very difficult to distinguish those ligand properties that are characteristic for one (or more) of those binding sites. Finally, protein as well as compound flexibility plays an important role in protein-ligand binding and

effects of slightly lengthier ligand substituents introduces a spatial component to a 2D-QSAR analysis that can be difficult to detect or interpret.

Even the highly predictive 3D-QSAR methods possess their limitations. The three-dimensional overlay of compounds suggests that all compounds bind to the receptor in exactly the same binding mode. This indicates that also for these methods, compound flexibility and induced fit are barely taken into account. To overcome such problems, the ligand overlap must be constructed with highest precision by taking into account multiple conformations, which can result in a very time-consuming task. In addition, the results of 3D-QSAR are best visualised using 3D colour-coded contour maps, but they are not as easily understood and exported as the multiple linear regression analysis from 2D QSAR.

### 5.1.2 QSAR Model Construction and Cross-validation

QSAR models are constructed using a training set to derive the weight constants and a validation set to derive the reliability of the QSAR model. One of the major concerns in constructing a QSAR model is the prevention of overfitting as well as overtraining. Overfitting is the occurrence where the training set uses too many variables to fit the data and does not recognise the underlying trends of the data. Instead, the data including the noise is learnt by heart, resulting in a poor reliability of the model. It is expressed in a very accurate model for the training set, yet a very inaccurate model for the validation set. Overfitting can also occur when descriptors are highly correlated with each other. In comparison, overtraining is also the occurrence where the data including the noise is learnt by heart, however, the cause lies in a poorly chosen training set that is often too restrictive and cannot be exported to the characteristics of the test set.

In general, the amount of compounds included in the analysis (**M**) needs to be much larger than the amount of properties (**N**), to prevent overfitting of the data (Equation 5–2). Similarly, the correlation between physico-chemical properties themselves needs to be low to prevent overfitting. This provides problems when small substituents are considered, because then certain physico-chemical properties tend to be correlated. An example is the mass of a small substituent that is highly correlated to both its volume and its surface area. If both properties are used in the analysis, there is a chance that their contribution will amplify each other causing them effectively to reduce the weights of other physico-chemical properties.

$$\text{Equation 5-2} \quad M \geq 2^{(N+1)}$$

In order to construct a meaningful QSAR, the results must be cross-validated, a process carried out to reduce overtraining. Constructing a cross-validated QSAR model requires the subdivision of the general compound dataset into several subsets. One subset, called the training set, is used to construct a QSAR model that determines the most important descriptors and their weights. Other subsets are the test set and the validation set. The test set can be derived from the training set (or both the training and test set can belong to a greater dataset), and it can be used to determine the internal predictive value of the model. The validation set is an independent dataset that is used to determine the general applicability of the QSAR model predictions. Sometimes the naming of the validation and

test sets are reversed indicating that the validation set is used to tune hidden parameters of the model, whereas the test set is used to predict novel compounds. The measure used to define the internal predictivity is the cross-validated  $R^2$ , called the  $q^2$ . The predictivity obtained by the validation set is called the  $R_{pred}^2$ . As a rule, the  $q^2$  is higher than the  $R_{pred}^2$ , since the data by which the training set is tested is correlated with the training data, and the data of the predicted set is generally uncorrelated. However, this is not necessarily true when the QSAR model is constructed using compounds that are highly similar to the compounds that are to be predicted.

An internal prediction of the constructed QSAR models can be obtained by cross-validation of the training data using various methods. One method is called K-fold cross-validation which partitions the training data (N samples) into K subsets (folds) of N/K samples. From these K subsets, one subset is retained as test set (sometimes also called the validation data), whereas the other K-1 subsets are used as training data. This method is then repeated K times until each fold has been used to test the constructed models. The results from the folds are then averaged to produce a single estimation.

Another method is called the leave-one-out cross-validation (LOO). This partitions the training data into N-1 values for training and 1 data point for validation (with N the total amount of data points). As for K-fold cross-validation, this method is repeated N times until each compound has been left out of the model construction. A variant of this method is the so-called leave-20%-out. Here the data is splice into 80% for the training set and 20% for the test set. This process is repeated several times to increase confidence. If the constructed leave-20%-out models continuously possess a high predictive value, the models are highly robust and can be used for the prediction of novel compounds. Thus, leave-20%-out can be used to produce a consensus scoring.

### 5.1.3 Receiver-Operating-Curves

One important method to test the usefulness of predictive models is to construct a contingency table and derive the characteristics for the receiver-operating-curve (ROC) for different model settings (Figure 5–1). The contingency table is constructed from a binary classification labelled as a positive or a negative prediction. Relating this to compound activity, the positive label would refer to a high potency, whereas the negative label would refer to a low potency. Thus, when a potent compound is predicted as positive, this is called a true positive (TP), however when it is predicted as negative, it is called a false negative (FN). The terms are similar for the inactive compounds, being true negative (TN) if correctly predicted and false positive (FP) if incorrectly predicted. Models can be optimised towards the best prediction of true positives or true negatives, but commonly it is desired to find a compromise between these two.

In order to find a good compromise between the correct prediction of positives and negatives, a ROC table can be defined using the sensitivity, which is the fraction of positives that is correctly predicted (Equation 5–3) and the specificity, which is the fraction of negatives that is correctly predicted (Equation 5–4). In the ROC table, a model constructed by random guess is a linear relationship from sensitivity of 1 to specificity of 1. A model that behaves better than the random guess must possess a higher sensitivity/specificity profile.



No real method exists for the determination of the correlation coefficient of a contingency table, however, the Matthews Correlation Coefficient (MCC [31]) can be determined to approximate the  $q^2$  and  $R_{\text{pred}}^2$  for QSAR (Equation 5–5). This correlation coefficient is -1 for total disagreement, 0 for a random prediction and 1 for perfect correlation. Hence, predictive models possess an MCC between 0 and 1.

$$\text{Equation 5-3} \quad \text{sensitivity} = \frac{TP}{FN + TP}$$

$$\text{Equation 5-4} \quad \text{specificity} = \frac{TN}{FP + TN}$$

$$\text{Equation 5-5} \quad \text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

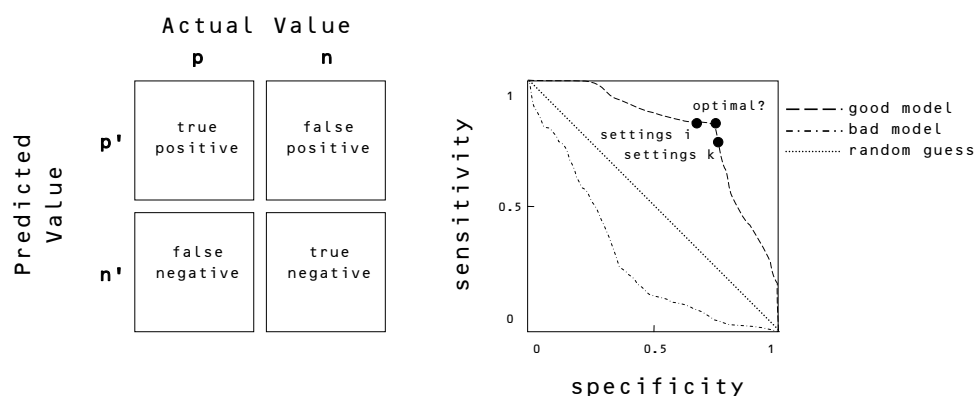


Figure 5–1 Contingency table (left) and its ROC table (right) in a sensitivity-specificity form

## 5.2 Problems Encountered By Performing QSAR on the Fadrazole Dataset

The development of a QSAR for our biological data is not straightforward. In the beginning, novel compounds had to be generated to guarantee a predictive QSAR model that does not overtrain the data. To illustrate the difficulties in constructing a QSAR for our compound database, we have analysed a small subset (Table 5–1), that comprise five series in total with two substituent positions. The QSAR models have been constructed with MOE-QuaSAR-Model [32]. The descriptors used for the construction of the models are three independent physico-chemical properties of the R5- and Rpara-substituents. These are volume, polar surface area (PSA) and the partitioning coefficient (logP) to represent the shape and the hydrophilic character of the substituent, respectively. Initially, the mass of the substituent was to be included, however, for these small substituents, the mass is highly correlated with the volume ( $R^2 > 0.8$ ). Because the number of compounds is in the order of 32, the maximum amount of descriptors that can be used is 4 (Equation 5–2). All QSAR models have been constructed with the same descriptor for both substituents.

Table 5–1 In vitro data for R5,-Rpara substituted benzyl-imidazoles. For each compound, the potencies of CYP11B1 as well as CYP11B2 are indicated.

Rpara	R5-H		R5-C <sub>3</sub> H <sub>6</sub> OH		R5-C <sub>2</sub> H <sub>4</sub> COOCH <sub>3</sub>		R5-Rb-Cyclohex		R5-Ph	
	H	11B1	11B2	11B1	11B2	11B1	11B2	11B1	11B2	11B1
CN	709.1	1799.0	180.5	285.7	326.0	192.9	36.3	159.4	4.8	10.7
OCH <sub>3</sub>	368.2	372.4	48.15	13.1	31.1	3.5	20.7	25.3	27.5	1.7
NO <sub>2</sub>	282.3	763.0	12.3	58.7	111.4	121.3			11.3	14.0
CH <sub>3</sub>	496.6	372.9	4.7	88.8	16.3	4.7			13.7	2.3
COOCH <sub>3</sub>	602.0	1932.5	48.1	23.6	79.1	15.1				
Br	218.3	> 300	5.8	24.4	111.3	21.5	2.0	129.5		
	211.6	478.5	18.4	6.21	57.4	4.0	30.5	60.4	7.3	5.1

### 5.2.1 Initial Trends

By examining the potency and selectivity of these series, several initial trends can be derived for a SAR. Unsubstituted R5-compounds possess the lowest potencies (circles, Figure 5–2a), whereas R5-phenyl substituted compounds possess a very high potency (plusses, Figure 5–2a). The exact contribution of the other R5-substituents groups is less clear, not to mention the difference between these the R5-substituents. Rpara-unsubstituted compounds also possess a lower potency than the substituted compounds, but again, a real trend is more difficult to derive.

The selectivity of the compounds is investigated by using a logarithmic selectivity plot (Figure 5–2b). Here, a compound is selective for CYP11B2 if the value of the compound is located above the dotted line and selective for CYP11B1 if it is below the dotted line. Furthermore, a value of 1 indicates a high selectivity value of 10. From this figure it can be concluded that, in general, three Rpara-substituents seem to confer selectivity for CYP11B2 (cyano, nitro, bromide), and three Rpara-substituents selectivity for CYP11B1 (unsubstituted, methoxy, methylacetate). There are several exceptions to these observations, such as the R5-propylol-Rpara-nitro compound (**triangle**) and the R5-unsubstituted-Rpara-bromide compound (**circle**). Therefore, it will be difficult for a 2D-QSAR to discriminate the effects of the substituent properties on the potency.

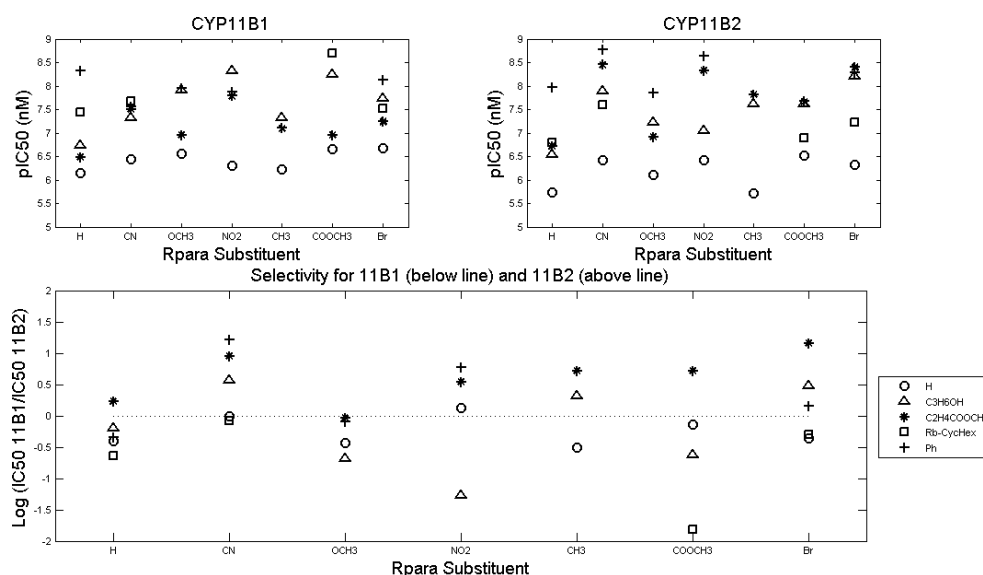


Figure 5–2 Potency and selectivity series for R5 substituted benzylimidazoles. a) compound potencies expressed in pIC50 values (nM) and b) compound selectivity expressed by a logarithmic scale. Compounds below the dotted line are CYP11B1 selective, compounds above the dotted line are CYP11B2 selective. Since the vertical scale is logarithmic, a value of 1 indicates a selectivity factor of 10.

### 5.2.2 QSAR Results: Why did we opt for Decision Tree Analysis?

All constructed QSAR models for the prediction of CYP11B1 possess a correlation coefficient ( $q^2$ ) below 0.6, indicating that their predictive value is very low (Table 5–2). Three of the QSAR models for CYP11B2 data are accurate ( $q^2 > 0.6$ ). Correlation of the volume descriptor (vol) of both the R5 and Rpara substituents leads to an acceptable prediction for the CYP11B2 pIC50 ( $q^2$  of 0.69). However, expanding this model with either the polar surface area (TPSA) or the partitioning coefficient (logP) only marginally raises the  $q^2$  to 0.72 and 0.70, respectively. These results suggest that only the volume of the R5 and Rpara substituents determine their contribution to the potency of the compound, which is not realistic. Apparently, different descriptors should be defined to increase the predictive value of the QSAR model.

Comparing the predicted pIC50 values to the experimental pIC50 values for CYP11B2 (Table 5–1, Figure 5–3), it is clear to see that for each substituent, the inhibitor potency is affected by R5 substitution in a different manner. Therefore, no clear-cut pattern can be observed despite the fact that R5 substitution, next to Rpara substitution clearly is of importance in determining ligand potencies and selectivity. The differences in potency are large for the R5-methyl acetate and R5-propylol series, yet small for the unsubstituted series. This behaviour can not be fit with the current QSAR model and it is anticipated that additional one-dimensional substituent descriptors will not improve the model.

Possible explanations for the seemingly uncorrelated structure-activity relationship are the flexibility of the compound scaffold and the possibility that compounds possess several different binding modes. To incorporate these aspects into a QSAR model, it would be advantageous to add three-dimensional descriptors to capture these influences. However, the construction of a reliable 3D-QSAR method is a difficult and time-consuming exercise and results can be difficult to interpret. Since we intend to perform a SAR analysis to guide compound synthesis, it is vital that the analysis can provide fast results that are easily understood. As such, we have employed the method of decision tree analysis for the prediction of the potency of new potential inhibitors.

Table 5–2 QSAR formulas

Property 1 and 2 (R5 and Rp)	Property 3 and 4 (R5 and Rp)	$q^2_{\text{CYP11B1}}$	$q^2_{\text{CYP11B2}}$	Formula CYP11B2 (pIC50)
TPSA	-	0.03	0.12	$6.93 + 0.023 P1 + 0.009 P2$
volume	-	0.45	0.69	$5.62 + 0.022 P1 + 0.013 P2$
log P(o/w)	-	0.32	0.18	$6.92 + 0.466 P1 + 0.164 P2$
TPSA	volume	0.53	0.72	$5.57 - 0.011 P1 + 0.006 P2 + 0.239 P3 + 0.012 P4$
TPSA	log P(o/w)	0.52	0.56	$5.89 + 0.045 P1 + 0.015 P2 + 0.781 P3 + 0.397 P4$
volume	log P(o/w)	0.53	0.70	$5.62 + 0.022 P1 + 0.013 P2 + 0.026 P3 + 0.181 P4$

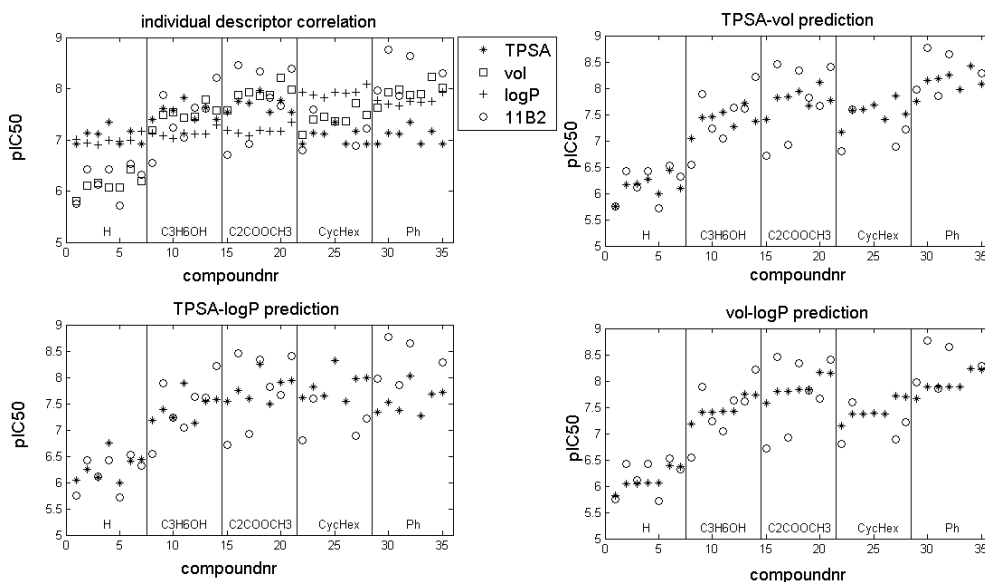


Figure 5–3 Prediction of CYP11B2 pIC50 values by different QSAR models. Indicated in each subplot are the CYP11B2 potency data (circles) with which the models are compared. The horizontal axis is subdivided into sections for each R5 substituent. The compound within these sections represents the potency of the following Rpara substituents: 'H', 'CN', 'OCH<sub>3</sub>', 'NO<sub>2</sub>', 'CH<sub>3</sub>', 'COOCH<sub>3</sub>', 'Br'. The upper left subplot displays the results for three different QSAR models that are based on one of the three descriptors for both substituents. The other figures display the results for QSAR models where two descriptors have been combined for both substituents.

### 5.3 Decision Tree Analysis

We have employed decision tree analysis using simple molecular descriptors to build an *in silico* prediction tool for novel CYP11B2 inhibitors. Decision tree analysis is a multivariate analysis by which it is possible to incorporate molecular descriptors to determine the substituent properties that are most determining for compound potency. The tree is a graph that consists of several nodes (Figure 5–4). The starting node is called the root and it contains all the compounds in the dataset. The tree is built by evaluating the descriptors in the dataset and making a decision to branch the root into subsequent nodes. Each of these nodes contains a unique compound subset of its parent node, and each compound of the parent node is present in one of the nodes. Subsequently, each node is branched further until a certain termination criterion has been met. The branched nodes are called decision nodes, whereas the unbranched nodes are called leaf nodes.

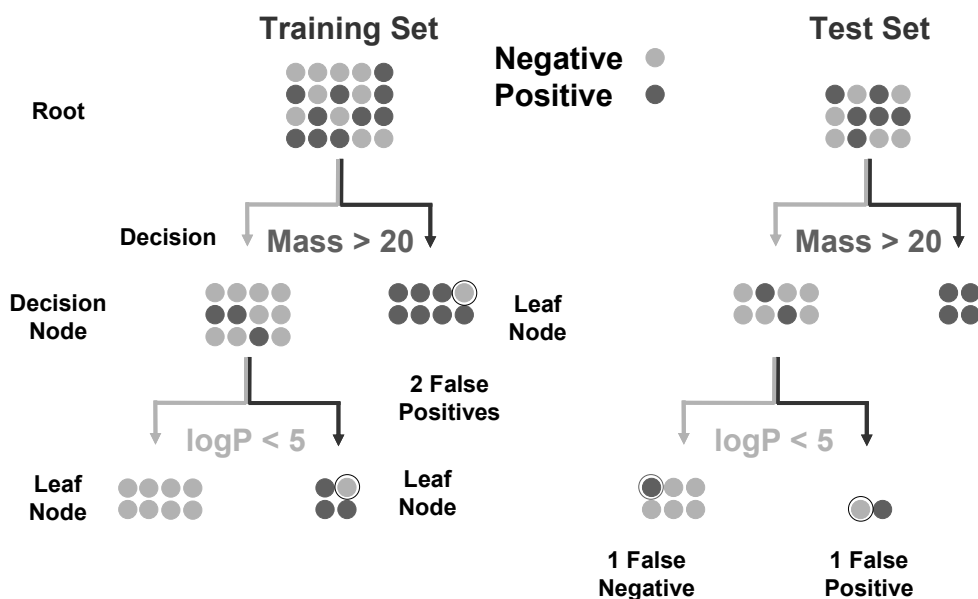


Figure 5–4 Decision tree analysis method for a training set and a test set. If the leaf nodes contain compounds that do not belong to the leaf classification, this compound is a false positive or false negative to the constructed tree.

For the prediction of potent inhibitors, the decision tree analysis requires the computation of molecular descriptors for the compounds (such as volume, logP and TPSA) and a classification of the compounds (such as active or inactive). The classification can consist of multiple labels, however, the trees are easier to validate using receiver-operating-curves (ROC tables) if only two labels are used (binary tree). The decision to branch a node is based on the value of a molecular descriptor and the classification of the compounds. The goal of the decision is to select the molecular descriptor that is most useful for classifying the

compounds in the dataset. If for example a large amount of compounds are active when a substituent volume is greater than  $3.0 \text{ \AA}^3$ , then this decision can be evaluated and two nodes will be created, one with compounds that possess a substituent volume greater than  $3.0 \text{ \AA}^3$ , and a second node that contains the other compounds. Dependent on how well the molecular descriptors correlate with the (in)activity of the compounds in the dataset, a branching can create either two new decision nodes, one decision and one leaf node, or two leaf nodes.

The branching of a tree stops when a termination criterion is met. Although user-defined criteria can be used, the general criteria are: (1) a maximum amount of tree depth, (2) a minimum amount of compounds in a node, or (3) a node for which no molecular descriptor is able to reliably separate a subset of either active or inactive compounds. These criteria are based on the prevention of overfitting as well as overtraining the data. Because every next branching defines a decision on top of the previous decision, a large amount of subsequent branching can overfit the data. Similarly, if each leaf contains only very few compounds, it is possible that the branching has been performed too often.

Branching can be optimised towards either splicing a node favouring the classification of positives or negatives by introducing a splicing cost. If the cost favours positive prediction, then the decision tree algorithm will introduce a high penalty towards the misclassification of negatives. Using this method of optimisation of splicing decision nodes decreases the amount of mistakes made by the training data for the prediction of positives.

### 5.3.1 Fadrazole analogue main set

The decision tree analysis described in this thesis has been continuously optimised using newly synthesised compounds. The dataset used for the final decision tree analysis described here, is composed of 128 compounds (Appendix B). The goal of the decision tree analysis is twofold, namely the prediction of CYP11B2 potent compounds, and the prediction of CYP11B2 selective compounds. The prediction of CYP11B2 potency can be performed by introducing a potency threshold for the compounds in the dataset, defining a potency label based on that threshold and training of decision trees using these settings.

The prediction of CYP11B2 selectivity is less evident. Each compound has been tested for CYP11B1 and CYP11B2 by in vitro screening, and the selectivity value can be obtained by the division of the CYP11B2 potency by the CYP11B1 potency. However, since both potency values are subject to error, the correct interpretation of a selectivity of for example 2, is not straightforward and the definition of a selectivity threshold for the decision tree analysis is not transparent. In addition, the construction of a selectivity-based decision tree results only in a description for the difference between CYP11B1 and CYP11B2, whereas the specific CYP11B2 characteristics are not determined. As a result, compounds that are predicted as selective CYP11B2 inhibitors may not be potent CYP11B2 inhibitors. Finally, there are few CYP11B2 selective compounds in the dataset and construction of a decision tree based on very few positive labels is difficult and will result in data-overtraining.

To overcome these problems, an alternative method can be used to approximate a prediction for the CYP11B2 selectivity. Decision trees are constructed for both CYP11B1 and CYP11B2 potency. The detail of the CYP11B1 analysis emphasises on the correct

prediction of all potent compounds, making certain that as few as possible CYP11B1 potent compounds are predicted incorrectly. This implies that the decision trees are pruned towards the optimisation of the sensitivity such that the analysis contains the least amount of false negatives. Gathering the results of these decision trees allows the prediction of CYP11B1 potent compounds, hence those compounds that are less likely to be selective for CYP11B2 and thus not interesting for synthesis.

The detail of the CYP11B2 analysis emphasises on the correct prediction of the non-potent compounds, ensuring that those compounds predicted as potent by the model will indeed be potent. Therefore, the models must contain little false positives, hence it implies an optimisation of the model specificity. These models result in the prediction of compounds that are certain to be potent, although a collection of potent compounds (false negatives) may be missed. Combining the results of both CYP11B1 and CYP11B2 models, those compounds predicted as non-potent on CYP11B1 and potent on CYP11B2 have the highest chance of being CYP11B2 selective compounds.

The program C5.0 [33] was used for the generation of the decision trees. Since every compound in the dataset is used in the training and test sets, an activity label must be defined for every compound. The potency threshold value that determines whether a compound is potent or non-potent has been chosen as 100 nM, e.g. a compound is potent with a value of less than 100 nM.

Because the compounds are structurally very similar, the leave-one-out cross-validation method will have problems validating the decision tree models. Therefore, the cross-validation method leave-20%-out has been used. The dataset of 128 compounds has been randomly divided into subsets comprising a training set (80%) and a test set (20%). This process of subdividing the data set into a training set and test set has been performed 5 times in total such that all compounds have been used in both the training set and the test set. Using these subsets, various decision tree models were constructed using different model settings based on the minimum leaf size and a cost setting. The minimum leaf size was varied from 1 to 5 at which point the model accuracy decreased. The cost settings used vary from a factor of  $\frac{1}{4}$  to 4, with increased detail around a cost value of 1. The values of  $\frac{1}{4}$  and 4 are the minimum values at which the decision tree starts to become overtrained. The additional detail around 1 has been investigated to increase confidence in the continuity of model generation. A cost value of 1 indicates that both potent and non-potent compounds are considered equally important. A cost value higher than 1 considers potent compounds more important than non-potent compounds (by that factor), and attempts to increase sensitivity by penalising the occurrence of false negatives. Similarly, a low cost factor values non-potent compounds as more important and attempts to increase specificity by penalising the occurrence of false positives. Continuity was deemed acceptable when the ROC table is smooth. The maximum decision tree depth has been set to 5 for all models to prevent overtraining as it disallows too many branches. The subset results have been subjected to consensus scoring for the prediction of novel CYP11B2 (selective) inhibitors.

### 5.3.2 Substituent descriptors

The selection of the correct molecular descriptors is important for the construction of the models. Many different physico-chemical properties can be computed that might relate to the activity of a substructure. As such, it is difficult to derive which descriptors can be used to describe the underlying trends of the data. In a way, less can sometimes result in more. The complexity of a descriptor does not always indicate that it is the correct one to use. Using few and interpretable molecular descriptors not only results in a more clear model, which can be interpreted, but it is also more likely to be of statistical significance [34,35]. Indeed, it has been proposed that the applicability of descriptors (whether simple or complex) to identify novel inhibitors heavily depends on the combination of the descriptor set [36].

Because the substituents on our compounds are small, we have opted that it is better to construct a decision tree analysis based on simple descriptors for hydrophilicity, shape and size. As a result, similar molecular descriptors have been computed for this dataset as discussed in the QSAR attempt above. The original parameters are: the substituent partition coefficient ( $\log P$ ), mass and polar surface area (TPSA). Additionally, the volume descriptor has been refined by using a set of descriptors called the Sterimol parameters [37] as defined in (Figure 5–5, B1, B2, B3, B4, B5 and L), which are more able to describe the substituent shape. These descriptors are also uncorrelated with the mass descriptor. Using these descriptors, three types of decision trees have been generated, the first one based only on the original descriptors, the second one only on the Sterimol parameters, and the last one based on both the original and the Sterimol parameters.

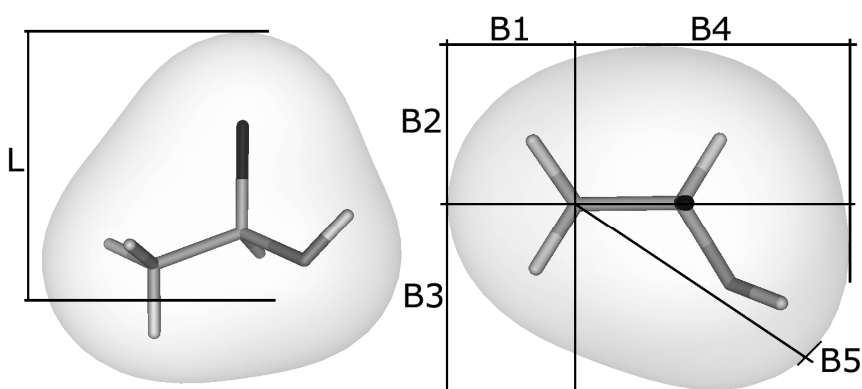


Figure 5–5 Sterimol Parameters for a substituent. L is the maximum extension of the substituent, B1 the minimum width and B5 the maximum width.



### 5.3.3 Results and Discussion

After decision tree generation, the models were sorted by performance based on their receiver-operating-curve (ROC) tables. All models continuously outperform the random guess sensitivity-specificity profile. An exemplary result of the analysis can be seen in (Figure 5–6). Every data point represents the average result of the 5 submodels, with the error bars indicating the standard deviation for both the sensitivity and specificity. Every data point also possesses a different value for the potency cost penalty. Because a change in model settings results in the generation of a new optimised decision tree, the points in the ROC table cannot be interconnected. The MCC calculation results for training set and test set as well as the sensitivity and specificity calculation for the test set can be found in Appendix D.

A high cost value results in a higher penalty for the misprediction of positive (potent) compounds and the sensitivity of these models will be highest. However, if the cost is too high, the penalty will be too severe such that the constructed decision tree model does not possess the ability to classify the training set into leaves that possess only potent compounds. As a result, the sensitivity of such models decreases again. An example of such a model is the CYP11B1 decision tree constructed with the original descriptors with a cost of 4 (Figure 5–6). The sensitivity of this model is lower than for a cost of 3, hence the reliability of this model can be considered as low. This can be shown by calculating the Matthews correlation coefficient (MCC) for these CYP11B1 decision tree models; for the cost 4 model the MCC are 0.45 ( $q^2$ , training set) and 0.30 ( $R^2$ , test set), whereas the other cost models possess MCC values that range from 0.69 to 0.76 ( $q^2$ , training set) and from 0.46 to 0.51 ( $R^2$ , test set).

Comparing the models that have been generated with the three different descriptor sets (Figure 5–6), it is observed that those constructed with the Sterimol descriptors perform best for all training sets ( $MCC_{\text{train}} = 0.80\text{--}0.95$  for both CYP11B1 and CYP11B2 potency). However, when investigating the results of the test set, the Sterimol models are accurate in the classification of potent compounds ( $MCC_{\text{test}} = 0.35\text{--}0.55$ ). Their sensitivity for low cost settings is comparable to that of high cost settings, but instead the sensitivity should have increased because a high cost value indicates that the amount of false negatives is decreased. Because the models are unable to discriminate the false negatives, it suggests that these models are overtrained.

The models constructed with only the original descriptors or a combination of the original and Sterimol descriptors are found to possess comparable MCC values as well as a comparable sensitivity/specificity profile. Only the model settings with a very low cost or a very high cost can be labelled as unreliable (italic Appendix D).

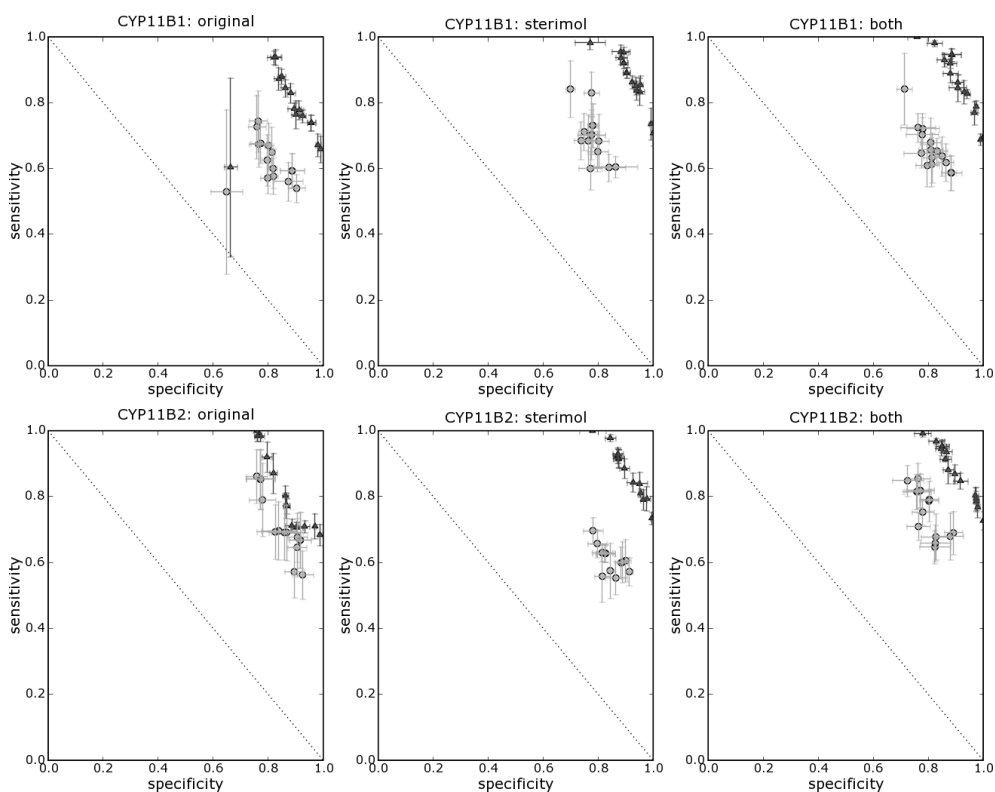


Figure 5–6 ROC table for the training set (triangles) and the test set (circles) for a minimum leaf size of 4 and a maximum decision tree depth of 5. The various data points indicate different cost settings. Cost settings for  $\frac{1}{4}$  and 4 are indicated. The original descriptors are the mass, TPSA, logP.

### 5.3.3.1 Prediction of CYP11B1 and CYP11B2 potency for novel compounds

By optimising the values for the MCC (Matthews correlation coefficient) and the sensitivity/specificity profile of the test set, two models were chosen for each leaf size setting (greyed, Appendix D). Using those decision tree settings, models were extracted for the original descriptors, the Sterimol descriptors and the combined descriptors. The models encompassing the different descriptor sets were tested independently on a validation set of 12 novel CYP11B inhibitors, followed by the calculation of the consensus score of the prediction.

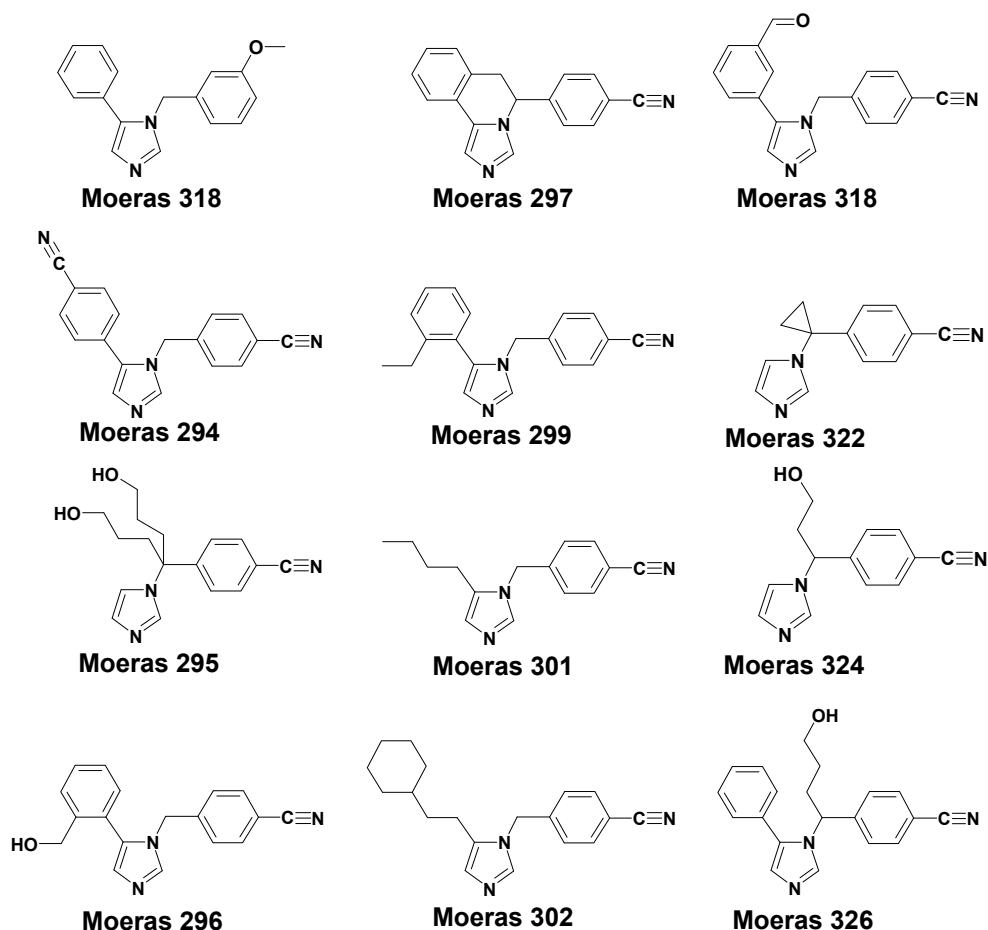


Figure 5-7 Compounds used for potency prediction using the decision tree models.

