Functional Domains of Aromatase Cytochrome P450 Inferred from Comparative Analyses of Amino Acid Sequences and Substantiated by Site-directed Mutagenesis Experiments*

Shiuan Chen and Dujin Zhou
From the Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, California 91010

Several functional domains, especially the active site regions, in aromatase cytochrome P450 were inferred by alignment of amino acid sequences of the enzyme from five species, human, rat, mouse, chicken, and trout, and that of Pseudomonas putida cytochrome P450cam, whose x-ray structure has been determined (Poulos, T. L., Finzel, B. C., and Howard, A. J. (1987) J. Mol. Biol. 195, 687–700). The predicted functions of these domains have been evaluated by site-directed mutagenesis. Eighteen mutants, including seven new mutants, have been generated in this laboratory. The seven newly prepared mutants are Q123E, Q123H, T310S, T310C, R365K, R365A, and NA20 (a mutant without the first 20 amino acids). The preparation and characterization of these new mutants are described. The structural model described in this paper should be very useful for future structure-function studies of aromatase by site-directed mutagenesis.

Aromatase, a cytochrome P450, catalyzes three consecutive hydroxylation reactions converting C19 androgens to aromatic C18 estrogenic steroids. This enzyme has received considerable attention because of the central importance of estrogens in many reproductive and metabolic processes. The synthesis of estrogens is required for the normal expression of secondary sexual characteristics and establishment and maintenance of pregnancy. Fetal expression in the brain is believed to determine male or female metabolic patterns expressed during adult development (1). Moreover, the abnormal expression of aromatase in a significant number of breast tumors (2–5), and the inhibition of the enzyme as part of a therapeutic approach to this disease, make the study of aromatase of paramount importance. In fact, since aromatase inhibitors are potentially useful drugs for treating breast cancer, the synthesis and screening of new aromatase inhibitors remains an active area of research in many laboratories and pharmaceutical companies. This enzyme is also of interest to a number of investigators because of the complexity of the reaction it catalyzes. Although it has been shown that three molecules of molecular oxygen and six reducing equivalents of NADPH are consumed during estrogen formation (6), the reaction mechanism is not yet completely elucidated.

The design of more effective aromatase inhibitor(s) for treatment of breast cancer will be benefited greatly by a more detailed understanding of the structure of the active site of aromatase. The availability of structural information at the active site of the enzyme should also help unravel the reaction mechanism of this enzyme. With these considerations in mind, we have carried out structure-function studies of aromatase by site-directed mutagenesis experiments (7, 8).

Aromatase belongs to the cytochrome P450 superfamily. Cytochrome P450 monooxygenases are a group of proteins that play a central role in the metabolism of a wide variety of foreign compounds in different organs and are involved in physiologically important pathways of synthesis of steroid hormones, prostaglandins, and vitamin D3 (9). The human and rodent genomes contain at least 50 cytochrome P450 genes, which are classified into 10 families according to the currently available protein sequence data (10). These proteins share amino acid sequence homology and contain common functional domains as indicated by comparisons of their amino acid sequence (11). A member of the cytochrome P450 superfamily is Pseudomonas putida cytochrome P450cam, whose x-ray structure has been determined (12). Although this bacterial cytochrome P450 exhibits less than 30% amino acid identity with mammalian cytochrome P450s, its x-ray structure has been widely used as a basis to generate models for other cytochrome P450s because it is the only available x-ray structure for cytochrome P450. Typically, the amino acid sequence of the cytochrome P450 of interest is initially aligned with that of cytochrome P450cam. The important regions of the interested cytochrome P450 are then postulated based on their sequence homology with the corresponding regions in cytochrome P450cam. A number of attempts have been made to align the amino acid sequences of different cytochrome P450s with that of cytochrome P450cam and to predict functional domains of these cytochrome P450s (11, 13–15). By this approach, both we (8) and Dr. E. Simpson’s group at Dallas (16) have prepared hypothetical models for the active site of aromatase. Based on these models, several aromatase mutants have been generated and analyzed (7, 8, 16). The studies have already provided useful information concerning the catalytic properties of the enzyme.

