Three-dimensional Structure of Steroid 21-Hydroxylase (Cytochrome P450 21A2) with Two Substrates Reveals Locations of Disease-associated Variants

Steroid 21-hydroxylase (cytochrome P450 21A2, CYP21A2) deficiency accounts for ~95% of individuals with congenital adrenal hyperplasia, a common autosomal recessive metabolic disorder of adrenal steroidogenesis. The effects of amino acid mutations on CYP21A2 activity lead to impairment of the synthesis of cortisol and aldosterone and the excessive production of androgens. In order to understand the structural and molecular basis of this group of diseases, the bovine CYP21A2 crystal structure complexed with the substrate 17-hydroxyprogesterone was determined to 3.0 Å resolution. An intriguing result from this structure is that there are two molecules of 17OHP bound to the enzyme, the distal one being located at the entrance of the substrate access channel and the proximal one bound in the active site. The substrate binding features locate the key substrate recognition residues not only around the heme but also along the substrate access channel. In addition, orientation of the skeleton of the proximal molecule is toward the interior of the enzyme away from the substrate access channel. The 17OHP complex of CYP21A2 provides a good relationship between the crystal structure, clinical data, and genetic mutants documented in the literature, thereby enhancing our understanding of congenital adrenal hyperplasia. In addition, the location of certain CYP21A2 mutations provides general understanding of structure/function relationships in P450s.

Cytochrome P450 (P450)\(^3\) 21-hydroxylase (CYP21A2) converts progesterone and 17-hydroxyprogesterone (17OHP) to 11-deoxycorticosterone and 11-deoxycortisol, respectively (Fig. 1) (1, 2). CYP21A2 is expressed in the adrenal cortex and plays an essential role in the biosynthesis of aldosterone and cortisol. Deficiency of CYP21A2 enzymatic activity is the most common cause of congenital adrenal hyperplasia (CAH), an inherited disorder of steroidogenesis (3, 4). There are six CYP proteins (CYP11A1, CYP11B1, CYP11B2, CYP17A1, CYP19A1, and CYP21A2) associated with the synthesis of the steroid hormones in humans. CYP11A1, CYP11B1, and CYP11B2 are located in the inner mitochondrial membrane, and the other CYP proteins, including CYP21A2, are microsomal monoxygenases.

CAH results from complete or partial loss of activity in one of several steroidogenic enzymes. Up to 95% of all cases of CAH are due to CYP21A2 deficiency, which reduces production of cortisol and aldosterone and may lead to excess production of androgenic androgens (3). CAH traditionally has three forms based on clinical phenotypes: a severe form with a concurrent defect in aldosterone biosynthesis (salt-wasting type) and a form with apparently normal aldosterone biosynthesis (simple virilizing type) together termed classic 21-hydroxylase deficiency. There is also a mild, non-classic form that may be asymptomatic or associated with signs of postnatal androgen excess. The worldwide incidence of CAH is 1:15,000 for the classical form and ~1:1,000 for the nonclassic form (3). In humans, there are two CYP21 genes located in the leukocyte

\(^a\) This work was supported, in whole or in part, by National Institutes of Health Grants GM69970 (to M. R. W. and F. P. G.), R37 CA090426 (to F. P. G.), and P30 ES000267 (to F. P. G. and M. R. W.).

\(^b\) This article contains supplemental Figs. S1–S3.

The atomic coordinates and structure factors (code 3QQ1) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

1 To whom correspondence may be addressed: Dept. of Biochemistry and Center in Molecular Toxicology, Vanderbilt University School of Medicine, 632 Robinson Research Bldg., 2200 Pierce Ave., Nashville, TN 37232-0146. Tel.: 615-343-4644; Fax: 615-343-0704; E-mail: bin.zhao@vanderbilt.edu.

2 To whom correspondence may be addressed: Dept. of Biochemistry and Center in Molecular Toxicology, Vanderbilt University School of Medicine, 632 Robinson Research Bldg., 2200 Pierce Ave., Nashville, TN 37232-0146. Tel.: 615-343-4644; Fax: 615-343-0704; E-mail: Michael.Waterman@vanderbilt.edu.

3 The abbreviations used are: P450, cytochrome P450; CAH, congenital adrenal hyperplasia; 17OHP, 17-hydroxyprogesterone; C3B21RA, C3B21T241R/L442A.
Structure of P450 21A2 and Disease Implications

Expression and Purification of Engineered CYP21A2 (C3B21RA)—In order to obtain a soluble form of bovine CYP21A2 for crystallographic experiments, two surface residues were mutated, Thr-241 (to Arg) and Leu-442 (to Ala). The synthetic oligonucleotides used for site-directed mutagenesis were as follows: C3B21T241R forward primer (5’-gaacagcagctCG-Ggccccagcag) and the reverse primer (5’-cctggaggCGcacgctc). The C3B21T241R/L442A forward primer was 5’-ctggtgccGCgactgctgc, and the reverse primer was 5’-ctcagacagtGCgaccagcag. C3B21 was used as template to obtain C3B21T241R, which was then used as the template for C3B21T241R/L442A (C3B21RA). The changes were confirmed by nucleotide sequence analysis (the PCR is described in Ref. 15). The reaction was carried out using the following cycling parameters: 95 °C for 30 s, 55 °C for 1 min, 68 °C for 7 min for 25 cycles and 68 °C for 6 min. Then 40 μl of the reaction (to which 1 μl of the restriction enzyme DpnI had been added) was incubated at 37 °C for 2 h, after which 2 μl of the DpnI-treated reaction was transformed into 50 μl of Escherichia coli DH5α cells. Following culture at 37 °C for 1 h, the cells were spread on an agar plate containing 100 μg/ml ampicillin.

