

## Structure of human steroid 5 $\alpha$ -reductase 2 with anti-androgen drug finasteride

Qingpin Xiao<sup>1,5</sup>, Lei Wang<sup>2</sup>, Shreyas Supekar<sup>3</sup>, Tao Shen<sup>4</sup>, Heng Liu<sup>2</sup>, Fei Ye<sup>4</sup>, Junzhou Huang<sup>4</sup>,  
Hao Fan<sup>3</sup>, Zhiyi Wei<sup>1\*</sup>, Cheng Zhang<sup>2\*</sup>

<sup>1</sup>Department of Biology, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China

<sup>2</sup>Department of Pharmacology and Chemical Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA15261, USA

<sup>3</sup>Bioinformatics Institute (BII), Agency for Science, Technology and Research (A\*STAR), Singapore 138671, Singapore

<sup>4</sup>Tencent AI Lab, Shenzhen, Guangdong 518000, China

<sup>5</sup>Faculty of Health Sciences, University of Macau, Macau SAR 999078, China.

\*Correspondence: Dr. Zhiyi Wei, [weizy@sustech.edu.cn](mailto:weizy@sustech.edu.cn) or Dr. Cheng Zhang, [chengzh@pitt.edu](mailto:chengzh@pitt.edu).

## Abstract

Human steroid 5 $\alpha$ -reductase 2 (SRD5 $\alpha$ 2) as a critical integral membrane enzyme in steroid metabolism catalyzes testosterone to dihydrotestosterone. Mutations on its gene have been linked to 5 $\alpha$ -reductase deficiency and prostate cancer. Finasteride and dutasteride as SRD5 $\alpha$ 2 inhibitors are widely used anti-androgen drugs for benign prostate hyperplasia, which have recently been indicated in the treatment of COVID-19. The molecular mechanisms underlying enzyme catalysis and inhibition remained elusive for SRD5 $\alpha$ 2 and other eukaryotic integral membrane steroid reductases due to a lack of structural information. Here, we report a crystal structure of human SRD5 $\alpha$ 2 at 2.8 Å revealing a unique 7-TM structural topology and an intermediate adduct of finasteride and NADPH as NADP-dihydrofinasteride in a largely enclosed binding cavity inside the membrane. Structural analysis together with computational and mutagenesis studies reveals molecular mechanisms for the 5 $\alpha$ -reduction of testosterone and the finasteride inhibition involving residues E57 and Y91. Molecular dynamics simulation results indicate high conformational dynamics of the cytosolic region regulating the NADPH/NADP<sup>+</sup> exchange. Mapping disease-causing mutations of SRD5 $\alpha$ 2 to our structure suggests molecular mechanisms for their pathological effects. Our results offer critical structural insights into the function of integral membrane steroid reductases and will facilitate drug development.

## Main

The membrane-embedded 5 $\alpha$ -reductase family in the human includes five members, SRD5 $\alpha$ 1-3 and much less characterized glycoprotein synaptic 2 (GSPN2) and GSPN2-like<sup>1</sup>. They mainly catalyze the irreversible reduction of the  $\Delta^{4,5}$  bond in  $\Delta^4$ -3-ketosteroids using reduced nicotinamide adenine dinucleotide phosphate (NADPH) as the hydride donor cofactor, though other lipid substrates have also been identified<sup>1-3</sup>. SRD5 $\alpha$ s are expressed differently in human body to play diverse functional roles despite their sequence similarity (**Extended Data Fig. 1**). SRD5 $\alpha$ 1 and SRD5 $\alpha$ 3 have been indicated to function in the metabolism of neurosteroids<sup>4,5</sup> and the protein *N*-linked glycosylation<sup>2,6</sup>, respectively. SRD5 $\alpha$ 2 is the most intensively investigated SRD5 $\alpha$  with well-characterized roles in androgen metabolism and androgen-related disorders<sup>1,7</sup>. All three SRD5 $\alpha$ s are located in the membrane of endoplasmic reticulum (ER) in the cells<sup>8</sup>.

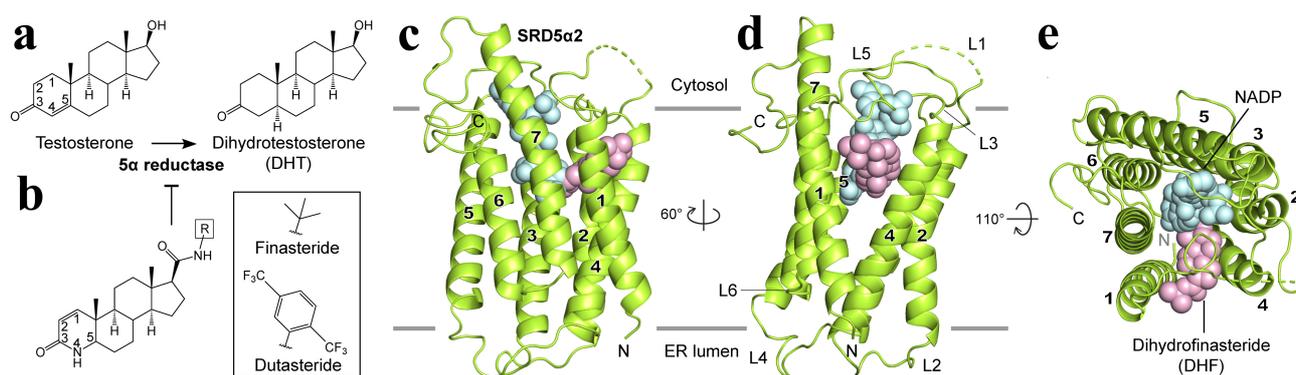
SRD5 $\alpha$ 2 is highly expressed in male reproductive systems<sup>7</sup> to convert testosterone to 5 $\alpha$ -dihydrotestosterone (DHT) (**Fig. 1a**), the major steroid hormone for androgen receptor<sup>1</sup>. A large number of mutations identified in the *SRD5A2* gene can result in inefficient levels of DHT, leading to an autosomal recessive disorder named 5 $\alpha$ -reductase deficiency associated with underdeveloped and atypical genitalia<sup>9-11</sup>. On the other hand, overproduction of DHT by SRD5 $\alpha$ 2 is associated with benign prostatic hyperplasia (BPH), androgenic alopecia and prostate cancer due to excessive androgen receptor signaling<sup>7,12</sup>. 5 $\alpha$ -reductase inhibitors (5ARIs) including finasteride and dutasteride (**Fig. 1b**), which mainly target SRD5 $\alpha$ 2 but also act on other SRD5 $\alpha$ s<sup>13</sup>, have been used as a major anti-androgenic class of drugs to treat BPH and androgenic alopecia<sup>1,7,12,14</sup> and indicated in the treatment of prostate cancer<sup>15</sup>. In particular, finasteride is among the top-100 most-prescribed drugs in the United States, which is associated with an irreversible action on SRD5 $\alpha$ 2<sup>16,17</sup>. Interestingly, androgen receptor signaling has recently been linked to COVID-19 disease severity, explaining why males are more prone to severe COVID-19 symptoms<sup>18</sup>. The 5ARI drugs have thus been suggested to hold the repurposing potential for treating COVID-19<sup>18,19</sup>.

SRD5 $\alpha$ s belong to a large group of eukaryotic membrane-embedded steroid reductases, which also include sterol reductases such as the 7-dehydrocholesterol reductase (DHCR7) that catalyzes the last step in cholesterol biosynthesis in humans<sup>20</sup>. Although these steroid/sterol reductases share very little sequence similarity, they all use NADPH as the cofactor to reduce specific carbon-carbon double bonds in their steroid substrates. So far, only one crystal structure of a bacterial membrane-embedded sterol reductase MaSR1 without any steroid substrate was reported for this group of reductases<sup>21</sup>. To further understand the molecular mechanisms underlying the function of eukaryotic steroid reductases and in particular the catalytic mechanism of SRD5 $\alpha$ s and the action of 5ARI drugs, we solved a crystal structure of human SRD5 $\alpha$ 2 in the presence of NADPH and finasteride. The structure revealed a topology of 7 transmembrane  $\alpha$ -helices (7 TMs), rather than the 10-TM topology of MaSR1, and an NADP-dihydrofinasteride intermediate adduct. The structure together with computational studies provided unprecedented molecular insights into the catalytic mechanism of SRD5 $\alpha$ 2, the irreversible action of finasteride on SRD5 $\alpha$ 2, and the molecular mechanisms underlying the pathological effects of disease-associated mutations.

## Results

## Structure determination and overall structure of human SRD5 $\alpha$ 2

Human SRD5 $\alpha$ 2 was expressed in insect Sf9 cells and purified in the presence of finasteride and crystallized in lipidic mesophase with a space group of *P622* (Extended Data Fig. 2)<sup>22</sup>. The structure was determined to 2.8-Å resolution by molecular replacement<sup>23</sup> based on a structural model from *de novo* prediction since we failed to use anomalous diffraction data to solve the structure (see Methods). Clear electron density allowed modeling of all 254 residues of SRD5 $\alpha$ 2 except for the first 4 residues and residues S39-A43 in a flexible loop region (Fig. 1c, d and Table 1). A dualsteric ligand was modeled as an adduct of finasteride and NADPH in the structure (Fig. 1e), which will be discussed later.