In light of the importance of such structure-function studies and because of the urgent need of a better model for detailed studies of the enzyme, we have decided to carefully examine the amino acid sequence of aromatase and have prepared new mutants to reevaluate the proposed model for the active site of the enzyme. We feel that such approaches are important and should provide a number of new insights concerning the structure-function relationship of aromatase. The important points derived from our recent amino acid sequence comparisons and the results from our new mutagenesis experiments are described and discussed in this paper.

*This research was supported by National Institutes of Health Grants CA 44735 and CA 53572. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡To whom correspondence should be addressed. Tel.: 818-359-8111 (ext. 2601); Fax: 818-301-8186.
**MATERIALS AND METHODS**

*Chemicals*—T4 kinase, T4 DNA ligase, and restriction endonucleases were obtained from Boehringer Mannheim and Bethesda Research Laboratories. Radiolabeled nucleotides, radiolabeled deoxyribose, and [35S]methionine were from Du Pont-New England Nuclear. DNA sequencing kits were from U. S. Biochemical Corp. The site-directed mutagenesis kits were from Bio-Rad.

**Stable Expression and Site-directed Mutagenesis Experiments—**

Aromatase and its mutant expression plasmids were constructed as described previously by Zhou et al. (7, 17). The transfection experiments were done according to the procedure of Zhou et al. (17). The introduction of specific base changes in the aromatase cDNA was accomplished by site-directed mutagenesis using the phagemid mutagenesis kit from Bio-Rad. A detailed procedure for the mutagenesis was described by Zhou et al. (7). Oligonucleotide primers used for the mutagenesis experiments were synthesized at the DNA synthesis facility at the City of Hope, and their sequences are shown in Fig. 1. These are reverse primers. In addition to changes at the specific bases, a silent mutation was introduced to create a unique restriction site in each primer. By doing this, the mutants could be easily detected upon restriction digestion of the plasmid DNAs using the proper restriction enzymes. All mutated cDNAs were sequenced completely to confirm that there were no other mutations introduced except those desired.

A primer-directed amplification method was used to generate the cDNA encoding amino acids 21-503 of human placental aromatase (i.e. the mutant N220). The 5' and 3' vector (17) was linearized with Sall and BamH, and the N220 cDNA was specifically amplified with the following two primers. The oligonucleotide 5'-ATTGTCGAC-CATTGCTGCTGGCCCATG-3' was designed for amplifying the aromatase cDNA beginning with the nucleotides encoding amino acid residue 21. The primer was also designed to contain the extra nucleotides ATTGTCGAC added to the 5'-end of the coding sequence to provide a Sall restriction site. The 3'-end of the amplified cDNA was established by using the oligonucleotide 5'-AGATCTGGTGTCCAGAACCTGGTC-3', which is complementary to the last 18 nucleotides of human aromatase coding sequence plus the additional sequence AGATCT to provide a BglII restriction site. The amplified cDNA fragment was ligated into PCR1000 vector from the TA cloning kits (Invitrogen Co., San Diego, CA) and subsequently subcloned into the mammalian expression vector, pCI, and subsequently cultured on six-well cell culture plates and washed twice with serum-free cell culture medium. The 3H-labeled substrate was dissolved in serum-free cell culture medium, filter-sterilized, and added to each well. After a 30-min incubation at 37 °C, followed by a 5-min incubation on ice, 1 ml of culture medium was withdrawn from each well. The culture medium was mixed with an equal volume of chloroform to extract the product. The aqueous phase was treated with dextran-treated charcoal and centrifuged, and the amount of the product, $^3$H$_2$O, was determined using a scintillation counter. The protein concentration was determined after dissolving cells with 0.5 M NaOH. This $^3$H$_2$O release assay for human aromatase expressed in CHO cells was previously validated by the product isolation assay (17). The Southern blot and Northern blot analyses were performed as described by Zhou et al. (17).

