DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Biosynthetic and Pharmacokinetic Approaches to Improve Steroid Therapeutics

Citation

Spady, Emma Sarah. 2019. Biosynthetic and Pharmacokinetic Approaches to Improve Steroid Therapeutics. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:42029556

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Biosynthetic and Pharmacokinetic Approaches to Improve Steroid Therapeutics

A dissertation presented

by

Emma Sarah Spady

to

The Committee on Higher Degrees in Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemical Biology

Harvard University

Cambridge, Massachusetts

April 2019

© 2019 – Emma Sarah Spady

All rights reserved.

Dissertation Advisor: Professor Pamela A. Silver Emma Sarah Spady

Biosynthetic and Pharmacokinetic Approaches to Improve Steroid Therapeutics

Abstract

Steroid hormone analogs are clinically important, but their use is limited by severe side effects. In my dissertation, I present two approaches to improve steroid drugs. The first approach, discussed in Chapter 2, aims to biosynthesize novel steroids. Steroid derivatives are difficult to make via traditional organic synthesis, but many enzymes regio- and stereo-selectively process a wide variety of steroid substrates. I expressed seventeen of these enzymes in mammalian cells, from which I selected the human cytochrome P450 CYP7B1 for use as a biocatalyst. HEK293 cells stably expressing CYP7B1 processed two non-native substrates into three novel products. The cells 7α- and 7β-hydroxylated 17α-hydroxypregnenolone, and 11α-hydroxylated 16α-hydroxyprogesterone. Two of these reactions were unanticipated, as CYP7B1 was thought to exclusively 7α-hydroxylate steroids. I explored how these products could have arisen in Chapter 3 using a Rosetta docking model. The model suggested that these substrates' D-ring hydroxyl groups prevents them from binding to CYP7B1 as neatly as the native substrate pregnenolone. This allows the non-native substrates to tilt, bringing different carbon atoms close to the active ferryl oxygen atom. In Chapter 4, I take a second approach to improving steroid drugs by considering whether a fusion protein could carry steroids to desired cells. I developed a multi-scale pharmacokinetics model to determine which features a glucocorticoid-binding antibody fusion protein requires to deliver steroid exclusively to leukocytes. The antibody's target antigen must endocytose quickly upon protein binding, but endosomal release of steroid or antibody is not helpful. The model also showed that the fusion protein could direct endogenous cortisol to leukocytes, achieving immunosuppression without any synthetic glucocorticoid.

Table of Contents

Acknowledgements v
Chapter 1: Introduction 1
1.1 The Importance of Steroid Drugs 1
1.2 Structural Features and Biological Activities of Endogenous and Synthetic Steroids 2
1.3 Steroid Signaling Mechanisms and their Consequences 9
1.4 The Promise of New Steroid Molecules 12
1.5 The Promise of Tissue-Targeted Glucocorticoids 17
1.5 Thesis Overview 21
Chapter 2: Mammalian Cells Engineered to Produce Novel Steroids 22
2.1 Attributions 22
2.2 Introduction 23
2.3 Results 24
2.4 Discussion 38
2.5 Methods 41
2.6 Acknowledgements 44
Chapter 3: Structural Models of CYP7B1 to Explain 7β- and 11α-Hydroxylations 45
3.1 Attributions 45
3.2 Introduction 45
3.3 Results 49
3.4 Discussion 55
3.5 Methods 57
3.6 Acknowledgements 59
Chapter 4: Pharmacokinetic Model of a Glucocorticoid-Binding Antibody Fusion Protein 60
4.1 Attributions 60
4.2. Introduction 60
4.3 Results 63
4.4 Discussion 71
4.5 Methods 73
Chapter 5: Conclusion 78
References 82
Appendix 1: Supplementary Tables, Figures, and File Index 96

Acknowledgements

First, I would like to thank my advisor, Pam Silver, who gave me the freedom to take my PhD in this direction. It was thanks to her patience that I was able to learn so much. Jeff Way has also been an enormous help; I hope I can one day emulate his protein structure intuition.

I would like to thank Tim Mitchison, the head of my dissertation advisory committee, who has given me excellent guidance since my arrival at Harvard. I also appreciate the advice from my committee members Kris Prather and Andrew Kruse. Thanks as well to the Laboratory of Systems Pharmacology, which funded me over the last four years, and to the collaborative, interdisciplinary community it fosters.

I am grateful for the colleagues I met in the Silver Lab, especially Cameron Myhrvold. From discussions during my rotation to proofreading my thesis, he helped with so many steps of my PhD journey and became a close friend along the way. Thanks also to Joe, Roger, Tobi, Shannon, Tim, Finn, Audam and all the Silverinos for thoughtful conversations and great times over the last six years.

I am indebted to all my friends for listening to me exclaim and complain about my research, and for taking my mind off it when needed. A special thanks to my girlfriend, Andi, whose love and cooking have nourished me throughout the writing process. Thanks as well to my best friend Lisa, my college friends Julia and Erika, and my Camberville friends Jules, Sarah, Mars, Molly, Jake and Gabe. Finally, I am lucky to have lived with the amazing women Lydia, Harlowe, and Laura during my time in Cambridge.

Most importantly, I must thank my family. Rose, you have always inspired me to do my best, whether by looking up to me as a child or by setting an example with your work ethic as a young adult. Dad, I am so proud to follow in your footsteps. Thank you for all your advice over the years. Mom, thank you for supporting and listening to me over innumerable dinners. Living near you has truly been a blessing. I love you all so much.

1

Introduction

1.1 The Importance of Steroid Drugs

Synthetic steroid drugs are critical in modern healthcare due to endogenous steroids' roles in the human body. Steroids are defined by their four-ringed structure, conventionally lettered A through D, and can be found in almost all eukaryotes. Cholesterol is the most common steroid in humans; it has 27 carbons, typically numbered as shown (Fig. 1.1). All animal cell membranes are 15-50% cholesterol, depending on the membrane location and cell type.[1,2] When this molecule's long fatty tail is oxidized or removed, the resulting amphiphilic steroid can diffuse through cell membranes. Hence, many animals use cholesterol-derived hormones to coordinate multi-tissue responses. Insects produce 20-hydroxyecdysone to regulate molting, while nematodes use dafachronic acid to regulate dauer formation.[3,4] Steroid hormones regulate sexual development, electrolyte balance, and long-term stress in most vertebrates, including humans.[5,6] Because of the many functions of endogenous steroid hormones, synthetic steroid drugs have diverse applications, from diuretics to anti-inflammatories to contraceptives. Of

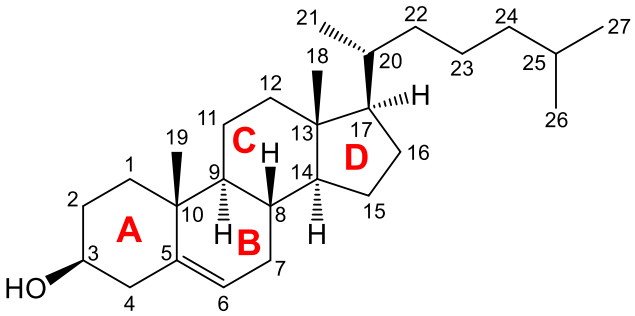


Figure 1.1. Conventionally-numbered structure of cholesterol. Red letters indicate rings A-D.

the 433 drugs on the WHO Essential Medicine List, 19 are steroids.[7] These drugs are incredibly potent, as they are active at concentrations below 10 nM.[8]

Glucocorticoids, also known as corticosteroids, are especially important as immunosuppressants. Around 1.2% of adults in the U.S. and 0.9 % of adults in the U.K. take oral glucocorticoids.[9,10] Most of these are long-term prescriptions to treat chronic conditions such as asthma (21% of total prescriptions), polymyalgia rheumatica / giant-cell arteritis (14%), chronic obstructive pulmonary disease (10%), and rheumatoid arthritis (6%).[10] Glucocorticoids are also used to treat lymphoid-cell-derived cancers, such non-Hodgkin's lymphoma, chronic lymphocytic leukemia, acute lymphoblastic leukemia, and multiple myeloma.[11] The mean duration of use in the U.S. is 4 years and 5 months, long enough for many severe side effects to manifest.[9] Osteoporosis and bone fractures are the most common, experienced by 21-30% of patients. Other common adverse events, affecting between 1% and 5% of glucocorticoid users, include cardiac conditions, cataracts, and gastrointestinal distress.[12] In addition, people on long-term glucocorticoid therapy are more than four times as likely to develop type 2 diabetes and hyperglycemia compared to controls.[12] Anti-inflammatory steroids with improved side effect profiles could therefore benefit millions of people worldwide.

1.2 Structural Features and Biological Activities of Endogenous and Synthetic Steroids

Humans produce five groups of tailless steroid hormones, usually categorized by the structural features that drive their binding to cognate receptors. Synthetic steroid drugs preserve these critical functional groups, while adding other useful features at positions tolerated by the

receptor. Progestins, exemplified by the endogenous progesterone, are structurally the simplest of these hormones. They have twenty-one carbons, with ketones at C-3 and C-20, as well as a C-4-5 double bond (Fig. 1.2A).[13] Progestins bind the progesterone receptor (PR), which primarily maintains pregnancy by coordinating endometrial tissue changes.[13] More recently, progestins have been found in the brain and spinal cord, where they play a role in neuroprotection.[14] As with all steroid hormones, functional group and ring structure stereochemistries are critical to receptor binding. As an example, 14-iso,17-isoprogesterone does not have gestational activity; PR binding evidently requires C-20 to be above the ring plane rather than below.[15] When a steroid is drawn in the conventional orientation, functional groups pointing away from the viewer are called α substituents, while groups pointing towards the viewer are β substituents. For example, one would describe the C-20 functional group of progesterone as having a β orientation. Synthetic drugs which activate PR are also called progestins, and PR antagonists are called antiprogestins. This naming pattern holds for all steroid receptor agonists and antagonists. Synthetic progestins,

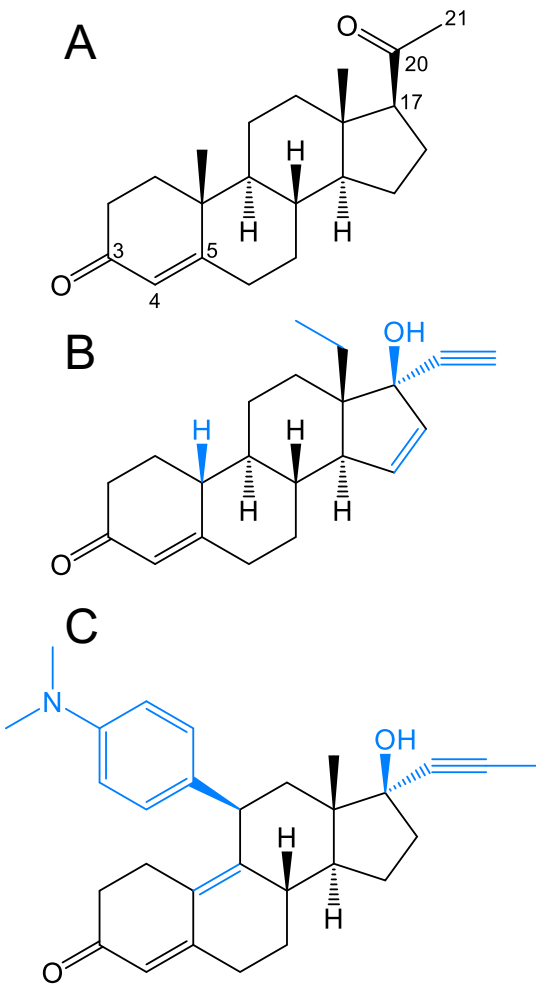


Figure 1.2. A) Progesterone, an endogenous human progestin. B) Gestodene, a synthetic progestin. C) Mifepristone, a synthetic antiprogestin. Differences between native progestins and analogues are in blue.

such as gestodene, are used as oral contraceptives. C-19 removal improves PR binding, and the D-ring modifications increase PR specificity over the other steroid receptors (Fig. 1.2B). Antiprogestins, such as mifepristone, can terminate early pregnancy by disrupting the uterine decidua (Fig. 1.2C).[16]

Androgens and estrogens govern sexual development in mammals. Dihydrotestosterone (DHT) is the most potent endogenous androgen in humans, though testosterone also binds to the androgen receptor (AR). Steroidal androgens generally have a β-hydroxyl at C-17, a ketone at C-3, and lack C-20 and C-21 (Fig. 1.3A). The removal of the C-3-4 double bond from testosterone to generate DHT improves AR binding.[17] Early in fetal development, AR activated by testosterone governs the differentiation of Wolffian ducts into the epididymis, vas deferens, and seminal vesicles. During male puberty, an increase in DHT drives secondary sexual characteristic development. Androgens also trigger erythropoetin synthesis, which increases red blood cell volume.[5,18] AR-activating anabolic androgens dominate the public image of steroids due to athletic doping, but have medical use in treating muscle wasting diseases and anemia. The differences between 7α-methyl-19-

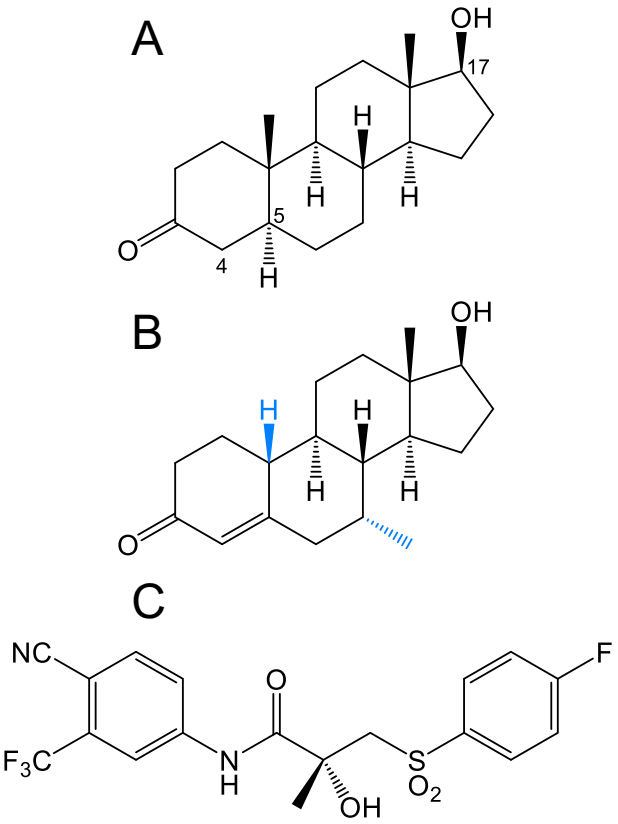


Figure 1.3. A) Dihydrotestosterone, an endogenous human androgen. B) 7α-methyl-19-nortestosterone, a synthetic androgen. C) Bicalutamide, a synthetic non-steroidal antiandrogen. Differences between native androgens and the analogue are in blue.

nortestosterone and testosterone prevent the former from converting to DHT or estrogens, decreasing its effect on prostate and breast tissue (Fig. 1.3B).[19] Antiandrogens are used in prostate cancer treatment, as AR activation typically drives cell proliferation. These antagonists, such as bicalutamide, are often non-steroidal (Fig 1.3C).[19]

Human estrogens are synthesized from androgens through the aromatization of the A-ring and removal of the C-19 carbon. The C-3 and C-17 hydroxyls and the flattened A ring are essential for 17β-estradiol binding to the two estrogen receptors α and β (ERα/ERβ) (Fig. 1.4A).[20] ERα is responsible for breast and uterine tissue proliferation, but both ERs are required for normal ovulation. ERs also play a role in prostate, lung, and brain morphogenesis.[21] Finally, ER activation is implicated in bone maintenance, as ER mutations often result in osteoporosis.[22] It should be noted that all human gonads and adrenal cortices make both androgens and estrogens, though males and females typically have differing ratios.[5] The most famous ER-modulating drugs act as activators in one cell type and repressors in another; the mechanisms behind this will be discussed shortly.[23] Raloxifene, used in treating breast cancer due to its anti-proliferative effects, inhibits ER in breast tissue while activating ER

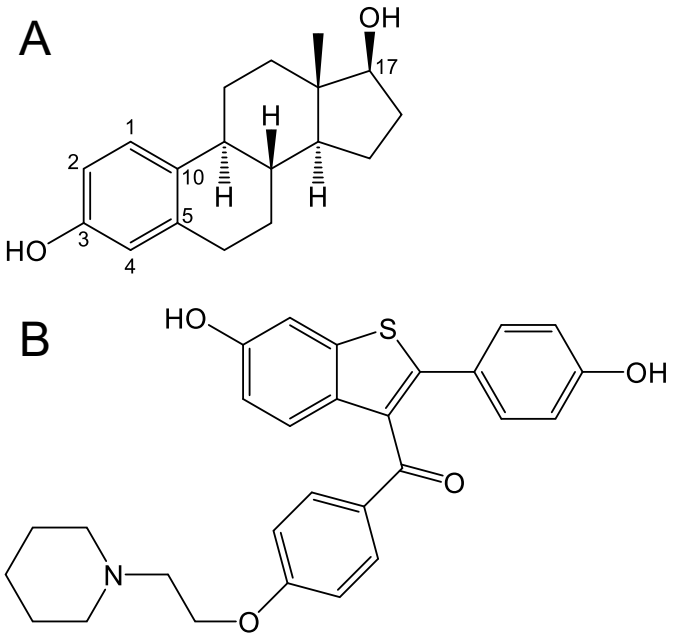


Figure 1.4. A) 17β-estradiol, an endogenous human estrogen. B) Raloxifene, a synthetic estrogen receptor modulator.

in bone (Fig. 1.4B).[24] The latter effect renders raloxifene helpful in preventing osteoporosis in postmenopausal women.[24]

Glucocorticoids are a class of 21-carbon steroids that alter human stress response through glucocorticoid receptor (GR) binding. Cortisol, the main endogenous GR activator, has 11β, 17α, and 21-hydroxylations in addition to the features of progesterone (Fig. 1.5A). GR activation causes increased blood glucose, decreased bone and muscle maintenance, and decreased inflammatory responses, all which are adaptive under stressful conditions in the short-term. Glucocorticoids are also essential to embryonic lung development, and have less-understood roles in cardiovascular health and fear behaviors.[25] As mentioned previously, synthetic glucocorticoid agonists are used as immunosuppressants. Dexamethasone features a C-1-2 double bond to make the A ring more rigid, preventing an entropy decrease upon steroid binding (Fig. 1.5B). The double bond also facilitates a hydrogen bond between GR and the C-3 ketone. Dexamethasone's methylation at C-16 decreases mineralocorticoid receptor affinity while increasing hydrophobic surface area, and its fluorination at C-9 slows degradation.[26]

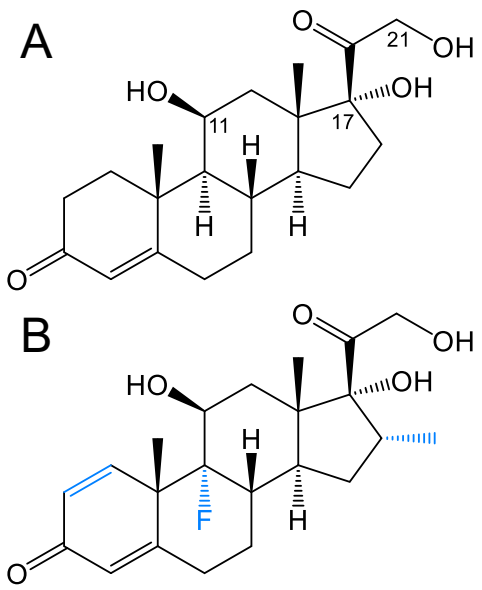


Figure 1.5. A) Cortisol, an endogenous human glucocorticoid. B) Dexamethasone, a synthetic glucocorticoid. Differences between native glucocorticoids and the analogue are in blue.

Endogenous mineralocorticoids are derived from glucocorticoids, and are best known for regulating blood pressure. Aldosterone, the primary human mineralocorticoid, prominently features a C-18 ketone. It lacks the 17α-hydroxylation found in cortisol, but retains the 11β- and 21-hydroxylations, which are essential for its mineralocorticoid receptor (MR) binding (Fig. 1.6A). 11-ketones and protruding groups below the D ring, which are found in endogenous glucocorticoids, decrease MR affinity.[27] The MR increases blood pressure when activated, primarily by stimulating sodium reabsorption in the kidney. It has effects on vasculature throughout the body, encouraging inflammatory and proliferatory states in blood vessel endothelial cells. Hence MR stimulation promotes vasoconstriction, atherosclerosis, and fibrosis.[28] Spironolactone, a synthetic MR antagonist, is used as a diuretic to treat hypertension and heart failure (Fig. 1.6B).[29] Its inhibitory properties can be traced to its 7α-thioester.[27] Despite its widespread use, spironolactone is not particularly selective; it inhibits PR and AR, but to a lesser extent than MR.[29]

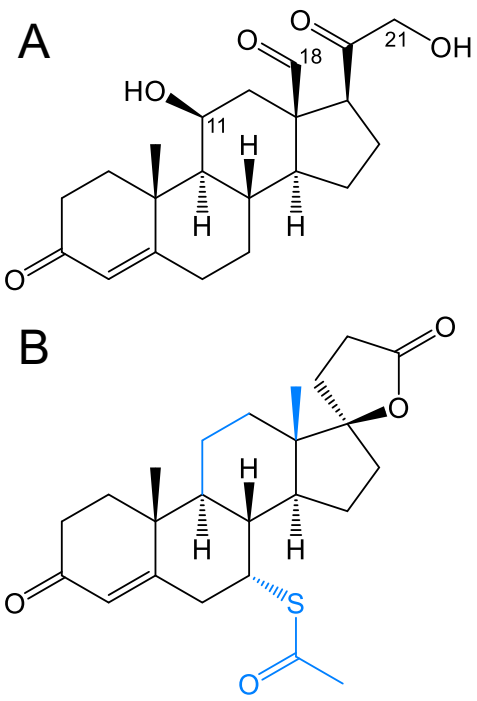


Figure 1.6. A) Aldosterone, an endogenous human mineralocorticoid. B) Spironolactone, a synthetic antimineralocorticoid. Differences between native mineralocorticoids and the analogue are in blue.

More recently, the signaling functions of sterols in humans have come to light. Generally, the term ‘sterol’ refers to tailed steroids, and ‘oxysterol’ to sterols with hydroxylations beyond the 3β-hydroxyl of cholesterol. Sterols with hydroxylations at carbons 7, 22, 24, 25, and/or 27 are found in human blood, the most abundant of which is 27-hydroxycholesterol (Fig. 1.7).[30,31]

Many of these oxysterols bind liver X receptors α and β (LXRα/LXRβ) and retinoic acid receptor-related orphan receptor α and γ (RORα/RORγ); they activate or antagonize the receptor depending on the steroid structure and cellular context. LXRα is primarily expressed in the liver, and initiates liver cholesterol uptake when activated to maintain lipid homeostasis.[30] LXRβ is expressed in many cell types, including epithelial tissues that transport water, leukocytes, and some neurons. It has neuroprotective and immunosuppressive effects when stimulated.[32,33] The RORs regulate cholesterol metabolic enzymes, and RORγ plays an unclear role in T helper 17 cells.[30] LXRβ-selective agonists are in development as immunosuppressive therapeutics, but all are non-steroidal.[33]

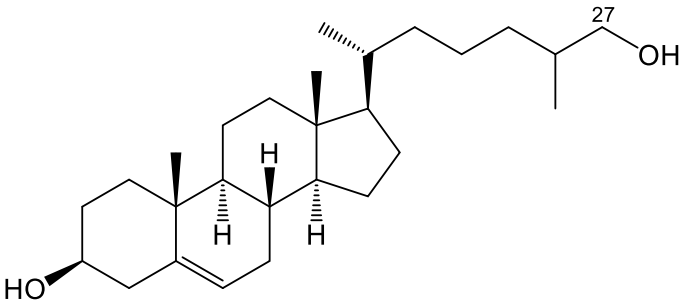


Figure 1.7. 27-hydroxycholesterol, an abundant human oxysterol.

Bile acids, originally thought to simply be cholesterol degradation products that aid in digestion, have important signaling roles. The primary human bile acids, cholic acid and chenodeoxycholic acid, are synthetized in the liver, amidated with taurine or glycine, and released into the intestines to solubilize fats (Fig. 1.8A).

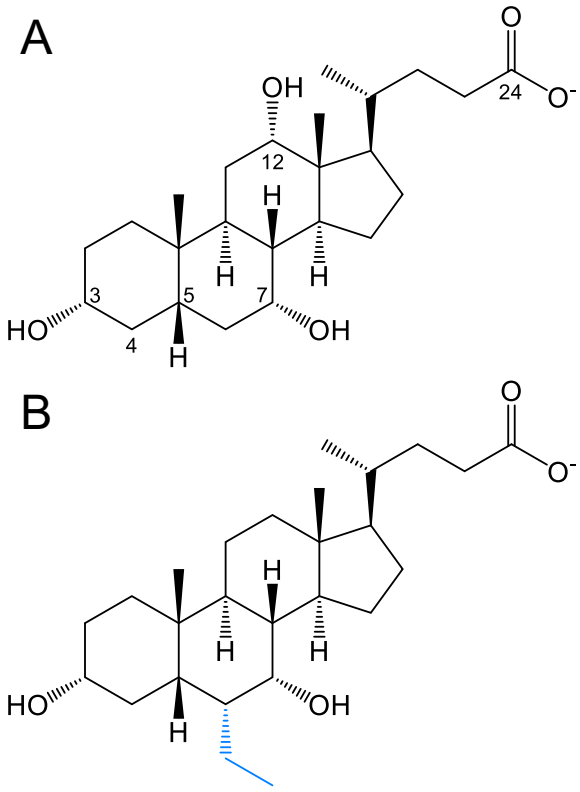


Figure 1.8. A) Cholic acid, a primary human bile acid. B) Obeticholic acid, a synthetic FXR agonist. Differences between native FXR agonists and the analogue are in blue.

These molecules have a hydrophobic β face and a hydrophilic α face, allowing them to act as detergents.[34] While the vast majority of bile acids remain in the digestive system, some leak into the bloodstream, where they signal through the farnesoid X receptors (FXRα/FXRβ) and TGR5. The FXRs govern lipid and glucose homeostasis; activation increases liver steatosis and blood triglyceride levels, and decreases blood glucose. TGR5 activation is immunosuppressive and drives thyroid hormone production, which stimulates energy expenditure.[35] The synthetic FXR agonist obeticholic acid was developed to treat primary biliary cholangitis, an autoimmune disease that destroys bile ducts in the liver (Fig. 1.8B). Given the other effects of bile acids, FXR agonists are being tested for treatment of non-alcoholic steatohepatitis, type 2 diabetes, and inflammatory bowel disease.[36]

1.3 Steroid Signaling Mechanisms and their Consequences

Canonically, endogenous and synthetic steroids exert their diverse effects by binding intracellular steroid receptors. Steroid hormones are amphiphilic, and thus can pass through cell membranes to reach their cytosolic receptors. These receptors are members of the nuclear receptor (NR) superfamily, and include the PR, AR, ERα/β, GR, MR, LXRα/β, and FXRα/β.[37] Upon steroid binding, cytosolic receptors are phosphorylated and enter the nucleus to act as transcription factors. They bind DNA sequences called hormone responsive elements, where they recruit coactivator proteins.[38] Coactivators bind using the receptor's NR box motif, which is only revealed upon ligand binding.[39] The coactivators, in turn, recruit transcriptional machinery or histone-modifying proteins to upregulate target gene transcription.[40] For some NRs, ligand-unbound receptors are also present in the nucleus, where they recruit corepressors to target DNA sequences. The ligand-unbound NR conformation reveals a CoRNR box motif, which the

corepressors bind.[39] In addition, steroids can act non-genomically, triggering cellular responses on the timescale of seconds rather than hours. These pathways are not as well-studied; some involve the usual activated steroid receptors interacting with kinases, while in others, steroids bind G protein-coupled receptors or other proteins.[41–43]

Different cell types respond differently to the same steroid through differences in coregulators, nuclear receptor phosphorylation, and other transcription factors. As one might expect from the canonical mechanism, the coactivators and corepressors available to steroid receptors in a cell affect the genes the receptor can regulate.[40,44] Post-translational modifications to these coregulators can even allow proteins generally thought of as corepressors, such as MTA1, to act as coactivators in different contexts.[45] Steroid hormone receptors can be phosphorylated independently of ligand binding, altering their intracellular location, chromatin binding, and coactivator recruitment. Thus changes in kinase levels under different conditions or in different cell types can result in different steroid activity profiles.[46] For example, GR serine 134 is phosphorylated by p38 mitogen-activated protein kinase under many cellular stressors, with or without glucocorticoid present. While phosphorylation-deficient mutant GR still translocates to bind chromatin upon glucocortioid activation, the lack of serine 134 phosphorylation dramatically altered its 14-3-3ζ cofactor recruitment and thus its transcriptional activity.[47] Finally, nuclear receptors may interact directly with other transcription factors, which vary dramatically between cells. The best-studied of these interactions is that of activated GR, which is generally anti-inflammatory, with the pro-inflammatory transcription factor NF-κB. GR can interfere with NF-κB binding DNA, and can disrupt the NF-κB transactivation domain to prevent NF-κB from recruiting transcriptional complexes.[48]

Though steroids diffuse easily through tissue, it is important to recall that endogenous steroid activity can be controlled by the small molecule's availability. Many steroid hormone receptors can bind non-cognate endogenous steroid ligands, which may agonize or antagonize the receptor's transcriptional action. Thus cells express enzymes that modify or degrade steroids that would trigger inappropriate receptor behavior.[49] Most famously, cortisol binds and activates the MR, and is much more abundant than the primary mineralocorticoid aldosterone. Kidney, heart, colon, and most other cells that express MR also express 11β-hydroxysteroid dehydrogenases, which oxidize the glucocorticoid 11β-hydroxyl to an 11-ketone. This disrupts cortisol MR binding while sparing the desired mineralocorticoid ligands.[50]

Because each class of steroid hormones coordinates a wide variety of biological activity, synthetic steroid drugs are plagued with on-target side effects. The mechanisms described above allow for one steroid binding to its cognate receptor to elicit very different responses in different tissues. However, steroid drugs are generally prescribed for only a handful of these responses; the rest become side effects. These can be called 'on-target side effects' because they stem from the small molecule drug binding its target protein, in this case a steroid hormone receptor.[51] For comparison, the more common 'off-target side effect' applies to small molecules that bind more than one protein. An example would be a non-specific kinase inhibitor binding both a desired kinase with positive effect and an undesirable kinase with a deleterious effect. On-target side effects are generally more difficult to combat, as changing the protein-small molecule interaction for greater target specificity cannot alleviate the problem.

These on-target side effects of steroid drugs limit their use in many applications. As previously mentioned, glucocorticoid-based immunosuppression brings with it increased blood glucose, decreased bone maintenance, and a variety of neurological effects.[25,52] Linking these

processes is adaptive under the natural stressful conditions that induce cortisol production. However, when synthetic glucocorticoids are used long-term, the immunosuppressive benefits are accompanied by type 2 diabetes and osteoporosis.[12] Similarly, endogenous androgens coordinate bone marrow stimulation, muscle gain, voice deepening, and body hair growth to trigger male puberty. Conventional androgen treatments help prevent muscle deterioration in children with Duchenne's muscular dystrophy, but their use is limited by their virilizing effects.[53] These issues, along with an increased risk of cardiovascular disease, prevent androgen use in other forms of wasting and in osteoporosis.[19] Progestins' effects on the uterus endometrium allow their use as female contraceptives and in prevention of endometrial hyperplasia. Nevertheless, progestins also increase breast cancer risk, decrease bone density, and increase blood pressure.[54]

1.4 The Promise of New Steroid Molecules

Steroids are a unique class of compounds in that a small 'chemical space' of structures can cover a great deal of 'biological space' in the cellular responses they elicit. Firstly, steroid receptors are very sensitive to structural differences in their ligands. There are few structural differences between endogenous steroids, such as cortisol or testosterone, and synthetic steroids with greatly improved properties, such as dexamethasone or 7α-methyl-19-nortestosterone. We can think of steroid hormones like these as being in a small 'chemical space' compared to all possible similarly-sized organic molecules; they have the same ring structure and many of the same functional groups. Nevertheless, these similar molecules will trigger very different responses in cells, even compared to the 'biological space' of all possible cellular drug responses.[55] A challenge in drug development is that while chemical libraries cover a large

'chemical space' of different structures, they may probe a smaller 'biological space' of similar cellular responses.[56] Thus synthesis of novel steroids brings with it an efficient exploration of 'biological space', and should be of pharmacological interest.

Thanks to the many processes regulated by steroid hormones, selective steroid receptor modulators (SSRMs) have been pursued for many years. The term SSRM is used for any small molecule that agonizes or antagonizes a steroid-binding nuclear receptor depending on the cell type in which it is found. Though humans have two estrogen, liver X, and farnesoid X genes, molecules that have different action on α versus β receptors are also considered SSRMs. In general, SSRMs that occupy a nuclear receptor's steroid binding site alter its coregulator recruitment. Upon binding, typical steroid receptor agonists allow the flexible alpha helix 12 (H12) to rest smoothly over them. This arranges the NR box to allow coactivator recruitment, resulting in the usual regulation of steroid-induced genes. An NR antagonist occupies the steroid binding site but blocks H12 from locking into place, exposing the corepressor-recruiting CoRNR box even in the presence of steroid agonists.[57] SSRMs partially distort the H12 position, which alters the exposure of the NR box and CoRNR box motifs. Different coregulators are more or less sensitive to these motif configurations. A cell in which the SSRM is an NR activator has plentiful or robust coactivators that are recruited to the receptor even with the distorted NR box position. Meanwhile, a cell in which the SSRM is an NR inhibitor is dominated by robust corepressors that recognize the partial exposure of the CoRNR box.[58] A third type of cell may have no robust regulators at all, which could result in a transcriptional state that is harder to characterize than active versus inactive receptor. One must keep in mind that this general SSRM mechanism does not apply to all of these molecules, which are defined by transcriptional profiles rather than by their biochemical mechanism.

Due to the importance of ER in breast cancer, more selective estrogen receptor modulators (SERMs) have been developed than for other steroid-binding nuclear receptors. This is partially facilitated by the existence of two ERs, α and β, with different expression in different tissues. The aim of most SERMs is to inhibit ER in breast and endometrial tissue, in order to prevent proliferation of cancers from these organs. However, the drug should spare or increase ER activity in bone, as otherwise patients may develop osteoporosis.[59] The oldest SERMs are tamoxifen and raloxifene, which are non-steroidal but bind to the same pocket as the native agonist estradiol. Tamoxifen is an ER antagonist in breast tissue and an agonist in bone and endometrium, while raloxifene is an ER antagonist in breast and endometrium, and only an agonist in bone. Both of these drugs activate ERβ, the dominant receptor in bone, upon binding. Their different effects are due to their ERα interaction. Tamoxifen activates ERα in endometrium due to its higher level of coactivator SRC-1; breast tissue has too little SRC-1 to enable this activity. Gene induction by raloxifene-bound ER is not affected by SRC-1 abundance, and thus raloxifene inhibits ER in both endometrium and breast.[60] Tamoxifen and raloxifene derivatives with improved side effect profiles have been developed, including clomifene, afimoxifine, bazedoxifene, and arzoxifene.[24] Fulvestrant is the only steroidal SERM, which acts exclusively as an ER downregulator.[61] It is an ERα antagonist, preventing its DNA binding and triggering its degradation via interaction with cytokeratins 8 and 18. However, fulverstrant-bound ERβ does not degrade, which results in a less thorough ER repression in bone.[62]

SSRMs for other steroid receptors, most of which are non-steroidal, are in clinical testing. They are far less studied than SERMs, and thus the precise mechanisms of their tissue selectivity are unclear. It suffices to say that they must differentially affect one of the many factors discussed above that play into steroid receptor signaling.[58] Selective androgen receptor

modulators (SARMs) are of interest in treating wasting conditions, and several are in clinical trials. They aim to activate AR in bone and muscle, but not in heart, prostate or liver tissue to avoid negative cardiovascular and virilizing side effects.[63] Examples include andarine and ostarine, which activate AR in all tissues, but much less so in prostate than in muscle.[19,64] Despite a large body of research, less progress has been made in selective glucocorticoid receptor modulators, which could act as immunosuppressants. These seek to activate GR in immune cells without liver or bone GR agonism.[65] This lack of progress is partially due to the inefficacy of GR agonists that prevent GR dimerization in the nucleus. These were developed due to the mistaken hypothesis that GR immunosuppression was primarily due to transcriptional repression, and that this repression takes place through monomeric DNA-bound receptor.[66] New SGRMs that are not dependent on this incorrect paradigm are in development, and include mapracorat, compound A, and ZK-245186.[67] Selective FXR and LXR modulators are in more preliminary research stages, but have promising applications in lipid diseases and obesity treatment.[36,58]

This lack of successful steroidal SSRMs may stem from the difficulty of synthesizing novel steroids using traditional organic chemistry methods. Given their biological properties, one might expect to find enormous combinatorial libraries of synthetic steroid drug candidates. However, this is not the case; the cyclohexane rings of most steroids make modification at many positions difficult.[68] Organic chemistry typically relies on functional groups to enable modification at specific carbons, with the groups' different properties permitting selective reactions. Steroids typically have many electronically similar secondary carbons on their cyclohexane and cyclopentane rings. Developing syntheses capable of discriminating between these sites to alter one C-H bond regio- and stereo-selectively can be done, but it is a nontrivial

task.[69,70] Once a steroid is determined to be of interest, it can be synthesized, but it is not worthwhile to undertake these complicated processes for an untested molecule. This is a prime opportunity to use biocatalytic techniques to diversify steroid reactions.

Despite the potential of biocatalytic processes in steroid synthesis, and the many characterized enzymes available, steroid-modifying enzymes are not used to make novel products. Enzyme expression can used to confirm biosynthetic pathways, as when cortisol was produced using six genes in mammalian cells.[71] Other research sought to prove that complex steroids could be synthesized *de novo* in non-native organisms. Heavily-engineered *Saccharomyces cerevisiae* can self-sufficiently produce hydrocortisone, pregnenolone, and progesterone.[72,73] This is remarkable, considering that yeasts have ergosterol in their membranes rather than cholesterol, the typical precursor to these human hormones. Human steroid-processing enzymes are sometimes expressed in yeast to characterize how naturally-occurring mutations affect enzyme activity.[74,75] Their purpose is generally not to synthesize interesting steroids, but to probe the enzyme. Certain enzymatic reactions are studied extensively, as they produce especially difficult functional groups required for known bioactive steroids. 11β-hydroxylation is most frequently pursued, as it is required for glucocorticoid receptor binding but cannot be produced synthetically. CYP11B1, the human steroid 11β-hydroxylase, has been expressed in *Schizosaccharomyces pombe* yeast for this purpose.[76] Fungal enzymes capable of 11-hydroxylation have also been expressed in model organisms, as current industrial methods use unmodified *Rhizopus*, *Aspergillus*, *Curvularia*, and *Cunninghamella* species as biocatalysts.[77,78] Unusual steroids are occasionally characterized as products from microorganisms, but these are not engineered systems.[79–82] Unfortunately, many of these enzymes are not pursued further, and thus remain unidentified.

1.5 The Promise of Tissue-Targeted Glucocorticoids

Steroid drug side effects can be limited by applying steroid directly to the desired tissue, as long as the steroid remains exclusively in this tissue. This approach is most commonly taken with anti-inflammatory glucocorticoids. Topical cortisol creams, which are called hydrocortisone creams, can take advantage of the skin's avascular nature to reduce inflammation. Glucocorticoids bind receptors in epidermal and dermal cells, and hence are soaked up before reaching the blood.[83] Steroid application to areas with more blood perfusion, such as the lungs or sinuses, can allow the drug to leak out. These tissues can be treated using inhaled steroids with labile functional groups, which quickly degrade into inactive metabolites after leaving the target.[84,85] Increased lipophilicity increases drug residence time in the lung, thus decreasing the dose required for effect. Examples of these drugs, from longest to shortest lung retention time, are fluticasone furoate, mometasone furoate, and ciclesonide.[86] Unfortunately, these approaches are not appropriate for many cell types where steroid drugs have therapeutic action. Rheumatoid arthritis treatment requires steroids to reach joints, and global immunosuppression requires steroids to remain in the blood to access leukocytes. Thus other methods of targeting glucocorticoids are needed.

