Published in final edited form as:
Nature. 2009 January 8; 457(7226): 219–223. doi:10.1038/nature07614.

Structural basis for androgen specificity and oestrogen synthesis in human aromatase

Debashis Ghosh[1,2], **Jennifer Griswold**[1], **Mary Erman**[1], and **Walter Pangborn**[1]

[1] Structural Biology, Hauptman-Woodward Medical Research Institute, 700 Ellicott Street, Buffalo, New York 14203, USA

[2] Pharmacology and Therapeutics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, New York 14263, USA

Abstract

Aromatase cytochrome P450 is the only enzyme in vertebrates known to catalyse the biosynthesis of all oestrogens from androgens[1–3]. Aromatase inhibitors therefore constitute a front-line therapy for oestrogen-dependent breast cancer[3,4]. In a three-step process, each step requiring 1 mol of O_2, 1 mol of NADPH, and coupling with its redox partner cytochrome P450 reductase, aromatase converts androstenedione, testosterone and 16α-hydroxytestosterone to oestrone, 17β-oestradiol and 17β,16α-oestriol, respectively[1–3]. The first two steps are C19-methyl hydroxylation steps, and the third involves the aromatization of the steroid A-ring, unique to aromatase. Whereas most P450s are not highly substrate selective, it is the hallmark androgenic specificity that sets aromatase apart. The structure of this enzyme of the endoplasmic reticulum membrane has remained unknown for decades, hindering elucidation of the biochemical mechanism. Here we present the crystal structure of human placental aromatase, the only natural mammalian, full-length P450 and P450 in hormone biosynthetic pathways to be crystallized so far. Unlike the active sites of many microsomal P450s that metabolize drugs and xenobiotics, aromatase has an androgen-specific cleft that binds the androstenedione molecule snugly. Hydrophobic and polar residues exquisitely complement the steroid backbone. The locations of catalytically important residues shed light on the reaction mechanism. The relative juxtaposition of the hydrophobic amino-terminal region and the opening to the catalytic cleft shows why membrane anchoring is necessary for the lipophilic substrates to gain access to the active site. The molecular basis for the enzyme's androgenic specificity and unique catalytic mechanism can be used for developing next-generation aromatase inhibitors.

Human aromatase is the product of the *CYP19A1* gene on chromosome 15q21.1 and consists of a haem group and a polypeptide chain of 503 amino-acid residues. Although aromatase has been extensively studied for more than 35 years[1–3,5–19], the mechanism of the aromatization step remains poorly understood. Many soluble bacterial P450s, such as P450cam20 and P450eryF21, as well as recombinant human microsomal P450s, such as 3A4 (ref. 22), 2D6 (ref. 23) and 2A6 (ref. 24), that metabolize drug/xenobiotics, have been

Correspondence and requests for materials should be addressed to D.G. (ghosh@hwi.buffalo.edu).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions J.G. and M.E. performed the purification and crystallization of aromatase. W.P. and J.G. contributed to diffraction data collection. D.G. was involved in diffraction data collection and processing. D.G. solved the structure, wrote the manuscript and was responsible for overall planning and supervision of the project.

Author Information Atomic coordinates and structure factor files have been deposited with the Protein Data Bank under the accession code 3EQM. Reprints and permissions information is available at www.nature.com/reprints.

crystallized and studied by X-ray crystallography. Several laboratories have reported the purification of aromatase from human placenta[7,8] and recombinant expression systems[14,18]. Nevertheless, attempts to crystallize either the placental or a recombinant or modified aromatase have been unsuccessful and an experimental aromatase structure has remained unknown. Numerous mechanistic and homology models based on known P450 structures and site-directed mutagenesis data have been proposed[5,6,9–13,15–18], leading to the identification of catalytically important residues and possible substrate-binding modes.

The 2.90-Å resolution crystal structure of aromatase purified from term human placenta[19] in complex with its natural substrate androstenedione (androst-4-ene-3,17-dione) exhibits the characteristic cytochrome P450 fold (Fig. 1a; see Methods, Supplementary Fig. 1 and Supplementary Table 1). Androstenedione binds with its β-face oriented towards the haem group and C19 4.0 Å from the Fe atom (Fig. 1b and Supplementary Fig. 2). To test the catalytic viability of the substrate-binding mode, the haem Fe is modelled as a hypothetical oxyferryl Fe(IV)=O moiety (Fig. 1c). The resulting binding geometry of the C19 methyl hydrogens closely resembles that of the reactants for hydroxylation by P450cam[25]. The residues comprising the catalytic cleft are Ile 305, Ala 306, Asp 309 and Thr 310 from the I-helix, Phe 221 and Trp 224 from the F-helix, Ile 133 and Phe 134 from the B–C loop, Val 370, Leu 372 and Val 373 from the K-helix–β3 loop, Met 374 from β3, and Leu 477 and Ser 478 from the β8–β9 loop (Fig. 1b). The 17-keto oxygen of the substrate makes a hydrogen bond (2.8 Å) with the backbone amide of Met 374 and a weak contact (3.4 Å) with NH1 of Arg 115 (Fig. 1b). The 3-keto oxygen is 2.6 Å from the carboxylate $O_{\delta 2}$ of the Asp 309 side chain (Figs 1b and 2a, b), indicating that the carboxylate moiety may be protonated. The hydrophobic residues and porphyrin rings of haem pack tightly against the steroid backbone, forming a cavity complementary in shape to the bound steroid (Fig. 2a). The side chains of residues Arg 115, Ile 133, Phe 134, Phe 221, Trp 224, Ala 306, Thr 310, Val 370, Val 373, Met 374 and Leu 477 make direct van der Waals contacts with the bound androstenedione. Ile 133, Phe 134, Phe 221, Trp 224 and Leu 477 approach the substrate from the α-face and follow the contour and puckering of the steroid backbone, while the side chains of Arg 115, Ala 306 and Met 374 make contacts at its edge, and Thr 310, Val 370 and Val 373 on the β-face. The combined surface creates a pocket that encloses the bound androstenedione snugly. The volume of the binding pocket is no more than 400 $Å^3$, considerably smaller than the volume of about 530 $Å^3$ of the active sites in 3A4 (ref. 22) and 2D6 (ref. 23), the two drug/xenobiotic-metabolizing human P450s with highest sequence identities (14–18%) to human aromatase.