**RESULTS AND DISCUSSION**

Difficulty definitely exists in aligning distantly related protein sequences of a mammalian cytochrome P450 such as aromatase and a procarboxylic cytochrome P450cam. Since amino acid sequences of aromatase of human (21-23), rat (24), mouse (25), chicken (26), and trout (27) have been deduced from their cDNA sequences, we have decided to align these sequences first. It is reasonable to assume that those regions having sequences conserved throughout the five species are of functional importance and the regions showing sequence variation are less important. The credibility of such analysis is increased by comparing sequences from distantly related species, such as human aromatase to chicken or trout aromatase. Fig. 2 shows the alignment of amino acid sequences of aromatase of human, mouse, rat, chicken, and trout aromatase. It is known that cytochrome P450s usually have rather similar patterns in the hydrophobic/hydrophilic partition, even though there is little primary sequence similarity among them (13). Hydrophobicity analysis was performed to unravel subdomains of aromatase proteins. These predicted subdomains are indicated as $A-Z$ in Fig. 3. These subdomains are speculative, but they are very helpful for further discussion of the structure-function relationship of aromatase. The difficulty has already been recognized in aligning certain regions of aromatase and cytochrome P450cam by hydrophobicity analysis. Therefore, we have not made an attempt to compare the predicted secondary structures of aromatase to that of cytochrome P450cam because we felt that the secondary structure prediction is much more complicated than the hydrophobicity analysis and the results might be difficult to evaluate. Secondary structure predictions have been used by Gotoh (30) for structural analysis of the cytochrome P450 family 2.

Our amino acid sequence comparison has allowed us to predict several functional subdomains of aromatase. The pre-
Fig. 2. Alignment of the amino acid sequences of aromatase of trout (tr), chicken (ck), rat (rt), mouse (ms), and human (hu) to that of cytochrome P450cam. This figure is a modification of Fig. 5 of Tanaka et al. (27). Amino acid residues which are identical for the five species of aromatase are indicated with asterisks. Amino acids with conservative changes are indicated with dots. Triangles indicate the location of the exon-intron boundaries for the human aromatase gene as reported by Means et al. (28) and Toda et al. (29). The amino acid residues studied by site-directed mutagenesis are underlined. Predicted subdomains of aromatase according to hydropathic index plots (see Fig. 3), are indicated by lines on the top of the sequences and named as A–Z.

dicted functions of these different subdomains are discussed below. The functions of some of the subdomains have been confirmed by our site-directed mutagenesis experiments. The results of recent site-directed mutagenesis experiments are also described below. Unless indicated, the numbering of amino acid residues is based on the sequence of human aromatase.

Amino Terminus—Aromatase has an atypical amino terminus when compared with other microsomal cytochrome P450s (22). Characteristically, microsomal cytochrome P450s contain a hydrophobic amino terminus indicative of a membrane anchor. The feature is also found in the aromatase amino terminus in subdomain C. However, this hydrophobic region in aromatase is preceded by the more hydrophilic subdomain B. This is a feature uncommon to microsomal cytochrome P450s. This region contains sequences which could form an amphiphilic helix (22). Interestingly, there is very little sequence conservation in this region in the five different species of aromatase described (see Fig. 2). For the trout aromatase, there are 19 additional amino acids at the amino terminus - subdomain A (27). Since the hydrophobic plot of trout aromatase reveals that subdomain A is hydrophobic (see Fig. 3), it has been previously proposed to function as an additional membrane insertion domain (27). However, since there is a positive charged amino acid residue (i.e., Arg-10) in this region, this segment may only partially insert itself into the membrane.

In order to explain these unusual structural features of the amino terminal regions of aromatase, two models have been postulated and are shown in Fig. 4. Model I suggests that subdomain C is a bitopic membrane-insertion domain. Subdomain B forms an amphiphilic helix which sits on the lumen side of the microsomal membrane, and in the case of trout aromatase, subdomain A loops back and reinserts itself into the membrane. Subdomain B contains two acidic amino acid residues and each of these residues carry one negative charge and may serve as the stop-transfer signal, preventing further translocation across the membrane and anchoring the protein in the membrane. Subdomain B was recently shown to be very important for the stability of aromatase protein. Without this subdomain (i.e. the mutant NAZO), the mutant protein was not detected by the immunoprecipitation analysis (Fig.
It is possible that an absence of this subdomain prevents anchoring of the protein in the membrane and results in a rapid degradation of the truncated protein in cytosol. Southern blot analysis has been performed to show the presence of the expression plasmid DNA (Fig. 6A) and Northern blot analysis has been performed to show the presence of the RNA message encoding for the NΔ20 mutant protein in the transfected cells (Fig. 6B). These results are important to indicate that the mutant NΔ20 is probably expressed, but degraded rapidly in the transfected cells. Our Northern blot analysis revealed the presence of 2.0- and 1.9-kb RNA transcripts encoding for the enzyme in cells transfected with the full-length and NΔ20 expression plasmids, respectively (Fig. 6B). A second transcript of 3–3.1 kb was also detected and had been observed in some of previous Northern blot analyses of cells transfected with aromatase expression plasmid (17). It was suggested to be caused by improper RNA processing in these cells.