Subsequently, the results for the different descriptor sets were compared. The results for the compound prediction by the different descriptor sets are listed in Table 5-3. Each descriptor set displays the fraction of all decision tree models that predict the compound as potent, i.e. a value of 1.00 indicates that all models for the descriptor set have predicted the compound as potent. In addition, a consensus label has been defined that combines the prediction performed by the different descriptor sets. The descriptors that are most important for the potency prediction of CYP11B1 and CYP11B2 are listed in Table 5-4. For the prediction of potent CYP11B1 substituents, the R5-substituent are heavier than approximately a chlorine atom, possess a length between 4.4 Å and 7.4 Å and possess a width larger than 2.45 Å (comparable to a phenyl group). The Rpara position can possess a maximum polar surface area comparably to that of a hydroxyl group and a substituent width of 1.5 Å. In addition, the Rortho position is preferably substituted.

The CYP11B2 substituents possess an R5-substituent with a mass more than the hydroxypropyl group (60 amu) and less than a phenyl group (80 amu). In addition the maximum width and length of the substituents must be smaller than that of a phenyl group (7.5 Å). In addition, Rb-substituted compounds preferably contain a polar atom and a size larger than a methyl group. Finally, Rortho-substituents should to be larger than a propyl chain and should possess a logP more than hydrogen.

Table 5–3 Predicted and real potency labels for the novel inhibitors using a potency cut-off value of 0.5. O: original descriptor set, S: Sterimol descriptor set, O+S: both descriptor sets. Values indicate the fraction of all models that predict the compound as potent (P). Consensus (Cons) is the combined prediction of all three descriptor sets as potent (P) and non-potent (N). The 'True' column indicates the real label defined by a cut-off value of 100nM (compare with the IC50). Compounds indicated with a star are correctly predicted by the models constructed with the original descriptor set, but not by the consensus.

	Original	Sterimol	Both	Consensus	True	IC50 (nM)	Correct?
<b>CYP11B1</b>							
183	1.00	0.72	0.76	Potent	Potent	32.4	Yes
294	0.40	0.08	0.22	Non-potent	Potent	15.1	No
295	0.08	0.32	0.56	Non-potent	Non-potent	> 1000	Yes
296	0.44	0.22	0.46	Non-potent	Non-potent	107.9	Yes
297	0.94	0.68	0.52	Potent	Potent	39.2	Yes
299	0.92	0.12	0.44	Non-potent	Potent	1.6	No *
301	0.94	0.82	0.72	Potent	Potent	0.8	Yes
302	0.92	0.62	0.60	Potent	Potent	2.9	Yes
318	0.44	0.92	0.76	Potent	Non-potent	125.4	No *
322	0.00	0.12	0.22	Non-potent	Potent	10.2	No
324	1.00	0.96	0.96	Potent	Potent	36.6	Yes
326	0.96	0.92	0.72	Potent	Potent	37.0	Yes
<b>CYP11B2</b>							
183	1.00	0.86	0.90	Potent	Potent	18.1	Yes
294	1.00	0.02	0.20	Non-potent	Potent	30.6	No *
295	0.38	0.28	0.26	Non-potent	Non-potent	>1000	Yes
296	1.00	0.90	0.62	Potent	Potent	20.4	Yes
297	1.00	0.94	0.86	Potent	Potent	41.2	Yes
299	1.00	0.90	0.64	Potent	Potent	0.9	Yes
301	0.70	0.96	0.84	Potent	Potent	1.1	Yes
302	1.00	0.96	0.66	Potent	Potent	1.2	Yes
318	0.92	0.96	0.84	Potent	Potent	58.3	Yes
322	0.00	0.02	0.06	Non-potent	Non-potent	101.4	Yes
324	0.38	0.02	0.26	Non-potent	Non-potent	292.6	Yes
326	0.08	0.12	0.22	Non-potent	Non-potent	403.4	Yes

The models constructed with the original descriptors were able to predict the CYP11B1 potency for 10 compounds indicating a high hit-rate (Table 5–3). However, using the Sterimol and the combination descriptor sets the prediction was only correct for 8 and 7 compounds, respectively. Moeras 294 and 322, which were incorrectly predicted by the original descriptors were also incorrectly predicted by the other two descriptor sets. Moeras 299 and 318 were only incorrectly predicted after inclusion of the Sterimol parameters and Moeras 295 was only incorrectly predicted by the combination of descriptors.

Moeras 294, 299 and 318 possess an R5-phenyl ring substituted at the para, ortho and meta positions, respectively. Although the results are not easy to interpret, it appears that the high R5 Sterimol shape parameter L of Moeras 294 (9.43) and Moeras 299 (8.76) results in a non-potent scoring, whereas the high R5 shape parameter B3 of Moeras 318 (3.64) results in a potent scoring (Table 5–4). Moeras 295 and 322 are both doubly substituted on the Rb position. Singly substituted Rb compounds generally result in a potent inhibitor, however, since the decision tree models are insufficiently trained with inhibitors that possess two substituents, the prediction fails for these compounds.

In general, the models on which the predictions are based possess a sensitivity of 0.75, hence, several false negatives were expected. Additionally, the sensitivity of the models based on the Sterimol parameters is even lower, indicating an uncertainty to use them for the prediction of novel compounds. Indeed, models generated with the Sterimol parameters may be slightly overtrained.

For the CYP11B2 potency, the models constructed with the original descriptor set were able to predict all compounds correctly, whereas the models constructed with the other descriptor sets contain one incorrect prediction. The compound that is incorrectly predicted is Moeras 294, which was also incorrectly predicted for the CYP11B1 models. The reason for its misprediction is also the non-potent scoring for a high value for the R5 shape parameter L (9.43) (Table 5–4). Even though the sensitivity of the CYP11B2 models is low (0.60) the results only contain Moeras 294 as false negative and only for the models constructed with Sterimol parameters. With a specificity of 0.90, there are also no false positives observed, however, the validation set is too small to draw any firm conclusions.

The validation set contains several compounds that are CYP11B2 selective (Moeras 183, 296, 299, 302 and 318). However, for three compounds their potencies fall within the chosen threshold for the potency label (100nM). The two CYP11B2 selective compounds that can be predicted by using a non-potent prediction for CYP11B1 and a potent prediction for CYP11B2 are Moeras 296 and 318. These compounds are indeed predicted correctly on both accounts using the original descriptors, hence, these compound are correctly predicted as selective. Unfortunately, this does not provide any indication towards using the decision trees for the prediction of CYP11B2 compound selectivity.

Re-evaluating the training and test sets, the amount of CYP11B2 selective compounds using a 100nM threshold is only 7 (Figure 5–8). For the utilisation of the decision tree analysis for CYP11B2 selectivity prediction, the selectivity threshold must be adjusted such that this amount is optimised for future research. The optimal value seems to be 30nM for which 15 compounds possess the desired CYP11B2 selectivity (Figure 5–8). However, this threshold is very tight as 5 of these compounds possess an *in vitro* CYP11B1 potency close to the threshold with a standard deviation passing the threshold. As a result, optimisation of the current decision tree analysis can hardly be performed by using a different potency threshold. Until new highly selective CYP11B2 selective inhibitors are synthesised, decision tree analysis must be performed in its current form.

Table 5–4 Reoccurring decisions for the CYP11B1 and CYP11B2 potency resulting from the decision tree analysis for the prediction of the potency label

Molecule	Original descriptor set	Sterimol descriptor set
hCYP11B1	R5_logP > 1.10	R5_B1 > 2.45
	R5_weight > 35.55	Rpara_B3 > 1.50
	R4_weight <= 15.04	R5_L > 4.38
	Rpara_PSA <= 26.30	R5_L < 7.48
	Rpara_weight > 19.00	R5_B3 > 3.55
	Rortho_weight > 1.008	
hCYP11B2	R5_weight > 57.12	R5_B1 > 2.10
	R5_weight <= 79.9	R5_L <= 7.66
	Rortho_logP > 0.13	Rortho_B5 > 2.49
	R5_weight > 1.008	R5_B4 <= 3.45
	Rb_PSA > 0	Rb_B1 > 1.2
	R4_weight <= 15.04	Rpara_L > 5.18
	Rpara_PSA <= 26.3	

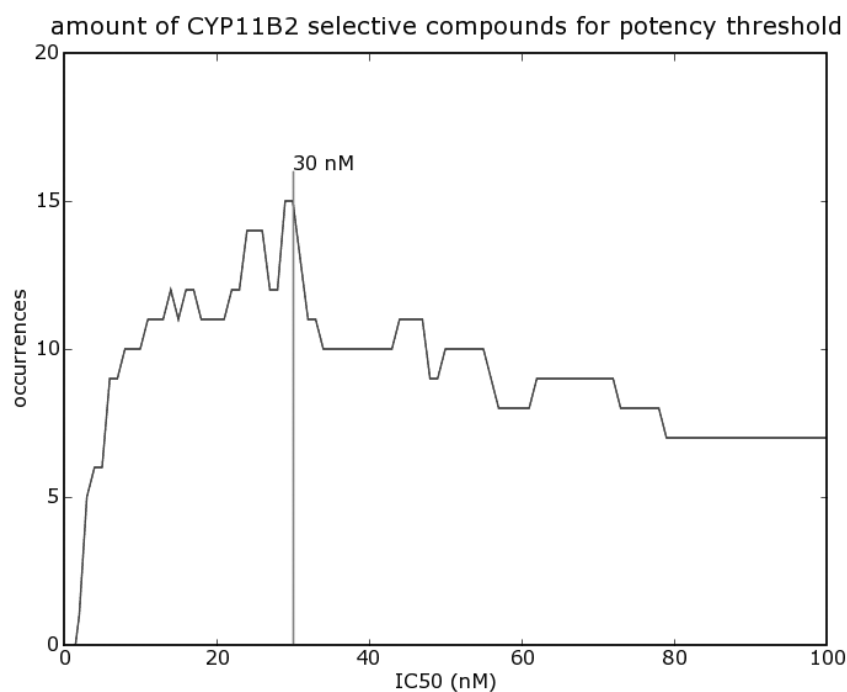


Figure 5–8 Distribution of CYP11B2 selective compounds for each potency threshold under 100nM. The optimal value is 30nM

## 5.4 Conclusions

The construction of a QSAR model for the various fadrazole analogues is a difficult task because multiple subgroups influence on the inhibitor potency of other substituents in a different manner. Therefore, we have employed decision tree analysis as a tool to predict CYP11B2 potency and selectivity. The results of the decision tree analyses showed that the models constructed with a descriptor set encompassing substituent mass, polar surface area and partitioning coefficient, yield the most reliable predictions for both the test set and the prediction set. For the 12 investigated compounds the models correctly predict 10 CYP11B1 potencies and 12 CYP11B2 potencies, defining good reliability even though the prediction set is small.

On the basis of decision tree analysis, we have been able to determine the descriptors that are important for the potency of CYP11B1 and CYP11B2. Although these provide an indicative measure for the inhibitor properties, they cannot be interpreted as a general rule set, since because of its nature, decision tree analysis evaluates sequential rather than parallel decisions. In general, the Sterimol descriptor set does not perform better than the decision trees constructed with the original, simpler descriptors. This is caused by overtraining of the models constructed with the Sterimol parameters, since they only contribute to the substituent shape. In addition, the combination of the original and Sterimol descriptor sets have not resulted in an increased accuracy of potency prediction. This indicates that perhaps the Sterimol parameters are not suitable for the expansion of the shape contributions.

The decision tree analysis uses an artificial potency threshold for the prediction of inhibitor potency. Using a combination of CYP11B1 and CYP11B2 potency predictions, a qualitative prediction can be obtained for the CYP11B2 selectivity. However, the threshold for determining CYP11B2 selectivity requires the training of compounds that possess a potency value above and below the threshold for CYP11B1 and CYP11B2, respectively. Since the decision tree models have not been trained with sufficient compounds that fit this requirement, no conclusions can be drawn for the performance on predicting CYP11B2 selectivity. Future research on the applicability of decision tree analysis in CYP11B2 selective drug design should involve the identification of more CYP11B2 selective compounds which can be used to either test or optimise the current decision tree models. In addition, potency thresholds and descriptor sets have to be defined that are better able to distinguish the selectivity of the compounds.

## Literature

- 1 C. Hansch, P.P. Maloney, T. Fujita, R.M. Muir, "Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients", *Nature*, 1962, 194, 4824,178-180
- 2 L.P. Hammett, "Some relations between reaction rates and equilibrium constants", *Chem. Revs.*, 1935, 17, 1, 125-136
- 3 L.P. Hammett, "The effect of structure upon the reactions of organic compounds. benzene derivatives", *J. Am. Chem. Soc.*, 1937, 58, 1, 96-103
- 4 H. Wiener, "Structural determination of paraffin boiling points", *J. Am. Chem. Soc.*, 1947, 69, 1, 17-20
- 5 R.W. Taft, "Polar and steric substituents for aliphatic and o-benzoate groups from rates on esterification and hydrolysis of esters", *J. Am. Chem. Soc.*, 1952, 74, 12, 3120-3128
- 6 H.H. Jaffe, "A reexamination of the hammett equation", *Chem. Rev.*, 1953, 53, 2, 191-261
- 7 C. Hansch, R.M. Muir, T. Fujita, P.P. Maloney, C.F. Geiger, M.J. Streich, "The correlation of biological activity of plant growth regulators and chloromycetin derivatives with hammett constants and partition coefficients", *J. Am. Chem. Soc.*, 1963, 85, 18, 2817-2824
- 8 C. Hansch, T. Fujita, " $\rho$ - $\sigma$ - $\pi$  analysis. A method for the correlation of biological activity and chemical structure", *J. Am. Chem. Soc.*, 1964, 86, 8,1616-1626
- 9 P.N. Craig, "Interdependence between physical parameters and selection of substituent groups for correlation studies", *J. Med. Chem.*, 1971, 14, 8, 680-684
- 10 P. Ertl, B. Rohde, P. Selzer, "Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties", *J. Med. Chem.*, 2000, 43, 20, 3714-3717
- 11 I. Gutman, N. Trinajstic, "Graph theory and molecular orbitals. total  $\phi$ -electron energy of alternant hydrocarbons", *Chem. Phys. Lett.*, 1972, 17, 4, 535-538
- 12 A.T. Balaban, "Highly discriminating distance-based topological index", *Chem. Phys. Lett.*, 1982, 89, 5, 399-404
- 13 L.H. Hall, L.B. Kier, "The molecular connectivity indices and kappa shape indices in structure-property modeling", *Reviews in Computational Chemistry*, K.B. Lipkowitz, D.B. Boyd (ed), ISBN 1-56081-515-9, 1991, 2, 367-422
- 14 A. Milicevic, S. Nikolic, N. Trinajstic, "On reformulated Zagreb indices", *Molecular Diversity*, 2004, 8, 4, 393-399
- 15 M. Wagener, J.P.M. Lommerse, "The quest for bioisosteric replacements", *J. Chem. Inf. Model.*, 2006, 46, 2, 677-685
- 16 A. Ghose, G. Crippen, "Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships I. partition coefficients as a measure of hydrophobicity", *J. Comp. Chem.* 1986, 7, 4, 565-577
- 17 A. Ghose, G.M. Crippen, "Use of physicochemical parameters in distance geometry and related three-dimensional quantitative structure-activity relationships: a demonstration using escherichia coli dihydrofolate reductase inhibitors", *J. Med. Chem.*, 1985, 28, 3, 333-346
- 18 R.D. Cramer, D.E. Patterson III, J.D. Bunce, "Comparative molecular field analysis (CoMFA). 1. effect of shape on binding of steroids to carrier proteins", *J. Am. Chem. Soc.*, 1988, 110, 18, 5959-5967
- 19 S.A. DePriest, D. Mayer, C.B. Naylor, G.R. Marshall, "3D-QSAR of angiotensin-converting enzyme and thermolysin inhibitors: a comparison of CoMFA models based on deduced and experimentally determined active site geometries", *J. Am. Chem. Soc.*, 1993, 115, 13, 5372-5384
- 20 G. Klebe, U. Abraham, "On the prediction of binding properties of drug molecules by comparative molecular field analysis", *J. Med. Chem.*, 1993, 36, 1, 70-80
- 21 G. Klebe, U. Abraham, T. Mietzner, "Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity", *J. Med. Chem.*, 1994, 37, 24, 4130-4146
- 22 M. Bohm, J. St Rzebecher, G. Klebe, "Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin and factor Xa", *J. Med. Chem.*, 1999, 42, 3, 458-477



- 23 H. Gohlke, G. Klebe, "Drugscore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein", *J. Med. Chem.*, 2002, 45, 19, 4153-4170
- 24 K. Silber, P. Heidler, T. Kurz, G. Klebe, "AFMoC enhances predictivity of 3D QSAR: a case study with DOXP-reductoisomerase", *J. Med. Chem.*, 2005, 48, 9, 3547-3563
- 25 B. Breu, K. Silber, H. Gohlke, "Consensus adaptation of fields for molecular comparison (AFMoC) models incorporate ligand and receptor conformational variability into tailor-made scoring functions", *J. Chem. Inf. Model.*, 2007, 47, 6, 2383-2400
- 26 H.J. Bremermann, "Optimization through evolution and recombination", *Self-organising systems*, 1962, M.C. Yovits, G.T. Jacobi, G.D. Goldstein (editors) Spartan Books, Washington DC. 1962, 93-106
- 27 T. Kohonen, "Self-organizing formation of topologically correct feature maps", *Biological Cybernetics*, 1982, 43, 1, 59-69
- 28 T. Kohonen, "Analysis of a simple self-organizing process", *Biological Cybernetics*, 1982, 44, 2, 135-140
- 29 D. Meyer, F. Leisch, K. Hornik, "The support vector machine under test", *Neurocomputing*, 2003, 55, 1-2, 169-186
- 30 R.M. Dawes, B. Corrigan, "Linear models in decision making", *Psychological Bulletin*, 1974, 81, 2, 93-106
- 31 B.W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme", *Biochim. Biophys. Acta.: Protein Structure*, 1975, 405, 2, 442-451
- 32 MOE (The Molecular Operating Environment) Version 2007.08, Chemical Computing Group Inc., Sherbrooke Street West, Suite 910, Montreal Canada H3A 2R7. <http://chemcomp.com>
- 33 C5.0, Rulequest Research Pty Ltd, 30 Athena Avenue, St Ives NSW 2075, Australia, <http://rulequest.com>
- 34 A. Bender, R.C. Glen, "A discussion of measures of enrichment in virtual screening: comparing the information content of descriptors with increasing levels of sophistication", *J. Chem. Inf. Model.*, 2005, 45, 5, 1369-1375
- 35 J. Hert, P. Willerr, D.J. Wilton, P. Acklin, K. Azzaoui, E. Jacoby, A. Schuffenhauer, "Comparison of fingerprint-based methods for virtual screening using multiple bioactive reference structures", *J. Chem. Inf. Comput. Sci.*, 2004, 44, 3, 1177-1185
- 36 A.C. Good, M.A. Hermsmeier, S.A. Hindle, "Measuring CAMD technique performance: a virtual screening case study in the design of validation experiments", *J. Comput.-Aided. Mol. Des.*, 2004, 18, 7, 529-536
- 37 A. Verloop, W. Hoogenstraaten, J. Tipker, in E.J. Ariens Ed., "Drug Design, Vol VII", Academic Press, New York, 1976, 180-185

## Chapter 6 Quantum Mechanics Calculations

In chapter 2 we have defined the protein-substrate binding mode related to substrate stabilisation, substrate conversion and the inhibitory action of the 'substrates' on the CYP11B isoforms that cannot convert those particular substrates. The elucidation of substrate conversion requires a more detailed description of the protein-ligand complex, since different protein and ligand ionisations may play an important role. To increase computation detail, we perform a ligand-based quantum mechanics study to investigate the reactivity and conformations of the substrates involved in the synthesis of aldosterone. We discuss the various reaction paths that result in the formation of the most dominant reaction product, corticosterone, the final CYP11B1 product, 18-hydroxycorticosterone, and the final CYP11B2 product, aldosterone. The substrate conformation required for aldosterone synthesis identified by this study is placed above the heme. Subsequently, the conformation of the protein-substrate transition state is determined using quantum mechanics calculations. The idea behind this was (1) to get mechanistic insight on the catalytic process of aldosterone formation, and (2) to be able to expand this knowledge for the purpose of designing transition state analogues that are highly selective CYP11B2 inhibitors.

Parts of this chapter are described in:

L. Roumen, B. van Hoof, K. Pieterse, P.A.J. Hilbers, R. Plate, E.M.G. Custers, M. de Gooyer, J.F.M. Smits, I. Beugels, J. Emmen, H.C.J. Ottenheijm, D. Leysen, J.J.R. Hermans, "**Quantum mechanical considerations on the mechanism of the multistep conversion of 11-deoxycorticosterone to aldosterone by cytochrome P450 isoenzymes**", *Submitted, July 2008*

## 6.1 Introduction

In the previous chapters, in order to investigate the interactions between protein and ligand we have addressed our problem using molecular mechanics and molecular dynamics simulations. Using such Newtonian-like mechanics, it is possible to accurately describe the behaviour of molecules as well as the intermolecular interactions of the system. However, using these mechanics, several assumptions have been made that limit the investigation of the actual chemical activity of our protein, i.e. the synthesis of aldosterone by CYP11B2. In the molecular dynamics method, all atoms are defined by particles that possess a position and a point charge that does not allow induced polarity by atoms in its neighbourhood. The atoms also possess predefined bonds with other atoms in the system, but no breaking of bonds and formation of new bonds is allowed. These factors play an important role in the conversion of a substrate to a product and are best investigated using electronic structure calculations using quantum mechanics (QM). We have included quantum mechanics calculations in our study for two reasons.

First, we have used quantum mechanics calculations on the different substrates involved in the biosynthesis of aldosterone to obtain mechanistic knowledge and derive possible differences between the activity of CYP11B1 and CYP11B2. To accomplish this, we have performed a conformation analysis and we have calculated a measure for atom reactivity, namely the Fukui index of the carbon atoms of the steroid skeleton [1,2,3]. Calculation of the Fukui index allows us to speculate on which particular conformations of the substrates are likely to occur in the CYP11B1 active site and the CYP11B2 active site that cause them to exert a different synthetic activity.

Second, we have applied quantum mechanics calculations to propose structural features for novel selective inhibitors based on the unique synthesis of aldosterone by CYP11B2. Since only CYP11B2 can catalyse the conversion of 18-hydroxycorticosterone to aldosterone, we propose that a compound derived from 18-hydroxycorticosterone or its reaction intermediate will be a very potent and selective CYP11B2 inhibitor. Using steroids as inhibitor is a limited approach because there is a high probability that they are converted by other cytochrome P450 enzymes or block important nuclear receptors. However, by mimicking the most CYP11B2 specific features of the steroid, highly selective inhibitors can be derived. In paragraph 2.7.5 we have suggested that the conversion of 18-hydroxycorticosterone features an internal hydrogen bond between the C<sub>18</sub>-hydroxyl group and the C<sub>20</sub>-ketone. This internal hydrogen bond allows the abstraction of a second hydrogen atom from C<sub>18</sub> by the iron-oxo species of CYP11B2, followed by the formation of an C<sub>18</sub>-gem-diol. By designing an analogue of this transition state, i.e. the conformation of the 18,18-dihydroxycorticosterone we opt to achieve CYP11B2 selective inhibition.

## 6.2 Quantum Mechanics

Quantum mechanics (QM) is a fundamental theory that provides insight on the behaviour of systems at the atomic and subatomic level. It is based on probability distributions to explain the likelihood of finding a particle, such as an electron, in a particular region around the atom nucleus at a particular time. Electrons are not seen as localised particles in space, but rather as clouds of negative charge spread out over the atom in patterns called atomic orbitals. In quantum theory, particles can be described as waves, and as such, every electron has its own unique wave function that describes its orbital designation, the shape of the orbital, and the magnetic moment of the orbital. The wave functions of a hydrogen atom are the well known s, p, d and f orbitals. The square of the wave function multiplied by a volume element yields the probability to find the electron inside the volume element. This quantity is also referred to as the electron density in the volume element and can be measured by X-ray diffraction (the type of structures on which the modelling work was based in Chapter 2).

### 6.2.1 Quantum Mechanics Calculation Methods

The properties of a molecule in its stationary form can be derived by solving the Schrödinger equation that uses the wave functions to relate a system Hamiltonian (often described using the kinetic and potential energy) to the total energy of the system [4,5]. However, in virtually all cases it is impossible to solve this equation analytically, which has given rise to the development of several approximation methods. The commonly used methods are the Hartree-Fock approximation (HF [6]) and the Density Functional Theory (DFT [7,8]).

The Hartree-Fock method is also known as the Self-Consistent Field method (SCF). In HF calculations, the one-electron wave functions are approximated by a linear combination of atomic orbitals (LCAO). The HF method is an iterative process where an initial guess for the atomic orbitals is followed by the calculation of the Fock operator (coulomb and exchange contributions) to obtain an optimised set of atomic orbitals. The HF method is finalised when the new orbitals are nearly identical to the preceding iteration, which is called convergence. Thus, using the HF method, the geometry of the molecular structure can be optimised.

One of the important parts neglected by the HF method is electron correlation which refers to the interaction between electrons in the system. DFT gives approximate solutions to both exchange and correlation energies (though HF does include accurate calculation of the exchange energy). The objective of DFT is also to replace the electronic wave function of the HF with the electronic density as the basic quantity. Whereas the wave function is dependent on  $3N$  variables (3 spatial variables for each of the  $N$  electrons), the density is a function of 3 variables and is a simpler quantity to use both conceptually and practically.

Currently, it is common to use a hybrid version of the two methods, hybrid functional methods. The most applied schematic are the BLYP and B3LYP methods [9,10,11,12,13,14,15]. These methods combine the exact HF exchange with a DFT exchange term and also add a correlation functional. In general, accurate QM calculations are performed by combining the B3LYP method with a split-valence basis set to generate the orbitals and initial electron density.

### 6.2.2 Basis Sets

In both of the methods mentioned to perform QM calculations, basis sets are used for the accurate calculation of the electron configuration and thereby molecular geometry. A basis set is a set of functions that describes the molecular orbital as a linear combination of the functions with the exact coefficients to be determined by the HF or DFT methods. These functions are usually atomic orbitals centred on atoms, but there are also functions centred in bonds and lone pairs. Various basis sets are used in practice, most of which are composed of either Slater-type orbital functions [16] or Gaussian-type orbital functions [17]. The simplest basis set is the STO-3g basis set, which is an acronym for Slater-Type-Orbitals simulated by the use of 3 Gaussian functions added together. This basis set is known as a *minimal* basis set, indicating that it has only as many orbitals as are needed to accommodate the electrons of neutral atoms and retain their spherical symmetry. Thus, the STO-3g basis set only has one basis function per hydrogen atom (1s), five basis functions per atom from Li to Ne (1s, 2s, 2px, 2py and 2pz), nine for second row elements Na to Ar and so on. The STO-3g basis set usually performs rather well at reproducing geometries of simple organic molecules, however, it does not perform well for the calculation of the energy, nor does it perform well in response to changing and anisotropic molecular environments as a result of inter-molecular interactions.

Since the valence electrons principally take part in molecular bonding, it is common to represent their orbitals by more than one basis function. The basis sets belonging to this type of modelling are called *split-valence* basis sets. For these basis sets, the atomic orbitals are split into two parts, an inner compact orbital and an outer diffuse orbital. The simplest split-valence basis set is called the 3-21g. Here, the 3 indicates that the core orbitals are represented by 3 Gaussian functions, whereas the inner and outer valence orbitals consist of 2 and 1 Gaussian functions, respectively. More advanced split-valence basis sets can involve next to the inner and outer orbital descriptions, an additional middle orbital description (ex. 6-311g).

Additional functions can be added to the basis set to increase accuracy. Examples are *polarisation* (\*) [18] and *diffuse* functions (+) [19], but also extra orbitals are to be considered (such as a d-function added to a basis set with p-type orbitals), as well as additional basis sets. Polarisation functions allow a p-orbital to be polarised away from the nucleus by mixing with a d-orbital of lower symmetry. This can improve the modelling of small rings and second-row elements, but it can also be applied to hydrogen atoms. Diffuse functions allow accurate computation of those orbitals that are most distant from the atomic nuclei, which can be important when considering anions or unshared electrons. Additional specialised basis sets are often used for the accurate modelling of heavy atoms such as metals.

In our study we have used the abovementioned B3LYP hybrid functional in combination with the 6-31g split-valence basis set [20,21,22,23,24,25,26,27,28,29]. For the modelling of heme structures, we have added the LANL2DZ basis set to the iron [30], polarisation functions to all atoms, and expanded the basis sets of sulphur, oxygen and nitrogen with diffuse functions. For the accurate calculation of the Fukui index of the substrates, we have expanded the 6-31g basis set with diffuse functions.

### 6.3 Catalytic cycle of cytochrome P450 enzymes

Before elaborating on the quantum mechanics study, we first explain the theory of substrate conversion performed by cytochrome P450 enzymes. The catalytic activity of cytochrome P450 enzymes involves different redox states of the oxygenated enzyme (Figure 6–1). The existence of the two-electron reduction resulting in several intermediate states of the iron-oxygen species was first discovered for CYP101 [31,32,33]. The different locations around the porphyrin structure are called proximal, where the heme is covalently anchored to the enzyme by the cysteine thiolate group (5th ligating position), and distal, where the iron interacts with either the solvent, oxygen or the substrates (6th ligating position).

In the resting state of the enzyme, the heme iron is coordinated with its 6th ligating position to a water molecule, with the iron atom in a low-spin ferric state (**1**,  $S=1/2$ ). In the first step of the substrate conversion cycle, the distal water molecule is displaced in favour of a substrate molecule (**RH**), upon which the spin-state of the ferric iron changes to high-spin (**2**,  $S=5/2$ ). The high-spin ferric iron is reduced to a ferrous state (**3**) after the supply of an electron. The shifting of spin-states is not typical for every cytochrome P450 enzyme, thus it is difficult to generalise these initial steps [34,35]. After oxygen binding, the last relatively stable intermediate in the cycle is produced (**4**). This complex is further reduced after the supply of an electron to the peroxo-ferric intermediate (**5a**). Protonation of this complex on the distal oxygen yields the hydroperoxo-ferric intermediate (**5b**), also known as **Compound 0**, or **Cpd 0**. This complex is protonated once more at the distal oxygen atom with subsequent hydrolysis of the O-O bond and formation of the iron-oxo intermediate, also known as **Compound I**, or **Cpd I** (**6**). The subsequent step in the cycle is called the **Oxygen Rebound** step, and is a process in itself [36]. This oxidation of the substrate yields a product complex (**7**). The product (**ROH**) is then displaced by water to yield the resting state [37].

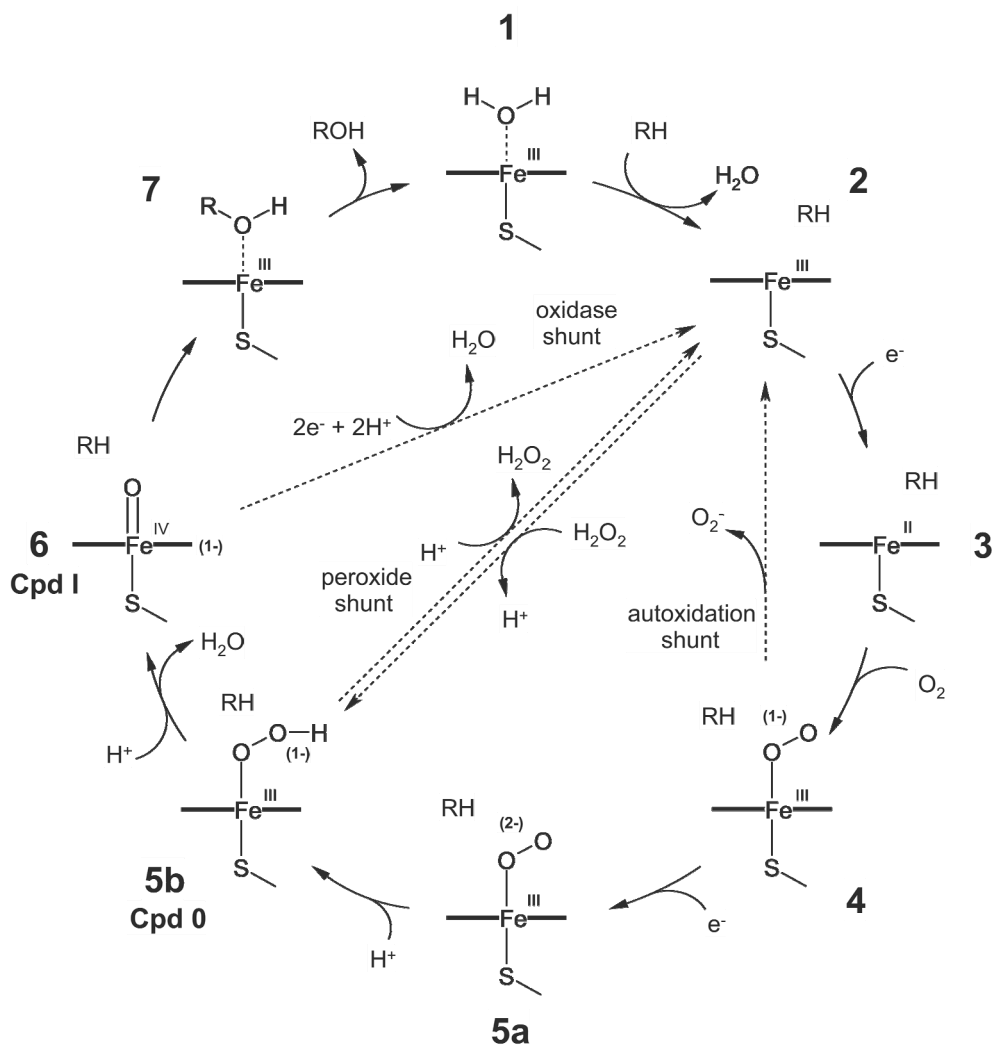


Figure 6–1 Catalytic cycle of cytochrome P450s.

### 6.3.1 Oxygen Rebound in detail

The consensus view of the oxygen rebound mechanism stated that an iron-oxygen species abstracts a hydrogen atom from the substrate, forming an iron-hydroxy species and an alkyl radical intermediate (**Figure 6–2**). Subsequently, the quick displacement of the hydroxyl group from the iron to the alkyl (the rebound) results in the final product.

The mechanism of abstraction of a hydrogen atom from an alkyl can be performed in a linear array of the atoms C–H–O [38]. During such a process, the distance between the carbon atom and the oxygen atom in the transition state of the C–H–O array is between 2.50 and 2.53 Å (Figure 6–2a) [39]. However, measurement of the life-time of the radical suggest

that this approach is non-plausible. The life-time of the radical has been determined to be only 70 fs [40], which is too short to bridge this distance. However, this time length is consistent with bond stretches and rotational processes [41], leaving the possibility of a 'side-on' approach for hydrogen abstraction (Figure 6–2b) [40]. Two main characteristics of the process support this 'side-on' approach. Firstly, the reaction is concerted, because the short life-time of the radical suggests that the reaction possesses no true intermediate structure [40]. Secondly, the reaction is non-synchronous, because it has been determined that the abstraction of the hydrogen atom precedes the collapse of the C–O bond [42]. The distance between the iron atom and the substrate carbon atom during the hydrogen abstraction mechanism reaches up to 5.0 Å [43].