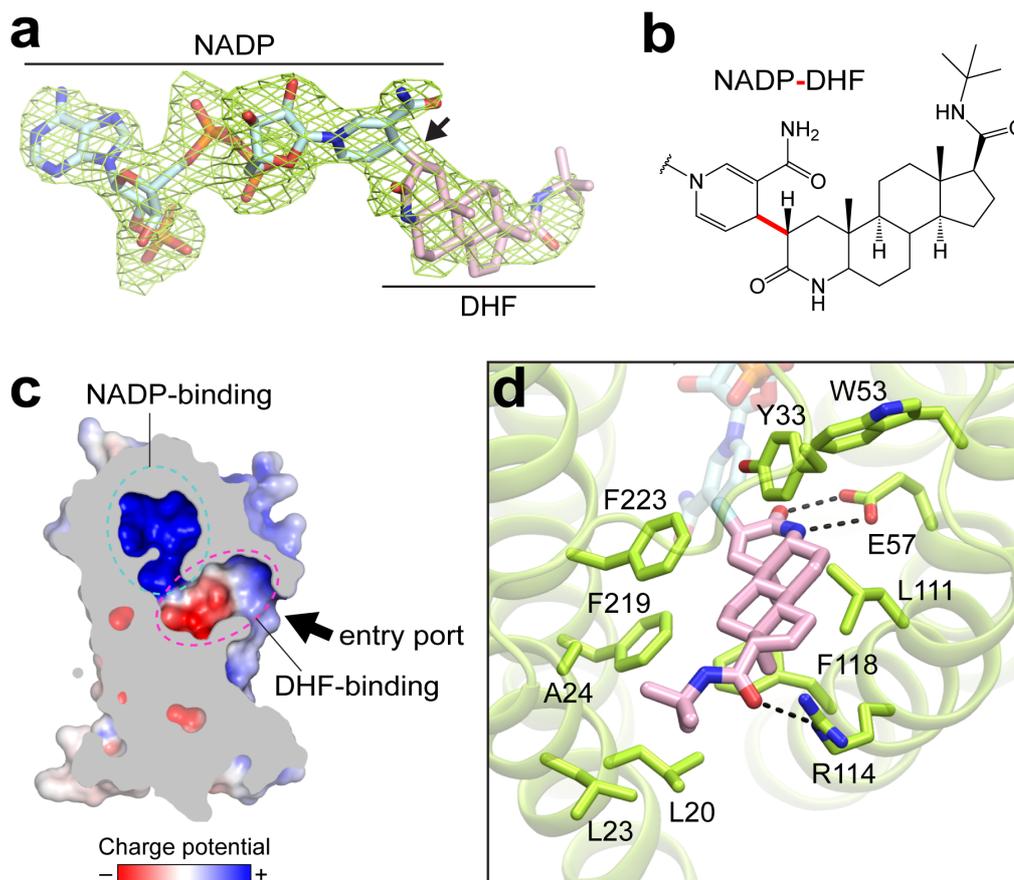


**Fig. 1: Overall structure of human SRD5 $\alpha$ 2.** (a) 5 $\alpha$ -reduction reaction of the  $\Delta^{4,5}$  double bond of testosterone catalyzed by SRD5 $\alpha$ 2 to generate dihydrotestosterone (DHT). (b) SRD5 $\alpha$ 2 inhibition by finasteride and dutasteride. The two inhibitors share the same ring structure with different R-groups connected in the tails. (c)-(e) Three views of the SRD5 $\alpha$ 2 structure. The NADP-DHF adduct was shown as spheres. L1-6 represent 6 loops connecting 7-TMs. The NADP and DHF moieties were colored in light cyan and light pink, respectively.

Unlike the bacterial sterol reductase MaSR1<sup>21</sup>, the structure of SRD5 $\alpha$ 2 contains 7 TMs (TM1-7) connected by 6 loops (L1-6) (Fig. 1c, d and Extended Data Fig. 3a). We assigned the carboxyl terminal (C-terminal) loop (C-loop) to face the cytosol and the amino terminal loop (N-loop) to face ER lumen according to the enriched positively charged residues at the C-terminal side<sup>24</sup> (Extended Data Fig. 3b). In addition, the N-terminal residue C5<sup>N</sup> (superscripts indicate location of the residues hereafter) forms a disulfide bond with C133<sup>L4</sup> in loop 4 (L4) (Extended Data Fig. 3a), suggesting an ER luminal location of the N-terminal side because of the reducing environment of the cytosol. The 7-TM topology is more commonly associated with G protein-coupled receptors (GPCRs)<sup>25</sup>, although the arrangement of TMs in SRD5 $\alpha$ 2 is distinct from that of GPCRs (Extended Data Fig. 3c).

### Intermediate adduct formed between finasteride and NADPH

The structure of SRD5 $\alpha$ 2 revealed a large cavity inside the 7-TM domain at the cytosolic side formed by all 7 TMs and L1, L3 and L5 (Fig. 1c, d). The cavity is completely occluded from the cytosol with only one opening on the side of 7-TM domain between TM1 and TM4. Clear electron density in the cavity revealed features of NADPH and finasteride, which allowed an unambiguous modeling of both ligands (Fig. 2a). Strikingly, after ligand fitting, the distance between the nicotinamide C-4 atom of NADPH and the C-2 atom of finasteride is shorter than 2Å, suggesting the formation of a covalent bond (Fig. 2a, b).



**Fig. 2: Formation and binding environment of NADP-DHF adduct.** (a)  $F_o-F_c$  electron omit map of the NADP-DHF adduct contoured at  $3\sigma$ . (b) Chemical structure of the NADP-DHF adduct. The covalent bond connecting the nicotinamide C-4 atom of NADPH and the C-2 atom of finasteride was highlighted in red. (c) Enclosed binding cavity for NADP-DHF with charge potentials. The potential entry port for the steroid substrates was indicated by an arrow. (d) Molecular details of the DHF binding pocket. Hydrogen bonds were indicated by dashed lines.

Strongly supporting our structural findings, a previous enzymological and mass spectrometric study on the mechanism of inhibition of SRD5 $\alpha$ 2 by finasteride identified an intermediate adduct as NADP-dihydrofinasteride (NADP-DHF)<sup>17</sup>. The same study indicated that SRD5 $\alpha$ 2 could catalyze the hydride transfer from NADPH to finasteride, leading to the formation of a covalent bond between the nicotinamide C-4 atom of NADPH and the C-2 atom of finasteride<sup>17</sup>, highly consistent with our structural observation. It is likely that SRD5 $\alpha$ 2 took endogenous NADPH during recombinant protein expression to catalyze the reaction with finasteride supplemented in protein buffers to generate the NADP-DHF intermediate, which was stable enough to be captured in our SRD5 $\alpha$ 2 crystals. In fact, NADP-DHF has been suggested to be one of the most potent non-covalent enzyme inhibitors in general with a very slow dissociation rate, explaining the irreversible action of finasteride on SRD5 $\alpha$ 2<sup>13</sup>. Therefore, we modeled NADP-DHF as the real ligand in our structure.

### Binding pockets for NADP-DHF

The substrate-binding cavity shows two relatively separate tunnel-like pockets for NADP and DHF (Fig. 2c). NADP adopts an extended anti-conformation to insert into the binding pocket with a positively charged environment inside the 7-TM bundle (Fig. 2c). Surprisingly, the binding pocket for

NADP is enclosed at the cytosolic side by the cytosolic loops, completely shielding NADP from the cytosol (**Fig. 1e**). The nicotinamide-ribose moiety is buried inside 7-TM and stabilized by extensive polar and hydrophobic interactions with residues mainly from TMs 2, 4-7, while the diphosphate moiety of NADP mainly forms polar interactions with L1 and TM6-7 (**Extended Data Fig. 4a-e**). The adenine-ribose phosphate moiety of NADP forms hydrogen bonds and salt bridges with residues mostly from three cytosolic loops (**Extended Data Fig. 4a-e**). Supporting such a cofactor-binding mode, mutations of residues that interact with NADP have been shown to either decrease or abolish the catalytic activity of SRD5 $\alpha$ 2 by previous enzymological studies (**Extended Data Fig. 4d**)<sup>26-28</sup>.

In contrast to the highly polar environment for NADP, the binding pocket of DHF is largely hydrophobic, where the core ring structure of DHF interacts with hydrophobic residues from TMs 1, 2, 4 and 7 (**Fig. 2d and Extended Data Fig. 4e**). A previous study showed that substitution of one aromatic residue F118<sup>TM4</sup> in the DHF-binding pocket to a leucine could dramatically decrease the SRD5 $\alpha$ 2 activity by disrupting the binding of testosterone<sup>29</sup>, suggesting an important role of this residue in the binding of steroid substrates. The polar groups located at each end of DHF engage in additional polar interactions with residues E57<sup>TM2</sup> and R114<sup>TM4</sup> of SRD5 $\alpha$ 2 (**Fig. 2d**).