An amino-terminal sequence extension was also found in microsomal cytochrome P450 lanosterol-14α-demethylase of yeast Saccharomyces cerevisiae (31). Since substitution of subdomain B of aromatase with the extra amino-terminal segment of yeast lanosterol demethylase failed in producing functionally active aromatase, it is thought that subdomain B may have a functional role rather than simply act as a D. Pompon, D. Zhou, and S. Chen, unpublished results.


Fig. 3. Kyte and Doolittle hydropathy analysis of the amino acid sequences of trout aromatase (A), human aromatase (B), and cytochrome P450cam (C). Subdomains of these proteins are predicted according to their hydrophobic/hydrophilic properties and named as A–Z.

Fig. 4. Two models of membrane topologies for aromatase.

Fig. 5. Immunoprecipitation analysis of the cells expressing aromatase and the mutants NΔ20, Q123E, Q123H, T310S, T310C, R365K, and R365A. The cell lysates (2 × 10⁷ dpm), prepared according to the procedure of Zhou et al. (7), were treated with a rabbit polyclonal antibody against human placental aromatase. The antibody-antigen complexes were collected using protein A-Sepharose and antigen was released from the complex by treating with sodium dodecyl sulfate-gel electrophoresis buffer. The sodium dodecyl sulfate-gel electrophoresis was performed with ¹⁴C-labeled antigen prepared from CHO cells expressing aromatase or its mutants. Std., ¹⁴C-labeled molecular weight standards; lane 1, cells transfected with the expression vector; lane 2, cells expressing NΔ20; lane 3, cells expressing the wild-type enzyme; lane 4, cells expressing the wild-type enzyme, but the experiments was done by treating the cell lysate with a nonspecific rabbit polyclonal antibody instead of an antibody against human placental aromatase; lane 5, cells expressing the wild-type enzyme; lane 6, cells expressing mutant T310C; lane 7, cells expressing mutant T310S; lane 8, cells expressing mutant Q123E; lane 9, cells expressing mutant Q123H; lane 10, cells expressing the wild-type enzyme; lane 11, cells expressing the wild-type enzyme, but the cell lysate was treated with a nonspecific rabbit polyclonal antibody; lane 12, cells expressing mutant R365A; and lane 13, cells expressing mutant R365K.
membrane anchoring signal, even though the sequence homology for subdomain B of aromatase from the five species is low. Taking these findings into consideration, model II for aromatase insertion in the microsomal membranes is postulated (Fig. 4). In this latter model, subdomains C and E insert in the membrane as a hairpin, a mechanism of membrane insertion for cytochrome P450 as proposed by several researchers (11,32). In this situation, subdomain B may position itself in such a way as to interact with certain regions of the proximal and distal helices (12). As discussed above, subdomain Y of aromatase is proposed to be the proximal helix. The regions in aromatase corresponding to the distal helix have also been identified, i.e., subdomains T and U. Since the sequence homology of this region to the distal helix of cytochrome P450 is not as obvious as that of the subdomain Y to the proximal helix of cytochrome P450, site-directed mutagenesis experiments have been performed to confirm the importance of this region (7, 8). Thr-310 (corresponding to Thr-292 of cytochrome P450cam) is the highly conserved threonine residue throughout known cytochrome P450s. Thr-292 of cytochrome P450cam has been suggested to be a part of the dioxygen-binding site (34), and when it is replaced with alanine, the efficiency of camphor hydroxylation drops to only 5–6%, and electrons are channeled to produce hydrogen peroxide and water instead of product (35, 36). The crystal structure of cytochrome P450cam mutant T252A suggests that the cause for uncoupling is the greater access of solvent to the active site in the mutant enzyme (34). It has been suggested that Thr-252 of cytochrome P450cam interacts with both Gly-248 and with water 687 in the internal solvent channel (34). The cytochrome P450cam mutants T252S and T252N hydroxylate camphor with 85 and 57% efficiencies, respectively (35, 37). Serine and asparagine have hydrophilic side chains as threonine and could form hydrogen bond with the carbonyl oxygen of Gly-248. In this laboratory, two new aromatase mutants, T310S and T310C, have been generated. T310S and T310C have 48.8 and 2.4% of the wild-type activity, respectively, when using [1β-3H]androstenedione as the substrate (Table I). Results from the immunoprecipitation analysis revealed that the mutant T310S was expressed at a level higher than that of the wild-type enzyme (Fig. 5). Therefore, the activity of T310S is probably less than 48.8% of the wild-type enzyme. Through “in-cell” assays, the Ke values for androstenedione for T310S and T310C were determined to be 32 and 22.4 nM, respectively. The Ke value for testosterone for T310S was determined to be 39.3 nM. Although T310S catalyzes androgen aromatization with a 34.1 or 48.8% efficiency of the wild-type enzyme (using testosterone or androstenedione as the substrate, respectively), the conformation of the active site has been modified by changing Thr to Ser so that T310S is therefore less sensitive to the treatment of 4-hydroxyandrostenedione (4-OHA) and more sensitive to the treatment of aminoglutethimide (AG) than the wild-type enzyme (Fig. 7). The changes in responses to these two inhibitors...
for T310S are in a similar fashion for those for D309A (8) and (also see later). Interestingly, the aromatase mutant P308F was found to be more sensitive to the treatment of 4-OHA (7). Thus, mutations at this region (Pro-308 to Thr-310) affect the wild-type enzyme (38) and a substrate analogue, suggesting that this region is a part of 4-OHA binding site. In addition, these results agree with the fact that 4-OHA and AG inhibit aromatase with different mechanisms. It was technologically difficult to determine accurately the inhibitory profiles of these inhibitors on T310C because this mutant has a very low activity. Immunoprecipitation analysis was performed to show that the T310C was indeed expressed in CHO cells (Fig. 5).