Liposomes and organic covalently-modified nanoparticles have been used to target glucocorticoids.[87] Many types of nanomaterials naturally accumulate at sites of inflammation, a phenomenon known as the enhanced permeability and retention (EPR) effect. These particles are too large to leak out of healthy blood vessels, but are smaller than the pores that form in inflamed tissue vasculature.[88] Hence the EPR effect can passively target liposomes, micelles, and nanoparticles that slowly release glucocorticoids. Unmodified liposomes can deliver

prednisolone to inflamed rat paws with improved pharmacokinetic properties compared to prednisolone alone.[89] Liposomes modified with polyethylene glycol (PEG) have a longer serum half-life because they cannot be removed by the mononuclear phagocyte system; unfortunately, the PEG renders the liposomes immunogenic.[90] Liposomes can also be modified with antibodies to bring particles to specific cells. An anti-CD163-functionalized liposome can deliver dexamethasone to macrophages in a Parkinson's disease model.[91] Polymeric nanoparticles can be covalently functionalized with glucocorticoids to deliver them via the EPR effect, with different pharmacokinetics than liposomes. Covalent drug-polymer linkers that cleave under low pH can ensure that the steroid is not released until it is internalized by a cell, extending the drug half-life. One such example is the N-(2-hydroxypropyl) methacrylamide (HPMA) dexamethasone-conjugated nanoparticle developed to treat rheumatoid arthritis.[92] A similar HPMA-dexamethasone conjugate treated inflammation after joint replacement with decreased side effects compared to free dexamethasone.[93]

Various antibody-glucocorticoid conjugate designs have been tested for cell-type-specific steroid delivery.[94] Antibody-drug conjugates (ADCs) consist of an antibody covalently bound to a bioactive small molecule. The antibody component binds a target surface antigen, often referred to simply as 'the target', which is present exclusively on desired cells. The ADC binding triggers antigen internalization and release of the small molecule inside the desired cell.[95–97] A handful of ADCs have been approved as cancer therapeutics, and their potential as anti-inflammatory drugs has been recognized.[98] An anti-CD163-dexamethasone and an anti-CD74-fluticasone derivative are particularly promising. The anti-CD163-dexamethasone targets macrophages, and can decrease LPS-mediated inflammation at lower doses in rats than free dexamethasone. Treatment with an effective dose of the ADC does not decrease thymus weight,

unlike the steroid alone, indicating lymphocytes are spared.[99] Anti-CD163-dexamethasone also alleviates fatty liver pathology in rats fed high-fructose diets; free dexamethasone worsens the condition, likely due to its effects on hepatocytes.[100] An anti-CD74-fluticasone selectively triggers B cell and not T cell transcriptional responses characteristic of glucocorticoid exposure. This holds even when these B cells are co-cultured with T cells, thanks to an innovation that prevents free steroid from escaping into the culture medium. The phosphonate fluticasone derivative was developed to have decreased membrane permeability at pH 7 and increased permeability at pH 5.5, permitting the molecule to diffuse out of the endosome but not out of the target cell.[101]

Antibody-drug conjugates are pharmacokinetically complex. These multi-part drugs have many parameters that can be independently altered, resulting in a plethora of choices for the developer. Firstly, macromolecule drug size and shape affects tissue exposure. Drug entry into tissues of interest varies depending on protein hydrodynamic radius and the nature of the capillary endothelium.[102] Proteins of different sizes are eliminated at different rates by the kidneys, particularly those on opposite sides of an approximate 50 kDa threshold.[103] Hence the choice of a nanobody, minibody, or full antibody for the protein component must be carefully considered. Half-life can also be altered by keeping an Fc region in the antibody, permitting FcRn-mediated recycling.[102]

Other pharmacokinetic parameters are driven by the nature of the ADC's target antigen. The term 'target-mediated drug distribution' encompasses all of the ways in which the antibody binding affects the ADC's pharmacokinetics. If the target protein is abundant, and the antibody is high-affinity, the antibody-drug conjugate may concentrate on the edge of a tissue. This saturates the cells bordering vasculature while failing to treat cells further away.[104] In addition, a good

target internalizes quickly to bring the drug inside the cell before the antibody releases. If the target is generally recycled after internalization, the antibody or the small molecule must be released in the endosome, or else the conjugate will return to the cell surface.[105]

Finally, the small molecule component of the ADC must be considered. The method of small molecule attachment to the ADC can alter its distribution. A non-degradable linker, which only releases the small molecule upon degradation of the antibody, may place the molecule in an inappropriate cellular compartment. However, a biodegradable linker has the risk of releasing the small molecule too soon, exposing off-target cells.[106] The pharmacokinetics of the released small molecule can play a role as well. If the small molecule permeates the cell membrane, it can escape from targeted cells and affect other nearby cells. Dying cells may discharge the small molecule as well, particularly in the case of cytotoxic drugs. This bystander effect can be helpful, as in the case of solid tumors where not every cell expresses the target marker. Unfortunately, if the affected cells are not the intended targets, this effect can defeat the purpose of the ADC.[104,105]

Pharmacokinetic models could help narrow the design choices for glucocorticoid-targeting antibody drugs, particularly in early stages of development. Ordinary differential equations (ODEs) are generally sufficient to understand most pharmacokinetic systems. A set of chemical rate equations, compartment volumes, and initial concentrations can generate a system of equations that can be solved for the concentration of each chemical species over time.[107–109] These ODEs are simple to solve in R, Matlab, or Python; the difficulty is in ensuring the simplified system setup accurately reflects the biology. Despite the advantages of using pharmacokinetic models during drug development, anti-inflammatory ADC research does not make full use of these tools.[94,99,101,110–112]

1.5 Thesis Overview

In this thesis, I approach the problems of novel steroid synthesis and steroid targeting. In Chapter 2, I biosynthesized novel steroids by feeding non-native substrates into the enzyme CYP7B1. When CYP7B1 had entirely novel actions on these substrates, I modeled CYP7B1 and substrate binding, as discussed in Chapter 3. In Chapter 4, I determined the necessary pharmacokinetic properties for a steroid-carrying antibody fusion protein using a multiscale mechanistic ODE model. Together, this work provides proofs-of-concept for improving steroid drugs, which, if fully implemented, could revolutionize their clinical use.

2

Mammalian Cells Engineered to Produce Novel Steroids

Emma S. Spady, Thomas P. Wyche, Jon Clardy, Jeffrey C. Way, and Pamela A. Silver

2.1 Attributions

This chapter is adapted from a paper published in ChemBioChem in September 2018.[113] It covers the biochemical and analytical aspects of "Mammalian Cells Engineered To Produce New Steroids". The structure models originally published in this paper will be presented in Chapter 3.

The project concept was developed by Dr. Jeffrey Way and myself, and I performed all the biological and design work. Dr. Thomas Wyche instructed me in the use of the mass spectrometer and advised in method development for the isolation of pure steroids. I operated the liquid chromatographs and mass spectrometer, and also developed the liquid chromatograph methods. Dr. Wyche took the NMR spectra of the purified steroids and coached me through the structure assignment process. Finally, I wrote the manuscript.

2.2 Introduction

Steroids represent an important class of therapeutically useful molecules, but they have significant dose-limiting side effects. Steroid hormone analogs take advantage of the wide-ranging effects of their endogenous counterparts, such as immunosuppression and increased bone density, and are used to treat conditions ranging from arthritis to anemia. Unfortunately, steroid receptors alter global responses, so steroid drugs have severe side effects. These include diabetes and osteoporosis for glucocorticoids and masculinizing effects for androgens.[19,25,114] Steroids are active at bloodstream concentrations in the nanomolar range and readily cross cell membranes.[8,115] New combinations of modifications around the central steroid scaffold could improve binding affinity, pharmacokinetic properties, and receptor specificity.[57]

Steroids must be modified stereo- and regio-specifically to result in useful drugs. Steroid receptor binding pockets are sensitive to minor modifications; a single hydroxylation or methylation can cause a steroid to bind a different receptor, dramatically changing its biological effects.[57] Due to its lack of activated carbons, the steroid ring structure is difficult to modify while preserving its integrity.[68] Traditional chemical processes for steroid modifications are elaborate and frequently less selective than their biological counterparts, to the extent that wild-type fungi are used as industrial biocatalysts.[68,116,117] These biocatalysts are used exclusively in producing known steroids, leaving novel steroid production to organic chemistry methods.

Biosynthesis can provide an alternative source of novel steroid therapeutics. The enzymes involved in steroid hormone synthesis modify a large variety of substrates, but still perform regio- and stereoselective reactions.[118] In principle, these enzymes could accept a wide variety of synthetic steroid substrates and modify them to create novel products. A single enzymatic step could replace a long organic synthesis, and thus allow for extensive exploration of the steroid

drug biochemical space. Exogenous expression of steroid enzymes has already been used to produce complicated steroids intracellularly in yeast and mammalian cells, with glucocorticoids being of particular interest.[71,72] However, intentional biosynthesis of chemically novel steroid products using engineered mammalian cells has not been deeply explored.

Here we report the synthesis of novel steroids by expressing various steroid-modifying enzymes in mammalian cells and providing them with non-native substrates. We isolated three steroids by expressing the steroid 7α-hydroxylase CYP7B1, two of which arose from the enzyme's previously unknown ability to 11α- and 7β-hydroxylate substrates. This approach could be used to synthesize a wide variety of steroids for which traditional organic synthesis is not feasible.

2.3 Results

Our strategy for novel steroid synthesis is to express genes encoding steroid-modifying enzymes in mammalian cells, and then supply the cells with non-native substrates that will result in novel products (Fig. 2.1). To identify candidates, we expressed a panel of enzymes in mammalian cells and confirmed their function on native substrates. For enzymes catalyzing multiple native reactions, the substrate was chosen by the commercial availability of the expected product, enabling quantification. We then selected one of

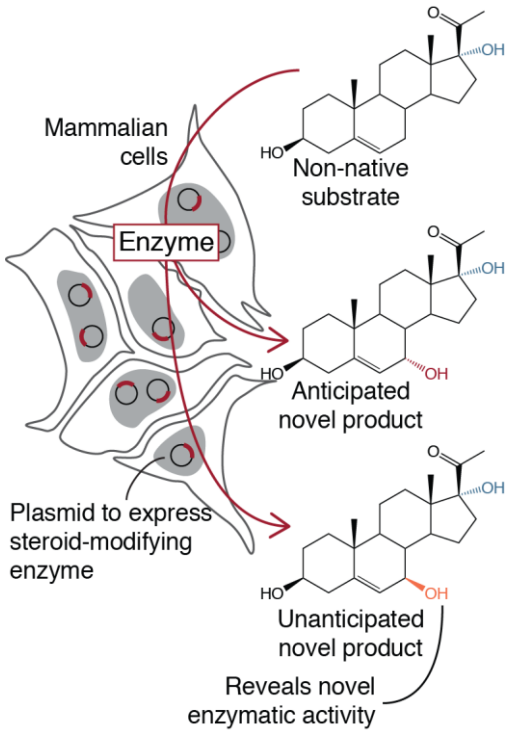


Figure 2.1. Overview of our method for novel steroid biosynthesis. Steroid-modifying enzymes are expressed heterologously in mammalian cells and exposed to a non-native substrate, resulting in novel steroid product(s).

the best-performing enzymes, CYP7B1, expressed it stably in cells, and supplied it with non-native substrates. These alternative substrates were commercially-available steroids that CYP7B1 could process into novel products, assuming that the enzyme modifies the substrates selectively. We isolated these products from the enzyme acting on the alternative substrates and confirmed the steroids' structures via NMR. Notably, we observed the 7α-hydroxylase CYP7B1 performing significant amounts of 11α- and 7β-hydroxylation on the alternative substrates.

Mammalian cells were used in this work because of several unique properties. Firstly, most of the enzymes tested were human, and cytochrome P450s can be difficult to express in heterologous hosts.[72,119] Mammalian cells also enable these enzymes to access cholesterol when required as a substrate. As there is no lipopolysaccharide in this system, any steroids in tissue culture supernatant could be tested rapidly on other mammalian cell lines without purification. While mammalian cells are unlikely to achieve the yields of yeast or bacterial cells, steroids are very potent drugs, and thus can be studied even in small quantities.[120,121]

Seventeen steroid-modifying enzymes were selected for expression in the human cell line HEK293. The enzymes were identified based on their gene sequence availability and the diverse reactions they catalyze (Table 3.1). Most of these enzymes are cytochrome P450s, and many are hydroxylases; the group also contains dehydrogenases, a methyltransferase, and a Baeyer-Villiger monooxygenase. Mammalian steroid-modifying enzymes are generally associated with the mitochondria or endoplasmic reticulum and may have signal sequences targeting them to these organelles; these sequences were retained. Bacterial enzymes were expressed without signal sequences and presumably localized to the cell cytoplasm. The HEK293 cell line was chosen because of its hardiness and ease of transfection.

Table 2.1. Steroid product yields from steroid-processing enzymes operating on native substrates in HEK293 cells. Substrates were chosen based on the literature and on the reaction resulting in a commercially available product.

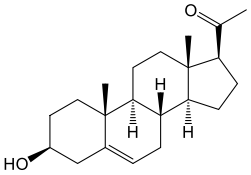
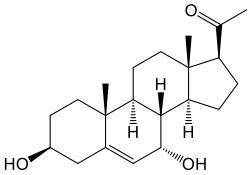
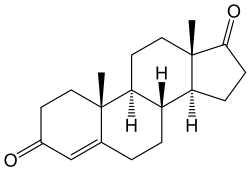
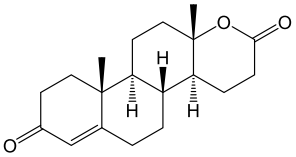
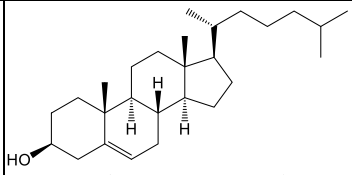
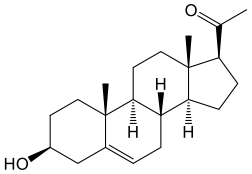
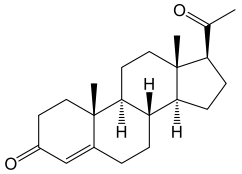
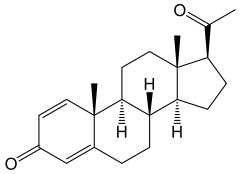
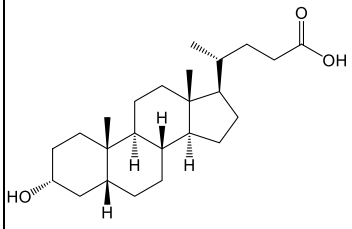
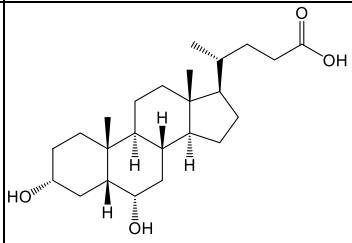
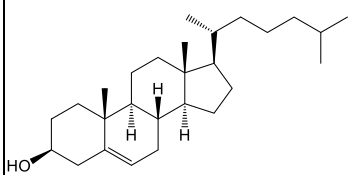
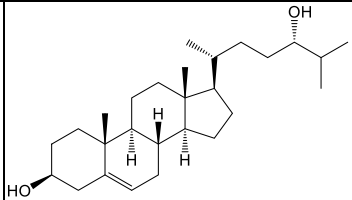
Enzyme Used	*Substrate Provided*	*Product Measured*	*40 μM Substr. Soln. in μg mL^{-1}*	*Amt. Product Detected (ng mL^{-1} medium)*	
				Enzyme ± SD	*Control ± SD*
CYP7B1[122]	 **1** (CAS 145-13-1)	 **2** (CAS 30626-96-1)	12.7	1400 ± 300	7.1 ± 0.9
CpdB[123,124]	 **9** (CAS 63-05-8)	 **10** (CAS 4416-57-3)	11.5	2600 ± 300	2 ± 2
CYP11A1[125]	 **11** (CAS 57-88-5)	 **1** (CAS 145-13-1)	15.5	150 ± 10 [a]	N.D.
Δ^1-KSTD2[126]	 **12** (CAS 57-83-0)	 **13** (CAS 1162-54-5)	12.6	54 ± 5	0.58 ± 0.05
CYP4A21[127]	 **14** (CAS 434-13-9)	 **15** (CAS 83-49-8)	15.1	30 ± 10 [b]	0.06 ± 0.09
CYP46A1[128]	 **11** (CAS 57-88-5)	 **16** (CAS 474-73-7)	15.5	8.0 ± 0.5 [a]	N.D.

Table 2.1 (Continued).

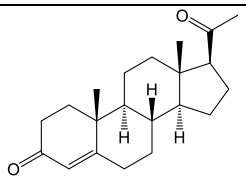
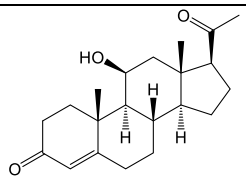
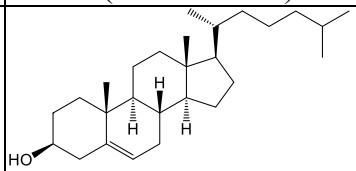
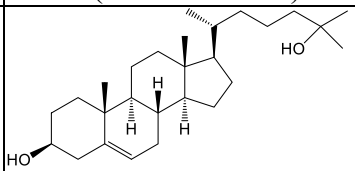
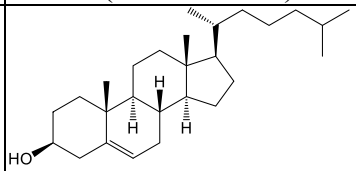
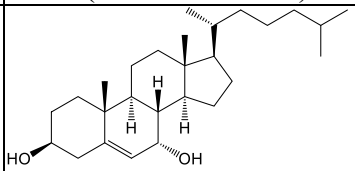
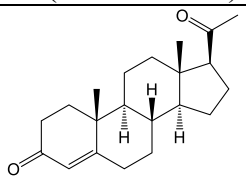
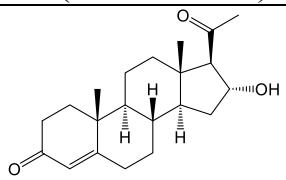
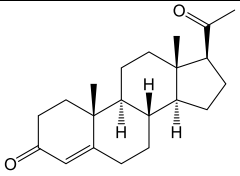
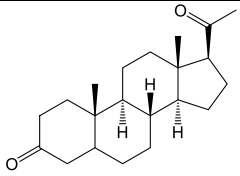
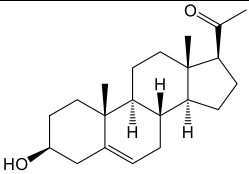
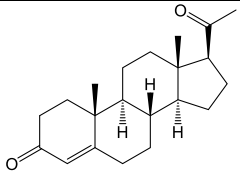
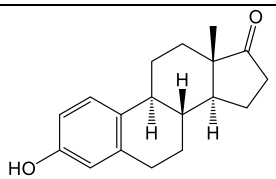
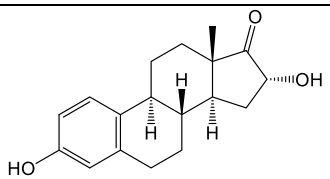
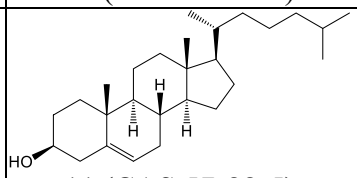
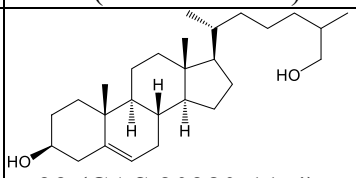
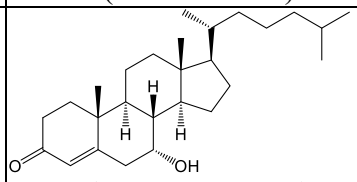
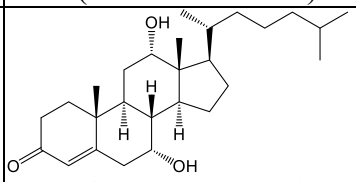
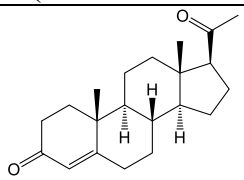
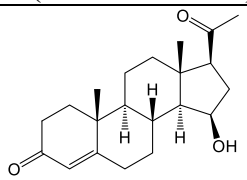
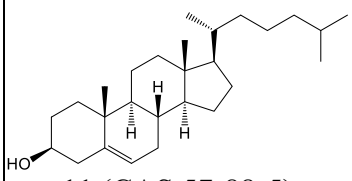
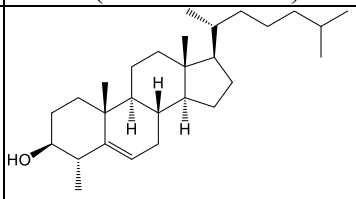
Enzyme Used	*Substrate Provided*	*Product Measured*	*40 μM Substr. Soln. in μg mL^{-1}*	*Amt. Product Detected (ng mL^{-1} medium)*	
				Enzyme ± SD	*Control ± SD*
CYP11B1 [76,129]	 **12** (CAS 57-83-0)	 **17** (CAS 600-57-7)	12.6	15 ± 2	12 ± 2
CH25H[128]	 **11** (CAS 57-88-5)	 **18** (CAS 2140-46-7)	15.5	4.6 ± 0.3 [a]	2 ± 1
CYP7A1[128]	 **11** (CAS 57-88-5)	 **19** (CAS 566-26-7)	15.5	4 ± 1 [a]	0.02 ± 0.07
CYP154C3 [130]	 **12** (CAS 57-83-0)	 **7** (CAS 438-07-3)	12.6	2.6 ± 0.3	0.42 ± 0.03
AKR1D1[131]	 **12** (CAS 57-83-0)	 **20** (CAS 566-65-4)	12.6	4 ± 1 [a]	3.9 ± 0.9
HSD3B7[131]	 **1** (CAS 145-13-1)	 **12** (CAS 57-83-0)	12.7	0.37 ± 0.07 [a]	0.5 ± 0.2

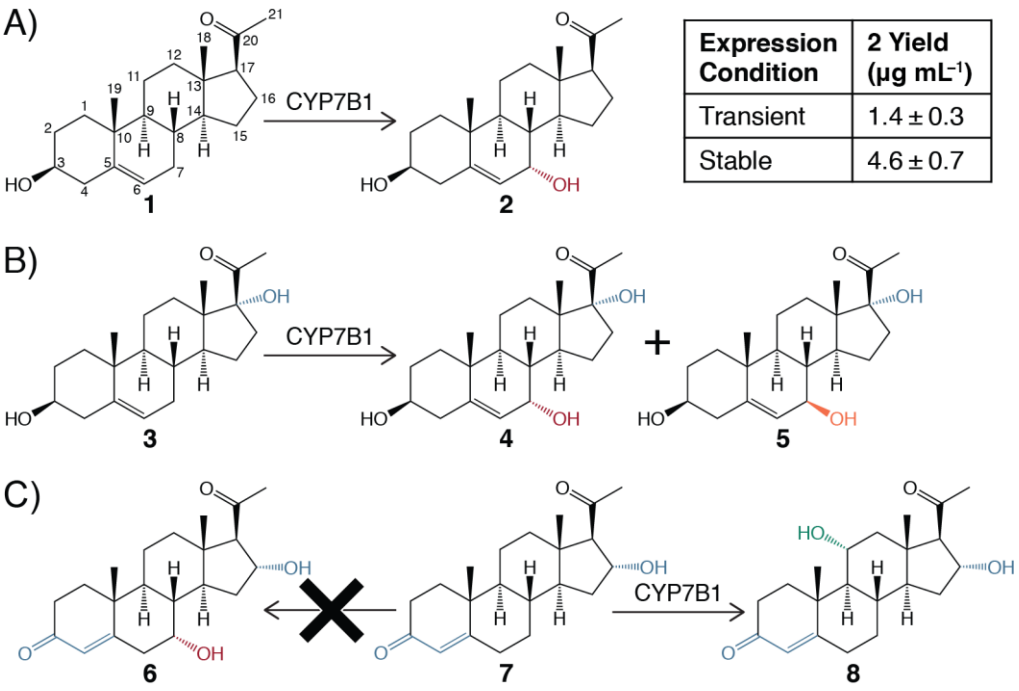
Table 2.1 (Continued).

Enzyme Used	*Substrate Provided*	*Product Measured*	*40 μM Substr. Soln. in μg mL^{-1}*	*Amt. Product Detected (ng mL^{-1} medium)*	
				Enzyme ± SD	*Control ± SD*
CYP3A4[132]	 **21** (CAS 53-16-7)	 **22** (CAS 566-76-7)	10.8	0.3 ± 0.3 [a]	N.D.
CYP27A1[128]	 **11** (CAS 57-88-5)	 **23** (CAS 20380-11-4)	15.5	N.D. [a]	N.D.
CYP8B1[128]	 **24** (CAS 3862-25-7)	 **25** (CAS 1254-03-1)	16.0	N.D. [a]	N.D.
CYP106A2 [133,134]	 **12** (CAS 57-83-0)	 **26** (CAS 600-72-6)	12.6	N.D. [a,c]	N.D.
STRM-1[135]	 **11** (CAS 57-88-5)	 **27** (CAS 481-25-4)	15.5	N.D. [a,c]	N.D.

[a] Product measured on an Agilent 6460 Triple Quad LC/MS with a Phenomenex Luna C18 column (5 μm, 250 x 10 mm) by the Harvard Small Molecule Mass Spectrometry Facility.
[b] Product measured on a Bruker Maxis Impact LC-q-TOF MS with a Phenomenex Synergi Polar column (50 x 2 mm) by the Harvard Small Molecule Mass Spectrometry Facility.
[c] Qualitative measurement due to lack of standard.

Expression constructs encoding each enzyme were transfected into HEK293 cells, and activity of each enzyme was tested on a native substrate and assayed via LC-MS. Negative enzyme activity controls consisted of transfecting the same plasmid backbone, but with a fluorescent reporter protein under the constitutive promoter. Cells were incubated for two days with a 40 μM solution of a native substrate for the relevant enzyme. The substrate concentration was chosen because of the low solubility of the less-hydroxylated steroids in media.[136] Lipids were extracted from the spent cells and medium and measured by LC-MS.[137,138] The yield per milliliter of spent medium was determined for each product when a standard was available.

Of the enzymes tested, the human 7α-hydroxylase CYP7B1 and the *Pseudomonas* cyclopentadecanone monooxygenase CpdB had the highest yields. CYP7B1 produced 1.4 μg of 7α-hydroxypregnenolone per mL (**2**) from a 12.7 μg mL^{-1} pregnenolone solution (**1**) (Fig. 2.2). CpdB made 2.6 μg of testololactone per mL (**10**) from an 11.5 μg mL^{-1} solution of androstenedione (**9**) (Fig. 2.3). Eight other steroid-modifying enzymes resulted in significantly more product than the fluorescent protein control when expressed (Table 2.1). These included two more bacterial enzymes, Δ^1-KSTD2 and CYP154C3, both of which used a 12.6 μg mL^{-1} progesterone (**12**) solution as substrate. From this, Δ^1-KSTD2 made 54 ng of 1-dehydroprogesterone (**13**) per mL, while CYP154C3 produced 2.6 ng of 16α-hydroxyprogesterone (**7**) per mL (Fig. 2.3).



Expression Condition	2 Yield (μg mL^{-1})
Transient	1.4 ± 0.3
Stable	4.6 ± 0.7

Figure 2.2. Steroid-modifying reactions observed from CYP7B1. A) CYP7B1 hydroxylates native substrate pregnenolone (**1**) to form 7α-hydroxypregnenolone (**2**). Yield is in μg of steroid per mL of medium extracted. B) CYP7B1 hydroxylates non-native substrate 17α-hydroxypregnenolone (**3**) to form novel product isomers 7α,17α-dihydroxypregnenolone **(4)** and 7β,17α-dihydroxypregnenolone (**5**). C) CYP7B1 hydroxylates non-native substrate 16α-hydroxyprogesterone (**7**) to form 11α,16α-dihydroxyprogesterone (**8**)**.** The anticipated product 7α,16α-dihydroxyprogesterone (**6**) was not observed.

Enzymes that failed entirely may have done so because of poor gene expression, but rather than troubleshooting, we focused on those enzymes that worked well immediately. CYP7B1, Δ^1-KSTD2 and CYP154C3 were pursued further, as they had potential to create novel products in a single step from commercially available substrates. While CpdB also had this potential, its high yield was found late in the project because its only commercially-available substrate-product pair are Schedule III controlled substances. The high 7α-pregnenolone (**2**) yield implied CYP7B1 had a good chance of processing alternative substrates into detectable quantities of novel steroids. In addition, CYP7B1 is a well-studied human enzyme.[139,140]

The CYP7B1-expressing cells' yield of **2** improved under different expression conditions. This ensured that the enzyme's action on alternative substrates would result in sufficient product for structure determination. The CYP7B1 electron transfer partner, cytochrome P450 oxidoreductase (POR), was expressed alongside CYP7B1 to prevent POR from limiting steroid hydroxylation. A modest 34% increase in **2** yield was achieved by transfecting CYP7B1 plasmid into stable POR-expressing HEK293 cells (Fig. 2.4A). A CYP7B1-expressing stable pool was constructed by puromycin selection to enable longer exposure of cells to substrate. These cells could be plated at low density and grown to confluency in the same medium, allowing for eight days of incubation with substrate instead of 48 hours. HEK293 cells were transfected with the CYP7B1 vector used previously and were selected on puromycin. This stable pool improved **2** yield by 229%, to 4.6 μg mL^{-1} (Fig. 2.4A). The CYP7B1-expressing stable pool was used in subsequent steroid syntheses, and the POR stable pool was not used.

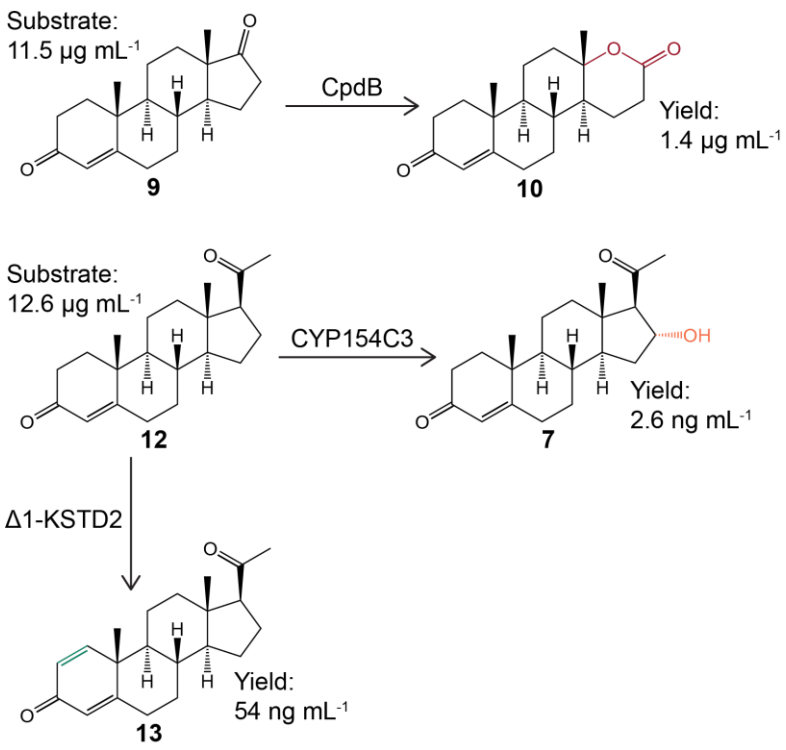


Figure 2.3. Steroids from bacterial enzymes successfully expressed in mammalian cells and provided with confirmed substrates. Yields are in mass of steroid measured per mL of spent cell medium extracted.

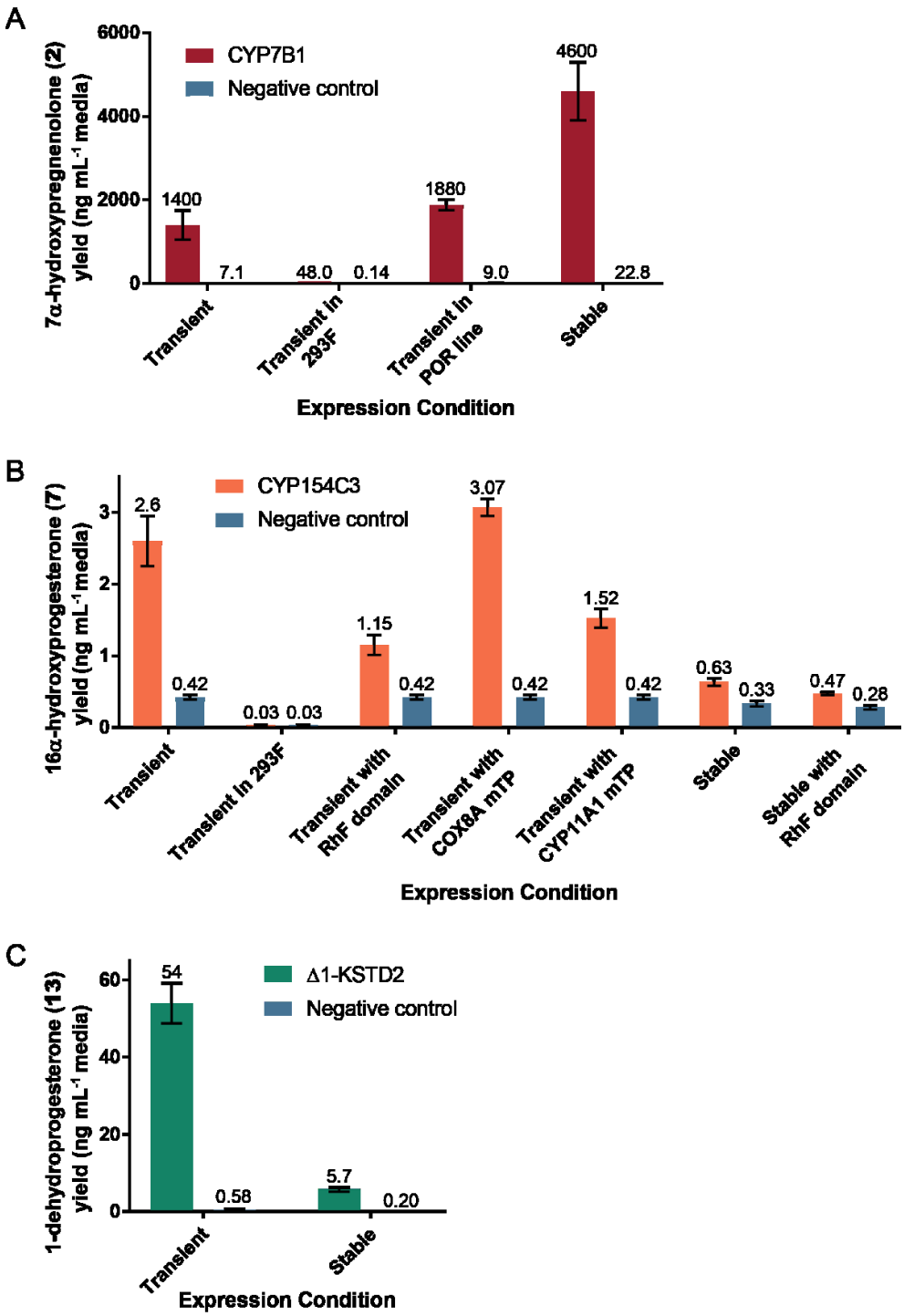


Figure 2.4. Expression methods tested to raise native product yield for CYP7B1, CYP154C3, and Δ1-KSTD2. A) Production of **2** from CYP7B1 expressed transiently, stably, or in the suspension cell line 293F. B) Production of **7** from CYP154C3 expressed transiently, stably, with the attached redox domain RhF, and with mitochondrial targeting peptides (mTPs). C) Production of **13** from Δ1-KSTD2 expressed transiently and stably.

The bacterial enzymes CYP154C3 and Δ^1-KSTD2 were also expressed in stable pools in an attempt to improve their yield (Fig. 2.4). Stable pool cell lines expressing these enzymes were made with the same method as the CYP7B1 line. However, the stable pools resulted in less product for both bacterial enzymes. CYP154C3's weeklong incubation with progesterone yielded only 0.63 ng mL^{-1} of **7**, compared to 2.6 ng mL^{-1} for transient transfectants (Fig 2.4B). The Δ^1-KSTD2 stable pool produced a mere 5.7 ng mL^{-1} of **13** when exposed to progesterone, compared to 54 ng mL^{-1} of product from transient transfectants (Fig 2.4C).

Alongside the aforementioned stable pool experiments, CYP154C3 and CYP7B1 were transiently transfected into suspension HEK293 Freestyle (293F) cells (Fig. 2.4). However, this dramatically decreased yield for both enzymes. CYP7B1 transfected into 293F cells only produced 48 ng mL^{-1} of **2**, a 97% decrease from its yield in adherent HEK293 cells. CYP154C3 transfected into 293F cells produced 0.03 ng mL^{-1} of **7**, a 99% decrease from the adherent cell yield. Fluorescent reporter plasmids transfected into 293F cells as negative enzyme controls indicated similar transfection efficiency to the adherent HEK293 cells. Thus the suspension cells were not pursued further as an enzyme expression system.

We considered other strategies to improve enzyme yield from CYP154C3 because the enzyme could, in principle, produce **7** from the inexpensive substrate **12**. CYP154C3 is a bacterial enzyme, and it typically requires two electron transfer partners to 16α-hydroxylate steroids. These partners can be replaced with a single redox domain (RhF) from the *Rhodococcus* sp. self-sufficient cytochrome P450, P450RhF, fused to CYP154C3's C-terminus[130]. This CYP154C3-RhF fusion protein performed poorly, with a **7** yield of 1.4 ng mL^{-1}, just over half of that of CYP153C3 alone (Fig. 2.4B). This was still significantly more than the control yield of 0.42 ng mL^{-1}, implying that CYP154C3 is compatible with the human homologues to its redox

partners, FDX1 and FDXR. FDX1 and FDXR are primarily located in the mitochondria, which suggests that CYP154C3 localization to the mitochondria could improve yield[129]. Adding the COX8A mitochondrial targeting peptide (mTP) to the N-terminus of CYP154C3 improved **7** yield to 3.07 ng mL^{-1}. Nevertheless, this yield was insufficient to consider CYP154C3 as a potential producer of novel steroids.

Stable CYP7B1-expressing cells were incubated with two alternative substrates that we predicted would result in novel products. We anticipated that CYP7B1 would convert 17α-hydroxypregnenolone (**3**) to 7α,17α-dihydroxypregnenolone (**4**) (Fig. 2.2B). Similarly, CYP7B1 was expected to hydroxylate 16α-hydroxyprogesterone (**7**) to produce 7α,16α-dihydroxyprogesterone (**6**) (Fig 2.2C). The CYP7B1 stable pool cell line was plated at low density and cultivated in the presence of a 40 μM alternative substrate solution for eight days. This corresponds to a 13.3 μg mL^{-1} solution of **3** and a 13.2 μg mL^{-1} solution of **7**. Lipids were extracted and detected by LC-MS as in the native substrate experiments. The lipids were compared to those from identically treated negative control stable pool cells, which expressed GFP instead of an enzyme. The desired hydroxylation product peaks were identified by having the correct mass and only being present with CYP7B1 expression. This separated the desired steroids from any hydroxylated isomers due to spontaneous degradation.

Peaks with masses corresponding to hydroxylation of the alternative substrates were detected exclusively in CYP7B1-expressing cells. CYP7B1 cells provided with alternative substrate **3** yielded two dihydroxypregnenolone peaks not present in the control (Fig. 2.5A). As dihydroxypregnenolones lose water during ESI-MS, the product peaks contained *m/z* = 331.2269, 313.2162, and 295.2057, corresponding to $[M+H-H_2O]^+$, $[M+H-2H_2O]^+$, and $[M+H-3H_2O]^+$ ions. The later, high signal molecule **4** eluted at 2.7 minutes, and the earlier, low signal

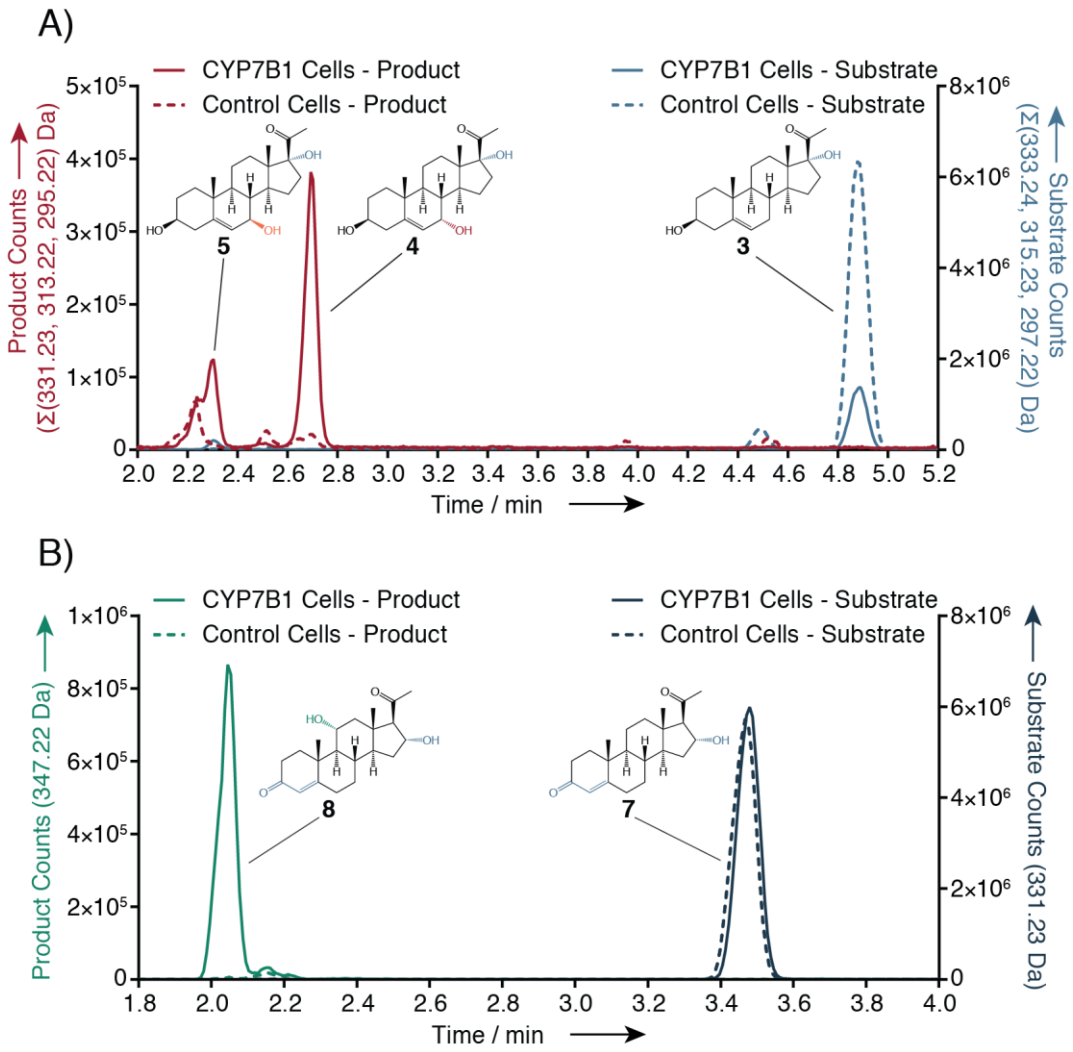


Figure 2.5. Extracted ion mass chromatograms comparing products of cells supplied with **3** and **7**. CYP7B1 cells stably expressed that enzyme, while the negative control cells stably expressed green fluorescent protein. CYP7B1 cell samples were diluted tenfold. A) Only CYP7B1 cells exposed to **3** produced peaks corresponding to the $[M+H-H_2O]^+$, $[M+H-2H_2O]^+$, and $[M+H-3H_2O]^+$ ions of **4** and **5**. B) Only CYP7B1 cells exposed to **7** produced a single peak corresponding to the $[M+H]^+$ ion of **8**.

molecule **5** eluted at 2.3 minutes. 0.7 mg of **4** and 0.4 mg of **5** were subsequently purified, implying a yield of 1.6 μg mL^{-1} and 0.91 μg mL^{-1}, respectively.