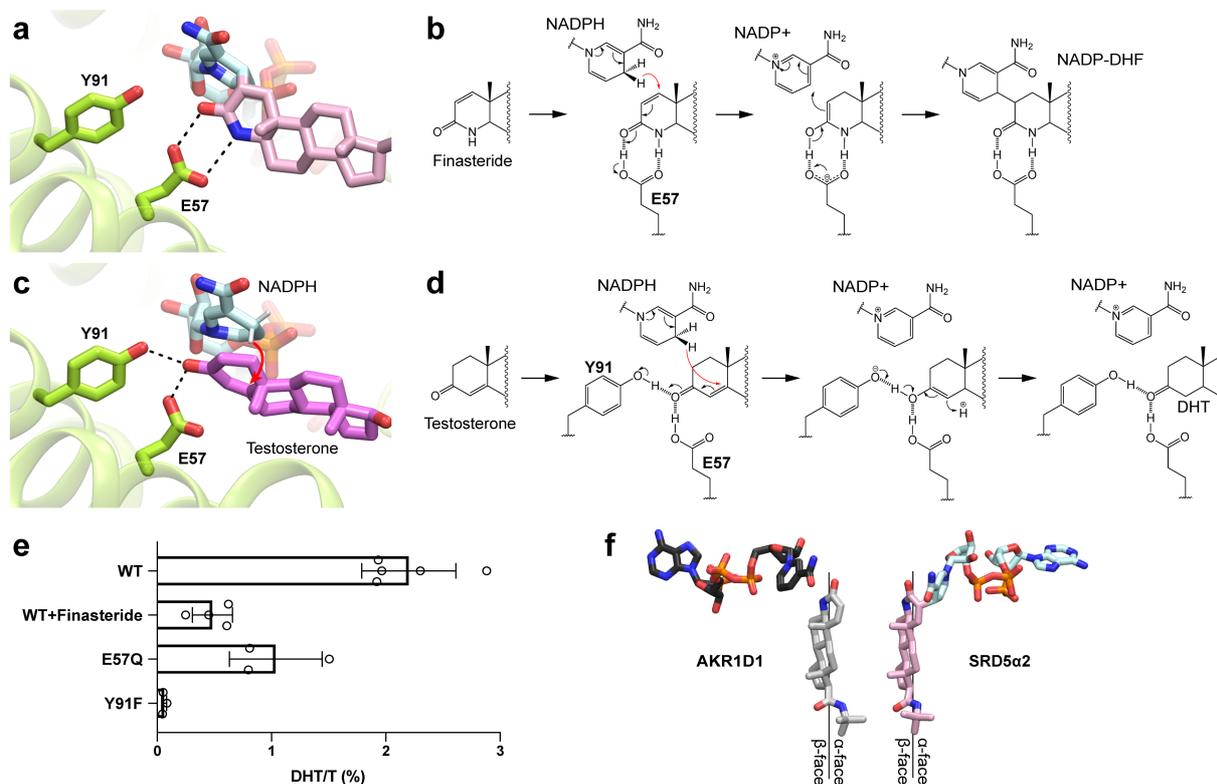
Sequence alignment analysis of SRD5 $\alpha$ s across different species including a plant homologue DET2 involved in the synthesis of phytohormones<sup>30</sup> indicated that while the residues involved in the binding of NADP are highly conserved, the residues in the DHF-binding pocket are much less conserved (**Extended Data Fig. 1**). This suggests that although all the SRD5 $\alpha$  family members use NADPH as the cofactor to reduce their substrates, they have evolved specific structural features to recognize different steroid or lipid substrates<sup>1,7</sup>. We further mapped the residue conservation to the SRD5 $\alpha$ 2 structure, which showed that the region around the tert-butylacetamide tail group of DHF is the least conserved part of the ligand-binding pocket (**Extended Data Fig. 4f**), explaining the selectivity of finasteride for human SRD5 $\alpha$ 1 over SRD5 $\alpha$ 2<sup>1,14</sup>.

### Potential mechanisms of SRD5 $\alpha$ 2 catalysis and inhibition

It has been proposed that two unknown residues in SRD5 $\alpha$ 2 point toward the C-3 carbonyl group in finasteride to facilitate the hydride transfer from NADPH to finasteride to form an enolized intermediate, which is followed by the formation of a covalent bond between finasteride and NADP<sup>17</sup>. In our structure, both the C-3 carbonyl and the N-4 amine groups of DHF form hydrogen bonds with E57<sup>TM2</sup> (**Fig. 3a**). We propose that through hydrogen-bonding interactions with E57<sup>TM2</sup>, finasteride is positioned in a way that the 4-pro-(*R*)-hydride of NADPH is in the proximity of the C-2 atom of finasteride to allow the hydride transfer to the  $\Delta^{1,2}$  bond in finasteride. Due to the presence of the N-4 amine group, the enolization of finasteride as the result of hydride transfer involves the C-3 carbonyl and the C-2 group, leading to the covalent bond formation between finasteride and NADP (**Fig. 3b**). Dutasteride contains the same core ring structure as finasteride (**Fig. 1b**) and therefore likely forms a similar adduct with NADPH. Indeed, dutasteride has been suggested to share the same irreversible inhibition mechanism as finasteride<sup>31</sup>.

Interestingly, finasteride also inhibits the activity of steroid 5 $\beta$ -reductase, a soluble steroid reductase that belongs to the NADPH-dependent aldo-keto reductase (AKR) superfamily<sup>32</sup>. Steroid 5 $\beta$ -reductase, also named AKR1D1, can reduce the  $\Delta^{4,5}$  bond in testosterone as SRD5 $\alpha$ 2 but to generate a

stereochemically different product, 5 $\beta$ -DHT<sup>3</sup>. A crystal structure of AKR1D1 with finasteride<sup>33</sup> showed that the relative position of finasteride to NADP<sup>+</sup> in AKR1D1 is opposite to that of DHF to NADP in SRD5 $\alpha$ 2 (**Extended Data Fig. 4g**). As a result, the 4-pro-(*R*)-hydride of NADPH is adjacent to the N-4 group instead of the C-2 group of finasteride (**Extended Data Fig. 4g**), thereby preventing hydride transfer<sup>33</sup>. Hence, the  $\Delta^{1,2}$  bond of finasteride cannot be reduced by AKR1D1, which accounts for the competitive and reversible action of finasteride on AKR1D1<sup>33</sup>.



**Fig. 3: Mechanisms for SRD5 $\alpha$ 2 catalysis and inhibition. (a)** Binding pose of DHF. Y91<sup>TM3</sup> is not directly hydrogen bonding to DHF. **(b)** Potential mechanism for finasteride inhibition and the covalent adduct formation between NADPH and finasteride. E57<sup>TM2</sup> facilitates the hydride transfer to the  $\Delta^{1,2}$  bond of finasteride, leading to the formation of a covalent bond. **(c)** Binding pose of testosterone based on our docking results. E57<sup>TM2</sup> and Y91<sup>TM3</sup> each forms a hydrogen bond with the substrate. **(d)** Potential mechanism the 5 $\alpha$ -reduction of testosterone. E57<sup>TM2</sup> and Y91<sup>TM3</sup> facilitate the hydride transfer to the  $\Delta^{4,5}$  bond of testosterone, leading to the formation of DHT. Hydrogen bonds are shown as dashed lines and the hydride transfer is shown as red curved arrows. **(e)** Catalysis of testosterone (T) to dihydrotestosterone (DHT) by wild type SRD5 $\alpha$ 2 (WT), WT with 500  $\mu$ M finasteride and two SRD5 $\alpha$ 2 mutants E57Q and Y91F. The ratios of DHT to T (DHT/T) were determined by mass spectrometry. All data are presented as mean  $\pm$  SEM of 3-5 independent experiments. **(f)** Distinct orientations of finasteride relative to NADPH in AKR1D1 and SRD5 $\alpha$ 2. The finasteride and NADPH conformations in the AKR1D1 structure (PDB ID 3G1R) were shown by align the core ring of finasteride to that of DHF in the SRD5 $\alpha$ 2 structure.

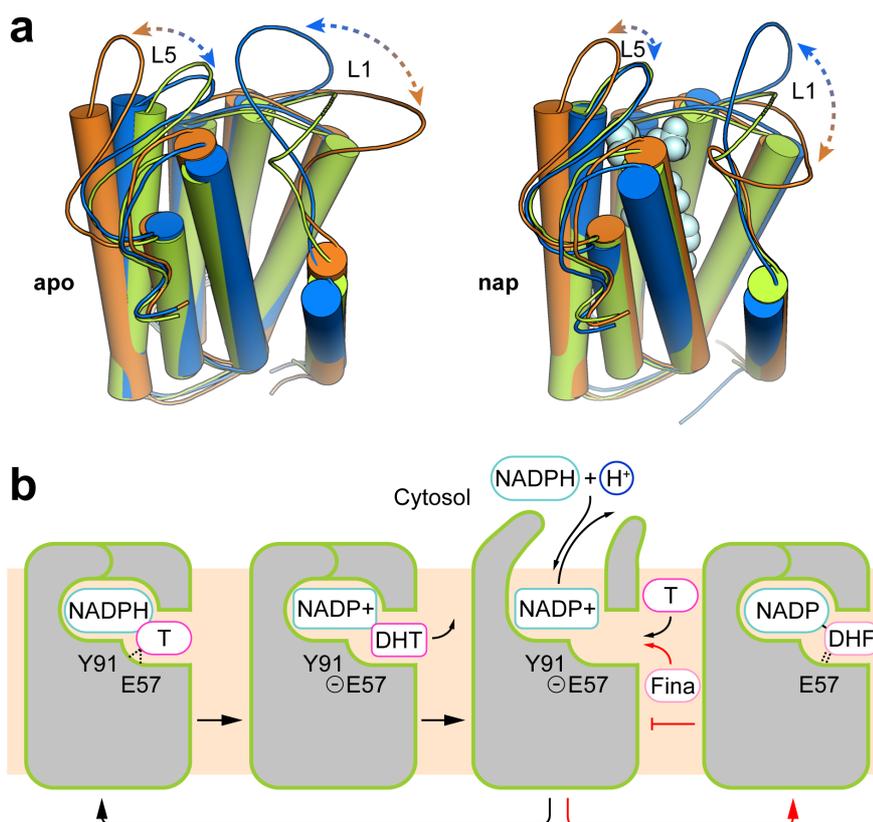
To investigate the binding pose of testosterone and the catalytic mechanism of SRD5 $\alpha$ 2, we docked NADPH and testosterone to our structure *in silico*. In the docked structure, while the NADPH molecule could be well aligned to the NADP moiety of NADP-DHF in the crystal structure, testosterone is