A proposed model of the active site of aromatase. The regions participated in substrate binding are shown. We have generated 17 mutants with changes in these regions, and they are identical are indicated with dots. The amino acid residues which have been subjected to mutagenesis study are underlined.
These results, obtained from site-directed mutagenesis studies, indicate that subdomains T and U are very important parts of the active site of aromatase.

**Steroid Binding Domains**—Subdomains W, X, and the carboxy-terminal portion of subdomain V are conserved in all five species of aromatase (see Fig. 2). Subdomains V and W have been suggested to be part of the steroid-binding site by comparing their amino acid sequences to those of steroid dependent enzymes and steroid binding proteins (40). According to the alignment of amino acid sequences of aromatase and cytochrome P450cam, subdomains V and W of aromatase is similar to the helix H of cytochrome P450cam. We are proposing that subdomains V and W are involved in interaction with the C and D rings of the steroid. Recently, mutations of rat 17α-hydroxylase at the corresponding region (see Fig. 8) resulted in selective loss of either hydroxylase or lyase activity of this enzyme (41). Since 17α-hydroxylase catalyzes reactions at the C-17 position of the steroid, it is logical to conclude that this region is involved in interaction with the C and D rings of the steroid. In addition, conversion of Lys-529 and Lys-531 of human estrogen receptor to Gln resulted in a great decrease of estradiol binding (42). This region of the estrogen receptor was found to be homologous to subdomains V and W of aromatase (see Fig. 8). Two aromatase mutants at the corresponding basic amino acid residue have recently been generated in this laboratory, i.e. R365A and R365K. These mutants were not active (Table I). The immunoprecipitation analysis revealed that R365A and R365K were expressed, but at a level lower than that of the wild-type enzyme (7) (see Figs. 2 and 3). The Km values for androstenedione and testosterone for the two mutants, Q123E and Q123H, were also generated recently (Table I). Interestingly, both of these mutants have slightly higher activity than the wild-type enzyme. This is the first time that aromatase mutants with a catalytic activity higher than the wild-type enzyme are generated. The immunoprecipitation analysis revealed that Q123E is expressed at a level slightly higher than that of the wild-type enzyme (Fig. 5). The Km values for androstenedione and testosterone for the two mutants, Q123H and Q123E, are very similar to those for the wild-type enzyme (Table I), suggesting that the binding affinities of two substrates are not affected by these mutations. While these subdomains H, I, and J are felt to be important regions, the critical amino acid residue(s) is yet to be identified.