Steroid product formulae were initially determined via high-resolution ESI-MS. 7α,17α-dihydroxypregnenolone (**4**) and 7β,17α-dihydroxypregnenolone (**5**): *m/z* 331.2269 $[M+H-H_2O]^+$ (calculated for $C_{21}H_{31}O_3^+$, *m/z* 331.2268), *m/z* 313.2162 $[M+H-2H_2O]^+$ (calculated for

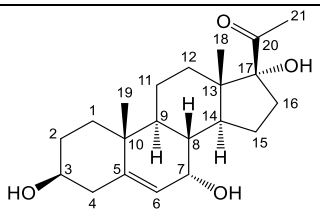
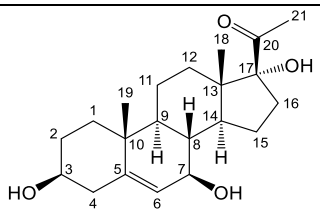
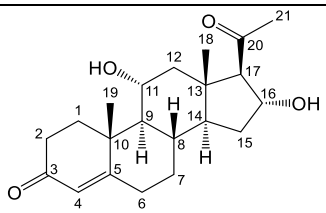
$C_{21}H_{29}O_2^+$, *m/z* 313.2162), and *m/z* 295.2057 $[M+H-2H_2O]^+$ (calculated for $C_{21}H_{27}O^+$, *m/z* 295.2056). These implied isomers with formula $C_{21}H_{32}O_4$. 11α,16α-dihydroxyprogesterone (**8**): *m/z* 347.2222 $[M+H]^+$ (calculated for $C_{21}H_{31}O_4^+$, *m/z* 347.2217). This suggests a molecule with formula $C_{21}H_{30}O_4$.

Analysis of 1D and 2D NMR data (gCOSY, gHSQC, gHMBC, ROESY, and TOCSY) allowed for structure determination of dihydroxypregnenolones **4** and **5**, corresponding to an expected 7α-hydroxylated product and an unexpected 7β-hydroxylated product (Table 2.2, Fig. 2.2B). COSY interactions between the hydrogen at C-6 (δ_H 5.57, 5.28) and a hydroxyl-adjacent hydrogen (δ_H 3.8, 3.79) in **4** and **5**, respectively, revealed that both products were 7-hydroxylated. The C-7 hydroxyl stereochemistry was determined by comparing the isomers to reference NMR for 7α-hydroxypregnenolone ($\delta_{C\text{-}7}$ 65.20) and 7β-hydroxypregnenolone ($\delta_{C\text{-}7}$ 73.14).[81,141] Compound **4** ($\delta_{C\text{-}7}$ 65.95) is therefore 7α,17α-dihydroxypregnenolone while **5** ($\delta_{C\text{-}7}$ 73.99) is 7β,17α-dihydroxypregnenolone.

The CYP7B1 stable pool cells provided with alternative substrate **7** exhibited a single dihydroxyprogesterone product peak (Fig. 2.5B). Dihydroxyprogesterones do not lose water as easily in ESI-MS, so the product peak was exclusively *m/z* = 347.2222, corresponding to the $[M+H]^+$ ion. A single isomer **8** with this *m/z* eluted at 2.05 minutes and was absent in the negative control lipids. 2.2 mg of **8** were eventually purified from tissue culture, with a yield of 6.9 μg mL^{-1}.

Compound **8**, the single dihydroxyprogesterone product, was determined to be 11α,16α-dihydroxyprogesterone after analysis of 1D and 2D NMR data (Table 2.2, Fig. 2.2C). No other masses that could correspond to the expected product **6** were detected. The newly hydroxylated carbon (δ_C 68.88, δ_H 3.95) was identified by locating its methylene neighbor (δ_C 50.89, δ_H 1.63,

Table 2.2. ^{1}H and ^{13}C NMR data (500 MHz for ^{1}H, 125 MHz for ^{13}C, in CD_3OD) for steroid products **4**, **5**, and **8**.

	 7α,17α-dihydroxypregnenolone (**4**)		 7β,17α-dihydroxypregnenolone (**5**)		 11α,16α-dihydroxyprogesterone (**8**)	
Position	**δ_C, type**	**δ_H, *J* in Hz**	**δ_C, type**	**δ_H, *J* in Hz**	**δ_C, type**	**δ_H, *J* in Hz**
1	38.1, CH_2	1.89. m; 1.20, td (13.6, 3.5)	38.2, CH_2	1.90, m; 1.09, m	38.6, CH_2	2.69, d (13.8); 2.06, t (13.2)
2	32.1, CH_2	1.82, m; 1.52, m	32.3, CH_2	1.83, d (12.4); 1.52, m	34.9, CH_2	2.48, m; 2.30, m
3	72.0, CH	3.50, m	72.1, CH	3.44, m	203.0, C	
4	42.9. CH_2	2.30, m	42.5, CH_2	2.26, m	124.7, CH	5.73, s
5	146.6, C		144.0, C		175.1, C	
6	124.9, CH	5.57, dd (5.3, 1.4)	127.4, CH	5.28, s	34.7, CH_2	2.49, m; 2.30, m
7	66.0, CH	3.80, t (4.0)	74.0, CH	3.79, d (7.7)	32.9, CH_2	1.86, d (12.7); 1.12, q (12.5)
8	39.3, CH	1.55, m	41.4, CH	1.51, m	36.3, CH	1.63, m
9	43.2, CH	1.37, td (12.1, 5.2)	49.7, CH	1.09, m	60.1, CH	1.22, t (10.1)
10	38.5, C		37.6, C		41.40, C	
11	21.6, CH_2	1.68, m; 1.53, m	21.9, CH_2	1.69, m; 1.51, m	68.9, CH	3.95, m
12	31.7, CH_2	1.87. t (3.4); 1.52, m	31.8, CH_2	1.90, m; 1.52, m	50.9, CH_2	2.24, m; 1.63, m
13	47.9, C		48.7, C		46.1, C	
14	45.4, CH	2.29, m	51.9, CH	1.91, m	54.4, CH	1.67, m
15	24.6, CH_2	1.93, m; 1.27, m	26.7, CH_2	1.99, m; 1.57, m	35.9, CH_2	1.78 (m)
16	34.1, CH_2	2.71, ddd (11.6, 3.2, 3.1); 1.49, m	34.4, CH_2	2.69, m; 1.45, m	74.2, CH	4.72, t (7.3)
17	91.5, C		91.0, C		72.8, CH	2.58, d (6.6)
18	15.1, CH_3	0.63, s	15.3, CH_3	0.63, s	15.9, CH_3	0.70, s
19	18.7, CH_3	1.02, s	19.5, CH_3	1.09, m	18.6, CH_3	1.35, s
20	213.6, C		213.6, C		210.2, C	
21	27.4. CH_3	2.20, s	27.5, CH_3	2.20, s	31.9, CH_3	2.19, s

2.24), with which it has a strong COSY correlation. This neighboring carbon has an HMBC with hydrogens at C-17 (δ_H 2.58) and C-18 (δ_H 0.70). This placed the neighboring carbon at C-12, as the D-ring hydroxyl had already been assigned and other locations were too far from C-17 and C-18. Therefore the CYP7B1 hydroxylation was at C-11. The C-11 hydroxyl stereochemistry was found through ROESY interactions. The hydrogen at C-11 (δ_H 3.95) interacts with hydrogens at C-18 (δ_H 0.70) and C-19 (δ_H 1.35), indicating that it is above the steroid ring plane. The hydroxyl at C-11 must face the opposite direction, and thus is an 11α-hydroxyl.

2.4 Discussion

We made the three steroids **4**, **5**, and **8** by expressing CYP7B1 in mammalian cells and providing it with non-native substrates **3** and **7**. **5** and **8** were the result of 7β- and 11α-hydroxylations, respectively, though CYP7B1 was thought to primarily perform 7α-hydroxylation. This work represents a proof of concept in using genetically engineered mammalian cells to produce novel steroids, and more broadly in using mammalian cells to produce novel small molecules. Our overall method can take advantage of inexpensive gene synthesis, diverse steroid-modifying enzymes, and mammalian cells' unique properties to quickly create steroids that are compatible with direct biological testing.

Steroids **4** and **5** are novel molecules, but **8** has a straightforward organic synthesis. Nevertheless, **8**'s chemical properties are not published; no NMR structural data were released when it was initially isolated.[142] **4** and **5** would have been difficult to produce via established methods. Both have been tentatively identified in mixtures of microbial steroid hydroxylation products, but no NMR data is available.[82,143] Selective 7β-hydroxylation of pregnenolones is possible with organic methods, but syntheses that directly favor 7α-hydroxylation do not appear

possible.[144] Nevertheless, C-7 is a relatively easy carbon to chemically oxidize, as it is adjacent to a double bond.[145] While microbial pregnenolone 7α-hydroxylation has been described, the organisms make a mixture of products, including non-7-hydroxylated steroids, and the enzymes have not been identified.[81,146,147] Hence our method for 7-hydroxylation fills a gap in steroid biosynthetic methods.

We observed unprecedented amounts of 11α-hydroxylated and 7β-hydroxylated products from CYP7B1. Prior to this work, CYP7B1 only performed 7α-hydroxylation, 6α-hydroxylation, or 7β-hydroxylation, depending on the substrate.[148] The lattermost product is minor, when observed at all; it comprised just 1% of the hydroxysteroid product from CYP7B1 action on DHEA.[122] In comparison, **5** was 36% of the total product isolated from CYP7B1 acting on **3**. 11α-hydroxylation by CYP7B1 had never been observed until we isolated **8**. These reactions demonstrate that even well-studied cytochrome P450s may process new substrates in unexpected ways; one cannot assume which isomers these enzymes produce based only on the available literature.

Our approach reveals that novel steroids can be generated by combining elements from the human metabolic repertoire. CYP7B1 is a human enzyme; its murine homolog is expressed primarily in the brain and at low levels in the liver.[149] Its 7α-hydroxylated steroid products are poorly understood, but improve memory in aged, memory-impaired mice.[148,150] While **7** is not present in the human body, **3** is an intermediate formed by CYP17A1 in the adrenal cortex and gonads.[151] However, human CYP17A1 preferentially processes **3** further into dehydroepiandrosterone.[152] Though the enzyme dissociates from the steroid after hydroxylation, relatively little **3** can escape into other tissues before CYP17A1 acts again.[153] While CYP7B1

likely does not encounter **3** in nature, in our system the enzyme is capable of processing this substrate into novel products **4** and **5**.

Here, we engineered mammalian cells to produce previously uncharacterized drug-like small molecules. Numerous protein drugs are already made in mammalian cells, typically due to species-specific post-translational modifications such as N-linked glycosylation. Mammalian cells and mammal-based cell-free systems have been engineered to make steroids, but these were only used to study the biosynthetic pathways of known molecules.[71,154] Engineered steroid biosynthesis, even in microbes, has been used to make known steroids; novel steroids are typically detected spontaneously from unmodified cells.[72,73,117,146] This work therefore opens the door both towards using mammalian cells for other small molecule products, and also towards directed biosynthesis of novel steroids.

Many novel molecules could be made with engineered cells expressing steroid-modifying enzymes, making more steroids accessible for biological testing. The decreasing cost of gene synthesis, the high frequency of mammalian cell co-transfection when plasmids are simply mixed together, and the high activity of steroids in biological assays could enable a means of identifying useful compounds. Our *in vivo* system could be modified to access all parts of the steroid ring system, unlike current organic synthesis methods. Specifically, we envision that libraries of steroids could be made by combinatorial transfection of mammalian cells, followed for example by direct biological testing in high-content cell-based assays to find new activities.[155]

2.5 Methods

2.5.1 Plasmids. The CYP7B1-containing plasmid had a pcDNA3.1 backbone, a CMV promoter driving intron-free CYP7B1 and puromyin resistance. This pcDNA3.1 backbone from Invitrogen includes a pUC origin and ampicillin resistance to enable maintenance in *Escherichia coli*. The human CYP7B1 isoform 1 cDNA sequence was from the NCBI CCDS database number 6180.1. The negative control plasmid was identical to the CYP7B1 plasmid but contained GFP in place of CYP7B1. This codon-optimized GFP sequence was originally from the pSF-CMV-Ub-daGFP plasmid (Oxford Genetics). A list of all steroid-modifying enzymes used in similar plasmids can be found in Table 2.1. Plasmids were constructed using Gibson assembly to insert IDT gBlocks containing the gene of interest into PCR-amplified plasmid backbones, which were then transformed into chemically-competent *E. coli* K12. Plasmids were prepared using the Qiagen PlasmidPlus MidiPrep or MaxiPrep kit.

2.5.2 Protein Expression in Mammalian Cells. HEK293 cells (ATCC) were grown in DMEM (Invitrogen) with 10% FBS (Gibco) and penicillin-streptomycin (Gibco). Cells were transfected with 31.5 μg of plasmid DNA per well in 6-well plates using the Lipofectamine 3000 kit (Thermo Fisher). Medium was replaced after six hours with fresh medium containing 40 μM steroid substrate. For concentrations of each substrate in ng mL^{-1}, refer to Table 2.1. Cells were incubated for 48 hours, after which the negative controls were checked for fluorescence, indicating acceptable transfection efficiency. Spent medium and cells were mixed by scraping, and the liquid was frozen for storage. Stable pools of HEK293 cells expressing steroid-modifying genes began with transfection of ~800,000 cells using the lipofectamine 3000 kit, as above. 48 hours after transfection, cells were selected via 1.5 μg mL^{-1} puromycin (Sigma). After eighteen days, antibiotic was decreased to 0.5 μg mL^{-1} puromycin. To modify steroids, stable

pool cells were plated at 7,000 cells per cm^2 with 40 μM of the desired substrate. The cells were incubated for eight days, after which adhered cells were scraped to mix with the medium and frozen for storage.

2.5.3 Steroid Product Isolation. Liquid-liquid separation with a 3:10 methanol : methyl-tert-butyl ether v/v organic phase removed lipids from the media[137]. The medium (2 mL) was extracted twice with solvent (6.5 mL and 3 mL), vortexing for three minutes each time. The upper layers were removed, combined, dried down, and resuspended in methanol (100 μL).

Lipids were analyzed with an Agilent 1200 series HPLC system and 6530 qTOF mass spectrometer. Steroids were separated with a Thermo Scientific Hypersil GOLD C18 column (1.9 μm, 50 x 2.1 mm) in an acetonitrile gradient in water with formic acid (0.1% v/v)[138]. 5 uL of the resuspended steroid was injected into the instrument. Concentrated steroid products, such as the CYP7B1-expressing cell extract in Figure 2.5, were diluted tenfold to maintain narrow peaks. The linear gradients of acetonitrile in water with 0.1% formic acid v/v are as follows: 0-1 min: 10-70% CH_3CN, 1-8 min: 30-50% CH_3CN, 8-9 min: 50-100% CH_3CN, 9-13 min: 100% CH_3CN, 13-13.5 min: 100-10% CH_3CN, 13.5-17 min: 10% CH_3CN. All percentages listed are v/v. A flow rate of 0.3 mL per min was used, and the column was maintained at 50°C. Known products were compared to standards for quantification. Please refer to Figure S1 for chromatograms of steroid products with standards and Table S1 for standard curve calculations used to quantify these steroids.

2.5.4 Steroid chemical sources. Substrates were purchased from the following vendors: Pregnenolone (**1**, CAS 145-31-1) from Sigma-Aldrich, 17α-hydroxypregnenolone (**3**, CAS 387-79-1) from Pfaltz and Bauer, 16α-hydroxyprogesterone (**7**, CAS 438-07-3) from Toronto Research Chemicals, androstenedione (**9**, CAS 63-05-8) from Crescent Chemical, and

progesterone (**12**, CAS 57-83-0) from Sigma-Aldrich. The standard 7α-hydroxypregnenolone (**2**, CAS 30626-96-1) was purchased from Toronto Research Chemicals, testololactone (**10**, CAS 4416-57-3) from Sigma-Aldrich CPR, 1-dehydroprogesterone (**13**, CAS 1162-54-5) from Sigma-Aldrich, and 16α-hydroxyprogesterone (**7**, CAS 438-07-3) from Toronto Research Chemicals.

2.5.5 Unknown Steroid Purification and NMR. Twenty-two T150 flasks of the CYP7B1-expressing cells were used to process substrate **3**, yielding 440 mL of medium. Sixteen similar flasks were used to process substrate **7**, resulting in 320 mL of medium. Liquid-liquid extraction was performed with the same proportions as above, with the addition of a sodium sulfate drying step. The product mixtures from **3** and **7** were resuspended in 3 and 5 mL of methanol, respectively.

An Agilent 1200 semi-preparative HPLC system with a Phenomenex Luna C18 column (5 μm, 250 x 10 mm) was used to purify the steroid products **4**, **5**, and **8**. 100 μL of solution was injected at a time, with a flow rate of 4 mL min^{-1}. The linear gradients of acetonitrile in water with 0.1% v/v formic acid are as follows: 0-24 min: 30-65% CH_3CN, 24-26 min: 65-100% CH_3CN, 26-29 min: 100% CH_3CN, 29-30 min: 100-10% CH_3CN, 30-35 min: 10% CH_3CN. All percentages listed are v/v. Under these conditions, **4** had a retention time of 8.7 min and **5** had a retention time of 10.3 min. Compound **8**, the product of CYP7B1 action on **7**, required two rounds of reverse-phase HPLC, using the same column, flow rate, and system as for **4** and **5**. The first HPLC used an acetonitrile gradient in water with 0.1% formic acid v/v with the following linear gradients: 0-24 min: 10-60% CH_3CN, 24-26 min: 60-100% CH_3CN, 26-29 min: 100% CH_3CN, 29-30 min: 100-10% CH_3CN, 30-35 min: 10% CH_3CN, yielding **8** in a mixture of products at 15.3 min. The second step of HPLC purification used a methanol gradient in water with 0.1% formic acid with the following linear gradients: 0-5 min: 50% CH_3OH, 5-20 min: 50-

55% CH_3OH, 20-22 min: 55-50% CH_3OH, 22-30 min: 50% CH_3OH. All percentages listed are v/v. This method produced pure **8** at 15.4 min. This resulted in 0.7 mg of **4**, 0.4 mg of **5**, and 2.2 mg of **8**.

NMR spectra were obtained in CD_3OD with a Bruker AVANCE 500 MHz spectrometer equipped with a $^{1}H\{^{13}C/^{15}N\}$ cryoprobe and a Bruker AVANCE 500 MHz spectrometer equipped with a $^{13}C/^{15}N\{^{1}H\}$ cryoprobe. ^{13}C and ^{1}H shifts and assignments are in Table 2.2. ^{13}C, ^{1}H, HSQC, HMBC, COSY, ROESY, and TOCSY spectra, can be found in the appendix figures S2-S24.

2.6 Acknowledgements

This work was supported by the Defense Advanced Research Projects Agency HR0011-12-C-0061 and the National Institute of Health P50 GM107618 (Laboratory of Systems Pharmacology) and R01 GM086258 (JC). We are grateful to the ICCB-Longwood Analytical Chemistry Core facility at Harvard Medical School for use of their mass spectrometer. We would also like to thank S. Trauger, K. Chatman, and G. Byrd at the Harvard Small Molecule Mass Spectrometry Core.

3

Structural Models of CYP7B1 to Explain 7β- and 11α-Hydroxylations

Emma S. Spady, Nathanael J. Rollins,
Jeffrey C. Way, and Pamela A. Silver

3.1 Attributions

This chapter is adapted from the paper "Mammalian Cells Engineered to Produce New Steroids", published in ChemBioChem in September 2018. Nathan Rollins was instrumental in teaching me how to use Rosetta. He developed the method for incorporating the QM-modeled heme into the protein structure and ran the commands to generate the protein structure predictions discussed. I studied the properties of cytochrome P450s to determine which energetically indistinguishable binding modes were most likely to have occurred. In addition, I made the figures and wrote the manuscript.

3.2 Introduction

The cytochrome P450 monooxygenase enzyme superfamily performs a wide variety of important oxidations. In humans, cytochrome P450s process about 75% of all small-molecule drugs, playing a critical role in their pharmacokinetics.[156] These enzymes are also interesting from a metabolic engineering standpoint due to their ability to oxidize un-activated carbons.[68,157]

Cytochrome P450s are essential for steroid hormone biosynthesis and metabolism. The steroid-modifying cytochrome P450s are less famously promiscuous than their xenobiotic-processing cousins, but still have the flexibility to accept several steroid isomers in their active sites.[158] Better understanding of cytochrome P450 substrate and product specificity is thus important in pharmacology, endocrinology, and metabolic engineering.

Steroid-processing cytochrome P450s may perform different reactions on relatively similar substrates. New functionalities are regularly discovered even for well-studied human enzymes. Recently, CYP17A1, typically known as a steroid 17α-hydroxylase and 17α-20-lyase, was found to catalyze 16α-hydroxylation and 6β-hydroxylation.[159] In the previous chapter, I demonstrated that CYP7B1, thought to exclusively 7α-hydroxylate steroids, is also capable of 11α- and 7β-hydroxylation on certain substrates.[113] Understanding how these new reactions arise is interesting from a biochemical standpoint, but there are also health implications for human enzymes. As even miniscule amounts of unusual steroid hormone isomers can have biological effects, these enzyme side products are important to understand.

The cytochrome P450 reaction mechanism enables oxidation of diverse substrates and positions within those substrates. Essentially, the enzyme's heme coordinates molecular oxygen and activates it by reduction (Fig. 3.1).[160] The necessary electrons pass from NADPH through a redox partner protein to reach the cytochrome P450 iron. The reduced oxygen then binds two protons, enabling one oxygen atom to depart as water. The remaining oxygen atom and the porphyrin form a π-radical ferryl intermediate.[161] The fate of this highly reactive species, known as compound I, depends on the enzyme under consideration. In the case of monooxygenases, such as CYP7B1 and CYP17A1, the ferryl oxygen will abstract a hydrogen from any nearby

carbon in the substrate (Fig. 3.2).[162] The ferryl hydroxyl quickly binds to the substrate's resulting carbanion, hydroxylating that position. This releases the iron in the heme, which returns to its initial Fe^{III} state.

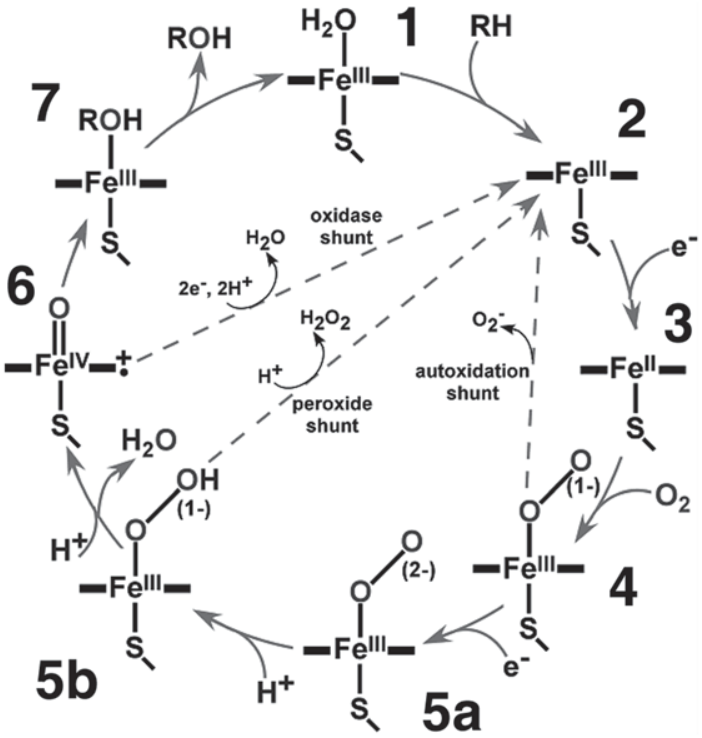


Figure 3.1. The cytochrome 450 catalytic cycle, from Denisov *et al.* 2005.[160]

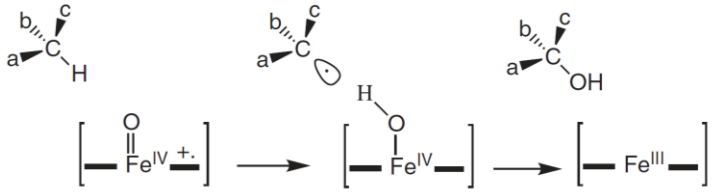


Figure 3.2. The carbon hydroxylation mechanism of cytochrome P450 compound I, from Ortiz de Montellano 2015. The configuration of substituents a, b, and c on the carbon is generally retained through the hydrogen abstraction and oxygen rebound.[162]

The oxidative power of the unstable ferryl oxygen enables cytochrome P450s to process many substrates. There are few chemical requirements of the substrate, save that it has an oxidizable carbon near the oxygen in the π-radical ferryl. The substrate and regio-specificity of a cytochrome P450's reaction depends solely on the flexibility of the hydrophobic active site. Cytochrome P450s that process diverse substrates, such as the xenobiotic oxidizers CYP3A4/5, CYP2C9, and CYP1A2, have active sites with few specific interactions (described amusingly as "hydrophobic custard" by Korzekwa).[163] The steroid-processing cytochromes tend to be more specific, accepting only steroid substrates and holding them at more precise angles for regio- and stereospecific modification.[158] Nevertheless, these

active sites are also hydrophobic slots, making it possible for different steroids to be modified differently if they sit off-kilter in the site. Hence if a steroid binds in a particular mode, it is appropriate to assume that the carbon nearest the ferryl oxygen will be hydroxylated.

I used a protein structure model to understand how the cytochrome P450 CYP7B1 performs different reactions on different substrates. This human enzyme was previously thought to be a steroid 7α-hydroxylase with very minor 7β-hydroxy side products.[140] Yet in the previous chapter, CYP7B1 7β- and 11α-hydroxylated the alternative substrates 17α-hydroxypregnenolone (**3**) and 16α-hydroxyprogesterone (**7**), respectively (Fig. 3.3). CYP7B1 7α-hydroxylated its native substrate pregnenolone (**1**) without detectable side products, and 7α-hydroxylated approximately twice as much **3** as it 7β-hydroxylated.[113] The cytochrome P450 mechanism dictates that the orientation of the substrate in the active site determines the regioselectivity of the reaction. Hence a docking model can show which residues may be relevant in driving these D-ring hydroxylated steroids to behave unusually. I used the macromolecular modeling software Rosetta to model native substrate **1** and alternative substrates **3** and **7** binding to CYP7B1, revealing how the expected 7α-hydroxylated and unexpected 7β- and 11α-hydroxylated products could arise.

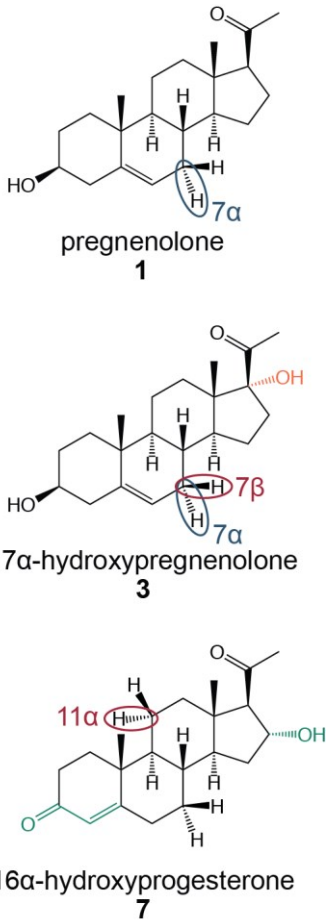


Figure 3.3. Substrates of CYP7B1 with anticipated hydroxylation sites circled in blue and unanticipated hydroxylation sites circled in red. Differences between alternative substrates 3 and 7 and the native substrate are in orange and green, respectively.

3.3 Results

The modeling software Rosetta can predict protein structures and dock small molecules into these predictions. It bases its structures on solved crystal structures of homologous proteins.[164] Rather than generating one computationally-intensive optimized model, Rosetta generates thousands of preliminary predictions by randomizing residue and ligand positions. Next, it adjusts the atoms slightly so no direct clashes form. Then it assigns each preliminary prediction a Rosetta score in Rosetta Energy Units (REUs). This score is intended to correlate with, but not directly represent, the total energy of protein folding and ligand binding. A lower score indicates the structure has optimal physical interactions, solvent accessible surface area, and bond angles.[165] In order to test enough protein conformations to be accurate, Rosetta often samples similar conformations more than once. Hence the preliminary structures are typically clustered by root-mean-square deviation of atomic positions (RMSD), and each cluster is represented by its best-scoring member. These representatives must meet a Rosetta score cutoff to be thought of as 'energetically realistic'.

As there are no CYP7B1 crystal structures, we used Rosetta comparative modeling to create a model crystal structure into which we could dock steroid substrates.[165] I searched for protein structures within 30% sequence identity and 10^{-4} BLAST confidence of CYP7B1. The structures were required to be bound to substrate analogs, in order to account for induced fit in the binding pocket. The two structures that fit these criteria were of CYP7A1 bound to cholest-4-en-3-one and 7-ketocholesterol. (PDB IDs 3sn5 and 3v8d).[166] CYP7A1, a human cholesterol 7α-hydroxylase, is 41% sequence identical to CYP7B1 with 95% coverage. The CYP7A1 structures were used to make rough models of the CYP7B1 apo-enzyme, into which a heme could be inserted. After compound I heme addition, the structures were optimized and scored. Structures

within fifty REU of the lowest-scoring structure were selected and clustered. One best-scoring representative protein from each of the twenty largest clusters was used in further analyses.

RosettaLigand docked each steroid substrate into this ensemble of CYP7B1 structures and identified energetically feasible binding modes.[164] Binding modes with acceptable overall protein fold were identified by their total Rosetta score falling within 65 REU of the lowest-scoring mode. Modes with reasonable steroid position were found by their ΔG-binding-analog Rosetta score. The ΔG-analog score is the difference in overall scores when the steroid is present versus absent in a given protein conformation. Structure models with ΔG-analog scores within 4 REU of the best mode were considered. This threshold is analogous to the score contribution of one hydrogen bond; differences lower than this do not correlate reliably with actual energy.[164] These structures were clustered at 3 Å RMSD between substrate atoms, and the cluster representative was chosen by the lowest ΔG-analog score. This process resulted in 89 binding modes for **1**, 71 binding modes for **3**, and 92 binding modes for **7** with indistinguishable binding energy (Fig. 3.4).

Final candidate modes were selected based on information on the cytochrome P450 mechanism. Because only specific product isomers were detected, the only reactive modes must be those that place the hydroxylation site close to the ferryl oxygen. In general, CYP450 crystal structures have substrates positioned with the hydroxylated carbon within 4-5 Å of the heme iron.[167] The CYP7A1 structures have distances between C-7 and the heme iron that are consistent with this assertion: 5.0 Å for cholest-4-en-3-one and 4.9 Å for 7-ketocholesterol.[168] The Rosetta model used a compound I heme and represents hydrogens, so the distance between the abstracted hydrogen and the ferryl oxygen could be measured for all binding modes. Binding modes with the relevant hydrogen less than 3.5 Å from the ferryl oxygen were selected for

further consideration. This narrowed the candidate positions to 17 for 7α-hydroxylation of **1**, 16 for 7α-hydroxylation of **3**, 23 for 7β-hydroxylation of **3**, and 22 for 11α-hydroxylation of **7**.

The substrate position in the CYP7A1 homologs determined which binding modes were most likely to occur. First, modes were eliminated if they placed non-reacting hydrogens dramatically closer to the ferryl oxygen than the observed hydroxylation sites. Correct steroid positions were assumed to occupy the same area over the heme as in the CYP7A1 crystal structures, in the direction of the propionate groups (Fig. 3.4A).[168] Given these constraints,

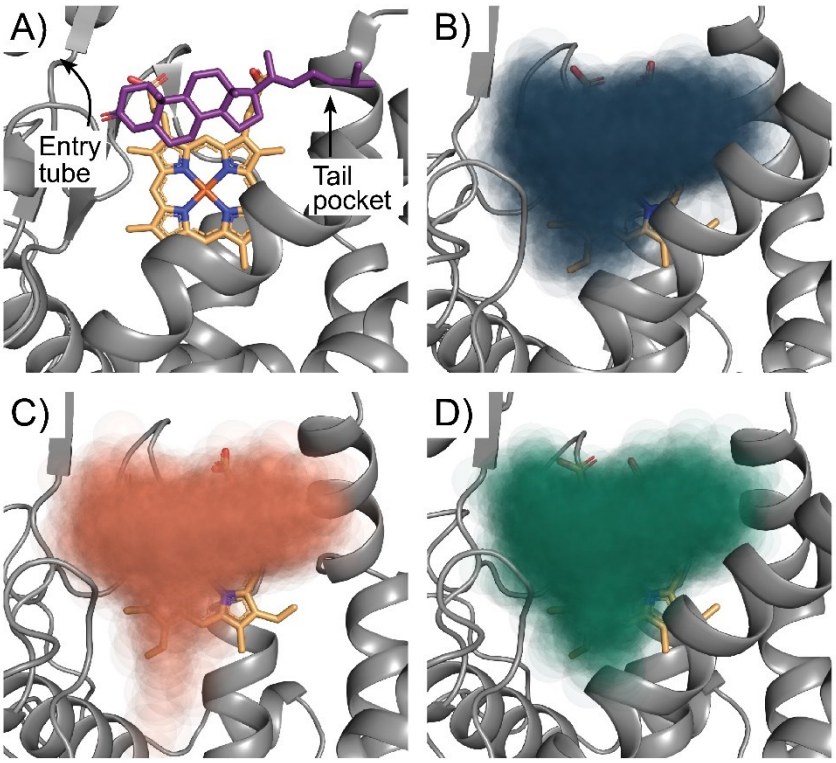


Figure 3.4. Binding modes generated from Rosetta for steroid substrates in CYP7B1. Substrate carbons and oxygens are represented as translucent spheres; color density approximates the number of modes occupying the space. Hemes are aligned for easy comparison of mode distributions. A) CYP7A1 bound to cholest-4-en-3-one, indicating direction of entry tube and tail pocket relative to the heme.[166] B) Native substrate **1** binding to CYP7B1. C) Alternative substrate **3** binding to CYP7B1. D) Alternative substrate **7** binding to CYP7B1.

unique structures for the 7α-hydroxylation reactions were found. However, two conformations for the 7β-hydroxylation and the 11α-hydroxylation were deemed possible.

The 'tilted' modes show the steroid rotating lengthwise to bring the relevant hydrogen close to the oxygen (Fig. 3.5AB). The 'flipped' modes show the steroid inverted in the binding site, switching the usual locations of the A and D rings (Fig. 3.5CD). Given the narrow entry tube to the enzyme homolog CYP7A1, the steroids cannot turn end-over-end after entering the protein. The flipped orientations would therefore require the steroids entering the tube A-ring-first, instead of the tail-first entry proposed for CYP7A1 substrates.[168] Yet A-ring-first entry is unlikely; otherwise, the native substrate **1** could presumably enter this way and become hydroxylated at positions other than 7α. Thus the 'tilted' binding modes are the best hypotheses for how CYP7B1 7β-hydroxylates **3** and 11α-hydroxylates **7**.

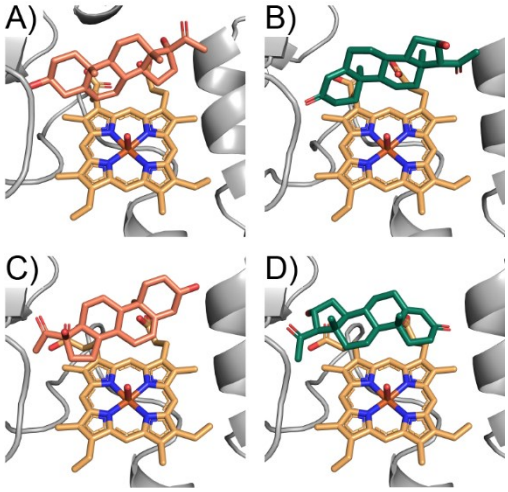


Figure 3.5. Candidate binding modes to explain 7β-hydroxylation of **3** and 11α-hydroxylation of **7** relative to the CYP7B1 heme. A) **3** binding in 'tilted' mode. B) **7** binding in 'tilted' mode. C) **3** binding in 'flipped' mode. D) **7** binding in 'flipped' mode.

These structure models suggest that unexpected CYP7B1 activities on **3** and **7** arise from these substrates shifting out of a tight pocket and tilting relative to the reactive ferryl oxygen. Cytochrome P450s typically hydroxylate the carbon closest to the ferryl oxygen; the mechanism does not require any activation of the substrate, though the transient carbanion intermediate formed favors carbons adjacent to double bonds.[162] As C-7 is adjacent to a double bond in native substrate **1** and the alternative substrate **3**, it is a particularly good hydroxylation site, and thus observed in their products (Fig. 3.6BC). However, **3**'s 17α-hydroxyl does not permit it to occupy

the same binding position as **1** in our model. The hydroxyl forces **3** into a wider channel to the left in the figure, where the steroid can tilt to result in 7β-hydroxylation (Fig. 3.6D). **7**'s 16α-hydroxylation would also result in a steric clash, which implies it will also shift out of **1**'s binding pocket. The severity of this clash may drive **7** to rotate dramatically, resulting in the observed 11α-hydroxylation (Fig. 3.6E).

Overall, the hydroxyls' steric clashes appear to drive steroid positioning more than their potential for hydrogen bonding. The D-ring hydroxyls of **3** and **7** do not form hydrogen bonds in any of the model structures, but nevertheless change these substrates' binding modes. The side view of CYP7B1's native substrate, pregnenolone (**1**), shows its tight fit in a hydrophobic binding pocket with little room beneath the D-ring (Fig. 3.6F). This fit may hold the substrate steady to ensure hydroxylation exclusively at position 7α. Functional groups at C-16 and C-17, such as those in **3** and **7**, can disrupt this fit to cause the observed loss of reaction specificity.

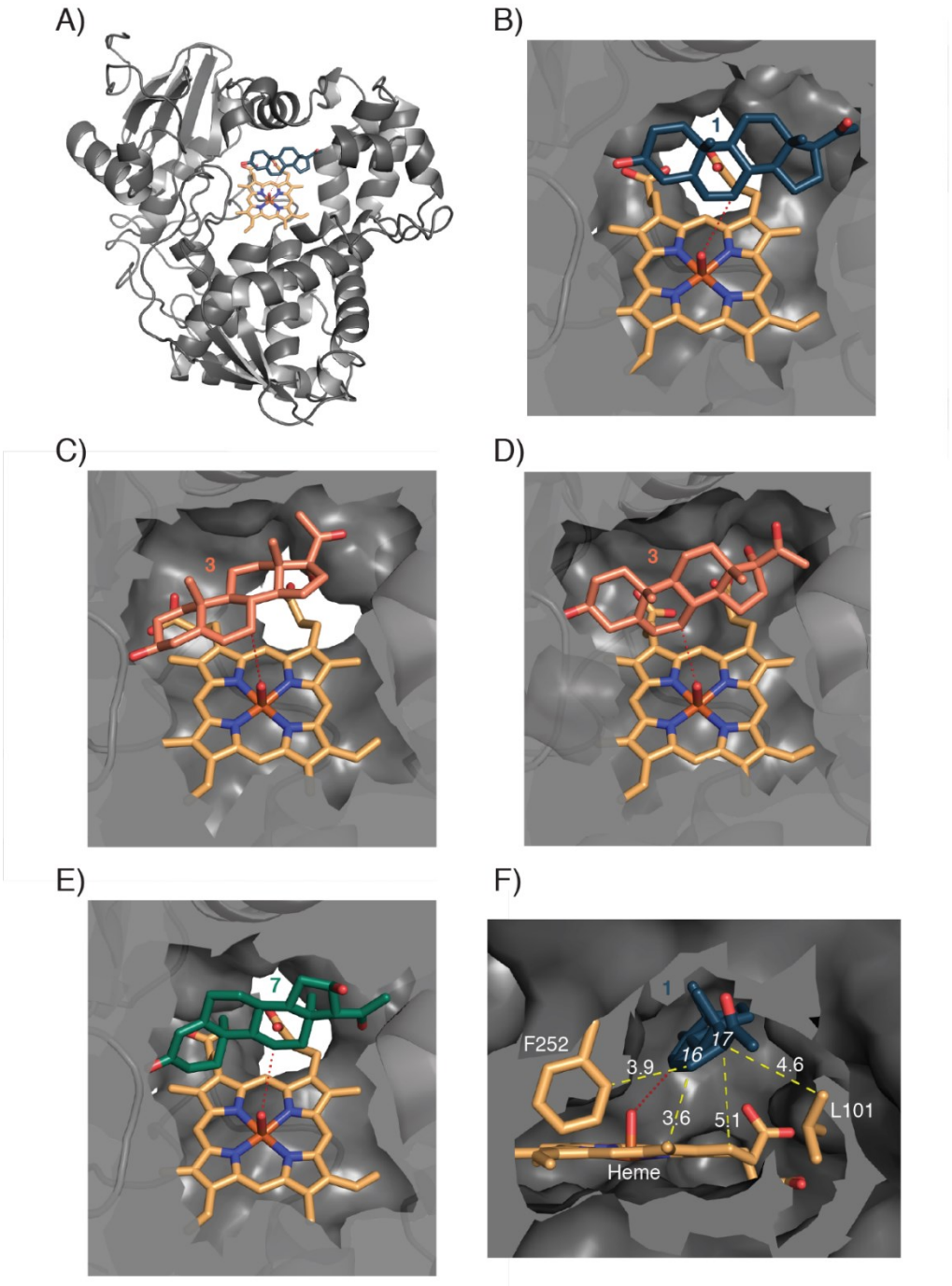


Figure 3.6. Predicted structures of the CYP7B1 active site with steroid substrates **1** (blue), **3** (orange), and **7** (green). A) Cartoon representation of holo-CYP7B1 with **1**. B-E) Structural models of CYP7B1 permitting 7α-hydroxylation of **1** (B) and **3** (C), 7β-hydroxylation of **3** (D), and 11α-hydroxylation of **7** (E). The red dashed line highlights the positions of the reactive ferryl oxygen and the observed hydroxylation site. F) Side view of native substrate **1** in active site. Yellow dashed lines indicate the proximity of steroid carbons C-16 and C-17 to the edge of the binding pocket. Distances are in Angstroms.