positioned more deeply into the cavity compared to DHF with its ring structure stacking parallelly to the nicotinamide ring of NADPH (**Fig. 3c**). This is likely due to the absence of the hydrogen bond between E57<sup>TM2</sup> and the N-4 group of DHF. As a result, the C-3 carbonyl group of testosterone forms hydrogen bonds with both E57<sup>TM2</sup> and Y91<sup>TM3</sup> and the 4-pro-(*R*)-hydride of NADPH is close to the  $\Delta^{4,5}$  bond of testosterone (~2.5 Å) (**Fig. 3c**). Previous structural studies on the soluble steroid reductase AKR1D1 have shown that the steroid C-3 carbonyl formed hydrogen bonds with residues E120 and Y58. In ARK1D1, Y58 has been suggested to function as the general acid-base catalytic group to polarize the steroid C-3 carbonyl group together with E120 to facilitate hydride transfer and substrate enolization<sup>34-36</sup>. We propose that SRD5 $\alpha$ 2 employs a similar catalytic mechanism, in which residues E57<sup>TM2</sup> and Y91<sup>TM3</sup> polarize the C-3 carbonyl of testosterone by hydrogen bonding to facilitate the hydride transfer from NADPH to the C-5 atom of testosterone, leading to the formation of an enolized intermediate followed by reduction of the  $\Delta^{4,5}$  bond in testosterone (**Fig. 3d**). Consistently, the optimum pH for the SRD5 $\alpha$ 2 catalytic activity is acidic (pH~5), which favors the protonation of E57<sup>TM2</sup> in order to form a hydrogen bond with the C-3 carbonyl of testosterone<sup>26</sup>. Our mutagenesis studies showed that substitution of E57<sup>TM2</sup> to a less acidic glutamine residue could compromise enzyme activity (**Fig. 3e and Extended Data Fig. 4h**). Likewise, the Y91F mutation, which eliminates the potential hydrogen bonding of Y91<sup>TM3</sup> to testosterone, essentially abolished the conversion of testosterone to DHT by SRD5 $\alpha$ 2 in our experiments (**Fig. 3e and Extended Data Fig. 4h**), supporting our proposed catalytic mechanism (**Fig. 3d**).

Despite sharing a potentially similar catalytic mechanism, the relative orientations of the steroid substrates to NADPH are distinct between SRD5 $\alpha$ 2 and AKR1D1. In our crystal structure, the nicotinamide ring of NADP is oriented toward the  $\alpha$ -face of DHF and residues E57<sup>TM2</sup> and Y91<sup>TM3</sup> are located on the other side of the core ring structure (**Fig. 3a, c**). In contrast, in AKR1D1, the nicotinamide ring of NADP<sup>+</sup> is oriented toward the  $\beta$ -face of steroid substrates (**Fig. 3f**)<sup>33-35</sup>. Such a steroid-binding mode in SRD5 $\alpha$ 2 suggests that the hydride is transferred from NADPH to the C-5 atom of testosterone at the  $\alpha$ -face, leading to the 5 $\alpha$ -stereochemistry of DHT generated by SRD5 $\alpha$ s.

### Structural dynamics of SRD5 $\alpha$ 2 for catalysis

The binding pocket of NADP-DHF only opens on the side of 7-TM, allowing the access of steroid substrates to SRD5 $\alpha$ 2 from the lipid bilayer (**Fig. 2c**). The cytosolic loops L1, L3 and L5 pack against each other to fully enclose the binding pocket for NADP (**Fig. 1d, e**), contrasting the highly exposed NADP<sup>+</sup>/NADPH-binding pockets in soluble AKRs and MaSR1<sup>21,37</sup>. All cytosolic loops are involved in the interactions with the adenine-ribose moiety of NADP. Such a conformation is compatible with the very tight binding of NADP-DHF to the enzyme<sup>13</sup> while imposing a physical barrier for the NADPH/ NADP<sup>+</sup> exchange during the reaction. It is unlikely that the nucleotides either enter or exit from the enzyme through the opening between TM1 and TM4 on the side of 7-TM considering the surrounding lipid environment and the highly polar nature of NADP<sup>+</sup>/NADPH (**Fig. 2c**). To lift the barrier, the cytosolic loops in SRD5 $\alpha$ 2 may undergo conformational changes during the reaction so that the cytosolic region can open up to expose the nucleotide-binding pocket to the cytosol before and after one reaction to allow the NADP<sup>+</sup>/NADPH exchange and thus efficient turnover of the reactions.

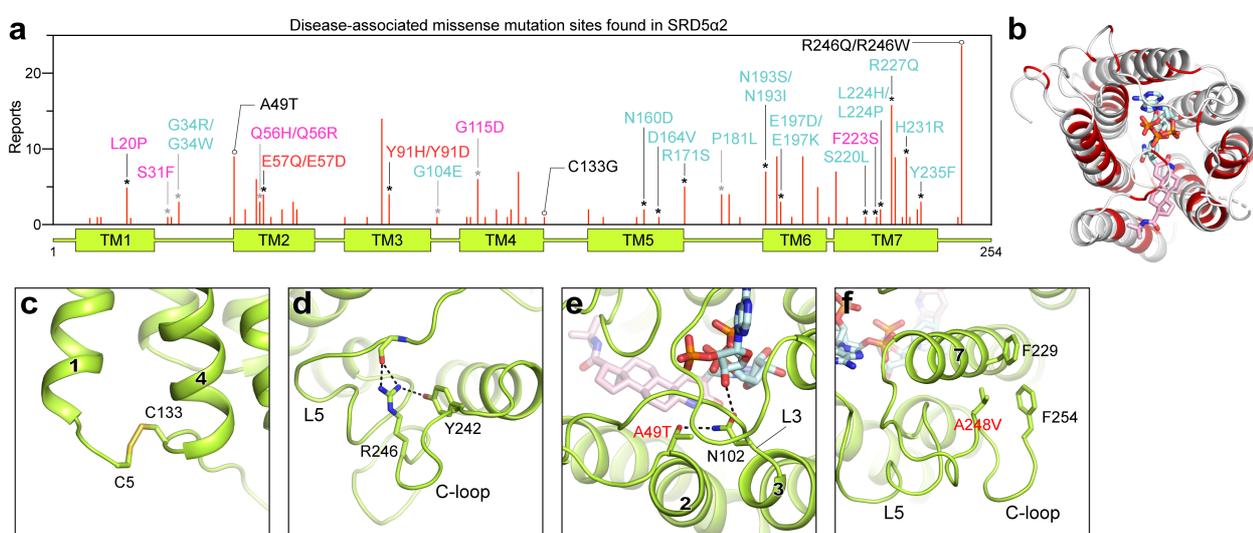


**Fig. 4: Dynamics of SRD5 $\alpha$ 2 during reaction.** (a) Dominant conformational motions in the SRD5 $\alpha$ 2 MD simulations using principal component analysis (PCA). To visualize the motions, two extreme conformations from the principal components for the apo and nap states are indicated in orange and blue respectively. The SRD5 $\alpha$ 2 crystal structure colored in green is superimposed as a reference. PCA clearly indicates large-scale motions in L1 and L5, suggesting opening up of the nucleotide binding cavity. (b) Model for the dynamics of SRD5 $\alpha$ 2 in one cycle of reaction. The reaction can be inhibited by finasteride by forming a stable adduct with the NADPH cofactor to stabilize the closed conformation of SRD5 $\alpha$ 2. Testosterone and finasteride were indicated as “T” and “Fina”, respectively. The catalysis of testosterone and the finasteride inhibition are shown with black and red arrows, respectively.

To further investigate the conformational dynamics of SRD5 $\alpha$ 2, we performed molecular dynamics (MD) simulations of SRD5 $\alpha$ 2 with NADP<sup>+</sup> (nap) and without NADP-DHF (apo) on microsecond timescales. Principle component analysis (PCA) of the MD trajectories clearly indicated high structural dynamics of the cytosolic loop L1 and to a lesser extent L5, which indeed resulted in the opening of the nucleotide-binding pocket during both nap and apo simulations (**Fig. 4a and Extended Data Fig. 5a**). In addition, in the apo simulations, L1 exhibited more pronounced conformational fluctuations as compared to it in the nap simulations (**Extended Data Fig. 5a**), presumably due to the stabilization of the cytosolic loops by NADP<sup>+</sup> in the nap simulations (**Fig. 2d and Extended Data Fig. 4a**). Since L1 is involved in the binding pockets for both NADP and finasteride in the crystal structure, our simulation results suggest that L1 may function as a ‘gate’ domain to control the NADPH/NADP<sup>+</sup> exchange and the binding of steroid substrates (**Fig. 4b**). Consistently, we observed higher B-factors in general for residues in L1 compared to residues in TMs and other cytosolic loops in the crystal structure, supporting the highly flexible feature of L1 (**Extended Data Fig. 5b**).

## Disease-related mutations of SRD5 $\alpha$ 2

To date, over 100 genetic mutations on the *SRD5 $\alpha$ 2* gene have been identified according to the Human Gene Mutation Database ([www.hgmd.cf.ac.uk](http://www.hgmd.cf.ac.uk)) to cause the rare autosomal recessive disorder 5 $\alpha$ -reductase deficiency<sup>9,10</sup>. Most mutations either abolish or reduce the activity of SRD5 $\alpha$ 2, leading to significantly reduced levels of DHT *in vivo*<sup>10</sup>. A majority of 5 $\alpha$ -reductase deficiency-causing mutations are missense mutations that locate throughout the whole protein (**Fig. 5a**). By mapping reported missense mutations of SRD5 $\alpha$ 2 protein in our structure, we found that most of the mutation sites are at the ligand-binding cavity (**Fig. 5b**), suggesting that many of those mutations compromise the activity of SRD5 $\alpha$ 2 by impairing the cofactor/substrate binding or the catalytic process. Examples include two of the founder mutations for the 5 $\alpha$ -reductase deficiency patients<sup>10</sup>, R227Q and R171S, which diminish the SRD5 $\alpha$ 2 activity likely by disrupting the hydrogen bonds with NADP (**Extended Data Fig. 4c-e**). Also supporting the critical roles of E57<sup>TM2</sup> and Y91<sup>TM3</sup> in our proposed catalytic mechanism, the mutations E57Q and Y91D have been shown to significantly reduce enzyme activity<sup>26,38</sup>. In addition, some mutations that are not in the ligand-binding cavity presumably impair protein folding/stability. For example, the C133G mutation eliminates the C5-C133 disulfide bridge that links TM1 and TM4, which may be important for the overall folding of SRD5 $\alpha$ 2 (**Fig. 5c**). As the most-frequently reported mutation site, R246<sup>C</sup> in the C-loop forms multiple hydrogen bonds with residues in L5 (**Fig. 5d**), potentially stabilizing the L5 to be in place for the cofactor binding. Consistently, the R246W mutation showed a decreased NADPH-binding affinity in previous studies<sup>9,26</sup>.