The C3B21RA protein was expressed and purified as reported earlier (16). In brief, cells co-transformed with the plasmids CYP21A2pET17b and pGro12 were cultured overnight in LB broth containing 100 μg/ml ampicillin and 50 μg/ml kanamycin. After inoculation (1:100, v/v) in 3 liters of Terrific Broth containing 100 μg/ml ampicillin and 50 μg/ml kanamycin, growth was carried out at 37 °C and 240 rpm for 6 h. Then 1 mM isopropyl 1-thio-β-D-galactopyranoside, 1 mM δ-aminovelinic acid, and 4 mg/ml arabinose were added, and culture was incubated for another 38 h at 27 °C. Cells were harvested by centrifugation (2850 × g, 20 min at 4 °C) and resuspended in 100 ml of TES buffer (250 mM sucrose, 50 mM Tris-HCl (pH 7.4), 0.5 mM EDTA) containing 1 mg/ml lysozyme. After incubation on ice for 30 min, the sample was centrifuged again. Cell pellets were sonicated in 250 ml of buffer A (50 mM potassium phosphate (pH 7.4), containing 20% glycerol (v/v), 0.1 mM dithiothreitol, 0.1 mM EDTA, 300 mM sodium chloride, 1.5% sodium cholate (w/v), 1% Tween 20 (w/v), and 0.1 mM phenylmethylsulfonyl fluoride). The supernatant (250 ml, 1200 nmol of CYP21A2) was applied on a nickel-NTA-agarose column (3 cm × 3 cm) equilibrated with buffer A. The column was washed with buffer A containing 20 mM imidazole, followed by buffer B (20 mM potassium phosphate (pH 7.4) containing 20% glycerol (v/v), 0.1 mM dithiothreitol, 0.1 mM EDTA, 1.0% sodium cholate (w/v), 1% Tween 20 (w/v), and 100 μM phenylmethylsulfonyl fluoride) and eluted with buffer C (50 mM potassium phosphate (pH 7.4) containing 20% glycerol (v/v), 0.1 mM dithiothreitol, 0.1 mM EDTA, and 1% sodium cholate, (w/v) containing 100 mM imidazole). The eluate was applied to a DEAE-Sepharose column (2 cm × 9 cm) and washed with 20 mM potassium phosphate buffer (pH 7.4) containing 20% glycerol (v/v), 0.1 mM dithiothreitol, 0.1 mM EDTA, and 1.0% sodium cholate (w/v). The flow-through fraction was applied to an SP-Sepharose column (2 cm × 6 cm) and washed with buffer D (50 mM potassium phosphate, pH 7.4, 20% glycerol, 0.1 mM DTT, 0.1 mM EDTA, 0.25% Cymal 5, and 50 mM NaCl), and the engineered form of CYP21A2 was eluted with a linear 50–300 mM NaCl gradient in the same buffer.

Catalytic Activity Assay—The procedure was modified from previous work (16). Preliminary experiments showed very similar rates of progesterone 21-hydroxylation by both a truncated P450 21A2 (16) and the C3B21RA mutant in 0.002% (w/v) Cymal 5 detergent and with 30 μM L-α-dilauroyl-sn-glycerol-3-phosphocholine vesicles, a commonly used P450 reconstitution system. For the determination of steady-state kinetic parameters, incubations contained 2 pmol of P450 21A2 (either the truncated version (16) or the C3B21RA mutant), 60 pmol of E. coli recombinant NADPH-P450 reductase (17, 18), 2 μg of L-α-dilauroyl-sn-glycerol-3-phosphocholine, and varying conc-
centrations of [4–14C]progesterone (51 mCi/mmol; American Radiolabeled Chemicals (St. Louis, MO); Lot 030610, 0.3–20 μM) in 85 μl of 50 mM potassium phosphate buffer. Preincubation was at 37 °C for 5 min, and reactions were initiated by the addition of 15 μl of an NADPH-generating system (final concentrations (in reaction) of 10 mM glucose 6-phosphate, 1 mM NADP+, and 2 μg ml−1 yeast glucose 6-phosphate dehydrogenase). After reaction for 2 min at 37 °C, reactions were quenched by the addition of 50 μl of CH3OH and chilled on ice.

A 50-μl aliquot of each reaction was spotted directly to the loading zone of a Whatman LK5DF silica gel TLC plate (20 cm × 20 cm). Authentic (unlabeled) standards of progesterone and 11-deoxycorticosterone were added to some of the lanes. The plates were dried in air in a fume hood (with the door open 15 cm to generate a draft), and the plates were developed in CH3OH:ethyl acetate (4:1, v/v), dried in the same way as described above, and visualized under 254-nm UV light. The zones containing 11-deoxycorticosterone and progesterone were removed by scraping (from individual lanes), and 1.0 ml of CH3OH was added to each (to extract the steroids), followed by 5.0 ml of ScintiVerse II liquid scintillation mixture (Fisher). Radioactivity was quantified using a Beckman LS6500 scintillation counter, and recovery was based upon [14C]progesterone samples eluted from the same plates.

The parameters kcat and km were derived from plots of reaction velocity versus substrate concentration fit to hyperbola using non-linear regression analysis (GraphPad Prism 5.0c, GraphPad, San Diego, CA).

Spectral Analyses of CYP21A2—Steady-state binding experiments were made using UV-visible spectroscopy using an Aminco 2Wa-OLIS spectrophotometer (On-Line Instrument Co., Bogart, GA). Both cuvettes (1.0-ml) contained 2.0 μM P450 21A2 in 50 mM potassium phosphate buffer. Aliquots (0.5 ml) of 400 μM progesterone or 21-hydroxyprogesterone (in CH3OH) were added to the sample cuvette, with an equivalent amount of ethanol added to the reference cuvette. Spectra (350–500 nm) were recorded at 23 °C. Data points were fit to a two-site model, with correction for enzyme depletion, using the program DynaFit (supplemental Fig. S2).