3.4 Discussion

Using Rosetta, I generated a structure model to explain the 7β-hydroxylation of 17α-hydroxypregnenolone (**3**) and the 11α-hydroxylation of 16α-hydroxyprogesterone (**7**) by CYP7B1. This human enzyme was previously thought to exclusively 7α-hydroxylate steroids, such as its native substrate pregnenolone (**1**). Based on the docking results, it is most likely that the hydroxylations below the D-ring in **3** and **7** push the steroid out of a tight binding pocket. This causes **3** and **7** to tilt relative to the native substrate **1**, permitting the unusual hydroxylations.

There are several residues in CYP7B1 that could be mutated to test whether the bulky residues near the pregnenolone D-ring influence hydroxylation regiospecificity. The heme is immediately below C-16 in the pregnenolone binding mode, but shrinking residues above the steroid could allow for more space. Leucine 182 and phenylalanine 252 could be mutated to alanine and leucine, respectively. These might enable 7α-hydroxylation of **7** and an increase in the 7α-hydroxylation of **3** relative to 7β-hydroxylation. However, many of the atoms bordering the D-ring are in the backbone of an alpha helix, and thus are difficult to intentionally shift. Another way to support this theory would be to replace small residues with larger amino acids. This might shift **1** to the wider area, enabling 7β- or 11α-hydroxylation. In particular, replacing glycine 251 with isoleucine could block a large part of the native substrate binding pocket.

While it is frustrating that Rosetta alone cannot distinguish a best binding mode, the set of equal modes may reflect actual positions the steroids occupy. As previously discussed, cytochrome P450s have flexible active sites resembling hydrophobic slots.[163] Thus the steroid might be capable of occupying many of the spaces indicated by the preliminary model set, at least temporarily. These model structures also show steroids may occupy space above and to the

left of the heme (Fig. 3.4). This is consistent with the established mechanism by which the steroid enters the active site. CYP7A1 has a hydrophobic tube in this position that extends from the heme to the top of the protein where it borders a endoplasmic reticulum membrane.[168] It is reasonable to believe that the homolog CYP7B1 would also take up its substrates from this direction, so Rosetta's placement of steroids in that area is rooted in a real phenomenon.

Rosetta's inability to distinguish steroid binding modes is not unusual, as it is extremely difficult to predict cytochrome P450 action *de novo*. Cytochrome P450 modeling is a field of intense study, largely because of its implications in identifying toxic drug metabolites. For those enzymes with large, extremely flexible active sites, it is helpful to consider which carbons in a substrate are most likely to react.[169] However, this is inappropriate for the more specific active sites of the steroid-modifying cytochrome P450s. Models using programs such as GOLD, AutoDock, or FlexX consider substrate orientation in the active site, typically making the assumption that carbons within 6 Å of the heme iron can be processed. Unfortunately, these only predict about 80% of modification sites within their top two results.[170] In addition, these docking models are heavily reliant on available crystal structures bound to similar substrates, in part due to active site flexibility.[171] As CYP7B1 had no available crystal structures, we doubt that these other methods could have outperformed Rosetta.

Our method for predicting alternative substrate binding modes in CYP7B1 could easily be repeated for other steroid-modifying cytochrome P450s. New enzyme activities can be found during indirectly related efforts, so their discoverers may not be interested in pursuing the reaction mechanism experimentally. Rosetta is not commonly used to dock ligands into existing cytochrome P450s, though it has been used to engineer the overall fold of novel

metalloenzymes.[172] This fast and relatively simple method for modeling an unexpected reaction will be helpful to researchers who cannot investigate it experimentally.

3.5 Methods

We generated comparative models for CYP7B1 binding steroids based on crystallized enzymes with Rosetta[165]. The algorithm selects for structures with optimal physical interactions, solvent accessible surface area, and bond angles by assigning structures a Rosetta score; lower Rosetta Energy Unit (REU) values suggest more energetically favorable conformations.[165] The two solved structures within 30% sequence identity and 10^{-4} BLAST confidence, and bound to chemically similar substrates, were PDB IDs 3sn5 and 3v8d[166]. We threaded the sequence of CYP7B1 onto the coordinates of those structures and combined fragments of each to create ten thousand hybrid models of CYP7B1[164,165]. The heme was inserted according to the original crystals' bond angles and obabel-derived partial charges; this was followed by all-atom optimization[173]. ~2500 reasonable structures were identified by their Rosetta score falling within fifty REU of the lowest-scoring structure and were clustered at 3 Å RMSD.

An ensemble of the twenty best-scoring CYP7B1 conformations were used in docking **1**, **3**, and **7**, generating ten thousand binding predictions for each steroid. Typical parameters for these RosettaLigand docking runs were used: the steroid was randomly placed within 5 Å of the binding pocket center, coarsely fit inside, and accommodated by sidechains within 5 Å[164]. Energetically feasible binding modes were identified by their overall Rosetta score within 65 REU of the best mode by overall score (to account for variations between the ensemble models), and their ΔG-binding analog score within 4 REU of the best mode (analogous to the score contribution of one hydrogen bond). These structures were clustered at 3 Å RMSD between

substrate atoms, and the cluster representative was chosen by the lowest ΔG-binding analog score.

RMSD clustering for protein structures was performed by first finding the distance between each structure and every other. PyMOL was used to align all structures, then we computed the root mean squared Euclidean distance between every carbon. The clustering was then performed by complete hierarchical linkage, extracting cluster branches with a cutoff RMSD of 3 Angstroms. When clustering binding modes by RMSD, the enzymes themselves are aligned according to the active site coordinates. The RMSD between each binding mode is then found as the root mean squared Euclidean distance between every atom shared in the substrate molecules.

This resulted in approximately seventy energetically comparable structures per steroid. As it was not possible to determine the best binding mode by energy alone, other information was incorporated to identify the modes shown. We selected conformations that placed the hydroxylation site within 3.5 Å of the ferryl oxygen, as these are the only modes capable of mediating the observed reactions. We avoided conformations that would require unlikely substrate trajectories into the active site, instead choosing those that resembled the orientations of substrates in the homologous crystals (as discussed in the results section). Thus we produced a binding mode for each hydroxylation mode for each substrate, for a total of four models. Please refer to the supplemental files for structure model files for all apo-enzyme conformations as well as the final binding modes.

3.6 Acknowledgements

This work was supported by the Defense Advanced Research Projects Agency HR0011-12-C-0061 and the National Institute of Health P50 GM107618 (Laboratory of Systems Pharmacology) and R01 GM086258 (JC).

4

Pharmacokinetic Model of a Glucocorticoid-Binding Antibody Fusion Protein

Emma S. Spady, Jeffrey C. Way, and Pamela A. Silver

4.1 Attributions

The anti-inflammatory antibody fusion protein was initially designed by Dr. Jeffrey Way. We agreed it would be interesting to model its action, so I set up the system and found the relevant constants from the literature. Some initial fusion protein constructs were expressed, but their pharmacokinetically-relevant parameters were never measured. This work is entirely unpublished.

4.2. Introduction

Anti-inflammatory steroids efficiently treat autoimmune conditions, but have numerous side effects due to activity in off-target tissues. These drugs mimic endogenous cortisol to bind to the glucocorticoid receptor (GR) located in the cytoplasm of leukocytes, hepatocytes, and osteoblasts. The GR then translocates to the nucleus, where it decreases immune response in leukocytes, but triggers glucose release in hepatocytes and decreases bone growth in osteoblasts.[25] This system prepares the body for stress, increasing blood sugar for fight-or-flight while decreasing costly maintenance functions. These global effects of glucocorticoids, also

known as corticosteroids, prevent their long-term use outside of local application.[12] External surfaces, such as the skin and respiratory tract, can be targeted by using formulations that prevent steroids from reaching the bloodstream or steroid derivatives that degrade quickly.[83,86] Yet many common and debilitating autoimmune diseases, such as lupus or rheumatoid arthritis, require glucocorticoids to reach cells deep inside the body.

Antibody fusion proteins, which can selectively bind desired cell types to deliver a cargo domain, have resolved these targeting problems for cytokines. The strategy taken for these 'chimeric activators' could improve many classes of therapeutics. Essentially, these drugs consist of a signaling domain fused to an antibody that exclusively binds a target cell. Mutations decrease the signaling domain affinity, so the relevant receptor only activates upon antibody binding. For example, erythropoietin signaling on red blood cell precursors alleviates anemia, but erythropoietin binding to megakaryocytes increases thrombosis.[174] To improve therapeutic index, Burrill *et al.* constructed an antibody fusion protein consisting of anti-glycophorin A, which exclusively binds hematopoietic cells, and an erythropoietin with decreased receptor affinity.[175] Similarly, Garcin *et al.* made an interferon alpha-2 with decreased affinity fused to an anti-PD-L2 nanobody, which decreases leukopenia in mice while triggering STAT1 activation in select splenocytes.[176]

Antibody-drug conjugates (ADCs) can carry small molecules to desired cell types, using covalent linkers to bind the molecule to the antibody. These face several obstacles that do not apply to the chimeric activators described above. Firstly, the small molecule component of an ADC must have a site which can be modified to attach to the antibody. The covalent linker must then be cleaved to release the drug, and the small molecule must usually enter the cell to take effect.[97] The current approved ADCs are used in cancer treatment, delivering cytotoxic

compounds to tumor cells.[177] However, research into anti-inflammatory ADCs is gaining traction.[98] ADCs can deliver promiscuous kinase inhibitors, such as dasatinib or PDE4 inhibitors, exclusively to leukocytes, preventing side effects.[111,112] Antibody-glucocorticoid conjugates have been studied in recent years. An anti-CD163-dexamethasone targets macrophages and is capable of sparing hepatocytes.[99,100] A carefully-designed fluticasone derivative conjugated to an anti-CD74 antibody selectively affects B cells and not T cells, even when they are co-cultured.[101]

With these concepts in mind, Jeff Way and I envisioned a glucocorticoid-carrying antibody fusion protein. This single polypeptide would consist of an anti-leukocyte single chain variable fragment (scFv) and a glucocorticoid-binding domain (Fig. 4.1). Glucocorticoids could be loaded onto the fusion protein, and the protein would expose only cells expressing the relevant antigen to the small molecule inside. The non-covalent glucocorticoid attachment could allow the protein to pick up stray steroid diffusing out of the desired cells, or even to carry endogenous cortisol to the target. While this drug is not a covalent ADC, it uses an antibody to deliver a small molecule, and thus ADC-related methods are relevant to its design.

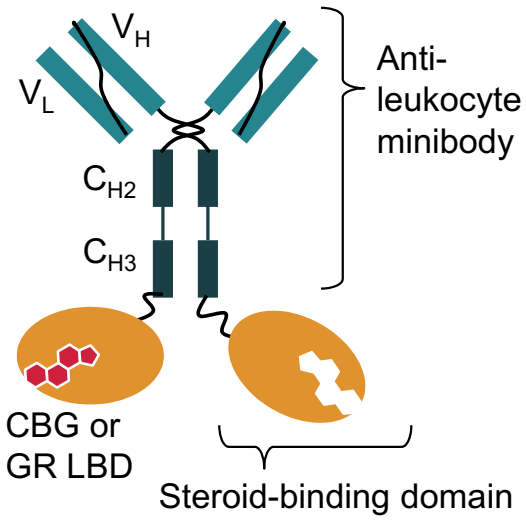


Figure 4.1. Schematic of a steroid targeting fusion protein.

Pharmacokinetic and molecular mechanism models can predict which features should be included in this antibody fusion protein drug. The basic design concept allows for many choices; the antibody target, the glucocorticoid, and the protein's steroid and antigen affinities could dramatically change the drug action. Modeling relevant compartment pharmacokinetics and

biochemical mechanisms could direct these choices, resulting in a more effective design. Rate constants, abundances for different molecules, and compartment volumes are available from literature, which can be assembled into an ordinary differential equation (ODE) model.[107,108] Pharmacokinetic and mechanistic models of ADCs are necessarily complex. The protein component's target binding affects the distribution of different drug subspecies. The small molecule component, when freed, also has its own pharmacokinetic properties.[105] Compartment-level distributions must be integrated with subcellular compartments and processes, resulting in a multiscale systems pharmacology model for even this relatively simple system.[178]

I used the dMod package in R to model the pharmacokinetics of a glucocorticoid-binding antibody fusion protein. The protein design is distinct from traditional ADCs in that free steroid can reassociate with the protein, adding to the system complexity. This model predicts several features that would improve drug efficacy, such as targeting a fast-internalizing antigen and loading a less permeable steroid cargo. It also suggests that protein without pre-loaded steroid could be used to carry endogenous cortisol to desirable cell types.

4.3 Results

This model considers a fusion protein consisting of an anti-CD45 minibody and a glucocorticoid binding domain (Fig. 4.2). CD45 was selected as the antibody target because it is found on all leukocytes, and no off-target cell types, such as hepatocytes or osteoblasts.[179] The BC8 variable region was used in this initial protein design, and was chosen based on its successful delivery of radioactive compounds to CD45-expressing cancers.[180] The IgG constant

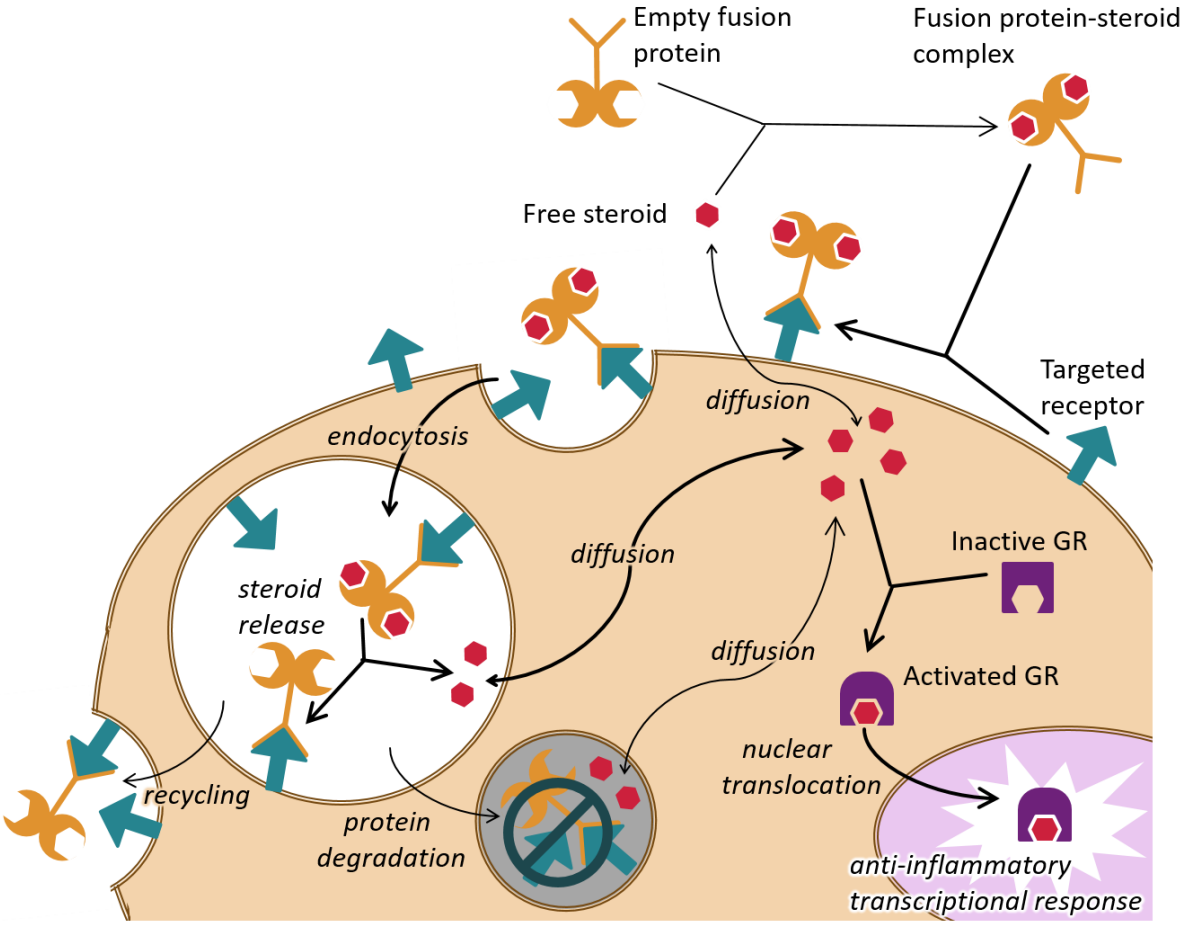


Figure 4.2. Schematic for relevant processes in steroid delivery via a glucocorticoid-binding antibody fusion protein.

fragment of the minibody enables neonatal Fc receptor recycling, and thus an extended serum half-life.[181] The steroid-binding domain could be based on corticosteroid binding globulin (CBG) or the steroid-binding domain of GR. The model uses the steroid on- and off-rates associated with GR, as its affinities to various synthetic glucocorticoids are better characterized.[182] This protein will deliver fluticasone propionate, chosen because of its high affinity to glucocorticoid receptor and its relatively short half-life in humans.[183] A wide variety of other synthetic glucocorticoids are available, and any could theoretically be loaded into this fusion protein if different properties are required. The engineered protein will form a homodimer, with each unit

binding two CD45 receptors and two glucocorticoid molecules. To simplify the model, each molecule of protein is represented as a unit capable of binding one receptor and one glucocorticoid molecule.

The R package dMod was used to generate and solve an ordinary differential equation (ODE) model, which provides a concentration over time for drug-derived molecules in on- and off-target cells.[184] The relevant cells and tissues are simplified into a system of four compartments, as is typical in pharmacokinetic models. The fusion protein begins in the 'blood' compartment and delivers steroid to 'on-target cytoplasm' via internalization into an 'on-target endosome'. Free steroid can diffuse from the 'blood' to an 'off-target cytoplasm', which has twice the volume of the 'on-target cytoplasm'. The fusion protein may bind steroid and antigen. The steroid can also bind the glucocorticoid receptor in the on- and off-target cytoplasm compartments. I used these constraints to construct the ODE model, in which eight chemical species interconvert via 39 chemical reactions with 22 rate constants (Fig. 4.3). The rate constant and volume parameter values have varying degrees of accuracy, internally and with respect to this application. Nevertheless, this simplified approach can suggest which aspects of a glucocorticoid-targeting antibody drug could be altered to improve its function. In the model, this efficacy is reflected in the difference in glucocorticoid receptor occupancy between on- and off-target cells.

An increased rate of target receptor endocytosis is critical to the fusion protein's ability to deliver glucocorticoid. CD45 receptor is not actively endocytosed upon antibody binding, resulting in a slow fusion protein uptake and extended half-life. Antibody fusion protein bound to CD45 spends too long in the 'blood' compartment, which allows the glucocorticoid to dissociate, diffuse across the off-target cell membrane, and bind to the off-target glucocorticoid

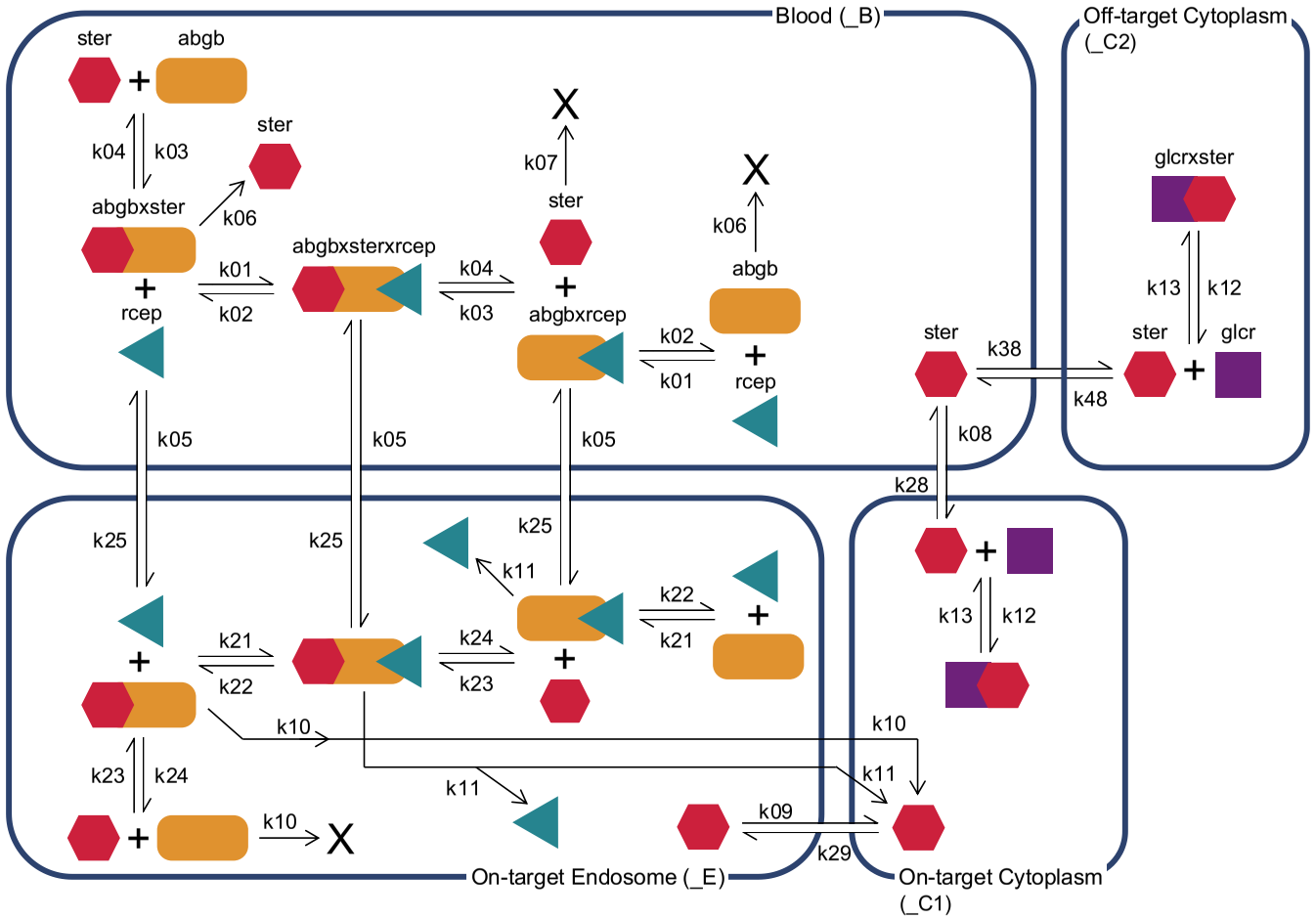


Figure 4.3. Diagram of the glucocorticoid-binding fusion protein pharmacokinetic model. Red hexagons represent steroid, yellow rounded rectangles represent fusion protein, teal triangles represent target receptor, and purple squares represent GR. Species may be depicted multiple times in the same compartment for clarity, but are in a shared pool in the model. A black X shows output for processes that remove all molecules involved. Labels denote the names of these species and their combinations in the code. Species names have appended suffixes denoting their compartment, as shown after the compartment name. Reversible processes are denoted by half-arrows, while irreversible degradation processes are shown as full arrows.

receptor. Fortunately, there are many receptors found on leukocyte subtypes that undergo clathrin-mediated endocytosis, such as CD163, CD11a and CD74. Antibody drugs have been designed to bind these receptors, so suitable scFv sequences exist that trigger internalization and degradation.[99,101,112] Rate constants from EGFR internalization were used to represent the speed of this process, as it is well-studied. Increasing the receptor internalization rate from 5.0×10^{-4} to 5.5×10^{-3} sec^{-1} and degradation rate from 2.26×10^{-5} to 1.65×10^{-4} sec^{-1} dramatically improves the maximum on-target glucocorticoid receptor occupancy (Fig. 4.4).

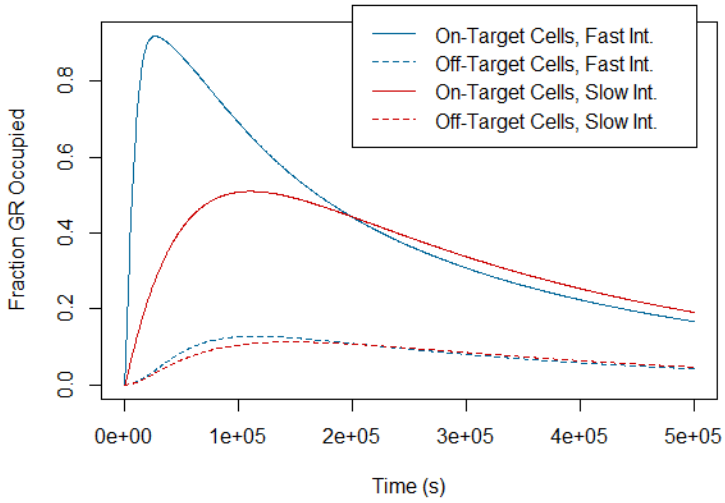


Figure 4.4. Increased receptor endocytosis and degradation improves the fraction of GR occupied in on-target versus off-target cells.

Histidine switching to release glucocorticoids in the endosome does not improve steroid targeting in this system. Histidine switching takes advantage of histidine protonation in the pH 5.5 environment of the endosome. Engineered proteins with correctly placed histidines will have increased dissociation rates for ligands in this compartment, potentially improving their pharmacokinetics.[185] In this system, release of antibody from the target receptor does not have much effect on GR occupancy, as the receptor-bound antibody degradation rate of 1.65×10^{-4} sec^{-1} is close to the free antibody degradation rate of 3.70×10^{-4} sec^{-1}. Thus it is only appropriate to consider whether histidine switching to release steroid from the fusion protein could improve targeting. Initially, the effect of a ten-fold increase in steroid-fusion protein dissociation rate was modeled. This is generous, as the largest affinity difference between

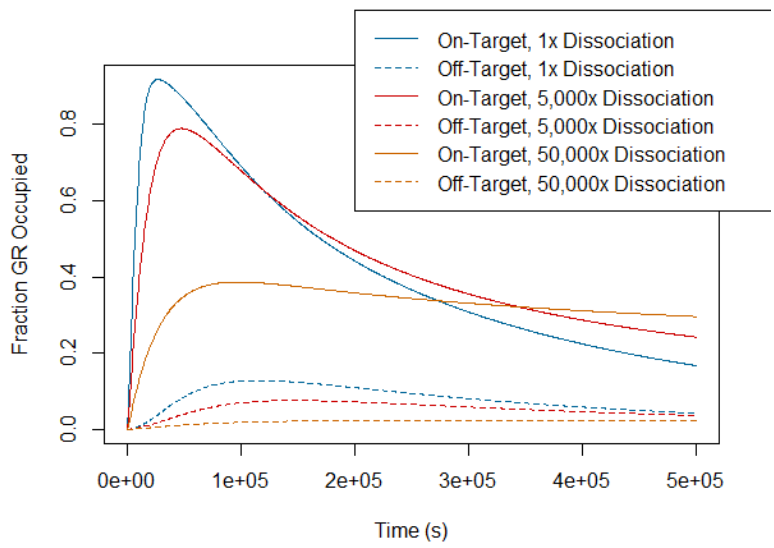


Figure 4.5. Increased dissociation of steroid in endosome does not improve GR occupancy in on-target cells.

pH 5.5 and pH 7.4 for a histidine switched protein is only 6.8-fold.[185] This ten-fold change made no difference in GR occupancy; a 5,000-fold increase in glucocorticoid off-rate is needed for an easily-observable change (Fig. 4.5). When the dissociation rate passes 1 sec^{-1}, at an increase of 50,000 from the original rate, the occupancy behavior shifts entirely, as shown. Unfortunately, a dissociation rate increase of even 100-fold would be hard to achieve without decreasing the steroid binding at pH 7. Thus histidine switching is unlikely to improve the fusion protein performance.

The small molecule-determined parameters of glucocorticoid receptor affinity, steroid-fusion protein affinity, and permeability can change system behavior substantially. This allows for potential improvements from a change in glucocorticoid. Decreasing steroid permeability increases targeting efficacy, such that a steroid that diffuses half as quickly increases on-target receptor occupancy (Fig. 4.6). However, changing steroid parameters might necessitate selecting a molecule with less affinity for the fusion protein and the glucocorticoid receptor. Thus I considered the system tolerance to a faster off-rate from both of these binding domains. A steroid with a two-fold or even four-fold faster GR and fusion protein off-rate is capable of accumulating in on-target cells, even with the original fluticasone permeability constant (Fig.

4.7). Hence this fusion protein could be loaded with different glucocorticoids if decreased permeability or other properties are required.

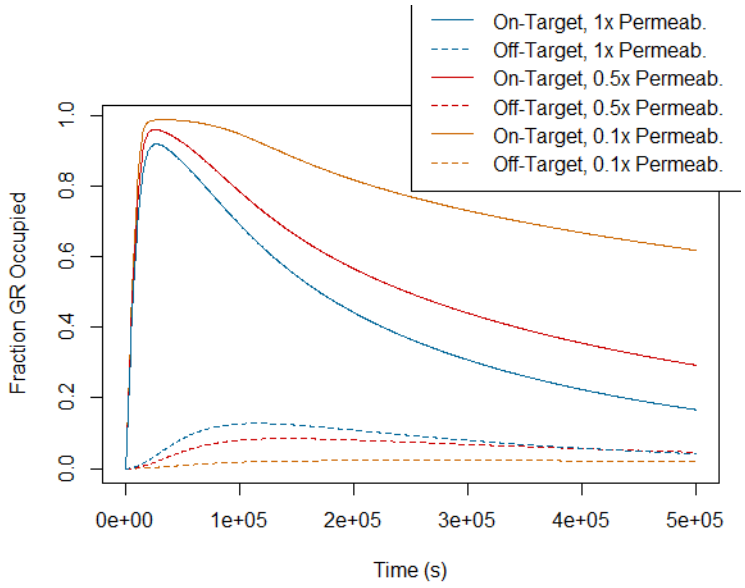


Figure 4.6. Decreased steroid permeability, resulting in a proportional decrease in diffusion rate, increases GR occupancy in on-target cells and decreases occupancy in off-target cells.

Fusion protein without pre-loaded steroid could use endogenous cortisol to increase on-target cell GR occupancy. This concept was considered by adding fusion protein without pre-bound steroid to a system with free cortisol in the 'blood' compartment. However, entering a 'blood' cortisol concentration of 4.14×10^{-8} M, the average in human plasma, occupied more than 60% of GR in on- or off-target cells (Fig. 4.8).[186] This is likely due to the oversimplified nature of the model. In order to consider this scenario further, an initial 'blood' compartment cortisol concentration of 4×10^{-9} was used. This occupies 16% of GR in all cells when fusion protein is absent. With the

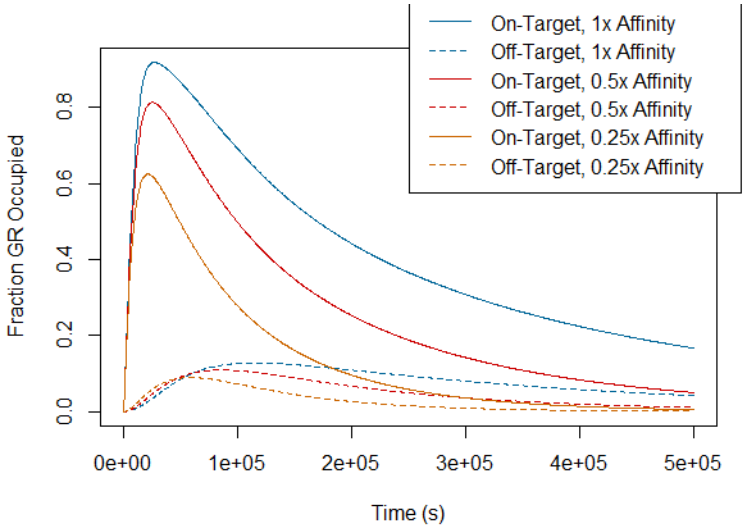


Figure 4.7. Targeted delivery decreases but is not eliminated with decreased steroid affinity, as represented by a faster steroid-fusion protein dissociation rate and a faster steroid-GR dissociation rate.

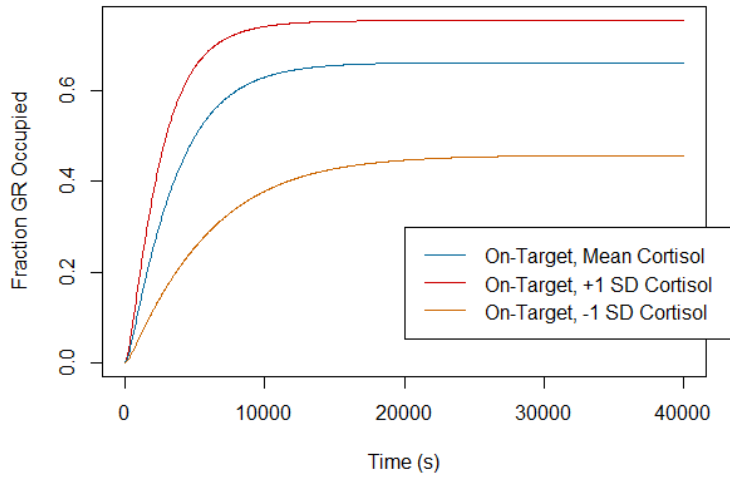


Figure 4.8. Literature values of cortisol in blood flood GR in all cells in this model, requiring approximate lower values for downstream use.

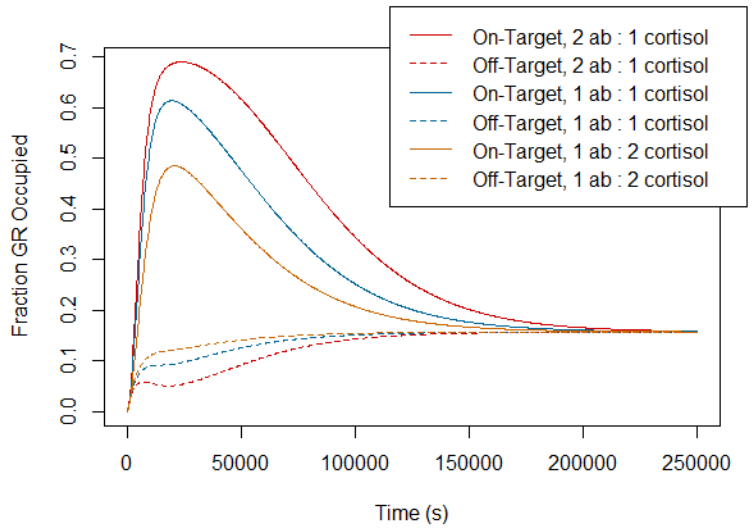


Figure 4.9. Fusion protein can carry endogenous cortisol to occupy on-target cells' GR, even when the cortisol concentration exceeds the fusion protein concentration.

assumption that total cortisol is in a steady state, the rate of free steroid degradation in the blood was set to zero. The steroid-GR on- and off-rates were adjusted for cortisol, which has a lower receptor affinity than fluticasone propionate.[187] GR occupancy in two-to-one, one-to-one, and one-to-two ratios of fusion protein to antibody are compared in Figure 4.9. Increased fusion protein results in higher maximum GR occupancy and longer occupancy difference between on- and off-target cells. While large quantities of fusion protein would be needed to bring cortisol selectively to the targeted cell type, it is theoretically feasible given these system parameters.

4.4 Discussion

This ODE model identified several important characteristics that should improve a glucocorticoid-binding antibody fusion protein. Firstly, the antigen chosen must be endocytosed quickly upon protein binding. A less-permeable glucocorticoid derivative could improve targeting, even if it had lower affinity for the receptor or the protein. Meanwhile, releasing glucocorticoid or fusion protein in the endosome via histidine switching does not increase selectivity. In addition, the fusion protein could shepherd endogenous cortisol to the antigen-bearing cell type.

The ODE model is relatively sensitive to the values of certain essential rate constants, and the available literature values might not be accurate for this system. In particular, the free steroid diffusion rate between compartments affects the GR activation cell-type specificity. This parameter depends on the surface area of the compartment, which requires accurate measures of the targeted cell type. It also requires an estimate of a permeability constant.[101,188] While there are references for all of these calculations, it is hard to ascertain if the diffusion rates are correct. In addition, the rates of target antigen endocytosis and free antibody degradation in the endosome are both rough estimates. It can be difficult to find even apparently simple characteristics, such as volume, surface area, and abundance of a given cell type. While the BioNumbers website aggregates this type of information, it was often missing the values this model required.[189]

Throughout this project, I noticed a distinct lack of software packages with support for multiscale pharmacology ODE models. The concept is not difficult, nor are the relevant mathematics, but most packages are not adapted for both separate compartments and many molecule types. The number of distinct chemical species from the fusion protein binding antigen

and steroid overwhelmed MATLAB's SimBiology GUI, despite its intuitive handling of multi-compartment systems. BioNetGen's RuleBender had excellent shortcuts for indicating combinatorial chemical species, but I had trouble with compartmentalization even with expert advice.[190] In addition, the RuleBender integrator behaved oddly when working with small numbers, which are necessary when working in terms of molarity rather than molecules. After switching to R, my strongest language, I found dMod, a flexible and powerful chemical ODE generator.[184] Despite the straightforward addReaction() function, I found the documentation inaccessible. dMod also has bug in which intercompartment transfer rates must be divided by the volume of the destination compartment. I found this through trial and error when glucocorticoid failed to distribute evenly between compartments, even without other molecules in the system. Given that multi-scale pharmacokinetic models are quite easy to imagine, and the mathematics is easier still, the lack of intuitive packages is surprising.

The fusion protein modeled here was considered before a recent wave of research on antibody-glucocorticoid conjugates. Jeff Way first proposed the protein design in spring 2016, and I developed the model that September. At the time, we were only aware of the Moestrup group's 2012 paper on an anti-CD163-dexamethasone ADC.[99] In September 2016 and March 2017, the Moestrup group published further research on this drug, finding its ability to treat liver inflammation.[91,100] The aforementioned anti-CD74-fluticasone phosphonate, from a team at Merck, was published in 2018.[101] While our project halted in 2017, considerations of antibody-based methods for glucocorticoid targeting were clearly prescient.

If this project had continued, the next step would be to measure relevant biochemical and physiological constants. Once protein designs were expressed, *in vitro* interferometry experiments using the ForteBio BLItz could accurately measure rate constants for antigen and

glucocorticoid binding. Rates of receptor internalization and antibody degradation after endocytosis could be measured in tissue culture by incorporating a fluorescent domain into the fusion protein. The dMod package can fit pharmacokinetic parameters, such as compartment volumes and diffusion rates, to results from *in vivo* experiments.[184] These measurements would reveal the weaknesses of the model, and accounting for these discrepancies would improve the model. The improved model could then direct new design features for the protein, creating a more efficient drug development process. Using these methods, this fusion protein could become the first non-covalent antibody-based small molecule delivery system.

4.5 Methods

The protein drug design was used to set up a pharmacokinetics compartment system for the ODE model (Table 4.1). The 'blood' compartment contains the total volume of blood and extracellular fluid in an average human, about 18 liters.[191] The antibody-steroid drug begins in this 'blood' compartment, as antibody drugs are typically injected. The cytoplasm of all leukocytes combined forms an 'on-target cytoplasm' compartment, which excludes the volumes of the cells' nuclei. The 'on-target cytoplasm' has a total volume of 0.059 L and a plasma membrane surface area of 13,000 dm^2, which faces the 'blood' compartment. These on-target cells have a relevant 'endosome' compartment, consisting of their pooled endosome volumes. This 'endosome' has 0.0033 L of volume and 3900 dm^2 of surface area, which separates it from the 'on-target cytoplasm'. Table S2 in the appendix contains details on the calculation of these leukocyte-related values. Finally, an 'off-target cytoplasm' was defined as representing cells in which glucocorticoids have an undesirable effect. The 'off-target cytoplasm' has twice the volume and surface area of the 'on-target cytoplasm', in order to simplify comparisons. The 'off-

target cytoplasm' has the same access to the 'blood' as the 'on-target cytoplasm'. There is no off-target endosome represented because without receptor, the endosome does not play a significant role in steroid distribution.

Table 4.1. Volumes, surface areas, and initial concentrations used in the ODE model of fusion protein steroid delivery. Refer to Table S2 in the appendix for calculations and original literature values.

Parameter	Value	Sources	Name
Total human extracellular volume (L)	18	Davies 1993[191]	VolB
Total leukocyte cytoplasm volume in human (L)	5.9E-02	Ting-Beall 1993[192], Segel 1981[193], Schmid-Schonbein 1980[194]	VolC1
Total leukocyte endosome volume in human (L)	3.3E-03	Corlier 2015[195], Sarkar 2003[196]	VolE
Total leukocyte plasma membrane surface area in human (dm^2)	1.3E+04	Schmid-Schonbein 1980[194]	SAplasma1
Total leukocyte endosome surface area in human (dm^2)	3.9E+03	Yogurtcu 2018[197]	SAendo
Ratio of off- to on-target cells	2	N/A	OfftoOnRatio
Initial concentration of CD45 receptors in blood (M)	3.99E-09	Peyron 1991[198], Bausch-Fluck 2015[199], Matthews 1991[200]	initamts["rcep_B"]
Initial concentration of CD45 receptors in endosome (M)	6.54E-06	Peyron 1991[198], Bausch-Fluck 2015[199], Matthews 1991[200]	initamts["rcep_E"]
Initial concentration of glucocorticoid receptors in leukocytes (M)	8.99E-08	Biggin 2011[201]	initamts["glcr_C1"], initamts["glcr_C2"]
Initial concentration of antibody drug in blood (M)	1E-10	N/A	initamts["abgbxster_B"]

Eight chemical species are present in these compartments. Four free molecules are shown: steroid, fusion protein, antigen receptor, and glucocorticoid receptor. The fusion protein can bind or release the antigen receptor and the steroid, resulting in three more chemical species. Glucocorticoid receptor can also bind steroid, creating an occupied glucocorticoid receptor. Initial concentrations of all species are listed in Table 4.1. Glucocorticoid receptor is located in the cytosol of on- and off-target cells; it is assumed to be entirely free of steroid ligand at the beginning of the simulation and equally abundant in both cytosolic compartments. The ratio of occupied to total glucocorticoid receptor in a cytosol compartment is used as a proxy for steroid response.