A distortion in the I-helix backbone resulting in a roughly 3.5-Å displacement of the helix axis is crucial for creating the androgen-specific binding pocket at the active site. This shift in the helix axis caused by Pro 308, a residue unique to aromatase among all P450s, is stabilized by a strong Ala 306CO···HOThr 310 (2.8 Å) hydrogen bond, as well as by an Asp 309 peptide CO···water (3.4 Å) interaction (Fig. 2b). Although an irregularity in the I-helix backbone at this region is observed in other P450s[20–24,26], this Pro 308-mediated axis shift precisely accommodates the 3-keto end of androstenedione near the fifth turn of the helix (Met 303 to Ala 307), allowing the Asp 309 side chain to hydrogen-bond to the 3-keto oxygen (Fig. 2b). Superposition of the aromatase backbone with bacterial P450cam and P450eryF (Supplementary Fig. 3) as well as with human P450 3A4 and P450 2D6 (Supplementary Fig. 4) shows that without this shift, the 3-keto end of the substrate would clash with the I-helices of P450cam, P450eryF, P450 2D6 and P450 3A4. The structure therefore confirms the critical roles of Pro 308 and Asp 309 predicted by mutagenesis and modelling[9,10,15].

The Asp 309 side chain also forms a hydrogen bond with a water molecule 3.0 Å from $O_{\delta 1}$ (Figs 1b and 2a and Supplementary Fig. 5). The geometries of these two hydrogen bonds

place the 3-keto oxygen and the water oxygen atom roughly in the carboxylate plane (Fig. 2b). Furthermore, this water molecule is situated at 3.6 Å from the guanidinium group of the Arg 192 side chain (Fig. 2b), which forms a salt bridge with Glu 483. An elongated electron density adjacent to the Ser 478 side chain (Supplementary Fig. 5) was modelled as two water molecules, hydrogen-bonded to each other and to the Ser 478 side chain OH, which in turn donates a hydrogen bond to His 480 $N_{\delta 1}$ farther from the active site. The Ser 478 side chain is linked through these two water molecules to Arg 192 by a weak hydrogen bond (3.4 Å; Fig. 2b and Supplementary Fig. 5). It is probably as a result of being linked to this network of proton donors on one side, and the keto group of a large hydrophobic substrate on the other, that the Asp 309 side chain remains protonated and engaged in the substrate-binding interaction. This network could also serve as the proton source for the proposed participation of the Asp 309 carboxylate moiety in the enolization process, as discussed below.

H2β of the A-ring of the bound androstenedione (Fig. 3a) that gets abstracted in the aromatization step is close to the Ala 306CO···HOγThr 310 pair (C=O···H2β-C2:3.7 Å and C2-H2β···OγH: 3.8 Å). Thr 310, highly conserved in P450s, has been implicated in the P450 hydroxylation steps. The mechanism of P450 hydroxylation has been studied extensively for two bacterial enzymes, P450cam20 and P450eryF21. In the dioxygen complex of P450cam, the residue pair Thr 252–Gly 248 carbonyl, and two catalytic water molecules, are involved in the activation of ferrous dioxygen to the hydroxylating Fe(IV)=O species by providing two protons20. A similar hydroxylation mechanism involving the corresponding Thr 310–Ala 306 carbonyl pair, and catalytic water molecules (the binding of which could be promoted by dioxygen binding as in P450cam), is probably at work for each of the first two steps catalysed by aromatase, yielding the C19-aldehyde derivative of androstene-dione through 19,19-*gem*-diol formation and retention of the *pro-S* hydrogen[1–3,5,6]. The same catalytic residues could also be responsible for the H2β abstraction of the 2,3-enolization processes in the aromatization step. To accomplish this, a nucleophilic attack on H2β-C by the Ala 306CO···HOγThr 310 moiety (perhaps along with a water) and a concerted electrophilic attack on the C3-keto oxygen by a protonated Asp 309 side chain could promote the H2β abstraction and 2,3-enolization, akin to $-H_2C2-$C3-keto to -HC2=C3-enol tautomerization (Fig. 3b). A bound water molecule (perhaps linked to the proton relay network) between Thr 310-Oγ and the iron-peroxy/hydroperoxy intermediate could lower the pK_a of Thr 310-OγH, rendering Ala 306C=O a more potent nucleophile; alternatively, it could itself act as a nucleophile, as shown in Fig. 3b. Asp 309 thus seems to have a direct participation in enolization, unlike the indirect roles of Asp 251 and Glu 244 in hydroxylation by P450cam and P450eryF, respectively[20,21]. A density-function-theory calculation for the final catalytic step of aromatase suggests a strikingly low energy barrier (less than 7 kcal mol^{-1}) for H1β abstraction when steroids are 2,3-enolized[27]. However, the 1β hydrogen is too far from this carbonyl (6.2 Å) to be abstracted in such a manner. It points at and is close to the haem Fe (4.2 Å), and is probably removed after the Fe-peroxy nucleophilic attack on the 19-aldehyde (Fig. 3b) as previously proposed[3,5,6].

To examine how a mechanism-based steroidal inhibitor could interfere with the aromatization process, an exemestane (Aromasin; one of the three aromatase inhibitors approved by the US Food and Drug Administration) molecule was built into the active site (Fig. 3c) by using the androstenedione backbone. The two steroids superimpose quite well (root mean square (r.m.s.) deviation about 0.2 Å), except for differences in puckering of the A-rings. The extra C6-methylidene group in exemestane is accommodated in a shallow hydrophobic crevice surrounded by the side-chain C atoms Thr 310-Cγ, Val 370-Cγ2 and Ser 478-Cβ, at the mouth of the active-site access channel (described below). The distance between the methylidene C and Cγ-Thr 310 is 3 Å, shorter than the van der Waals contact distance. Indeed, a slight adjustment of these side chains on the binding of exemestane is

highly likely. The clamping of C6-methylidene in a hydrophobic surrounding, resulting in entropic gain and a lowering of the free energy and the dissociation constant, could greatly reduce the mobility of the Thr 310 side chain and/or interfere with its ability to interact with the catalytic waters for the creation of the active oxyferryl moiety. Exemestane would thus remain tightly bound in the pocket without being hydroxylated at C19.