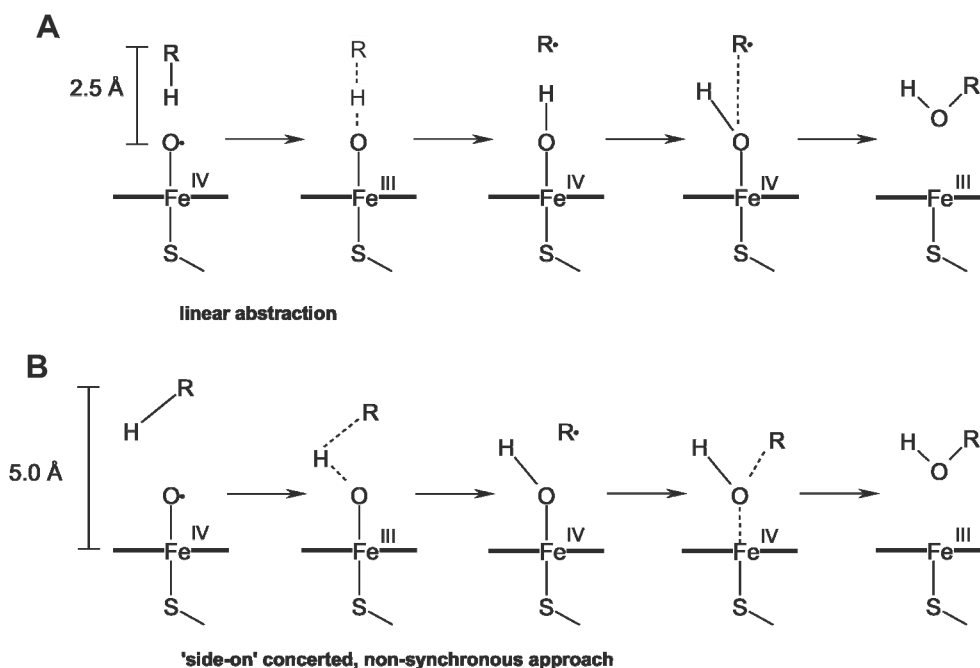


Figure 6–2 a) Mechanism for linear hydrogen abstraction by the heme (which is non-plausible for cytochrome P450 enzymes) b) The originally suggested oxygen rebound mechanism that is concerted and non-synchronous

It has been suggested by Newcomb et al [44,45], that the oxygen rebound mechanism is not the only mechanism possible for the conversion of substrates. The mechanism Newcomb et al have suggested is the so-called 'two-oxidant mechanism', where both **Cpd 0** and **Cpd I** can perform substrate conversion. This statement has been based on the experimental results using specific probes that distinguish between radical conversion (oxygen rebound) and cationic conversion (**Figure 6–3**) [44,46]. The cationic conversion of the substrate involves the electrophilic insertion of an OH<sup>+</sup> group by **Cpd 0** that leads to the formation of a carbocation in the substrate. During this insertion, it is proposed that the catalytic threonine



residue in the centre of the I-helix forms a stabilising interaction for the insertion of the  $\text{OH}^\cdot$ . The carbocation is able to rearrange itself to form several different rearranged products, which are not found for the radical oxygen rebound, since the radical life-time is too short. It has been proposed that this process may even be completed through the  $\text{H}_2\text{O}_2$  shunt (**2** to **5b**, Figure 6–1) [44].

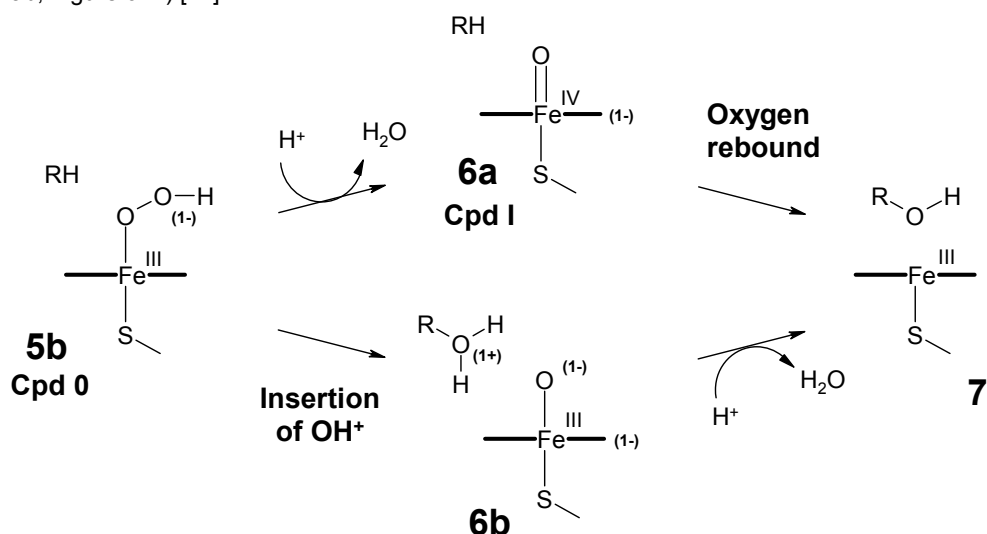


Figure 6–3 Two-oxidant mechanism suggested by Newcomb et al [44]. The original oxidant performs the oxygen rebound, whereas the suggested second oxidant involves the insertion of an  $\text{OH}^\cdot$ -group.

Recently, the mechanism of the oxygen rebound step (Figure 6–2b), has been expanded to the 'two-state reactivity' mechanism (TSR Figure 6–4) [47,48,49,50]. This theory implies that **Cpd I (5b)** is in fact a two-state reagent; one state is low-spin (doublet state), and the other high-spin (quartet state). The energy levels of these spin-states are energetically very close, and product formation is a result from the interplay of both these reactive states. Both reaction paths start with hydrogen abstraction from the substrate (**S**) to form an intermediate ferryl-hydroxo and substrate radical state (**C<sub>1</sub>**). The difference between the two spin-states is that high-spin state forms a radical that possesses a significant barrier for oxygen rebound ( ${}^4\text{C}_{\text{reb}}$ ,  ${}^4\text{TS}_{\text{reb}}$ ) and the low-spin state possesses a virtually barrierless oxygen rebound ( ${}^2\text{C}_{\text{reb}}$ ). Since the low-spin reaction contains a small barrier, it is effectively concerted, corresponding to the measurement results of the radical life-time [40]. The life-time of the barrier in the high-spin is significantly higher than that of the low-spin, allowing rearrangements to take place (Figure 6–4). Resulting from these findings, Kumar et al have suggested that the emergence of rearranged product from experiments by Newcomb et al [44], previously explained by the 'two-oxidant' mechanism, can also be explained by the two-state reactivity mechanism [50]. Furthermore, since the orbital composition of **Cpd I** involves several closely lying orbitals and becomes denser throughout the reaction path, other states may also participate to give rise to a multi-state reactivity mechanism [48,51,52].

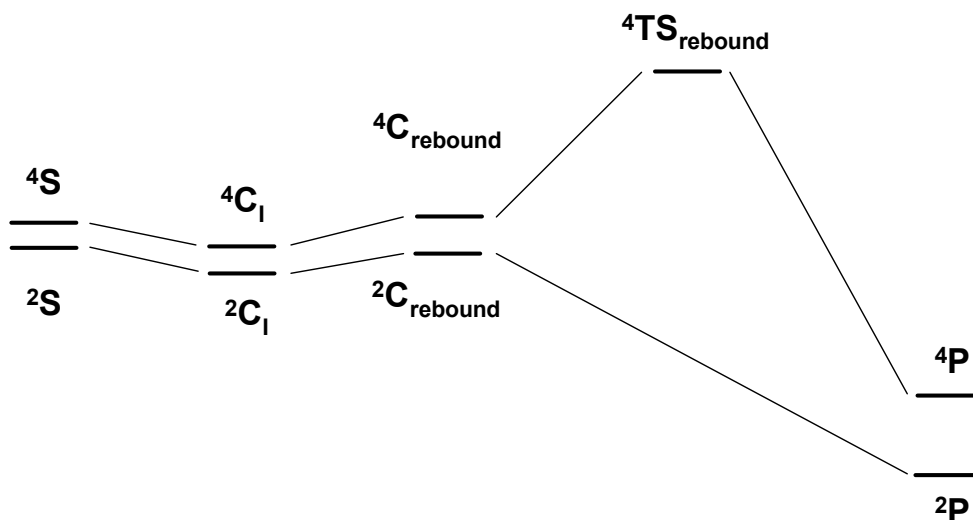


Figure 6–4 Two-State Reaction Mechanism: schematic energy profiles of the high-spin quartet and low-spin doublet states of Cpd I. S represents the heme-substrate complex.  $C_1$  represents the stable heme-substrate radical state.  $C_{\text{rebound}}$  and  $TS_{\text{rebound}}$  represent the oxygen rebound. P represents the heme-product state. The Arabic number 2 and 4 indicate the doublet and quartet states, respectively.

### 6.3.2 Uncoupling

Uncoupling is a process that allows the cytochrome P450 enzyme to return to its non-oxygenated state. These non-productive shunts are the result of  $O_2$  conversion to peroxide, superoxide or water (Figure 6–1) [37]. The collapse can occur if electron or proton delivery is not timely or if the substrate is resistant to oxidative attack [53]. The availability of solvent near the molecular oxygen in the active site plays an important role in destabilising the complex of the iron-oxygen species and the substrate [54,55]. The presence of solvent can act as a supply of protons to facilitate the dissociation of di-oxygen and the formation of  $H_2O_2$ .

Complex instability can be caused by ligands that do not fit the active site cavity very well, or ligands that possess no strong stabilising interactions in the active site, allowing them to move. A good example of this scenario is described for the conversion of camphor analogues by CYP101. The instability of the camphor analogues in the active site is caused either by a lack of electronic and/or hydrophobic interactions. The consequences of the weaker binding is both an increased amount of different hydroxylated products due to substrate rotation, and a decreased efficiency of the substrate conversion due to uncoupling [54,56,57,58].

## 6.4 Mechanistic Knowledge

The conversion cascade of 11-deoxycorticosterone (DOC) to aldosterone involves the synthesis of several intermediate structures (Figure 6–5) [59,60,61]. From *in vitro* measurements, it has been derived that the main product of both CYP11B1 and CYP11B2 is the formation of corticosterone (B). CYP11B1 is able to form 18-hydroxy-11-deoxycorticosterone (18OH-DOC) as well as small amounts of 18-hydroxycorticosterone (18OH-B). The rat CYP11B1 isoform is even able to synthesise 19-hydroxy-11-deoxycorticosterone (19OH-DOC) [60]. CYP11B2 is also able to synthesise 18OH-DOC and 18OH-B, yet its distinction from CYP11B1 lies in its unique ability to oxidise the C<sub>18</sub>-hydroxyl group into a C<sub>18</sub>-oxo group, characteristic for aldosterone. The *in vitro* measurements suggest that CYP11B1 and CYP11B2 possess a preferential hydroxylation for C<sub>11β</sub>, followed by C<sub>18</sub>. Rat CYP11B1 also possesses the ability to hydroxylate the C<sub>19</sub>, although with the lowest preference, since 19OH-DOC is formed the least.

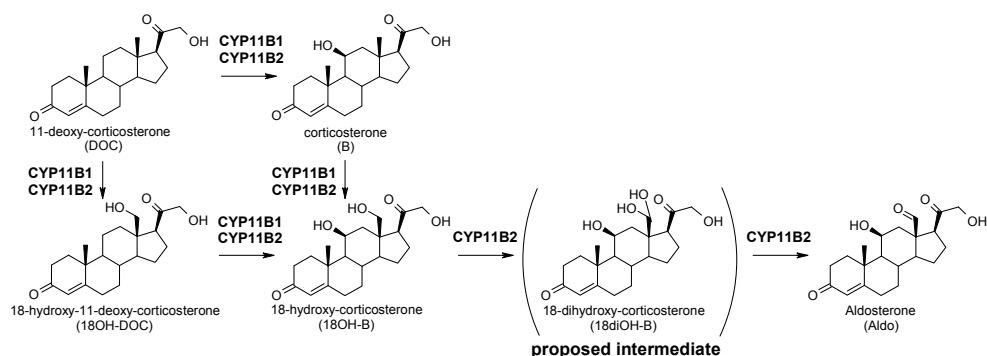


Figure 6–5 Biosynthesis of aldosterone from DOC through either B or 18OH-DOC.

When evaluating the substrate hydroxylation reactions performed by the CYP11B family [59,60], little is known about the exact conversion mechanism. However, recent explorations have been applied to provide new insights. First, in correspondence to other research groups we have proposed that the conversion of 18OH-B to aldosterone by CYP11B2 features the formation of a C<sub>18,18</sub>-gemdiol (Figure 6–5) [62,63,64]. Possibly, CYP11B1 is not able to oxidise the same carbon twice which may explain why CYP11B1 cannot catalyse aldosterone formation, in contrast to CYP11B2. This hypothesis implies that the conversion of 18OH-B features a second hydrogen atom extraction followed by oxygen rebound, rather than a direct oxidation mechanism. Indeed, the presence of gemdiol intermediates has been explained by the experimental work on CYP2E1 by Guengerich et al [65,66,67], and has also been hypothesised for other cytochrome P450 enzymes (CYP19 [43,63,68,69], and CYP27 [70]). Furthermore, the aromatisation of androgens has been detailed by a quantum mechanical study by Hackett et al, which emphasises the involvement of a gemdiol in the biosynthesis of estrogens [43].

Second, measurements and calculations on other cytochrome P450 enzymes have confirmed that the multistep conversion of substrates involves intermediate structures that do not dissociate from the enzyme and exchange with the medium (CYP2E1 [65,66,67], CYP17 [71], CYP19 [43,68,69], CYP24 [72] and CYP27 [70]). If the conversion of DOC to aldosterone is a multistep conversion, the intermediate structures do not exchange with the medium until the formation of the 18-gemdiol (Figure 6–5). When this gemdiol exchanges with the medium, it is dehydrated to aldosterone.

Third, from *in vitro* measurements it is debatable which of the possible conversion pathways featuring either B or 18OH-DOC as intermediate structure, is dominant. In humans, it has been shown that 18OH-DOC is a poor substrate for CYP11B1 and rather acts as a competitive inhibitor for the conversion of DOC into B [59]. For CYP11B2, 18OH-DOC also decreases the conversion of DOC into B, yet it increases the formation of 18OH-B and aldosterone. These results suggest that 18OH-DOC is only a substrate for CYP11B2. However, this does not necessarily hold true when the conversion of DOC is a multistep conversion, as an unfavourable conformation of 18OH-DOC may be converted into 18OH-B, although at a low rate. As a result, B may even be a by-product when the multistep conversion into 18OH-B and aldosterone involves the synthesis of 18OH-DOC. This proposition challenges the general assumption that CYP11B1 and CYP11B2 are C<sub>11</sub>-C<sub>18</sub>-hydroxylases. We propose that CYP11B2 may even be a C<sub>18</sub>-C<sub>11</sub>-hydroxylase, and we will test this hypothesis accordingly.

To elucidate the conversion mechanism of DOC to aldosterone, we have performed a quantum mechanics analysis that can provide viable information for the understanding of the biosynthesis performed by CYP11B1 and CYP11B2. Extending the theory that the conversion intermediates do not leave the active site, we propose that CYP11B2 requires one dominant conformation of the initial substrate during the multistep conversion before the final product is released to the medium.

To investigate this hypothesis, we have performed a ligand-based quantum mechanical analysis of the substrate conformations that occur during conversion from DOC to aldosterone. As a measure for carbon atom reactivity we have used the Mulliken population analysis derived Fukui index (Equation 6–1) [1,2,3]. Here,  $\rho$  is the atomic orbital density on a carbon atom and N is the amount of electrons in the model system.  $F^+$  is the Fukui index of a nucleophilic attack by the oxygen species on a carbon atom. The calculation of this particular Fukui index requires the calculations of the ligand with a net charge of 0 and the ligand with a net charge of -1. A high positive  $F^+$  indicates a higher preference for carbon atom oxidation. The cut-off value for the Fukui index has been chosen as 0.03 comparable to normal fluctuations observed for hydrogen atoms.

$$\text{Equation 6-1} \quad F^+ = \rho_{N+1} - \rho_N$$

#### 6.4.1 Steroid Conformation Generation

For this study, we have used a systematic conformation analysis as shown in (Figure 6–6). Beginning with a conformational analysis of DOC, we continued to generate conformations from the energetically favourable structures of DOC for every succeeding substrate in the cascade. The initial conformations of DOC were generated using MOE 2007.08 [73]. The conformations within a root mean square distance (RMSD) of 0.3 Å were labelled as duplicate geometries. Input files were generated with Molden v4.6 [74] for the resulting 10 DOC structures following geometry optimisation using the quantum mechanics package Gaussian03 [75]. A Mulliken population analysis was performed to determine the electron density distribution of the atomic orbitals. For the accurate calculation of the Fukui index a combination of the B3LYP functional and the 6-31+g basis set was chosen. The neutral and negatively charged states of the structures were calculated according to the Fukui index for a nucleophilic attack of the oxygen species.

The geometry and electron distribution of the steroid conformations were compared using a root mean square error (RMS). The initial DOC geometries that converged to similar conformation minima were pruned and the lowest energy conformations were used as input for the conformation analysis of 18OH-DOC and B. Two possible conformations exist for the C<sub>11β</sub>-hydroxyl group. These conformations were labelled **u** and **v**, indicating that the hydroxyl group is pointing away from C<sub>8</sub> and towards C<sub>8</sub>, respectively (Figure 6–7). Four generic directions for the C<sub>18</sub>-hydroxyl group exist. Two of these directions are towards C<sub>8</sub> and C<sub>12</sub>, away from the hydroxyacetyl group, labelled **w** and **x**, respectively. The other two directions that were analysed allow the C<sub>18</sub>-hydroxyl group to approach the C<sub>20</sub>-carbonyl and form an internal hydrogen bond. These directions are roughly towards C<sub>16</sub> and C<sub>20</sub> and labelled **y** and **z**, respectively (Figure 6–7). Conformations for 18OH-B and 18diOH-B were derived from the optimal conformations found for 18OH-DOC and B.

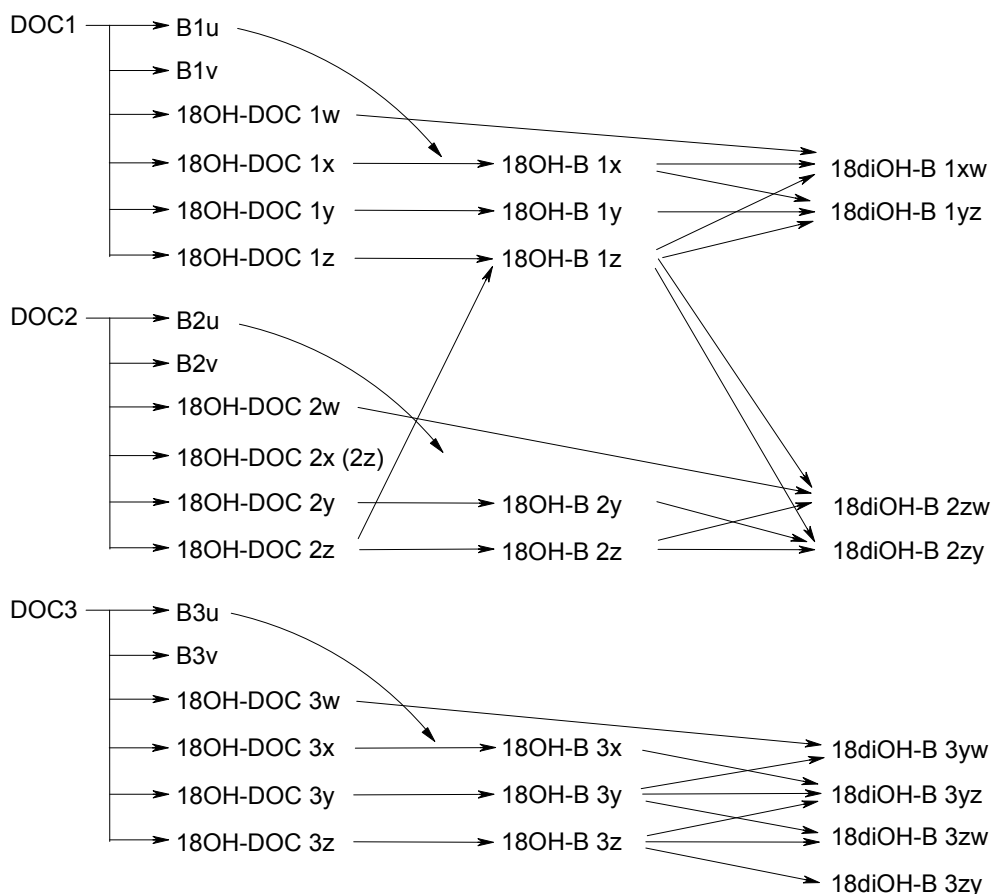


Figure 6–6 Flowchart of the conformations involved in aldosterone synthesis that are used in our analyses. Crossover between conformations is portrayed with a dashed arrow. The steroids are 11-deoxycorticosterone (DOC), corticosterone (B), 18-hydroxy-11-deoxycorticosterone (18OH-DOC), 18hydroxycorticosterone (18OH-B) and 18-dihydroxycorticosterone (18diOH-B). Conformation nomenclature is determined by a numeral referring to the initial DOC conformation as well as one or two Arabic letters referring to different orientations of the hydroxyl groups.

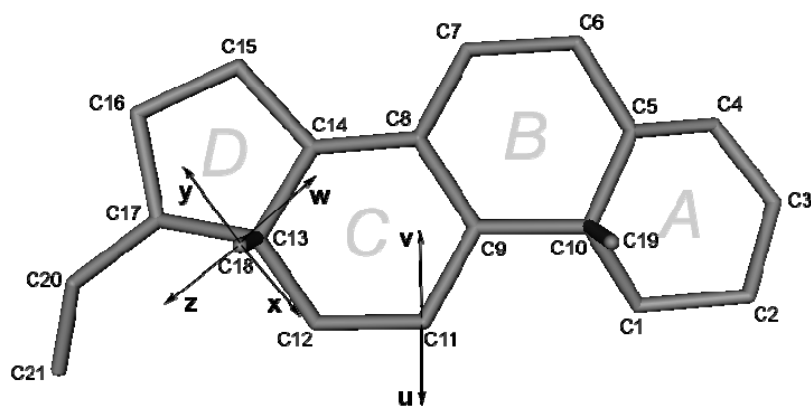


Figure 6–7 Steroid carbon atom enumeration, direction of the  $C_{118}$ -hydroxyl group of B (blue) and direction of  $C_{18}$ -hydroxyl groups (red). The enumeration of the rings is shown in italic letters.

## 6.4.2 Results and Discussion: Mechanistic Knowledge

### 6.4.2.1 DOC Conformations

For the 10 initial DOC conformations, several structures have converged to an identical minimal energy conformation. The number of unique energy minima for DOC has been proven to be 6, and is attributed to conformational changes in the hydroxyacetyl group (Figure 6–8a). The  $C_{20}$  carbonyl group points either towards  $C_{12}$  or  $C_{16}$ . Furthermore, three favourable types of conformations have been identified for the hydroxyacetyl group; eclipsed, gauche and antiperiplanar (Table 6–1). In the eclipsed conformations, a very strong internal hydrogen bond is formed that results in a lower potential energy than observed for the gauche and antiperiplanar conformations. For the  $C_{16}$ -oriented conformations, one gauche conformation optimises to an antiperiplanar conformation due to steric hindrance (conformation DOC6). The antiperiplanar conformation is higher in energy than the gauche conformations and, therefore, has been discarded from further analysis. The three lowest energy conformations (DOC1, DOC2 and DOC3) have been used for the conformational analysis of B and 18OH-DOC. DOC3 has been used for two additional reasons. Firstly, both DOC1 and DOC2 feature an internal hydrogen bond for the hydroxyacetyl group, that is absent in DOC3, and secondly, in Chapter 2 we have proposed the hydroxyacetyl conformation of DOC3 as a requirement for aldosterone synthesis [64]. The three conformations of DOC are displayed in (Figure 6–9).

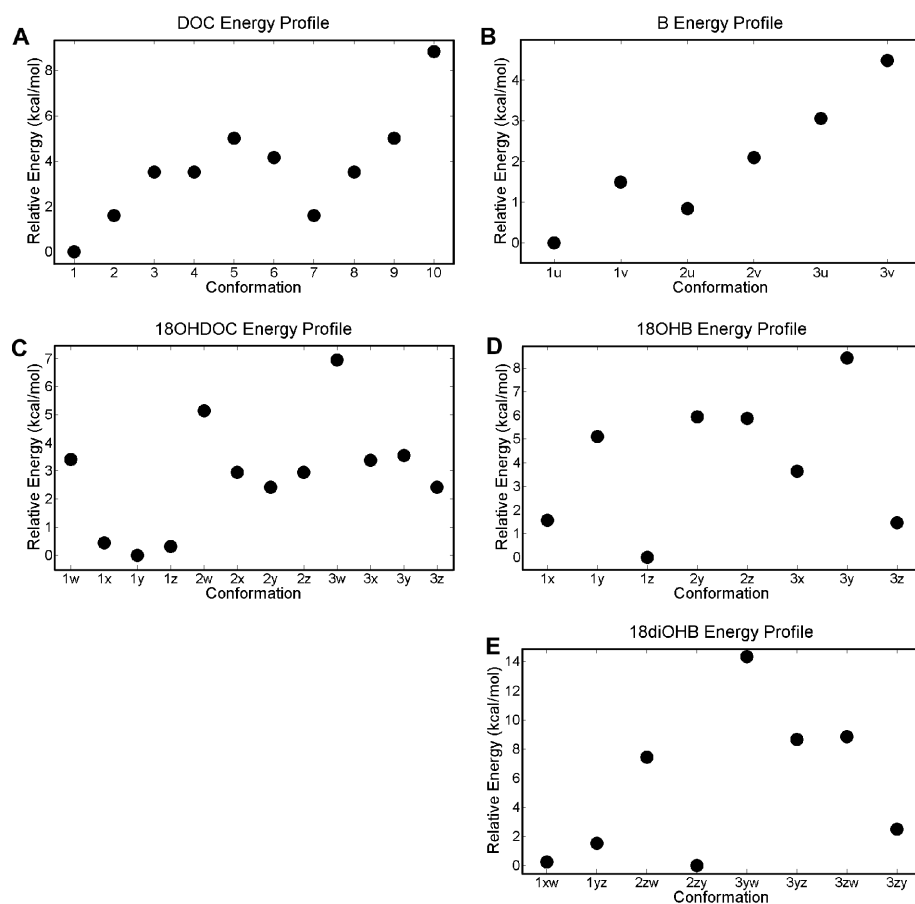


Figure 6-8a-e Relative energies derived for the conformations of a) DOC b) B c) 18OH-DOC d) 18OH-B and e) 18diOH-B. The horizontal axis indicates the conformation enumeration.



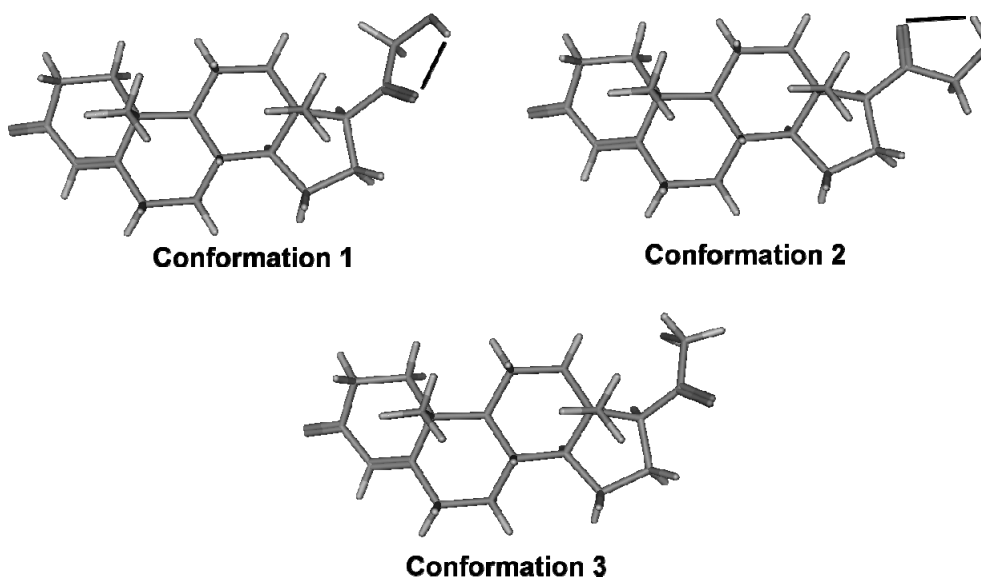


Figure 6–9 The three lowest energy conformations for the hydroxyacetyl group of DOC. Hydrogen bonds are indicated with lines. In Chapter 2 we have proposed that conformation 3 is required for the synthesis of aldosterone.

Table 6–1 Conformation of the Hydroxyacetyl group of DOC

Conformation	Dihedral Angle O20–C20–C21–O21	Type	Orientation C20 Carbonyl
1	-4.9°	eclipsed	C16
3	155.5°	gauche	C16
4 (same as 3)	155.5°	gauche	C16
8 (same as 3)	154.3°	gauche	C16
6	-177.9°	antiperiplanar	C16
2	-3.4°	eclipsed	C12
7 (same as 2)	-2.5°	eclipsed	C12
5	-144.2°	gauche	C12
9 (same as 5)	-145.3°	gauche	C12
10	116.2°	gauche	C12

#### 6.4.2.2 B Conformations

Since the conformation of the hydroxyacetyl group has remained consistent, the conformations of B1, B2 and B3 have been found to be identical to the corresponding conformations of DOC1, DOC2 and DOC3. For each of the three hydroxyacetyl group conformations, the lowest energy contains the C<sub>11</sub>-hydroxyl group pointing away from C<sub>8</sub> (B1u, B2u, B3u in Figure 6–7 and Figure 6–8b).

#### 6.4.2.3 18OH-DOC Conformations

For 18OH-DOC, four conformations were generated for the C<sub>18</sub>-hydroxyl group labelled **w** to **z** (Figure 6–7 and Figure 6–8c). An internal hydrogen bond between the C<sub>18</sub>-hydroxyl and the C<sub>20</sub>-carbonyl groups has been found for conformations **1y**, **1z**, **3y** and **3z**. For these structures, the conformation of the hydroxyacetyl group has slightly reoriented to optimise the hydrogen bond. No hydrogen bond has been observed for conformations **2y** and **2z**. If the hydroxyacetyl group had reoriented itself to form a hydrogen bond, these conformations would have become identical to conformations **1y** and **1z**, respectively. Since the conformation **2z** does not form a hydrogen bond between the C<sub>18</sub>-hydroxyl group and the C<sub>20</sub>-carbonyl group, it converges to an identical conformation as **2x**. The energy of the **w** conformations are much higher than those of the **x**, **y** and **z** conformations. Therefore, the **w** conformations have been discarded from further analysis.

#### 6.4.2.4 18OH-B Conformations

Formation of 18OH-B can take place from either B or 18OH-DOC. Therefore, starting conformations of 18OH-B were based on the lowest energy conformations of both B and 18OH-DOC. Because the lowest energy conformations of B all possess their C<sub>11</sub>-hydroxyl group pointing away from C<sub>8</sub>, we have chosen to generate all 18OH-B conformations with the same direction for the C<sub>11</sub>-hydroxyl group, the **u** conformation. The following 18OH-DOC structures were considered; **1x**, **1y**, **1z**, **2y**, **2z**, **3x**, **3y** and **3z** (Figure 6–8d). The C<sub>11</sub>- and C<sub>18</sub>-hydroxyl groups of these conformations possess different hydrogen bonding patterns that are summarised in (Table 6–2).

The introduction of the C<sub>11</sub>-hydroxyl group to 18OH-B results in very stable internal hydrogen bonds. The most optimal electronic conformations (**1x**, **1z**, **3x** and **3z**) all possess hydrogen bond pointing from the C<sub>11</sub>-hydroxyl group to the C<sub>18</sub>-hydroxyl group. The presence of this specific hydrogen bond limits the rotational freedom of the hydroxyacetyl group and has an impact on the orientation of the C<sub>20</sub>-carbonyl group. Investigation of the conformations of **1z** and **3z** reveals that the C<sub>20</sub>-carbonyl group has changed its orientation from C<sub>16</sub> to C<sub>12</sub>.

Comparing the 18OH-B conformations to those of 18OH-DOC, it can be derived that a rotation of around 60° of the hydroxyacetyl group allows crossover between structure 18OH-B **2z** to 18OH-B **1z** (Figure 6–6).

Table 6–2 Internal hydrogen bonds for the 18OH-B conformations

Conformation	Hydrogen bonds	Type	Orientation C20-Carbonyl
1w	C21-OH to C20-carbonyl C18-OH to C11-OH	eclipsed	C16
1x	C21-OH to C20-carbonyl C11-OH to C18-OH	eclipsed	C16
1y	C21-OH to C20-carbonyl C18-OH to C20-carbonyl	eclipsed	C16-C12
1z	C21-OH to C20-carbonyl C18-OH to C20-carbonyl C11-OH to C18-OH	eclipsed	C12
2y	C21-OH to C20-carbonyl	eclipsed	C12
2z	C21-OH to C20-carbonyl C18-OH to C11-OH	eclipsed	C12
3x	C11-OH to C18-OH	antiperiplanar	C16
3y	C18-OH to C20-carbonyl	antiperiplanar	C16
3z	C11-OH to C18-OH C18-OH to C20-carbonyl	antiperiplanar	C12

#### 6.4.2.5 18diOH-B Conformations

Several conformations of 18diOH-B can originate from multiple 18OH-B conformations and there are also many crossover possibilities (Figure 6–6, Figure 6–8e, Figure 6–10). The stabilisation of the additional C<sub>18</sub>-hydroxyl group is most favourable for each of the **1yz** combinations, as well as the **1xw** conformation due to an interplay of both a C<sub>11</sub>-C<sub>18</sub> and a C<sub>18</sub>-C<sub>20</sub> hydrogen bond. Additional structures can be determined if the 18OH-DOC conformation **w** is investigated, however, several of these extrapolated structures are already covered (Figure 6–6), and the additional conformations are anticipated to be less likely products of the oxidation sequence, due to the unfavourable potential energy of their intermediate structures.

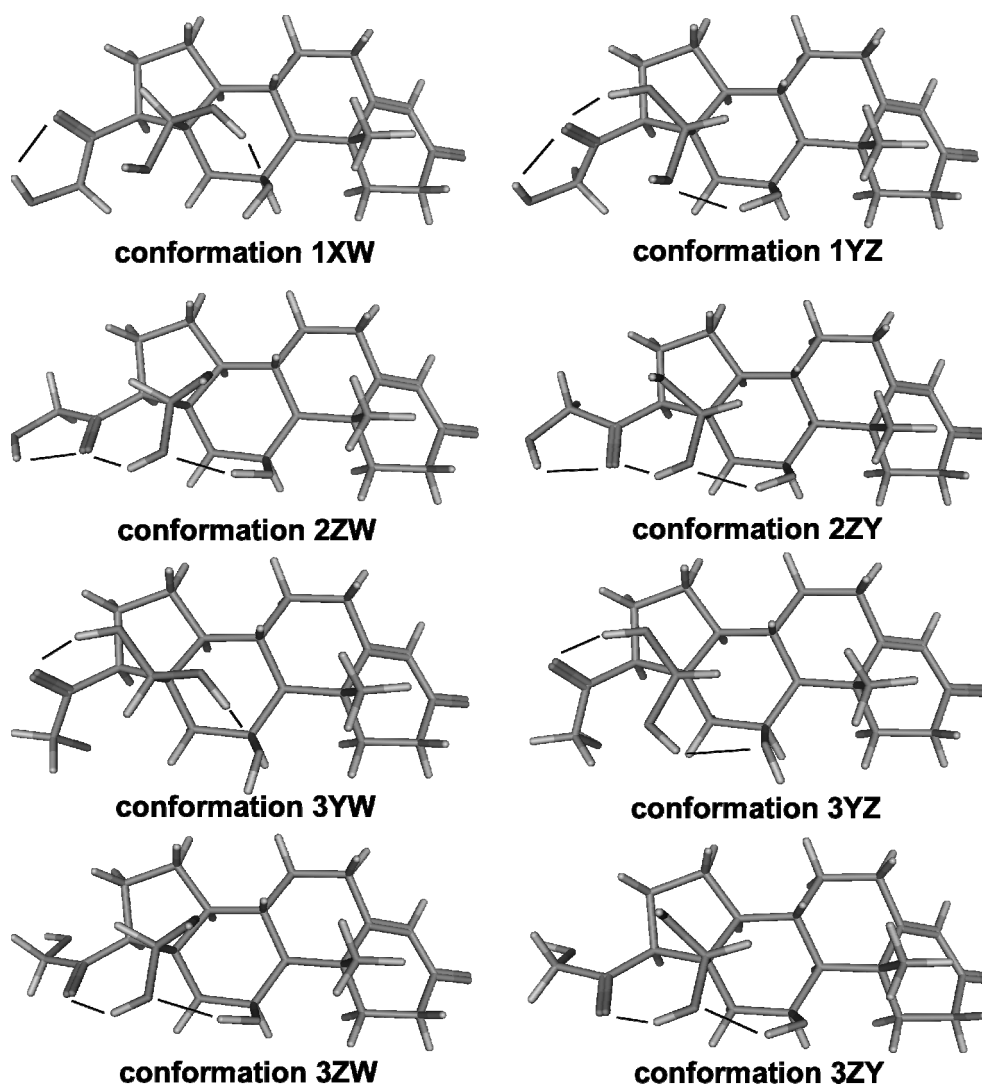


Figure 6–10 The investigated conformations of 18-dihydrocorticosterone. Hydrogen bonds are indicated with lines.