These chemical species interconvert and move between compartments via 39 reactions, which were entered into the R package dMod.[184] The reactions use 22 rate constants, which can be found in Table 4.2. Some of these constants were used as-is or required only a unit conversion. For calculated values, parameter inputs and calculation methods are shown in Table S2 in the appendix. The code used to generate the model and compare the effects of parameter changes can be found in the supplemental files. The chemical concentrations are in terms of molarity, volume is in liters, surface area is in dm^2, and time is in seconds. Rate constants are in sec^{-1} or $M^{-1}sec^{-1}$, as appropriate. The dMod package solves the ODE system generated by the input reactions, giving the concentrations of each species in each compartment over time.

Table 4.2. Rate constants used in the ODE model of fusion protein steroid delivery. Asterisks indicate calculations more complex than unit conversions from the literature; refer to Table S2 in the appendix for more information.

Rate Constant	Value	Units	Source	Name
Antibody-receptor association rate	1.00E+06	$L\ mol^{-1}\ sec^{-1}$	Raman 1992[202]	k01, k21
Antibody-receptor dissociation rate	3.86E-04	sec^{-1}	Matthews 1991[200], Lin 2006[180]	k02, k22
Glucocorticoid binding domain-steroid association rate for fluticasone propionate	3.98E+04	$L\ mol^{-1}\ sec^{-1}$	Hogger 1994[182]	k03, k23
Glucocorticoid binding domain-steroid dissociation rate for fluticasone propionate	2.20E-05	sec^{-1}	Hogger 1994[182]	k04, k24
Glucocorticoid binding domain-steroid association rate for cortisol	3.76E+04	$M^{-1}\ sec^{-1}$	Eliard 1984[187]	k03, k23
Glucocorticoid binding domain-steroid dissociation rate for cortisol	7.71E-04	sec^{-1}	Eliard 1984[187]	k04, k24
Receptor internalization rate for actively endocytosed receptor	5.50E-03	sec^{-1}	Sigismund 2008[203]	k05 (fast)
Receptor externalization rate for actively endocytosed receptor	1.83E-02	sec^{-1}	Sigismund 2008[203]*	k25 (fast)
Receptor internalization rate for passively endocytosed receptor	5.00E-04	sec^{-1}	Alberts 2002[204]	k05 (slow)
Receptor externalization rate for passively endocytosed receptor	1.67E-03	sec^{-1}	Alberts 2002[204]*	k25 (slow)
Antibody drug degradation rate in blood compartment	2.01E-06	sec^{-1}	Keizer 2010[181]	k06
Steroid degradation rate in blood compartment for fluticasone propionate	3.08E-05	sec^{-1}	Allen 2013[183]	k07
Steroid degradation rate in blood compartment for cortisol	0	sec^{-1}	Assuming constant total cortisol amount	k07
Steroid diffusion rate from blood to on-target cytoplasm	1.61E-06	$M^{-1}\ sec^{-1}$	Brandish 2018[101]*	k08
Steroid diffusion rate from on-target cytoplasm to blood	4.91E-04	$M^{-1}\ sec^{-1}$	Brandish 2018[101]*	k28
Steroid diffusion rate from blood to off-target cytoplasm	6.44E-06	$M^{-1}\ sec^{-1}$	Brandish 2018[101]*	k38
Steroid diffusion rate from off-target cytoplasm to blood	9.83E-04	$M^{-1}\ sec^{-1}$	Brandish 2018[101]*	k48
Steroid diffusion rate from on-target cytoplasm to endosome	2.70E-08	$M^{-1}\ sec^{-1}$	Brandish 2018[101]*	k09

Table 4.2 (Continued).

Rate Constant	Value	Units	Source	Name
Steroid diffusion rate from endosome to on-target cytoplasm	4.83E-07	M^{-1} sec^{-1}	Brandish 2018[101]*	k29
Free antibody degradation rate in endosome	3.70E-04	sec^{-1}	Huotari 2011[205]	k10
Receptor degradation rate in endosome for actively endocytosed receptor	1.65E-04	sec^{-1}	Sigismund 2008[203]	k11 (fast)
Receptor degradation rate in endosome for passively endocytosed receptor	2.26E-05	sec^{-1}	Minami 1991[206]	k11 (slow)
Steroid to glucocorticoid receptor association rate for fluticasone propionate	3.98E+04	M^{-1} sec^{-1}	Hoffer 1994[182]	k12
Steroid to glucocorticoid receptor dissociation rate for fluticasone propionate	2.20E-05	sec^{-1}	Hoffer 1994[182]	k13
Steroid to glucocorticoid receptor association rate for cortisol	3.76E+04	M^{-1} sec^{-1}	Eliard 1984[187]	k12
Steroid to glucocorticoid receptor dissociation rate for cortisol	7.71E-04	sec^{-1}	Eliard 1984[187]	k13

5

Conclusion

In this brief conclusion, I will suggest how we can make good on the promise of steroid drugs. In the introduction, I cover two approaches towards the goal of developing steroid-based drugs with fewer side effects. The first approach is to develop more selective steroid receptor modulators (SSRMs). Steroidal SSRMs are underexplored due to challenges in regio- and stereo-selective synthesis. Chapter 2 demonstrates a potential solution by using steroid biosynthetic enzymes to intentionally generate novel molecules. This could make new modifications around the steroid ring structure accessible for medicinal chemistry optimization. The second approach is to use sophisticated macromolecules that deliver steroids exclusively to on-target cells. These macromolecules are essentially altering the steroid pharmacokinetics, so modeling these properties should accelerate drug development. Chapter 4 shows a pharmacokinetic model for an innovative antibody fusion protein with reversible steroid binding, which could carry synthetic or endogenous glucocorticoids to leukocytes. The biosynthetic and pharmacokinetic techniques used in this dissertation are not recent developments, but had never been applied to steroids.

This naturally leads to a question: Why are there not improved steroid drugs already? Few innovations in SSRMs, outside of SERMs, have reached the market. This is surprising, given the clinical importance of androgens, progestins, and especially glucocorticoids. Part of this can be attributed to an emphasis on cancer research at the expense of other conditions and diseases. However, I believe much of the problem lies in a dearth of interdisciplinary research. If

organic chemists and biochemists were to collaborate more frequently, the use of biocatalysts as tools to synthesize novel steroids would be obvious. Similarly, multi-scale pharmacokinetics models would always be used to steer macromolecule drug development if more biologists and biochemists were trained in computer programming.

It is not a coincidence that all the SSRMs on the market are selective estrogen or androgen receptor modulators (SERMs / SARMs) with applications in cancer treatment. These SERMs and SARMs are primarily developed to treat breast and prostate cancer, respectively, which are serious public health concerns. 2,015,013 new cases of breast cancer and 2,128,974 cases of prostate cancer were diagnosed in Americans between 1999 and 2008.[207] Nevertheless, systemic glucocorticoids are used by a similar number of Americans. Between 1999 and 2008, 2,513,259 Americans were prescribed oral anti-inflammatory steroids.[9] Cancer research generally receives more funding in the United States than research on autoimmune disorders, many of which could be treated with improved glucocorticoids. The National Institutes of Health spent $5,389 million on cancer research in 2015, compared to only $822 million on autoimmune disease research.[208] However, these conditions affect similar numbers of people. 10.3 million people alive today in the U.S. have had a cancer diagnosis in the past thirteen years, and cancer killed 9.6 million people between 1999 and 2015.[207] Meanwhile, autoimmune diseases affect 23.5 million people in the U.S.[208] Some funding discrepancy can be attributed to cancer's potential lethality, whereas many of the inflammatory conditions degrade quality of life long before they kill. Nevertheless, autoimmune disorders cause a great deal of suffering, which could be alleviated by improved steroid drugs. Increased funding for the treatment of autoimmune diseases would help selective glucocorticoid receptor modulators get the attention and research they deserve.

Collaboration between biochemists and organic chemists will allow biocatalysts to become widely adopted in steroid synthesis. The steroid biosynthesis concepts demonstrated in this dissertation should be applied to more important steroid-modifying reactions. My reading during this project covered dozens of steroid-modifying enzymes; between them, this set of enzymes could theoretically alter any carbon in the steroid ring structure or tail. However, it was unclear which enzymes would be the most relevant to making interesting products. A collaborator with a background in organic chemistry might have known which atoms in steroids are particularly challenging to alter with organic methods. They could also recognize which enzymatically-added functional groups are conducive to further modification by synthetic means. As I could not distinguish which reactions would be the most interesting, I tested as many steroid-modifying enzymes as I could. I chose to pursue CYP7B1 partially because there were commercially-available substrates that could become novel products with a single 7α-hydroxylation. Unfortunately, 7α-hydroxylation is not as chemically interesting as some of the other enzymes' reactions. Many steroids have a C-5-6 double bond, which allows the adjacent C-7 to be selectively oxidized.[145] I did not know of this method to add a ketone at C-7 until a reviewer informed me. Any future steroid biosynthesis projects should involve an organic chemist collaborator to identify which reactions, and thus which enzymes, would truly improve the field.

If multiscale pharmacokinetic models are to become standard in drug design, training for biochemists and biologists must involve basic programming. Coursework in drug design should introduce pharmacokinetics software; I was introduced to MATLAB's SimBiology package in such a class. While I ultimately used R instead of MATLAB, I would not have considered a modeling approach at all without this exposure. A mathematics background is not required to use

these techniques, as I built this model with no courses in differential equations. As an added benefit, PhD students with modeling experience are better prepared for jobs in the pharmaceutical industry. Industry groups appreciate pharmacokinetics more than their academic counterparts; as an example, Brandish *et al.* consider steroid permeability in their drug design, while Graversen *et al.* do not.[99,101] Academic drug development would clearly profit from integrating more pharmacokinetics techniques into its process. This requires more accessible tools for multiscale pharmacokinetics models. The development of better-quality software packages could be an easy first step. Most tools either struggle with many chemical species or handle compartment volumes incorrectly. A database with accurate values of common pharmacokinetics constants would be more challenging to build, but also more rewarding. It could help academic groups embarking on new projects, particularly those unfamiliar with the animal experiments needed to measure these parameters. With these tools, any biochemistry student could make a multiscale kinetic model to better understand their drug delivery system of choice.

I have been incredibly fortunate to work in the collaborative environments of the Harvard Systems Biology department, the Laboratory of Systems Pharmacology, and the Chemical Biology PhD program. Their infrastructure allowed me to integrate biology, chemistry, and programming methods to address these problems in steroid drugs. While I have grown very fond of steroids, they are not the only class of molecules that could benefit from biosynthetic and pharmacokinetic approaches. I look forward to seeing how future interdisciplinary students apply this arsenal of techniques to other therapeutics at Harvard and beyond.

References

1. Van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: Where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9**, 112–124 (2008).

2. Maxfield, F. R. & van Meer, G. Cholesterol, the central lipid of mammalian cells. *Curr. Opin. Cell Biol.* **22**, 422–429 (2010).

3. Rewitz, K. F., Rybczynski, R., Warren, J. T. & Gilbert, L. I. Identification, characterization and developmental expression of Halloween genes encoding P450 enzymes mediating ecdysone biosynthesis in the tobacco hornworm, Manduca sexta. *Insect Biochem. Mol. Biol.* **36**, 188–199 (2006).

4. Liu, J. & Hekimi, S. The impact of mitochondrial oxidative stress on bile acid-like molecules in. *Worm* 1–7 (2013). doi:10.1371/journal.pgen.1002553

5. Bentley, P. Sex Hormones in Vertebrates. *Encycl. Life Sci.* 1–5 (2006). doi:10.1038/npg.els.0001846

6. Patrick, G. History of Cortisone and Related Compounds. *eLS* 1–5 (2013). doi:10.1002/9780470015902.a0003627.pub2

7. World Health Organization Medicines. WHO Model List of Essential Medicines. *World Heal. Organ.* 1–43 (2013).

8. Mager, D. E., Moledina, N. & Jusko, W. J. Relative immunosuppressive potency of therapeutic corticosteroids measured by whole blood lymphocyte proliferation. *J. Pharm. Sci.* **92**, 1521–1525 (2003).

9. Overman, R. A., Yeh, J. Y. & Deal, C. L. Prevalence of oral glucocorticoid usage in the United States: A general population perspective. *Arthritis Care Res.* **65**, 294–298 (2013).

10. Fardet, L., Petersen, I. & Nazareth, I. Prevalence of long-term oral glucocorticoid prescriptions in the UK over the past 20 years. *Rheumatology (Oxford).* **50**, 1982–1990 (2011).

11. Lin, K. T. & Wang, L. H. New dimension of glucocorticoids in cancer treatment. *Steroids* **111**, 84–88 (2016).

12. Rice, J. B., White, A. G., Scarpati, L. M., Wan, G. & Nelson, W. W. Long-term Systemic Corticosteroid Exposure: A Systematic Literature Review. *Clin. Ther.* **39**, 2216–2229 (2017).

13. Chabbert-Buffet, N., Meduri, G., Bouchard, P. & Spitz, I. M. Selective progesterone receptor modulators and progesterone antagonists: Mechanisms of action and clinical applications. *Hum. Reprod. Update* **11**, 293–307 (2005).

14. Guennoun, R. *et al.* Progesterone and allopregnanolone in the central nervous system: Response to injury and implication for neuroprotection. *J. Steroid Biochem. Mol. Biol.* **146**, 48–61 (2015).

15. Haberlandt, L. Ueber hormonale Sterilisierung des weiblichen Tierkorpers. Ein Beitrag zur Lehre von der inneren Sekration des Eierstockes und der Placenta. *Fortschritte der Naturwissenschaftlichen Forsch. Berlin* **12**, 1–70 (1924).

16. Madauss, K. P., Stewart, E. L. & Williams, S. P. The evolution of progesterone receptor ligands. *Med. Res. Rev.* **27**, 374–400 (2007).

17. Cadilla, R. & Turnbull, P. Selective Androgen Receptor Modulators in Drug Discovery: Medicinal Chemistry and Therapeutic Potential. *Curr. Top. Med. Chem.* **6**, 245–270 (2006).

18. Pihlajamaa, P., Sahu, B. & Jänne, O. A. Determinants of receptor- and tissue-specific actions in androgen signaling. *Endocr. Rev.* **36**, 357–384 (2015).

19. Haendler, B. & Cleve, A. Recent developments in antiandrogens and selective androgen receptor modulators. *Mol. Cell. Endocrinol.* **352**, 79–91 (2012).

20. Anstead, G. M., Carlson, K. E. & Katzenellenbogen, J. A. The estradiol pharmacophore: Ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site. *Steroids* **62**, 268–303 (1997).

21. Heldring, N. *et al.* Estrogen Receptors: How Do They Signal and What Are Their Targets. *Physiol. Rev.* **87**, 905–931 (2007).

22. Gennari, L., De Paola, V., Merlotti, D., Martini, G. & Nuti, R. Steroid hormone receptor gene polymorphisms and osteoporosis: a pharmacogenomic review. *Expert Opin. Pharmacother.* **8**, 537–553 (2007).

23. Nilsson, S., Koehler, K. F. & Gustafsson, J. Å. Development of subtype-selective oestrogen receptor-based therapeutics. *Nat. Rev. Drug Discov.* **10**, 778–792 (2011).

24. Komm, B. S. & Mirkin, S. An overview of current and emerging SERMs. *J. Steroid Biochem. Mol. Biol.* **143**, 207–222 (2014).

25. Kadmiel, M. & Cidlowski, J. A. Glucocorticoid receptor signaling in health and disease. *Trends Pharmacol. Sci.* **34**, 518–530 (2013).

26. He, Y. *et al.* Structures and mechanism for the design of highly potent glucocorticoids. *Cell Res.* **24**, 713–726 (2014).

27. Huyet, J., Pinon, G. M., Fay, M. R., Rafestin-Oblin, M. E. & Fagart, J. Structural determinants of ligand binding to the mineralocorticoid receptor. *Mol. Cell. Endocrinol.* **350**, 187–195 (2012).

28. McCurley, A. & Jaffe, I. Z. Mineralocorticoid receptors in vascular function and disease. *Mol. Cell. Endocrinol.* **350**, 256–265 (2012).

29. Jaisser, F. & Farman, N. Emerging Roles of the Mineralocorticoid Receptor in Pathology: Toward New Paradigms in Clinical Pharmacology. *Pharmacol. Rev.* **68**, 49–75 (2015).

30. Ma, L. & Nelson, E. R. Oxysterols and nuclear receptors. *Mol. Cell. Endocrinol.* **484**, 42–51 (2019).

31. Griffiths, W. J. & Wang, Y. Analysis of oxysterol metabolomes. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1811**, 784–799 (2011).

32. Gabbi, C., Warner, M. & Gustafsson, J. Å. Action mechanisms of Liver X Receptors. *Biochem. Biophys. Res. Commun.* **446**, 647–650 (2014).

33. Fessler, M. B. The challenges and promise of targeting the Liver X Receptors for treatment of inflammatory disease. *Pharmacol. Ther.* **181**, 1–12 (2018).

34. Hofmann, A. F. & Hagey, L. R. Bile acids: Chemistry, pathochemistry, biology, pathobiology, and therapeutics. *Cell. Mol. Life Sci.* **65**, 2461–2483 (2008).

35. Donkers, J. M., Roscam Abbing, R. L. P. & van de Graaf, S. F. J. Developments in bile salt based therapies: A critical overview. *Biochem. Pharmacol.* **161**, 1–13 (2019).

36. Massafra, V., Pellicciari, R., Gioiello, A. & van Mil, S. W. C. Progress and challenges of selective Farnesoid X Receptor modulation. *Pharmacol. Ther.* **191**, 162–177 (2018).

37. Laudet, V. Nuclear Receptor Genes. *Encycl. Life Sci.* 1–6 (2006). doi:10.1038/npg.els.0006154

38. Busillo, J. M., Rhen, T. & Cidlowski, J. A. Steroid Hormone Action. in *Yen & Jaffe's Reproductive Endocrinology: Physiology, Pathophysiology, and Clinical Management: Eighth Edition* (eds. Whirledge, S. & Cidlowski, J. A.) 115-131.e4 (Elsevier, 2018). doi:10.1016/B978-0-323-47912-7.00005-6

39. Weikum, E. R., Liu, X. & Ortlund, E. A. The nuclear receptor superfamily: A structural perspective. *Protein Sci.* **27**, 1876–1892 (2018).

40. Millard, C. J., Watson, P. J., Fairall, L. & Schwabe, J. W. R. An evolving understanding of nuclear receptor coregulator proteins. *J. Mol. Endocrinol.* **51**, T23–T36 (2013).

41. Schwartz, N., Verma, A., Bivens, C. B., Schwartz, Z. & Boyan, B. D. Rapid steroid hormone actions via membrane receptors. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863**, 2289–2298 (2016).

42. Wilkenfeld, S. R., Lin, C. & Frigo, D. E. Communication between genomic and non-genomic signaling events coordinate steroid hormone actions. *Steroids* **133**, 2–7 (2018).

43. Levin, E. R. & Hammes, S. R. Nuclear receptors outside the nucleus: extranuclear signalling by steroid receptors. *Nat. Rev. Mol. Cell Biol.* **17**, 783–797 (2016).

44. Pathways, R. Cancer revoked: oncogenes as therapeutic targets. *Nat. Rev. Cancer* **3**, 375–380 (2003).

45. Billette, C. *et al.* Genome sequence of the button mushroom Agaricus bisporus reveals mechanisms governing adaptation to a humic-rich ecological niche. *Proc. Natl. Acad. Sci.* **109**, 17501–17506 (2012).

46. Treviño, L. S. & Weigel, N. L. Phosphorylation: A fundamental regulator of steroid receptor action. *Trends Endocrinol. Metab.* **24**, 515–524 (2013).

47. Galliher-Beckley, A. J., Williams, J. G. & Cidlowski, J. A. Ligand-Independent

Phosphorylation of the Glucocorticoid Receptor Integrates Cellular Stress Pathways with Nuclear Receptor Signaling. *Mol. Cell. Biol.* **31**, 4663–4675 (2011).

48. McKay, L. I. & Cidlowski, J. A. Molecular control of immune/inflammatory responses: Interactions between nuclear factor-κB and steroid receptor-signaling pathways. *Endocr. Rev.* **20**, 435–459 (1999).

49. Lathe, R. & Kotelevtsev, Y. Steroid signaling: Ligand-binding promiscuity, molecular symmetry, and the need for gating. *Steroids* **82**, 14–22 (2014).

50. Odermatt, A. & Kratschmar, D. V. Tissue-specific modulation of mineralocorticoid receptor function by 11β-hydroxysteroid dehydrogenases: An overview. *Mol. Cell. Endocrinol.* **350**, 168–186 (2012).

51. Rudmann, D. G. On-target and off-target-based toxicologic effects. *Toxicol. Pathol.* **41**, 310–314 (2013).

52. Judd, L. L. *et al.* Adverse consequences of glucocorticoid medication: Psychological, cognitive, and behavioral effects. *Am. J. Psychiatry* **171**, 1045–1051 (2014).

53. Ponnusamy, S. *et al.* Androgen receptor agonists increase lean mass, improve cardiopulmonary functions and extend survival in preclinical models of Duchenne muscular dystrophy. *Hum. Mol. Genet.* **26**, 2526–2540 (2017).

54. Africander, D., Verhoog, N. & Hapgood, J. P. Molecular mechanisms of steroid receptor-mediated actions by synthetic progestins used in HRT and contraception. *Steroids* **76**, 636–652 (2011).

55. Dobson, C. M. Chemical space and biology. *Nature* **432**, 824–828 (2004).

56. Haggarty, S. J. The principle of complementarity: Chemical versus biological space. *Curr. Opin. Chem. Biol.* **9**, 296–303 (2005).

57. Huang, P., Chandra, V. & Rastinejad, F. Structural Overview of the Nuclear Receptor Superfamily: Insights into Physiology and Therapeutics. *Annu. Rev. Physiol.* **72**, 247–272 (2010).

58. Burris, T. P. *et al.* Nuclear Receptors and Their Selective Pharmacologic Modulators. *Pharmacol. Rev.* **65**, 710–778 (2013).

59. Riggs, B. L. & Hartmann, L. C. Selective estrogen receptor modulators (SERMs) - Mechanisms of Action and Application to Clinical Practice. *N. Engl. J. Med.* **348**, 618–29 (2003).

60. Shang, Y. & Brown, M. Molecular determinants for the tissue specificity of SERMs. *Science.* **295**, 2465–2468 (2002).

61. Fawell, S. E. *et al.* Inhibition of estrogen receptor-DNA binding by the 'pure' antiestrogen ICI 164,384 appears to be mediated by impaired receptor dimerization. *Proc. Natl. Acad. Sci.* **87**, 6883–6887 (1990).

62. Long, X., Fan, M. & Nephew, K. P. Estrogen receptor-α-interacting cytokeratins potentiate the antiestrogenic activity of fulvestrant. *Cancer Biol. Ther.* **9**, 389–396 (2010).

63. Solomon, Z. J. *et al.* Selective Androgen Receptor Modulators: Current Knowledge and Clinical Applications. *Sex. Med. Rev.* **7**, 84–94 (2019).

64. Narayanan, R., Coss, C. C. & Dalton, J. T. Development of selective androgen receptor modulators (SARMs). *Mol. Cell. Endocrinol.* **465**, 134–142 (2018).

65. Meijer, O. C., Koorneef, L. L. & Kroon, J. Glucocorticoid receptor modulators. *Ann. Endocrinol. (Paris).* **79**, 107–111 (2018).

66. Clark, A. R. & Belvisi, M. G. Maps and legends: The quest for dissociated ligands of the glucocorticoid receptor. *Pharmacol. Ther.* **134**, 54–67 (2012).

67. Sundahl, N., Bridelance, J., Libert, C., De Bosscher, K. & Beck, I. M. Selective glucocorticoid receptor modulation: New directions with non-steroidal scaffolds. *Pharmacol. Ther.* **152**, 28–41 (2015).

68. Pellissier, H. & Santelli, M. Chemical and biochemical hydroxylations of steroids. A review. *Org. Prep. Proced. Int.* **33**, 1–58 (2001).

69. Newhouse, T. & Baran, P. S. If C-H bonds could talk: Selective C-H bond oxidation. *Angew. Chemie - Int. Ed.* **50**, 3362–3374 (2011).

70. Salvador, J., Silvestre, S. & Moreira, V. Catalytic Oxidative Processes in Steroid Chemistry: Allylic Oxidation, beta-Selective Epoxidation, Alcohol Oxidation and Remote Functionalization Reactions. *Curr. Org. Chem.* **10**, 2227–2257 (2006).

71. Mathew, P. A., Mason, J. I., Trant, J. M. & Waterman, M. R. Incorporation of steroidogenic pathways which produce cortisol and aldosterone from cholesterol into nonsteroidogenic cells. *Mol. Cell. Endocrinol.* **73**, 73–80 (1990).

72. Szczebara, F. M. *et al.* Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nat. Biotechnol.* **21**, 143–149 (2003).

73. Duport, C., Spagnoli, R., Degryse, E. & Pompon, D. Self-sufficient biosynthesis of pregnenolone and progesterone in engineered yeast. *Nat. Biotechnol.* **16**, 186–189 (1998).

74. Dhir, V. *et al.* Steroid 17α-hydroxylase deficiency: Functional characterization of four mutations (A174E, V178D, R440C, L465P) in the CYP17A1 gene. *J. Clin. Endocrinol. Metab.* **94**, 3058–3064 (2009).

75. Bleicken, C. *et al.* Functional characterization of three CYP21A2 sequence variants (p.A265V, p.W302S, p.D322G) employing a yeast co-expression system. *Hum. Mutat.* **30**, 443–450 (2009).

76. Hakki, T., Zearo, S., Drăgan, C. A., Bureik, M. & Bernhardt, R. Coexpression of redox partners increases the hydrocortisone (cortisol) production efficiency in CYP11B1 expressing fission yeast Schizosaccharomyces pombe. *J. Biotechnol.* **133**, 351–359 (2008).

77. Petrič., Š., Hakki, T., Bernhardt, R., Žigon, D. & Črešnar, B. Discovery of a steroid 11α-hydroxylase from Rhizopus oryzae and its biotechnological application. *J. Biotechnol.* **150**, 428–437 (2010).

78. Chen, X. *et al.* Molecular cloning, expression of CPR gene from Rhizopus oryzae into Rhizopus nigericans and its application in the 11α-hydroxylation of 16α, 17-epoxy-progesterone. *Enzyme Microb. Technol.* **66**, 28–34 (2014).

79. Manuel Cruz Silva, M., F. Carvalho, J., Riva, S. & Luisa Sa e Melo, M. Biocatalytic Transformations of Steroids: Focus on Hydrolase-Catalyzed Reactions. *Curr. Org. Chem.* **15**, 928–941 (2011).

80. Kristan, K. & Rižner, T. L. Steroid-transforming enzymes in fungi. *J. Steroid Biochem. Mol. Biol.* **129**, 79–91 (2012).

81. Wang, Y., Sun, D., Chen, Z., Ruan, H. & Ge, W. Biotransformation of 3β-hydroxy-5-en-steroids by Mucor silvaticus. *Biocatal. Biotransformation* **31**, 168–174 (2013).

82. Andrushina, V. A., Druzhinina, A. V., Yaderets, V. V., Stitsenko, T. S. & Voishvillo, N. E. Hydroxylation of steroids by Curvularia lunata mycelium in the presence of methyl-β-cyclodextrine. *Appl. Biochem. Microbiol.* **47**, 42–48 (2011).

83. Wiedersberg, S., Leopold, C. S. & Guy, R. H. Bioavailability and bioequivalence of topical glucocorticoids. *Eur. J. Pharm. Biopharm.* **68**, 453–466 (2008).

84. Derendorf, H., Nave, R., Drollmann, A., Cerasoli, F. & Wurst, W. Relevance of pharmacokinetics and pharmacodynamics of inhaled corticosteroids to asthma. *Eur. Respir. J.* **28**, 1042–1050 (2006).

85. Derendorf, H. & Meltzer, E. O. Molecular and clinical pharmacology of intranasal corticosteroids: Clinical and therapeutic implications. *Allergy Eur. J. Allergy Clin. Immunol.* **63**, 1292–1300 (2008).

86. Daley-Yates, P. T. Inhaled corticosteroids: Potency, dose equivalence and therapeutic index. *Br. J. Clin. Pharmacol.* **80**, 372–380 (2015).

87. Lühder, F. & Reichardt, H. M. Novel drug delivery systems tailored for improved administration of glucocorticoids. *Int. J. Mol. Sci.* **18**, (2017).

88. Metselaar, J. M., Mastrobattista, E. & Storm, G. Liposomes for intravenous drug targeting: design and applications. *Mini Rev. Med. Chem.* **2**, 319–29 (2002).

89. Hosseini, S. H., Maleki, A., Eshraghi, H. R. & Hamidi, M. Preparation and in vitro/pharmacokinetic/pharmacodynamic evaluation of a slow-release nano-liposomal form of prednisolone. *Drug Deliv.* **23**, 3008–3016 (2016).

90. Van Den Hoven, J. M. *et al.* Complement activation by PEGylated liposomes containing prednisolone. *Eur. J. Pharm. Sci.* **49**, 265–271 (2013).

91. Tentillier, N. *et al.* Anti-Inflammatory Modulation of Microglia via CD163-Targeted Glucocorticoids Protects Dopaminergic Neurons in the 6-OHDA Parkinson's Disease Model. *J. Neurosci.* **36**, 9375–9390 (2016).

92. Liu, X. M. *et al.* Synthesis and evaluation of a well-defined HPMA copolymer-dexamethasone conjugate for effective treatment of rheumatoid arthritis. *Pharm. Res.* **25**, 2910–2919 (2008).

93. Ren, K. *et al.* Macromolecular prodrug of dexamethasone prevents particle-induced peri-implant osteolysis with reduced systemic side effects. *J. Control. Release* **175**, 1–9 (2014).

94. Everts, M. *et al.* Selective Intracellular Delivery of Dexamethasone into Activated Endothelial Cells Using an E-Selectin-Directed Immunoconjugate. *J. Immunol.* **168**, 883–889 (2002).

95. Carter, P. J. & Lazar, G. A. Next generation antibody drugs: Pursuit of the 'high-hanging fruit'. *Nat. Rev. Drug Discov.* **17**, 197–223 (2018).

96. Perez, H. L. *et al.* Antibody-drug conjugates: current status and future directions. *Drug Discov. Today* **19**, 869–81 (2014).

97. Beck, A., Goetsch, L., Dumontet, C. & Corvaïa, N. Strategies and challenges for the next generation of antibody-drug conjugates. *Nat. Rev. Drug Discov.* **16**, 315–337 (2017).

98. Liu, R., Wang, R. E. & Wang, F. Antibody-drug conjugates for non-oncological indications. *Expert Opin. Biol. Ther.* **16**, 591–593 (2016).

99. Graversen, J. H. *et al.* Targeting the hemoglobin scavenger receptor CD163 in macrophages highly increases the anti-inflammatory potency of dexamethasone. *Mol. Ther.* **20**, 1550–1558 (2012).

100. Svendsen, P. *et al.* Antibody-Directed Glucocorticoid Targeting to CD163 in M2-type Macrophages Attenuates Fructose-Induced Liver Inflammatory Changes. *Mol. Ther. - Methods Clin. Dev.* **4**, 50–61 (2017).

101. Brandish, P. E. *et al.* Development of Anti-CD74 Antibody-Drug Conjugates to Target Glucocorticoids to Immune Cells. *Bioconjug. Chem.* **29**, 2357–2369 (2018).

102. Tibbitts, J., Canter, D., Graff, R., Smith, A. & Khawli, L. A. Key factors influencing ADME properties of therapeutic proteins: A need for ADME characterization in drug discovery and development. *MAbs* **8**, 229–245 (2016).

103. Maack, T. Renal Handling of Proteins and Polypeptides. in *Comprehensive Physiology* 2039–2082 (American Cancer Society, 2010). doi:10.1002/cphy.cp080244

104. Khera, E. & Thurber, G. M. Pharmacokinetic and Immunological Considerations for Expanding the Therapeutic Window of Next-Generation Antibody–Drug Conjugates. *BioDrugs* **32**, 465–480 (2018).

105. Sadekar, S., Figueroa, I. & Tabrizi, M. Antibody Drug Conjugates: Application of Quantitative Pharmacology in Modality Design and Target Selection. *AAPS J.* **17**, 828–836 (2015).

106. Kamath, A. V. Pharmacokinetics of Antibody – Drug Conjugates Pharmacokinetic Characteristics of an ADC. in *Antibody-drug conjugates: fundamentals, drug development, and clinical outcomes to target cancer* (eds. Olivier, K. J. & Hurvitz, S. A.) 245–266 (Wiley, 2016).

107. Singh, A. P., Shin, Y. G. & Shah, D. K. Application of Pharmacokinetic-Pharmacodynamic Modeling and Simulation for Antibody-Drug Conjugate Development.

Pharm. Res. **32**, 3508–3525 (2015).

108. Schmidt, B. J. *et al.* Nonclinical Pharmacology and Mechanistic Modeling of Antibody-Drug Conjugates in Support of Human Clinical Trials. *Antibody-Drug Conjug.* 207–243 (2016). doi:10.1002/9781119060727.ch9

109. *Principles of clinical pharmacology*. (Elsevier Academic Press, 2012).

110. Coyne, C. P. & Narayanan, L. Dexamethasone-(C21-phosphoramide)-[anti-EGFR]: Molecular design, synthetic organic chemistry reactions, and antineoplastic cytotoxic potency against pulmonary adenocarcinoma (A549). *Drug Des. Devel. Ther.* **10**, 2575–2597 (2016).

111. Wang, R. E. *et al.* An immunosuppressive antibody-drug conjugate. *J. Am. Chem. Soc.* **137**, 3229–3232 (2015).

112. Yu, S. *et al.* Targeted delivery of an anti-inflammatory PDE4 inhibitor to immune cells via an antibody-drug conjugate. *Mol. Ther.* **24**, 2078–2089 (2016).

113. Spady, E. S. *et al.* Mammalian Cells Engineered To Produce New Steroids. *ChemBioChem* **19**, 1827–1833 (2018).

114. Sarnes, E. *et al.* Incidence and US Costs of Corticosteroid-Associated Adverse Events: A Systematic Literature Review. *Clin. Ther.* **33**, 1413–1432 (2011).

115. Keski-Rahkonen, P., Huhtinen, K., Poutanen, M. & Auriola, S. Fast and sensitive liquid chromatography-mass spectrometry assay for seven androgenic and progestagenic steroids in human serum. *J. Steroid Biochem. Mol. Biol.* **127**, 396–404 (2011).

116. Fernandes, P., Cruz, A., Angelova, B., Pinheiro, H. M. & Cabral, J. M. S. Microbial conversion of steroid compounds: Recent developments. *Enzyme Microb. Technol.* **32**, 688–705 (2003).

117. Swizdor, A., Kolek, T., Panek, A. & Milecka, N. Selective Modifications of Steroids Performed by Oxidative Enzymes. *Curr. Org. Chem.* **16**, 2551–2582 (2012).

118. Nes, W. D. Biosynthesis of Cholesterol and Other Sterols. *Chem. Rev.* **111**, 6423–6451 (2011).

119. Parikh, A., Gillam, E. M. J. & Guengerich, F. P. Drug metabolism by Escherichia coli expressing human cytochromes P450. *Nat. Biotechnol.* **15**, 784–788 (1997).

120. Gerber, A., Milhim, M., Hartz, P., Zapp, J. & Bernhardt, R. Genetic engineering of Bacillus megaterium for high-yield production of the major teleost progestogens 17α,20β-di- and 17α,20β,21α-trihydroxy-4-pregnen-3-one. *Metab. Eng.* **36**, 19–27 (2016).

121. Wriessnegger, T. & Pichler, H. Yeast metabolic engineering - Targeting sterol metabolism and terpenoid formation. *Prog. Lipid Res.* **52**, 277–293 (2013).

122. Rose, K. A. *et al.* Cyp7b, a novel brain cytochrome P450, catalyzes the synthesis of neurosteroids 7alpha-hydroxy dehydroepiandrosterone and 7alpha-hydroxy pregnenolone. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4925–4930 (1997).

123. Beneventi, E. *et al.* Enzymatic Baeyer-Villiger oxidation of steroids with cyclopentadecanone monooxygenase. *J. Mol. Catal. B Enzym.* **58**, 164–168 (2009).

124. Iwaki, H. *et al.* Pseudomonad cyclopentadecanone monooxygenase displaying an uncommon spectrum of Baeyer-Villiger oxidations of cyclic ketones. *Appl. Environ. Microbiol.* **72**, 2707–2720 (2006).

125. Slominski, A. T. *et al.* Novel activities of CYP11A1 and their potential physiological significance. *J. Steroid Biochem. Mol. Biol.* **151**, 25–37 (2015).

126. Knol, J., Bodewits, K., Hessels, G. I., Dijkhuizen, L. & van der Geize, R. 3-Keto-5α-steroid Δ 1 -dehydrogenase from Rhodococcus erythropolis SQ1 and its orthologue in Mycobacterium tuberculosis H37Rv are highly specific enzymes that function in cholesterol catabolism . *Biochem. J.* **410**, 339–346 (2008).

127. Lundell, K., Hansson, R. & Wikvall, K. Cloning and Expression of a Pig Liver Taurochenodeoxycholic Acid 6α-Hydroxylase (CYP4A21). *J. Biol. Chem.* **276**, 9606–9612 (2002).

128. Russell, D. W. The Enzymes, Regulation, and Genetics of Bile Acid Synthesis. *Annu. Rev. Biochem.* **72**, 137–174 (2003).

129. Miller, W. L. Steroid hormone synthesis in mitochondria. *Mol. Cell. Endocrinol.* **379**, 62–73 (2013).

130. Makino, T., Katsuyama, Y., Otomatsu, T., Misawa, N. & Ohnishi, Y. Regio- and stereospecific hydroxylation of various steroids at the 16α position of the D ring by the Streptomyces griseus cytochrome P450 CYP154C3. *Appl. Environ. Microbiol.* **80**, 1371–1379 (2014).

131. Auchus, R. J. & Adams, C. M. Cholesterol, Steroid and Isoprenoid Biosynthesis. in *Encyclopedia of Life Sciences* (John Wiley & Sons, Ltd, 2006). doi:10.1038/npg.els.0001393

132. Thomas, M. P. & Potter, B. V. L. The structural biology of oestrogen metabolism. *J. Steroid Biochem. Mol. Biol.* **137**, 27–49 (2013).

133. Lee, G. Y., Kim, D. H., Kim, D., Ahn, T. & Yun, C. H. Functional characterization of steroid hydroxylase CYP106A1 derived from Bacillus megaterium. *Arch. Pharm. Res.* **38**, 98–107 (2015).

134. Schmitz, D., Zapp, J. & Bernhardt, R. Steroid conversion with CYP106A2 - production of pharmaceutically interesting DHEA metabolites. *Microb. Cell Fact.* **13**, 81 (2014).

135. Hannich, J. T. *et al.* Methylation of the Sterol Nucleus by STRM-1 Regulates Dauer Larva Formation in Caenorhabditis elegans. *Dev. Cell* **16**, 833–843 (2009).

136. Kabasakalian, P., Britt, E. & Yudis, M. D. Solubility or some steroids in water. *J. Pharm. Sci.* **55**, 642–642 (1966).

137. Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl- tert -butyl ether for high-throughput lipidomics . *J. Lipid Res.* **49**,

1137–1146 (2008).

138. Surowiec, I., Koc, M., Antti, H., Wikström, P. & Moritz, T. LC-MS/MS profiling for detection of endogenous steroids and prostaglandins in tissue samples. *J. Sep. Sci.* **34**, 2650–2658 (2011).

139. Stiles, A. R., McDonald, J. G., Bauman, D. R. & Russell, D. W. CYP7B1: One cytochrome P450, two human genetic diseases, and multiple physiological functions. *J. Biol. Chem.* **284**, 28485–28489 (2009).

140. Yantsevich, A. V. *et al.* Human steroid and oxysterol 7α-hydroxylase CYP7B1: Substrate specificity, azole binding and misfolding of clinically relevant mutants. *FEBS J.* **281**, 1700–1713 (2014).

141. Schmitz, D., Zapp, J. & Bernhardt, R. Steroid conversion with CYP106A2 - production of pharmaceutically interesting DHEA metabolites. *Microb. Cell Fact.* **13**, 81 (2014).

142. Zaretskii, V. I., Wulfson, N. S., Kogan, L. M., Voishvillo, N. E. & Torgov, I. V. Mass spectrometry of steroid systems—V. *Tetrahedron* **22**, 1399–1405 (1966).

143. Schaaf, O. & Dettner, K. Transformation of steroids by Bacillus strains isolated from the foregut of water beetles (Coleoptera: Dytiscidae): II. Metabolism of 3b-hydroxypregn-5-en-20-one (pregnenolone). *J. Steroid Biochem. Mol. Biol.* **75**, 187–199 (2000).

144. Schenck, G. O., Neumuller, O. A. & Eisfeld, W. Zur photosensibilisierten Autoxydation der Steroide: Delta5-Steroid-7a-Hydroperoxyde und -7-ketone durch Allylumlagerung von Delta6-Steroid-5a-Hydroperoxyden. *European J. Org. Chem.* **579**, 202–210 (1958).

145. Horn, E. J. *et al.* Scalable and sustainable electrochemical allylic C-H oxidation. *Nature* **533**, 77–81 (2016).