Pre-steady-state binding kinetics were recorded at 37 °C with an OLIS RSM-1000 stopped-flow spectrophotometer, operating in the full scanning mode (collecting spectra from 310 to 530 nm). One syringe contained 4.0 μM P450 21A2 in 50 mM potassium phosphate buffer (pH 7.4) containing 100 mM NaCl, 0.1 mM dithiothreitol, and 0.1 mM EDTA. The other syringe contained 40 μM progesterone or 17-hydroxyprogesterone in the same buffer. Spectral (scanning) data were collected for 20 s, and the kinetic traces at 390 and 418 nm (20-nm scan width, single variable decomposition) were fit to biexponential equations using the manufacturer’s software.

Crystalization, Data Collection, and Structure Determination—0.2 mM CYP21A2 (C3B21RA) in buffer D was co-crystallized with 17OHP (0.4 mM, containing 2% (v/v) CH3OH and 17-deoxycorticosterone) grown by the hanging drop vapor diffusion method using 5–15% (w/v) polyethylene glycol 3350, 0.5 M ammonium sulfate, and 0.1 M HEPES (pH 7.0) as the mother liquor. At 20 °C, the ferric 17OHP-bound crystals appeared within a few days. Cryoprotection was achieved by soaking the crystals in 30% (v/v) ethylene glycol prior to flash freezing in liquid nitrogen.

All diffraction data were collected at 100 K at the Southeast Regional Collaborative Access Team (SER-CAT) 21-ID beamline and the Life Science Collaborative Access Team (LS-CAT) beamline, both at the Advanced Photon Source, Argonne National Laboratory. The x-ray data were processed and scaled with the HKL package programs HKL2000 (19). The crystals belong to the monoclinic space group P21, with the following unit cell parameters: a = 67.87 Å, b = 167.99 Å, c = 111.84 Å, β = 90.08°. The resolution is 3.38 Å with a 2.0 Å/σ cut-off, which is extended to 3 Å due to the completeness over 90% in the highest resolution shell. The structure was determined by molecular replacement using the program PHASER (20) and the rabbit CYP2C5 structure as a search model because CYP2C5 has progesterone 21-hydroxylase activity and 30% sequence identity to bovine CYP21A2. The initial model was built in COOT (21), and refinement was performed using CNS1.3 (22) in space group P21212, and CNS parameter and topology files were generated by PRODRG (23). There were two molecules of CYP21A2 in the asymmetric unit. However, the refinement did not progress well in space group P21212, the R-free did not improve substantially beyond 38% at 3.0 Å resolution, and the Ramachandran plot diverged during the refinement.

The analysis of intensity statistics using CNS1.3 (22) and twinning server (24) revealed that the data sets are affected by perfect hemihedral twinning. The proper space group for CYP21A2 crystals is P21 rather than P21212. The twinning is probably coupled with non-crystallographic symmetry in the P21 space group and indicates there are four CYP21A2 molecules in the asymmetric unit. The data were reprocessed in the P21 space group, and refinement was performed against intensities using the protocols for twinned data (twinning operator −h,k,−l), a fixed twin fraction of 0.5. The R-free dropped to 33% after the rigid body refinement in the P21 space group at 3.0 Å resolution. After several rounds of positional, B-factor, and slow annealing refinements alternated with manual model building, the water molecules were added automatically in CNS. Finally, the non-crystallographic symmetry restraint was removed, and with several rounds of refinement, the R values were reduced to 28.3% (R-work) and 29.6% (R-free) at 3.0 Å. Final refinement statistics are given in Table 1. All structural figures were prepared with the program PyMOL (25).

RESULTS

Substrate Binding and Catalytic Activity of CYP21A2—Wild-type bovine CYP21A2 has been efficiently expressed in E. coli (16). At lower concentrations, the protein showed the expected catalytic activity, which suggests that the protein is stable and properly folded. However, the poor solubility of the protein disrupted attempts at crystallization. To enhance the protein solubility (26), we engineered hydrophobic residues exposed to solvent, based on multiple mammalian CYP21A2 sequences, including human and bovine (supplemental Fig. S1). The greatly improved solubility properties of the mutant containing the T241R and L442A substitutions led to successful crystallization. In order to test whether C3B21RA has the same functional activity as the wild-type protein, both titration and enzy-
Structure of P450 21A2 and Disease Implications

TABLE 1
Data collection and refinement of CYP21A2 (C3B21RA) structure (17OHP-bound CYP21A2 complex)

| Data collection statistics | P2₁ |
|----------------------------|-----|
| Space group                |     |
| Unit cell constants        |     |
| Wavelength (Å)             | a = 67.87 Å, b = 167.99 Å, c = 111.84 Å, β = 90.08° |
| Total observations         | 0.97 |
| Unique reflections         | 183,838 |
| Completeness (%)           | 47,485 |
| l/σ(l)                     | 95.1 (91.9) |
| Rmerge (%)                 | 4.5 (1.2) |

| Refinement statistics      |     |
|----------------------------|-----|
| Resolution range (Å)       | 30–3.0 |
| No. of reflections used in refinement | 47,168 |
| No. of water molecules     | 100 |
| Protein atoms⁶              | 14,144 |
| Heme atoms                 | 172² |
| Ligand atoms               | 112¹⁴ |
| Rmerge/Rmerge (%)           | 28.3/29.6 |
| Root mean square deviation, bond lengths (Å) | 0.01 |
| Root mean square deviation, bond angles (degrees) | 1.63 |

⁶ Values for the highest resolution shell are shown in parentheses.
¹ Residues 29–130/133–268/274–409/413–483 (A), 29–130/133–263/274–326/332–483 (B), 29–130/133–263/274–326/332–482 (C), 29–130/134–263/274–326/333–481 (D).
² Four heme molecules.
¹⁴ Eight 17OHP molecules.

matic activities were measured. For substrate binding, both biological substrates, progesterone and 17OHP, produced a typical type I P450 binding spectrum in which the Soret band shifted from 418 to 390 nm upon binding, indicating that progesterone and 17OHP can access the heme in the active site. The kcat and Kₘ values determined for the C3B21RA mutant were 13 ± 2 min⁻¹ and 2.4 ± 0.9 μM, respectively. These values are slightly lower than the corresponding values measured for the truncated wild-type protein (24 ± 3 min⁻¹, 1.7 ± 0.7 μM), previous measurements with the unmutated protein (kcat 13 ± 6 min⁻¹, Kₘ 1.9 ± 0.9 μM), and earlier preparations isolated from bovine adrenals (16).