An access channel links the active site to the outer surface. Figure 4a is a view of the interior of a semi-transparent solvent-excluded surface[28] that also excludes the active-site region, consisting of the steroid-binding pocket and haem, from the protein interior by forming a 'pouch'-like cleft that opens only to the exterior through the channel, at the arrowhead. The inset shows a view along this channel, revealing the locations of three water molecules within the channel and a glimpse of the opening to the active-site cavity. The salt-bridging Arg 192–Glu 483 pair as well as Asp 309 and Ser 478 line the channel that hosts the putative proton relay network and is also probably the major transport route to and from the active site for water, oxygen and steroid molecules. This channel seems to be a confluence of what was previously described as channels 2a, 2ac and 2c for other P450s[29]. Although it becomes narrower at points, the channel is probably flexible to permit the passage of molecules such as steroids.

Having seven cysteines in the reduced form, the bulk of aromatase probably resides in the reducing environment of the cytoplasm. A hydrophobicity plot of the aromatase sequence suggests lipid integration for residues 21–42 and 49–71, thereby placing the N terminus with glycosylation at Asn 12 on the opposite side; that is, the lumen[30]. It is possible that the transmembrane segment of residues 21–42, which is too short to traverse the bilayer as a regular α-helix, is at least partly an extended polypeptide devoid of secondary structure. This is consistent with the observed weak electron density for the polypeptide chain beyond residue 45 towards the N terminus. It is also likely that the hydrophobic helix A′ (residues 57–68) and part of helix A (residues 69–80) are embedded in the membrane. This arrangement positions several arginine (Arg 64, Arg 79 and Arg 86) and tryptophan (Trp 67 and Trp 88; Trp 239 from the F–G loop) residues at the lipid/protein interface, a telltale sign of lipid integration. Besides, electron densities for at least two detergent molecules were identified near Trp 67 of helix A′. We therefore propose that lipid integration of aromatase begins with these helices as the N terminus traverses farther into the bilayer towards the lumen side. A model (Fig. 4b) based on these concepts places the entrance to the active-site access channel (Fig. 4a) on the membrane surface. Although other possible entry and exit routes cannot be excluded, this arrangement allows the lipophilic substrate to enter the aromatase active site directly from within the membrane, travelling between and across the F–G loop and the β8–β9 loop, roughly along the path indicated by the white arrow in Fig. 4b. The structure of aromatase therefore provides a rationale for its crucial membrane integration.

The crystal structure of human aromatase reveals a finely tuned molecular machine that makes oestrogens from androgens. Use of the molecular basis for enzyme–substrate and enzyme–drug interactions could lead to more efficacious intervention of oestrogen production.

METHODS SUMMARY

Aromatase was purified from term human placenta by immuno-affinity chromatography in a highly active form. It was complexed with androstenedione and crystallized at 4 °C in the oxidized high-spin ferric state of the haem iron with poly(ethylene glycol) 4000 as the precipitant. The space group was $P3_221$ and the unit cell parameters are $a = b = 140.2$ Å, $c = 119.3$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$, having one aromatase molecule in the asymmetric unit.

Diffraction data at about 100 K were collected initially at the Cornell High Energy Synchrotron Source (CHESS) and then to 2.90 Å resolution at the Advanced Photon Source, Argonne National Laboratory, with glycerol as a cryoprotectant. Two data sets at the Fe absorption edge were also collected at the CHESS. The structure was solved by the molecular replacement method coupled with Bijvoet difference Fourier synthesis for identifying the correct solution. Model building and refinement were performed with Coot and Refmac5, respectively. The final model contained 452 amino acid residues; 44 N-terminal and 7 C-terminal residues could not be built because of weakness of their electron densities. The final *R* factor for all reflections between 38 and 2.90 Å resolution was 0.214, and the *R*-free value was 0.244. The r.m.s. deviations of bond lengths and angles from ideal values were 0.009 Å and 1.32°, respectively. The average isotropic thermal factor for all atoms was 77.3 $Å^2$. There were only two violations in the backbone torsion angle Ramachandran plot, both in the loop regions. The oxyferryl Fe(IV)=O moiety was generated by adding an oxygen atom to Fe with the modelling software MOE (Chemical Computing Group) The exemestane molecule was built into the active site by superimposing it on the experimentally derived androstenedione atomic positions with MOE.

METHODS

Purification and crystallization

The enzyme was purified to homogeneity from the microsomal fraction of homogenized fresh human placenta by immunoaffinity chromatography (Supplementary Fig. 1a). The purified enzyme was highly active: the specific activity with androstenedione as the substrate ranged between about 10 and about 100 nmol min^{-1} mg^{-1} over a large number of purification experiments. Details of this procedure have been described previously[19,31]. The absorption spectrum of the androstenedione complex had a Soret band at 394 nm (Supplementary Fig. 1b), which is characteristic of the oxidized high-spin ferric (Fe^{3+}) state of the haem iron and is suggestive of the formation of the androstenedione complex[32]. Freshly purified aromatase in 100 mM potassium phosphate buffer pH 7.4 containing 20% glycerol, 20 mM dithiothreitol, 0.1 mM androstenedione and 1 mM n-dodecyl-β-D-maltopyranoside (BDM) was mixed with reservoir cocktails of 24–30% poly(ethylene glycol) 4000 and 0.5 M NaCl in 0.05 M Tris-HCl buffer pH 8.5 and vapour-diffused in sealed 24-well sitting-drop plates against corresponding reservoir solution. The purification and crystallization experiments were all conducted at 4 °C. Reddish-brown hexagonal rod-shaped crystals appeared in 7–10 days and continued to grow up to 14–16 days. Typically, the crystals were about 0.05–0.50 mm in length, with a hexagonal cross-section of about 0.01–0.12 mm (Supplementary Fig. 1c).