146. Shan, L. H. *et al.* Synthesis of 3β, 7α, 11α-trihydroxy-pregn-21-benzylidene-5-en-20-one derivatives and their cytotoxic activities. *Bioorganic Med. Chem. Lett.* **19**, 6637–6639 (2009).

147. Choudhary, M. I., Batool, I., Shah, S. A. A., Nawaz, S. A. & Atta-ur-Rahman. Microbial hydroxylation of pregnenolone derivatives. *Chem. Pharm. Bull. (Tokyo).* **53**, 1455–9 (2005).

148. Rose, K. *et al.* Neurosteroid hydroxylase CYP7B: Vivid reporter activity in dentate gyrus of gene-targeted mice and abolition of a widespread pathway of steroid and oxysterol hydroxylation. *J. Biol. Chem.* **276**, 23937–23944 (2001).

149. Stapleton, G. *et al.* A novel cytochrome P450 expressed primarily in brain. *J. Biol. Chem.* **270**, 29739–29745 (1995).

150. Yau, J. L. W., Noble, J., Graham, M. & Seckl, J. R. Central administration of a cytochrome P450-7B product 7 alpha-hydroxypregnenolone improves spatial memory retention in cognitively impaired aged rats. *J. Neurosci.* **26**, 11034–11040 (2006).

151. Chung, B. C. *et al.* Cytochrome P450c17 (steroid 17 alpha-hydroxylase/17,20 lyase): cloning of human adrenal and testis cDNAs indicates the same gene is expressed in both

tissues. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 407–11 (1987).

152. Petrunak, E. M., DeVore, N. M., Porubsky, P. R. & Scott, E. E. Structures of human steroidogenic cytochrome P450 17A1 with substrates. *J. Biol. Chem.* **289**, 32952–32964 (2014).

153. Gonzalez, E. & Guengerich, F. P. Kinetic processivity of the two-step oxidations of progesterone and pregnenolone to androgens by human cytochrome P450 17A1. *J. Biol. Chem.* **292**, 13168–13185 (2017).

154. Kominami, S., Nishida, N. & Takemori, S. Reconstitution of the steroidogenic pathway from cholesterol to aldosterone in liposome membranes. *Biochim. Biophys. Acta - Lipids Lipid Metab.* **1301**, 199–206 (1996).

155. Kau, T. R. *et al.* A chemical genetic screen identifies inhibitors of regulated nuclear export of a Forkhead transcription factor in PTEN-deficient tumor cells. *Cancer Cell* **4**, 463–476 (2003).

156. Guengerich, F. P. Human cytochrome P450 enzymes. in *Cytochrome P450: Structure, Mechanism, and Biochemistry, Fourth Edition* (ed. Ortiz De Montellano, P. R.) 523–785 (2015). doi:10.1007/978-3-319-12108-6_9

157. Gillam, E. & Hayes, M. The Evolution of Cytochrome P450 Enzymes as Biocatalysts in Drug Discovery and Development. *Curr. Top. Med. Chem.* **13**, 2254–2280 (2013).

158. Auchus, R. J. & Miller, W. L. P450 enzymes in steroid processing. in *Cytochrome P450: Structure, Mechanism, and Biochemistry, Fourth Edition* (ed. Ortiz De Montellano, P. R.) 851–879 (Springer International Publishing, 2015). doi:10.1007/978-3-319-12108-6_12

159. Yoshimoto, F. K., Gonzalez, E., Auchus, R. J. & Guengerich, F. P. Mechanism of 17α,20-Lyase and New Hydroxylation Reactions of Human Cytochrome P450 17A1. *J. Biol. Chem.* **291**, 17143–17164 (2016).

160. Denisov, I. G., Makris, T. M., Sligar, S. G. & Schlichting, I. Structure and Chemistry of Cytochrome P450. *Chem. Rev.* **105**, 2253–2278 (2005).

161. Denisov, I. G. & Sligar, S. G. Activation of molecular oxygen in cytochromes P450. in *Cytochrome P450: Structure, Mechanism, and Biochemistry, Fourth Edition* 69–109 (Springer International Publishing, 2015). doi:10.1007/978-3-319-12108-6_3

162. Ortiz De Montellano, P. R. Substrate Oxidation by Cytochrome P450 Enzymes. in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (eds. Ortiz De Montellano, P. R. & Voss, J. J. De) 183–245 (2005). doi:10.1007/b139087

163. Korzekwa, K. Enzyme Kinetics of Oxidative Metabolism: Cytochromes P450. in *Enzyme Kinetics in Drug Metabolism: Fundamentals and Applications* (eds. Nagar, S., Argikar, U. A. & Tweedie, D. J.) 149–166 (Springer International Publishing, 2014).

164. Combs, S. A. *et al.* Small-molecule ligand docking into comparative models with Rosetta. *Nat. Protoc.* **8**, 1277–1298 (2013).

165. Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**,

1735–1742 (2013).

166. Tempel, W. *et al.* Structural characterization of human cholesterol 7α-hydroxylase. *J. Lipid Res.* **55**, 1925–1932 (2014).

167. Poulos, T. L. & Johnson, E. F. Structures of Cytochrome P450 Enzymes. in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (ed. Ortiz De Montellano, P. R.) 3–32 (Springer International Publishing, 2015).

168. Tempel, W. *et al.* Structural characterization of human cholesterol 7α-hydroxylase. *J. Lipid Res.* **55**, 1925–1932 (2014).

169. Rydberg, P., Gloriam, D. E., Zaretzki, J., Breneman, C. & Olsen, L. SMARTCyp: A 2D method for prediction of cytochrome P450-mediated drug metabolism. *ACS Med. Chem. Lett.* **1**, 96–100 (2010).

170. Olsen, L., Oostenbrink, C. & Jørgensen, F. S. Prediction of cytochrome P450 mediated metabolism. *Adv. Drug Deliv. Rev.* **86**, 61–71 (2015).

171. Oostenbrink, C., de Ruiter, A., Hritz, J. & Vermeulen, N. Malleability and Versatility of Cytochrome P450 Active Sites Studied by Molecular Simulations. *Curr. Drug Metab.* **13**, 190–196 (2012).

172. Heinisch, T. *et al.* Improving the catalytic performance of an artificial metalloenzyme by computational design. *J. Am. Chem. Soc.* **137**, 10414–10419 (2015).

173. O'Boyle, N. M. *et al.* Open Babel: An Open chemical toolbox. *J. Cheminform.* **3**, 1–14 (2011).

174. Bunn, H. F. Erythropoiethin. *Cold Spring Harb. Perspect. Med.* **3**, 1–20 (2013).

175. Burrill, D. R., Vernet, A., Collins, J. J., Silver, P. A. & Way, J. C. Targeted erythropoietin selectively stimulates red blood cell expansion in vivo. *Proc. Natl. Acad. Sci.* **113**, 5245–5250 (2016).

176. Garcin, G. *et al.* High efficiency cell-specific targeting of cytokine activity. *Nat. Commun.* **5**, 3016 (2014).

177. Hedrich, W. D., Fandy, T. E., Ashour, H. M., Wang, H. & Hassan, H. E. Antibody–Drug Conjugates: Pharmacokinetic/Pharmacodynamic Modeling, Preclinical Characterization, Clinical Studies, and Lessons Learned. *Clin. Pharmacokinet.* **57**, 687–703 (2018).

178. Sorger, P. K. *et al. Quantitative and Systems Pharmacology in the Post-genomic Era: New Approaches to Discovering Drugs and Understanding Therapeutic Mechanisms*. (2011). doi:10.1049/cp:20050775

179. Saunders, A. E. & Johnson, P. Modulation of immune cell signalling by the leukocyte common tyrosine phosphatase, CD45. *Cell. Signal.* **22**, 339–348 (2010).

180. Lin, Y. *et al.* A genetically engineered anti-CD45 single-chain antibody-streptavidin fusion protein for pretargeted radioimmunotherapy of hematologic malignancies. *Cancer Res.* **66**, 3884–3892 (2006).

181. Keizer, R. J., Huitema, A. D. R., Schellens, J. H. M. & Beijnen, J. H. Clinical Pharmacokinetics of Therapeutic Monoclonal Antibodies. *Clin. Pharmacokinet.* **49**, 493–507 (2010).

182. Högger, P. & Rohdewald, P. Binding kinetics of fluticasone propionate to the human glucocorticoid receptor. *Steroids* **59**, 597–602 (1994).

183. Allen, A., Bareille, P. J. & Rousell, V. M. Fluticasone furoate, a novel inhaled corticosteroid, demonstrates prolonged lung absorption kinetics in man compared with inhaled fluticasone propionate. *Clin. Pharmacokinet.* **52**, 37–42 (2013).

184. Kaschek, D., Mader, W., Fehling-Kaschek, M., Rosenblatt, M. & Timmer, J. Dynamic Modeling, Parameter Estimation and Uncertainty Analysis in R. *bioRxiv* 085001 (2016). doi:10.1101/085001

185. Sarkar, C. A. *et al.* Rational cytokine design for increased lifetime and enhanced potency using pH-activated "histidine switching". *Nat. Biotechnol.* **20**, 908–913 (2002).

186. Westphal, U. Steroid Protein Interactions II. *Monogr Endocrinol* **27**, 1–603 (1986).

187. Eliard, P. H. & Rousseau, G. G. Thermodynamics of steroid binding to the human glucocorticoid receptor. *Biochem. J.* **218**, 395–404 (1984).

188. Faassen, F., Kelder, J., Lenders, J., Onderwater, R. & Vromans, H. Physicochemical properties and transport of steroids across Caco-2 cells. *Pharm. Res.* **20**, 177–186 (2003).

189. Milo, R., Jorgensen, P., Moran, U., Weber, G. & Springer, M. BioNumbers: The database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38**, 750–753 (2009).

190. Harris, L. A. *et al.* BioNetGen 2.2: Advances in rule-based modeling. *Bioinformatics* **32**, 3366–3368 (2016).

191. Davies, B. & Morris, T. Physiological parameters in laboratory animals and humans. *Pharmaceutical research* **10**, 1093–5 (1993).

192. Ting-Beall, H. P., Needham, D. & Hochmuth, R. M. Volume and osmotic properties of human neutrophils. *Blood* **81**, 2774–80 (1993).

193. Segel, G. B., Cokelet, G. R. & Lichtman, M. A. The measurement of lymphocyte volume: importance of reference particle deformability and counting solution tonicity. *Blood* **57**, 894–899 (1981).

194. Schmid-Schonbein, G. W., Shih, Y. Y. & Chien, S. Morphometry of human leukocytes. *Blood* **56**, 866–875 (1980).

195. Corlier, F. *et al.* Modifications of the endosomal compartment in peripheral blood mononuclear cells and fibroblasts from Alzheimer's disease patients. *Transl. Psychiatry* **5**, (2015).

196. Sarkar, C. A. Cell-Level Pharmacokinetic Model of Granulocyte Colony-Stimulating Factor: Implications for Ligand Lifetime and Potency in Vivo. *Mol. Pharmacol.* **63**, 147–158 (2003).

197. Yogurtcu, O. N. & Johnson, M. E. Cytosolic proteins can exploit membrane localization to trigger functional assembly. *PLoS Comput. Biol.* **14**, 1–28 (2018).

198. Peyron, J. F. *et al.* The CD45 protein tyrosine phosphatase is required for the completion of the activation program leading to lymphokine production in the Jurkat human T cell line. *Int. Immunol.* **3**, 1357–66 (1991).

199. Bausch-Fluck, D. *et al.* A mass spectrometric-derived cell surface protein atlas. *PLoS One* **10**, 1–22 (2015).

200. Matthews, D. C. *et al.* Radiolabeled anti-CD45 monoclonal antibodies target lymphohematopoietic tissue in the macaque. *Blood* **78**, 1864–74 (1991).

201. Biggin, M. D. Animal Transcription Networks as Highly Connected, Quantitative Continua. *Dev. Cell* **21**, 611–626 (2011).

202. Raman, C. S., Jemmerson, R., Nall, B. T. & Allen, M. J. Diffusion-limited rates for monoclonal antibody binding to cytochrome c. *Biochemistry* **31**, 10370–10379 (1992).

203. Sigismund, S. *et al.* Clathrin-Mediated Internalization Is Essential for Sustained EGFR Signaling but Dispensable for Degradation. *Dev. Cell* **15**, 209–219 (2008).

204. Alberts, B. *et al.* Transport into the cell from the plasma membrane: endocytosis. in *Molecular Biology of the Cell* 695–752 (2015).

205. Huotari, J. & Helenius, A. Endosome maturation. *EMBO J.* **30**, 3481–3500 (2011).

206. Minami, Y., Stafford, F. J., Lippincott-Schwartz, J., Yuan, L. C. & Klausner, R. D. Novel redistribution of an intracellular pool of CD45 accompanies T cell activation. *J. Biol. Chem.* **266**, 9222–9230 (1991).

207. U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on November 2017 submission data (1999-2015). *U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute* (2018).

208. *Report of the Director - National Institutes of Health: Fiscal Years 2014-2015*. (2016).

209. Athens, J. W. *et al.* Leukokinetic studies. IV. The total blood, circulating and marginal granulocyte pools and the granulocyte turnover rate in normal subjects. *J. Clin. Invest.* **40**, 989–995 (1961).

210. Trepel, F. Number and distribution of lymphocytes in man. A critical analysis. *Klin. Wochenschr.* **52**, 511–515 (1974).

211. Vander, A. & Sherman, J. *Human physiology: the mechanisms of body function*. (William C Brown Pub, 2001).

212. Matthews, D. C. *et al.* Selective radiation of hematolymphoid tissue delivered by anti-CD45 antibody. *Cancer Res.* **52**, 1228–1234 (1992).

A1

Supplementary Tables, Figures and File Index

Figure S1. Extracted ion chromatograms (EICs) of steroid products and standards used in the quantification reported in Table S1. In each graph, the black trace is from a standard, while the red trace is from an extract of cells transiently expressing an enzyme and supplied with a substrate. The blue trace is from an extract of cells transiently expressing a fluorescent reporter instead of an enzyme, but supplied with the same substrate. A) EIC of 299.2 Da ions, corresponding to the $[M+H-H_2O]^+$ ion of **1**. B) EIC of 297.22 Da ions, corresponding to the $[M+H-2H_2O]^+$ ion of **2**. C) EIC of 331.23 Da ions, corresponding to the $[M+H]^+$ ion of **7**. D) EIC of 303.20 Da ions, corresponding to the $[M+H]^+$ ion of **10**. E) EIC of 315.2 Da ions, corresponding to the $[M+H]^+$ ion of **12**. F) EIC of 313.21 Da ions, corresponding to the $[M+H]^+$ ion of **13**. G) EIC of 391.28 Da ions, corresponding to the $[M]^-$ ion of **15**. H) EIC of 367.3 Da ions, corresponding to the $[M+H-2H_2O]^+$ ion of **16**. I) EIC of 331.23 Da ions, corresponding to the $[M+H]^+$ ion of **17**. J) EIC of 385.2 Da ions, corresponding to the $[M+H-H_2O]^+$ ion of **18**. K) EIC of 385.2 Da ions, corresponding to the $[M+H-H_2O]^+$ ion of **19**. L) EIC of 317.2 Da ions, corresponding to the $[M+H]^+$ ion of **20**. M) EIC of 287.2 Da ions, corresponding to the $[M+H]^+$ ion of **22**.

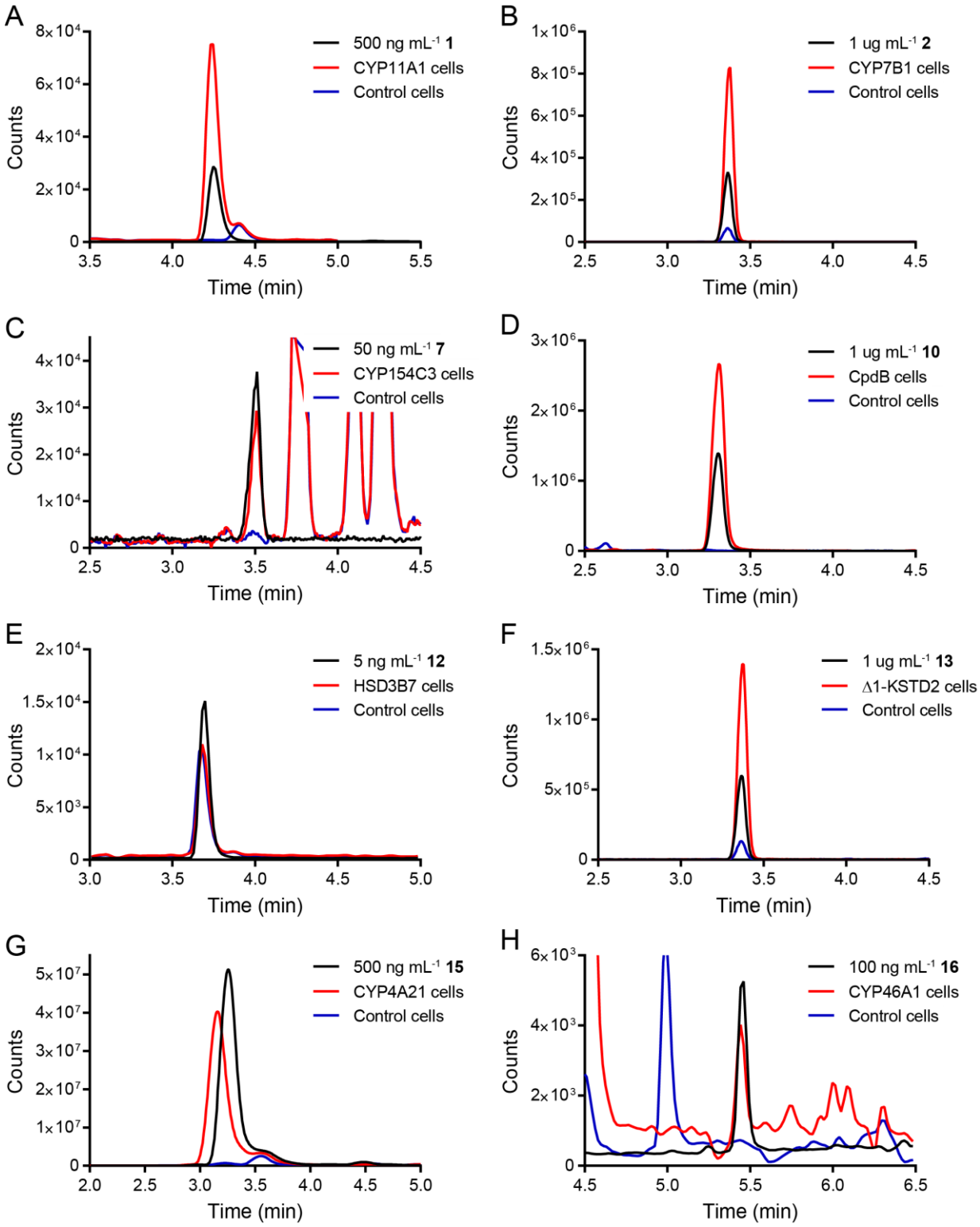


Figure S1 (Continued).

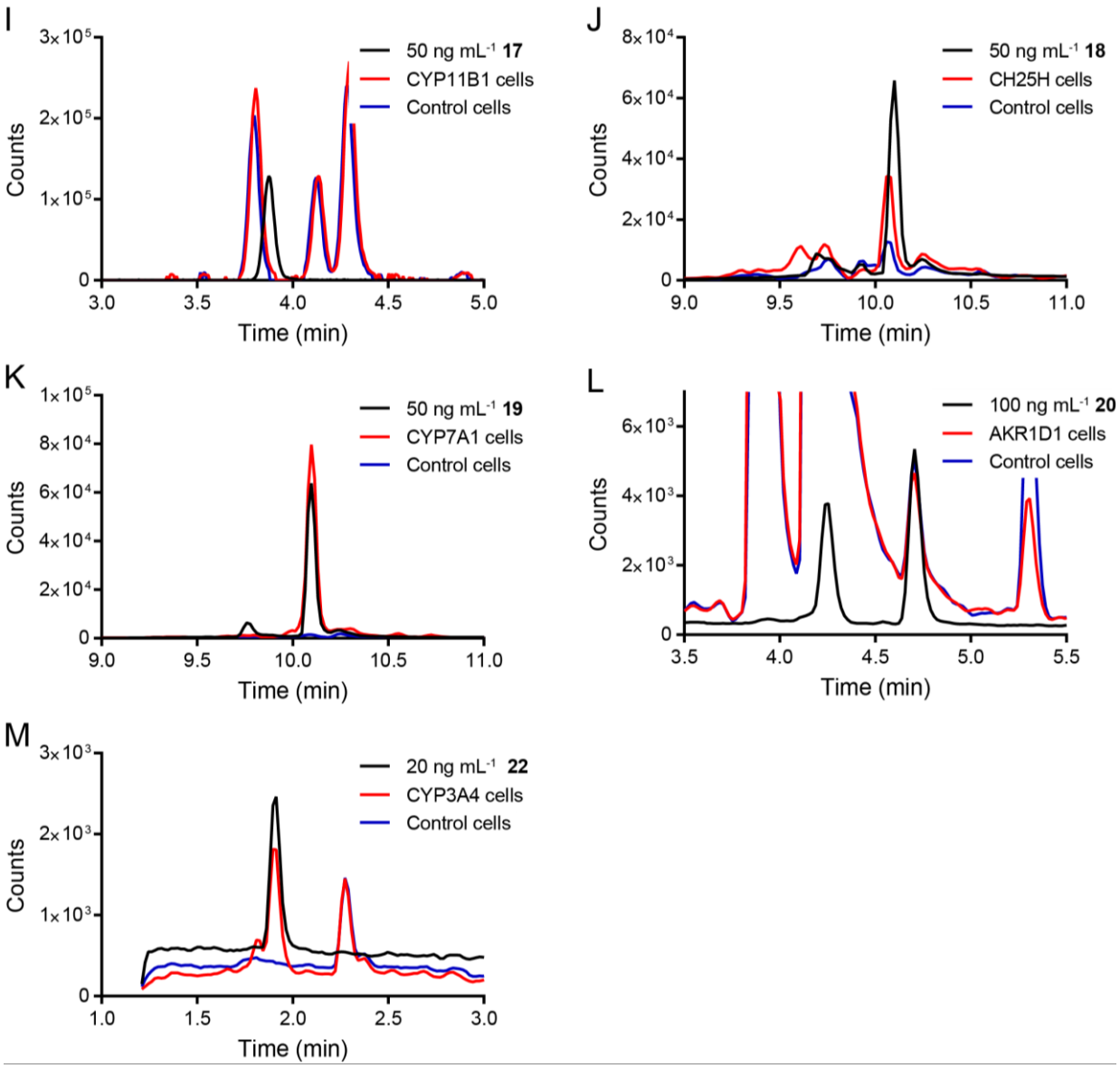


Figure S1 (Continued).

Table S1. Standard curve construction for steroid quantification. Linear regressions were used to calculate the relationship between concentration and peak area for each molecule. These were used to calculate the amount of product in cell culture extracts. Equations are of the form a*(concentration) + b = peak area or a*(concentration)^2 + b*(concentration) + c = peak area.

Molecule: **1**	*m/z* $[M+H-H_2O]^+$: 299.2	RT: 4.3 min

Standard Conc. (ng/mL)	**Peak Area**
5	2431
20	5783
100	24498
500	117871

Param.	**Value**
a	233.02
b	123.99
R^2	0.99997

Sample Name	**Peak Area**	**Calculated Conc. (ng/mL)**
CYP11A1_1	393940	1685
CYP11A1_2	354625	1517
CYP11A1_2	330552	1413
Control_1	N.D.	N.D.
Control_2	N.D.	N.D.
Control_3	N.D.	N.D.

Molecule:	**2**		*m/z* Σ($[M+H]^+$, $[M+H-H_2O]^+$, $[M+H-2H_2O]^+$) :	Σ(333.24, 315.23, 297.22)	RT:	3.4 min

Standard Conc. (ng/mL)	**Peak Area**
10	16658950
5	9195007
1	2334576
0.5	1163717
0.1	223134
0	24715

Param.	**Value**
a	1677380
b	223969
R^2	0.9969

Sample Name	**Peak Area**	**Calculated Conc. (ng/mL)**
CYP7B1_1 dil*	5477487	3.1
CYP7B1_2 dil*	5604191	3.2
CYP7B1_3 dil*	4914167	2.8
Control_1	621646	0.24
Control_2	507744	0.17
Control_3	479499	0.15

Table S1 (Continued).

Molecule: **7**	*m/z* $[M+H]^+$: 331.23	RT: 3.5 min

Standard Conc. (ng/mL)	**Peak Area**
0	3477
0	4137
10	39524
50	165886
100	325295
500	1612174
1000	3161516

Param.	**Value**
a	3163750
b	8598
R^2	0.999927

Sample Name	**Peak Area**	**Calculated Conc. (ng/mL)**
CYP154C3_1	105413	31
CYP154C3_2	119161	35
CYP154C3_3	110917	32
Control_1	17767	2.9
Control_2	14694	1.9
Control_3	15730	2.3

Molecule: **10**	*m/z* $[M+H]^+$: 303.20	RT: 3.2 min

Standard Conc. (ng/mL)	**Peak Area**
Enzyme Curve	
5	24540422
1	6553987
0.5	3417817
0.1	705753
0	4083
Control Curve	
1	3989516
0.5	2101268
0.1	422434
0.05	265399
0.01	41395
0	1827
0	14322

Param.	**Value**
Enzyme Curve	
a	4817780
b	684938
R^2	0.99526
Control Curve	
a	3999030
b	22852
R^2	0.99929

Sample Name	**Peak Area**	**Calculated Conc. (ng/mL)**
CpdB_1 dil*	11278897	2.2
CpdB_2 dil*	13463995	2.7
CpdB_3 dil*	14887658	2.9
Control_1	120252	0.02
Control_2	73614	0.01
Control_3	441322	0.10

Table S1 (Continued).

Molecule: **12**	*m/z* $[M+H]^+$: 315.2	RT: 3.7 min

Standard Conc. (ng/mL)	**Peak Area**
5	54573
20	140818
100	650642
500	2489639

Param.	**Value**
a	-3.1985
b	6540.9
c	20139
Weight	1/x
R^2	0.99965

Sample Name	**Peak Area**	**Calculated Conc. (ng/mL)**
HSD3B7_1 dil†	38568	2.8
HSD3B7_2 dil†	43039	3.5
HSD3B7_3 dil†	50824	4.7
Control_1 dil†	38507	2.8
Control_2 dil†	44858	3.8
Control_3 dil†	64496	6.8

Molecule: **13**	*m/z* $[M+H]^+$: 313.21	RT: 6.6 min

Standard Conc. (ng/mL)	**Peak Area**
Enzyme Curve	
100	765274
500	3675985
750	4907704
1000	6309987
2500	13285317
5000	24181333
0	2241
0	2480
Control Curve	
100	895559
10	149164
1	59459
0.01	44899
0.1	46806
0	33522

Param.	**Value**
Enzyme Curve	
a	4803.5
b	726938
R^2	0.9936
Control Curve	
a	8497.1
b	47550
R^2	0.9991

Sample Name	**Peak Area**	**Calculated Conc. (ng/mL)**
Δ1-KSTD2_1	6335700	1168
Δ1-KSTD2_2	5245124	941
Δ1-KSTD2_3	6183075	1136
Control_1	150780	12.1
Control_2	134751	10.3
Control_3	154921	12.6

Table S1 (Continued).

Molecule: **15**	*m/z* $[M]^-$: 391.28	RT: 3.2 min

Standard Conc. (ng/mL)	**Peak Area**
20	3.620E+07
100	1.170E+08
500	5.457E+08
2000	1.769E+09

Param.	**Value**
a	-115.986
b	1110570
c	13072500
Weight	1/x
R^2	0.9998

Sample Name	**Peak Area**	**Calc. Conc. (ng/mL)**
CYP4A21_1 dil†	4.767E08	437.48
CYP4A21_2 dil†	2.319E08	201.24
CYP4A21_3 dil†	3.426E08	306.57
Control_1 dil†	N.D.	N.D.
Control_2 dil†	3.43E08	1.87
Control_3 dil†	N.D.	N.D.

Molecule: **16**	*m/z* $[M+H-2H_2O]^+$: 367.34	RT: 5.5 min

Standard Conc. (ng/mL)	**Peak Area**
Enzyme Curve	
20	5251
100	16072
500	56470
1000	88875
Control Curve	
20	1327
100	2290
2000	24941
10000	75474

Param.	**Value**
Enzyme Curve	
a	-0.04688
b	132.68
c	2703.3
Weight	1/x
R^2	0.99957
Control Curve	
a	-5.586E-04
b	13.029
c	1054.8
Weight	1/x
R^2	0.99994

Sample Name	**Peak Area**	**Calculated Conc. (ng/mL)**
CYP46A1_1 dil†	12496	76
CYP46A1_2 dil†	9561	53
CYP46A1_3 dil†	11586	69
Control_1 dil†	N.D.	N.D.
Control_2 dil†	N.D.	N.D.
Control_3 dil†	N.D.	N.D.

Table S1 (Continued).

Molecule: **17**	*m/z* $[M+H]^+$: 331.23	RT: 3.9 min

Standard Conc. (ng/mL)	Peak Area
10	93710
50	541265
50	460129
100	883652
500	3952188

Param.	Value
a	7750620
b	85601
R^2	0.9984

Sample Name	Peak Area	Calculated Conc. (ng/mL)
CYP11B1_1	1008806	119
CYP11B1_2	1304978	157
CYP11B1_3	1410937	171
Control_1	1063593	126
Control_2	1209550	145
Control_3	838570	97

Molecule: **18**	*m/z* $[M+H-2H_2O]^+$: 367.34	RT: 9.9 min

Standard Conc. (ng/mL)	Peak Area
5	1750
20	5694
50	11358
200	43733
500	115384
1000	227498

Param.	Value
a	220.3
b	863.2
Weight	1/x
R^2	0.9959

Sample Name	Peak Area	Calculated Conc. (ng/mL)
Control_1	14598	62.4
Control_2	9392	38.7
Control_3	11488	48.2
CH25H_1	20973	91.3
CH25H_2	22182	96.8
CH25H_3	20366	88.5

Molecule: **19**	*m/z* $[M+H-2H_2O]^+$: 367.34	RT: 10.1 min

Standard Conc. (ng/mL)	Peak Area
5	16172
20	67430
50	150172
200	761017
500	1273335

Param.	Value
a	3638.15
b	-3465.15
Weight	1/x
R^2	0.986

Sample Name	Peak Area	Calculated Conc. (ng/mL)
CYP7A1_1	223751	63
CYP7A1_2	205123	57
CYP7A1_3	210350	59
Control_1	23798	7
Control_2	6070	3
Control_3	25150	8

Table S1 (Continued).

Molecule: **20**	*m/z* $[M+H]^+$: 317.2	RT: 4.7 min

Standard Conc. (ng/mL)	**Peak Area**
20	5321
100	13312
500	51495
2000	149765

Param.	**Value**
a	-0.015293
b	103.87
c	3226.1
Weight	1/x
R^2	0.999979

Sample Name	**Peak Area**	**Calculated Conc. (ng/mL)**
AKR1D1_1	6734	34
AKR1D1_2	8510	51
AKR1D1_3	6510	32
Control_1	6913	36
Control_2	8629	52
Control_3	6222	29

Molecule: **22**	*m/z* $[M+H]^+$: 287.2	RT: 1.9 min

Standard Conc. (ng/mL)	**Peak Area**
20	6439
100	13589
500	60541
2000	214585
10000	769799

Param.	**Value**
a	-0.003586
b	112.46
c	3932.0
Weight	1/x
R^2	0.99964

Sample Name	**Peak Area**	**Calculated Conc. (ng/mL)**
CYP3A4_1	4625	6.2
CYP3A4_2	4631	6.2
CYP3A4_3	4592	5.9
Control_1	N.D.	N.D.
Control_2	N.D.	N.D.
Control_3	N.D.	N.D.

*1:10 dilution of sample before measurement
†1:1 dilution of sample before measurement

Table S2. Calculations for volumes, surface areas, initial amounts, and select rate constants used in the ODE model for glucocorticoid-binding fusion protein pharmacokinetics. Replacement of a source with equations indicate the ID letters of the relevant constants. References to 'k' numbers indicate a rate constant or set of constants that can take multiple values due to variation in inputs; this avoids showing repetitive arithmetic. All of the final 'k' values are in main text Table 4.2.

ID	Parameter	Value	Units	Source
A	Number of granulocytes in human body	4.55E+10	cells	Athens 1961[209]
B	Number of lymphocytes in human body	4.60E+11	cells	Trepel 1974[210]
C	Number of monocytes in human body	4.40E+09	cells	Vander 2001[211] and Alberts 2002[204]; assuming half of total population is in the blood
D	Volume of single granulocyte	300	μm^3	Ting-Beall 1993[192]
E	Volume of single lymphocyte	210	μm^3	Segel 1981[193]
F	Volume of single monocyte	240	μm^3	Schmid-Schonbein 1980[194]
G	Fraction of granulocyte volume in the cytoplasm	0.627		Schmid-Schonbein 1980[194]
H	Fraction of lymphocyte volume in the cytoplasm	0.516		Schmid-Schonbein 1980[194]
I	Fraction of monocyte volume in the cytoplasm	0.681		Schmid-Schonbein 1980[194]
J	Total leukocyte cytoplasm volume in body	5.91E-02	L	((A*D*G)+(B*E*H)+(C*F*I)) *1E-15
K	Endosome volume of single granulocyte	6.435	μm^3	Corlier 2015[195]
L	Endosome volume of single lymphocyte	6.435	μm^3	Corlier 2015[195]
M	Endosome volume of single monocyte	10	μm^3	Sarkar 2003[196]
N	Total leukocyte endosome volume in body	3.30E-03	L	((A*K)+(B*L)+(C*M))*1E-15
O	Surface area of a single granulocyte	300	μm^2	Schmid-Schonbein 1980[194]
P	Surface area of a single lymphocyte	250	μm^2	Schmid-Schonbein 1980[194]
Q	Surface area of a single monocyte	450	μm^2	Schmid-Schonbein 1980[194]
R	Total leukocyte plasma membrane surface area in body	1.31E+04	dm^2	((A*O)+(B*P)+(C*Q))*1E-10
S	Ratio of plasma membrane surface area of leukocyte to endosome surface area	0.3		Yogurtcu 2018[197] (estimate)
T	Total leukocyte endosome surface area in body	3.92E+03	dm^2	R*S

Table S2 (Continued).

ID	Parameter	Value	Units	Source
U	Apparent permeability coefficient for fluticasone propionate in lymphocytes	2.10E-09	$dm*sec^{-1}$	Brandish 2018[101]
V	Diffusion rate formula for fluticasone propionate	See k08, k28, k38, k48, k09, and k29	sec^{-1}	U*surface area *destination compart vol
W	Formula for receptor externalization rate (assuming constant total amount on surface and endosome)	See k25	sec^{-1}	Internalization rate*(R/T)
X	Number of glucocorticoid receptors in a single macrophage	1.30E+04	molecules	Biggin 2011[201]
Y	Concentration of glucocorticoid receptors in leukocyte cytoplasm	8.99E-08	M	(X*1E15)/(F*6.022E23)
Z	Number of CD45 receptors on a single cell from a leukocyte cell line	1.00E+05	molecules	Peyron 1991[198], Bausch-Fluck 2015[199], Matthews 1991[212]
α	CD45 receptors per um^2 of membrane	3.33E+02	molecules $*um^{-2}$	Z/D
β	Concentration of CD45 receptors in blood	3.99E-09	M	(α*1E10*R)/(6.022E23*18)
γ	Concentration of CD45 receptors in endosome	6.54E-06	M	(α*1E10*T)/(6.022E23*N)

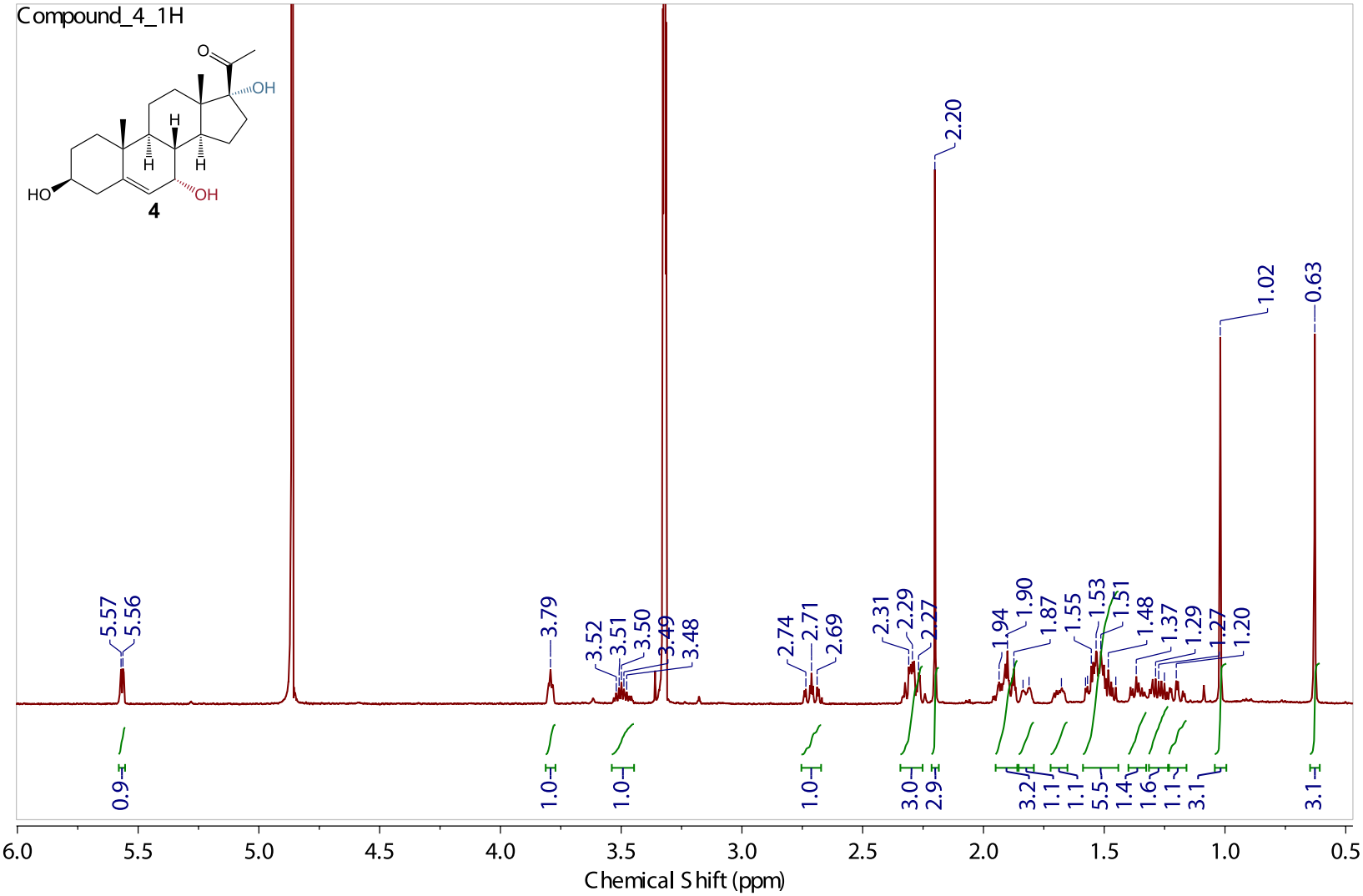


Figure S2. ^{1}H NMR spectrum of 7α,17α-dihydroxypregnenolone (compound **4**) (CD_3OD, 500 MHz).

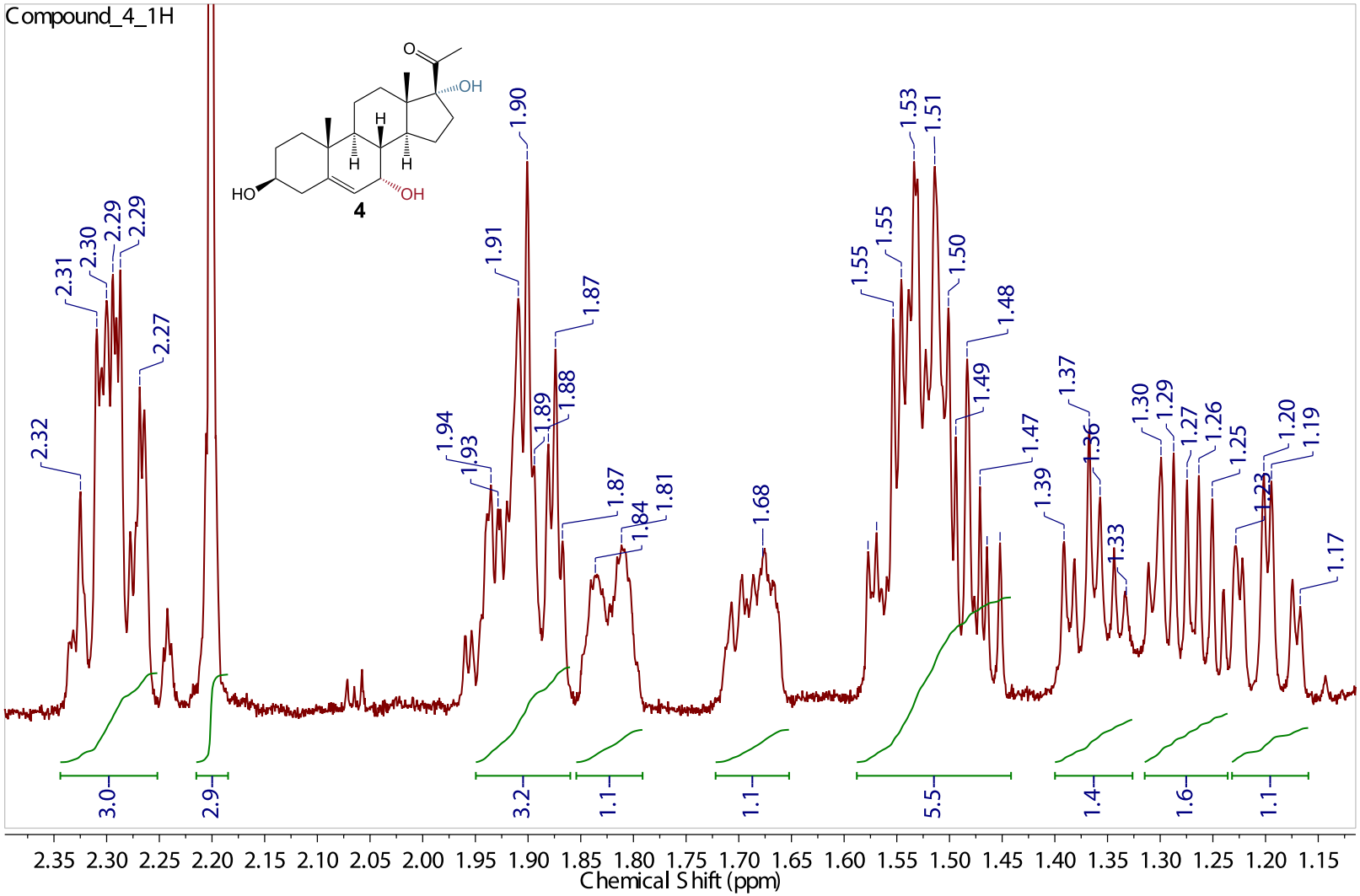


Figure S3. ^{1}H NMR spectrum of 7α,17α-dihydroxypregnenolone (compound **4**) (CD_3OD, 500 MHz).