Overall Structure of Bovine CYP21A2—In order to understand the structural relationship with CAH, an engineered CYP21A2 (C3B21RA) structure complex with 17OHP was determined at a maximum resolution of 3.0 Å. The overall structure of the engineered CYP21A2 exhibits the typical P450 fold consisting of α-helical and β sheet domains as seen in all other known P450 structures (Fig. 2). In detail, the tertiary structure consists of 12 major α-helices, which are similar to those found in most other P450s. However, there are only two groups of sheets with six β-strands in the structure instead of four groups of sheets with 10 β-strands as in most of the other P450s (27, 28). The β sheets in CYP21A2 only include a mixed four-stranded sheet 1 (β1, 58–62; β2, 69–73, β3, 364–366; and β4, 381–386) close to the N-terminal region and two anti-parallel strands in sheet 2 (β5, 451–453; β6, 475–478) at the C terminus. This feature indicates that there are more loop structural elements in CYP21A2 than in many of the other P450 structures. For example, the loop structural residues in CYP21A2 are about 52% in total structural residues compared with 45% loop residues in the CYP11A1 structure (29). The other interesting feature is a kink present in each of the B, E, and J helices. One example is the hydrogen bond between the side chain of Lys-75 in the N terminus of the B helix and the hydroxyl of Ser-416 in the meander region (30). Another hydrogen bond between the side chain of Glu-80 in the C terminus of the B helix and the carbonyl of Ser-371 in a β loop (β3 and β4) together cause stretching, which breaks the B helix into two pieces. These hydrogen bonds also connect the meander region and β sheet 1 through the B helix. Another example is that three residues (Thr-160, Gln-162, and Glu-164) located in the N terminus of the E helix form hydrogen bonds with Gln-477 in β6, His-190 in the F helix, and Ser-300 in the I helix, resulting in a kink in the E helix. In addition, there is a break in the J helix. The side chains of Arg-315 and Gln-317 in the N terminus of the J helix engage in hydrogen bonds with the side chains of Asp-406 in the meander region and Gln-477 in β6. The side chain of Asp-321 in the C terminus of the J helix hydrogen bonds with the side chain of Arg-483 in β6. Thus, these interactions between the meander region, the J helix, and β6 maintain the secondary structural elements in the correct positions to keep the proper local folding of the CYP21A2. Also, the J helix is stretched into two shorter helices to meet the overall structural requirements. Arg-483 and Gln-481 can launch an extensive

FIGURE 2. Ribbon diagram of the 17OHP-bound engineered CYP21A2 (C3B21RA) structure. The heme and the two 17OHP molecules are colored green in stick models. The F helix is stretched into two short helices, and the F/G helices swing away from the structural core.
hydrogen-bonding network in the region of the C terminus and extend the network to the J helix and the meander region, which may stabilize the local structure, including the C terminus and the meander region.

Another feature in the secondary structure of CYP21A2 is that the F helix is twisted into two different helices, which are observed in only a few other P450s (e.g. human CYP51 (31)) although the F’ helix in CYP51 is not associated with the ligand. The F’ helix with the F’G loop swings away from the axis of the F helix and is bent toward the B’ helix. As discussed below, this region of the structure contributes significantly to substrate binding. In addition, the N-terminal residues 29–33 form close contacts with residues 373–376, which are located on a β-hairpin loop between β3 and β4. Although there are no hydrogen bonds among these residues, the distance between the backbone amide nitrogen of Leu-30 and the side chain of Tyr-375 is 3.5 Å, and the distance between the carbon of Pro-31 and the side chain of Tyr-375 is 2.9 Å, which conserves close contacts between the N-terminal loop and the β-hairpin loop (β3 and β4). In general, the secondary structural elements in the engineered CYP21A2 are well defined among loops, sheets, and helices. They stabilize each other through hydrogen bonding networks and van der Waals interactions. In addition, mutant T241R is located on the C-terminal end of the G helix, and the side chain of the arginine substitution is exposed to solvent. L442A is at the end of the L helix, and there is space left for the leucine. Thus, the original amino acids from the wild type would not be expected to affect the structure of the engineered protein.

Active Site Cavity—The most striking feature of the tertiary CYP21A2 structure is that there are two molecules of 17OHP present in the active site of each of the four CYP21A2 structures per crystallographic asymmetric unit (Fig. 3A). The electron density maps in the region of the active site clearly reveal two bound substrate molecules (Fig. 4). The overall shape of the electron densities agrees with the structure of substrate added in the co-crystallization experiments. One 17OHP ligand is situated in the proximal portion of the CYP21A2 active site over the heme; the other 17OHP is bound at the substrate entrance located between loops β1/β2 and β3/β4 and the F’ helix in a cleft on the distal surface of the protein. This distal 17OHP is partially exposed to the bulk solvent and is ∼10 Å away from the proximal substrate molecule. The superimposition of the four molecules (Fig. 3B) indicates that all four proximal 17OHPs are in the same orientation and that the substrates are stable in the active site of the enzyme. The four distal 17OHP molecules show a small variation in the binding modes in the substrate access channel because the corresponding secondary structures, such as β1/β2 and β3/β4 loops and F’ helices, are located in slightly different positions. This may suggest that substrate entry structural elements are more dynamic and that the distal substrates may be in different binding stages. The two bound molecules in one enzyme clearly define the substrate access path from the protein surface into the buried active site (Fig. 5). The proximal 17OHP lies over the heme plane and forms an angle of about 45°, thereby tilting toward the G helix. The 3-keto oxygen of the A-ring in the proximal 17OHP establishes a hydrogen bond (2.4 Å) with the guanidinium group of the Arg-232 side chain, located in the middle of the G helix. Thus, the substrate is anchored by the G helix at the A-ring, thereby positioning the carbon of the C21 methyl group over

FIGURE 3. Evidence for the presence of four molecules present in the asymmetric unit cell. A, the four CYP21A2 structures are colored in green, yellow, blue, and orange. The two molecules of 17OHP are seen in cyan, pink, green, and yellow as a sphere model. Heme is in a red stick model. B, superimposed overall structures of 17OHP-bound CYP21A2 as a backbone stick model. All four proximal 17OHP molecules are in the same orientation, and the locations of the four distal 17OHP molecules show small variations in binding modes in the substrate access channel.