Data collection

Initially, diffraction data sets to about 3.3 Å resolution were collected at the A-1 station of the Cornell High Energy Synchrotron Source (CHESS). The crystal was flash-cooled in a stream of liquid nitrogen with about 40% glycerol as the cryoprotectant, and maintained at about 100 K during data collection. Additionally, two data sets, each to 4.2 Å resolution, were measured at the CHESS F-2 station by tuning the beam to the peak and the inflection point of the iron absorption edge (inflection (1.7433 Å): total 31,833, unique 9,814, completion 97%, $I/\sigma I$ highest 4.1, R_{merge} 0.159; remote (1.7284 Å): total 32,671, unique 9,903, completion 98%, $I/\sigma I$ highest 4.6, R_{merge} 0.126). Finally, a 2.9-Å diffraction data set used for solution and refinement of the structure was gathered at cryogenic temperature at beamline 19-ID-D (0.979 Å) of the Advanced Photon Source, Argonne National Laboratory. The data were recorded on an ADSC Q315 CCD detector and processed with the HKL3000 software package[33]. The space group was $P3_221$ and the unit cell parameters were $a = b =$ 140.2 Å, $c = 119.3$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$. There was one aromatase molecule in the

asymmetric unit, with a solvent content of about 79%. A total of 184,295 diffraction intensities were measured, yielding 30,371 unique reflections. The diffraction data were 99.4% complete to 2.90 Å resolution with an overall intensity to sigma ratio of 31.1 and an R_{merge} of 0.067. The intensity to sigma ratio was 2.8 in the highest-resolution shell. Supplementary Table 1 summarizes the diffraction data statistics.

Structure solution and refinement

The structure was solved by the molecular replacement method (at 3.3 Å), coupled with Bijvoet difference Fourier synthesis with the Fe-absorption edge data sets (at 4.5 Å). The latter helped in identifying the correct molecular replacement solution, at the same time confirming that there was one molecule in the asymmetric unit. Extensive rotation and translation function searches were conducted with a large number of P450 coordinates from the Protein Data Bank (PDB ID codes: 1PQ2, 1R90, 1TQN, 1Z10, 1W0E, 1Z11, 1ZO4, 1ZOA, 2F9Q, 2FDV, 2FDY, 2FDU, 2FDW, 2HI4, 2J0C, 2J0D, 2OJD and 2P85) with the AMORE and MOLREP routines in the CCP4 software package34. Only two search models yielded the correct molecular replacement solution: 2F9Q (human P450 2D6; ref. 23) and 1W0E (human P450 3A4; ref. 22), the two human cytochrome P450s with highest sequence identities with aromatase (14–18%). Model building and refinement were performed with Coot35 and Refmac5 (ref. 36) routines, respectively, running on either a dual-CPU G5 or a Powerbook G4 with the Mac OS 10.4 operating system. The final model contained 452 amino-acid residues, a haem group, one androstenedione molecule, 35 solvent waters and 2 phosphate ions (3,767 total atoms). Models for 44 N-terminal and 7 C-terminal residues could not be built because of weakness of the their electron densities. Apart from these residues, the electron density for the rest of the molecule was mostly well defined, except for two short loop/turn regions. The fit between the experimental electron density of side chains and the corresponding sequence was excellent except for a few exposed charged amino acids, such as lysines. In the space of missing N-terminal residues, an isolated patch of weak electron density was identifiable and could be fitted to a four-turn helix. However, because of the lack of side-chain identities, the helix could not be modelled unequivocally. Additionally, electron densities for two detergent molecules were located near the Trp 67 side chain, the presumed transmembrane region of the enzyme, but were not included in the final refinement because of the possibility of multiple orientations of the sugar moiety or the alkyl chain. The final *R* factor for all reflections between 37.8 and 2.90 Å resolutions was 0.214, and the *R*-free value was 0.244. The r.m.s. deviations of bond lengths and angles from ideal values were 0.009 Å and 1.32°, respectively. The average isotropic thermal factor (*B*) for all atoms was 77.3 $Å^2$, whereas the Wilson plot *B*-value was 94.5 $Å^2$. Of 407 non-glycine and non-proline residues, there were two violations in the backbone torsion angle Ramachandran plot, all in weaker loop regions. Overall random coordinate errors were 0.33 Å (based on *R*-working), 0.26 Å (on *R*-free) and 0.19 Å (on maximum likelihood). A summary of all numbers from refinement is provided in Supplementary Table 1.

The overall structure of aromatase

The structure of aromatase consists of 12 major α-helices (A to L) and 10 β-strands (1 to 10) distributed into one major and three minor sheets, and follows the characteristic cytochrome P450 fold (Fig. 1a). The major β-sheet is a mixed four-stranded sheet that begins near the N terminus (β1:83–88 and β2:93–97) but ends in two strands from the C-terminal half of the polypeptide chain (β3:373–376 and β6:393–396). The N-terminal residues 47–50 make one backbone hydrogen bond with β1 and add an extra β-strand-like element to this sheet in aromatase. Each of the three minor sheets consists of two antiparallel strands scattered over the polypeptide chain (sheet2: β4:381–383 and β5:386–388; sheet3: β8:473–475 and β9:479–481; sheet4: β7:458–461 and β10:491–494). Of the 12 major helices, I (293–324), F (210–227), G (242–267), H (278–287), C (138–152), D (155–174), E (187–205), J (326–

341), K (354–366) and L (440–455) are similar to those found in most of the cytochrome P450s. Helices A′ (57–68), A (69–80), B′ (100–109), B (119–126), G′ (232–236), H′ (271–274), J′ (346–349), K′ (398–404) and K″ (414–418) are one to four turns long and have more variability among P450s. The residues in aromatase involved in haem coordination are Arg 115, Trp 141, Arg 145, Arg 375 and Arg 435.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Y. Osawa, who pioneered aromatase research and its purification from human placenta at the institute, for many discussions and encouragement; past graduate students, postdoctoral scientists and research associates for contributions to the initial purification and crystallization efforts; H. Davies for discussions; D. Gewirth, V. Cody, G. DeTitta and J. Griffin for critically reading the manuscript; staff at the Women's and Children's Hospital of Buffalo for providing the placenta used in this work; and staffs of the Cornell High Energy Synchrotron Source and the Advanced Photon Source, Argonne National Laboratory, for helping with the synchrotron X-ray data collection. The research is supported in part by grants GM62794 and GM59450 (to D.G.) from the National Institutes of Health.