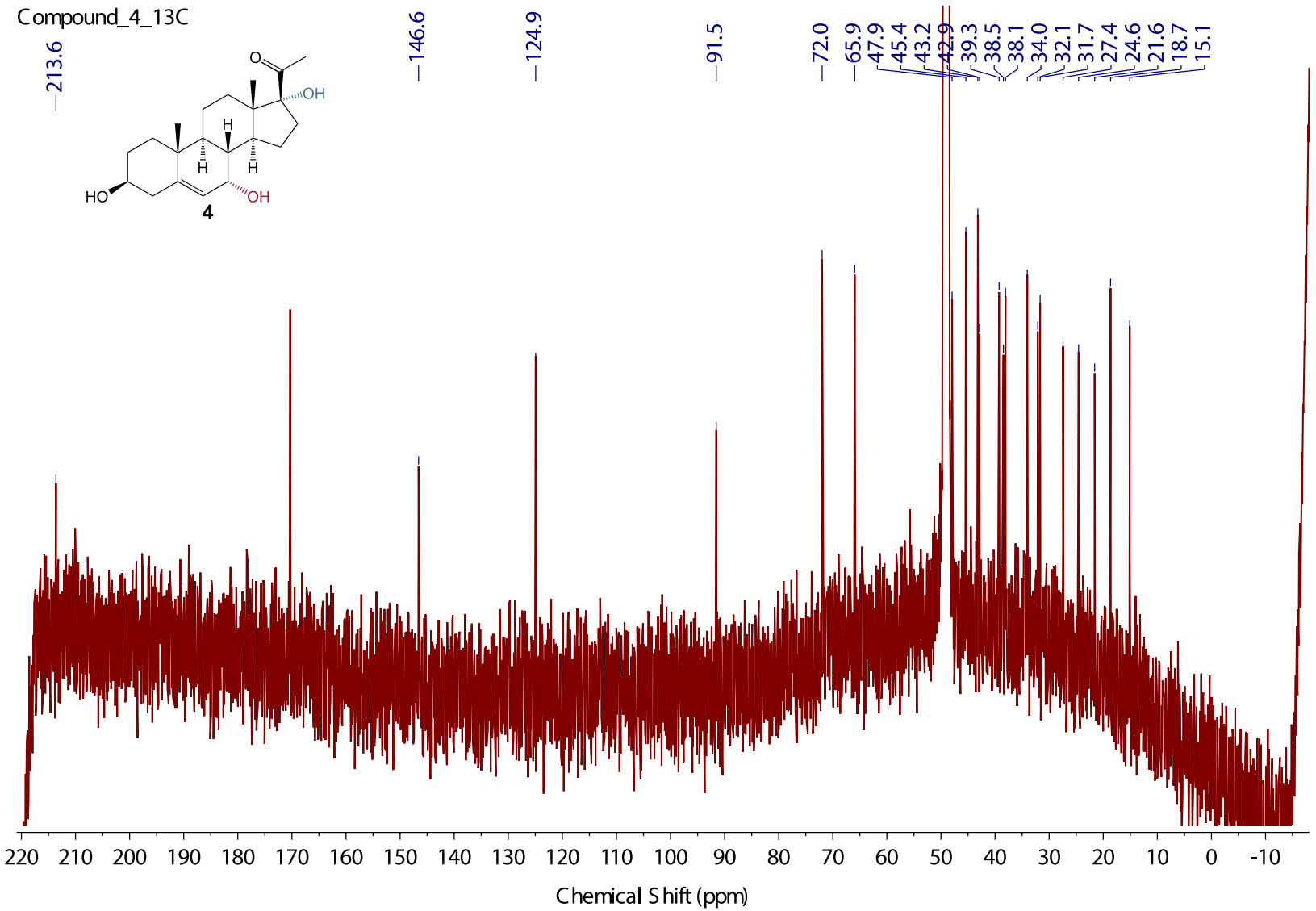


Figure S4. ^{13}C NMR spectrum of 7α,17α-dihydroxypregnenolone (compound **4**) (CD_3OD, 125 MHz).

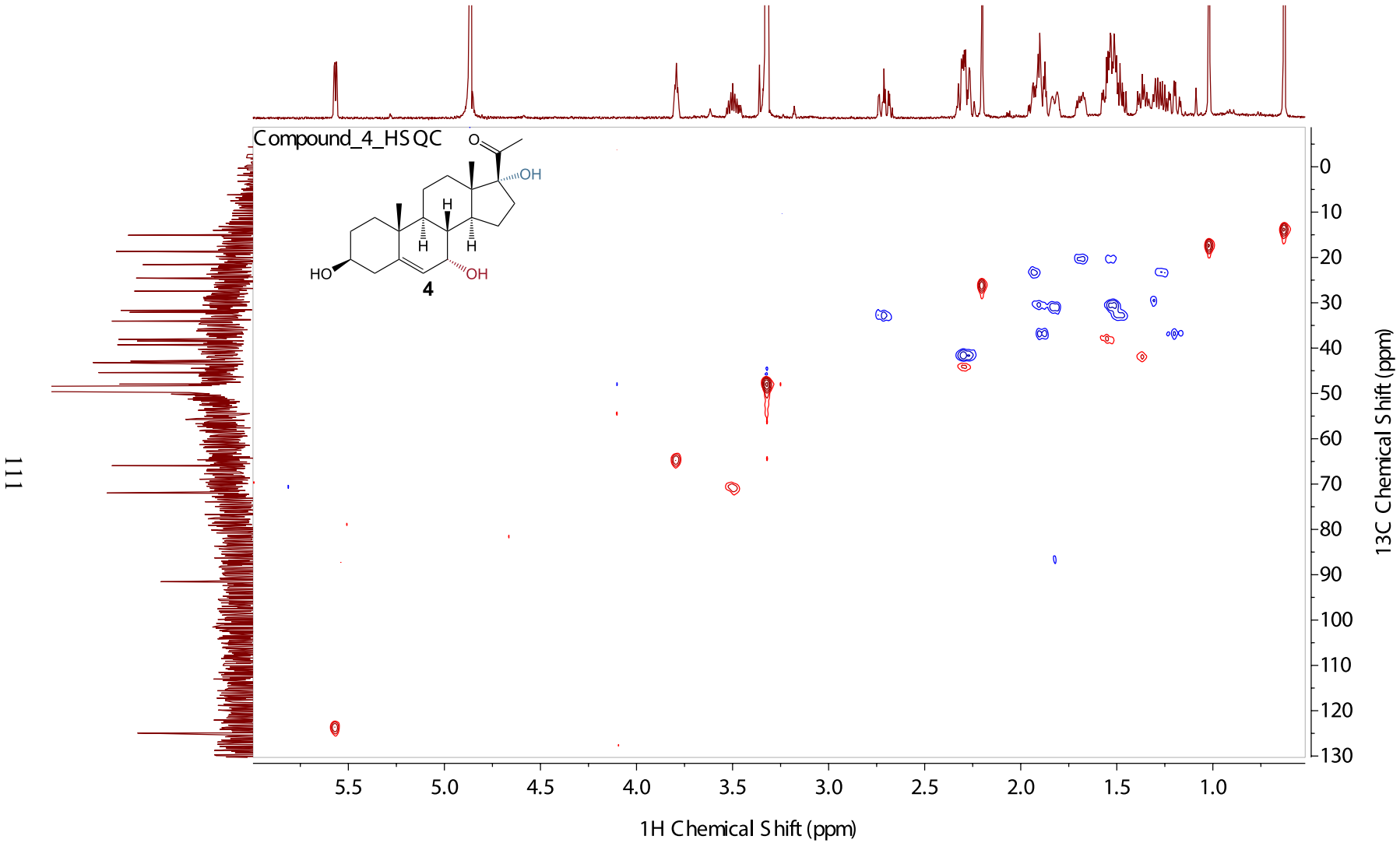


Figure S5. ^{1}H-^{13}C HSQC spectrum of 7α,17α-dihydroxypregnenolone (compound **4**) (CD_3OD, 500 MHz).

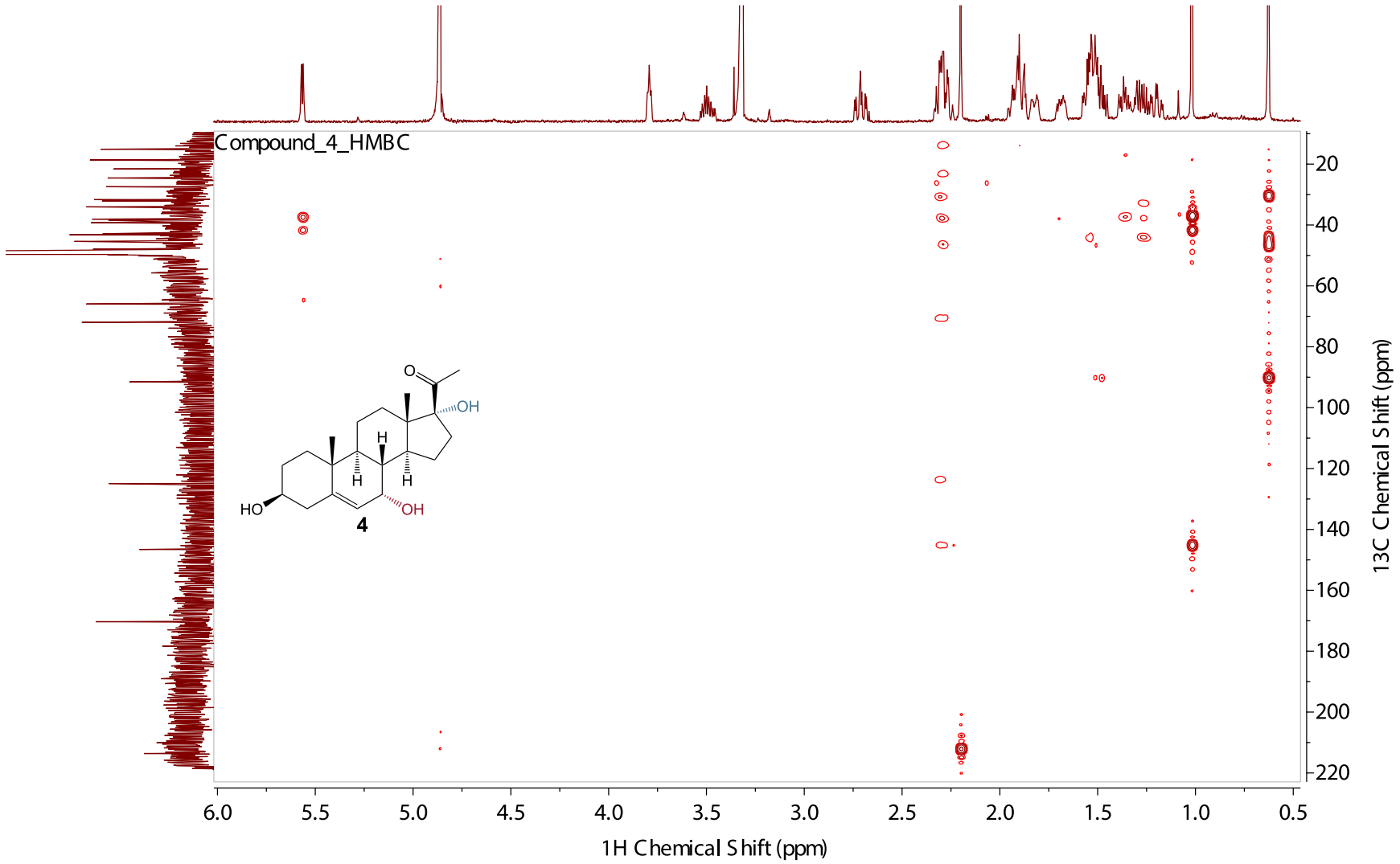


Figure S6. ^{1}H-^{13}C HMBC spectrum of 7α,17α-dihydroxypregnenolone (compound **4**) (CD_3OD, 500 MHz).

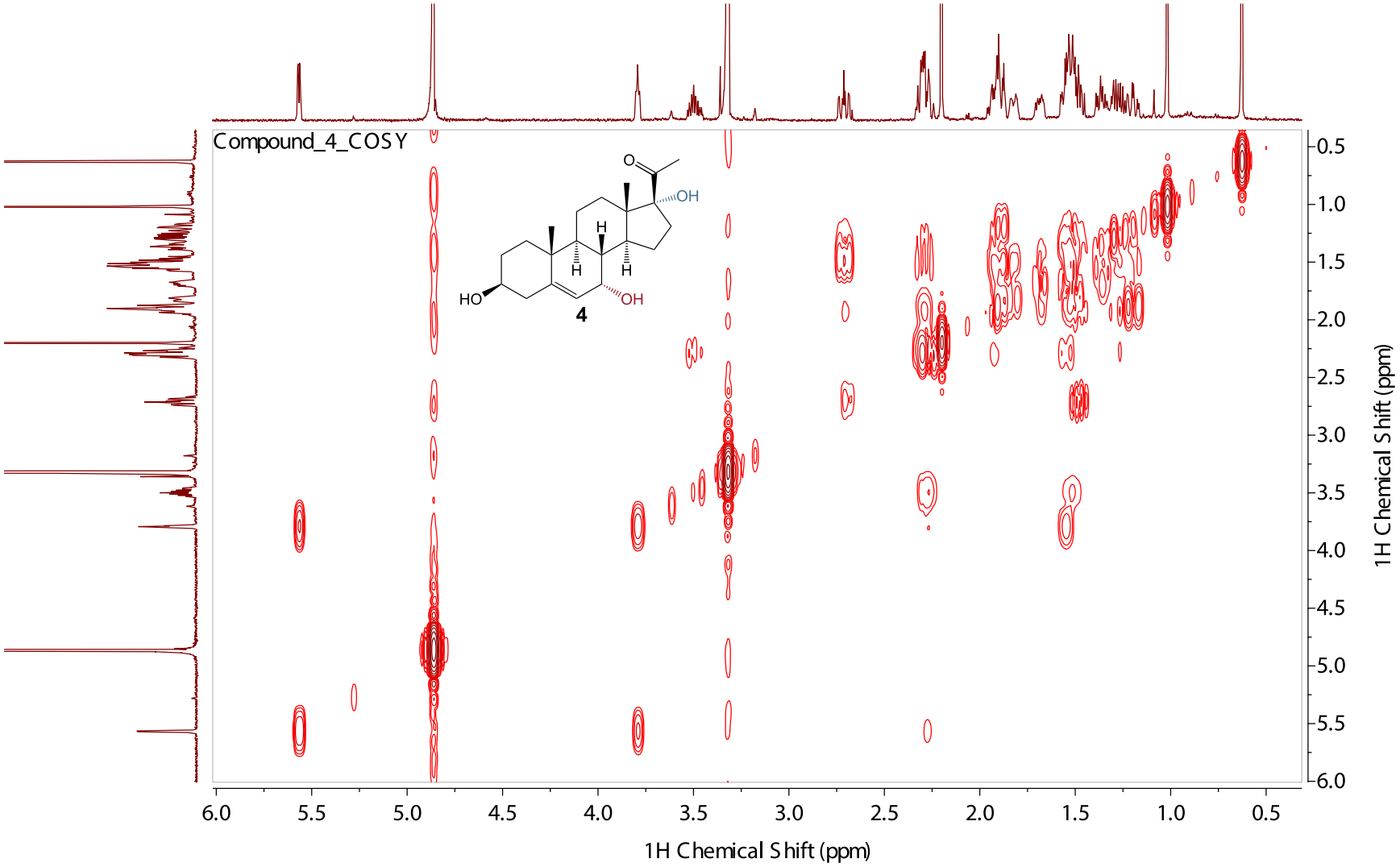


Figure S7. ^{1}H-^{1}H COSY spectrum of 7α,17α-dihydroxypregnenolone (compound **4**) (CD_3OD, 500 MHz).

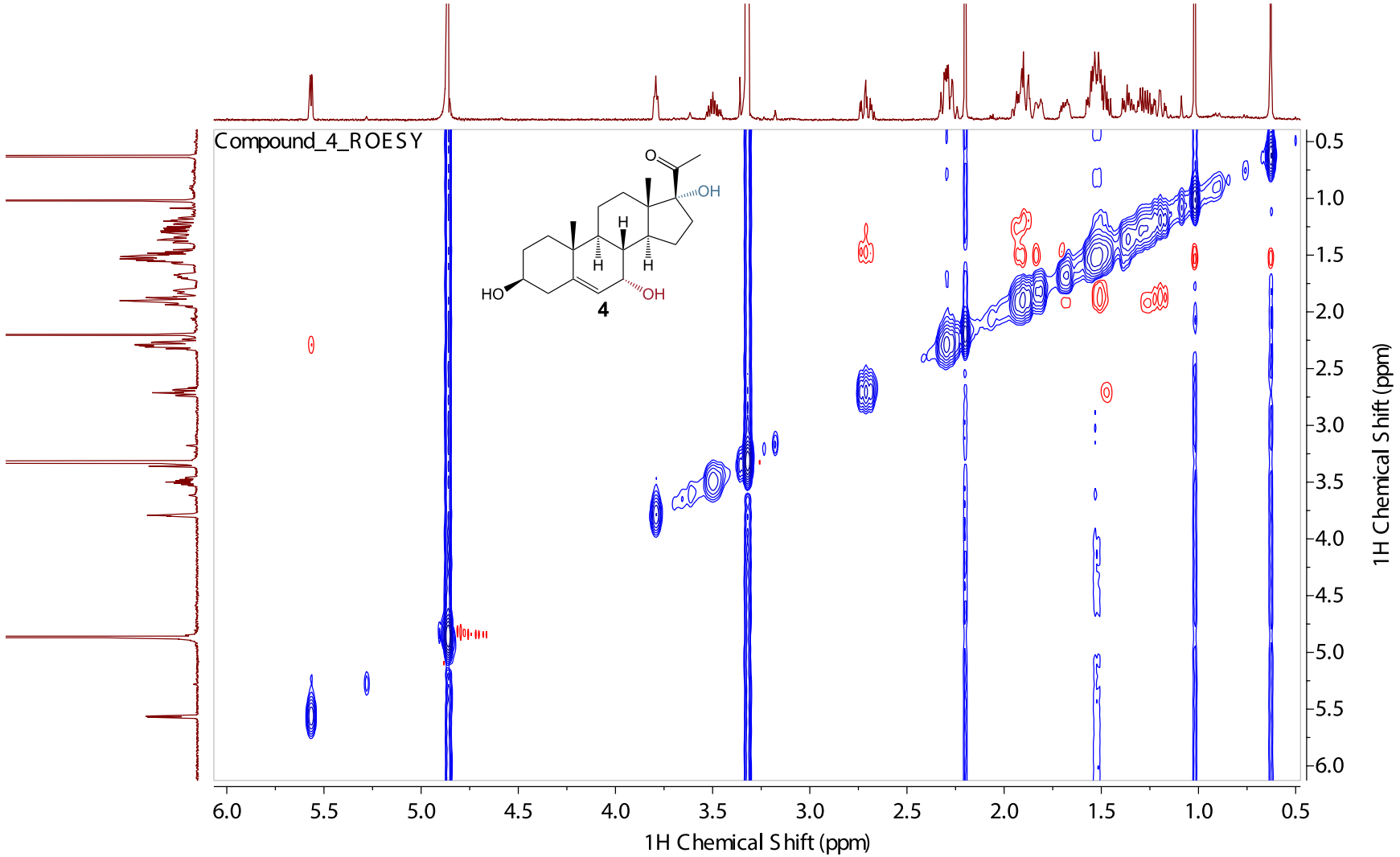


Figure S8. ^{1}H-^{1}H ROESY spectrum of 7α,17α-dihydroxypregnenolone (compound **4**) (CD_3OD, 500 MHz).

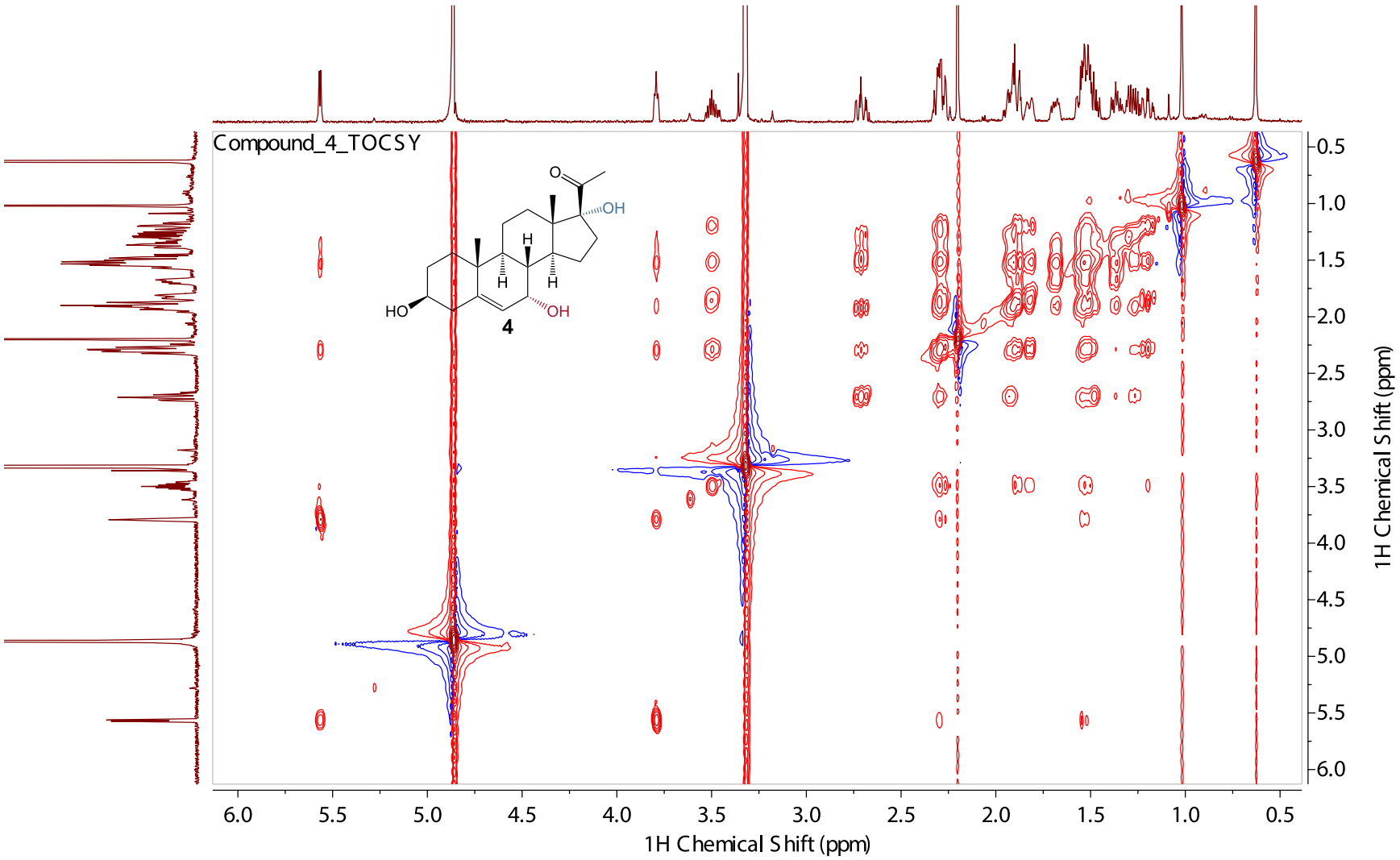


Figure S9. ^{1}H-^{1}H TOCSY spectrum of 7α,17α-dihydroxypregnenolone (compound **4**) (CD_3OD, 500 MHz).

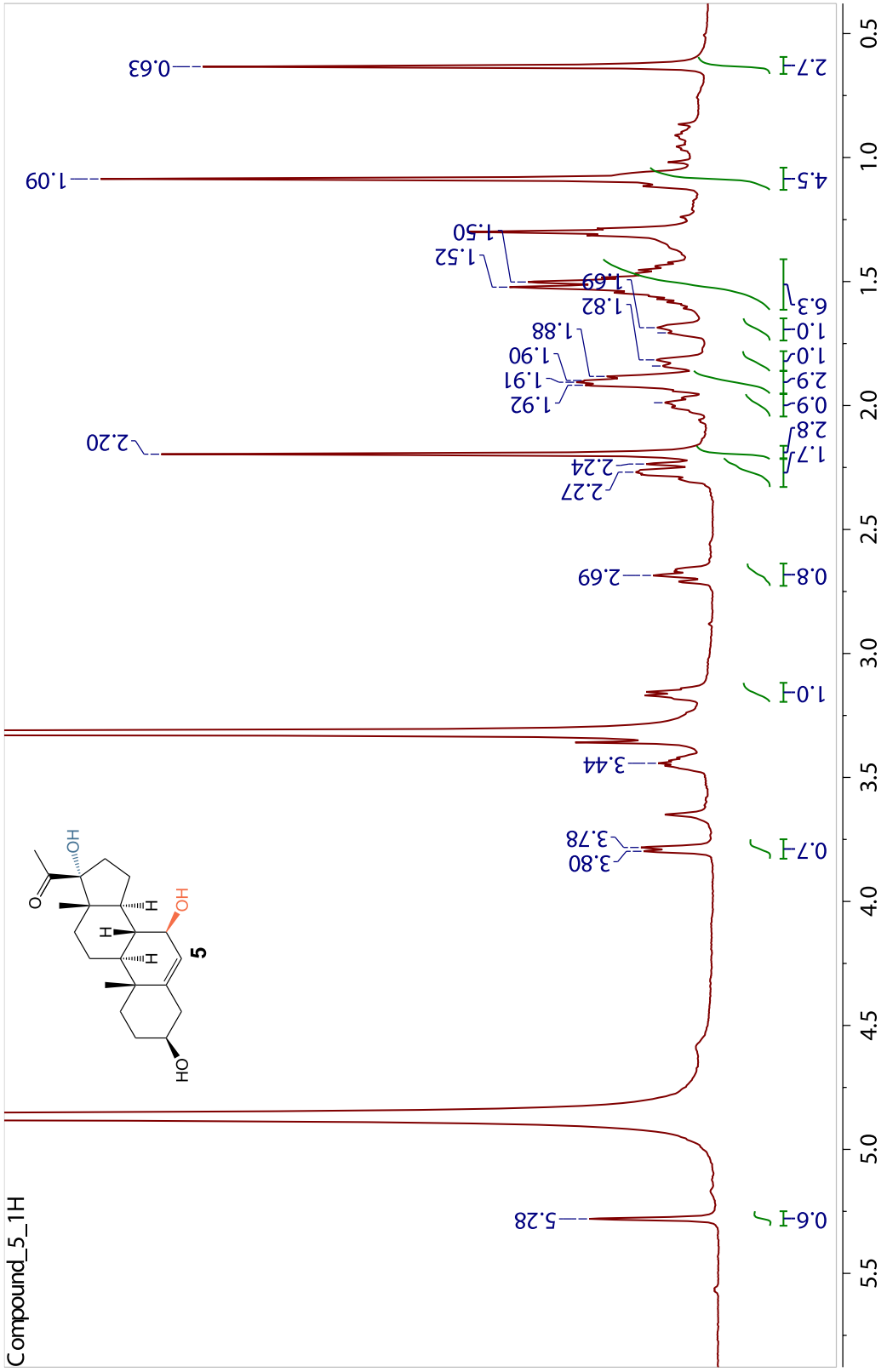


Figure S10. ^{1}H NMR spectrum of 7β,17α-dihydroxypregnenolone (compound **5**) (CD_3OD, 500 MHz).

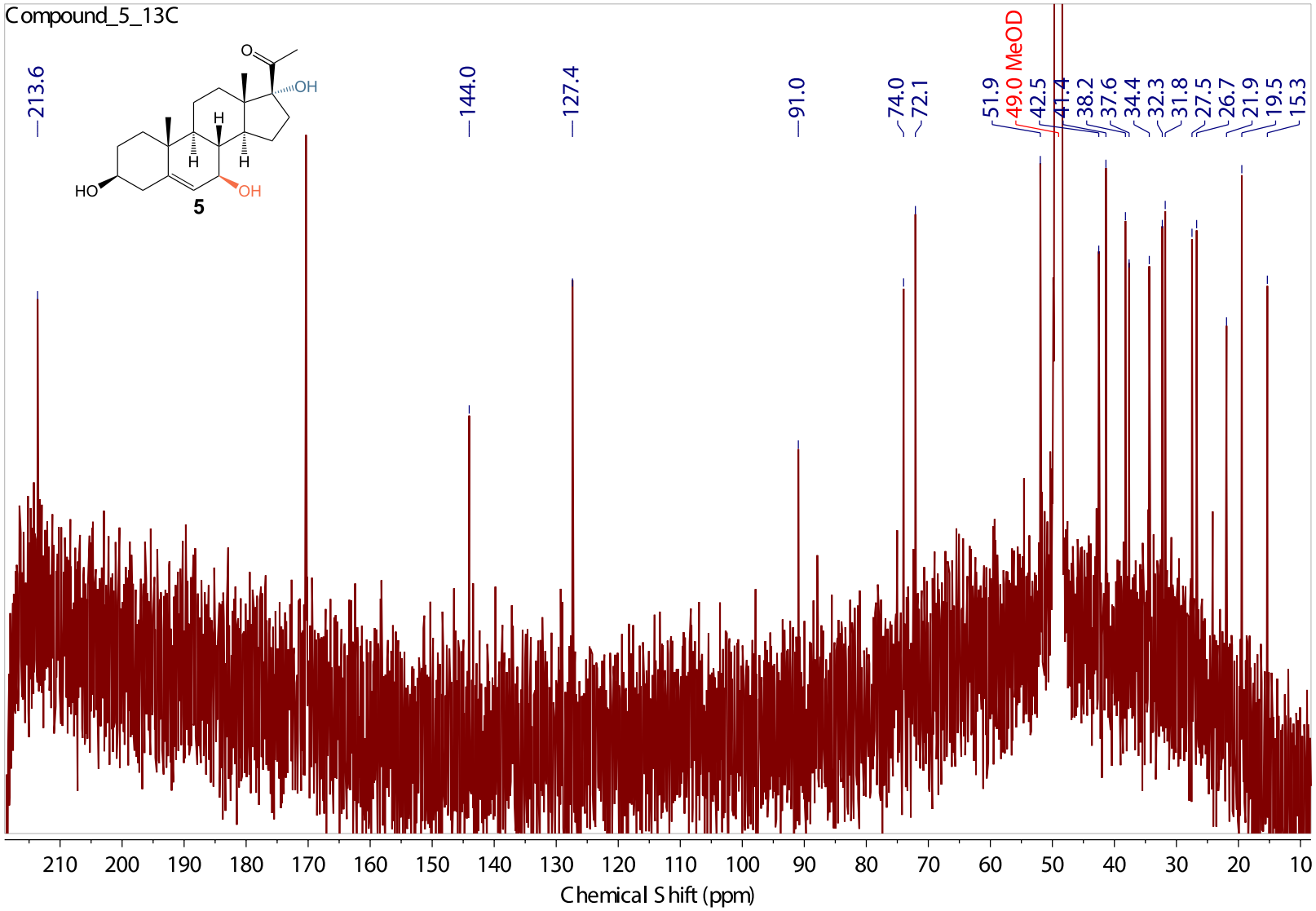


Figure S11. ^{13}C NMR spectrum of 7β,17α-dihydroxypregnenolone (compound **5**) (CD_3OD, 125 MHz).

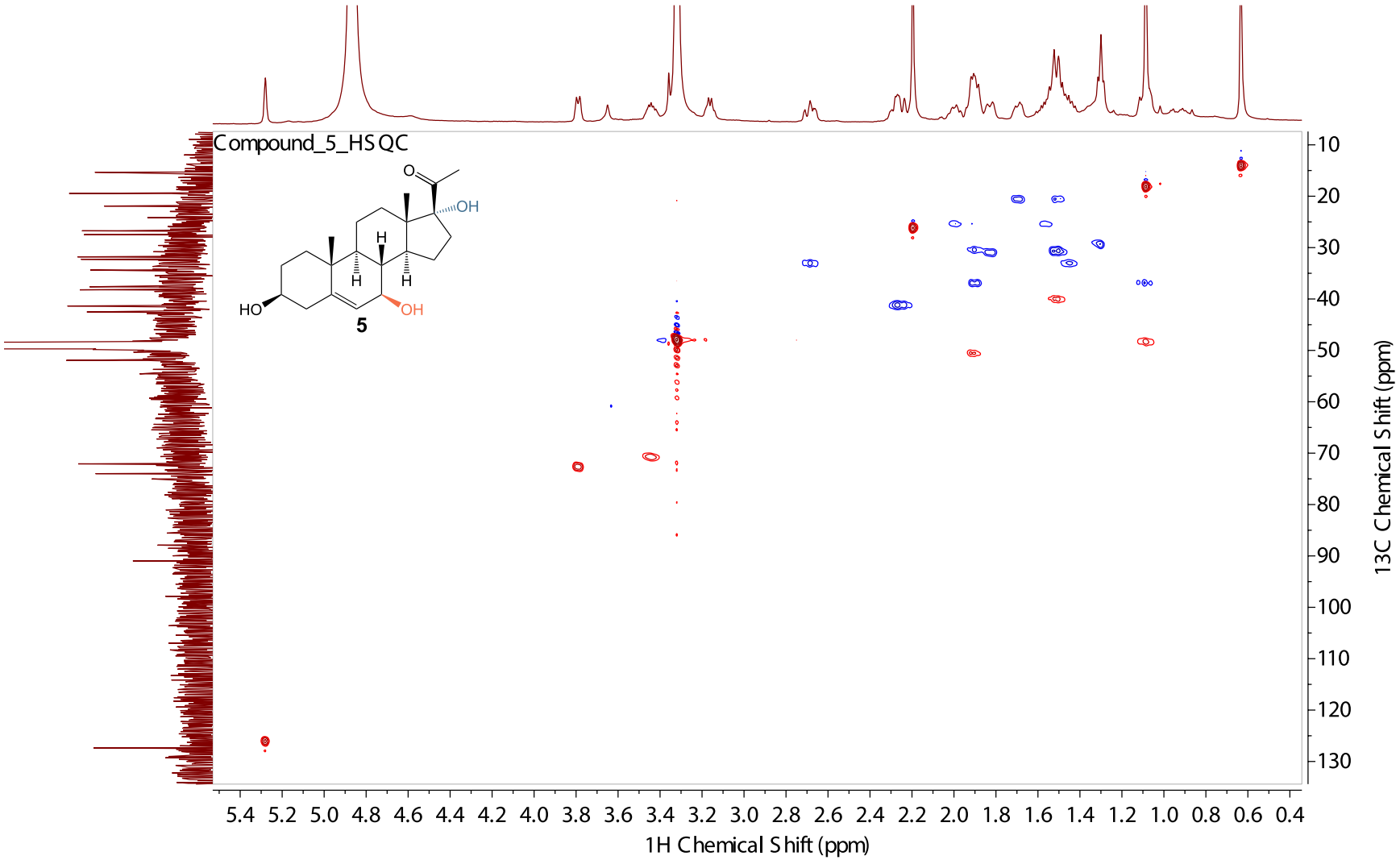


Figure S12. ^{1}H-^{13}C HSQC spectrum of 7β,17α-dihydroxypregnenolone (compound **5**) (CD_3OD, 500 MHz).

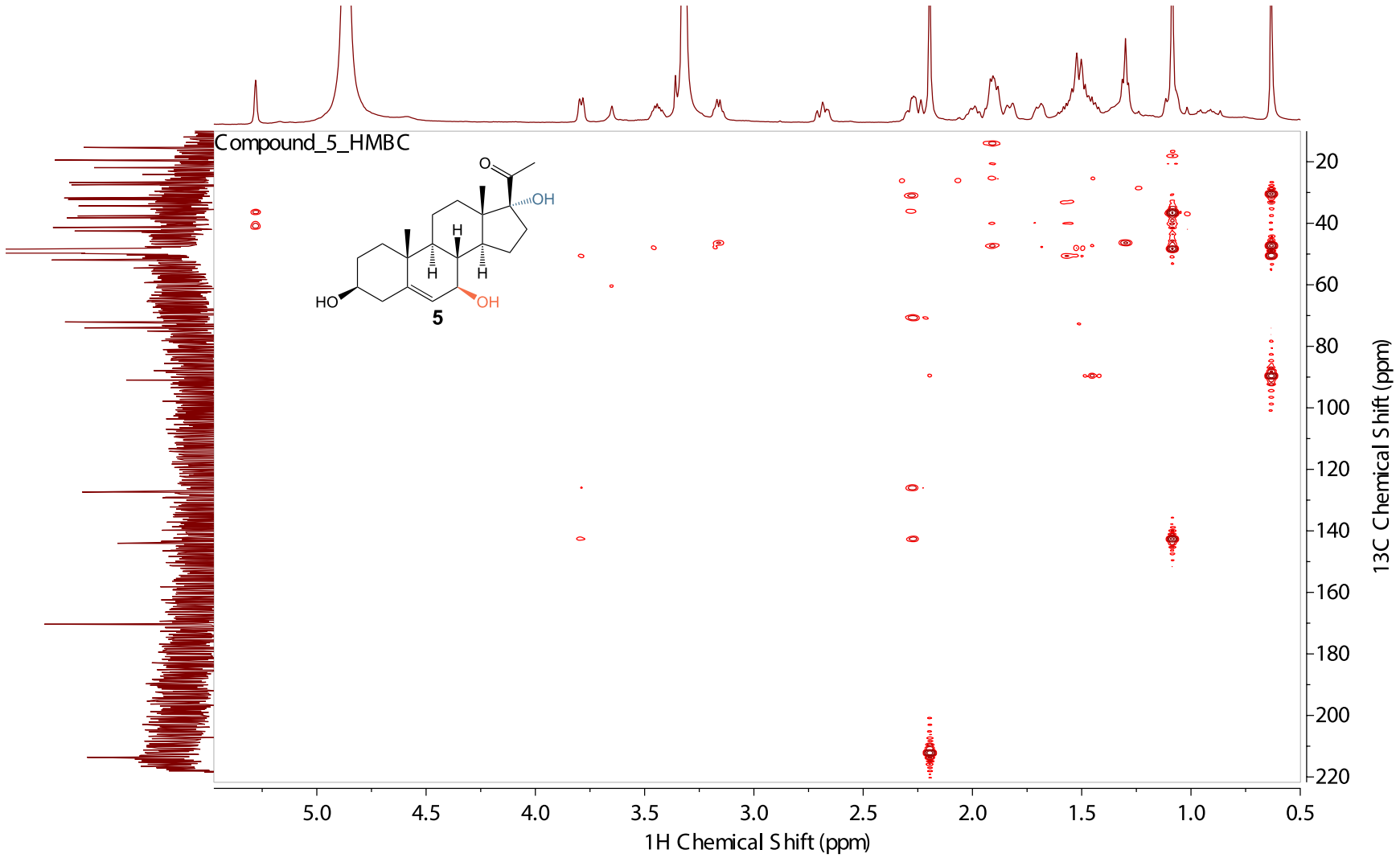


Figure S13. ^{1}H-^{13}C HMBC spectrum of 7β,17α-dihydroxypregnenolone (compound **5**) (CD_3OD, 500 MHz).