FIGURE 4. Electron density for 17OHP in the structure of 17OHP-bound CYP21A2. The electron density map was calculated using A-weighted 2Fo – Fo coefficients and is contoured at 1.0 s. The proximal 17OHP and distal 17OHP molecules in the green stick models (left) are denoted S1 and S2, respectively. Potential hydrogen bonds and the distances between the corresponding carbon and oxygen atoms to the iron atom are shown as dotted lines (right).
The heme iron at a distance of 4.4 Å. Furthermore, the 20-keto oxygen of the substrate is situated at a distance of 4.6 Å from the heme iron, which may allow a hydrogen bonding interaction with the oxygen atom to help binding or proton transfer and dioxygen scission during the CYP21A2 catalytic reaction cycle (32, 33). Thus, the 20-keto group might be part of a pharmacophore of the substrates for their biological activities.

Surprisingly, the distal 17OHP in CYP21A2 is bound at the substrate entrance. The average crystallographic temperature factors of the proximal and distal 17OHP molecules in CYP21A2 are 64.8 and 72.5 Å³ compared with 56.4 Å³ for heme factors of the proximal and distal 17OHP molecules in CYP21A2 structures, further indicating that each 17OHP substrate molecule in the four CYP21A2 structures occupies a slightly different position (Fig. 3).

The proximal substrate binding pocket is defined by several hydrophobic residues over the heme. The residues within 4 Å around the proximal 17OHP create a functional catalytic pocket. At the end of carbon C21 of 17OHP, Thr-294 in the F helix and put C21 in a proper position in the reaction center. The distal 17OHP binding region is composed of Leu-64, Gly-65, and Leu-66 in the β1/β2 loop, Ile-95 in the B-B’ loop, Gln-206 and Met-210 in the F’ helix, Leu-360 and Ala-361 in the K-β4 loop, and Cys-467 and Gly-468 in the β hairpin of β5/β6, all of which contact with the distal 17OHP molecule within 4 Å. These residues form a cavity in the substrate access channel (Fig. 5). It is worth noting that the A-ring of the distal 17OHP sits deep inside the channel, with the other end of the structure being partially exposed to solvent. In addition, the 3-keto oxygen forms hydrogen bonds with the carbonyl oxygen of Ala-361 at 3.3 Å and the carbonyl oxygen of Gly-468 at 3.6 Å. These interactions may indicate increased movement of the corresponding structural elements and also indicate that this cavity is triggered upon ligand binding to select the ligand on the distal surface of the protein. Furthermore, the weak hydrogen bonding contacts between the distal 17OHP and Ala-361/Gly-468 could be of importance for distal 17OHP binding and release.

Spectral Measurements—The structures provide evidence for the presence of two ligands in the active site (i.e. the substrate 17OHP). We carried out spectroscopic assays to provide additional evidence for the presence of multiple ligands.

Titration of CYP21A2 with 17OHP provided a plot that, upon visual inspection, showed very tight binding of a single molecule of substrate (to 2 µM), followed by a further increase in the heme spectral changes (Fig. 6). Fitting required very tight binding of a single molecule of the first molecule (Kd ≈ 50 nM), with a second Kd of 0.7 µM. The fitting is not ideal in the absence of information about what the actual extinction coefficient is. In the case of progesterone (not shown), the break following binding of one ligand was not clear, and the results could be explained with a simple hyperbolic, unimolecular binding isotherm, corrected for the enzyme concentration (Kd = 0.40 ± 0.07 µM), although a more complex relationship may exist.

Pre-steady-state binding also showed multiphasic behavior (Fig. 7). The spectral changes observed upon mixing of 20 µM (10-fold excess over CYP21A2) were relatively slow for the
malian CYP21A2 shows that >85% of the human CAH mutants occur in highly conserved residues. Thus, the conserved residues are expected to be crucial in maintaining the enzyme structure and function among species during the evolution. In addition, the locations of mutants are distributed throughout the structure, from the N-terminal region to the C-terminal loop, on all structural elements through loop regions, β sheets, and all α helices except the A and F’ helices. Helices I and G show the most hits among helical elements. About 55% of the mutants are located in loop structures, and many severe mutations are located in the meander region.

As noted above, the most intriguing result from this structure is that there are two bound molecules of 17OHP and therefore two distinct binding cavities located in the protein (Fig. 5). One is on the surface with the substrate access channel for the distal 17OHP and the other is buried into the heme pocket for the proximal 17OHP. There is a narrow U-shaped bottle neck channel linked between these two cavities. Both the structural and spectral/kinetic data may suggest that 17OHP could occupy two positions in the active site and that the distal 17OHP might be bifunctional in ensuring that substrate is bound to the distal binding site, subsequently dissociating and moving into the active site. Thus, it may help close the active site when the proximal substrate triggers the conformational changes.