References

1. Thompson EA, Siiteri PK. Utilization of oxygen and reduced nicotinamide adenine dinucleotide phosphate by human placental microsomes during aromatization of androstenedione. J Biol Chem 1974;249:5364–5372. [PubMed: 4153532]
2. Simpson ER, et al. Aromatase cytochrome P450, the enzyme responsible for estrogen biosynthesis. Endocr Rev 1994;15:342–355. [PubMed: 8076586]
3. O'Neal Johnston J. Aromatase inhibitors. Crit Rev Biochem Mol Biol 1998;33:375–405. [PubMed: 9827706]
4. Eisen A, Trudeau M, Shelley W, Messersmith H, Pritchard KI. Aromatase inhibitors in adjuvant therapy for hormone receptor positive breast cancer: A systematic review. Cancer Treat Rev 2008;34:157–174. [PubMed: 18164821]
5. Akhtar M, Calder DL, Corina DL, Wright JN. Mechanistic studies on C19-demethylation in oestrogen biosynthesis. Biochem J 1982;201:569–580. [PubMed: 7092812]
6. Akhtar M, Njar VC, Wright JN. Mechanistic studies on aromatase and related C–C bond cleaving P-450 enzymes. J Steroid Biochem Mol Biol 1993;44:375–387. [PubMed: 8476751]
7. Nakajin S, Shinoda M, Hall PF. Purification to homogeneity of aromatase from human placenta. Biochem Biophys Res Commun 1986;134:704–710. [PubMed: 3947346]
8. Kellis JT, Vickery LE. Purification and characterization of human placental aromatase cytochrome P-450. J Biol Chem 1987;262:4413–4420. [PubMed: 3104339]
9. Zhou D, Pompon D, Chen S. Structure–function studies of human aromatase by site-directed mutagenesis: kinetic properties of mutants Pro-308–Phe, Tyr-361–Phe, Tyr-361–Leu, and Phe-406–Arg. Proc Natl Acad Sci USA 1991;88:410–414. [PubMed: 1988941]
10. Kadohama N, Yarborough C, Zhou D, Chen S, Osawa Y. Kinetic properties of aromatase mutants Pro 308Phe, Asp 309Asn, and Asp 309Ala and their interactions with aromatase inhibitors. J Steroid Biochem Mol Biol 1992;43:693–701. [PubMed: 1472461]
11. Chen S, et al. Structure–function studies of human aromatase. J Steroid Biochem Mol Biol 1993;44:347–356. [PubMed: 8476748]
12. Laughton CA, Zvelebil MJ, Neidle S. A detailed molecular model for human aromatase. J Steroid Biochem Mol Biol 1993;44:399–407. [PubMed: 8476753]
13. Oh SS, Robinson CH. Mechanism of human placental aromatase: a new active site model. J Steroid Biochem Mol Biol 1993;44:389–397. [PubMed: 8476752]

14. Amarneh B, Simpson ER. Expression of a recombinant derivative of human aromatase P450 in insect cells utilizing the baculovirus vector system. Mol Cell Endocrinol 1995;109:R1–R5. [PubMed: 7664973]
15. Graham-Lorence S, Amarneh B, White RE, Peterson JA, Simpson ER. A three-dimensional model of aromatase cytochrome P450. Protein Sci 1995;4:1065–1080. [PubMed: 7549871]
16. Kao YC, Korzekwa KR, Laughton CA, Chen S. Evaluation of the mechanism of aromatase cytochrome P450. A site-directed mutagenesis study. Eur J Biochem 2001;268:243–251. [PubMed: 11168357]
17. Chen S, et al. Structure–function studies of aromatase and its inhibitors: a progress report. J Steroid Biochem Mol Biol 2003;86:231–237. [PubMed: 14623516]
18. Hong Y, Cho M, Yuan YC, Chen S. Molecular basis for the interaction of four different classes of substrates and inhibitors with human aromatase. Biochem Pharmacol 2008;75:1161–1169. [PubMed: 18184606]
19. Lala P, et al. Suppression of human cytochrome P450 aromatase activity by monoclonal and recombinant antibody fragments and identification of their stable antigenic complex. J Steroid Biochem Mol Biol 2004;88:235–245. [PubMed: 15120417]
20. Nagano S, Poulos TL. Crystallographic study on the dioxygen complex of wild-type and mutant cytochrome P450cam. Implications for the dioxygen activation mechanism. J Biol Chem 2005;280:31659–31663. [PubMed: 15994329]
21. Nagano S, Cupp-Vickery JR, Poulos TL. Crystal structures of the ferrous dioxygen complex of wild-type cytochrome P450eryF and its mutants, A245S and A245T: investigation of the proton transfer system in P450eryF. J Biol Chem 2005;280:22102–22107. [PubMed: 15824115]
22. Williams PA, et al. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. Science 2004;305:683–686. [PubMed: 15256616]
23. Rowland P, et al. Crystal structure of human cytochrome P450 2D6. J Biol Chem 2006;281:7614–7622. [PubMed: 16352597]
24. Sansen S, Hsu MH, Stout CD, Johnson EF. Structural insight into the altered substrate specificity of human cytochrome P450 2A6 mutants. Arch Biochem Biophys 2007;464:197–206. [PubMed: 17540336]
25. Guallar V, Baik MH, Lippard SJ, Friesner RA. Peripheral heme substituents control the hydrogen-atom abstraction chemistry in cytochromes P450. Proc Natl Acad Sci USA 2003;100:6998–7002. [PubMed: 12771375]
26. Podust LM, Poulos TL, Waterman MR. Crystal structure of cytochrome P450 14α-sterol demethylase (CYP51) from *Mycobacterium tuberculosis* in complex with azole inhibitors. Proc Natl Acad Sci USA 2001;98:3068–3073. [PubMed: 11248033]
27. Hackett JC, Brueggemeier RW, Hadad CM. The final catalytic step of cytochrome p450 aromatase: a density functional theory study. J Am Chem Soc 2005;127:5224–5237. [PubMed: 15810858]
28. Pettersen EF, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612. [PubMed: 15264254]
29. Cojocaru V, Winn PJ, Wade RC. The ins and outs of cytochrome P450s. Biochim Biophys Acta 2007;1770:390–401. [PubMed: 16920266]
30. Shimozawa O, et al. Core glycosylation of cytochrome P-450(arom). Evidence for localization of N terminus of microsomal cytochrome P-450 in the lumen. J Biol Chem 1993;268:21399–21402. [PubMed: 8407981]
31. Yoshida N, Osawa Y. Purification of human placental aromatase cytochrome P-450 with monoclonal antibody and its characterization. Biochemistry 1991;30:3003–3010. [PubMed: 2007137]
32. Sato, R.; Omura, T. Cytochrome P-450. Kodansha/Academic; 1978.
33. Otninowski, Z.; Minor, W. HKL Manual. Yale University; 1995.
34. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D 1994;50:760–763. [PubMed: 15299374]
35. Emsley P, Cowtan K. Coot: model building tools for molecular graphics. Acta Crystallogr D 2004;60:2126–2132. [PubMed: 15572765]

36. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D 1997;53:240–255. [PubMed: 15299926]

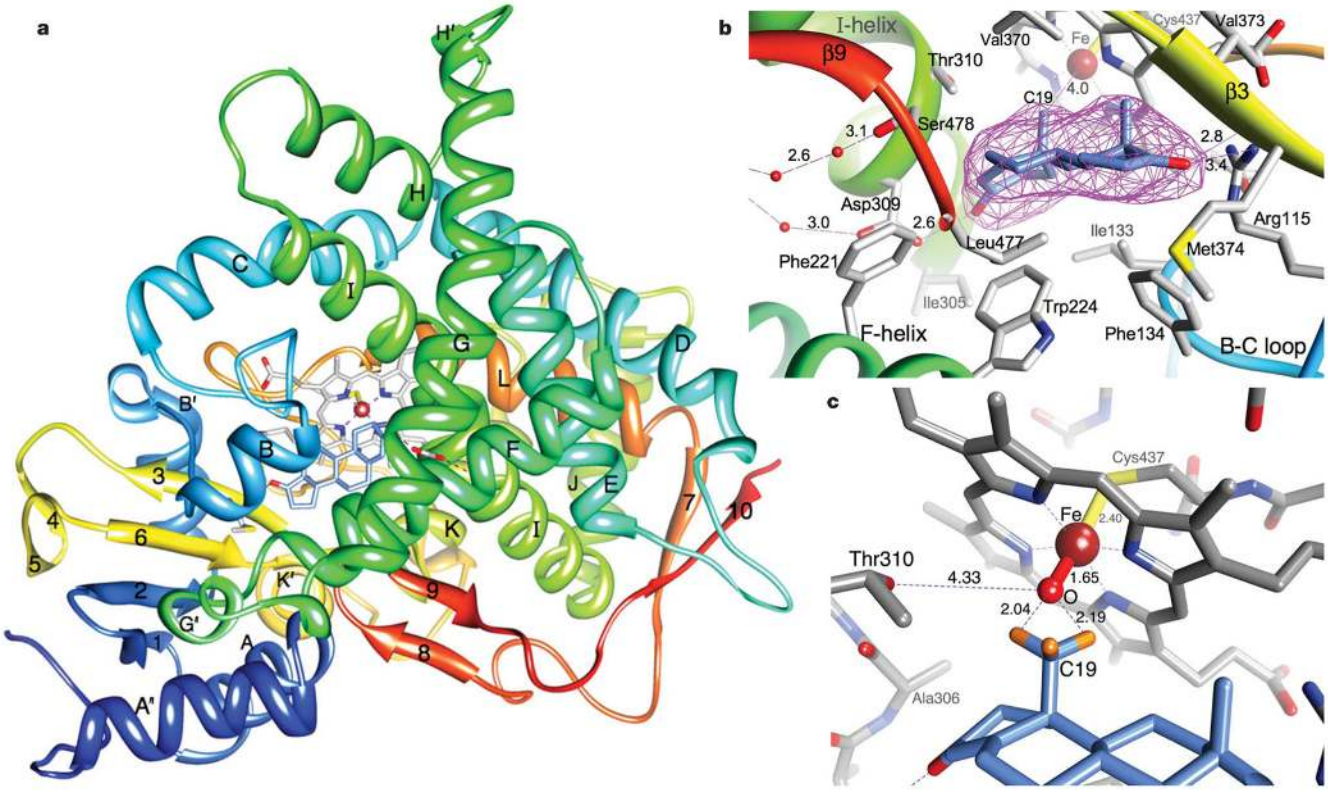

Figure 1. The structure of aromatase

a, A ribbon diagram showing the overall structure. The N terminus, starting at residue 45, is coloured dark blue and the C terminus ending at residue 496 is coloured red. The α-helices are labelled from A to L and β-strands are numbered from 1 to 10. The haem group, the bound androstenedione molecule at the active site and its polar interactions are shown. **b**, A close-up view of the active site showing the bound androstenedione molecule in unbiased difference ($F_{obs} - F_{cal}$) electron density contoured at 4.5 times the standard deviation. **c**, Modelling of Fe(III) as an oxyferryl Fe(IV)=O moiety. The C19-methyl hydrogen atoms are shown at the calculated ideal positions. Important side chains, haem and water molecules are depicted in element colours: grey, C; blue, N; red, O; yellow, S; firebrick, Fe; orange, H. The C atoms of androstenedione are coloured cornflower blue. This colour code is used in all figures. Distances are in ångströms. The directions of view into the active site are roughly similar in all panels. Unless otherwise noted, all three-dimensional illustrations were prepared with Chimera[28].