### 6.4.3 Fukui index analysis

Because the steroid C<sub>11</sub>, C<sub>18</sub> and C<sub>19</sub> are known hydroxylation sites for the CYP11B family, the Fukui index calculations have been restricted to these atoms. A high positive Fukui index indicates a higher preference for carbon atom hydroxylation. A general observation on the Fukui indices is the frequently high value for C<sub>19</sub> (Figure 6–11). For all 6 DOC conformations, only **DOC3** possesses a higher Fukui index for C<sub>11</sub> than for C<sub>19</sub>, and none of the DOC conformations possess a higher C<sub>18</sub> Fukui index than the C<sub>19</sub>. The presence of the C<sub>20</sub>-hydroxyl group is unfavourable for the Fukui index of C<sub>11</sub> and C<sub>18</sub> because it decreases the electron density in the negatively charged conformation. As a result, the C<sub>19</sub> Fukui index is higher.

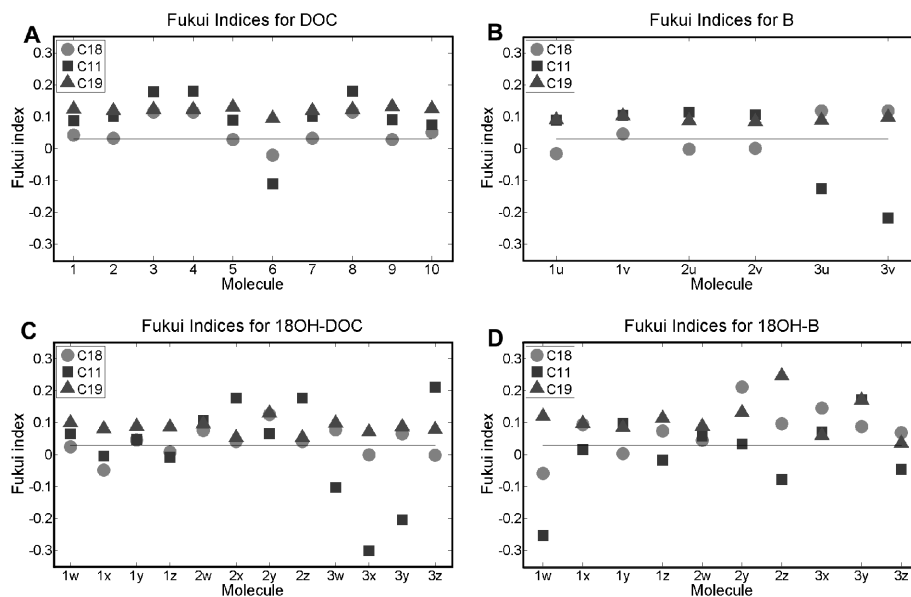


Figure 6–11a-d Fukui indices for a) DOC b) B c) 18OH-DOC and d) 18OH-B. Indicated with the horizontal line is the Fukui index threshold value.

### 6.4.4 Fukui index analysis of sequential paths

The most likely multistep conversion paths for hydroxylation of DOC to aldosterone were investigated. Although C<sub>18</sub>-hydroxylation of DOC is not preferred in accordance to the low C<sub>18</sub> Fukui indices (Figure 6–11a), the index is high enough to result in C<sub>18</sub>-hydroxylated products of **DOC1** and **DOC3**. For **DOC2**, the index falls within the cut-off of 0.03, however, this path has been investigated for completeness. The paths to which the initial C<sub>18</sub>-hydroxylation applies are Path 4, Path 5 and Path 6, which were investigated in accordance to the assumption that the Fukui index is high enough to allow hydroxylation. Because it is unknown whether 18diOH-B can be synthesised via 18OH-DOC, these paths have not been investigated.

The main consideration for all paths is that the substrates are sequentially converted allowing only limited movement of the intermediate substrates in the active site before the final product is released from the protein active site to the environment. The Fukui index for each substrate conformation can be reviewed in Figure 6–11a-e. A brief summary of the investigated paths including success and failure is listed in Figure 6–12. Despite having discarded them from further analysis, the calculation of the multistep paths required the calculation of the Fukui indices for the 18OH-B **1w** and **2w** conformations.

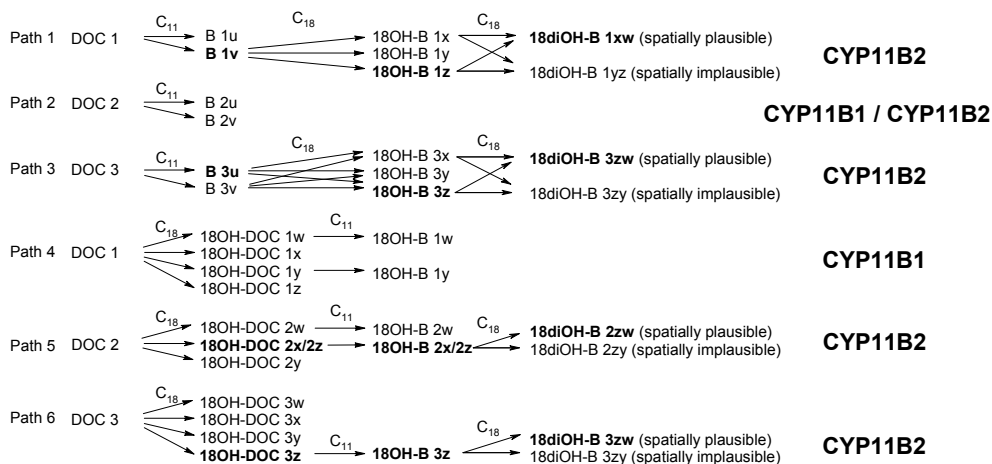


Figure 6–12 All possible multistep paths from DOC to 18diOH-B based on the conformations DOC1, DOC2 and DOC3. For each path, indicated in Bold are the most likely routes taken inside that path. In addition, each path has been given a label indicating whether we deem the path preferred by CYP11B1 or CYP11B2.

#### 6.4.4.1 Path 1

The first path is the conversion of **DOC1** through  $C_{11}$ -hydroxylation, followed by two consecutive  $C_{18}$ -hydroxylations to 18-diOH-B. The Fukui indices of **DOC1** indicate a preference of  $C_{19}$  over  $C_{11}$  and finally  $C_{18}$ . Assuming that  $C_{19}$ -hydroxylation is prohibited by enzyme's active site influences, the path progresses to **B1u** and **B1v**. From these two conformations, **B1u** has the lowest energy configuration, but it possesses a negative  $C_{18}$  Fukui index that does not allow further hydroxylation in the path. The  $C_{18}$  Fukui index of **B1v** does allow hydroxylation, although the indices of  $C_{11}$  and  $C_{19}$  are higher. If **B1v** is hydroxylated on  $C_{18}$ , the resulting conformations of 18OH-B possess a stabilised  $C_{11}$ -hydroxyl group through a hydrogen bonding interaction with the  $C_{18}$ -hydroxyl group (Table 6–2). This conformation possessed a more **u** conformation. A second  $C_{18}$ -hydroxylation is allowed for the two conformations **18OH-B1x** and **18OH-B1z** to form 18di-OH-B. The resulting conformations are **18diOH-B1xw** and **18diOH-B1yz** (Figure 6–10).

#### 6.4.4.2 Path 2

The second path is the conversion of **DOC2** via  $C_{11}$ -hydroxylation, followed by two consecutive  $C_{18}$ -hydroxylations to 18diOH-B. Investigating the Fukui indices of **DOC2**, there is again a preference of  $C_{19}$  over  $C_{11}$  and finally  $C_{18}$ . Even if we assume that  $C_{19}$ -hydroxylation is blocked and **B2u** or **B2v** is synthesised, the path stops. The  $C_{18}$  Fukui indices of the conformations are very close to 0 indicating that no further  $C_{18}$ -hydroxylation is likely. The path stops at the formation of B, which could be an explanation for the much higher concentrations of B compared to the other CYP11B1 and CYP11B2 products.

#### 6.4.4.3 Path 3

The third path is the conversion of **DOC3** through  $C_{11}$ -hydroxylation, followed by two consecutive  $C_{18}$ -hydroxylations to 18diOH-B. Here, the Fukui index of  $C_{11}$  is higher than the  $C_{19}$ , which in turn is nearly equal to that of the  $C_{18}$ . This is the only DOC conformation that possesses a Fukui index that clearly favours the formation of B. The energy for these complexes shows that **B3u** has the lowest energy configuration. Following the formation of **B3u** or **B3v**, the Fukui indices of these conformations of B both indicate another clear preference for further  $C_{18}$ -hydroxylation. From the 18OH-B conformations that can be synthesised, **18OH-B3x** and **18OH-B3z** (lowest energy) both prefer further hydroxylation on  $C_{18}$  to form **18diOH-B3zw** or **18dOH-B3zy** (Figure 6–10).

#### 6.4.4.4 Path 4

The fourth path is the conversion of **DOC1** via  $C_{18}$ -hydroxylation, followed by  $C_{11}$ -hydroxylation and  $C_{18}$ -hydroxylation. The Fukui index of  $C_{18}$  is high enough to allow conversion to 18OH-DOC. After formation of the **18OH-DOC1** conformations, only the **18OH-DOC1w** and **18OH-DOC1y** conformations possess a high enough Fukui index for further  $C_{11}$ -hydroxylation. However, the subsequent  $C_{18}$ -hydroxylation of **18OH-B1w** and **18OH-B1y** seems impossible, because the Fukui indices for  $C_{18}$  are too low; the path stops at the formation of 18OH-B. Because CYP11B1 can synthesise 18OH-B and not aldosterone, this path seems to explain the enzymatic activity of CYP11B1.

#### 6.4.4.5 Path 5

The fifth path is the conversion of **DOC2** through  $C_{18}$ -hydroxylation, followed by  $C_{11}$ -hydroxylation and  $C_{18}$ -hydroxylation. Again, the Fukui index of  $C_{18}$  is less than that of  $C_{11}$  and  $C_{19}$ , but high enough to allow conversion into 18OH-DOC. Here, the conformations of **18OH-DOC2w** and **18OH-DOC2x/2z** possess a favourable  $C_{11}$  Fukui index for further oxidation to **18OH-B2w** and **18OH-B2z**. From these two conformations, only **18OH-B2z** possesses a favourable enough  $C_{18}$  Fukui index to form **18diOH-B2zw** or **18diOH-B2zy** (Figure 6–10).

#### 6.4.4.6 Path 6

The sixth and last path is the conversion of **DOC3** via C<sub>18</sub>-hydroxylation, followed by C<sub>11</sub>-hydroxylation and C<sub>18</sub>-hydroxylation. Although **DOC3** possesses a strong preference for the formation of B (Path 3), the C<sub>18</sub> Fukui index is very high (higher than the C<sub>11</sub> indices for **DOC1** and **DOC2**) and is almost the same as the C<sub>19</sub> index. It is plausible that 18OH-DOC is formed from this DOC conformation. After this conversion, only **18OH-DOC3z** possesses a high C<sub>11</sub>-Fukui index that allows formation of **18OH-B3z**. As seen in Path 3, formation of this structure can result in the formation of **18diOH-B3zw** and **18diOH-B3zy** (Figure 6–10).

#### 6.4.5 Structural investigation of the hydroxylation paths

Although the Fukui index provides an indicative measure for the reactivity of the carbon atoms, it does not provide information on the spatial restrictions that may be present for each of the conversion steps. We have suggested in Chapter 2 that the  $\beta$ -side of the corticosteroids is oriented at an angle of around 45° to the heme [64]. This is the only conformation that presents all hydroxylation sites in close proximity to the heme iron (cut-off 5.0 Å, paragraph 2.7.3 ). Assuming that this orientation is correct, hydrogen atom extraction from C<sub>18</sub> is most plausible from the orientation types **x** and **z**. This has an implication for the paths discussed above.

Path 1 resulted in the conversion of **18OH-B1x** and **18OH-B1z** into either **18diOH-B1xw** or **18diOH-B1yz**. Both **18OH-B1x** and **18OH-B1z** possess an internal hydrogen bond between the C<sub>18</sub>-hydroxyl group and the C<sub>20</sub>-ketone group and are very closely related in structure (Figure 6–10). Because of this interaction, the '**w**' C<sub>18</sub>-hydrogen is in reality more an '**x**' C<sub>18</sub>-hydrogen and can be oxidised by the heme iron. The '**y**' C<sub>18</sub>-hydrogen is blocked from oxidation by the internal hydrogen bond. As a result, the conversion of **18OH-B1x** and **18OH-B1z** into **18diOH-B1xw** is spatially favourable, whereas the formation of **18diOH-B1yz** is not. Similar conclusions can be drawn for the results of Path 3, Path 5 and Path 6. For each of these paths, the oxidation of **18OH-B x** or **z** results in a spatially plausible C<sub>18</sub> hydrogen abstraction on the **w** side and not the **y** side (Figure 6–12).

#### 6.4.6 18OH-DOC as an in vitro inhibitor for CYP11B1 and CYP11B2

Investigation of the 18OH-DOC species may provide insight on the inhibitory action it possesses on both CYP11B1 and CYP11B2, and the substrate function it possesses on the synthesis of 18OH-B and aldosterone by CYP11B2. For the inhibitory action on CYP11B1, which is a poor catalyst for the formation of 18OH-B and aldosterone, we suggest that the binding mode of 18OH-DOC must feature a poor Fukui index for further C<sub>11</sub>-hydroxylation. The conformations of 18OH-DOC that match this query are **1x**, **1z**, **3w**, **3x**, and **3y**. The affinity of one of these conformations to the active site of CYP11B1 must be close to that of DOC itself if 18OH-DOC is to act as an inhibitor, but in this study we cannot estimate binding affinities to the active site.

To explain the activity of 18OH-DOC on CYP11B2, we suggest that the effect of 18OH-DOC on CYP11B2 is three-fold: (1) the binding affinity of 18OH-DOC to the active site is close to



that of DOC, (2) the binding orientation of 18OH-DOC features a C<sub>11</sub>-hydroxylation distance within 5.0 Å and (3), the conformation of 18OH-DOC in CYP11B2 can act as a substrate for the formation of 18OH-B and aldosterone, i.e. the C<sub>11</sub> Fukui index of the 18OH-DOC conformation is high. There are two conformations of 18OH-DOC that might match these features, namely **18OH-DOC2z** and **18OH-DOC3z**. The **18OH-DOC2z** and **18OH-DOC3z** conformations possess a high C<sub>11</sub> Fukui index and can be converted into 18OH-B (Path 5 and Path 6, respectively). At the same time, they can compete with **DOC2** and **DOC3** for the binding site. The **18OH-DOC2z** conformation seems most plausible for the inhibition of DOC, since **DOC2** does not seem to be a great substrate for the conversion into **B**. This implies that **18OH-DOC2z** may have a greater affinity for CYP11B2 than **DOC2** and may halt the synthesis of B in favour of the synthesis of 18OH-B and aldosterone.

#### 6.4.7 Study Limitations: Mechanistic Knowledge

The heme cofactor plays a prime role in the conversion of substrates and inclusion of the heme in the calculations can increase accuracy and confidence. However, since the substrates are converted differently by CYP11B1 and CYP11B2, it would still be impossible to derive a clear picture for the regio-specificity of the two isoforms. Therefore, inclusion of the heme in these calculations would not add more information to describe the differences in substrate conversion steps for aldosterone synthesis. For the inclusion of active site influences, the calculations can be expanded to QM/MM, which are time-consuming and require an accurate description of the protein active site. Unfortunately, this required detail is unattainable for homology models.

The calculation of the Fukui index is best used for synthesis of compounds that are not shielded from the environment, i.e. they do not take into account the stabilisation or hindrance by proteins. Therefore, interpretation of our results in comparison with observed substrate conversions by the cytochrome P450 enzymes is not that straightforward, as active site influences are important for the regio-specific hydroxylation of the substrates. In this respect, the high C<sub>19</sub> Fukui indices observed from the calculations may indicate that protein-ligand interactions play an important role in discriminating between the three different hydroxylation sites of DOC.

It is uncertain whether the substrates actually assume the conformations researched in our investigations. Moreover, the optimal energy conformation of the substrate is not necessarily the optimal conformation in the protein-substrate complex. To overcome these problems, it is necessary to generate multiple conformations of DOC to investigate the influence of the substrate conformation on the likelihood for oxidation.

The investigation of the possible sequential paths also depends on the assumption that the conversion of the substrates does not exchange the intermediate products with the environment. Although it has been determined that other cytochrome P450 enzymes work or likely work according to this theory [65-70], we are uncertain whether it is also applicable for the cytochrome P450 11B family. However, because the conversion pathway is very regio-specific, we believe that it is correct to assume that structures are largely restricted in their movement in the protein active site.

### 6.5 Transition state analogues of the 18-hydroxycorticosterone conversion

Using Fukui index calculations, we have been able to derive the most likely multistep conversion from DOC to aldosterone as performed by CYP11B2. The conformations that are most likely to be involved in the last conversion step are **18OH-B3z** and **18diOH-B3zw** (Path 3, paragraph 6.4.4.3 ). To obtain a basis for the design of transition state analogues, we have derived the transition state that occurs during the conversion of 18-hydroxycorticosterone to aldosterone for this particular conformation of the gemdiol.

Since quantum mechanics calculations are computationally costly, it is imperative that the model system is reduced sufficiently, yet still retains an accurate description of the enzymatic activity. The catalytic site of the enzyme can be simplified in many ways. In general, the system can be simplified at either the level of the heme or at the level of the substrate and the enzyme environment. The final model system used during our calculations of the transition state of 18-hydroxycorticosterone can be seen in (Figure 6–13).

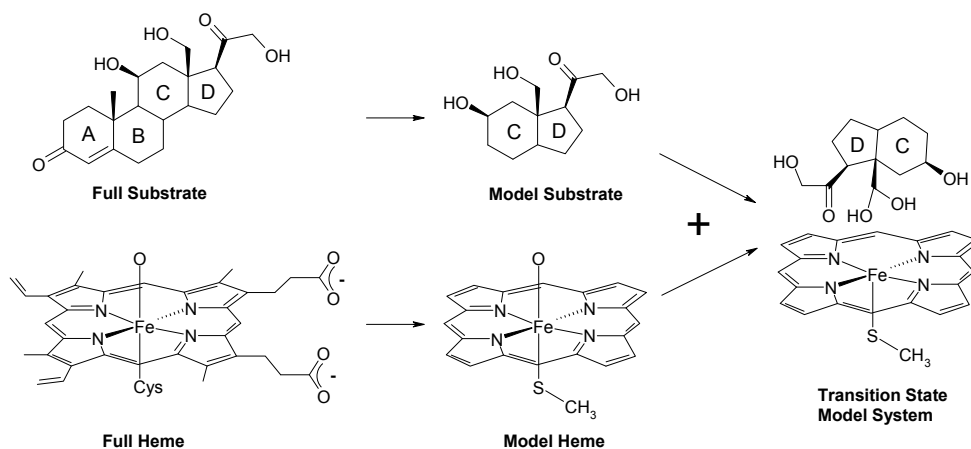


Figure 6–13 Model system for the Transition State that occurs during the conversion of 18-hydroxycorticosterone to aldosterone. left) original substrate and heme. centre) model systems for the substrate and heme. right) model system of the transition state.

#### 6.5.1 Reduction of Heme Complexity

The active species of the heme prosthetic group have been described in (paragraph 6.3 ). The simplest model that has been used as a representative for the modelling of the heme system in the **Cpd 0** state is an iron-oxo species. This model consists solely of an iron atom, an oxygen atom, and a substrate. Modelling the conversion of methane to methanol with this model system has clearly shown the two-state reaction of the oxygen rebound (paragraph 6.3.1 ) [76,77,78], however, the approach of the substrate to the iron-oxo species is not plausible. If the atomic positions of this model system are superposed with the full heme system, the position of the substrate is on the wrong side of the heme (Figure 6–14a,d).

A slightly more complete model of the heme includes the sulphur and nitrogen atoms stabilised by hydrogen atoms as  $[\text{FeO}(\text{SH})-(\text{NH}_3)_4]^+$  (Figure 6–14b) [79]. This model is claimed to include all the features that are essential for the accurate modelling of the hydrogen abstraction steps; the oxyferryl species, the thiolate group, and the octahedral coordination of the iron centred between four nitrogen donors. This model supports the two-state rebound mechanism and the recorded energy barriers of the rebound mechanism are close to computations performed with the full heme system [79].

Higher precision can be obtained by including the exact ring systems of the heme group and adjustments to the thiolate group [43,50,52,80,81]. Generally, these models truncate the heme to porphine by removing all side chains from the porphyrin group. A less severe truncation can be performed by using octamethyl porphine where the side chains are exchanged by methyl groups [82]. The cysteine residue can be modelled as a thiolate ( $\text{SH}^-$ ), methyl mercaptide ( $\text{SCH}_3^-$ ), cysteine anion ( $\text{CysS}^-$ ), or expanded with NH groups to mimic the hydrogen bonding network of the cysteine (Figure 6–14c) [83]. The current consensus for modelling the heme with quantum mechanics calculations is to use at least the porphine group in combination with the thiolate [83,84].

To keep the computational effort within bounds, we have chosen to use the unsubstituted porphine group in combination with a methyl mercaptide group for the accurate modelling of the heme system in this thesis. The structure of this complex is shown in (Figure 6–14c), with hydrogen atoms as Y-substituents, and a methyl group as X-substituent (see also Figure 6–13, centre).

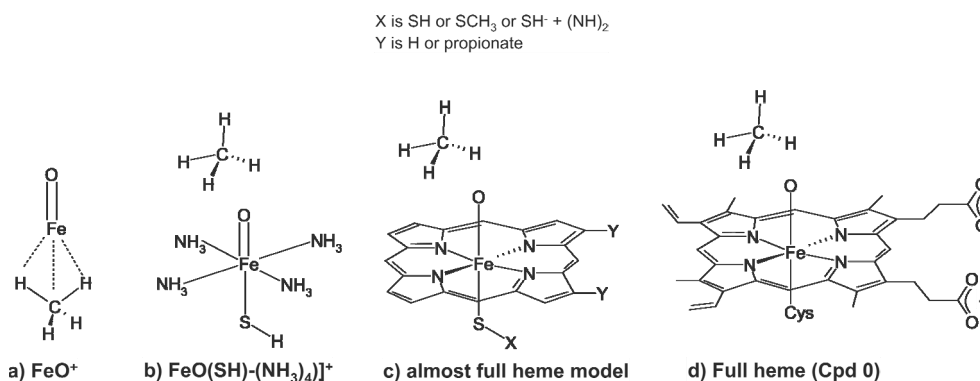


Figure 6–14 Comparison of the different heme model systems compared to the actual system.

### 6.5.2 Reduction of Substrate Complexity

In early quantum mechanics studies on the catalytic activity of monooxygenases, the calculations have been performed using methane as the substrate [47,76,85,86]. The results from the calculations on methane have been expanded to conclude general features of the catalytic activity for larger molecules, in particular alkyl moieties. Since the general features of the conversion of methane have been clarified, the complexity of the model systems has been expanded to involve more and more detail. In recent studies, larger substrates have

been used [3,43,79,50,52,87] as well as additional active site properties such as water molecules or amino acid side chains [79,88,89,90]. These implementations have provided insight on the stabilisation of **Cpd 0** and **Cpd I**, the stability of the substrate in the active site, and the delivery of protons to the active site through water channels.

In this thesis, the steroidal substrates are much larger than the substrates that have been computed in current literature. To reduce computation cost, the substrate size can be reduced by removing regions of the substrate that play no apparent role during the conversion steps. Hackett et al. have shown that reduction of androgens by neglecting the contribution of the steroid C and D rings can be used to explain their aromatisation by CYP19 [43] (for steroid enumeration see Figure 1–3). Using a comparable approach, the complexity of 18-hydroxycorticosterone has been reduced by neglecting the A and B-rings of the steroid skeleton (Figure 6–13, centre).

To obtain an initial pose for the reduced 18-hydroxycorticosterone structure during hydrogen abstraction, we have first explored transition state 1 by using ethanol as a substrate (Figure 6–15). Next, the reduced 18-hydroxycorticosterone structure has been superimposed on the heme-ethanol complex. Subsequently, the conformations of the substrate-bound reaction complex (**1**) and product-bound complex (**13**) have been optimised, and the hydrogen abstraction pathway has been investigated using single point calculations (Figure 6–16).

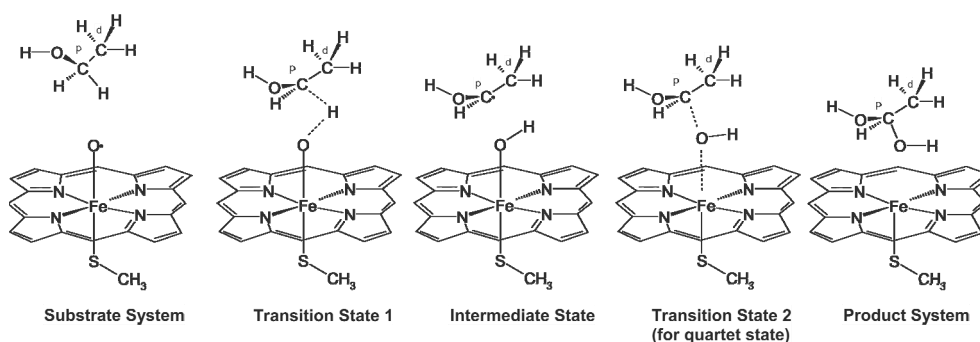


Figure 6–15 Schematic for the structures of the reaction complex during the conversion of ethanol to ethanediol. Transition state 2 only occurs in the quartet state, similar to (Figure 6–2b and Figure 6–4)

## 6.6 Results and Discussion: Transition state analogues

The rate-determining step of the reaction is the hydrogen abstraction of transition state 1, which results in the intermediate state where the alkyl-radical is in complex with the hydroxylated heme. During this process, the characteristic of the proximal carbon atom changes from an  $sp^3$  hybridised conformation to an  $sp^2$  hybridised conformation (Figure 6–15, Figure 6–16 conf 6). Subsequently, the reaction of the intermediate state cation (Figure 6–16 conf 9) is swift and smooth for the formation of the product (oxygen rebound, see also Figure 6–4).

Our calculations on the substrate conversion of the 18-hydroxycorticosterone (reduced form) are slightly different than the results published by other groups for the substrate camphor [79,89]. We have calculated a relative energy barrier for Transition State 1 as 38.6 kcal/mol,

whereas the barrier for camphor has been determined to be as low as 20-25 kcal/mol. The intermediate state has been computed 5.7 kcal/mol higher than the substrate-bound reaction complex, whereas the intermediate state of camphor is 15 kcal/mol. We have not detected any Transition State for the rebound reaction, and the stable gemdiol product of the reduced 18-hydroxycorticosterone structure has a low energy minimum as -75.4 kcal/mol, whereas the product of camphor has been determined in the range of -50 kcal/mol. These energetic differences are caused by our method of calculation, as we have derived the complexes using single point calculations, which do not optimise the geometry. Nevertheless, the calculated conformation of the complexes during this reaction state provides information for the design of transition state analogues.

The exact positioning of the substrate above the heme is an important feature that determines the selectivity of substrate conversion by CYP11B1 and CYP11B2 (paragraph 2.7.5 ). It is anticipated that this selectivity between CYP11B1 and CYP11B2 cannot be distinguished when only the heme-substrate complex is computed with quantum mechanics calculations, rather, for this to be achieved, amino acids in close proximity to the substrate must be also included in the calculations. It is possible to implement these active site properties to the calculations, however, since the homology models have been based on cytochrome structures that are a maximum of 30% identical (paragraph 2.4.2 ), it is likely that the required detail for the determination of substrate selectivity can never be reached. Resolving the substrate selectivity of CYP11B1 and CYP11B2 may prove to be fruitful if a combined method such as QMMM (Quantum Mechanics/Molecular Mechanics) is used.

Fortunately, our main research interest in quantum mechanics calculations on the synthesis of aldosterone lies only in the uncovering of the transition state and constructing analogues of the substrate conformation in this particular state. The other steps of substrate need not be resolved in order to design the CYP11B2 selective analogues, and the exact active site differences near the substrate need not be taken into account. Since we are only interested in the transition state of the complex, it is a reasonable assumption that the other protein-substrate states do not need to be computed.

Several analogues have been identified based on their steric and electrostatic overlap with the reduced form of 18-hydroxycorticosterone (Figure 6–17). Because the exact overlap of the steroidal rings is difficult to obtain, the steric overlap of most suggested structures is limited to a spherical shape featuring a possible interaction site with the heme iron. Compounds (1) and (6) may function as a good basis for further inhibitor development, since compound (1) has been used as a substructure for drugs, and compound (6) is related to oxazoles and imidazoles, that are also used as substructures of drugs. The chemical nature of the iron-ligating nitrogen of compounds (2) and (5) is doubtful as they may prefer a protonated form. Likewise, compound (4) is not stable and will be subject to ring opening. Lastly, the camphor-like compound (3) provides the best steric overlap and a good electrostatic overlap, and although it is theoretically possible, the acidic group does not interact with a heme iron.

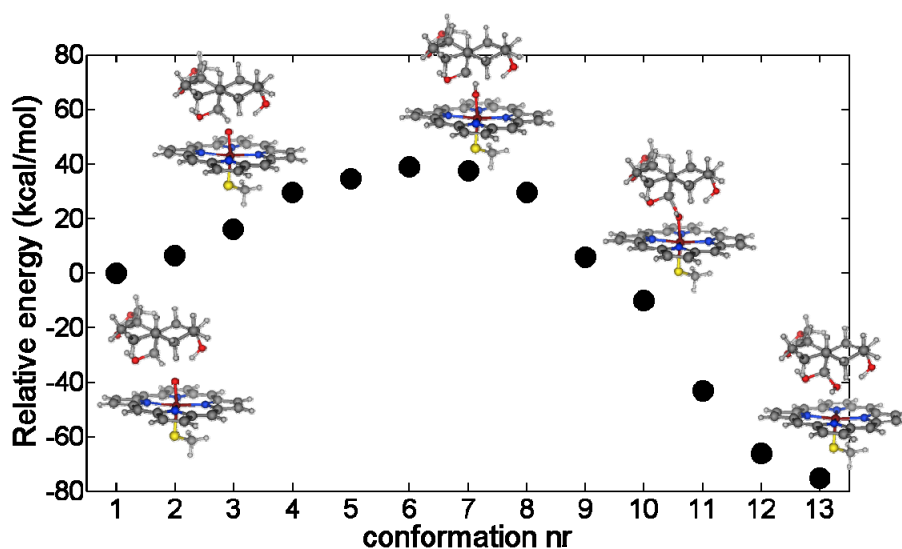


Figure 6-16 Reaction pathway for the conversion of the reduced 18-hydroxycorticosterone to the reduced 18-dihydroxycorticosterone for the doublet state. For the doublet state, no second Transition State is observed during product formation.

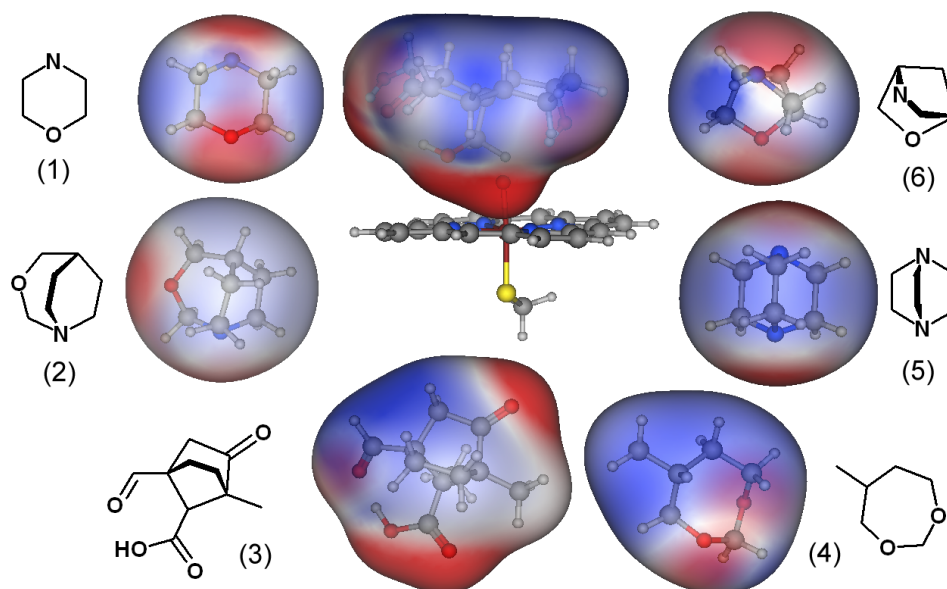


Figure 6-17 Electrostatic and steric properties of the Transition State (centre top) and several analogues. Negatively charged regions are coloured red, whereas positively charged regions are coloured blue.

## 6.7 Conclusions: Transition State Analogues and Mechanistic Knowledge

Using quantum mechanical calculations, we have attempted to elucidate the substrate conversion mechanism performed by CYP11B1 and CYP11B2. To accomplish this we have performed both a conformation analysis as well as an analysis towards hydroxylation site reactivity in accordance with the Fukui index. For each of the possible substrate conversion paths, we have calculated the carbon atom reactivity and thus an indicative measurement for the likelihood of substrate conversion.

We find that the determination of the Fukui index for these structures is justifiable for the gathering of mechanistic knowledge on the conversion of 11-deoxycorticosterone (DOC) to aldosterone, as we have derived plausible multistep conversion paths for different initial substrate conformations. In addition, we have been able to discriminate between paths that are likely taken by either CYP11B1 or CYP11B2. **Path 2** ends in the synthesis of corticosterone (B) which may explain its large concentration measured in *in vitro* experiments [59,60,61]. **Path 4** ends with the synthesis of 18-hydroxycorticosterone (18OH-B) and may be representative for the substrate conversion by CYP11B1. The other **paths** end with the successful synthesis of aldosterone and may explain the substrate conversion mechanism of CYP11B2.

Based on the Fukui index, the most plausible path for aldosterone synthesis is **path 3**, which features the initial conformation **DOC3**. This path clearly possesses favourable Fukui indices for each subsequent hydroxylation step resulting in the synthesis of 18-dihydroxycorticosterone (18diOH-B). In the modelling study of Chapter 2, we have defined the conformations involved in this path to be determinant for the synthesis of aldosterone [64]. Indeed, the results from our quantum mechanics study can confirm and emphasise our homology modelling findings.

We expanded the investigations on the hydroxylation paths to uncover the cause for the inhibitory effect of 18OH-DOC on the conversion of DOC by CYP11B1 and CYP11B2, and the increased synthesis of aldosterone by CYP11B2. We propose that the inhibitory effect on CYP11B1 can be attributed to a binding mode of 18OH-DOC creating an unfavourable C<sub>11</sub> Fukui index, thereby prohibiting further oxidation. The conformation of 18OH-DOC which increases the synthesis of aldosterone by CYP11B2 may itself be converted into 18OH-B and aldosterone. This suggests that the involvement of active site conformation plays a large role in substrate stabilisation and the regio-specific substrate conversion performed by both members of the CYP11B family.

In a next step, the conformations that are most likely involved in the last conversion step of aldosterone synthesis (**path 3**) have been used to determine the transition state of the conversion of 18OH-B. This investigation has been completed by calculating the reaction path of the reduced structure of 18OH-B to 18diOH-B. Because we only determined the doublet state rather than the quartet and sextet states, only the first (desired) transition state was observed, and not the oxygen rebound. Although the magnitude of the potential energy diagram of our reaction path is high, the generic structure of the gemdiol transition state has been identified and can be used for the construction of highly CYP11B2-selective substrate

analogues. Whether the exact sp<sup>2</sup> conformation of the substrate in this hydrogen abstraction transition state can be accurately mimicked remains to be tested.