Another interesting feature of CYP21A2 is that the geometries of the proximal 17OHP substrates in CYP21A2 are different from those in CYP11A1 (29). In CYP21A2, the A-ring end of the substrate is pointed toward the G helix. In contrast, the long side chain of the cholesterol in CYP11A1 is located above the heme in the reaction center, and the steroid backbone remains in the substrate access channel. In CYP11A1, the substrate access and product release probably share the same substrate access channel. However, CYP21A2 may have a different path for product exit because of the substrate orientation in the active site and the distal substrate blocking the channel. This raises the question of how the product would release from the CYP21A2 active site. An opening is formed by the N terminus of the I helix with the B’ and G helix. This leads to a channel that exits to the exterior of the protein surface. The Met-282, Val-285, and Asp-286 residues in the I helix, residues Ile-229, Arg-232, and Asp-233 in the G helix, and Val-101 and Gln-103 in the B’ helix line the channel. It is possible that this channel, although too narrow at certain points to let steroid products pass through, permits the passage of the product when the flexibility of the surrounding tertiary structure is taken into account.

**DISCUSSION**

CYP21A2 is the cytochrome P450 with the largest number of naturally occurring mutants involved in steroidogenic hormone biosynthesis and is directly responsible for the largest number of cases of CAH (3, 5, 6). In general, the severity of the disease correlates well with the level of CYP21A2 enzymatic deficiency (35). The analysis of primary sequences of all mam-

---

**FIGURE 7. Pre-steady-state binding of progesterone and 17OHP to CYP21A2.** Spectra were recorded in an OLIS-RSM 1000 instrument at 37 °C. One syringe contained 4 μM P450 21A2 in 50 mM potassium phosphate buffer (pH 7.4), 100 mM NaCl, 0.1 mM dithiothreitol, and 0.1 mM EDTA. The other syringe contained 40 μM progesterone or 17OHP in the same buffer. The components (equal volumes) were mixed, and spectra were acquired every 1 ms. Some of the early traces and the final trace (16 s) are shown in A (fully changed). A control data points were fit to biexponential plots using the manufacturer’s software (GlobalWorks). B, progesterone binding. k1 = 1.8 ± 0.7 s⁻¹, k2 = 0.66 ± 0.06 s⁻¹. C, 17OHP binding. k1 = 2.5 ± 0.3 s⁻¹, k2 = 0.26 ± 0.10 s⁻¹. Rapid phase (1.8 and 2.5 s⁻¹ for progesterone and 17OHP, respectively), followed by a slower phase with a rate 1 order of magnitude slower. The faster rate is considerably slower than expected for a diffusion-controlled process (i.e. 2 s⁻¹/20 μM = 10⁸ M⁻¹ s⁻¹). The slow, multiphasic binding is consistent with the multisite binding of ligands to CYP21A2, although this observation alone cannot be considered a proof.
Structure of P450 21A2 and Disease Implications

enzyme activities. The mutants cause CAH through impairment of the CYP21A2 catalytic activity in multiple ways, such as interruption of substrate binding, breaking the proton relay system, inhibiting NADPH-P450 reductase interaction, affecting heme ligation, and reducing protein structure solubility, stability for substrate oxidation, and product release. All of these key functional residues are located in different secondary structural elements and generally cannot be deduced from homology modeling. Examples include the following.

**Substrate Access**—Gly-65 is located in the β1/β2 loop and is associated with the distal 17OHP (Fig. 8A). In a van der Waals model, the backbone carbon from Gly-65 is the only possible amino acid to fit the narrow substrate access opening. The long side chain of the variant G65E may block the substrate.

**Substrate Binding**—The α face of the proximal 17OHP is directed toward the I helix, and the 17α hydroxyl points into Gly-290 and Gly-291 located in the middle of the I helix so that a helical kink is created. Thus, human variants G291C, G291S, and G292D might have altered substrate binding in the active site in such a way as to eliminate activity. In addition, Tyr-98 is located in the B' helix, and its aromatic side chain points to the active site. Tyr-98 is about 3.8 Å away from the side chain of Trp-200, which is involved in the proximal binding site. Tyr-98 may play a role in helping to close the active site and assist Trp-200 to inhibit movement of the solvent molecule into the heme pocket. Therefore, it is not surprising that Tyr-98 mutated to hydrophilic residues decreased the activity significantly (36).

**Proton Transfer**—Thr-294, in the I helix, is a highly conserved residue found in most P450s. The mutant T294N probably disrupts substrate binding because the hydroxyl group in Thr-294 is only 2.4 Å away from the C21 methyl group of the proximal 17OHP. Also, the asparagine might break the proton transfer path during the catalytic cycle (34).

**Protein Solubility**—Gln-227 is located at the third turn of the G helix. This residue is identified as a surface residue, and its side chain is exposed to solvent. In the structure, two water molecules (WAT90 and WAT98) could be nearby. This suggests that Gln-227 might change the protein hydrophilicity if changed to a nonpolar residue and lose its function as a structural stabilizer in CYP21A2.

**Heme Binding**—a common feature of all P450s is that the P450 folding and activity requires heme as prosthetic group. Several residues in CYP21A2 participate in heme binding. The heme group makes ionic and hydrogen bonding interactions (via its propionate moieties) with side chains of Arg-92, Arg-425, His-364, Trp-117, and Ser-109. The mutation R425H is likely to affect the heme ligation with the protein. In addition, both Ile-108 and Ala-361 are directed toward the heme at a distance 3.6 Å. The bulky side chain mutants for I108R and A361V may sterically clash with the heme, thus pushing it in the