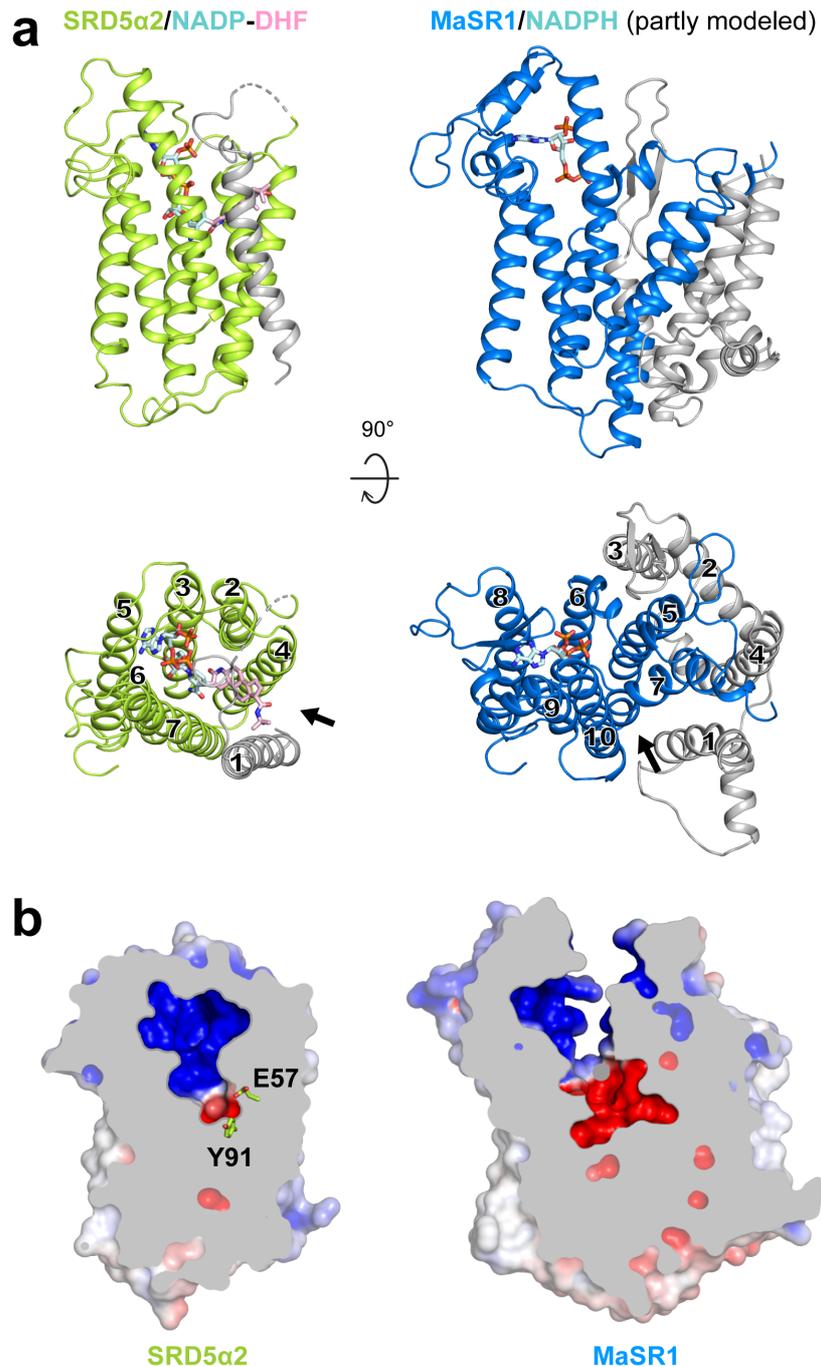
**Fig. 5: Structural analysis of disease-associated missense mutations of SRD5 $\alpha$ 2.** (a) Distribution of identified missense mutations of SRD5 $\alpha$ 2. The red bar length for each mutation site indicates how many times it was reported based on the data collected from HGMD and literature. The sites that are involved in the binding of NADP-DHF and the formation of the binding cavity are indicated by black and grey asterisks respectively. The mutations presumably disrupting the steroid binding, the cofactor binding, and the catalysis are labeled in pink, cyan and red respectively. (b) Mapping of the mutation sites onto the SRD5 $\alpha$ 2 structure. The mutation sites are colored in red. (c)-(d) Environment of C133 and R246 suggesting their roles in protein folding. (e) Potential hydrogen-bonding interactions caused by the A49T mutation. (f) Potential hydrophobic interactions caused by the A248V mutation.

On the other hand, two recurrent somatic mutations of SRD5 $\alpha$ 2, A49T and A248V, in human prostate cancer have been shown to significantly increase the activity of SRD5 $\alpha$ 2<sup>29</sup>. Especially, the A49T mutation has been linked to significantly increased risks of prostate cancer for certain populations<sup>39</sup>. A49<sup>TM2</sup> is located at the cytosolic end of TM2 to face L3 and the adenine-ribose phosphate moiety of NADPH. The A49T mutation may lead to an additional hydrogen bond with residues from L3 to further stabilize the conformation of L3 (**Fig. 5e**). As for A248<sup>C</sup>, it tightly packs against L5. The A248V mutation likely results in additional hydrophobic interactions with nearby residues F229<sup>TM7</sup> and F254<sup>C</sup> to stabilize the C-loop, which in turn stabilizes the conformation of L5. Since both L3 and L5 are involved in the cofactor binding, the mutations of these residues may enhance the cofactor binding by stabilizing its binding pocket to increase the activity of SRD5 $\alpha$ 2.

## DISCUSSION

In contrast to the well-understood mechanisms for the function of soluble steroid reductases, for which numerous structures have been reported<sup>37</sup>, the molecular mechanisms governing the function of eukaryotic membrane-embedded steroid reductases have remained enigmatic due to limited structural information. To our knowledge, the reported structure of human SRD5 $\alpha$ 2 with the dualsteric ligand NADP-DHF represents the first structure of eukaryotic membrane-embedded steroid reductase. Together with computational studies, our structure unveils the binding cavity inside the 7-TM bundle for NADPH and steroid substrates with flexible cytosolic loops. Structural analysis and mutagenesis studies suggest the molecular mechanisms for enzyme catalysis and inhibition involving newly identified residues E57<sup>TM2</sup> and Y91<sup>TM3</sup>. Because of the chemical differences in the steroid core ring, testosterone and finasteride adopt different binding poses relative to NADPH so that the  $\Delta^{4,5}$  bond in testosterone and the  $\Delta^{1,2}$  bond in finasteride can be reduced by SRD5 $\alpha$ 2 to generate distinct end products. Our results well explain the 5 $\alpha$ -reduction reactions catalyzed by SRD5 $\alpha$ s, contrasting the 5 $\beta$ -reduction reactions catalyzed by soluble steroid reductases. The mechanism-based irreversible action of finasteride and dutasteride led to their successful use as anti-androgen drugs, which may be repurposed for treating COVID-19 patients with excessive androgen receptor signaling<sup>18,19</sup>. Mapping disease-related mutations of SRD5 $\alpha$ 2 to our structure also provides feasible molecular mechanisms for the effects of those mutations in the 5 $\alpha$ -reductase deficiency and the prostate cancer.

The SRD5 $\alpha$ 2 structure together with simulation studies reveal unexpected structural features for the binding of NADP<sup>+</sup>/NADPH and steroid substrates. The cytoplasmic loops L1, L3 and L5 enclose the binding pocket of NADPH inside the 7-TM bundle of SRD5 $\alpha$ 2 to position it close to testosterone for catalysis. To our knowledge, such a structure feature has not been observed in other enzymes using pyridine nucleotides including NADH and NADPH as cofactors. Our MD simulation results suggest that the cytoplasmic loop L1 undergoes dramatic conformational changes during the reaction to allow the NADP<sup>+</sup>/NADPH exchange (**Fig. 4a**). The potential large energy barrier for the cytosolic region to overcome to open up the NADPH-binding pocket may hinder NADPH binding. As a result, the reported dissociation constant of NADPH for SRD5 $\alpha$ 2 ( $\sim$ 3-10 $\mu$ M)<sup>1</sup> is higher than that of NADPH for the soluble steroid reductase AKR1D1 with a highly exposed NADPH-binding pocket ( $\sim$ 0.5 $\mu$ M)<sup>35</sup>. As for the steroid substrates, they likely access the ligand-binding pocket of SRD5 $\alpha$ 2 from the lipid bilayer through the opening between TM1 and TM4 (**Fig. 2c, 6a**), which is analogue to the lateral ligand entrance mechanism for several GPCRs with lipid ligands<sup>40,41</sup>.



**Fig. 6: Structural comparison of SRD5 $\alpha$ 2 and MaSR1.** (a) Structural superposition of the transmembrane regions of SRD5 $\alpha$ 2 and MaSR1. The core transmembrane regions in SRD5 $\alpha$ 2 and MaSR1 that can be structurally aligned were shown as colored ribbons while the rest parts were shown as grey ribbons. The potential substrate entry ports in SRD5 $\alpha$ 2 and MaSR1 were indicated by arrows. (b) Occluded and exposed ligand-binding cavities in SRD5 $\alpha$ 2 and MaSR1, respectively.