Based on the amino acid sequence comparisons and results from the site-directed mutagenesis experiments as described from the above sections, a hypothetical model of the active site of aromatase is generated (Fig. 9). The heme disc is embedded between the proximal (subdomain Y) and distal (subdomains T and U) helices. The steroid substrate is situated in the active site as indicated. The C and D rings of the steroid interact with subdomains V and W. Subdomain X connects subdomain W and subdomain Y (the proximal helix). A part of subdomains H, I, and J is situated in such a way as to interact with the A ring of the steroid. The 17 mutants with changes in these subdomains, which are generated in this laboratory, are indicated in this figure. All of these subdomains contain amino acid sequences conserved throughout the five species.

In addition to these domains which are thought to be involved in substrate binding, several structural domains of aromatase are also proposed. It is difficult to align subdomains E, F, and G of aromatase with those of cytochrome P450cam. The hydrophatic analysis also revealed significant differences at these regions between aromatase and cytochrome P450cam. These subdomains are currently thought to play a role in maintaining the structure of the aromatase protein. Furthermore, using the x-ray structure of cytochrome P450cam as a model, subdomains K, L, P, Q, R, and S are positioned near the surface of aromatase protein. Therefore, these subdomains may play little role in enzyme catalysis. Subdomains M, N, and O have low sequence homology and even low conservation of hydrophatic properties among the five species of aromatase (see Figs. 2 and 3). The alignment of these regions with cytochrome P450cam has been difficult. The function of these regions is not clear, although according to our alignment, Lys-216 and Lys-230 are thought to form hydrogen binding with Asp-309 in maintaining proper conformation of the protein.

---

1 Two mutants C124S and C124A have been generated recently using a yeast expression system. The preliminary studies have revealed that these two mutants have very low activities, suggesting that Cys-124 may be an important amino acid residue for enzyme catalysis (M.-A. Sari, D. Mansuy, D. Zhou, and S. Chen, unpublished results).
Finally, the amino terminal portion of subdomain V shows little conservation among the five species of aromatase. According to our alignment, the corresponding region does not exist in cytochrome P450cam. No function role is assigned to this region at the present time.

Thus, through comparison of amino acid sequences of aromatase from five species and that of cytochrome P450cam, and evaluation of the catalytic properties of our aromatase mutants, function domains of aromatase have been proposed. The intron-exon boundaries of the human aromatase gene have been identified (28, 29). It is very interesting to find that each exon of human aromatase represents a domain with a unique function (see Table II), assigned according to our analysis as described above. While the exact meaning of this finding is not clear, it may be that exon shuffling and modular construction are involved during the course of evolution of the aromatase protein.

In conclusion, by careful examination, important functional domains of aromatase cytochrome P450 are predicted. Without a three-dimensional structure of aromatase, the structural model described in this paper should be very useful for future structure-function studies of aromatase by site-directed mutagenesis. The model will be evaluated further by proper mutagenesis experiments and refined when new information is available. Through this approach, very useful protein structural information, that is essential for unraveling the reaction mechanism of aromatase and for designing more effective aromatase inhibitors for the treatment of breast cancer, will be generated.

Acknowledgment—We thank Dr. Patrick Clarke for useful comments on this manuscript.

REFERENCES
Functional domains of aromatase cytochrome P450 inferred from comparative analyses of amino acid sequences and substantiated by site-directed mutagenesis experiments.
S Chen and D Zhou


Access the most updated version of this article at [http://www.jbc.org/content/267/31/22587](http://www.jbc.org/content/267/31/22587)

Alerts:
- When this article is cited
- When a correction for this article is posted

[Click here](http://www.jbc.org/content/267/31/22587.full.html#ref-list-1) to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at [http://www.jbc.org/content/267/31/22587.full.html#ref-list-1](http://www.jbc.org/content/267/31/22587.full.html#ref-list-1)