## Literature

- 1 K. Fukui, "Role of frontier orbitals in chemical reactions", *Science*, 1982, 218, 19, 747-754
- 2 R.G. Parr, W. Yang, "Density functional approach to the frontier-electron theory of chemical reactivity", *J. Am. Chem. Soc.*, 1984, 106, 15, 4049-4050
- 3 M.E. Beck, "Do Fukui function maxima relate to sites of metabolism? a critical case study", *J. Chem. Inf. Model.*, 2005, 45, 2, 273-282
- 4 E. Schrödinger, "An undulatory theory of the mechanics of atoms and molecules", *The Physical Review*, 1926, 28, 6, 1049-1070
- 5 E. Schrödinger, "Über das Verhältnis der Heisenberg-Born-Jordanschen Quantenmechanik zu der meinen", *Annalen der Physik*, 1926, 384, 8, 734-756
- 6 J.C. Slater, "A simplification of the Hartree-Fock method", *Phys. Rev.* 1951, 81, 3, 385-390
- 7 W. Kohn, L.J. Sham, "Self-consistent equations including exchange and correlation effects", *Phys. Rev.* 1965, 140, 4A, 1133-1138
- 8 P. Hohenberg, W. Kohn, "Inhomogeneous electron gas", *Phys. Rev.* 1964, 136, 3B, 864-871
- 9 A.D. Becke, "Correlation energy of an inhomogeneous electron gas: a coordinate-space model", *J. Chem. Phys.*, 1988, 88, 2, 1053-1062
- 10 A.D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behaviour", *Phys. Rev. A.*, 1988, 38, 6, 3098-3100
- 11 A.D. Becke, "Density-functional thermochemistry. I. The effect of exchange-only gradient correction", *J. Chem. Phys.*, 1992, 96, 3, 2155-2160
- 12 A.D. Becke, "Density-functional thermochemistry. II. The effect of the Perdew-Wang generalized-gradient correlation correction", *J. Chem. Phys.*, 1992, 97, 12, 9173-9177
- 13 A.D. Becke, "Density-functional thermochemistry. III. The role of exact exchange", *J. Chem. Phys.*, 1993, 98, 7, 5648-5652
- 14 C. Lee, W. Yang, R.G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density", *Physical Review B*, 1988, 37, 2, 785-789
- 15 B. Miehlich, A. Savin, H. Stoll, H. Preuss, "Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr", *Chemical Physics Letters*, 1989, 157, 3, 200-206
- 16 J.C. Slater, "Atomic shielding constants", *Phys. Rev.*, 1930, 36, 1, 57-64
- 17 S.F. Boys, "Electronic wave functions. I. a general method of calculation of the stationary states of any molecular system", *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 1950, 200, 1063, 542-554
- 18 M.J. Frisch, J.A. Pople, J.S. Binkley, "Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets", *J. Chem. Phys.*, 1984, 80, 7, 3265-3269
- 19 T. Clark, J. Chandrasekhar, G.W. Spitznagel, P. von Rague Schleyer, "Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+g basis set for first-row elements Li-F", *J. Comp. Chem.*, 1983, 4, 3, 294-301
- 20 R. Ditchfield, W.J. Hehre, J.A. Pople, "Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules", *J. Chem. Phys.* 1971, 54, 2, 724-728
- 21 W.J. Hehre, R. Ditchfield, J.A. Pople, "Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules", *J. Chem. Phys.*, 1972, 56, 5, 2257-2261
- 22 P.C. Hariharan, J.A. Pople, "The influence of polarization functions on molecular orbital hydrogenation energies", *Theo. Chim. Acta*, 1973, 28, 3, 213-222
- 23 P.C. Hariharan, J.A. Pople, "Accuracy of AHn equilibrium geometries by single determinant molecular orbital theory", *Mol. Phys.*, 1974, 27, 1, 209-214
- 24 M.S. Gordon, "The isomers of silacyclopropane", *Chem. Phys. Lett.*, 1980, 76, 1, 163-168
- 25 J.-P. Blaudeau, M.P. McGrath, L.A. Curtiss, L. Radom, "Extension of Gaussian-2 (G2) theory to molecules containing third-row atoms K and Ca", *J. Chem. Phys.*, 1997, 107, 13, 5016-5021
- 26 M.M. Francl, W.J. Pietro, W.J. Hehre, J.S. Binkley, D.J. DeFrees, J.A. Pople, M.S. Gordon, "Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements", *J. Chem. Phys.*, 1982, 77, 7, 3654-3665
- 27 R.C. Binning Jr., L.A. Curtiss, "Compact contracted basis sets for third-row atoms: Ga-Kr", *J. Comp. Chem.*, 1990, 11, 10, 1206-1216

- 28 V.A. Rassolov, J.A. Pople, M.A. Ratner, T.L. Windus, "6-31G\* basis set for atoms K through Zn", *J. Chem. Phys.*, 1998, 109, 4, 1223-1229
- 29 V.A. Rassolov, M.A. Ratner, J.A. Pople, P.C. Redfern, L.A. Curtiss, "6-31G\* basis set for third-row atoms", *J. Comp. Chem.*, 2001, 22, 9, 976-984
- 30 P.J. Hay, W.R. Wadt, "Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg", *J. Chem. Phys.*, 1985, 82, 1, 270-283
- 31 H.E. Conrad, K. Lieb, I.C. Gunsalus, "Mixed function oxidation. III. an electron transport complex in camphor ketolactonization", *J. Biol. Chem.*, 1965, 240, 10, 4029-4037
- 32 J. Hedegaard, I.C. Gunsalus, "Mixed function oxidation. IV. an induced methylene hydroxylase in camphor oxidation", *J. Biol. Chem.*, 1965, 240, 10, 4038-4043
- 33 M. Katagiri, B.N. Ganguli, I.C. Gunsalus, "A soluble cytochrome P-450 functional in methylene hydroxylation", *J. Biol. Chem.*, 1968, 243, 3543-3546
- 34 S.G. Sligar, "Coupling of spin, substrate, and redox equilibriums in cytochrome P450", *Biochemistry*, 1976, 15, 24, 5399-5406
- 35 F.P. Guengerich, W.W. Johnson, "Kinetics of ferric cytochrome P450 reduction by NADPH-cytochrome p450 reductase: rapid reduction in the absence of substrate and variation among cytochrome P450 systems", *Biochemistry*, 1997, 36, 48, 14741-14750
- 36 J.T. Groves, G.A. McClusky, "Aliphatic hydroxylation via oxygen rebound. oxygen transfer catalyzed by iron", *J. Am. Chem. Soc.*, 1976, 98, 3, 859-861
- 37 I.G. Denisov, T.M. Makris, S.G. Sligar, I. Schlichting, "Structure and chemistry of cytochrome P450", *Chem. Rev.*, 2005, 105, 6, 2253-2277
- 38 G.L. Fox, H.B. Schlegel, "An ab initio study of hydrogen atom abstractions from substituted methanes by substituted methyl radicals", *J. Phys. Chem.*, 1992, 96, 1, 298-302
- 39 H. Tachikawa, N. Hokari, H. Yoshida, "An ab initio MO study on hydrogen abstraction from methanol by methyl radical", *J. Phys. Chem.*, 1993, 97, 39, 10035-10041
- 40 M. Newcomb, M.-H. Le Tadic-Biadatti, D.L. Chestney, E.S. Roberts, P.F. Hollenberg, "A nonsynchronous concerted mechanism for cytochrome P-450 catalyzed hydroxylation", *J. Am. Chem. Soc.*, 1995, 117, 49, 12085-12091
- 41 B.K. Carpenter, "Dynamic matching: the cause of inversion of configuration in the [1,3] sigmatropic migration?", *J. Am. Chem. Soc.*, 1995, 117, 23, 6336-6344
- 42 M.H. Gelb, D.C. Heimbrook, P. Malkonen, S.G. Sligar, "Stereochemistry and deuterium isotope effects in camphor hydroxylation by the cytochrome P450cam monooxygenase system", *Biochemistry*, 1982, 21, 2, 370-377
- 43 J.C. Hackett, R.W. Brueggemeier, C.M. Hadad, "The final catalytic step of cytochrome P450 aromatase: a density functional theory study", *J Am Chem Soc*, 2005, 127, 14, 5224-5237
- 44 M. Newcomb, P.H. Toy, "Hypersensitive probes and the mechanisms of cytochrome P450-catalyzed hydroxylation reactions", *Acc. Chem. Rev.*, 2000, 33, 7, 449-455
- 45 M. Newcomb, R. Shen, S.-Y. Choi, P.H. Toy, P.F. Hollenberg, A.D.N. Vaz, M.J. Coon, "Cytochrome P450-catalyzed hydroxylation of mechanistic probes that distinguish between radicals and cations. evidence for cationic but not for radical intermediates", *J. Am. Chem. Soc.*, 2000, 122, 12, 2677-2686
- 46 A. Wilks, P.R. Ortiz de Montellano, "Rat liver heme oxygenase. high level expression of a truncated soluble form and nature of the meso-hydroxylating species", *J. Biol. Chem.*, 1993, 268, 30, 22357-22362
- 47 F. Ogliaro, N. Harris, S. Cohen, M. Filatov, S.P. de Visser, S. Shaik, "A model 'rebound' mechanism of hydroxylation by cytochrome P450: stepwise and effectively concerted pathways, and their reactivity patterns", *J. Am. Chem. Soc.*, 2000, 122, 37, 8977-8989
- 48 S.P. de Visser, F. Ogliaro, N. Harris, S. Shaik, "Multi-state epoxidation of ethene by cytochrome P450: a quantum chemical study", *J. Am. Chem. Soc.*, 2001, 123, 13, 3037-3047
- 49 S. Shaik, S.P. de Visser, F. Ogliaro, H. Schwarz, D. Schroder, "Two-state reactivity mechanisms of hydroxylation and epoxidation by cytochrome P-450 revealed by theory", *Curr. Opin. Chem. Biol.*, 2002, 6, 5, 556-567
- 50 D. Kumar, S.P. de Visser, K. Sharma, S. Cohen, S. Shaik, "Radical clock substrates, their C-H hydroxylation mechanism by cytochrome P450, and other reactivity patterns: what does theory reveal about the clock's behaviour?", *J. Am. Chem. Soc.*, 2004, 126, 6, 1907-1920

- 51 S. Shaik, S.P. de Visser, D. Kumar, "One oxidant, many pathways: a theoretical perspective of monooxygenation mechanisms by cytochrome P450 enzymes", *J. Biol. Inorg. Chem.*, 2004, 9, 6, 661-668
- 52 S.P. de Visser, D. Kumar, S. Shaik, "How do aldehyde side products occur during alkene epoxidation by cytochrome P450? theory reveals a state-specific multi-state scenario where the high-spin component leads to all side products", *J. of Inorganic Biochemistry*, 2004, 98, 7, 1183-1193
- 53 A.W. Munro, H.M. Girvan, K.J. McLean, "Variations on a (t)heme-novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily", *Nat. Prod. Rep.*, 2007, 24, 3, 585-609
- 54 R. Raag, T.L. Poulos, "Crystal structures of cytochrome P-450cam complexed with camphane, thiocamphor, and adamantane: factors controlling P-450 substrate hydroxylation", *Biochemistry*, 1991, 30, 10, 2674-2684
- 55 T.L. Poulos, R. Raag, "Cytochrome P450cam: crystallography, oxygen activation, and electron transfer", *the FASEB journal*, 1992, 6, 2, 674-679
- 56 W.M. Atkins, S.G. Sligar, "Metabolic switching in cytochrome P-450cam: deuterium isotope effects on regioselectivity and the monooxygenase/oxidase ration", *J. Am. Chem. Soc.*, 1987, 109, 12, 3754-3760
- 57 W.M. Atkins, S.G. Sligar, "The roles of active site hydrogen bonding in cytochrome P-450cam as revealed by site-directed mutagenesis", *J. Biol. Chem.*, 1988, 263, 35, 18842-18849
- 58 W.M. Atkins, S.G. Sligar, "Molecular recognition in cytochrome P-450: alteration of regioselective alkane hydroxylation via protein engineering", *J. Am. Chem. Soc.*, 1989, 111, 7, 2715-2717
- 59 A. Fisher, E.C. Friel, R. Bernhardt, C. Gomez-Sanchez, C. Connell, J.M.C. Fraser, E. Davies, "Effects of 18-hydroxylated steroids on corticosteroid production by human aldosterone synthase and 11 $\beta$ -hydroxylase", *J Clin Endocrinol Metab*, 2001, 86, 9, 4326-4329
- 60 Y. Nonaka, M. Okamoto, "Functional expression of the cDNAs encoding rat 11 $\beta$ -hydroxylase [cytochrome P450(11 $\beta$ )] and aldosterone synthase [cytochrome P450(11 $\beta$ , Aldo)]", *Eur J Biochem*, 1991, 202, 3, 897-902
- 61 M. Lauber, J. Muller, "Purification and characterization of two distinct forms of rat adrenal cytochrome P450<sub>11 $\beta$</sub> : functional and structural aspects", *Archives of Biochemistry and Biophysics*, 1989, 274, 1, 109-119
- 62 S. Takemori, S. Kominami, "The role of cytochromes P-450 in adrenal steroidogenesis", *Trends In Biochemical Sciences*, 1984, 9, 9, 393-396
- 63 J.O. Johnston, C.L. Wright, G.W. Holbert, "Enzyme-activated inhibitors of steroidal hydroxylases", *J. Steroid Biochem. Molec. Biol.*, 1995, 52, 1, 17-34
- 64 L. Roumen, M.P.A. Sanders, K. Pieterse, P.A.J. Hilbers, R. Plate, E. Custers, M. de Gooyer, J.F.M. Smits, I. Beugels, J. Emmen, H.C.J. Ottenheijm, D. Leysen, J.J.R. Hermans, "Construction of 3D models of the CYP11B family as a tool to predict ligand binding characteristics", *J Comput-Aided Mol Des*, 2007, 21, 8, 455-471
- 65 F.P. Guengerich, "Oxidation cleavage of carboxylic esters by cytochrome P-450", *The Journal of Biological Chemistry*, 1987, 262, 18, 8459-8462
- 66 L.C. Bell, F.P. Guengerich, "Oxidation kinetics of ethanol by human cytochrome P450 2E1", *The Journal of Biological Chemistry*, 1997, 272, 47, 29643-29651
- 67 L.C. Bell-Parikh, F.P. Guengerich, "Kinetics of cytochrome P450 2E1-catalyzed oxidation of ethanol to acetic acid via acetaldehyde", *The Journal of Biological Chemistry*, 1999, 274, 34, 23833-23840
- 68 S. Miyairi, J. Fishman, "Radiometric analysis of oxidative reactions in aromatization by placental microsomes", *The Journal of Biological Chemistry*, 1985, 260, 1, 320-325
- 69 Y. Watanabe, Y. Ishimura, "Aromatization of tetralone derivatives by Fe<sup>III</sup>PFP(Cl)/PhIO and cytochrome P450<sub>cam</sub>: a model study on aromatase cytochrome P-450 reaction", *J Am Chem Soc*, 1989, 111, 1, 410-411
- 70 J.J. Cali, D.W. Russell, "Characterization of human sterol 27-hydroxylase", *The Journal of Biological Chemistry*, 1991, 266, 12, 7774-7778

- 71 T. Yamazaki, T. Ohno, T. Sakaki, M. Akiyoshi-Shibata, Y. Yabusaki, T. Imai, S. Kominami, "Kinetic analysis of successive reactions catalyzed by bovine cytochrome P45017 $\alpha$ lyase", *Biochemistry*, 1998, 37, 9, 2800-2806
- 72 M. Akiyoshi-Shibata, T. Sakaki, Y. Ohyama, M. Noshiro, K. Okuda, Y. Yabusaki, "Further oxidation of hydroxycalcitol 24-hydroxylase. a study with the mature enzyme expressed in *Escherichia coli*", *Eur. J. Biochem.*, 1994, 224, 2, 335-343
- 73 MOE (The Molecular Operating Environment) Version 2007.08, software available from Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Canada H3A 2R7. <http://www.chemcomp.com>
- 74 G. Schaftenaar, J.H. Noordik, "Molden: a pre- and post-processing program for molecular and electronic structures", *J. comput.-Aided. Mol. Des.*, 2000, 14, 2, 123-134
- 75 Gaussian 03, Revision C.02, M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, and J.A. Pople, Gaussian, Inc., Wallingford CT, 2004.
- 76 K. Yoshizawa, Y. Shiota, T. Yamabe, "Methane-methanol conversion by MnO<sup>+</sup>, FeO<sup>+</sup>, and CoO<sup>+</sup>: a theoretical study of catalytic selectivity", *J. Am. Chem. Soc.*, 1998, 120, 3, 564-572
- 77 K. Yoshizawa, Y. Shiota, T. Yamabe, "Abstraction of the hydrogen atom of methane by iron-oxo species: the concerted reaction path is energetically more favorable", *Organometallics*, 1998, 17, 13, 2825-2831
- 78 K. Yoshizawa, Y. Shiota, T. Yamabe, "Intrinsic reaction coordinate analysis of the conversion of methane to methanol by an iron-oxo species: a study of crossing seams of potential energy surfaces", *J. Chem. Phys.*, 1999, 111, 2, 538-545
- 79 J.C. Schoneboom, S. Cohen, H. Lin, S. Shaik, W. Thiel, "Quantum mechanical/molecular mechanical investigation of the mechanism of C-H hydroxylation of camphor by cytochrome P450cam: theory supports a two-state rebound mechanism", *J. Am. Chem. Soc.*, 2004, 126, 12, 4017-4034
- 80 M.J. de Groot, R.W.A. Havenith, H.M. Vinkers, R. Zwaans, N.P.E. Vermeulen, J.H. Lenthe, "Ab initio calculations on iron-porphyrin model systems for intermediates in the oxidative cycle of cytochrome P450s", *J. Comput.-Aided. Mol. Des.*, 1998, 12, 2, 183-193
- 81 P.K. Sharma, S.P. de Visser, F. Ogliaro, S. Shaik, "Is the ruthenium analogue of compound I of cytochrome P450 an efficient oxidant? a theoretical investigation of the methane hydroxylation reaction", *J. Am. Chem. Soc.*, 2003, 125, 10, 2291-2300
- 82 F. Ogliaro, S.P. de Visser, S. Cohen, J. Kaneti, S. Shaik, "The experimentally elusive oxidant of cytochrome P450: a theoretical 'trapping' defining more closely the 'real' species", *Chem. Bio. Chem.*, 2001, 2, 11, 848-851
- 83 S. Shaik, D. Kumar, S.P. de Visser, A. Altun, W. Thiel, "Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes", *Chem. Rev.*, 2005, 105, 6, 2279-2328
- 84 I.G. Denisov, T.M. Makris, S.G. Sligar, I. Schlichting, "Structure and chemistry of cytochrome P450", *Chem. Rev.*, 2005, 105, 6, 2253-2277
- 85 F. Ogliaro, M. Filatov, S. Shaik, "Alkane hydroxylation by cytochrome P450: is kinetic isotope effect a reliable probe of transition state structure?", *Eur. J. Inorg. Chem.*, 2000, 2000, 12, 2455-2458
- 86 N. Harris, S. Cohen, M. Filatov, F. Ogliaro, S. Shaik, "Two-state reactivity in the rebound step of alkane hydroxylation by cytochrome P-450: origins of free radicals with finite lifetimes", *Angew. Chem. Int. Ed.*, 2000, 39, 11, 2003-2007

- 87 V. Guallar, M. Baik, S.J. Lippard, R.A. Friesner, "Peripheral heme substituents control the hydrogen-atom abstraction chemistry in cytochromes P450", *Proc. Natl. Acad. Sci. USA*, 2003, 100, 12, 6998-7002
- 88 V. Guallar, R.A. Friesner, "Cytochrome P450cam enzymatic catalysis cycle: a quantum mechanics/molecular mechanics study", *J. Am. Chem. Soc.*, 2004, 126, 27, 8501-8508
- 89 T. Kamachi, K. Yoshizawa, "A theoretical study on the mechanism of camphor hydroxylation by compound I of cytochrome P450", *J. Am. Chem. Soc.*, 2003, 125, 15, 4652-4661
- 90 V. Guallar, D.L. Harris, V.S. Batista, W.H. Miller, "Proton-transfer dynamics in the activation of cytochrome P450eryF", *J. Am. Chem. Soc.*, 2002, 124, 7, 1430-1437

## Chapter 7 Concluding Remarks

The molecular modelling work presented in this thesis has been guided by synthesis carried out by SyMO-Chem NV and by Organon NV, as well as *in vitro* measurements carried out at the University of Maastricht. The close interaction between the three disciplines has led to feedback for compound synthesis based on molecular modelling insights, and feedback for molecular modelling based on *in vitro* measurement data. An important aspect of the molecular modelling work comprised the elucidation of inhibitor interactions followed by the prioritisation of subsequent compound synthesis. Thus, molecular modelling has provided an integral link between compound synthesis and *in vitro* screening.

Using molecular modelling we have constructed protein models for CYP11B1 and CYP11B2. Their active site conformations have provided insights on the binding modes of both substrates and inhibitors, and have allowed identification of the protein-ligand interactions most important for binding affinity. Various computational approaches have been explored for the prediction of ligand binding affinity as well as the design of novel CYP11B2 selective inhibitors, and have opened up new areas for future investigations. The most important conclusions that may stimulate further research are summarised below.

Cytochrome P450 enzymes all share the same general protein fold. Therefore, it is difficult to construct homology models that can correctly predict the selectivity of ligands towards different cytochromes. The structural resemblance of CYP11B1 and CYP11B2 is especially high featuring a sequence identity of more than 90%. Our homology modelling work suggests that even though the structural overlap between proteins is high, it is still possible to identify the structural features that determine the conformational variations between the proteins. During the construction of the homology models there were no amino acid differences detected between the active sites of the two isoforms. Instead, the structural active site differences are caused by amino acids lining the active site. Although the modelling has been performed in detail for the active site regions as well as the structural core of the proteins, the amino acid replacement method of homology modelling is not perfect. Homology models often contain amino acid chains possessing unfavourable chain rotations and it is possible that some conformational differences between our CYP11B1 and CYP11B2 models are caused by such disparities. Therefore, it is uncertain whether the conformational differences found are indeed related to the amino acid differences lining the active site or to conformational inaccuracies of other amino acids in close proximity to the active site. Ideally, a CYP11B2 crystal structure should be resolved to validate our model results. However, due to the hydrophobic character of the membrane-bound region of the cytochrome P450 enzymes, these crystal structures are difficult to obtain. In addition, the available mammalian crystal structures have been modified extensively to solve insolubility problems, hence these modifications may affect the protein fold (such as for CYP2C9 and CYP2D6). Despite the modelling limitations, the constructed homology models can be used as a tool to specify protein-ligand interactions. For better elucidation of active site differences between CYP11B1 and CYP11B2, we recommend a method such as site-

directed mutagenesis to investigate which amino acids play an important role in either (1) substrate recognition, (2) protein conformation stabilisation, (3) control of the active site cavity size, and (4) active site entrance, which can be substantiated by *in vitro* measurements such as performed by Bernhardt et al (1996).

To investigate their dynamic behaviour, the constructed homology models of CYP11B1 and CYP11B2 have been subjected to a study which combines molecular dynamics with molecular docking. The combination of these disciplines has been applied to rationalise experimental data from *in vitro* measurements. The experimental data is an ensemble average over many different protein and ligand conformations as well as orientations, and therefore, the homology models need to be assessed to determine whether they provide enough information for the prediction of protein-ligand interactions. One of the challenges involved with molecular docking and molecular dynamics is using the correct algorithms and parameter sets. Although many such combinations have been described, their accuracy varies and the best applicable algorithm-parameter set is different based on the physico-chemical properties of the protein, the ligands and the combination thereof. Another challenge concerning molecular dynamics simulations is that it is difficult to determine whether the conformational changes are caused by the dynamic behaviour of the protein or by inaccuracies in the dynamics parameter sets resulting in the loss of stabilising interactions. In particular, all molecular dynamics simulations will at some point result in protein structures that deviate so much from the starting structures, such that the carefully refined spatial arrangements deemed important in the homology modelling methods are lost. Whereas these changes from the carefully refined interactions may help elucidate the subtle differences between the binding of various ligands, observations from far advanced molecular dynamics states may be misinterpreted. Nevertheless, the careful investigation using molecular dynamics allows the inclusion of the flexible behaviour of a protein active site relating to ligand specificity and induced binding. As such, we have been able to determine that our homology models are stable throughout the molecular dynamics simulation, as is demonstrated by the conservation of protein-ligand and protein-water interactions. Therefore, we are confident that the structural integrity of our protein models is sufficient for further investigation on protein-ligand interactions. In future, molecular dynamics simulations can be expanded by calculating protein-ligand binding free energies that can be used to evaluate *in vitro* measured binding affinities.

To obtain a predictive model for inhibitor binding affinity, we have performed molecular docking of our compound dataset. Ideally, the docking study is validated using a training set of protein-ligand interaction points. Since this is not available for either CYP11B1 or CYP11B2, a training set has been constructed devised of the *in vitro* measurement data, followed by a validation using a small test set of newly synthesised compounds. Investigation of this docking validation method showed that the homology models were among the best performing protein states sampled by the molecular dynamics simulations. It not only suggests that the homology models are best suitable for the docking of novel CYP11B inhibitors, but also that these structures, which have been constructed and refined

mostly at the early stages of the project, have been performing at an optimal level until the end of the project. However, to be able to predict binding affinity in a reliable way, we suggest that homology models are updated at regular intervals during molecular modelling research, and that discrepancies of the protein structure are chartered such that the misprediction of the original models can be adjusted using a correction term. This allows the original homology models to be used for a prolonged time during which the *in silico* models can be optimised using novel *in vitro* measurement data.

In addition to the complexity in the estimation of ligand binding affinities, there are also several limitations of molecular docking. The protein homology model structure is treated as a static composition of the protein. Protein-ligand interactions are rather a result of the interplay between protein and ligand characteristics, indicating that ligand-induced binding modes will not be discovered by molecular docking. In comparison, the limitation of our homology models is that they fail to dock certain types of substructures due to protein-ligand clashes. We have attempted to compensate for ligand-induced binding effects by including the dynamic properties of the active site using protein states sampled by the molecular dynamics simulations. Indeed, different active site conformations are able to relate the potency of ligand in different ways, emphasising that the inclusion of the flexible behaviour is important for the evaluation of such ligand-induced binding modes.

Next to these limitations, our protein models possess a poor description for the active site entrance. The compounds possessing a substituent that points out of the active site cavity are continuously mispredicted, indicating that our models require further structural refinements. The active site cavity currently opens up into solvent, however, structural assessment of the cytochrome P450 fold indicates that this region is membrane bound. In comparison, the crystal structures of most cytochrome P450 enzymes have been made soluble for structure elucidation, especially in the regions contacting the membrane (helices F and G). Using the conformation of these crystal structures during modelling has not lead to an optimal description for this active site region, even with the most questionable crystal structures discarded from the analysis (CYP2C9, CYP2D6). Therefore, we propose that it is questionable whether the modified characteristics of these mutant regions still represent the *in vivo* folding of the protein. To our knowledge, no studies are currently available in which these mutation influences have been derived. As a result, these regions will continue to possess structural inaccuracies for all future homology modelling attempts on cytochromes. Perhaps, for the correct elucidation of the protein fold, protein-membrane interactions need to be taken into account, which opens up a new challenging field of research.

Next to protein-inhibitor interactions, we have described the interactions between the proteins and their natural substrates. From this analysis, we have related the CYP11B2 specific conversion of the steroid 18-hydroxycorticosterone to a delicate difference in binding mode for the two isoforms. In the CYP11B1 active site the steroid forms an inhibitory interaction with the iron-oxygen species preventing its conversion, whereas in CYP11B2 it forms a stabilising internal hydrogen bond that allows its conversion into aldosterone. Indeed, in a quantum mechanical ligand-based exploration of substrate reactivity, we have found that this CYP11B2 specific steroid conformation is the most plausible for conversion



into aldosterone by the iron-oxygen species. Further exploration of the steroid reactivity suggests that other paths in the conversion of 11-deoxycorticosterone to aldosterone exist, including a path that ends with the *in vivo* dominant product, corticosterone, and a path that ends with the *in vivo* CYP11B1 final product 18-hydroxycorticosterone.

Although these promising results are a first step towards the understanding of the substrate conversion mechanism, several key interactions have been neglected. Firstly, the iron-oxygen species can assume different oxidation states, influencing the stabilisation and conversion of the ligands in the protein active site. Secondly, the ligand-based approach only considers the nucleophilic attack of the oxygen atom to the substrate and neglects the stabilising interactions of amino acids in the protein active site. Thirdly, the investigated substrate conversion paths are all based on the assumption that the ligands do not leave the active site until the final product has been formed. In order to compensate for these assumptions, a method such as QMMM can be performed to increase the modelling detail of the iron-oxygen species as well as to take into consideration important active site interactions. Even if the homology model active site detail required for the accurate performance of QMMM on CYP11B1 and CYP11B2 may be insufficient to determine their conversion differences, the calculation of, for instance, only CYP11B2 may provide new insights on the conversion mechanism. The investigation using QMMM has not been performed due to study time limitations, but is recommended for future research. Currently, the quantum mechanics calculations performed in this thesis have yielded plausible explanations for the preferred substrate conversion paths and these considerations can be used as a starting point for further explorations.

A parallel investigation towards the transition state of 18-hydroxycorticosterone during its conversion to aldosterone has elucidated the heme-substrate complex and has provided insight into the construction of transition state inhibitors that are able to mimic the CYP11B2 specific conformation of the substrate. The elucidation of the transition state has been approximated using single point calculations, whereas this normally requires a careful investigation using quantum mechanical transition state detection methods that are very difficult to converge. Thus, deriving conformations using quantum mechanics calculation is difficult, yet the simplified method can be applied for the derivation of the transition state. The transition state features the presence of an  $sp^2$  hybridised carbon that must be specifically mimicked. Any large deviations from the transition state may result in mimicking the binding of other CYP11B substrates, which probably results in the inhibition of both isoforms. Our approximation of the 18-hydroxycorticosterone transition state is accurate for the construction of analogues, but whether the exact conformation of the transition state can be accomplished by a CYP inhibitor remains to be seen.

A final approach to the prediction of binding affinity is to use ligand-based methods such as (3D)QSAR. The construction of a predictive QSAR model can be difficult and time-consuming. This is especially true for our compound set where changes in different substructures influence the potency of other substructures in a different manner. This indicates that the compounds are too flexible or may bind in different binding modes, making

it difficult for a QSAR to be constructed. To circumvent these problems, we have used a simplified type of structure-activity elucidation method, namely decision tree analysis. Using decision tree analyses allows the quick evaluation of ligand structures and the prediction of their potency. However, the method employed in this thesis still involves quality control and requires the definition of different training and test sets. The inaccurate distribution of the compound in the various training or test sets can result in the overtraining of the decision tree analysis. Moreover, if the compounds are very similar, it can be difficult to construct a reliable decision tree that can distinguish the different contributions to inhibitor potency. Hence, the analysis is both limited by the compounds that are used for model construction and compounds that can be accurately predicted by the models. Taking these considerations into account, a decision tree analysis model is more easily updated than a QSAR model each time new compounds are available, and it can provide a strong tool for the prediction of ligand potency. Since the prediction set for our final decision tree models is relatively small, the power of the decision tree analysis is difficult to determine. However, we have been able to determine substrate properties that are vital for inhibitor potency for CYP11B1 and CYP11B2, which can be used as discriminants for the design of selective inhibitors.

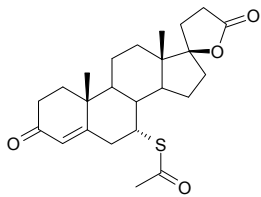
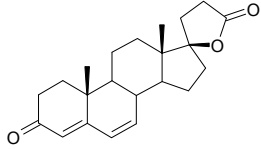
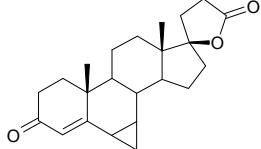
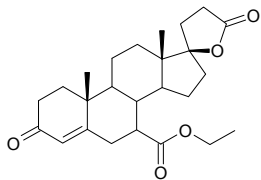
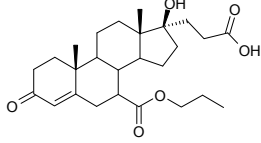
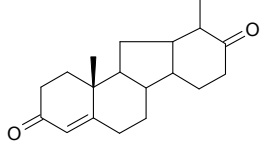
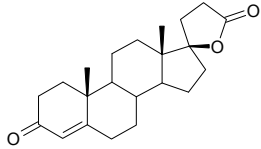
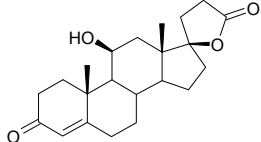
In conclusion, using the molecular modelling techniques investigated in our study, we have been able to explain protein-ligand interactions that contribute to binding affinity and inhibitor selectivity. Pending future investigations, our *in silico* models can be refined by performing site-directed mutagenesis, ligand microcalorimetry measurements for unravelling the contributions to binding affinities, crystal structure elucidation, and further optimisation using *in vitro* measurement data from newly synthesised compounds. Although this project has been aimed towards development of a drug for the treatment of heart failure, the methods discussed in this thesis are independent from the drug target and can be applied to any drug project.

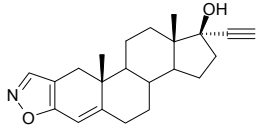
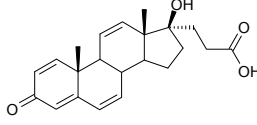
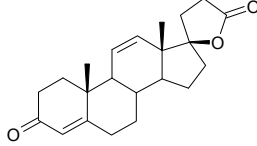
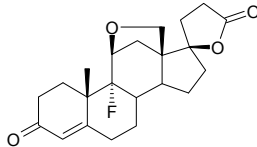
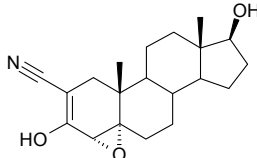
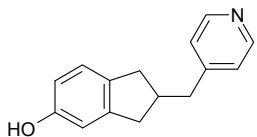
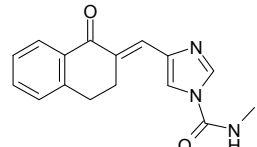
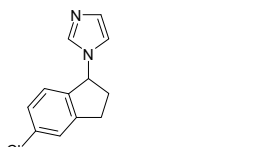
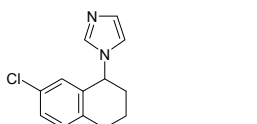


## Appendix A: Known inhibitors of the CYP11B family

The following structures were identified at the start of the project to possess CYP11B1 or CYP11B2 inhibition. The cut-off for the activity separations are 10 $\mu$ M (--), 1 $\mu$ M (-), 100nM (+) <10nM (++) and 1nM (+++).