Despite sharing very little sequence similarity and having different structural topology, human SRD5 $\alpha$ 2 and bacterial MaSR1 can be aligned in their core structure of six transmembrane helices (TM2-7 in SRD5 $\alpha$ 2) that participates in the binding of NADPH and substrates (Fig. 6a). There is also an opening between TM7 and TM10 in MaSR1 (Fig. 6a), which may serve as the potential entry port for its substrates. Such conserved structural features imply that these two enzymes and likely other

membrane-embedded steroid reductases take NADPH from the cytosol through cytoplasmic loops and steroid substrates from the lipid bilayer into ligand-binding cavities inside the cell membrane for catalysis. In contrast to the buried NADP in SRD5 $\alpha$ 2, the partially modeled NADPH in the structure of MaSR1 occupies a different binding pocket that is exposed to the cytosol (**Fig. 6a, b**)<sup>21</sup>, suggesting diverse NADPH recognition mechanisms for membrane-embedded steroid reductases. We speculate that these enzymes are likely to adopt distinct structural features within their TM bundles so NADPH can be appropriately positioned towards their steroid substrates for site-specific carbon-carbon double bond reduction. Further structural investigation on other important eukaryotic membrane-embedded steroid reductases including SRD5 $\alpha$ 3 and DHCR7 is needed to understand the potentially different mechanisms by which these enzymes reduce specific carbon-carbon double bonds at different positions of chemically similar steroid substrates using NADPH.

**TABLE 1. X-ray data collection and refinement statistics.**

<b>Data collection</b>	
Space group	<i>P</i> 6 2 2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	107.449, 107.449, 103.372
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 120
Resolution (Å)	50–2.8 (2.85–2.8)
<i>R</i> <sub>merge</sub> <sup>a</sup>	0.248 (1.274)
<i>I</i> / $\sigma$ <i>I</i>	13.0 (1.05)
<i>CC</i> <sub>1/2</sub> <sup>b</sup>	1.0 (0.415)
Completeness (%)	99.7 (100.0)
Redundancy	9.6 (6.8)
<b>Refinement</b>	
Resolution (Å)	50–2.8 (3.19–2.8)
No. reflections	9155 (2857)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> <sup>c</sup>	0.239 (0.297) / 0.265 (0.336)
No. atoms	2036
Protein	1926
Ligand	110
Mean <i>B</i> (Å <sup>2</sup> )	75.8
r.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.77
Ramachandran analysis	
Favored region (%)	95.9
Allowed region (%)	4.1
Outliers (%)	0
The numbers in parentheses represent values for the highest resolution shell.	
<sup>a</sup> <i>R</i> <sub>merge</sub> = $\sum  I_i - I_m  / \sum I_i$ , where <i>I</i> <sub>i</sub> is the intensity of the measured reflection and <i>I</i> <sub>m</sub> is the mean intensity	
<sup>b</sup> <i>CC</i> <sub>1/2</sub> is the correlation coefficient of the half datasets.	
<sup>c</sup> <i>R</i> <sub>work</sub> = $\sum  F_{obs}  -  F_{calc}  / \sum  F_{obs} $ , where <i>F</i> <sub>obs</sub> and <i>F</i> <sub>calc</sub> are observed and calculated structure factors.	
<i>R</i> <sub>free</sub> = $\sum_T  F_{obs}  -  F_{calc}  / \sum_T  F_{obs} $ , where T is a test data set of about 5 % of the total reflections	

## Methods

### SRD5 $\alpha$ 2 expression and purification

cDNA of human full-length wild-type SRD5 $\alpha$ 2 was synthesized (IDT) and cloned into pFastBac vector with an N-terminal signal peptide followed by a Flag epitope and an 8xhistidine tag. One tobacco-etch virus (TEV) protease cleavage was introduced after the His tag. SRD5 $\alpha$ 2 protein was expressed in the insect Sf9 cells using the Bac-to-Bac baculovirus system (ThermoFisher). Cells were infected by baculovirus at a density of  $4.0 \times 10^6$  cells/mL and harvested after 48 h at 27 °C. To stabilize the protein, all purification steps were accomplished in the presence of the inhibitor finasteride (Tocris). Sf9 cells were lysed in buffer containing 20 mM Tris pH7.5, 2.0 mg ml<sup>-1</sup> iodoacetamide, 0.2  $\mu$ g ml<sup>-1</sup> leupeptin, 100  $\mu$ g ml<sup>-1</sup> benzamidine and 0.5  $\mu$ M finasteride. The cell pellet was further resuspended and homogenized using the glass dounce homogenizer in buffer containing 20 mM HEPES pH7.5, 750 mM NaCl, 1.0  $\mu$ M finasteride, 20% glycerol, 1% dodecyl maltoside (DDM), 0.1% cholesterol hemisuccinate (CHS), 0.2% sodium cholate, 2.0 mg ml<sup>-1</sup> iodoacetamide, 0.2  $\mu$ g ml<sup>-1</sup> leupeptin, 100  $\mu$ g ml<sup>-1</sup> benzamidine and salt active nuclease (ArcticZymes) and incubated at 4°C for 90 min. After centrifugation, the supernatant was collected and incubated with Ni-NTA resin (GE healthcare) in the presence of 8 mM imidazole to prevent the non-specific binding at 4°C overnight. The resin was pelleted and washed with 20 mM HEPES pH7.5, 500 mM NaCl, 1.0  $\mu$ M finasteride, 0.1% DDM, 0.02% CHS and 40 mM imidazole. The bound protein was eluted using the same buffer with 400 mM imidazole and then loaded onto the anti-Flag M1 affinity column after supplementing with 2 mM CaCl<sub>2</sub>. After extensively and slowly exchanging detergent to 0.1% lauryl maltose neopentyl glycol (LMNG) (Anatrace), the protein was eluted using buffer containing 20 mM HEPES pH7.5, 100 mM NaCl, 1.0  $\mu$ M finasteride, 0.01% LMNG, 0.001% CHS, 200  $\mu$ g ml<sup>-1</sup> synthesized Flag peptide (GL Biochem) and 5 mM EDTA. TEV protease and PNGase F were added to the eluted protein and incubated at 4°C overnight. The treated sample was reloaded onto Ni-NTA column to remove TEV protease and the flow-through fraction was collected. The protein was further purified by size-exclusion chromatography using a Superdex S200 increase column (GE healthcare) in buffer containing 20 mM HEPES pH7.5, 100 mM NaCl, 1.0  $\mu$ M finasteride, 0.01% LMNG and 0.001% CHS. The monodispersed fractions were pooled together and concentrated to 50 mg ml<sup>-1</sup> for crystallization.

### Crystallization

Purified protein in complex with finasteride was reconstituted into the lipidic cubic phase (LCP)<sup>22</sup> through mixing protein with monoolein/cholesterol lipid mixture (10:1 w/w) (Avanti) at a weight ratio of 2:3 (protein: lipid) using coupled syringes (ArtRobbins). Using a Gryphon LCP robot (ArtRobbins), the LCP mixture was dispensed onto 96-well glass sandwich plates in 20 nL drops and then overlaid with 700 nL precipitant conditions. Plates were incubated at 20 °C and crystals grew in the condition of 100 mM tris-sodium citrate pH5.0, 28-34% PEG600, 100-150 mM NaCl and 100 mM Li<sub>2</sub>SO<sub>4</sub>. Crystals were collected from the LCP matrix and frozen in liquid nitrogen for data collection.

### *De novo* structure modeling

To predict the 3D model of human SRD5 $\alpha$ 2, several alternative multi-sequence alignments were generated using a diverse set of sequence databases under different sequence search engines<sup>42</sup> and then fused by implementing a deep residual network with a *strip pooling* module<sup>43</sup> to effectively

capture long-range relationship of residual pairs. Following trRosetta<sup>44</sup>, we generated the 3D model of SRD5 $\alpha$ 2 from the predicted distance and orientation using constrained minimization. Specifically, the predicted distance- and orientation-probabilities were first converted into potentials, which are then used as restraints to be fed into Rosetta together with coarse-grained energy optimization. Finally, the top 5 folded structures satisfying the restraints were selected according to Rosetta energy as initial models for structure determination.

### **X-ray data collection and structure determination**

X-ray diffraction data was collected at the beamline 23ID-B, GM/CA of Advanced Photon Source (APS) in the Argonne National Laboratory at Chicago. Each crystal was exposed with a 10  $\mu$ m x 10  $\mu$ m beam for 0.2 s and 0.2-degree oscillation per frame to collect 20 degrees of rotation data. Data sets from 5 crystals were processed and merged in HKL2000 software<sup>45</sup>.

To determine the phase of the crystal structure, the *de novo* structure models were used as the search models for molecular replacement. One of the models was successfully used for searching the solution by Molrep in CCP4 package<sup>46</sup>. The initial phase was largely improved by using density modification methods in PHENIX<sup>47</sup>. COOT<sup>48</sup> was used for model rebuilding based on the improved electron density. The rebuilt model was further refined in PHENIX with an additional TLS refinement was performed. The model quality was check by MolProbity<sup>49</sup>. The final refinement statistics are listed in **Table 1**. All structure figures were prepared by PyMOL (<http://www.pymol.org/>).