---

**TABLE 2**

Proposed structural basis of disease-causing mutants of CAH

| Human        | Bovine | Assigned structural basis | Location   |
|--------------|--------|---------------------------|------------|
| P30Q         | P31Q   | Stability                 | N-terminal loop |
| K54X         | K55X   | Stability                 | AB loop    |
| G64E         | G65E   | Substrate                 | B1β2 loop  |
| K74X         | K75X   | Reductase binding         | B helix    |
| G90V         | G91V   | Stability                 | BB' loop   |
| Y97X         | Y98X   | Solubility                | B' helix   |
| L107R        | L108R  | Heme                      | B' C loop  |
| V139E        | V140E  | Stability                 | D helix    |
| L142P        | L143P  | Stability                 | D helix    |
| L167P        | L168P  | Stability                 | E helix    |
| Q228X        | Q227X  | Solubility                | G helix    |
| L261P        | L260P  | Stability                 | H helix    |
| Q262X        | Q261X  | Stability                 | H I loop   |
| G291C        | G290C  | Substrate                 | I helix    |
| G291S        | G290S  | Substrate                 | I helix    |
| G292D        | G291D  | Substrate                 | I helix    |
| T295N        | T294N  | Proton transfer           | I helix    |
| W302R        | W301R  | Stability                 | I helix    |
| R316X        | R315X  | Stability                 | J helix    |
| L353R        | L352R  | Stability                 | K helix    |
| R354C        | R353C  | EXXR                      | Kβ4 loop   |
| R354H        | R353H  | EXXR                      | Kβ4 loop   |
| R356P        | R355P  | Stability                 | Kβ4 loop   |
| R356W        | R355W  | Stability                 | Kβ4 loop   |
| A362V        | A361V  | Heme                      | Kβ4 loop   |
| G375S        | G374S  | Substrate                 | Bβ34 loop  |
| Y376X        | Y375X  | Stability                 | Bβ34 loop  |
| E380D        | E379D  | Stability                 | Bβ34 loop  |
| F404S        | F403S  | Stability                 | KL loop    |
| W405X        | W404X  | Stability                 | KL loop    |
| R408C        | R407C  | Stability                 | KL loop    |
| R426H        | R425H  | Heme                      | KL loop    |
| L446P        | L445P  | Stability                 | L helix    |
| T450P        | T449P  | Substrate                 | Lβ6 loop   |
| Q481P, P482S | Q481P, P482S | Stability | C terminus |
| R483W        | R483W  | Stability                 | C terminus |

---

**FIGURE 8.** Detailed local structural features of CYP21A2. A, Gly-65 (in a van der Waals surface model) is located at the β-hairpin turn and is less than 4 Å from the distal substrate. It serves as a gating residue at the mouth of the substrate access channel; B, Gly-379 is directed toward the surface of the protein and at the loop between β1-3 and β1-4. The carbonyl group of Glu-379 can form a hydrogen bond with the oxygen of Thr-367, thus locking the two β sheets in the structure. Thr-367 makes an additional hydrogen bond with Asp-88, which is in the loop near the B helix. These hydrogen bond interactions play an important role in maintaining the structural fold. Dotted lines show the presumed hydrogen bonding network.
opposite direction so that heme may not bind to the protein properly or not at all.

**EXXR Structural Motif**—The EXXR motif is highly conserved in essentially all P450 sequences (15) and is located in the K helix. The Glu and Arg and residues from the meander region can form a set of salt bridge/hydrogen bonding interactions that participate in the formation of the tertiary structure. There is evidence that suggests EXXR is linked somehow to heme association with the P450 polypeptide (30, 37). In CYP21A2, Glu-350 and Arg-353 of the EXXR motif show standard ion pair interaction and also make hydrogen bonding interactions with residues Pro-400, Glu-402, Arg-404, Thr-347, and His-391 from the meander region. Substitution of Arg-353 to either histidine or cysteine in this position could disrupt the naturally occurring interactions with Glu-350, Pro-400, and His-391, essential for maintaining the folded structure.

**Structural Stability**—Some residues can extend the hydrogen bonding network or other molecular interactions between secondary structural elements. These residues, like a bridge, play important roles in properly integrating the secondary structural elements to stabilize the scaffold of the protein. The side chain of Glu-379 is directed toward the surface of the protein at the β4 loop (Fig. 8B). The carboxylic acid of Glu-379 can form a hydrogen bond with the carbonyl oxygen of Thr-367 located at the β3 loop, which locks the β3 and β4 sheets. Thr-367 extends hydrogen bonding interaction with Asp-88 at the BB’ loop via its backbone amide nitrogen. Thus, the three loops lie together via Glu-379, Thr-367, and Asp-88. The side chain of Glu-379 mutated to aspartic acid has one less carbon so that this mutation could break the interaction with Thr-367 to interrupt the β3, β4, and BB’ loop structure. Another example is that both Arg-483 and Gln-481 lie in β6 at the C terminus of the protein. The side chain Arg-483 can make a hydrogen bond with Asp-321 in the J helix (Fig. 9A). Gln-481 is able to make hydrogen bond connections with Gln-317 in the J helix and Gln-446 in the Lβ6 loop. Arg-315 in the J helix can connect with Asp-406 in the meander region through a hydrogen bond. These hydrogen bond networks may connect the C terminus, J helix, and meander region to maintain the structural elements in the correct positions. Any mutation on these residues probably results in the interruption of the networks. Thus, the human mutants R483W and Q481P/P482S lead to CAH. A third example of an interaction that may affect protein stability concerns the two aromatic rings of Trp-301 and Phe-305 from the C-terminal end of the I helix, which form π-electron stacking interactions with the side chain of Phe-403 in the meander region (these side chains all lie within 5 Å from each other; Fig. 9B). This structural feature stabilizes the meander region and the I helix. The stacking interactions require all three aromatic side chains. Replacement of any one of these amino acids by a non-aromatic residue will probably interrupt these interactions. This is the case for mutants W301R and F403S, both of which are clinically associated with CAH (38). Based on multiple sequence alignments, Trp-302 and Phe-403 are strictly conserved in all CYP21A2 sequences. Phe-305 exists in all mammals except in swine, where tyrosine is present, a residue that also features an aromatic side chain that may play an important role in bridging Trp-301 and Phe-402. It is interesting that aromatic residues at 301 and 402 in CYP21A2 can also be found at the corresponding residue in some other P450 enzymes, such as CYP2C5 (Tyr-305 and Phe-404), CYP11A1 (Trp-298 and Phe-396), and CYP19A1 (Phe-317 and Phe-413). What is not known, however, is whether non-aromatic substitution of these corresponding residues in other P450s would significantly reduce the catalytic activities and therefore might play similar roles as in CYP21A2.