Compound Name	Structure	CYP11B1 Activity	CYP11B2 Activity	Reference
18-ethynyl-11-deoxycorticosterone		-	+	[1,2]
18-vinyl-11-deoxycorticosterone		+	+	[2]
18-ethynylprogesterone		--	-	[2,3]
18-vinylprogesterone		+	++	[2,3,4]
20-hydroxyimino-pregna-5,14-diene-3 $\beta$ -ol		-	-	[5]
Mespirenone		-	-	[6,7]

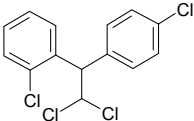
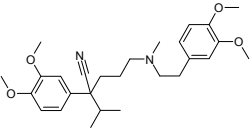
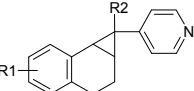
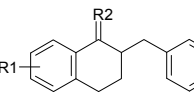
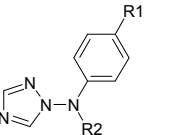
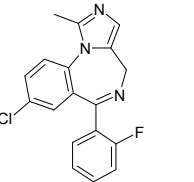
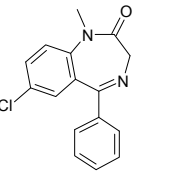
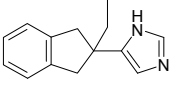
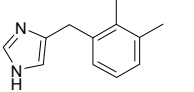
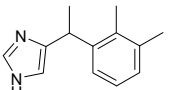
Compound Name	Structure	CYP11B1 Activity	CYP11B2 Activity	Reference
Spirolactone		-	-	[7]
Canrenone		-	-	[7,8,9]
SC 23133, Prorenone		+	+	[8]
SC 26304		-	-	[8]
SC 27169		-	-	[8]
SC 19886		+	+	[8]
SC 5233, Spirolactam		--	-	[10]
SC 5233, 11-hydroxylated		--	-	[10]

Compound Name	Structure	CYP11B1 Activity	CYP11B2 Activity	Reference
Danazole		--	--	[11]
84-0029-05		--	--	[12]
23-0586		--	--	[12]
23-0563		--	-	[12]
Trilostane		--	--	[13,14,15]
5-hydroxy-2-(4-pyridylmethane)indane		+	++	[5]
4'-methoxy-4-(1-(1H-imidazole)ethyl)biphenyl		-	-	[5]
5-chloro-1-(1-imidazolyl)indane		-	-	[5]
7-chloro-1-(1-imidazolyl)tetralin		+++	+++	[5]

Compound Name	Structure	CYP11B1 Activity	CYP11B2 Activity	Reference
7-bromo-1-(1-imidazolyl) tetralin		+++	+++	[5]
1-(4-pyridyl(methyl)) tetralin		+	+	[5]
3-(4-aminophenyl)-3-cyclohexylpiperidine-2,6-dione		+	+	[5]
Clotrimazole		++	++	[12]
Miconazole		+	+	[12]
CGS-16949A (fadrazole)		+	+++	[16,17]
aminoglutethimide		++	++	[15,18,19,20]
3-(4-aminophenyl)-3-cyclohexylpiperidine-2,6-dione		+	+	[18]
R-Etomidate		+++	+++	[12,21,22,23,24]

Compound Name	Structure	CYP11B1 Activity	CYP11B2 Activity	Reference
R 8110 (4-Fluoro-etomidate)		++	++	[23]
Ketoconazole		++	++	[12,20,24,25,26]
Fluconazole		-	-	[25]
Metyrapone		++	++	[20,24,27,28]
Metyrapone Analogues		++	++	[28,29]
CGS-20267 (Ietrozole)		-	-	[30,31]
CGS-18320B		+	+++	[30]
3-MeSO2-DDE		-	-	[20,32,33,34]
4-Me-SO2-2,3,6,4'-tetra-chlorobiphenyl (4-CB64)		-	-	[32]



Compound Name	Structure	CYP11B1 Activity	CYP11B2 Activity	Reference
o,p'-DDD		-	-	[33]
Verapamil		+	++	[35]
4-pyridyl substituted tetrahydrocyclopropa(a)naphthalenes		+++	+++	[36,37]
Tetrahydronaphthalenes		+++	+++	[37]
1-N,N-disubstituted amino-1-H-1,2,4,-triazoles		+++	+++	[38]
Midazolam		--	--	[39]
Diazepam		--	--	[39]
Atipamezole		+	+	[40]
Detomidine		+	+	[40]
Medetomidine		+	+	[40]

## Literature

- 1 C.E. Gomez-Sanchez, E.P. Gomez-Sanchez, M.F. Foecking, M.Y. Zhou, "Inhibition of steroidogenesis in rat adrenal cells by 18-ethynyldeoxycorticosterone: evidence for an alternative pathway of aldosterone biosynthesis", *J. Steroid Biochem. Molec. Biol.*, 1997, 62, 2/3, 207-212
- 2 E. Davioud, A. Piffeteau, C. Delorme, S. Coustal, A. Marquet, "18-vinyldeoxycorticosterone: a potent inhibitor of the bovine cytochrome P-450 11 $\beta$ ", *Bioorganic and Medicinal Chemistry*, 1998, 6, 10, 1781-1788
- 3 C. Delorme, A. Piffeteau, A. Viger, A. Marquet, "Inhibition of bovine cytochrome P-45011 $\beta$  by 18-unsaturated progesterone derivatives", *Eur. J. Biochem.*, 1995, 232, 1, 247-256
- 4 G. Defaye, A. Piffeteau, C. Delorme, A. Marquet, "Specific inhibition of the last step of aldosterone biosynthesis by 18-vinylprogesterone in bovine adrenocortical cells", *J. Steroid. Biochem. Molec. Biol.*, 1996, 57, 1, 141-147
- 5 P.B. Ehmer, M. Bureik, R. Bernhardt, U. Muller, R.W. Hartmann, "Development of a test system for inhibitors of human aldosterone synthase (CYP11B2): screening in fission yeast and evaluation of selectivity in V79 cells", *Journal of Steroid Biochemistry and Molecular Biology*, 2002, 81, 2, 173-179
- 6 K. Weindel, S. Lewicka, P. Vecsei, "Inhibitory effects of the novel anti-aldosterone compound mespirenone on adrenocortical steroidogenesis in vitro", *Arzneimittelforschung.*, 1991, 41, 9, 946-949
- 7 K. Weindel, S. Lewicka, P. Vecsei, "Comparative dose-effect studies with four C-17-spirosteroidal mineralocorticoid synthesis inhibitors", *J. Steroid Biochem.*, 1989, 34, 1-6, 455-459
- 8 P. Netchitailo, C. Delarue, I. Perroteau, F. Leboulenger, M.-H. Capron, H. Vaudry, "Relative inhibitory potency of five mineralocorticoid antagonists on aldosterone biosynthesis in vitro", *Biochemical Pharmacology*, 1985, 34, 2, 189-194
- 9 H. Naganuma, N. Sasano, "The effects of canrenone K (soldactone) on the adrenocortical steroidogenic enzymes in bovine adrenal gland", *Folia Endocrinol.*, 1986, 62, 3, 149-157
- 10 H.C. Erbler, "Selective inhibition of aldosterone synthesis by 11-hydroxylated spiro lactone in rat adrenals", *Arch. Pharmacol.*, 1973, 280, 3, 331-337
- 11 R.L. Barbieri, R. Osathanondh, J.A. Canick, R.J. Stillman, K.J. Ryan, "Danazol inhibits human adrenal 21- and 11 $\beta$ -hydroxylation in vitro", *Steroids*, 1980, 35, 3, 251-263
- 12 A. Wada, T. Ohnishi, Y. Nonaka, M. Okamoto, "Inhibition of bovine adrenocortical mitochondrial cytochrome P-45011 $\beta$ -mediated reactions by imidazole derivatives and mineralocorticoid analogs", *J. Steroid. Biochem.*, 1988, 31, 5, 803-808
- 13 G.O. Potts, J.E. Creange, H.R. Hardinmg, H.P. Schane, "Trilostane, an orally active inhibitor of steroid biosynthesis", *Steroids*, 1978, 32, 2, 257-267
- 14 H.R. Harding, J.E. Creange, G.O. Potts, H.P. Schane, "Inhibition of furosemide-induced kaliuresis in the rat by trilostane, an inhibitor of adrenal steroidogenesis (41961)", *Proceedings of the Society for Experimental Biology and Medicine*, 1984, 177, 3, 388-391
- 15 F. Lambert, F. Corcelle-Cerf, J. Lammerant, J. Kolanowski, "On the specificity of the inhibitory effect of trilostane and aminoglutethimide on adrenocortical steroidogenesis in guinea pig", *Molecular and Cellular Endocrinology*, 1984, 37, 1, 115-120
- 16 S.W.J. Lamberts, H.A. Bruining, H. Marzouk, J. Zuiderwijk, P. Uitterlinden, J.J. Blijd, W.H.L. Hackeng, F.H. de Jong, "The new aromatase inhibitor CGS-16949A suppresses aldosterone and cortisol production by human adrenal cells in vitro", *Journal of Clinical Endocrinology and Metabolism*, 1989, 69, 4, 896-901
- 17 L.E. Demers, J.C. Melby, T.E. Wilson, A. Lipton, H.A. Harvey, R.J. Santen, "The effects of CGS 16949A, an aromatase inhibitor on adrenal mineralocorticoid biosynthesis", *Journal of Clinical Endocrinology and Metabolism*, 1990, 70, 4, 1162-1166
- 18 R.W. Hartmann, G. Grun, U. Bartz, M. Palzer, "Evaluation of the racemate and the enantiomers of a new highly active and selective aromatase inhibitor of the aminoglutethimide type", *J. Steroid. Biochem. Molec. Biol.*, 1992, 43, 7, 641-648
- 19 R.J. Santen, R.I. Misbin, "Aminoglutethimide: review of pharmacology and clinical use", *Pharmacotherapy*, 1981, 1, 2, 95-120

- 20 M.K. Johansson, J.T. Sanderson, B.-O. Lund, "Effects of 3-MeSO<sub>2</sub>-DDE and some CYP inhibitors on glucocorticoid steroidogenesis in the H295R human adrenocortical carcinoma cell line", *Toxicology in Vitro*, 2002, 16, 2, 113-121
- 21 I. Varga, K. Racz, R. Kiss, L. Futo, M. Toth, O. Sergev, E. Glaz, "Direct inhibitory effect of etomidate on corticosteroid secretion in human pathologic adrenocortical cells", *Steroids*, 1993, 58, 2, 64-68
- 22 H.M. Schulte, G. Benker, D. Reinwein, W.G. Sippell, B. Allolio, "Infusion of low dose etomidate: correction of hypercortisolemia in patients with Cushing's Syndrome and dose-response relationship in normal subjects", *Journal of Clinical Endocrinology and Metabolism*, 1990, 70, 5, 1426-1430
- 23 R. De Coster, W. Wouters, D. Beerens, C. Haelterman, R. Doolaeghe, N. Goeminne, M. Krekels, "Comparative effects of etomidate and its fluoro analogue, R8110 on testicular, adrenal and ovarian steroid biosynthesis", *J. Vet. Pharmacol. Ther.*, 1988, 11, 4, 345-355
- 24 M. Fassnacht, S. Hahner, F. Beuschlein, A. Klink, M. Reincke, B. Allolio, "New mechanisms of adrenostatic compounds in a human adrenocortical cancer cell line", *European Journal of Clinical Investigation*, 2000, 30, S3, 76-82
- 25 C. Eckhoff, W. Oelkers, V. Bahr, "Effect of two oral antimycotics, ketoconazole and fluconazole, upon steroidogenesis in rat adrenal cells in vitro", *J. Steroid Biochem.*, 1988, 31, 5, 819-823
- 26 R.M. Couch, J. Muller, Y.S. Perry, J.S.D. Winter, "Kinetic analysis of inhibition of human adrenal steroidogenesis by ketoconazole", *Journal of Clinical Endocrinology and Metabolism*, 1987, 65, 3, 551-554
- 27 M. Watanuki, B.E. Tilley, P.F. Hall, "Cytochrome P-450 for 11 $\beta$ - and 18-hydroxylase activities of bovine adrenocortical mitochondria: one enzyme or two?", *Biochemistry*, 1978, 17, 1, 127-130
- 28 S.J. Hays, M.C. Tobes, D.L. Gildersleeve, D.M. Wieland, W.H. Beierwaltes, "Structure-activity relationship study of the inhibition of adrenal cortical 11 $\beta$ -hydroxylase by new metyrapone analogues", *J. Med. Chem.*, 1984, 27, 1, 15-19
- 29 M.C. Tobes, S.J. Hays, D.L. Gildersleeve, D.M. Wieland, W.H. Beierwaltes, "Adrenal cortical 11 $\beta$ -hydroxylase and side-chain cleavage enzymes. requirement for the A- and B-pyridyl ring in metyrapone for inhibition", *J. Steroid Biochem.*, 1985, 22, 1, 103-110
- 30 A.S. Bhatnagar, A. Hausler, K. Schieweck, M. Lang, R. Bowman, "Highly selective inhibition of estrogen biosynthesis by CGS20267, a new non-steroidal aromatase inhibitor", *J. Steroid. Biochem. Molec. Biol.*, 1990, 37, 6, 1021-1027
- 31 L.M. Demers, A. Lipton, H.A. Harvey, K.B. Kambic, H. Grossberg, C. Brady, R.J. Santen, "The efficacy of CGS 20267 in suppressing estrogen biosynthesis in patients with advanced stage breast cancer", *J. Steroid. Biochem. Molec. Biol.*, 1993, 44, 4-6, 687-691
- 32 M. Johansson, C. Larsson, A. Bergman, B.-O. Lund, "Structure-activity relationship for inhibition of CYP11B1-dependant glucocorticoid synthesis in Y1 cells by aryl methyl sulfones", *Pharmacology & Toxicology*, 1998, 83, 5, 225-230
- 33 O. Lindhe, B. Skogseid, I. Brandt, "Cytochrome P450-catalyzed binding of 3-methylsulfonyl-DDE and o,p'-DDD in human adrenal zona fasciculata/reticularis", *The Journal of Clinical Endocrinology & Metabolism*, 2002, 87, 3, 1319-1326
- 34 B.-O. Lund, J. Lund, "Novel involvement of a mitochondrial steroid hydroxylase (P450c11) in xenobiotic metabolism", *The Journal of Biological Chemistry*, 1995, 270, 36, 20895-20897
- 35 N. Blanchouin-Emeric, M. Zenatti, G. Defaye, B. Aupetit, "Verapamil directly inhibits aldosterone synthesis by adrenal mitochondria in vitro", *J. Steroid. Biochem.*, 1988, 30, 1-6, 453-456
- 36 R.W. Hartmann, H. Bayer, G. Grun, T. Seergejew, U. Bartz, M. Mitrenga, "Pyridyl-substituted tetrahydrocyclopropa[a]naphthalenes: highly active and selective inhibitors of P450 arom", *J. Med. Chem.*, 1995, 38, 12, 2103-2111
- 37 G.A. Wachter, R.W. Hartmann, T. Sergejew, G.L. Grun, D. Ledergerber, "Tetrahydronaphthalenes: influence of heterocyclic substituents on inhibition of steroidogenic enzymes P450 arom and P450 17", *J. Med. Chem.*, 1996, 39, 4, 834-841
- 38 M. Okada, T. Yoden, E. Kawaminami, Y. Shimada, M. Kudoh, Y. Isomura, "Studies on aromatase inhibitors. II. synthesis and biological evaluation of 1-amino-1H-1,2,4-triazole derivatives", *Chem. Pharm. Bull.*, 1997, 45, 2, 333-337

- 39 I. Thomson, R. Fraser, C.J. Kenyon, "Inhibition of bovine adrenocortical steroidogenesis by benzodiazepines: a direct effect on microsomal hydroxylation or an inhibition of calcium uptake?", *Journal of Endocrinology*, 1992, 135, 2, 361-369
- 40 L.P. Jager, G.J. De Graaf, H.C. Aura Widjaja-Greefkes, "Effects of atipamezole, detomidine, and medetomidine on release of steroid hormones by porcine adrenocortical cells in vitro", *European Journal of Pharmacology*, 1998, 346, 1, 71-76

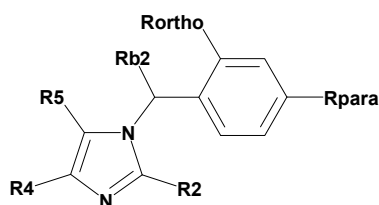


## Appendix B: Chemical Structures of Fadrazole Analogues

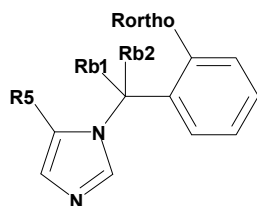
Chemical structures and *in vitro* potency values for fadrazole analogues (Moeras numbering). Potency is shown in nM. The various substituent positions can be found in each of their generic scaffolds below.

The selectivity factor has been determined by dividing the CYP11B1 potency by the CYP11B2 potency, thus a higher value indicates a higher selectivity for CYP11B1B2.

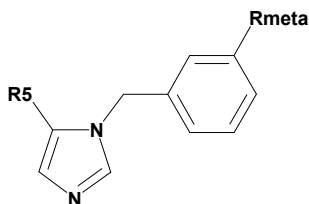
Compounds used in the molecular docking study have a Yes/No label indicating that their docking pose contained the imidazole group pointing towards the heme ('Yes') or not ('No'). Compounds used which have been docked and predicted using the regression in Chapter 4, possess the 'Prediction' label. Compounds used for the training and test set in the decision tree analysis are indicated with a 'Yes' label, whereas the compounds with the 'Prediction' label have been used for prediction.



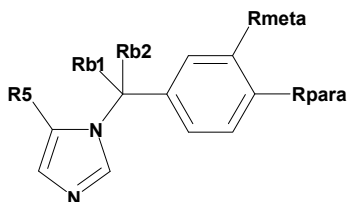
Moeras	R2	R4	R5	Rb2	Rortho	Rpara	11B1 IC50 (nM)	11B2 IC50 (nM)	Selectivity (11B1/11B2)	Docking pose ok?	DT used?
44	CH2OH	H	H	H	H	OCH3	1000+	1000+	-----	Yes	Yes
46	CH2OH	H	H	Rb-Ro-Cyhex	H	H	254.3 ± 26.5	1628.2 ± 433.4	0.16	--	No
133	CH3	H	H	H	H	H	300+	300+	-----	Yes	Yes
138	CH3	H	H	H	H	CN	522.2 ± 107.4	203.5 ± 30.0	2.57	Yes	Yes
159	NH2	R4-R5-Ph	H	H	H	H	300+	300+	-----	No	Yes
160	SH	H	H	H	H	H	300+	300+	-----	Yes	Yes



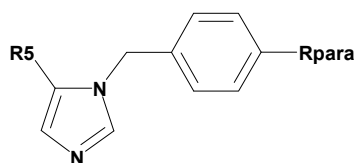
Moeras	R5	Rb1	Rb2	Rortho	11B1 IC50 (nM)	11B2 IC50 (nM)	Selectivity (11B1/11B2)	Docking Pose ok?	DT used?
45	H	H	Rb-Ro-Cyhex		16.3 ± 2.8	44.5 ± 5.3	0.37	Yes	Yes
249	H	Ph	2Ph-C2H4-2Ph		890.09 ± 161.1	1000+	-----	No	Yes
250	H	H	2Ph-CH3-Ph		47.3 ± 7.4	52.2 ± 5.0	0.91	Prediction	Yes
290	Ph	H	H	CN	49.3 ± 35.8	139.8 ± 43.8	0.35	Prediction	Yes



Moeras R5	Rmeta	11B1		11B2		Selectivity (11B1/11B2)	Docking Pose ok?	DTused?
		IC50 (nM)		IC50 (nM)				
103	t-But	NHCOCH3	6.1 ± 1.4	117.8 ± 9.3	0.05	Yes	Yes	
139	H	CN	300+	300+	-----	Yes	Yes	
141	C3H6OH	CN	168.2 ± 19.4	300+	-----	Yes	Yes	
142	H	COOH	300+	1070 ± 378.5	-----	No	Yes	
144	H	F	205.7 ± 40.1	815.1 ± 23.3	0.25	Yes	Yes	
145	H	Br	142.7 ± 98.6	365.2 ± 99.8	0.39	Yes	Yes	
146	C3H6OH	F	56.6 ± 9.1	46.1 ± 2.0	1.23	Yes	Yes	
148	H	Cl	151.3 ± 19.5	456.6 ± 28.5	0.33	Yes	Yes	
150	H	COOCH3	730.23 ± 63.7	300+	-----	No	Yes	
152	C3H6OH	Br	34.6 ± 3.9	28.9 ± 4.0	1.20	Yes	Yes	
154	H	Br	58.9 ± 1.9	156.3 ± 7.1	0.38	Yes	Yes	
155	H	NH2	390.0 ± 90.6	600.0 ± 124.5	0.65	Yes	Yes	
156	H	CH2OH	206.3 ± 1.5	1174.4 ± 221.1	0.18	Yes	Yes	
157	H	HCO	398.0 ± 48.3	393.2 ± 143.9	1.01	Yes	Yes	
158	C3H6OH	Cl	50.2 ± 15.4	94.7 ± 3.4	0.53	Yes	Yes	
183	Ph	OCH3	32.4 ± 3.2	18.1 ± 1.8	1.79	Yes	Prediction	
272	Ph	CN	32.1 ± 13.1	28.3 ± 6.6	1.13	Yes	Yes	
275	Ph	F	2.6 ± 1.7	16.7 ± 1.2	0.16	Yes	Yes	
276	Ph	Br	1.8 ± 1.0	7.4 ± 2.1	0.24	Yes	Yes	
278	Ph	Cl	1.9 ± 0.3	14.8 ± 4.0	0.13	Yes	Yes	
280	Ph	NO2	13.0 ± 9.2	11.1 ± 3.8	1.17	Yes	Yes	

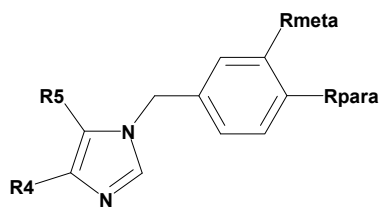


Moeras R5	Rb1	Rb2	Rmeta	Rpara	11B1		11B2		Selectivity (11B1/11B2)	Docking Pose ok?	DTused?
					IC50 (nM)		IC50 (nM)				
5	R5-Rb-Cyhex	H	H	2-Pyr	6.8 ± 1.24	139.2 ± 49.2	0.05	Yes	Yes		
6	R5-Rb-Cyhex	H	H	HCO	9.8 ± 4.2	93.4 ± 30.7	0.11	Yes	Yes		
7	R5-Rb-Cyhex	H	H	CH2OH	2.2 ± 1.1	68.0 ± 18.0	0.03	Yes	Yes		
8	R5-Rb-Cyhex	H	H	COCH3	4.7 ± 0.3	23.1 ± 6.6	0.20	Yes	Yes		
11_R	COOC2H5	CH3*	H	H	0.5 ± 0.2	1.7 ± 0.9	0.29	Yes	Yes		
26	R5-Rb-Cyhex	H	H	COOH	2145.9 ± 410.6	871.1 ± 110.2	2.46	Yes	Yes		
27_R	R5-Rb-Cyhex*	H	H	CN	118.6 ± 8.9	6.0 ± 1.9	19.77	Yes	Yes		
27_S	R5-Rb-Cyhex*	H	H	CN	39.5 ± 4.4	171.2 ± 51.7	0.23	Yes	Yes		
52	R5-Rb-Cyhex	H	H	COOCH3	2.0 ± 0.6	129.5 ± 21.4	0.02	Yes	Yes		
60	R5-Rb-Cyhex	H	H	CONH2	70.52 ± 20.1	446.5 ± 3.1	0.16	Yes	Yes		
107	R5-Rb-Cyhex	H	H	H	36.3 ± 5.99	159.4 ± 19.1	0.23	Yes	Yes		
108	R5-Rb-Cyhex	H	H	Br	30.5 ± 4.31	60.4 ± 11.2	0.50	Yes	Yes		
181	H	C3H6OH	H	CN	7.4 ± 0.9	2.4 ± 0.3	3.08	Yes	Yes		
182	H	CH2PH	H	CN	189.6 ± 18.0	237.6 ± 9.9	0.80	Yes	Yes		
202	H	Ph	H	H	10.81 ± 1.9	271.91 ± 85.5	0.04	Yes	Yes		
227	R5-Rb-Cyhex	H	F	H	74.21 ± 6.3	357.79 ± 58.4	0.21	Yes	Yes		
228	R5-Rb-Cyhex	H	F	F	23.6 ± 1.8	252.3 ± 97.6	0.09	Yes	Yes		
229	R5-Rb-Cyhex	H	Br	H	45.04 ± 0.7	252.69 ± 87.0	0.18	Yes	Yes		
235	H	3CN-Ph	H	CN	380.63 ± 16.7	1480.25 ± 657.8	0.26	No	Yes		
236	H	C3H7	H	CN	85.41 ± 7.6	186.83 ± 63.8	0.46	Yes	Yes		
237	H	C3H7	C3H7	CN	85.57 ± 3.5	181.24 ± 8.1	0.47	Yes	Yes		
251	H	Ph	H	H	88.7 ± 2.1	473.8 ± 101.1	0.19	Yes	Yes		
252	Cycprop-CH2OH	Ph	H	H	258.9 ± 2.3	120.5 ± 35.5	2.15	No	Yes		
293	Ph	CH3	H	CN	2.3 ± 0.7	8.9 ± 3.5	0.26	Prediction	Yes		
295	H	C3H6OH	C3H6OH	CN	1000+	1000+	-----	Prediction	Prediction		
297	R5-2-Ph-CH2-Rb	H	H	CN	39.2 ± 3.4	41.2 ± 0.9	0.95	Prediction	Prediction		
322	H	Rb-Rb-Cycprop	H	CN	10.2 ± 2.2	101.4 ± 5.9	0.10	Prediction	Prediction		
324	H	C2H4OH	H	CN	36.6 ± 6.9	292.6 ± 14.6	0.13	Prediction	Prediction		
325	Ph	C2H4OH	H	CN	N.D.	N.D.	-----	--	No		
326	Ph	C3H6OH	H	CN	37.0 ± 0.7	403.4 ± 46.7	0.09	Prediction	Prediction		

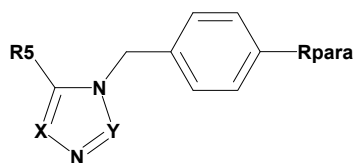


Moeras R5	Rpara		11B1 IC50 (nM)	11B2 IC50 (nM)	Selectivity (11B1/11B2)	Docking Pose ok?	DT used?
9	2-Naph	COOH	1463.1 ± 347.1	363.6 ± 103.2	4.03	No	Yes
10	2-Naph	CN	20.3 ± 10.4	24.8 ± 24.8	0.82	Yes	Yes
24	C3H6OH	CN	48.16 ± 19.76	13.1 ± 4.9	3.68	Yes	Yes
25	C3H6Cl	CN	4.75 ± 3.0	1.4 ± 0.2	3.39	Yes	Yes
42	H	CN	461.5 ± 156.0	465.5 ± 258.0	0.99	Yes	Yes
43	H	OCH3	456.5 ± 152.3	1952.0 ± 111.7	0.23	Yes	Yes
49	C2H4COOCH3	OCH3	111.4 ± 31.1	121.3 ± 44.5	0.92	Yes	Yes
50	C3H6OH	OCH3	12.3 ± 2.0	58.7 ± 2.8	0.21	Yes	Yes
51	C3H6OH	NO2	4.7 ± 1.0	88.8 ± 9.0	0.05	Yes	Yes
53	C3H6OH	CH3	48.1 ± 4.1	23.6 ± 7.6	2.04	Yes	Yes
54	C3H6OH	COOCH3	5.8 ± 2.5	24.4 ± 11.9	0.24	Yes	Yes
55	H	CN	368.2 ± 93.0	372.4 ± 245.2	0.99	Yes	Yes
56	H	COOH	300+	300+	-----	No	Yes
57	H	OCH3	282.3 ± 112.7	763.0 ± 102.8	0.37	Yes	Yes
58	H	CONH2	557.8 ± 31.7	3543.6 ± 534.9	0.16	Yes	Yes
59	H	NO2	496.6 ± 250.2	372.9 ± 0.8	1.33	Yes	Yes
61	H	NH2	235.5 ± 49.6	5705.5 ± 276.7	0.04	Yes	Yes
109	H	Br	211.6 ± 13.97	478.5 ± 24.2	0.44	Yes	Yes
110	H	H	709.14 ± 67.9	1799.0 ± 322.0	0.39	Yes	Yes
114	CH3	CN	140.5 ± 42.9	12.3 ± 4.3	11.42	Yes	Yes
115	Ph	CN	27.5 ± 3.5	1.7 ± 0.6	16.18	Yes	Yes
116	HCO	CN	477.9 ± 372.5	61.8 ± 5.4	7.73	Yes	Yes
117	COOCH3	CN	47.8 ± 21.2	7.1 ± 6.7	6.73	Yes	Yes
118	CH2OH	CN	285.2 ± 29.4	28.5 ± 3.7	10.01	Yes	Yes
119	CH3OCOCH3	CN	212.5 ± 17.2	43.55 ± 9.5	4.88	Yes	Yes
120	C2H4COOCH3	CN	31.1 ± 6.1	3.5 ± 0.6	8.89	Yes	Yes
121	C2H4COOCH3	CH3	79.1 ± 0.03	15.1 ± 2.3	5.24	Yes	Yes
122	C2H4COOCH3	NO2	16.3 ± 0.7	4.7 ± 2.1	3.47	Yes	Yes
123	C2H4COOCH3	H	326.0 ± 18.0	192.9 ± 17.87	1.69	Yes	Yes
124	C2H4COOCH3	Br	57.4 ± 5.6	4.0 ± 0.48	14.35	Yes	Yes
125	C2H4COOCH3	COOCH3	111.3 ± 6.1	21.52 ± 2.14	5.17	No	Yes
126	C3H6OH	H	180.5 ± 18.6	285.73 ± 26.00	0.63	Yes	Yes
127	C3H6OH	Br	18.4 ± 4.0	6.21 ± 0.91	2.96	Yes	Yes
128	CHCHCH3	CH3	3.5 ± 0.9	3.79 ± 2.21	0.92	Yes	Yes
129	C3H7	CH3	5.6 ± 0.1	2.83 ± 0.19	1.98	Yes	Yes
130	C3H6Cl	H	2.16 ± 0.73	6.20 ± 0.15	0.35	Yes	Yes
131	C3H6Cl	Br	1.1 ± 0.01	3.8 ± 0.2	0.29	Yes	Yes
134	H	CH3	602.0 ± 182.3	1932.5 ± 210.6	0.31	Yes	Yes
137	Br	CN	73.2 ± 6.9	23.1 ± 6.5	3.17	Yes	Yes
143	H	F	494.3 ± 31.2	329.8 ± 230.8	1.50	Yes	Yes
147	H	Cl	300+	300+	-----	Yes	Yes
149	H	COOCH3	218.3 ± 50.9	300+	-----	Yes	Yes
151	C3H6OH	F	115.5 ± 6.3	15.5 ± 3.2	7.45	Yes	Yes
153	C3H6OH	Cl	43.6 ± 0.4	54.5 ± 1.2	0.80	Prediction	Yes
173	H	COOCH2Ph	322.2 ± 82.0	1000+	-----	No	Yes
174	H	CH2OH	695.1 ± 73.3	1000+	-----	Yes	Yes
184	Ph	OCH3	11.3 ± 0.9	14.0 ± 2.1	0.81	Yes	Yes
201	CH2OH	H	311.78 ± 6.4	1000+	-----	Yes	Yes
203	HCO	H	580.52 ± 205.5	1039.83 ± 208.7	0.56	Yes	Yes
204	CCON(CH3)2	H	1000+	1000+	-----	Yes	Yes
230	3-CH3-Ph	CN	129.88 ± 79.3	116.7 ± 145.6	1.11	Yes	Yes
231	C2H4COOCH3	F	104.42 ± 7.8	49.6 ± 78.5	2.11	Yes	Yes
232	C2H4COOCH3	Cl	123.38 ± 30.0	176.75 ± 53.3	0.70	Yes	Yes
233	3-Pyr	CN	517.0 ± 13.3	N.d.	-----	Yes	Yes
238	4-Pyr	CN	307.06 ± 23.0	407.41 ± 37.6	0.75	Yes	Yes
268	Ph	Br	7.3 ± 1.8	5.1 ± 2.0	1.43	Yes	Yes
269	Ph	F	15.9 ± 5.3	10.7 ± 1.5	1.49	Yes	Yes
270	2-CH3-Ph	CN	5.7 ± 2.2	3.7 ± 3.4	1.54	Yes	Yes
271	3-CH3-Ph	CN	6.2 ± 2.4	5.2 ± 5.2	1.19	Yes	Yes
273	Ph	NO2	13.7 ± 5.4	2.3 ± 0.8	5.96	Yes	Yes
274	4-F-Ph	CN	27.4 ± 8.6	24.8 ± 5.1	1.10	Yes	Yes
279	Ph	Cl	25.4 ± 15.4	5.8 ± 1.6	4.38	Yes	Yes
281	3-F-Ph	CN	31.8 ± 8.5	5.5 ± 1.0	5.78	Yes	Yes
282	2-F-Ph	CN	19.9 ± 3.4	2.3 ± 0.5	8.65	Yes	Yes
291	Ph	H	4.8 ± 0.6	10.7 ± 2.7	0.45	Prediction	Yes
292	2-OCH3-Ph	CN	5.3 ± 4.3	12.7 ± 2.9	0.42	Prediction	Yes
294	4-CN-Ph	CN	15.1 ± 9.6	30.6 ± 22.3	0.49	Prediction	Prediction
296	2-CH2OH-Ph	CN	107.9 ± 10.7	20.4 ± 2.2	5.29	Prediction	Prediction
299	2-C2H5-Ph	CN	1.6 ± 0.2	0.9 ± 0.1	1.78	Prediction	Prediction
301	C4H9	CN	0.8 ± 0.2	1.1 ± 0.1	0.73	Prediction	Prediction
302	C2H4Cycchex	CN	2.9 ± 0.8	1.2 ± 0.2	2.42	Prediction	Prediction
318	3-HCO-Ph	CN	125.4 ± 24.5	58.3 ± 6.6	2.15	Prediction	Prediction





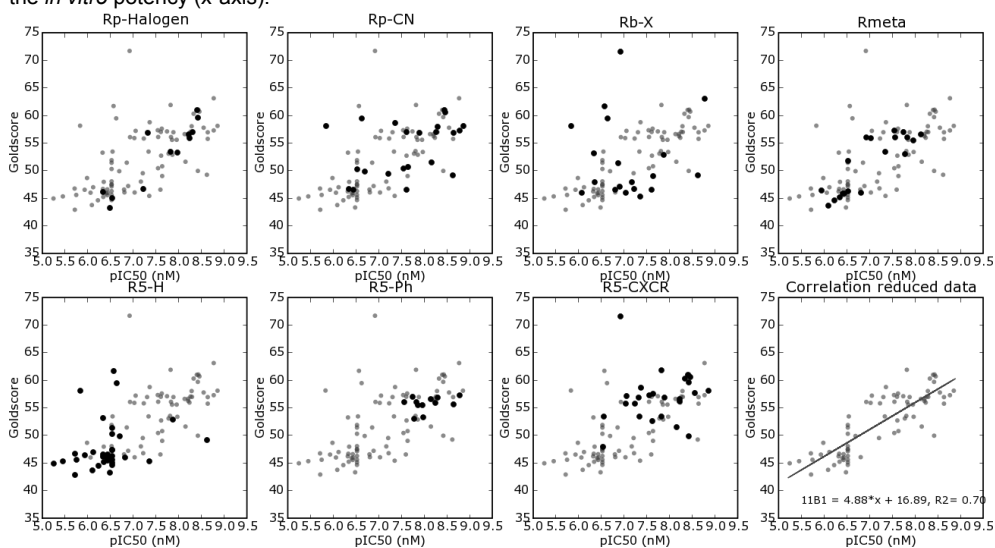
Moeras	R4	R5	Rmeta	Rpara	11B1 IC50 (nM)	11B2 IC50 (nM)	Selectivity (11B1/11B2)	Docking Pose ok?	DT used?
136	R4-R5-Ph	H	H	Cl	503.5 ± 47.5	346.1 ± 70.1	1.45	No	Yes
140	CH3	H	H	CN	300+	300+	-----	Yes	Yes
161	R4-R5-Ph	H	H	COOH	300+	300+	-----	No	Yes
170	R4-R5-Ph	H	H	COOCH3	219.1 ± 3.7	901.21 ± 17.8	0.24	No	Yes
171	R4-R5-Ph	H	H	COOCH2H5	889.7 ± 88.2	1000+	-----	No	Yes
172	R4-R5-Ph	H	H	COOCH2Ph	227.8 ± 78.9	1000+	-----	No	Yes
175	R4-R5-Ph	H	H	CH2OH	302.7 ± 14.5	1000+	-----	No	Yes
213	Cl	Cl	C(NH2)NOH	H	1000+	1000+	-----	Yes	Yes



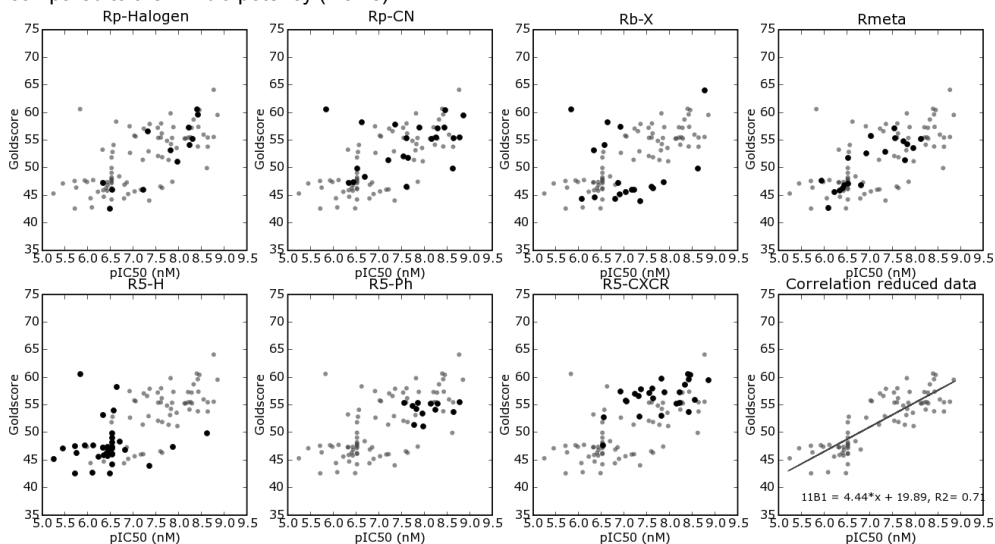
Moeras	X	Y	R5	Rpara	11B1 IC50 (nM)	11B2 IC50 (nM)	Selectivity (11B1/11B2)	Docking Pose ok?	DT used?
267	N	CH	2-Pyr	CN	1942.0 ± 864.2	1027.2 ± 106.8	1.89	No	No
283	N	N	Ph	CN	1000+	528.3 ± 111.9	-----	No	No
289	N	CH	Ph	CN	1000+	461.9 ± 31.7	-----	No	No

### Appendix C.1.1: CYP11B1 Docking Results using GoldScore

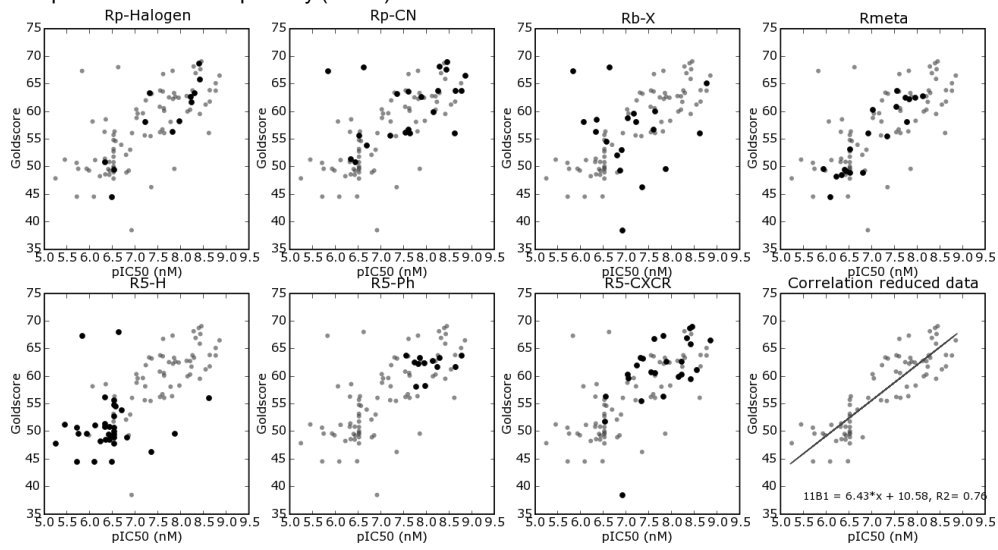
Group docking results (y-axis) in models equilibrated with Fadrazole (homology model), compared to the *in vitro* potency (x-axis).