### **Molecular docking**

The NADP-DHF adduct from the crystal structure was removed and an analogous NADP-testosterone adduct with NADP covalently bound to the C-5 atom of testosterone at the  $\alpha$ -face was docked to the crystal structure. In the docked pose, the covalent link was removed to obtain NADPH and testosterone. This was followed by optimizing the ligands and their surrounding 5  $\text{\AA}$  protein residues. Thereafter, NADPH was redocked to the optimized protein, followed by redocking testosterone to the NADPH docked protein. The docking poses were filtered to be within 10  $\text{\AA}$  RMSD of the X-ray ligand densities to obtain the pose for NADPH and testosterone docked to SRD5 $\alpha$ 2. Docking calculations were performed using Glide<sup>50</sup>.

### **Molecular modelling and MD simulations**

SRD5 $\alpha$ 2 models were built based on the solved crystal structure to incorporate unresolved residues 1-4 (N-terminus) and 39-43 (loop L1) using Modeller<sup>51</sup>. We generated 5000 models and selected top 10 scoring models based on DOPE scoring function<sup>52</sup>. From the top 10 models, two most structurally diverse protein configurations were selected to start molecular dynamics (MD) simulations for unliganded (apo0, apo1) and NADP<sup>+</sup>-bound (nap0, nap1) states. NADP<sup>+</sup> was placed at the NADP density in the crystal structure. NADP<sup>+</sup> was modelled at a molecular charge of -2. Ionization states of protein residues were calculated at pH=7 by solving Poisson-Boltzmann equation in continuum electrostatics models ( $\epsilon_{\text{protein}}=20$ ;  $\epsilon_{\text{solvent}}=80$ ) using APBS<sup>53</sup>. All residues were found to be in their standard ionization states. Based on the crystal structure, a disulfide bond was modelled between Cys-5 and Cys-133. The protein models were embedded in a pre-equilibrated 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer followed by solvation (TIP3P water model) and

neutralization using potassium and chloride ions at 150 mM. The simulation setups comprised *ca.* 44000 atoms each. CHARMM36 forcefield was employed for the MD simulations including protein, ligands, lipids, water and ions<sup>54</sup>. The systems were first energy minimized for 10000 steps and then heated gradually from 0 K to 310 K for 250 ps using a Langevin thermostat with heavy atoms restrained at 10 kcal mol<sup>-1</sup> Å<sup>-2</sup> in an *NVT* ensemble. The heated systems were subjected to 8 successive rounds of 1 ns equilibration steps. During the equilibration, protein and ligand heavy atoms were subjected to harmonic restraints and lipids were subjected to planar restraints to maintain bilayer planarity. The harmonic restraints for each step were relaxed progressively going from 10 kcal mol<sup>-1</sup> Å<sup>-2</sup> to 0.1 kcal mol<sup>-1</sup> Å<sup>-2</sup>. The equilibrations were performed at a 1 fs timestep at  $T = 310$  K and  $P = 1$  bar using the Langevin thermostat and Nosé-Hoover Langevin barostat in *NPT* ensemble. The production runs were performed with a hydrogen-mass repartitioning (HMR) scheme with a timestep of 3.6 fs with a non-bonded cutoff at 12 Å<sup>55</sup>. Long range electrostatics were evaluated with the Particle Mesh Ewald method. Protein and lipid bond-lengths were constrained with the SHAKE algorithm. Each system was simulated for *ca.* 1.25 μs, giving a total simulation time of *ca.* 5.4 μs (2.7 μs apo and 2.7 μs nap). Trajectory snapshots were saved at every 50 ps. The simulations were performed with NAMD 2.13<sup>56</sup>. The simulation setup was constructed using CHAMM-GUI<sup>57</sup>. In order to analyze the multidimensional conformational landscape, we performed principal component analysis (PCA) of MD trajectories to identify the dominant modes of protein motions (principal components) in the apo and nap simulation states. PCA was performed with pyPcazip<sup>58</sup>. Visual Molecular Dynamics (VMD)<sup>59</sup> was employed for visualization and for performing RMSF analysis to probe protein mobility.

### Enzyme activity measurement

We used insect cell membranes overexpressing different constructs of SRD5α2 in the measurement of enzyme activity. Sf9 cells expressing wild type SRD5α2 (WT) and mutants E57Q and Y91F were harvested after 48 h transfection. The expression level of each construct was determined by flow cytometry using FITC-labeled anti-FLAG M2 antibody (Sigma Aldrich) in the presence of 0.5% tritonX-100. For membrane preparation, the cell pellets were resuspended in lysis buffer containing 20 mM Tris pH7.5, 1 mM EDTA, 0.2 μg ml<sup>-1</sup> leupeptin and 100 μg ml<sup>-1</sup> benzamidine. The lysed and homogenized samples were centrifuged at 1,200xg for 8 min. The supernatant was further centrifugated at 300,000xg for 50 min to get the membrane pellets, which were resuspended in buffer containing 20 mM HEPES pH7.5, 100mM NaCl, 1 mM EDTA, 0.2 μg ml<sup>-1</sup> leupeptin and 100 μg ml<sup>-1</sup> benzamidine and stored at -80°C.

The enzyme activity was assayed using the prepared membrane fractions (0.3 mg each reaction) in the presence of 0.5 mM NADPH and 0.5 mM testosterone in 0.1 M sodium-citrate pH 5.0 buffer. Finasteride at 0.5 mM was used as a negative control. After incubation for 4 hours at 37°C, the steroids were extracted with chloroform, evaporated by centrifugal concentrator and re-dissolved in methanol for liquid chromatography–mass spectrometry (LC-MS) analysis using buffer A containing 1 mM ammonium formate and buffer B containing 100% methanol. The ratios of the peak area of dihydrotestosterone to the peak area of testosterone were calculated as indicators of enzyme activity. The data were processed by Prism8 (GraphPad). The experiments on the wt SRD5α2 (WT) were repeated 5 times. The experiments on WT with 0.5 mM finasteride were repeated 4 times. The

experiments on each mutant, E57Q or Y91F, were repeated 3 times. The data were represented as mean  $\pm$  SEM.

### **Data availability**

The coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes PDB 7BW1. Distribution of research materials generated in this study including plasmids will need appropriate Material Transfer Agreements (MTAs).

## **Acknowledgements**

We thank the staff of the GM/CA of APS at the Argonne National Laboratory in Chicago for assistance with X-ray diffraction data collection and Southern University of Science and Technology Core Research Facilities for assistance with LC-MS analysis. This work was supported by the National Institutes of Health grants 1R35GM128641 (to C.Z.), the National Natural Science Foundation of China 31971131 and 31770791, Shenzhen-Hong Kong Institute of Brain Science, Shenzhen Fundamental Research Institutions 2019SHIBS0002 (to Z.W.), and funding support from the Biomedical Research Council of A\*STAR (to S.S. and H.F.). The molecular dynamics simulations were performed on resources of the National Supercomputing Centre, Singapore (<https://www.nscg.sg>).

## **Author contributions**

C.Z. and Z.W. supervised the project and designed the research with H.F., L.W. and Q.X.. Q.X. performed the experiments for protein expression, purification and crystallization with the assistance from L.W.. L.W. performed x-ray diffraction data collection with the assistance from H.L.. C.Z. processed the x-ray diffraction data. T.S., F.Y. and J.H. performed *de novo* structure modeling. Z.W. determined the crystal structure. S.S. and H.F. performed computational docking and simulation studies. C.Z., Z.W., S.S. and H.F. prepared the manuscript with the assistance from Q.X..

## **Competing interests**

The authors declare no competing interests.

## References

- 1 Azzouni, F., Godoy, A., Li, Y. & Mohler, J. The 5 alpha-reductase isozyme family: a review of basic biology and their role in human diseases. *Adv Urol* **2012**, 530121, doi:10.1155/2012/530121 (2012).
- 2 Cantagrel, V. *et al.* SRD5A3 is required for converting polyprenol to dolichol and is mutated in a congenital glycosylation disorder. *Cell* **142**, 203-217, doi:10.1016/j.cell.2010.06.001 (2010).
- 3 Langlois, V. S., Zhang, D., Cooke, G. M. & Trudeau, V. L. Evolution of steroid-5alpha-reductases and comparison of their function with 5beta-reductase. *Gen Comp Endocrinol* **166**, 489-497, doi:10.1016/j.ygcen.2009.08.004 (2010).
- 4 Compagnone, N. A. & Mellon, S. H. Neurosteroids: biosynthesis and function of these novel neuromodulators. *Front Neuroendocrinol* **21**, 1-56, doi:10.1006/frne.1999.0188 (2000).
- 5 Dong, E. *et al.* Brain 5alpha-dihydroprogesterone and allopregnanolone synthesis in a mouse model of protracted social isolation. *Proc Natl Acad Sci U S A* **98**, 2849-2854, doi:10.1073/pnas.051628598 (2001).
- 6 Kousal, B. *et al.* Review of SRD5A3 Disease-Causing Sequence Variants and Ocular Findings in Steroid 5alpha-Reductase Type 3 Congenital Disorder of Glycosylation, and a Detailed New Case. *Folia Biol (Praha)* **65**, 134-141 (2019).
- 7 Robitaille, J. & Langlois, V. S. Consequences of steroid-5alpha-reductase deficiency and inhibition in vertebrates. *Gen Comp Endocrinol* **290**, 113400, doi:10.1016/j.ygcen.2020.113400 (2020).
- 8 Scaglione, A. *et al.* Subcellular localization of the five members of the human steroid 5alpha-reductase family. *Biochim Open* **4**, 99-106, doi:10.1016/j.biopen.2017.03.003 (2017).
- 9 Thigpen, A. E. *et al.* Molecular genetics of steroid 5 alpha-reductase 2 deficiency. *J Clin Invest* **90**, 799-809, doi:10.1172/JCI115954 (1992).
- 10 Mendonca, B. B. *et al.* Steroid 5alpha-reductase 2 deficiency. *J Steroid Biochem Mol Biol* **163**, 206-211, doi:10.1016/j.jsbmb.2016.05.020 (2016).
- 11 Imperato-McGinley, J., Guerrero, L., Gautier, T. & Peterson, R. E. Steroid 5alpha-reductase deficiency in man: an inherited form of male pseudohermaphroditism. *Science* **186**, 1213-1215, doi:10.1126/science.186.4170.1213 (1974).
- 12 Marks, L. S. 5alpha-reductase: history and clinical importance. *Rev Urol* **6 Suppl 9**, S11-21 (2004).
- 13 Schmidt, L. J. & Tindall, D. J. Steroid 5 alpha-reductase inhibitors targeting BPH and prostate cancer. *J Steroid Biochem Mol Biol* **125**, 32-38, doi:10.1016/j.jsbmb.2010.09.003 (2011).
- 14 Aggarwal, S., Thareja, S., Verma, A., Bhardwaj, T. R. & Kumar, M. An overview on 5alpha-reductase inhibitors. *Steroids* **75**, 109-153, doi:10.1016/j.steroids.2009.10.005 (2010).
- 15 Liss, M. A. & Thompson, I. M. Prostate cancer prevention with 5-alpha reductase inhibitors: concepts and controversies. *Curr Opin Urol* **28**, 42-45, doi:10.1097/MOU.0000000000000464 (2018).
- 16 Tian, G., Chen, S.-Y., Facchine, K. L. & Prakash, S. R. Chemical Mechanism of the Covalent Modification of 5.alpha.-Reductases by Finasteride as Probed by Secondary Tritium Isotope Effects. *Journal of the American Chemical Society* **117**, 2369-2370, doi:10.1021/ja00113a034 (1995).
- 17 Bull, H. G. *et al.* Mechanism-Based Inhibition of Human Steroid 5 $\alpha$ -Reductase by Finasteride: Enzyme-Catalyzed Formation of NADP-Dihydrofinasteride, a Potent Bisubstrate Analog Inhibitor. *Journal of the American Chemical Society* **118**, 2359-2365, doi:10.1021/ja953069t (1996).
- 18 Wambier, C. G. *et al.* Androgen sensitivity gateway to COVID-19 disease severity. *Drug Dev Res*, doi:10.1002/ddr.21688 (2020).