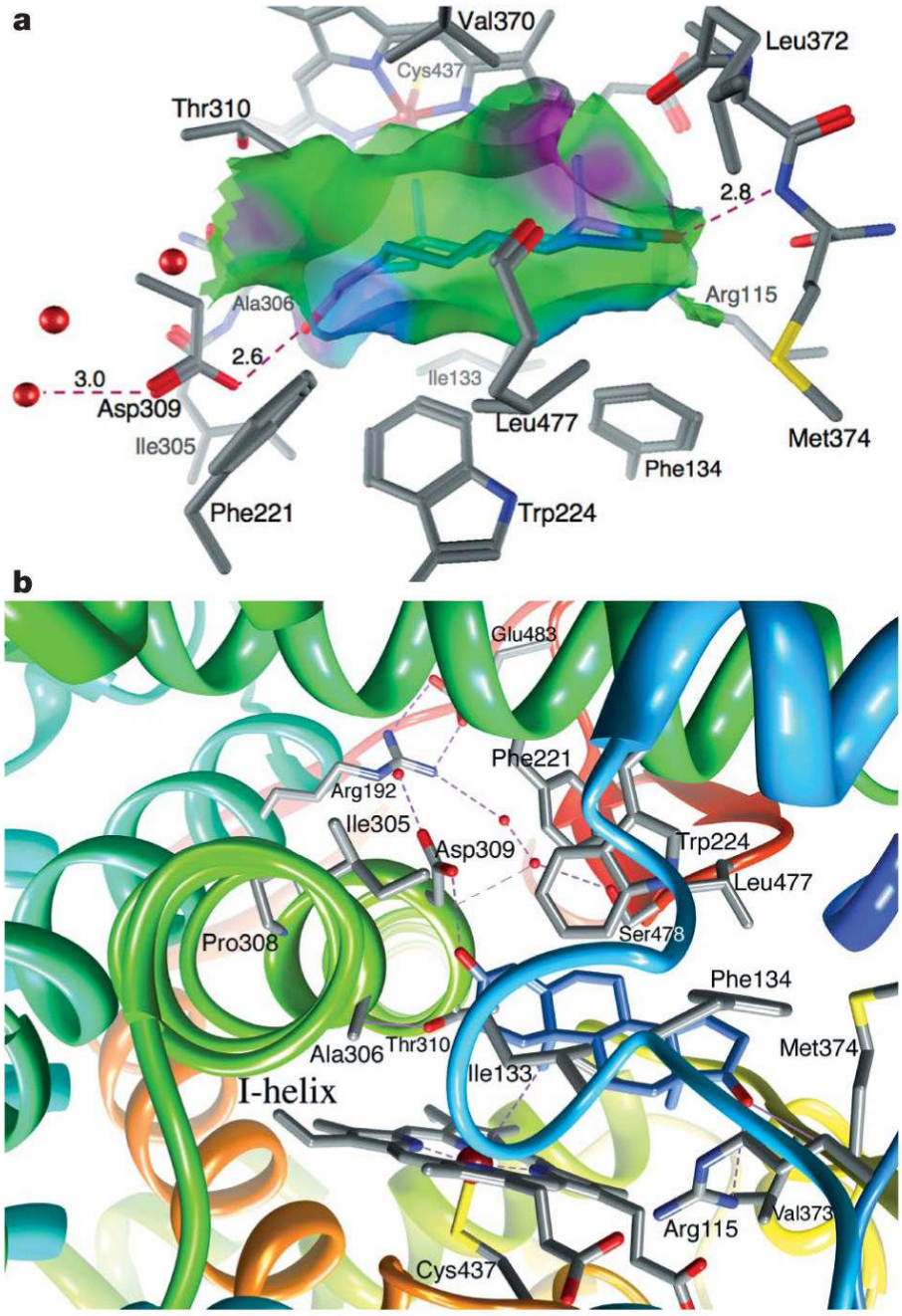

Figure 2. Views of the active site of aromatase

a, A van der Waals interaction surface cast by the protein and haem atoms at the active site. The semi-transparent surface, coloured green for hydrophobic interactions and magenta for polar interactions, closely resembles the shape, size and puckering of the steroid backbone. This figure was prepared with MOE. **b**, A view along the I-helix axis from its N-terminal end. The disruption to the helicity of the backbone at residues Pro 308-Asp 309-Thr 310 causes the helix axis to displace by about 3.5 Å, allowing the side chain of Asp 309 to interact with the 3-keto oxygen of the steroid. The deviation from helicity could be stabilized by a strong Ala 306CO···HOThr 310 (2.8 Å) hydrogen bond, as well as by Asp 309 peptide CO···water (3.4 Å) interaction as indicated.