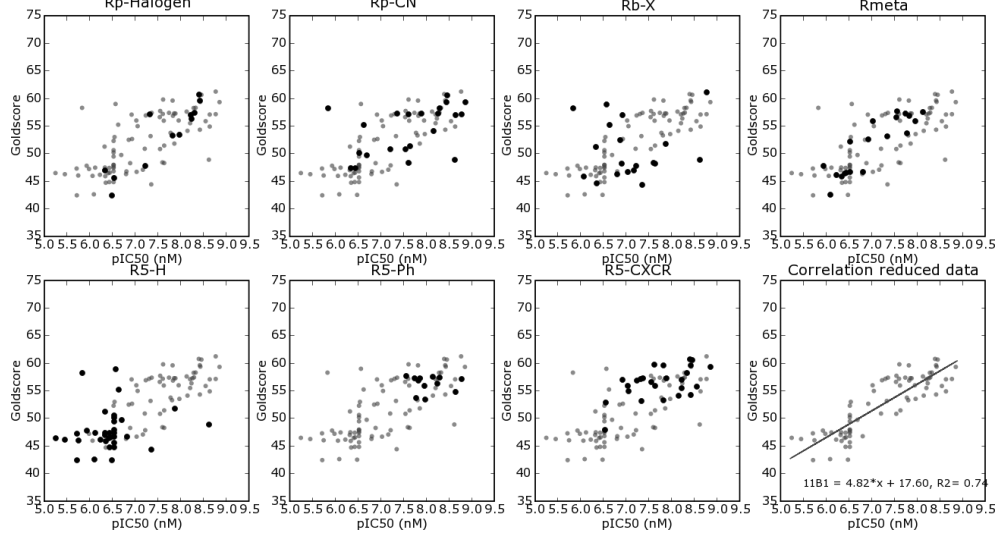
Group docking results (y-axis) in models equilibrated with Fadrazole (average results protein states), compared to the *in vitro* potency (x-axis).



Group docking results (y-axis) in models equilibrated with 18-hydroxycorticosterone (homology model), compared to the *in vitro* potency (x-axis).

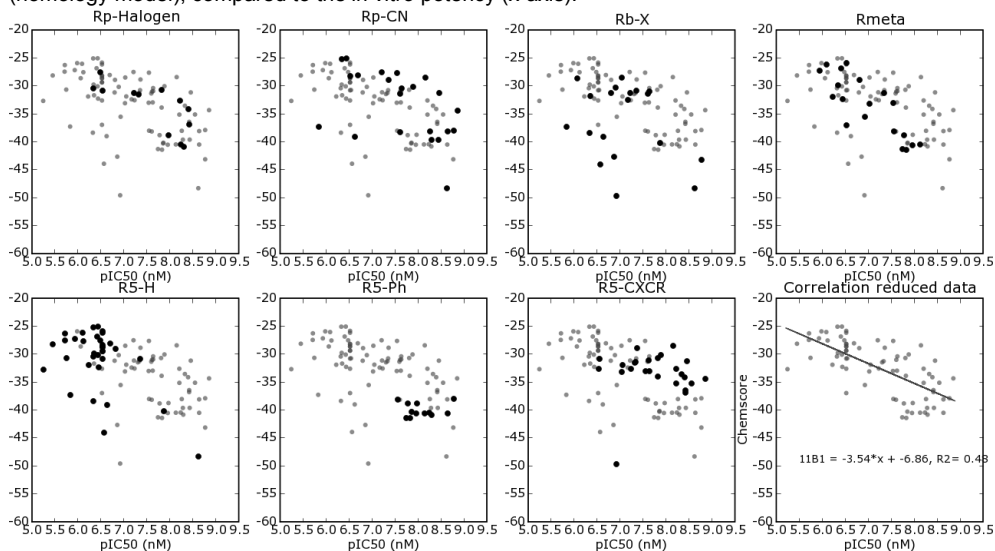


Group docking results (y-axis) in models equilibrated 18-hydroxycorticosterone (average results protein states), compared to the *in vitro* potency (x-axis).

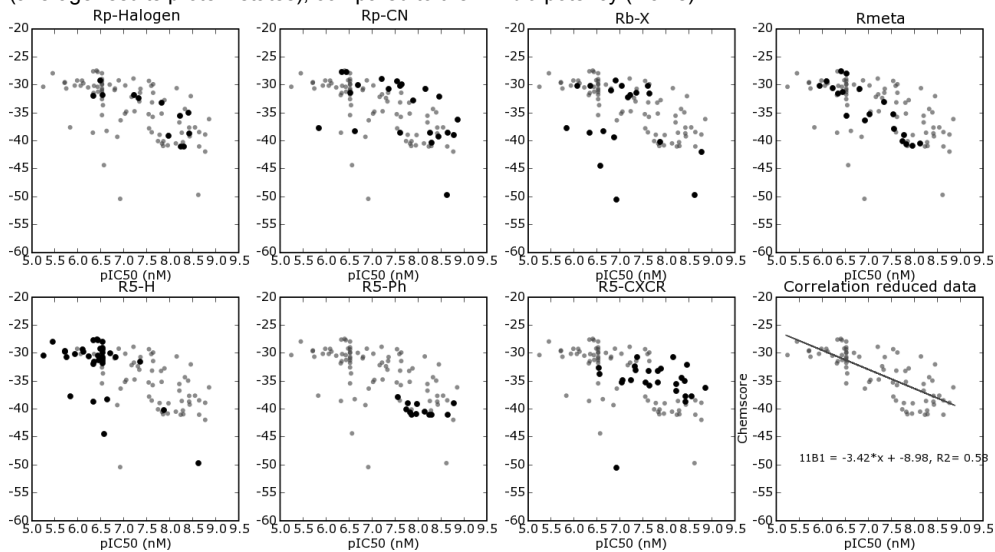


## Appendix C.1.2: CYP11B1 Docking Results using ChemScore

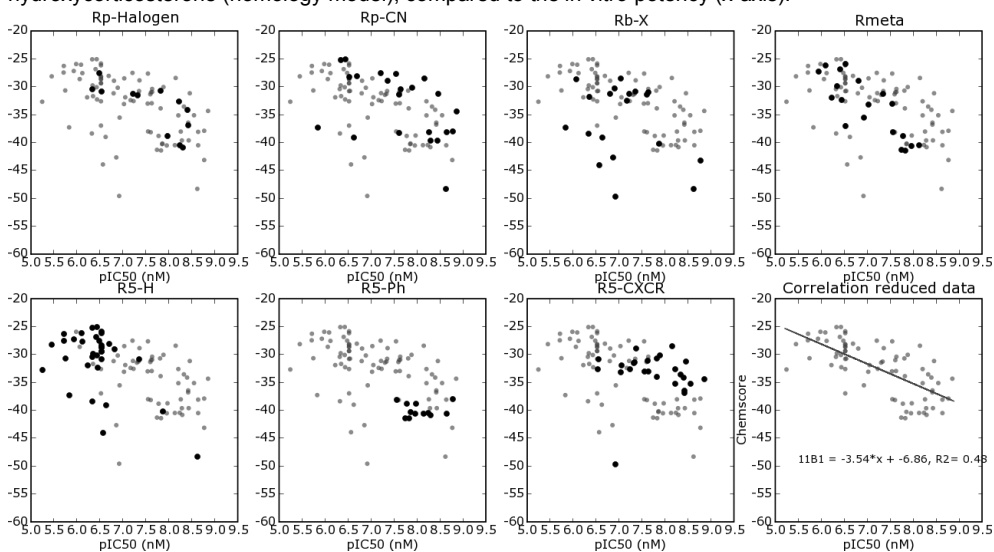
Group docking results (calculated free energy of binding, y-axis) in models equilibrated with Fadrazole (homology model), compared to the *in vitro* potency (x-axis).



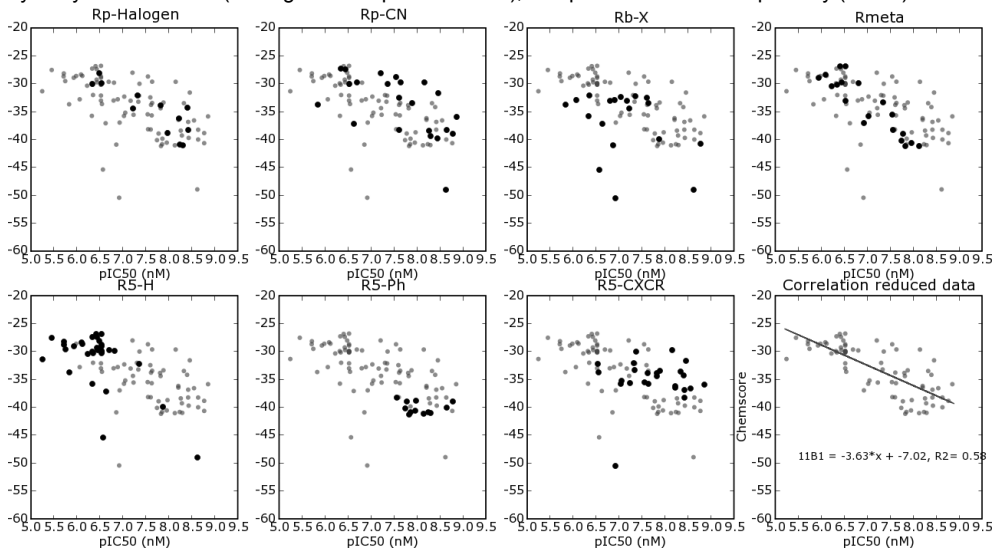
Group docking results (calculated free energy of binding, y-axis) in models equilibrated with Fadrazole (average results protein states), compared to the *in vitro* potency (x-axis).



Group docking results (calculated free energy of binding, y-axis) in models equilibrated with 18-hydroxycorticosterone (homology model), compared to the *in vitro* potency (x-axis).

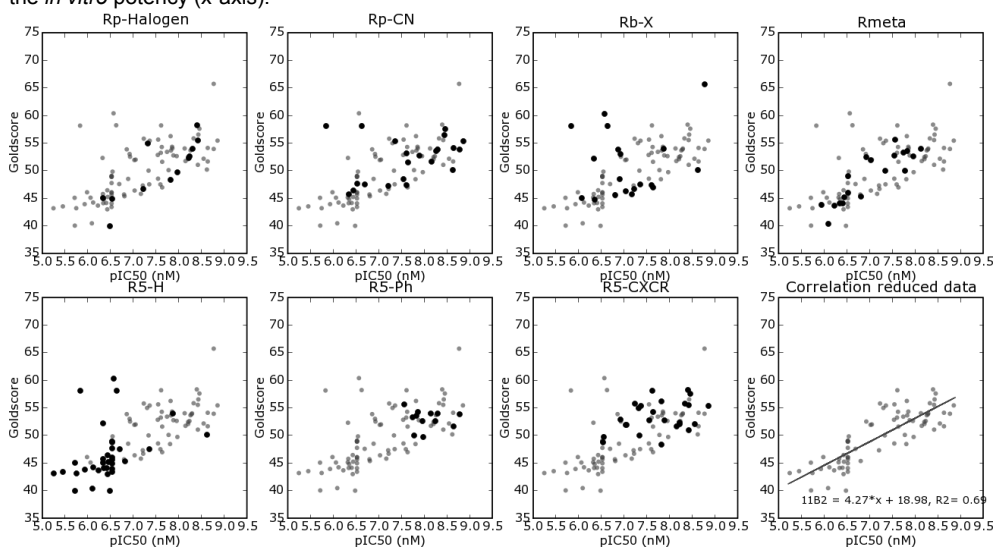


Group docking results (calculated free energy of binding, y-axis) in models equilibrated with 18-hydroxycorticosterone (average results protein states), compared to the *in vitro* potency (x-axis).

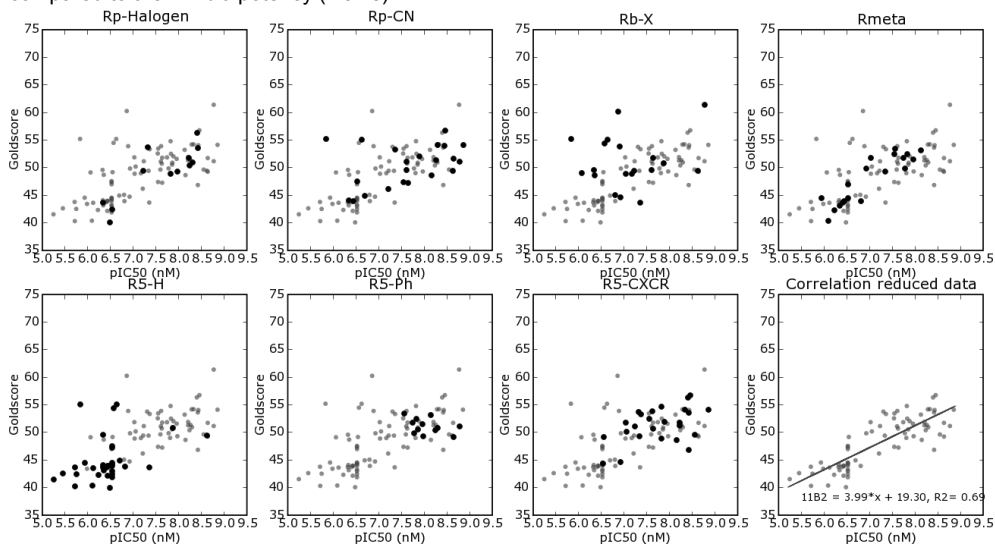


### Appendix C.2.1: CYP11B2 Docking Results using GoldScore

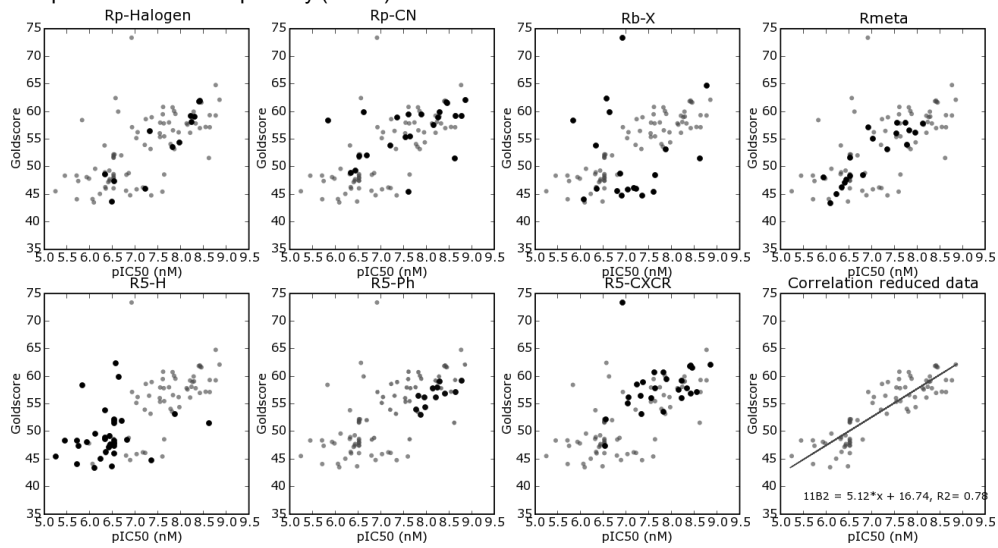
Group docking results (y-axis) in models equilibrated with Fadrazole (homology model), compared to the *in vitro* potency (x-axis).



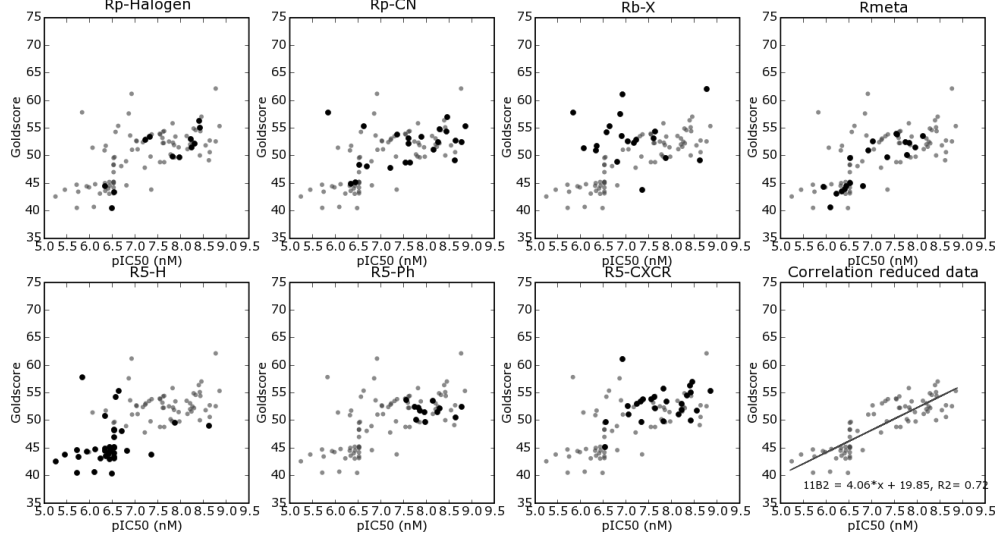
Group docking results (y-axis) in models equilibrated with Fadrazole (average results protein states), compared to the *in vitro* potency (x-axis).



Group docking results (y-axis) in models equilibrated with 18-hydroxycorticosterone (homology model), compared to the *in vitro* potency (x-axis).

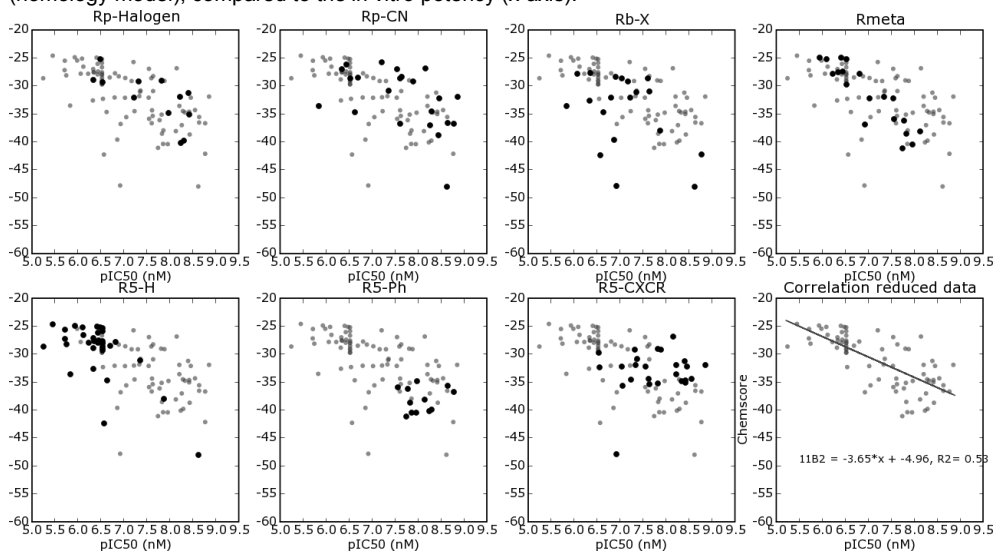


Group docking results (y-axis) in models equilibrated with 18-hydroxycorticosterone (average results protein states), compared to the *in vitro* potency (x-axis).

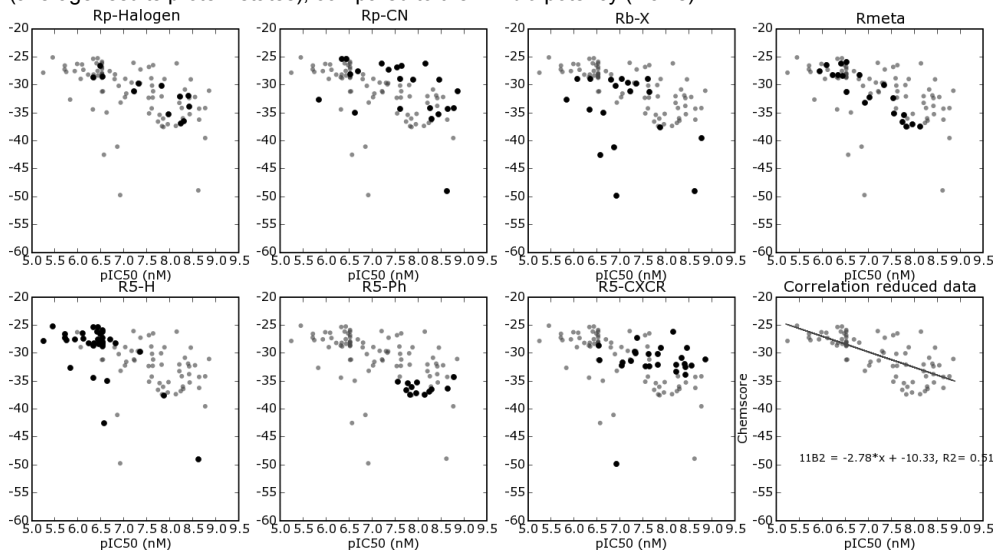


## Appendix C.2.2: CYP11B2 Docking Results using ChemScore

Group docking results (calculated free energy of binding, y-axis) in models equilibrated with Fadrazole (homology model), compared to the *in vitro* potency (x-axis).

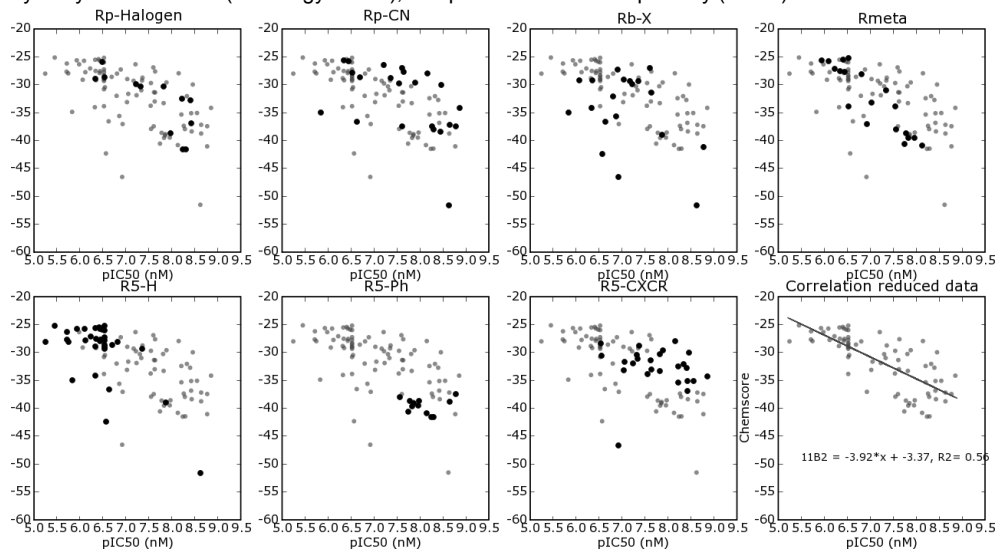


Group docking results (calculated free energy of binding, y-axis) in models equilibrated with Fadrazole (average results protein states), compared to the *in vitro* potency (x-axis).

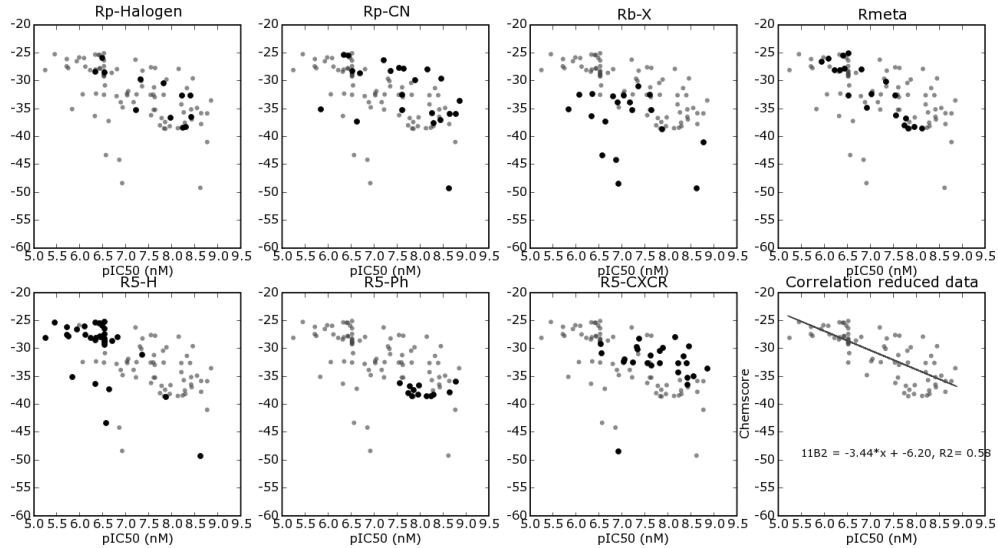




Group docking results (calculated free energy of binding, y-axis) in models equilibrated with 18-hydroxycorticosterone (homology model), compared to the *in vitro* potency (x-axis).



Group docking results (calculated free energy of binding, y-axis) in models equilibrated with 18-hydroxycorticosterone (average results protein states), compared to the *in vitro* potency (x-axis).



### Appendix D: Decision Tree Analysis Results

Decision tree model results are displayed in matrix form. Herein, the results are vertically ordered by cost factor (0.25 to 4) and horizontally by leaf size settings (1 to 5), as well as descriptor sets. *Italic values indicate models that possess cost settings that lead to the overtraining of these models. Greyed values indicate the model settings that have been used for the potency prediction of novel compounds.*

Calculated Matthews's correlation coefficient (MCC) for the decision tree models as per training set (q<sup>2</sup>) and test set (R<sup>2</sup>). An MCC higher than 0 indicates that the models behave better than a random guess and an MCC of 1.0 would indicate a perfect model.

CYP11B1	Minimum Leafsize 1					Minimum Leafsize 2					Minimum Leafsize 3					Minimum Leafsize 4					Minimum Leafsize 5									
	Original		Sterilab		Both	Original		Sterilab		Both	Original		Sterilab		Both	Original		Sterilab		Both	Original		Sterilab		Both					
	Cost	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	Cost	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	Cost	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	Cost	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	Cost	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>
0.25	<i>0.68</i>	<i>0.43</i>	<i>0.78</i>	<i>0.43</i>	<i>0.72</i>	<i>0.50</i>	<i>0.68</i>	<i>0.43</i>	<i>0.78</i>	<i>0.43</i>	<i>0.72</i>	<i>0.49</i>	<i>0.68</i>	<i>0.43</i>	<i>0.78</i>	<i>0.44</i>	<i>0.70</i>	<i>0.49</i>	<i>0.72</i>	<i>0.48</i>	<i>0.75</i>	<i>0.43</i>	<i>0.70</i>	<i>0.47</i>	<i>0.63</i>	<i>0.42</i>	<i>0.70</i>	<i>0.46</i>	<i>0.69</i>	<i>0.46</i>
0.33	<i>0.75</i>	<i>0.47</i>	<i>0.87</i>	<i>0.44</i>	<i>0.81</i>	<i>0.47</i>	<i>0.74</i>	<i>0.49</i>	<i>0.82</i>	<i>0.46</i>	<i>0.84</i>	<i>0.50</i>	<i>0.74</i>	<i>0.51</i>	<i>0.89</i>	<i>0.48</i>	<i>0.77</i>	<i>0.47</i>	<i>0.69</i>	<i>0.48</i>	<i>0.74</i>	<i>0.43</i>	<i>0.72</i>	<i>0.46</i>	<i>0.64</i>	<i>0.42</i>	<i>0.73</i>	<i>0.43</i>	<i>0.67</i>	<i>0.46</i>
0.50	<i>0.84</i>	<i>0.50</i>	<i>0.95</i>	<i>0.59</i>	<i>0.91</i>	<i>0.48</i>	<i>0.81</i>	<i>0.50</i>	<i>0.92</i>	<i>0.53</i>	<i>0.87</i>	<i>0.48</i>	<i>0.81</i>	<i>0.50</i>	<i>0.92</i>	<i>0.53</i>	<i>0.87</i>	<i>0.48</i>	<i>0.74</i>	<i>0.46</i>	<i>0.84</i>	<i>0.41</i>	<i>0.79</i>	<i>0.44</i>	<i>0.70</i>	<i>0.47</i>	<i>0.81</i>	<i>0.46</i>	<i>0.77</i>	<i>0.43</i>
0.56	<i>0.83</i>	<i>0.50</i>	<i>0.95</i>	<i>0.46</i>	<i>0.91</i>	<i>0.45</i>	<i>0.76</i>	<i>0.42</i>	<i>0.87</i>	<i>0.43</i>	<i>0.87</i>	<i>0.45</i>	<i>0.73</i>	<i>0.42</i>	<i>0.85</i>	<i>0.40</i>	<i>0.82</i>	<i>0.47</i>	<i>0.72</i>	<i>0.51</i>	<i>0.83</i>	<i>0.43</i>	<i>0.76</i>	<i>0.47</i>	<i>0.70</i>	<i>0.47</i>	<i>0.79</i>	<i>0.43</i>	<i>0.74</i>	<i>0.48</i>
0.63	<i>0.83</i>	<i>0.48</i>	<i>0.95</i>	<i>0.42</i>	<i>0.90</i>	<i>0.47</i>	<i>0.77</i>	<i>0.43</i>	<i>0.89</i>	<i>0.40</i>	<i>0.85</i>	<i>0.46</i>	<i>0.74</i>	<i>0.45</i>	<i>0.86</i>	<i>0.38</i>	<i>0.82</i>	<i>0.47</i>	<i>0.74</i>	<i>0.52</i>	<i>0.81</i>	<i>0.35</i>	<i>0.78</i>	<i>0.44</i>	<i>0.69</i>	<i>0.39</i>	<i>0.79</i>	<i>0.36</i>	<i>0.77</i>	<i>0.49</i>
0.71	<i>0.86</i>	<i>0.52</i>	<i>0.96</i>	<i>0.46</i>	<i>0.90</i>	<i>0.47</i>	<i>0.77</i>	<i>0.45</i>	<i>0.91</i>	<i>0.43</i>	<i>0.86</i>	<i>0.43</i>	<i>0.81</i>	<i>0.46</i>	<i>0.87</i>	<i>0.42</i>	<i>0.81</i>	<i>0.48</i>	<i>0.73</i>	<i>0.47</i>	<i>0.82</i>	<i>0.35</i>	<i>0.80</i>	<i>0.46</i>	<i>0.70</i>	<i>0.41</i>	<i>0.78</i>	<i>0.44</i>	<i>0.77</i>	<i>0.44</i>
0.83	<i>0.86</i>	<i>0.50</i>	<i>0.95</i>	<i>0.48</i>	<i>0.93</i>	<i>0.49</i>	<i>0.81</i>	<i>0.48</i>	<i>0.88</i>	<i>0.46</i>	<i>0.87</i>	<i>0.49</i>	<i>0.76</i>	<i>0.49</i>	<i>0.85</i>	<i>0.41</i>	<i>0.81</i>	<i>0.44</i>	<i>0.74</i>	<i>0.50</i>	<i>0.81</i>	<i>0.42</i>	<i>0.80</i>	<i>0.41</i>	<i>0.67</i>	<i>0.36</i>	<i>0.78</i>	<i>0.47</i>	<i>0.76</i>	<i>0.49</i>
1.00	<i>0.84</i>	<i>0.56</i>	<i>0.92</i>	<i>0.47</i>	<i>0.94</i>	<i>0.58</i>	<i>0.79</i>	<i>0.49</i>	<i>0.84</i>	<i>0.46</i>	<i>0.88</i>	<i>0.58</i>	<i>0.73</i>	<i>0.45</i>	<i>0.84</i>	<i>0.42</i>	<i>0.84</i>	<i>0.48</i>	<i>0.69</i>	<i>0.44</i>	<i>0.78</i>	<i>0.48</i>	<i>0.76</i>	<i>0.38</i>	<i>0.67</i>	<i>0.42</i>	<i>0.78</i>	<i>0.46</i>	<i>0.74</i>	<i>0.40</i>
1.20	<i>0.89</i>	<i>0.55</i>	<i>0.92</i>	<i>0.47</i>	<i>0.96</i>	<i>0.55</i>	<i>0.80</i>	<i>0.55</i>	<i>0.84</i>	<i>0.44</i>	<i>0.89</i>	<i>0.60</i>	<i>0.76</i>	<i>0.50</i>	<i>0.84</i>	<i>0.42</i>	<i>0.84</i>	<i>0.50</i>	<i>0.73</i>	<i>0.45</i>	<i>0.78</i>	<i>0.45</i>	<i>0.79</i>	<i>0.45</i>	<i>0.71</i>	<i>0.45</i>	<i>0.78</i>	<i>0.46</i>	<i>0.75</i>	<i>0.44</i>
1.40	<i>0.86</i>	<i>0.50</i>	<i>0.90</i>	<i>0.48</i>	<i>0.96</i>	<i>0.48</i>	<i>0.77</i>	<i>0.50</i>	<i>0.84</i>	<i>0.48</i>	<i>0.88</i>	<i>0.45</i>	<i>0.74</i>	<i>0.48</i>	<i>0.83</i>	<i>0.44</i>	<i>0.82</i>	<i>0.35</i>	<i>0.73</i>	<i>0.48</i>	<i>0.83</i>	<i>0.45</i>	<i>0.80</i>	<i>0.33</i>	<i>0.70</i>	<i>0.46</i>	<i>0.79</i>	<i>0.49</i>	<i>0.77</i>	<i>0.48</i>
1.60	<i>0.82</i>	<i>0.45</i>	<i>0.90</i>	<i>0.48</i>	<i>0.91</i>	<i>0.49</i>	<i>0.77</i>	<i>0.51</i>	<i>0.83</i>	<i>0.53</i>	<i>0.86</i>	<i>0.50</i>	<i>0.74</i>	<i>0.49</i>	<i>0.83</i>	<i>0.53</i>	<i>0.81</i>	<i>0.44</i>	<i>0.71</i>	<i>0.48</i>	<i>0.82</i>	<i>0.41</i>	<i>0.82</i>	<i>0.39</i>	<i>0.68</i>	<i>0.42</i>	<i>0.79</i>	<i>0.49</i>	<i>0.75</i>	<i>0.40</i>
1.80	<i>0.82</i>	<i>0.46</i>	<i>0.87</i>	<i>0.44</i>	<i>0.90</i>	<i>0.50</i>	<i>0.76</i>	<i>0.48</i>	<i>0.85</i>	<i>0.45</i>	<i>0.83</i>	<i>0.51</i>	<i>0.73</i>	<i>0.47</i>	<i>0.85</i>	<i>0.44</i>	<i>0.79</i>	<i>0.44</i>	<i>0.70</i>	<i>0.46</i>	<i>0.80</i>	<i>0.41</i>	<i>0.80</i>	<i>0.46</i>	<i>0.70</i>	<i>0.43</i>	<i>0.79</i>	<i>0.40</i>	<i>0.79</i>	<i>0.46</i>
2.00	<i>0.80</i>	<i>0.45</i>	<i>0.87</i>	<i>0.46</i>	<i>0.90</i>	<i>0.49</i>	<i>0.78</i>	<i>0.56</i>	<i>0.86</i>	<i>0.46</i>	<i>0.87</i>	<i>0.48</i>	<i>0.76</i>	<i>0.50</i>	<i>0.85</i>	<i>0.46</i>	<i>0.77</i>	<i>0.45</i>	<i>0.75</i>	<i>0.50</i>	<i>0.84</i>	<i>0.46</i>	<i>0.77</i>	<i>0.44</i>	<i>0.71</i>	<i>0.44</i>	<i>0.82</i>	<i>0.43</i>	<i>0.80</i>	<i>0.45</i>
3.00	<i>0.73</i>	<i>0.43</i>	<i>0.86</i>	<i>0.58</i>	<i>0.78</i>	<i>0.45</i>	<i>0.73</i>	<i>0.43</i>	<i>0.87</i>	<i>0.54</i>	<i>0.78</i>	<i>0.45</i>	<i>0.69</i>	<i>0.46</i>	<i>0.85</i>	<i>0.58</i>	<i>0.74</i>	<i>0.49</i>	<i>0.66</i>	<i>0.45</i>	<i>0.80</i>	<i>0.54</i>	<i>0.73</i>	<i>0.51</i>	<i>0.70</i>	<i>0.46</i>	<i>0.80</i>	<i>0.54</i>	<i>0.72</i>	<i>0.46</i>
4.00	<i>0.49</i>	<i>0.33</i>	<i>0.75</i>	<i>0.42</i>	<i>0.71</i>	<i>0.46</i>	<i>0.59</i>	<i>0.43</i>	<i>0.75</i>	<i>0.42</i>	<i>0.70</i>	<i>0.46</i>	<i>0.45</i>	<i>0.50</i>	<i>0.68</i>	<i>0.37</i>	<i>0.67</i>	<i>0.44</i>	<i>0.45</i>	<i>0.50</i>	<i>0.68</i>	<i>0.37</i>	<i>0.67</i>	<i>0.44</i>	<i>0.32</i>	<i>0.20</i>	<i>0.63</i>	<i>0.68</i>	<i>0.62</i>	<i>0.45</i>