- 19 Ghazizadeh, Z. *et al.* Androgen Regulates SARS-CoV-2 Receptor Levels and Is Associated with Severe COVID-19 Symptoms in Men. *bioRxiv*, 2020.2005.2012.091082, doi:10.1101/2020.05.12.091082 (2020).
- 20 Porter, F. D. & Herman, G. E. Malformation syndromes caused by disorders of cholesterol synthesis. *J Lipid Res* **52**, 6-34, doi:10.1194/jlr.R009548 (2011).
- 21 Li, X., Roberti, R. & Blobel, G. Structure of an integral membrane sterol reductase from *Methylomicrobium alcaliphilum*. *Nature* **517**, 104-107, doi:10.1038/nature13797 (2015).
- 22 Caffrey, M. Crystallizing membrane proteins for structure determination: use of lipidic mesophases. *Annu Rev Biophys* **38**, 29-51, doi:10.1146/annurev.biophys.050708.133655 (2009).
- 23 Rossmann, M. G. The molecular replacement method. *Acta Crystallogr A* **46 ( Pt 2)**, 73-82, doi:10.1107/s0108767389009815 (1990).
- 24 von Heijne, G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**, 456-458, doi:10.1038/341456a0 (1989).
- 25 Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356-363, doi:10.1038/nature08144 (2009).
- 26 Wigley, W. C. *et al.* Natural mutagenesis study of the human steroid 5 alpha-reductase 2 isozyme. *Biochemistry* **33**, 1265-1270, doi:10.1021/bi00171a029 (1994).
- 27 Wang, M., Bhattacharyya, A. K., Taylor, M. F., Tai, H. H. & Collins, D. C. Site-directed mutagenesis studies of the NADPH-binding domain of rat steroid 5alpha-reductase (isozyme-1) I: analysis of aromatic and hydroxylated amino acid residues. *Steroids* **64**, 356-362, doi:10.1016/s0039-128x(99)00010-0 (1999).
- 28 Makridakis, N. M., di Salle, E. & Reichardt, J. K. Biochemical and pharmacogenetic dissection of human steroid 5 alpha-reductase type II. *Pharmacogenetics* **10**, 407-413, doi:10.1097/00008571-200007000-00004 (2000).
- 29 Makridakis, N., Akalu, A. & Reichardt, J. K. Identification and characterization of somatic steroid 5alpha-reductase (SRD5A2) mutations in human prostate cancer tissue. *Oncogene* **23**, 7399-7405, doi:10.1038/sj.onc.1207922 (2004).
- 30 Li, J., Nagpal, P., Vitart, V., McMorris, T. C. & Chory, J. A role for brassinosteroids in light-dependent development of *Arabidopsis*. *Science* **272**, 398-401, doi:10.1126/science.272.5260.398 (1996).
- 31 Tian, G., Mook, R. A., Jr., Moss, M. L. & Frye, S. V. Mechanism of time-dependent inhibition of 5 alpha-reductases by delta 1-4-azasteroids: toward perfection of rates of time-dependent inhibition by using ligand-binding energies. *Biochemistry* **34**, 13453-13459, doi:10.1021/bi00041a024 (1995).
- 32 Sanli, G., Dudley, J. I. & Blaber, M. Structural biology of the aldo-keto reductase family of enzymes: catalysis and cofactor binding. *Cell Biochem Biophys* **38**, 79-101, doi:10.1385/CBB:38:1:79 (2003).
- 33 Drury, J. E., Di Costanzo, L., Penning, T. M. & Christianson, D. W. Inhibition of human steroid 5beta-reductase (AKR1D1) by finasteride and structure of the enzyme-inhibitor complex. *J Biol Chem* **284**, 19786-19790, doi:10.1074/jbc.C109.016931 (2009).
- 34 Di Costanzo, L., Drury, J. E., Penning, T. M. & Christianson, D. W. Crystal structure of human liver Delta4-3-ketosteroid 5beta-reductase (AKR1D1) and implications for substrate binding and catalysis. *J Biol Chem* **283**, 16830-16839, doi:10.1074/jbc.M801778200 (2008).
- 35 Faucher, F., Cantin, L., Luu-The, V., Labrie, F. & Breton, R. Crystal structures of human Delta4-3-ketosteroid 5beta-reductase (AKR1D1) reveal the presence of an alternative binding site responsible for substrate inhibition. *Biochemistry* **47**, 13537-13546, doi:10.1021/bi801276h (2008).

- 36 Di Costanzo, L., Drury, J. E., Christianson, D. W. & Penning, T. M. Structure and catalytic mechanism of human steroid 5beta-reductase (AKR1D1). *Mol Cell Endocrinol* **301**, 191-198, doi:10.1016/j.mce.2008.09.013 (2009).
- 37 Barski, O. A., Tipparaju, S. M. & Bhatnagar, A. The aldo-keto reductase superfamily and its role in drug metabolism and detoxification. *Drug Metab Rev* **40**, 553-624, doi:10.1080/03602530802431439 (2008).
- 38 Vilchis, F. *et al.* Novel compound heterozygous mutations in the SRD5A2 gene from 46,XY infants with ambiguous external genitalia. *J Hum Genet* **53**, 401-406, doi:10.1007/s10038-008-0274-2 (2008).
- 39 Makridakis, N. M. *et al.* Association of mis-sense substitution in SRD5A2 gene with prostate cancer in African-American and Hispanic men in Los Angeles, USA. *Lancet* **354**, 975-978, doi:10.1016/S0140-6736(98)11282-5 (1999).
- 40 Wang, L. *et al.* Structures of the Human PGD2 Receptor CRTH2 Reveal Novel Mechanisms for Ligand Recognition. *Mol Cell* **72**, 48-59 e44, doi:10.1016/j.molcel.2018.08.009 (2018).
- 41 Luginina, A. *et al.* Structure-based mechanism of cysteinyl leukotriene receptor inhibition by antiasthmatic drugs. *Sci Adv* **5**, eaax2518, doi:10.1126/sciadv.aax2518 (2019).
- 42 Wang, S., Sun, S., Li, Z., Zhang, R. & Xu, J. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLoS Comput Biol* **13**, e1005324, doi:10.1371/journal.pcbi.1005324 (2017).
- 43 Hou, Q., Zhang, L., Cheng, M. & Feng, J. Strip Pooling: Rethinking Spatial Pooling for Scene Parsing. *arXiv* (2020).
- 44 Yang, J. *et al.* Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci U S A* **117**, 1496-1503, doi:10.1073/pnas.1914677117 (2020).
- 45 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).
- 46 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).
- 47 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
- 48 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
- 49 Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315, doi:10.1002/pro.3330 (2018).
- 50 Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47**, 1739-1749, doi:10.1021/jm0306430 (2004).
- 51 Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815, doi:10.1006/jmbi.1993.1626 (1993).
- 52 Shen, M. Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* **15**, 2507-2524, doi:10.1110/ps.062416606 (2006).
- 53 Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112-128, doi:10.1002/pro.3280 (2018).
- 54 Huang, J. & MacKerell, A. D., Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **34**, 2135-2145, doi:10.1002/jcc.23354 (2013).
- 55 Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* **11**, 1864-1874, doi:10.1021/ct5010406 (2015).

- 56 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802, doi:10.1002/jcc.20289 (2005).
- 57 Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **12**, 405-413, doi:10.1021/acs.jctc.5b00935 (2016).
- 58 Shkurti, A. *et al.* pyPcazip: A PCA-based toolkit for compression and analysis of molecular simulation data. *SoftwareX* **5**, 44-50, doi:10.1016/j.softx.2016.04.002 (2016).
- 59 Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33-38, doi:10.1016/0263-7855(96)00018-5 (1996).