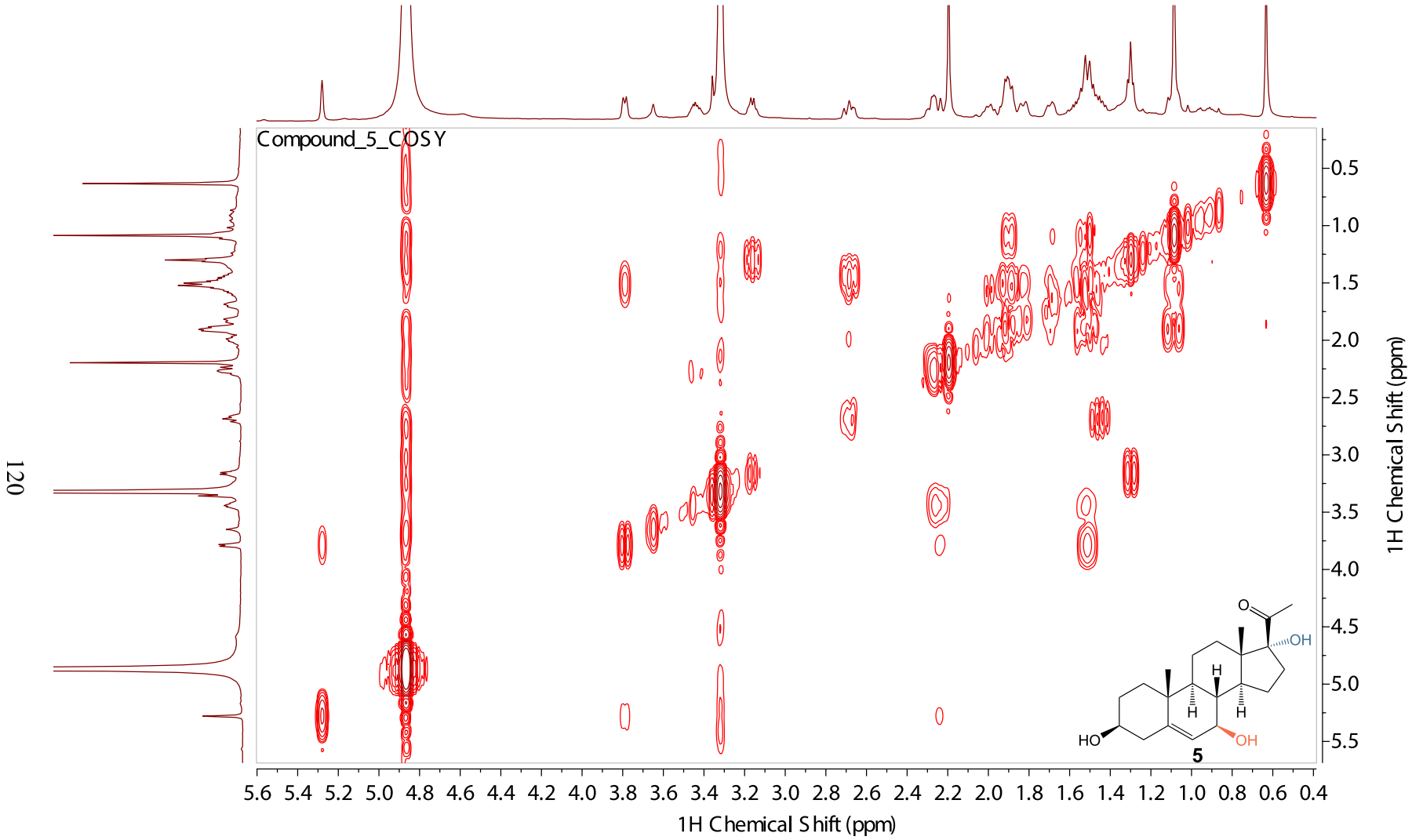


Figure S14. ^{1}H-^{1}H COSY spectrum of 7β,17α-dihydroxypregnenolone (compound **5**) (CD_3OD, 500 MHz).

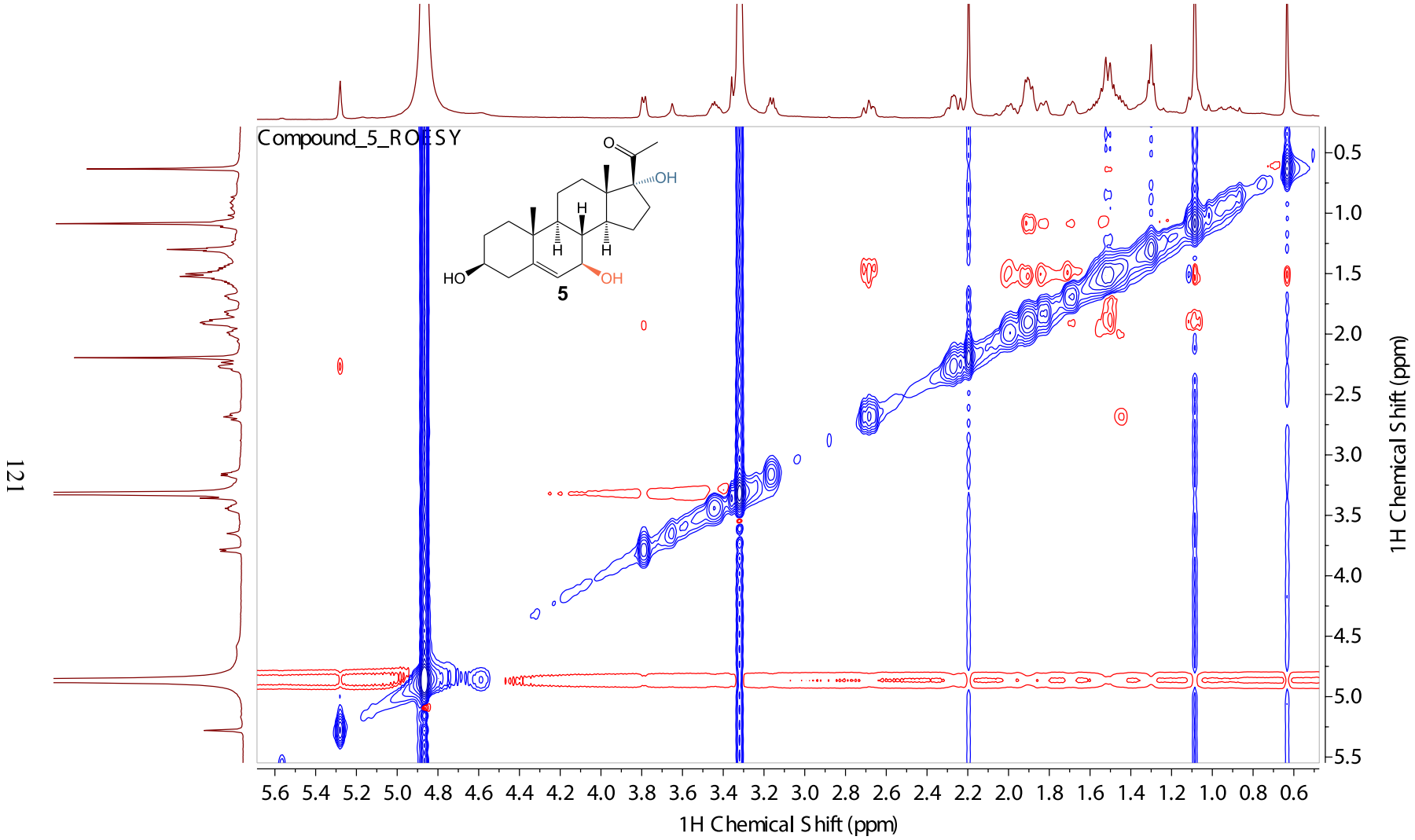


Figure S15. ^{1}H-^{1}H ROESY spectrum of 7β,17α-dihydroxypregnenolone (compound **5**) (CD_3OD, 500 MHz).

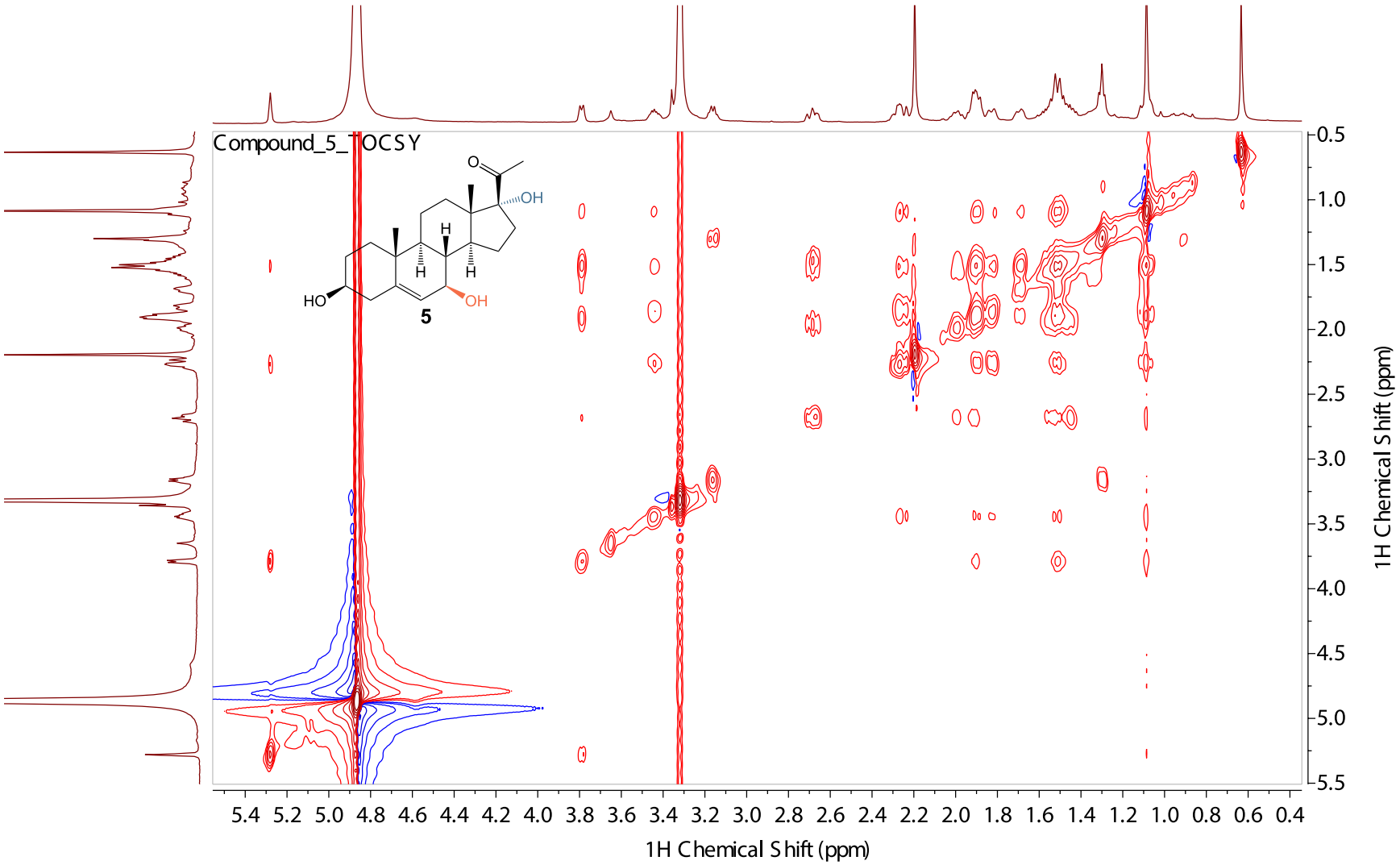


Figure S16. ^{1}H-^{1}H TOCSY spectrum of 7β,17α-dihydroxypregnenolone (compound **5**) (CD_3OD, 500 MHz).

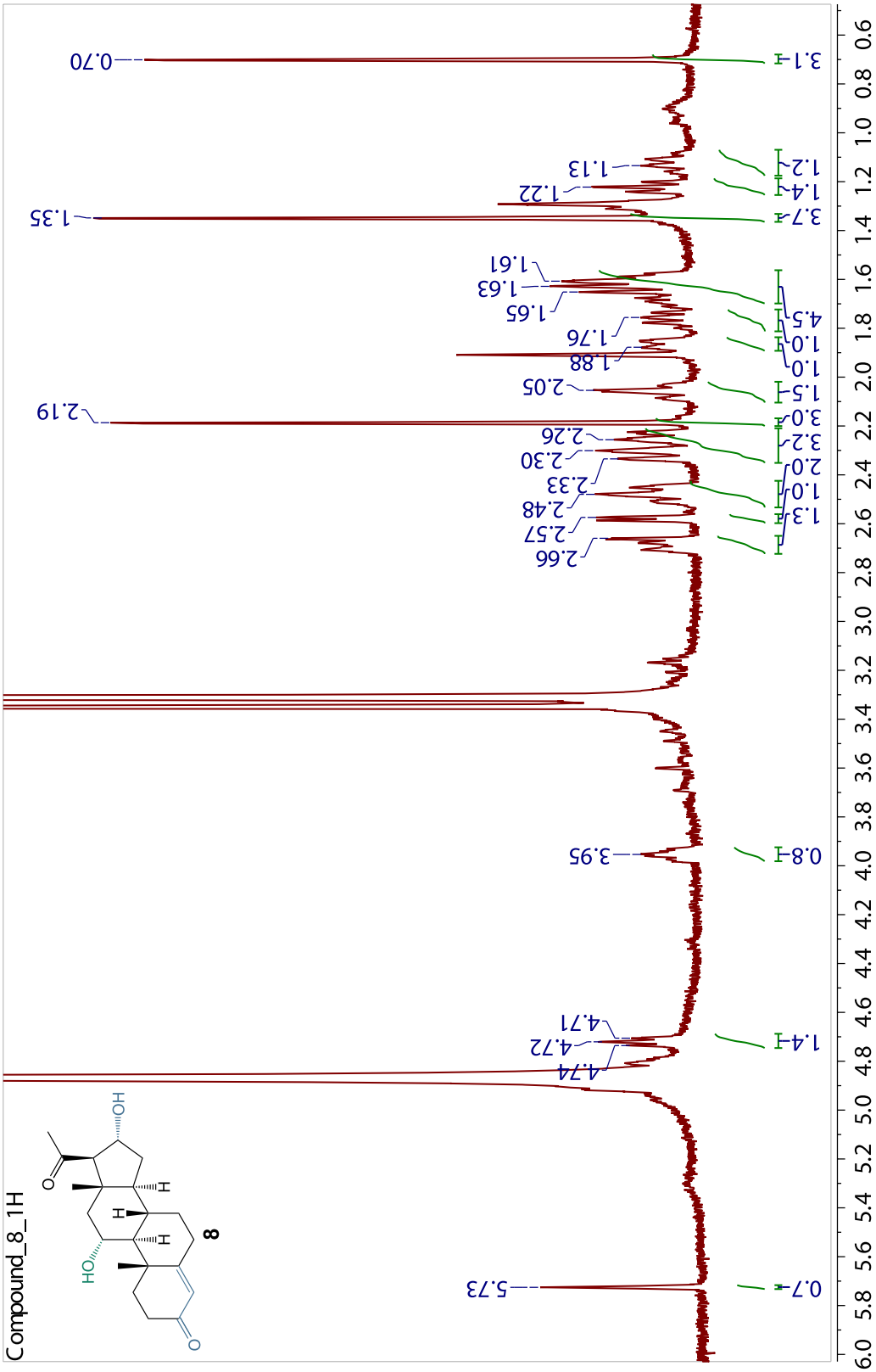


Figure S17. ^{1}H NMR spectrum of 11α,16α-dihydroxyprogesterone (compound **8**) (CD_3OD, 500 MHz).

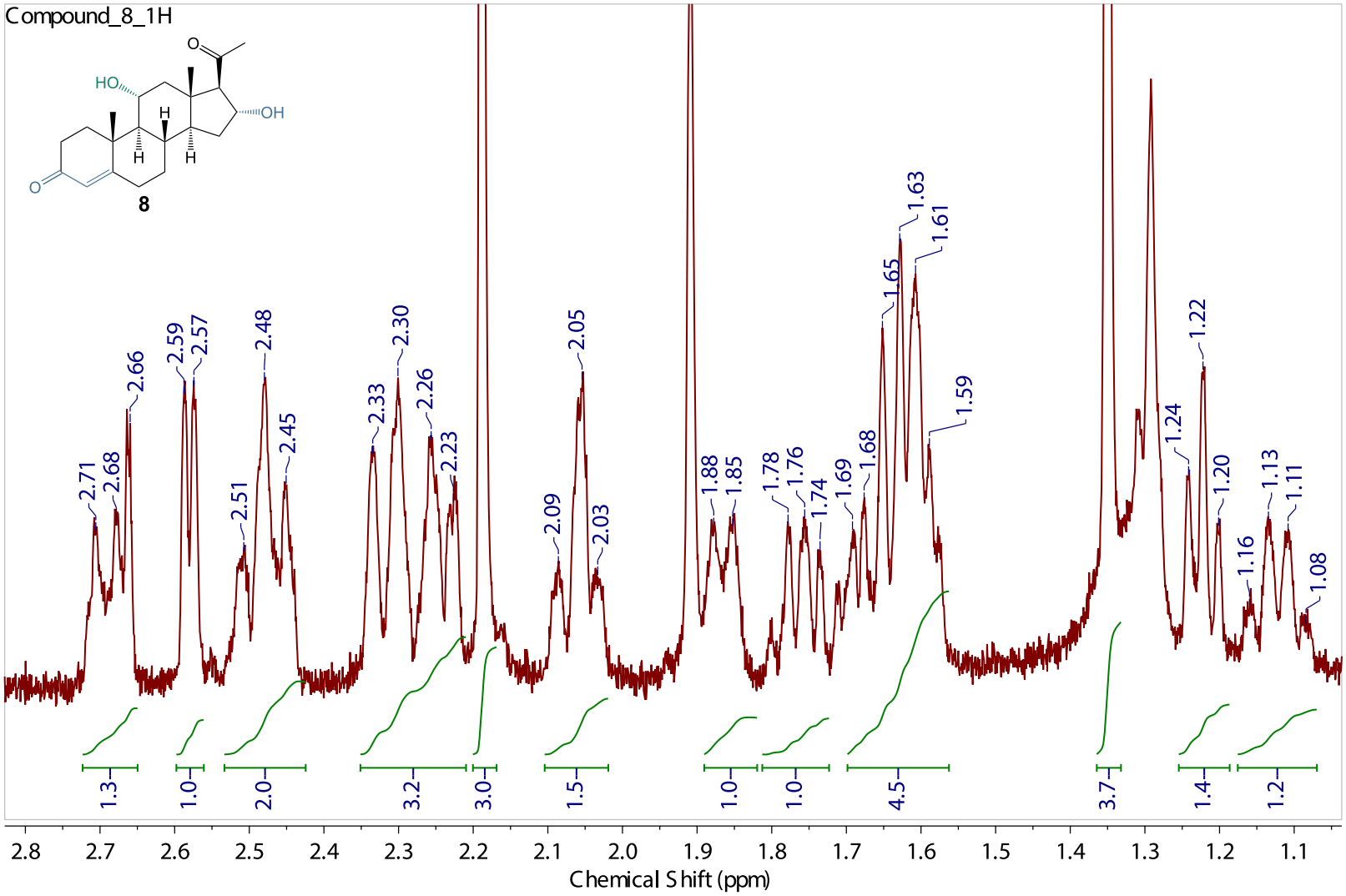


Figure S18. ^{1}H NMR spectrum of 11α,16α-dihydroxyprogesterone (compound **8**) (CD_3OD, 500 MHz).

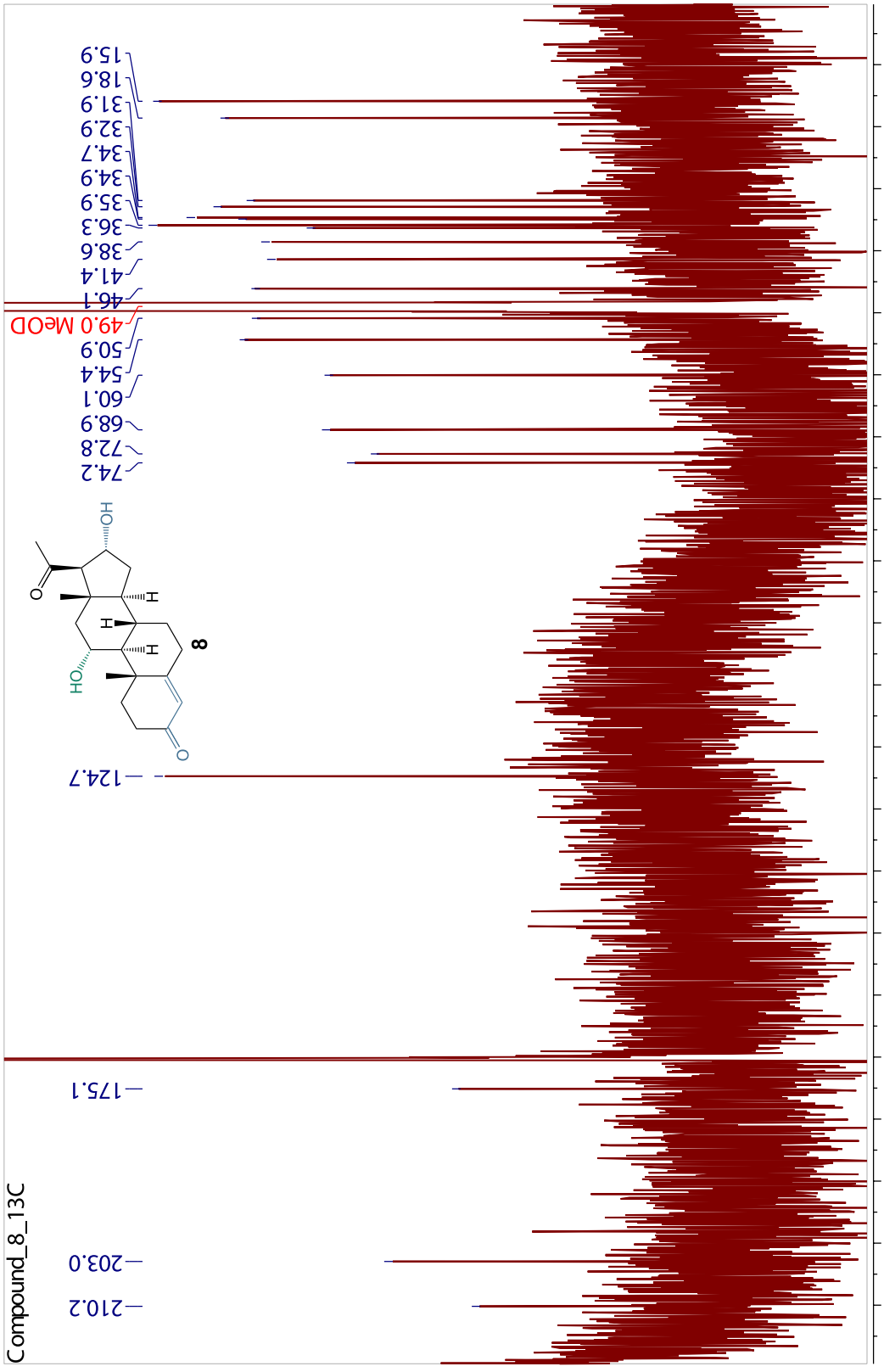


Figure S19. ^{13}C NMR spectrum of 11 α,16α-dihydroxyprogesterone (compound **8**) (CD_3OD, 125 MHz).

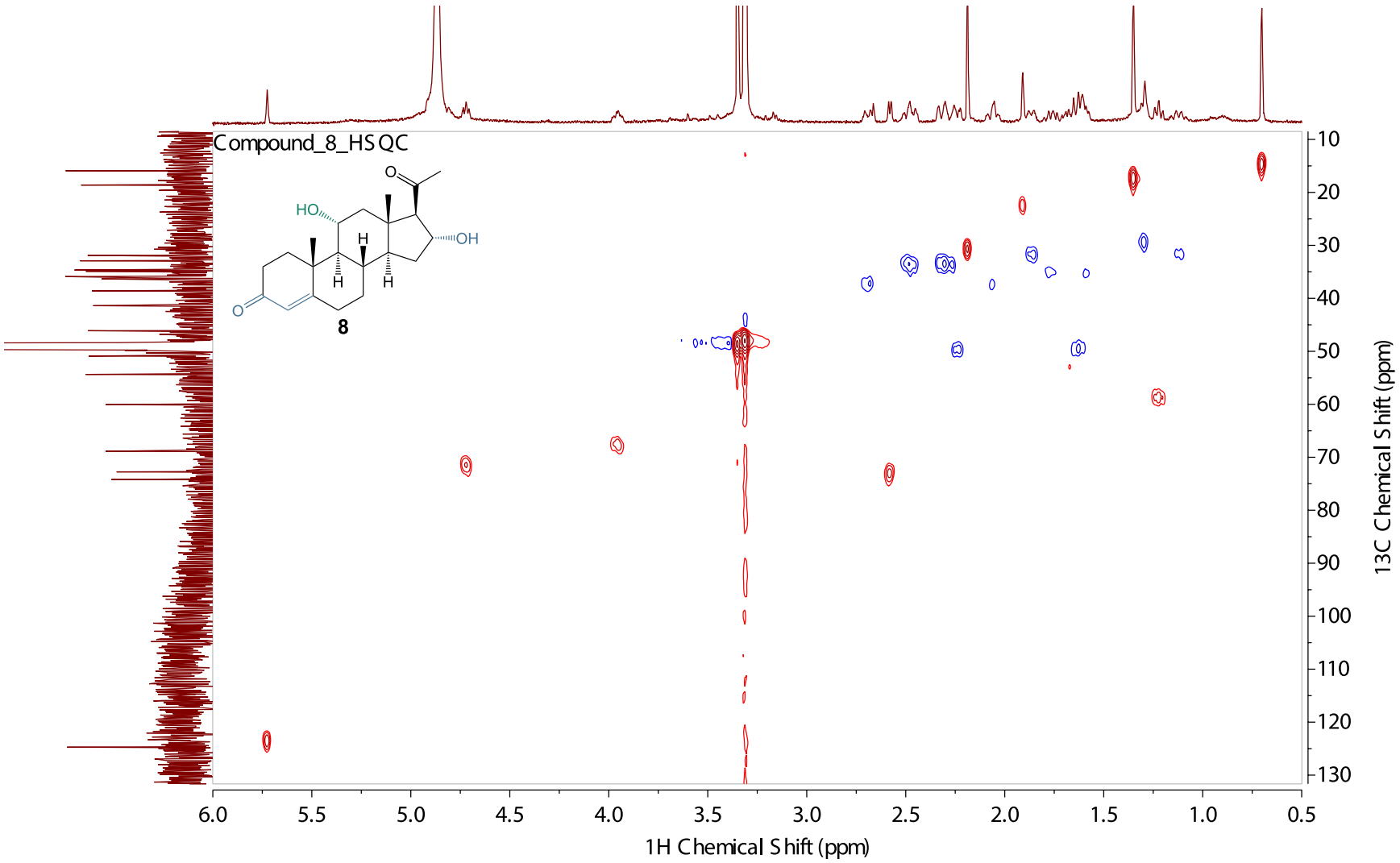


Figure S20. ^{1}H-^{13}C HSQC spectrum of 11α,16α-dihydroxyprogesterone (compound **8**) (CD_3OD, 500 MHz).

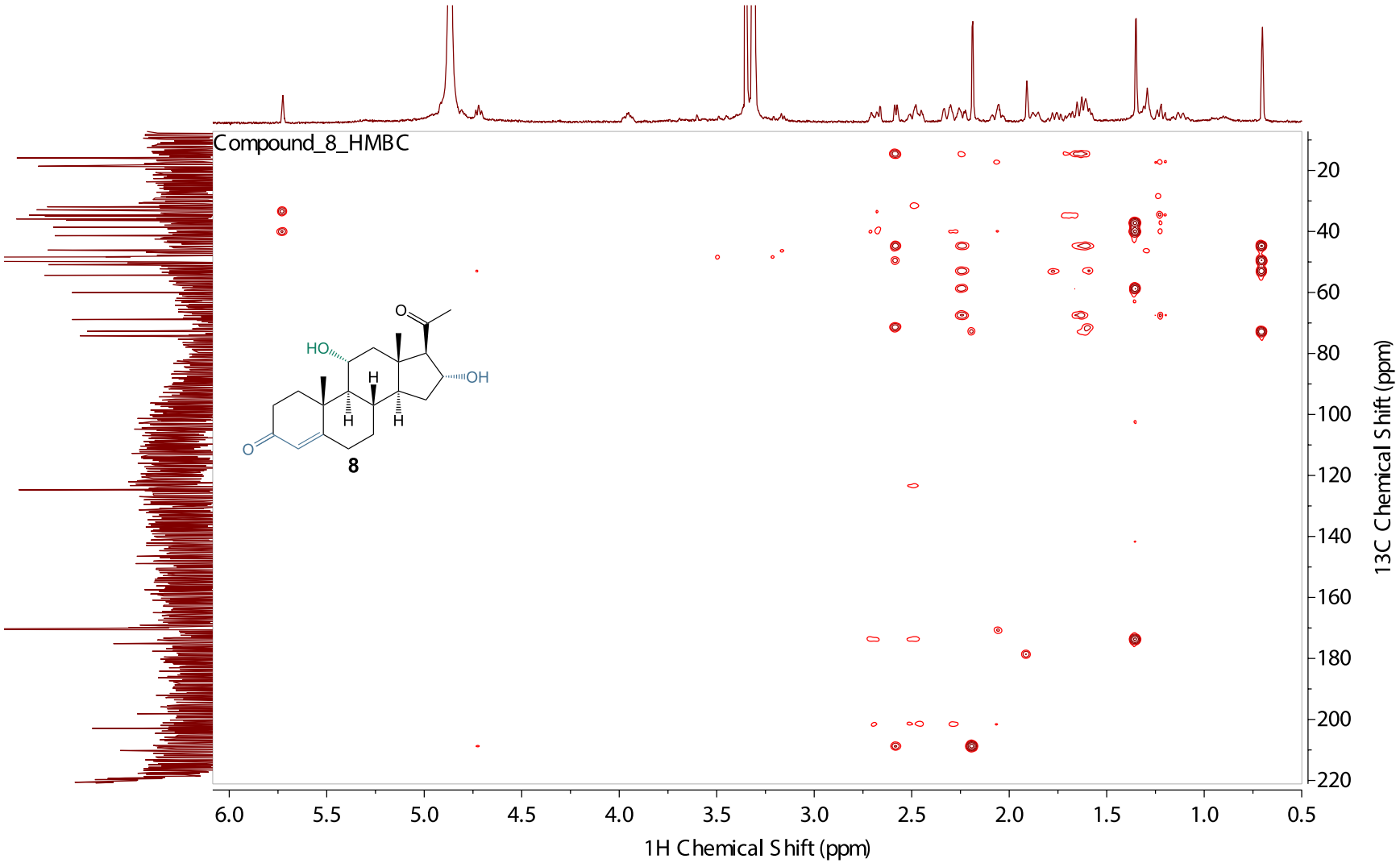


Figure S21. ^{1}H-^{13}C HMBC spectrum of 11α,16α-dihydroxyprogesterone (compound **8**) (CD_3OD, 500 MHz).

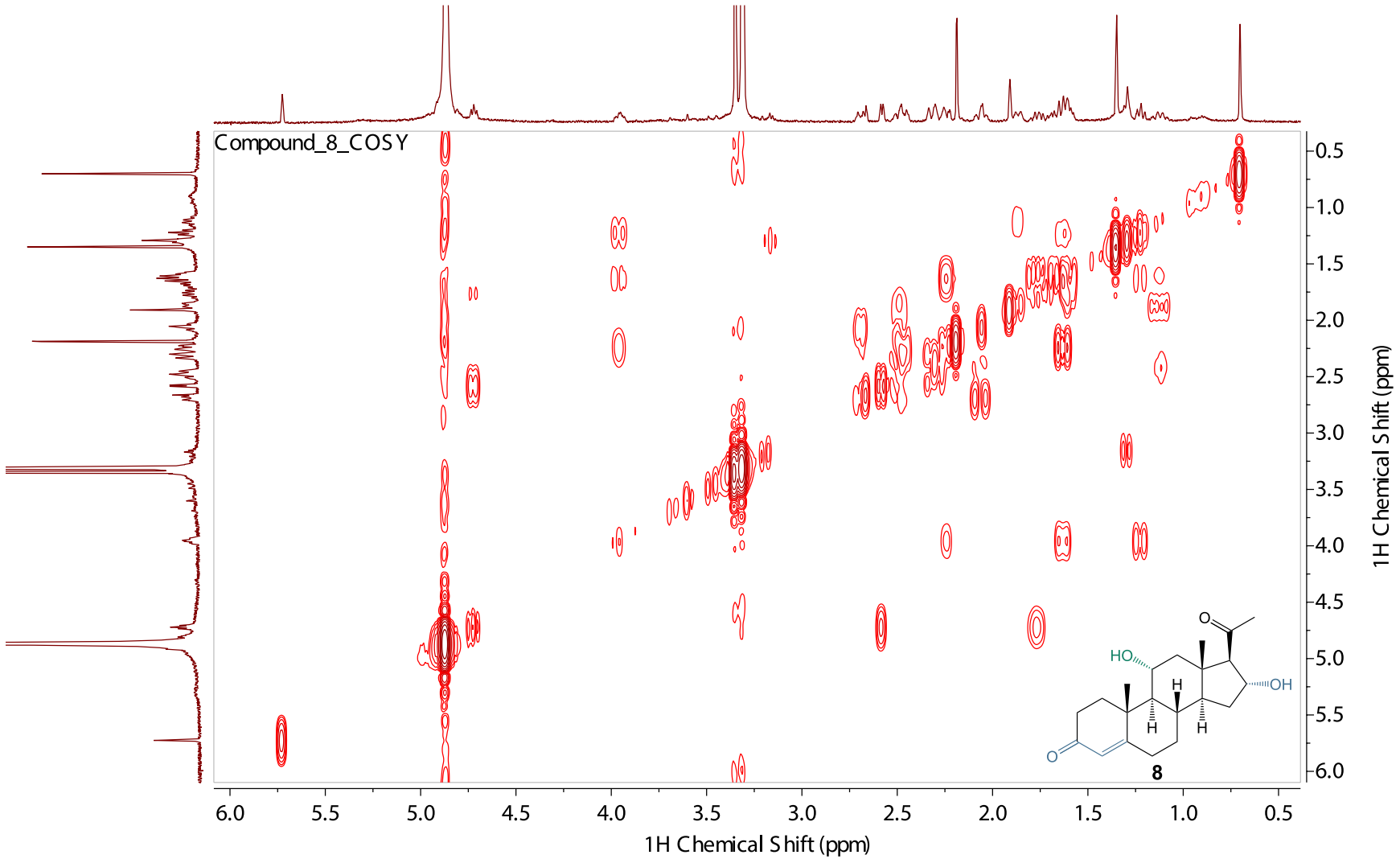


Figure S22. ^{1}H-^{1}H COSY spectrum of 11α,16α-dihydroxyprogesterone (compound **8**) (CD_3OD, 500 MHz).

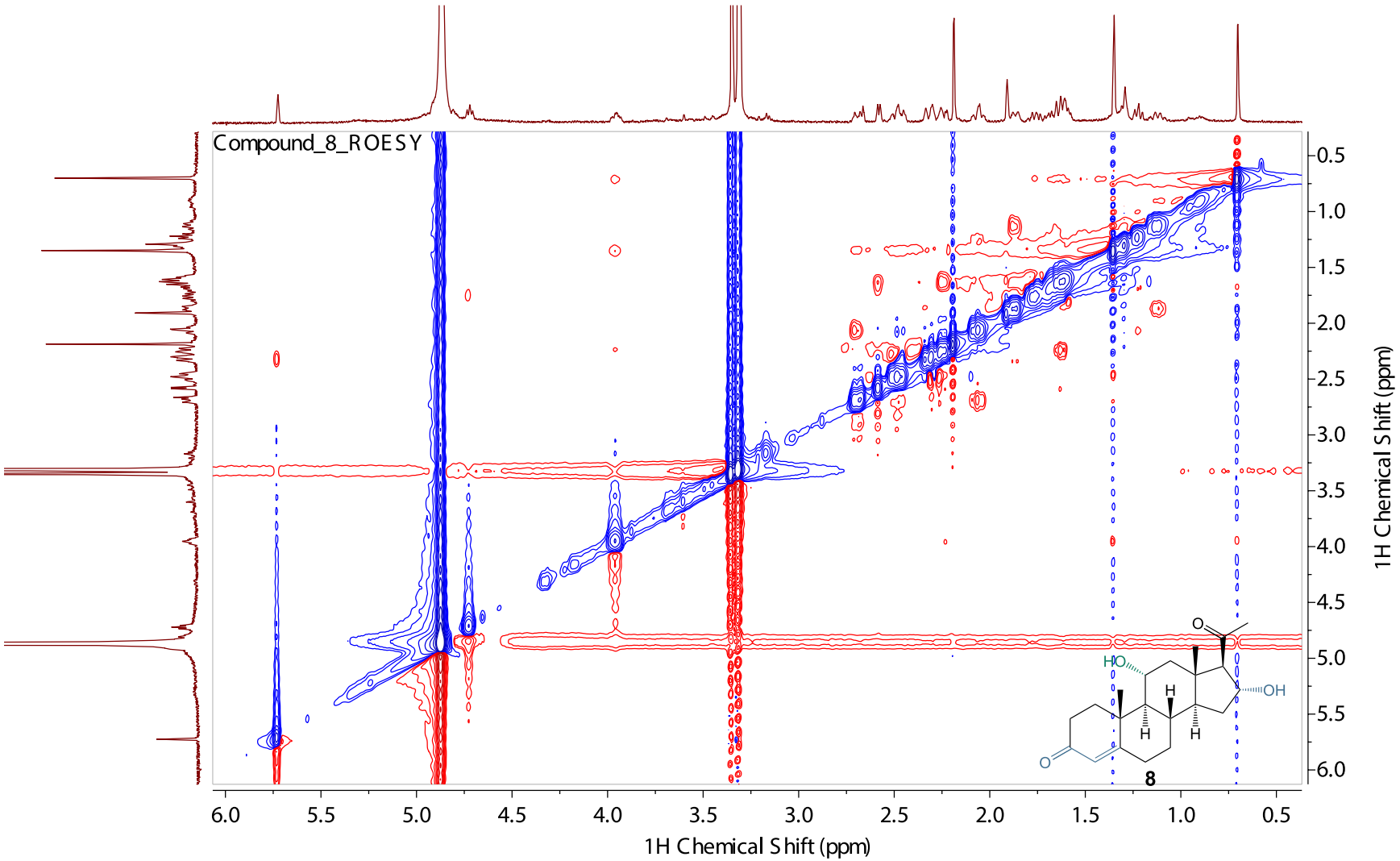


Figure S23. ^{1}H-^{1}H ROESY spectrum of 11α,16α-dihydroxyprogesterone (compound **8**) (CD_3OD, 500 MHz).

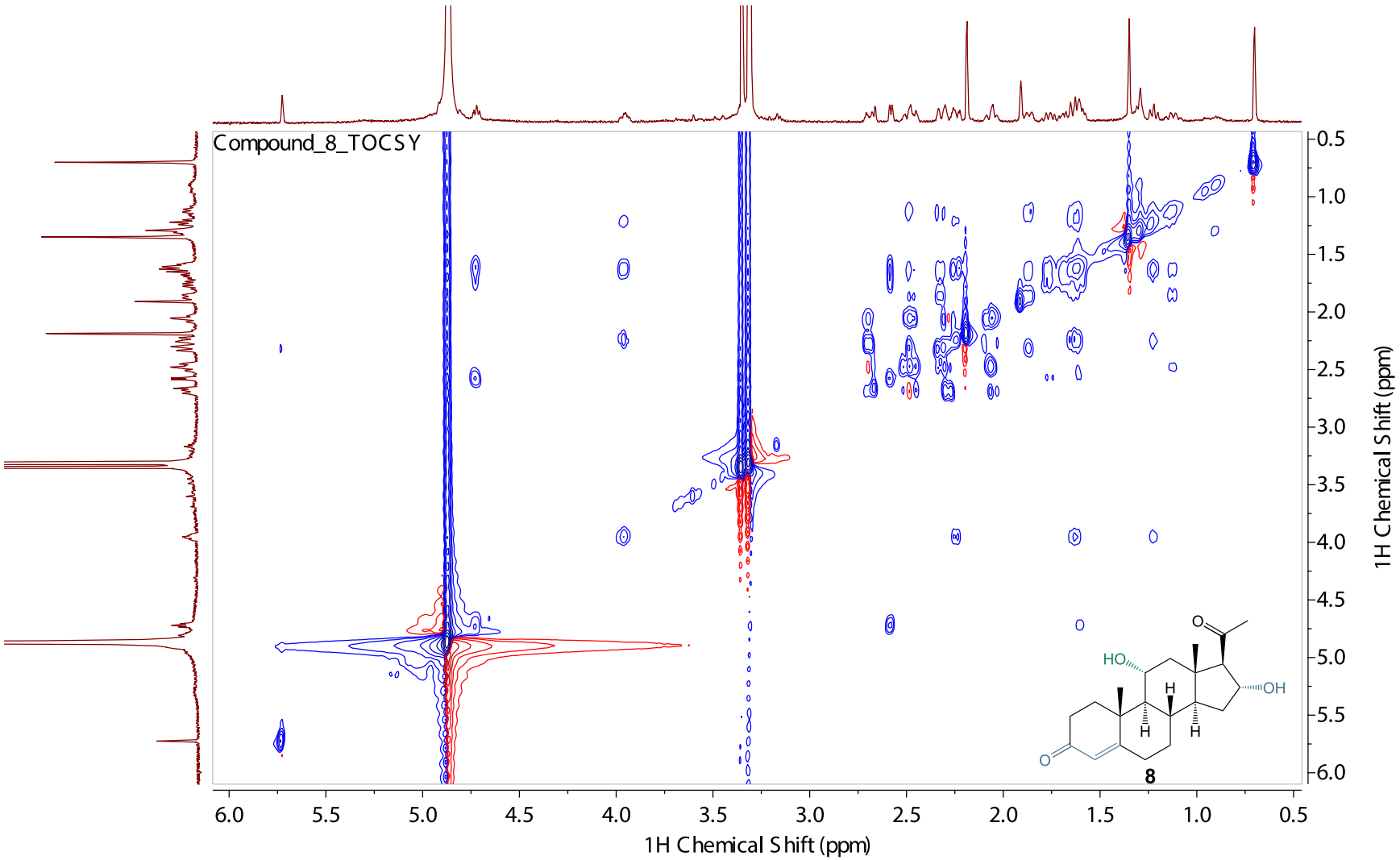


Figure S24. ^{1}H-^{1}H TOCSY spectrum of 11α,16α-dihydroxyprogesterone (compound **8**) (CD_3OD, 500 MHz).

Supplemental File Index

Three items accompany this thesis in the file "Spady_2019_Supplemental.zip".

"CYP7B1_Model_Components": Directory containing model structures of apo-enzyme CYP7B1, heme, and substrate structures used in Chapter 3. Files are in .pdb format.

"CYP7B1_Substrate_Binding_Modes": Directory containing the four final CYP7B1 binding modes for steroid substrates **1**, **3**, and **7**, as discussed in Chapter 3. Files are in .pdb format.

"SteroidProteinPKModel.R": Code for generating a pharmacokinetic model of a steroid-binding antibody fusion protein, as discussed in Chapter 4. The file is an R script that requires the dMod package.