CYP11B2	Minimum Leafsize 1					Minimum Leafsize 2					Minimum Leafsize 3					Minimum Leafsize 4					Minimum Leafsize 5									
	Original		Sterilab		Both	Original		Sterilab		Both	Original		Sterilab		Both	Original		Sterilab		Both	Original		Sterilab		Both					
	Cost	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	Cost	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	Cost	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	Cost	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>	R <sup>2</sup>	Cost	q <sup>2</sup>	R <sup>2</sup>	q <sup>2</sup>
0.25	<i>0.74</i>	<i>0.55</i>	<i>0.77</i>	<i>0.53</i>	<i>0.79</i>	<i>0.59</i>	<i>0.73</i>	<i>0.55</i>	<i>0.77</i>	<i>0.53</i>	<i>0.81</i>	<i>0.54</i>	<i>0.73</i>	<i>0.56</i>	<i>0.76</i>	<i>0.59</i>	<i>0.78</i>	<i>0.56</i>	<i>0.74</i>	<i>0.52</i>	<i>0.76</i>	<i>0.53</i>	<i>0.78</i>	<i>0.56</i>	<i>0.68</i>	<i>0.50</i>	<i>0.74</i>	<i>0.51</i>	<i>0.79</i>	<i>0.56</i>
0.33	<i>0.77</i>	<i>0.50</i>	<i>0.84</i>	<i>0.52</i>	<i>0.83</i>	<i>0.51</i>	<i>0.76</i>	<i>0.51</i>	<i>0.83</i>	<i>0.51</i>	<i>0.83</i>	<i>0.52</i>	<i>0.75</i>	<i>0.57</i>	<i>0.83</i>	<i>0.53</i>	<i>0.80</i>	<i>0.52</i>	<i>0.73</i>	<i>0.52</i>	<i>0.79</i>	<i>0.53</i>	<i>0.77</i>	<i>0.58</i>	<i>0.69</i>	<i>0.47</i>	<i>0.78</i>	<i>0.50</i>	<i>0.77</i>	<i>0.56</i>
0.50	<i>0.83</i>	<i>0.47</i>	<i>0.86</i>	<i>0.58</i>	<i>0.84</i>	<i>0.50</i>	<i>0.79</i>	<i>0.49</i>	<i>0.85</i>	<i>0.56</i>	<i>0.84</i>	<i>0.47</i>	<i>0.72</i>	<i>0.50</i>	<i>0.79</i>	<i>0.56</i>	<i>0.84</i>	<i>0.50</i>	<i>0.69</i>	<i>0.49</i>	<i>0.78</i>	<i>0.54</i>	<i>0.80</i>	<i>0.52</i>	<i>0.66</i>	<i>0.55</i>	<i>0.77</i>	<i>0.53</i>	<i>0.78</i>	<i>0.49</i>
0.56	<i>0.80</i>	<i>0.48</i>	<i>0.88</i>	<i>0.59</i>	<i>0.86</i>	<i>0.48</i>	<i>0.75</i>	<i>0.46</i>	<i>0.86</i>	<i>0.54</i>	<i>0.83</i>	<i>0.54</i>	<i>0.70</i>	<i>0.50</i>	<i>0.85</i>	<i>0.54</i>	<i>0.83</i>	<i>0.54</i>	<i>0.68</i>	<i>0.52</i>	<i>0.81</i>	<i>0.51</i>	<i>0.82</i>	<i>0.51</i>	<i>0.62</i>	<i>0.60</i>	<i>0.77</i>	<i>0.52</i>	<i>0.78</i>	<i>0.46</i>
0.63	<i>0.80</i>	<i>0.47</i>	<i>0.89</i>	<i>0.52</i>	<i>0.90</i>	<i>0.47</i>	<i>0.79</i>	<i>0.44</i>	<i>0.86</i>	<i>0.53</i>	<i>0.88</i>	<i>0.53</i>	<i>0.73</i>	<i>0.47</i>	<i>0.82</i>	<i>0.52</i>	<i>0.85</i>	<i>0.58</i>	<i>0.71</i>	<i>0.48</i>	<i>0.81</i>	<i>0.51</i>	<i>0.83</i>	<i>0.53</i>	<i>0.62</i>	<i>0.60</i>	<i>0.79</i>	<i>0.43</i>	<i>0.79</i>	<i>0.47</i>
0.71	<i>0.83</i>	<i>0.45</i>	<i>0.88</i>	<i>0.51</i>	<i>0.91</i>	<i>0.48</i>	<i>0.76</i>	<i>0.44</i>	<i>0.85</i>	<i>0.54</i>	<i>0.88</i>	<i>0.53</i>	<i>0.69</i>	<i>0.48</i>	<i>0.81</i>	<i>0.46</i>	<i>0.85</i>	<i>0.43</i>	<i>0.61</i>	<i>0.59</i>	<i>0.77</i>	<i>0.45</i>	<i>0.76</i>	<i>0.46</i>	<i>0.61</i>	<i>0.59</i>	<i>0.77</i>	<i>0.45</i>	<i>0.76</i>	<i>0.46</i>
0.83	<i>0.82</i>	<i>0.42</i>	<i>0.90</i>	<i>0.46</i>	<i>0.95</i>	<i>0.46</i>	<i>0.76</i>	<i>0.44</i>	<i>0.85</i>	<i>0.55</i>	<i>0.88</i>	<i>0.54</i>	<i>0.71</i>	<i>0.42</i>	<i>0.82</i>	<i>0.43</i>	<i>0.85</i>	<i>0.44</i>	<i>0.70</i>	<i>0.50</i>	<i>0.78</i>	<i>0.47</i>	<i>0.80</i>	<i>0.52</i>	<i>0.63</i>	<i>0.54</i>	<i>0.76</i>	<i>0.38</i>	<i>0.75</i>	<i>0.51</i>
1.00	<i>0.79</i>	<i>0.38</i>	<i>0.93</i>	<i>0.46</i>	<i>0.92</i>	<i>0.50</i>	<i>0.70</i>	<i>0.48</i>	<i>0.83</i>	<i>0.48</i>	<i>0.86</i>	<i>0.48</i>	<i>0.66</i>	<i>0.46</i>	<i>0.80</i>	<i>0.47</i>	<i>0.82</i>	<i>0.54</i>	<i>0.66</i>	<i>0.53</i>	<i>0.77</i>	<i>0.47</i>	<i>0.75</i>	<i>0.56</i>	<i>0.64</i>	<i>0.54</i>	<i>0.76</i>	<i>0.45</i>	<i>0.72</i>	<i>0.54</i>
1.20	<i>0.78</i>	<i>0.41</i>	<i>0.93</i>	<i>0.46</i>	<i>0.90</i>	<i>0.45</i>	<i>0.69</i>	<i>0.45</i>	<i>0.83</i>	<i>0.48</i>	<i>0.83</i>	<i>0.44</i>	<i>0.67</i>	<i>0.45</i>	<i>0.80</i>	<i>0.47</i>	<i>0.78</i>	<i>0.57</i>	<i>0.67</i>	<i>0.49</i>	<i>0.77</i>	<i>0.47</i>	<i>0.75</i>	<i>0.56</i>	<i>0.65</i>	<i>0.51</i>	<i>0.76</i>	<i>0.45</i>	<i>0.73</i>	<i>0.50</i>
1.40	<i>0.79</i>	<i>0.41</i>	<i>0.92</i>	<i>0.47</i>	<i>0.88</i>	<i>0.40</i>	<i>0.68</i>	<i>0.51</i>	<i>0.84</i>	<i>0.48</i>	<i>0.85</i>	<i>0.45</i>	<i>0.65</i>	<i>0.49</i>	<i>0.81</i>	<i>0.47</i>	<i>0.78</i>	<i>0.57</i>	<i>0.65</i>	<i>0.49</i>	<i>0.77</i>	<i>0.42</i>	<i>0.75</i>	<i>0.49</i>	<i>0.62</i>	<i>0.49</i>	<i>0.76</i>	<i>0.45</i>	<i>0.75</i>	<i>0.57</i>
1.60	<i>0.75</i>	<i>0.47</i>	<i>0.85</i>	<i>0.43</i>	<i>0.86</i>	<i>0.49</i>	<i>0.68</i>	<i>0.52</i>	<i>0.82</i>	<i>0.45</i>	<i>0.85</i>	<i>0.46</i>	<i>0.67</i>	<i>0.52</i>	<i>0.76</i>	<i>0.44</i>	<i>0.80</i>	<i>0.56</i>	<i>0.65</i>	<i>0.52</i>	<i>0.76</i>	<i>0.44</i>	<i>0.76</i>	<i>0.45</i>	<i>0.61</i>	<i>0.50</i>	<i>0.76</i>	<i>0.45</i>	<i>0.72</i>	<i>0.50</i>
1.80	<i>0.73</i>	<i>0.43</i>	<i>0.86</i>	<i>0.45</i>	<i>0.83</i>	<i>0.37</i>	<i>0.68</i>	<i>0.52</i>	<i>0.82</i>	<i>0.41</i>	<i>0.83</i>	<i>0.44</i>	<i>0.67</i>	<i>0.52</i>	<i>0.77</i>	<i>0.44</i>	<i>0.80</i>	<i>0.50</i>	<i>0.63</i>	<i>0.53</i>										

Potency Prediction table for the 12 Moeras compounds. Moeras number is shown in the first column. The tables are horizontally ordered by descriptor set. For each descriptor set, 10 models have been used for the prediction of the potency, encompassing 5 different settings for the minimum leaf size (1 to 5) and two different cost settings per leaf size. These settings coincide with the greyed models in the tables on the previous page. The prediction by each model is a value ranging from 0 to 5 indicating the amount of submodels that predict the compound as potent (there are 5 submodels per leaf size and cost setting). The final prediction value is a fraction value, for which 1.00 indicates that all submodels have predicted the compound as potent.

CYP11B1											CYP11B2											
Original descriptor set											Original descriptor set											
Leaf	1	2	2	3	3	4	4	5	5	Fraction	Leaf	1	2	2	3	3	4	4	5	5	Fraction	
Cost	1.2	1.2	2	1.2	2	0.63	2	1.4	2		Cost	0.83	3	1.6	3	1.6	3	1	3	1.6	3	Fraction
183	5	5	5	5	5	5	5	5	5	1.00	5	5	5	5	5	5	5	5	5	5	1.00	
294	2	2	1	4	2	2	2	2	1	0.40	5	5	5	5	5	5	5	5	5	5	1.00	
295	1	3	0	0	0	0	0	0	0	0.08	2	2	2	2	2	3	3	3	0	0	0.38	
296	2	2	2	4	2	2	3	2	1	0.44	5	5	5	5	5	5	5	5	5	5	1.00	
297	5	5	5	5	5	5	4	5	4	0.94	5	5	5	5	5	5	5	5	5	5	1.00	
299	5	5	5	5	5	5	4	5	4	0.92	5	5	5	5	5	5	5	5	5	5	1.00	
301	5	5	5	5	5	5	4	5	4	0.94	4	4	4	4	3	5	4	3	2	2	0.70	
302	5	5	5	5	5	5	4	5	4	0.92	5	5	5	5	5	5	5	5	5	5	1.00	
318	2	2	2	4	2	2	3	2	1	0.44	5	4	5	4	5	4	5	4	5	5	0.92	
322	0	0	0	0	0	0	0	0	0	0.00	0	0	0	0	0	0	0	0	0	0	0.00	
324	5	5	5	5	5	5	5	5	5	1.00	2	2	2	2	2	3	3	3	0	0	0.38	
326	5	5	5	5	5	5	4	5	4	0.96	1	3	0	0	0	0	0	0	0	0	0.08	
Sterimol descriptor set											Sterimol descriptor set											
Leaf	1	2	2	3	3	4	4	5	5	Fraction	Leaf	1	2	2	3	3	4	4	5	5	Fraction	
Cost	0.83	3	1.6	3	1.6	3	1	3	1.6	3	Cost	1	1.2	1	1.2	1	1.2	0.71	1.8	0.83	1.8	Fraction
183	4	3	4	4	4	3	4	3	4	0.72	5	5	5	3	4	3	4	4	5	5	0.86	
294	1	1	0	1	0	1	0	0	0	0.08	0	1	0	0	0	0	0	0	0	0	0.02	
295	2	2	1	2	1	2	2	1	2	0.32	0	1	1	1	1	4	2	1	2	1	0.28	
296	3	0	1	1	1	0	3	0	2	0.22	4	4	4	4	5	4	5	5	5	5	0.90	
297	3	3	4	3	4	3	4	3	4	0.68	5	4	5	4	5	4	5	5	5	5	0.94	
299	3	0	0	1	0	0	1	0	1	0.12	4	4	4	4	5	4	5	5	5	5	0.90	
301	4	5	3	5	3	5	3	5	3	0.82	5	5	5	4	5	4	5	5	5	5	0.96	
302	3	4	2	4	2	4	2	4	2	0.62	5	5	5	4	5	4	5	5	5	5	0.96	
318	5	5	4	5	4	5	4	5	4	0.92	5	5	5	4	5	4	5	5	5	5	0.96	
322	1	1	0	1	0	1	1	0	1	0.12	0	1	0	0	0	0	0	0	0	0	0.02	
324	5	5	5	4	5	4	5	5	5	0.96	0	1	0	0	0	0	0	0	0	0	0.02	
326	3	5	4	5	4	5	5	5	5	0.92	1	1	0	1	0	1	1	0	1	0	0.12	
Both descriptor sets											Both descriptor sets											
Leaf	1	2	2	3	3	4	4	5	5	Fraction	Leaf	1	2	2	3	3	4	4	5	5	Fraction	
Cost	0.83	3	1.6	3	1.6	3	1	3	1.6	3	Cost	1	1.2	1	1.2	1	1.2	0.71	1.8	0.83	1.8	Fraction
183	5	5	4	4	4	4	5	2	3	0.76	5	5	5	5	5	4	5	3	5	3	0.90	
294	1	1	1	1	2	0	3	1	1	0.22	3	1	0	1	1	0	3	0	1	0	0.20	
295	3	3	3	3	3	3	2	2	4	0.56	5	2	0	1	1	0	4	0	0	0	0.26	
296	3	2	1	1	2	1	5	3	2	0.46	5	0	3	0	2	3	5	4	5	4	0.62	
297	2	1	3	2	4	2	4	2	4	0.52	4	2	5	4	5	4	5	4	5	5	0.86	
299	3	1	2	1	2	2	4	2	2	0.44	3	1	3	1	3	3	5	4	5	4	0.64	
301	4	4	4	4	4	4	4	2	3	0.72	4	4	5	4	5	3	5	4	4	4	0.84	
302	3	1	2	1	2	3	5	4	5	0.60	3	2	3	1	3	3	5	4	5	4	0.66	
318	4	4	2	3	3	3	5	5	4	0.76	5	3	4	3	3	4	5	5	5	5	0.84	
322	1	1	1	1	1	0	1	3	1	0.22	0	1	0	1	1	0	0	0	0	0	0.06	
324	5	4	5	5	5	5	4	5	5	0.96	5	2	0	1	1	0	4	0	0	0	0.26	
326	4	4	4	4	4	3	4	2	5	0.72	1	1	1	1	1	1	0	1	3	1	0.22	

## Summary

### *A Computational Study of the Substrate Conversion and Selective Inhibition of Aldosterone Synthase*

When a functional or structural impairment of cardiac output has occurred, the cardiovascular system will attempt to compensate for the reduced blood flow. Unfortunately, many of the resulting processes, such as the renin angiotensin aldosterone system, will progressively weaken the heart, resulting in the condition called heart failure. The renin angiotensin aldosterone regulatory system is currently targeted with medicine for heart failure. Many successes for the prolongation of patient age have been achieved by inhibition of angiotensin II synthesis and action. It has become apparent that this approach is suboptimal. Antagonists of aldosterone have provided better treatment options, however, side-effects are still observed. In the search for an alternative therapeutic application, we have studied a novel treatment involving the selective inhibition of aldosterone biosynthesis. The scope of this study has involved the *in silico* design and prediction of novel inhibitors, the synthesis of these inhibitors and analogues, and finally the *in vitro* measurement of their potency.

The biosynthesis of aldosterone is performed by two cytochrome p450 enzymes, 11B1 and 11B2, denoted as CYP11B1 and CYP11B2, respectively. From these two family members, only CYP11B2 can perform the final synthesis step that converts 18-hydroxycorticosterone into aldosterone. CYP11B1 performs the synthesis of glucocorticoids that are responsible for metabolic, immunologic and homeostatic functions. Because these glucocorticoid actions should not be inhibited, the newly designed medicine must be CYP11B2 selective. Since CYP11B1 is highly homologous to CYP11B2, we have performed an *in silico* study that allows us to model the interactions of substrates and inhibitors in both the active sites of CYP11B1 and CYP11B2.

Using comparative modelling, we have constructed models for the three dimensional architecture of both proteins. These models have been validated by investigating the torsional properties of the protein backbone and residue side chains, the overall protein packing and the dynamic behaviour of the protein models. Subsequently, the models have been used to evaluate the binding mechanisms and conversion mechanisms for the natural steroidal ligands of CYP11B1 and CYP11B2. A hypothetical binding mode has been proposed for 18-hydroxycorticosterone in CYP11B2, featuring the presence of stabilising hydrogen bonding interactions required for its conversion. Quantum mechanical analyses on the conversion of the steroids involved have shown a favourable conversion for this conformation, thereby supporting our hypothesis. In addition, the quantum mechanical analyses have provided insights on steroid conformations in the active sites during conversion.

The suitability of the protein models for inhibitor design has been tested by subjecting the models to a case study with four known inhibitors of CYP11B1 and CYP11B2. Using molecular dynamics and molecular docking, the inhibitor potencies for CYP11B1 and CYP11B2 have been predicted, and their interactions with the proteins have been evaluated. The trends in inhibitor potency found by these computational methods have been confirmed by *in vitro* inhibition measurements. As a next step, the molecular docking study has been expanded to improve the confidence in the predictive power of the models. Using the protein states evaluated by the molecular dynamics study, the molecular docking results of inhibitor analogues have been investigated and the predictive power of the models has been qualitatively improved.

In a final approach, we have performed a ligand-based investigation of the inhibitor analogues to determine which ligand characteristics are important for the potency for CYP11B1 and CYP11B2. To this end, we have conducted decision tree analyses on the physico-chemical properties of inhibitor substituents, resulting in a collection of descriptors that can be used for the prediction and design of novel inhibitors.

We have shown that a combination of synthesis, molecular modelling and experimental measurements form a promising approach towards the design of potentially new inhibitors.

## Samenvatting

Wanneer de functie of structuur van het hart wordt verstoord, verlaagt de lichaamsdoorbloeding. Als reactie hierop zal het cardiovasculaire systeem de verminderde bloedtoevoer proberen te compenseren door herstelmechanismen in te schakelen zoals het renine angiotensine aldosteron systeem. Helaas leiden de herstelmechanismen er bij grote en langdurige stoornissen toe dat de werking van het hart geleidelijk aan wordt verzwakt, wat resulteert in de aandoening genaamd hartfalen. Huidige therapieën voor hartfalen zijn dan ook gericht op de inhibitie van het renine angiotensine aldosteron systeem. Er zijn al veel successen geboekt met de inhibitie van angiotensine II biosynthese en activiteit, maar het is ook aangetoond dat deze aanpak niet optimaal is. Antagonisten van de aldosteron activiteit hebben geleid tot betere behandelingsopties, maar bijwerkingen worden nog steeds waargenomen. Daarom hebben we een alternatieve therapeutische toepassing onderzocht die zich richt op de inhibitie van de aldosteron biosynthese. Het doel van deze studie omvat het *in silico* ontwerp en voorspelling van nieuwe inhibitoren, de synthese van deze stoffen, en ten slotte de *in vitro* activiteitsbepaling van deze stoffen.

De biosynthese van aldosteron wordt uitgevoerd door twee cytochroom P450 eiwitten, 11B1 en 11B2, aangeduid als CYP11B1 en CYP11B2. De eerste stappen in de synthese van aldosteron kunnen door allebei de familieleden worden uitgevoerd. Echter, de laatste stap, de omzetting van 18-hydroxycorticosterone tot aldosteron, kan alleen worden uitgevoerd door het familielid CYP11B2. CYP11B1 is verantwoordelijk voor de synthese van glucocorticoiden die belangrijk zijn in metabole, immunologische en homeostatische processen. Omdat deze glucocorticoid functies niet belemmerd mogen worden, moet het nieuw ontworpen medicijn selectief zijn voor CYP11B2.

CYP11B1 en CYP11B2 vertonen een grote homologie in de aminozuursequentie. Daarom hebben we de driedimensionale eiwitstructuur van CYP11B1 en CYP11B2 gemodelleerd met behulp van de *in silico* techniek 'comparative modelling'. De 3D modellen zijn gevalideerd op basis van de torsie-eigenschappen van het eiwitskelet en de aminozuren, de totale eiwitstructuur en het dynamische gedrag van de eiwitmodellen. Vervolgens zijn de binding van substraten en inhibitoren in de active site van beide eiwitten onderzocht.

Met de resultaten van dit onderzoek naar het bindingsmechanisme van de natuurlijke substraten hebben we een CYP11B2 specifieke binding van 18-hydroxycorticosterone voorgesteld die nodig is voor de voltooiing van de synthese van aldosteron. Deze zogenaamde binding mode bevat een stabiliserende interne waterstofbrug die de steroïde in een belangrijke conformatie vasthoudt zodat het verder kan worden omgezet. Kwantum mechanische berekeningen aan de steroïden en hun verschillende mogelijke conformaties hebben een gunstige omzetting voorspeld voor de door ons voorgestelde conformatie. Daarbij ondersteunen deze berekeningen onze hypothese van de CYP11B2 specifieke aldosteron synthese. Bovendien hebben de kwantum mechanische berekeningen ons inzicht in de interactie tussen eiwit en substraat vergroot.

De geschiktheid van de eiwitmodellen voor het ontwerp van nieuwe inhibitoren is getest door de interacties te onderzoeken tussen de eiwitten en vier bekende inhibitoren van CYP11B1 en CYP11B2. Met behulp van moleculaire dynamica en moleculaire docking zijn de eiwit-inhibitor interacties geëvalueerd en is de interactie sterkte van elke inhibitor voorspeld. De *in vitro* activiteitsbepalingen hebben de trend in de voorspelde inhibitie bevestigd. Vervolgens is de docking studie uitgebreid om vertrouwen te krijgen in de voorspellende kracht van de modellen. Verschillende eiwitconformaties, verkregen uit de moleculaire dynamica studie, zijn meegenomen in de docking van nieuwe inhibitoren. Hierdoor is de voorspellende kracht van de modellen kwalitatief verbeterd.

Als laatste hebben we de inhibitoren onderzocht om te bepalen welke kenmerken van belang zijn voor hun activiteit op CYP11B1 en CYP11B2. Daarvoor is een zogenaamde beslisboom analyse uitgevoerd naar de fysisch-chemische eigenschappen van de inhibitor substituenten. Hieruit volgde een verzameling van eigenschappen en hun gewenste waardes, welke gebruikt kunnen worden voor de voorspelling en het ontwerp van nieuwe inhibitoren.

Deze studie laat zien dat een combinatie van synthese, moleculair modelleren en experimentele metingen, een veelbelovende en succesvolle aanpak is voor het ontwerp van nieuwe inhibitoren.

## Dankwoord

Bij dezen zou ik graag iedereen willen bedanken die een bijdrage heeft geleverd aan het tot stand komen van mijn proefschrift. Allereerst is dit de gehele CYP project groep waarmee we het STW project hebben uitgevoerd. Vanuit de chemie hoek zijn dit Henk en Joris van SyMO-Chem, en Ralf en Erica van Organon, die mij tijdens de maandelijkse project-bijeenkomsten meer inzicht hebben verschaft in de mogelijkheden op het gebied van synthese. Daarnaast zijn dit vanuit de farmacologie hoek Jos, Judith, Ilona en Eveline van de UM en Marcel van Organon. Mede dankzij hen heb ik geleerd de waarde van *in vitro* en *in vivo* testsystemen in te schatten. Verder wil ik Koen en Peter van de TUE bedanken voor hun begeleiding van het modelleerwerk. Van de UM kunnen uiteraard Dirk, Harry en vooral Rob niet onvermeld blijven, aangezien zij de grootste invloed hebben gehad in de ontwikkeling van mijn wetenschappelijke carrière. Als laatste wil ik de gebruikerscommissie bedanken voor hun waardevolle suggesties tijdens de halfjaarlijkse bijeenkomsten.

Naast de CYP project groep zijn er ook mensen geweest die mij geholpen hebben in het toepassen van de methodes die gebruikt zijn in dit proefschrift. Ik zou graag Rita en Jan van Organon en Huub van de TUE willen bedanken voor hun hulp met de beslisboom analyse. Voor de MOBbers kan ik alleen maar zeggen dat hun input tijdens de 'wekelijkse' discussies zeer nuttig zijn gebleken. Volgens mij is er trouwens nog steeds geen consensus over waar de afkorting MOB nu eigenlijk voor staat. Van deVGen wil ik Sandro bedanken voor zijn tips met betrekking tot mijn eerste publicatie, en 'De Jan' voor zijn begeleiding tijdens onze samenwerking op het modellingsgebied.

Ik heb ook het genoegen gehad om studenten te begeleiden die op hun beurt hun steentje hebben bijgedragen aan het project. Bedankt Pieter, voor de opzet van het QM werk en Bram voor het uitzoeken van de Fukui analyse. Daarnaast wil ik ook Marijn bedanken voor zijn aandeel in het MD stuk. Ik wens hun allen een leerzame academische toekomst toe.

Op de TUE heb ik mij ook kostelijk vermaakt tijdens de lunchtafel gesprekken. Mogen de vele quotes nergens herhaald worden, ook al zijn ze opgeschreven. Bedankt collega's voor de interesse en ondersteuning. Koen, nogmaals bedankt voor het maken van de cover. In het bijzonder heb ik veel kunnen lachen met mijn collega's en vrienden, Sander en René. Ik vind het heel tof dat jullie mijn paranimfen willen zijn en ik hoop dat we nog veel pret zullen beleven. Hopelijk verliezen we elkaar niet uit het oog.

Verder wil ik mijn familie en vrienden bedanken voor hun interesse en ondersteuning tijdens mijn onderzoek. Als allerliefste, wil ik mijn vriendin Diana bedanken voor het geduld dat ze zich heeft weten op te brengen in de periodes dat ik door Word gefrustreerd ben geraakt. Ook heeft ze altijd veel belangstelling getoond in mijn werk en het tot in de details met mij bediscussieerd.





## Publications

### Refereed Journal Publications

*The work described in Chapters 4 and 5 will be submitted to the Journal of Medicinal Chemistry before the promotion date (October 1, 2008).*

*The work described in Chapter 6 has been submitted to the Journal of Physical Chemistry B as:*

L. Roumen, B. van Hoof, K. Pieterse, P.A.J. Hilbers, R. Plate, E.M.G. Custers, M. de Gooyer, J.F.M. Smits, I. Beugels, J. Emmen, H.C.J. Ottenheijm, D. Leysen, J.J.R. Hermans, "Quantum mechanical considerations on the mechanism of the multistep conversion of 11-deoxycorticosterone to aldosterone by cytochrome P450 isoenzymes"

M. Minnaard-Huiban, J.M.A. Emmen, L.Roumen, I.P.E. Beugels, G.M.S. Cohuet, H. van Essen, E. Ruijters, K. Pieterse, P.A.J. Hilbers, H.C.J. Ottenheijm, R. Plate, M.E. de Gooyer, J.F.M. Smits, J.J.R. Hermans, "Fadrozole reverses cardiac fibrosis in SHHF rats: discordant enantioselectivity versus reduction of plasma aldosterone", J. Endocrinol., 2008, 149, 1, 28-31

L. Roumen, M.P.A. Sanders, K. Pieterse, P.A.J. Hilbers, R. Plate, E. Custers, M. de Gooyer, J.F.M. Smits, I. Beugels, J. Emmen, H.C.J. Ottenheijm, D. Leysen, J.J.R. Hermans, "Construction of 3D models of the CYP11B family as a tool to predict ligand binding characteristics", J Comput-Aided Mol Des, 2007, 21, 8, 455-471

### Refereed Conference Proceedings

L. Roumen, K. Pieterse, P.A.J. Hilbers, R. Plate, E. Custers, M. de Gooyer, J.F.M. Smits, I. Beugels, J. Emmen, H.C.J. Ottenheijm, D. Leysen, J.J.R. Hermans, "Benzyl-imidazoles: Potent and Selective Inhibitors of Aldosterone Synthase (CYP11B2)"  
34th Meeting of the International Aldosterone Conference, San Francisco, USA, 2008

J.M.A. Emmen, I.P.E. Beugels, L. Roumen, J.J. Debets, H. van Essen, P.A.J. Hilbers, K. Pieterse, H.M. Janssen, J.W. Peeters, H.C.J. Ottenheijm, R. Plate, M.E. de Gooijer, J.F.M. Smits, J.J.R. Hermans, "Are CYP11B2 (Aldosterone-synthase) Inhibitors Anti-fibrotic Agents in the Heart? Studies on a Novel CYP11B2 Inhibiting Compound in SHHF Rats"  
34th Meeting of the International Aldosterone Conference, San Francisco, USA, 2008

### Conference Posters

J.M.A. Emmen, I.P.E. Beugels, L. Roumen, J.J. Debets, H. van Essen, P.A.J. Hilbers, K. Pieterse, H.M. Janssen, J.W. Peeters, H.C.J. Ottenheijm, R. Plate, M.E. de Gooijer, J.F.M. Smits, J.J.R. Hermans, "*Are CYP11B2 (Aldosterone-synthase) Inhibitors Anti-fibrotic Agents in the Heart? Studies on a Novel CYP11B2 Inhibiting Compound in SHHF Rats*"  
Annual Endocrine Conference, ENDO08, San Francisco, USA, 2008

L. Roumen, B. van Hoof, K. Pieterse, P.A.J. Hilbers, R. Plate, E.M.G. Custers, M. de Gooyer, J.F.M. Smits, I. Beugels, J. Emmen, H.C.J. Ottenheijm, D. Leysen, J.J.R. Hermans, "*Quantum Mechanical Considerations on the Mechanism of the P450 Conversion of 11-deoxycorticosterone to Aldosterone*", Drug Metab. Revs. 2008, 40, S56-S57  
10th ISSX European Regional Meeting, Vienna, Austria, 2008

L. Roumen, K. Pieterse, P.A.J. Hilbers, R. Plate, M. de Gooyer, J.F.M. Smits, H.C.J. Ottenheijm, D. Leysen, J. Emmen, J.J.R.M. Hermans, "*Construction of 3D models for Cyp11B1 and Cyp11B2 based on substrate regio-selectivity and inhibitor enantioselectivity*"  
FIGON Dutch Medicine Days, Lunteren, The Netherlands, 2007

P.W. van Grootel, L. Roumen, P.A.J. Hilbers, G. Schaftenaar, J. de Vlieg, "*Theoretical Analysis of hydroxylation by Cytochrome P450 enzymes*"  
7th ICCS, Noordwijkerhout, The Netherlands, 2005

L. Roumen, E. Ruijters, P.A.J. Hilbers, J.F.M. Smits, D. Leysen, J.J.R. Hermans, "*Predicted stereoselectivity of the inhibition of Aldosterone Synthase and Cortisol Synthase*"  
FIGON Dutch Medicine Days, Lunteren, The Netherlands, 2005

L. Roumen, E. Ruijters, P.A.J. Hilbers, J.F.M. Smits, D. Leysen, J.J.R. Hermans, "*Predicted stereoselectivity of Fadrazole inhibiting Aldosterone Synthase and Cortisol Synthase*"  
Holst Memorial Lecture, High Tech Campus Eindhoven, The Netherlands, 2005

L. Roumen, E. Ruijters, J.J.R.M. Hermans, P.A.J. Hilbers, D. Leysen, H.C.J. Ottenheijm, J.F.M. Smits, M. de Gooyer, "*The Development of aldosterone synthase inhibitors: a concept to treat heart failure*"  
FIGON Dutch Medicine Days, Lunteren, The Netherlands, 2004

### Publication related to this work

J. van Buijtenen, B.A.C. van As, M. Verbruggen, L. Roumen, J.A.J.M. Vekemans, K. Pieterse, P.A.J. Hilbers, L.A. Hulshof, A.R.A. Palmans, E.W. Meijer, "*Switching from S- to R-selectivity in the Candida antarctica lipase B-catalyzed ring-opening of  $\omega$ -methylated lactones: tuning polymerizations by ring size*", J Am. Chem. Soc., 2007, 129, 23, 7393-7398

## Curriculum Vitae

Luc Roumen was born on August 15, 1979, in Geldrop, The Netherlands. He attended the Scholengemeenschap Philips van Horne, in Weert, where he obtained his Gymnasium diploma in 1997.

He studied Biomedical Engineering at the Eindhoven University of Technology from 1997 until August 2003. As part of his study he completed a six months traineeship and his graduation project at the Department of Molecular Design and Informatics at Organon NV, Oss, The Netherlands. The traineeship involved an investigation of the three-dimensional molecular structure of G-protein Coupled Receptors (GPCRs). The graduation project involved the construction of molecular models for the three-dimensional architecture of Cytochrome P450 (CYP) enzymes, as well as the *in silico* (virtual) screening for CYP inhibitors.

From October 2003 until February 2004, he worked as a scientist at the Department of Computer-Aided Molecular Design at deVGen NV, Ghent-Zwijnaarde, Belgium, investigating the binding characteristics of protein kinases.

On February 2004, he was invited to start his Ph.D. project at the Department of BioModeling and bioinformatics at the Eindhoven University of Technology. The work is part of an STW funded research project, which aims to develop selective inhibitors for Aldosterone Synthase, a Cytochrome P450 enzyme. The research project has been performed in close collaboration with Maastricht University (Department of Pharmacology and Toxicology), Organon NV (Department of Medicinal Chemistry, Department of Pharmacology) and SyMO-Chem BV.

This research has resulted in this thesis, on which he expects to receive his Ph.D. on October 1, 2008. In addition, the compounds designed in this study have been filed in a patent entitled "N-Benzyl Imidazole Derivatives", patent number 2008.053USPP, application number 61/032282. The filing date of this patent is February 28 2008 by N.V. Organon for the inventors, Luc Roumen, Joris Wilhelmus Peeters and Johannes Joseph Robertus Maria Hermans.

Starting September 2008, he will become a postdoctoral researcher at the Department of Medicinal Chemistry, Vrije Universiteit, Amsterdam.

## Notes