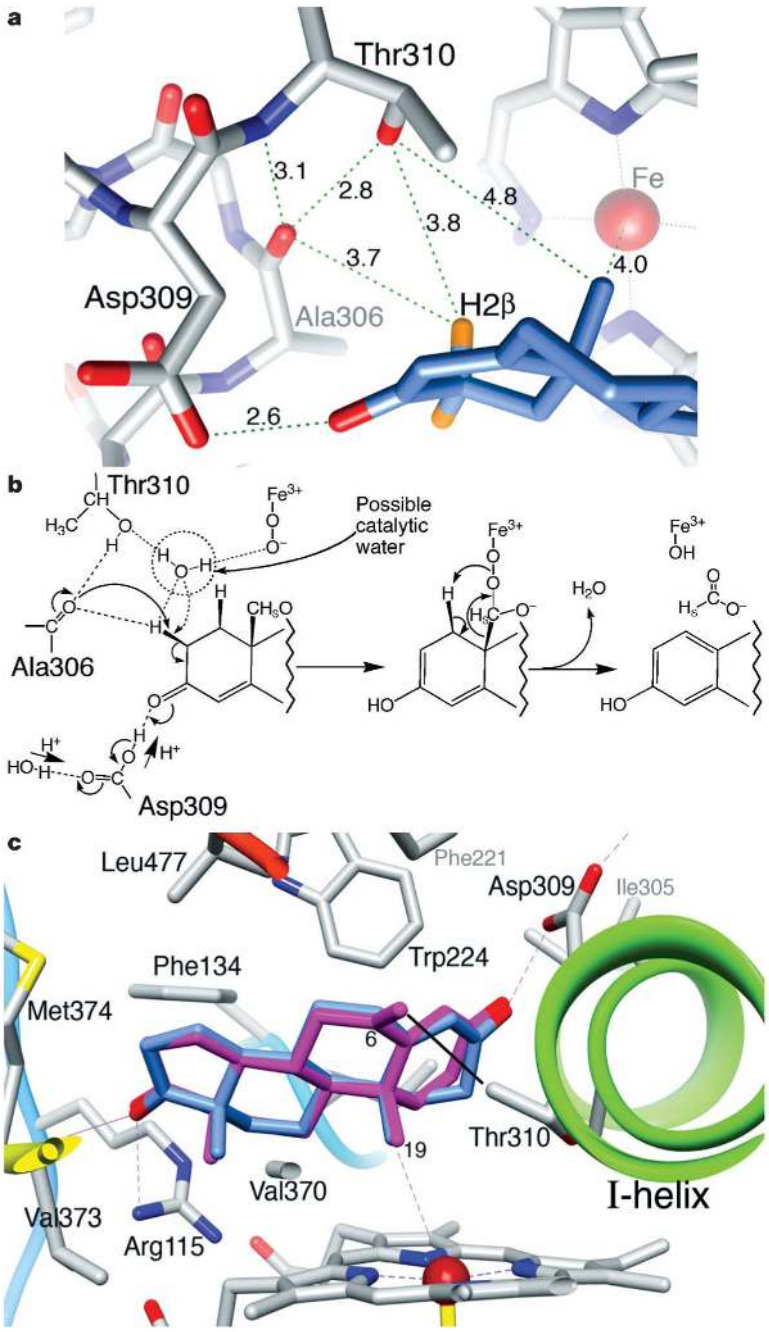

Figure 3. Steroid–protein interactions and mechanistic implications

a, A close-up view of the Ala 306CO···HOThr 310 pair that may function in the aromatization of the A-ring of androstenedione. Calculated hydrogen-atom positions of C2 of the bound androstenedione are shown. Distances are in ångströms. **b**, A possible mechanism for H2β abstraction and 2,3-enolization that could be initiated by a nucleophilic attack on C2-H2β by the Ala 306CO···HOThr 310 pair, along with an electrophilic attack on the C3 carbonyl by a protonated Asp 309 side chain. The direction of proton flow from the proton relay network through Asp 309 carboxylate to the substrate is indicated by arrows. Involvement of a catalytic water in H2β abstraction is a possibility. The backbone carbonyl of the Ala 306CO···HOThr 310 pair aided by a potential catalytic water molecule, or the

water oxygen itself (as indicated by dotted arrow) could act as the nucleophile. H1β abstraction is drawn as proposed previously[6]. **c**, Modelling of an exemestane molecule (C atoms in magenta) after the experimental positioning of androstenedione. The short van der Waals contact distance (3 Å) between the C6-methylidene carbon and Cγ of Thr 310 is indicated by a black line.

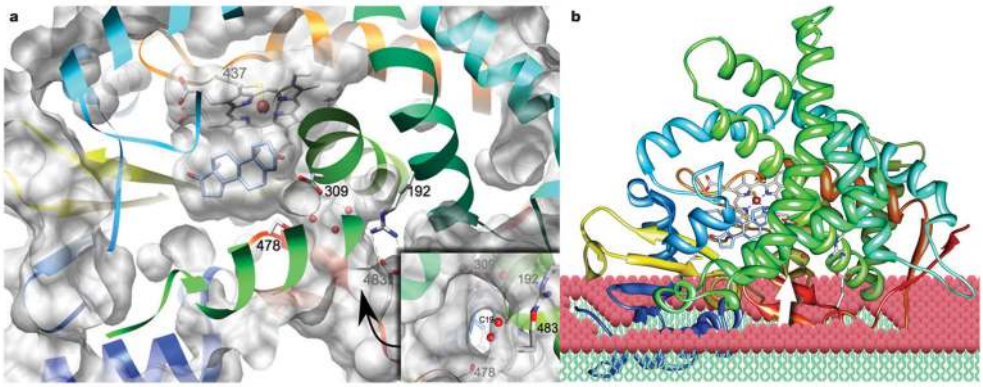

Figure 4. A putative active-site access channel from within the lipid bilayer

a, The solvent-excluded surface[28] of aromatase excludes the steroid-binding pocket and haem from the protein interior by forming a 'pouch'-like cleft that has the only opening to the protein exterior through a channel, roughly at the arrowhead. The course of the polypeptide chain is shown in rainbow colour. Residues Arg 192, Asp 309, Ser 478 and Glu 483 border this channel from the protein interior; three water molecules, part of the proton relay network, are within the channel. The inset is a view along this channel at the arrowhead, showing the locations of water molecules and opening to the active site. **b**, In a proposed membrane integration model, the opening to the active-site access channel rests on the lipid bilayer surface, allowing the steroids to enter the aromatase active site directly from within the bilayer, roughly along the arrow shown. The model suggests lipid integration/association of the N terminus up to helix A, and other loops near the C terminus. The orientation of aromatase is roughly the same in both panels.