**NADPH-P450 Reductase Binding**—Electron transfer through redox proteins is essential for P450 activity. It is believed that basic residues on the proximal face of P450s interact with acidic residues on the surface of the redox protein (39). Mutation of a cluster of basic residues on the proximal surface of CYP21A2 results in CAH. The basic residues consist of Arg-338, Arg-340, Lys-74, Arg-407, and Arg-368 and may be involved in a redox protein interaction according to the P450BM3 structure (40). Perhaps mutations of these residues can cause CAH by an inability to interact correctly with the redox partner. In summary, the crystal structure of CYP21A2 not only offers valuable information about CAH-associated residues in CYP21A2 and the etiology of CAH but also offers an enhanced understanding of the detailed structure/function relationships of P450s.
Structure of P450 21A2 and Disease Implications

Acknowledgments—We thank Dr. Martin Egli for reading the manuscript and the personnel at the SER-CAT 22-BM and LS-CAT beamlines at APS/Argonne National Laboratory for technical assistance, M. V. Martin and L. M. Folkmann for preparation of NADPH-P450 reductase, and K. Trisler for help in the preparation of the manuscript.

REFERENCES

1. Kagawa, N., and Waterman, M. R. (1991) Evidence that an adrenal-specific nuclear protein regulates the cAMP responsiveness of the human CYP21B (P450C21) gene. J. Biol. Chem. 266, 11199–11204
2. Kagawa, N., and Waterman, M. R. (1990) cAMP-dependent transcription of the human CYP21B (P-450C21) gene requires a cis-regulatory element distinct from the consensus cAMP-regulatory element. J. Biol. Chem. 265, 11299–11305
3. White, P. C., and Speiser, P. W. (2000) Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Clin. Chem. 46, 330–339
4. White, P. C., Waterman, M. R., and Dupont, B. (1990) cAMP-dependent transcription of human 21-hydroxylase genes. Proc. Natl. Acad. Sci. U.S.A. 87, 5111–5115
5. Kagawa, N., and Waterman, M. R. (1990) cAMP-dependent transcription of human 21-hydroxylase genes. Proc. Natl. Acad. Sci. U.S.A. 87, 5111–5115
6. White, P. C., Grossberger, D., Onufrej, B. J., Chaplin, D. D., New, M. L., Dupont, B., and Strominger, J. L. (1985) Two genes encoding steroid 21-hydroxylase are located near the genes encoding the fourth component of complement in man. Proc. Natl. Acad. Sci. U.S.A. 82, 1089–1093
7. Igashia, Y., Tanae, A., Inoue, H., Hiromasa, T., and Fuji-Kuriyama, Y. (1988) aberrant splicing and missense mutations cause steroid 17α-hydroxylase (P-450C21) deficiency in humans. Possible gene conversion products. Proc. Natl. Acad. Sci. U.S.A. 85, 7486–7490
8. Krone, N., and Arlt, W. (2009) Genetics of congenital adrenal hyperplasia. Best Pract. Res. Clin. Endocrinol. Metab. 23, 181–192
9. Robins, T., Carlsson, J., Sunnerhagen, M., Wedell, A., and Persson, B. (2006) Molecular mechanism of human CYP21 based on mammalian CYP2C5. Structural features correlate with clinical severity of mutations causing congenital adrenal hyperplasia. Mol. Endocrinol. 20, 2946–2964
10. Mornet, E., and Gibrat, J. F. (2000) A 3D model of human P450c21. Study of the putative effects of steroid 21-hydroxylase gene mutations. Hum. Genet. 106, 330–339
11. Tauber, A., Kellogg, M., and Majzoub, J. A. (2010) Monitoring of therapy in congenital adrenal hyperplasia. Clin. Chem. 56, 1245–1251
12. Concolino, P., Mello, E., Zuppi, C., and Capoluongo, E. (2010) Molecular diagnosis of congenital adrenal hyperplasia due to 21-hydroxylase deficiency. An update of new CYP21A2 mutations. Clin. Chem. Lab. Med. 48, 1057–1062
13. Shaw, A. M. (2010) 21-hydroxylase deficiency congenital adrenal hyperplasia. Neonatal Netw. 29, 191–196
14. Wedell, A. (2011) Molecular genetics of 21-hydroxylase deficiency. Endocr. Dev. 20, 80–87
15. Rupasinghe, S., Schuler, M. A., Kagawa, N., Yuan, H., Lei, L., Zhao, B., Kelly, S. L., Waterman, M. R., and Lamb, D. C. (2006) The cytochrome P450 gene family CYP157 does not contain EXXR in the K-helix reducing the absolute conserved P450 residues to a single cysteine. FEBS Lett. 580, 6338–6342
16. Arase, M., Waterman, M. R., and Kagawa, N. (2006) Purification and characterization of bovine steroid 21-hydroxylase (P450c21) efficiently expressed in Escherichia coli. Biochem. Biophys. Res. Commun. 344, 400–405
17. Shen, A. L., Porter, T. D., Wilson, T. E., and Kasper, C. B. (1989) Structural analysis of the FMN binding domain of NADPH-cytochrome P-450 oxi-doreductase by site-directed mutagenesis. J. Biol. Chem. 264, 7584–7589
18. Hanno, I. H., Reed, J. R., Guengerich, F. P., and Hollenberg, P. F. (2000) Expression of human cytochrome P450 286 in Escherichia coli. Characterization of catalytic activity and expression levels in human liver. Arch. Biochem. Biophys. 376, 206–216
19. Otrwiowski, Z., Borek, D., Majewski, W., and Minor, W. (2003) Multiparametric scaling of diffraction intensities. Acta Crystallogr. A 59, 228–234