Journal of Computer-Aided Molecular Design (2006) 19: 857–870
DOI 10.1007/s10822-005-9024-0

# Active site acidic residues and structural analysis of modelled human aromatase: A potential drug target for breast cancer

J. Narashima Murthy[a,b], M. Nagaraju[a], G. Madhavi Sastry[a], A. Raghuram Rao[b] & G. Narahari Sastry[a,*]
[a]*Molecular Modelling Group, Organic Chemical Sciences, Indian Institute of Chemical Technology, 500007, Hyderabad, Andhra Pradesh, India;* [b]*Medicinal Chemistry Research Division, University College of Pharmaceutical Sciences, Kakatiya University, 506009, Warangal, Andhra Pradesh, India*

Received 8 July 2005; accepted 18 October 2005


*Key words:* hydrogen bonding, information entropy, polar atoms, solvent accessibility, steroidal substrates

**Summary**

This study sheds new light on the role of acidic residues present in the active site cavity of human aromatase. Eight acidic residues (E129, D222, E245, E302, D309, E379, D380 and D476) lining the cavity are identified and studied using comparative modeling, docking, molecular dynamics as well as statistical techniques. The structural environment of these acidic residues is studied to assess the stability of the corresponding carboxylate anions. Results indicate that the environment of the residues E245, E302 and D222 is most suitable for carboxylate ion formation in the uncomplexed form. However, the stability of D309, D222 and D476 anions is seen to increase on complexation to steroidal substrates. In particular, the interaction between D309 and T310, which assists proton transfer, is found to be formed following androgen/nor-androgen complexation. The residue D309 is found to be clamped in the presence of substrate which is not observed in the case of the other residues although they exhibit changes in properties following substrate binding. Information entropic analysis indicates that the residues D309, D222 and D476 have more conformational flexibility compared to E302 and E245 prior to substrate binding. Interaction similar to that between D476 and D309, which is expected to assist androgen aromatization, is proposed between E302 and E245. The inhibition of aromatase activity by 4-hydroxy androstenedione (formestane) is attributed to a critical hydrogen bond formation between the hydroxy moiety and T310/D309 as well as the large distance from D476. The results corroborate well with earlier site directed mutagenesis studies.

## Introduction

Aromatase is a membrane bound microsomal cytochrome P450 enzyme involved in the specific conversion of androgens to estrogens by aromatization of *A*-ring of steroids. Further, these estrogens and their carcinogenic metabolites are responsible for progression of breast cancer [1, 2]. Uniqueness of this enzyme is it's ability to catalyze three successive oxidative transformations of the same substrate molecule before releasing the product of the reaction outside the enzyme pocket. All three reactions are stereo specific and utilize 3 mol of oxygen and 3 mol of NADPH [3–5]. In particular, acidic amino acid residues in the active site of human aromatase play a vital role in determining the efficacy of proton transport essential for aromatization of androgens [6]. Nevertheless, the

*To whom correspondence should be addressed. Fax: +91-40-2716-0512; E-mail: gnsastry@iict.res.in

enzymatic mechanism of estrogen formation and the role of these acidic residues is not totally understood. Earlier site directed mutagenesis studies on D309 of human aromatase implied the probable role of this aspartate in decarboxylation of androstenedione or testosterone and attracting a proton at C2 of the steroids during aromatization [7–9]. This process is assumed to be assisted by basic residues H475 and K473 present in proximity to C3 [10]. Previous site directed mutagenesis studies suggested that the repulsive force between D476 and E309 may be important for efficient aromatization of C19-androgens. However, the D309A mutant which induced activity loss with androstenedione as well as testosterone, exhibited activity with nor-testosterone. In contrast, the mutant E302A was found to be inactive with all the three substrates [10]. This suggests that E302 has an important part to play in the enzymatic mechanism involving both androgens as well as nor-androgens. However, previous studies demonstrated it's role in aromatization of nor-androgens through interaction with C2. Another residue whose mutation affected the enzymatic activity is T310. Some of the speculations on it's probable role indicate it's involvement in stabilizing the I-helix, forming hydrogen bonds, donating proton to the substrate as well as interacting with the C3 carbonyl group of the steroid substrates [11–13]. In addition, the residue I133 is thought to be involved in the optimal orientation of the neighboring residues helping the interaction with substrates in an indirect fashion [8]. Hence, understanding the properties of residues in the substrate binding pocket is imperative to comprehend the reaction mechanism unambiguously. However, absence of the 3-D crystal structure of human aromatase renders it elusive to comprehend these properties of the enzyme based on structural aspects and one has to resort to comparative modeling techniques. Previous studies derived knowledge largely from the available crystal structures of the bacterial cytochrome P-450s (CYP) (prokaryotic) which however shared low sequence homology with the human aromatase (eukaryotic) [10, 14, 15].

To date, crystal structures of thirteen prokaryotic CYPs and five eukaryotic CYPs have been reported [16]. Prokaryotic CYPs are soluble proteins and easy to crystallize, while eukaryotic CYPs are membrane associated enzymes and difficult to crystallize. Prokaryotic CYPs differ significantly in details that confer specificity of substrate despite their similar topological organization. Although the topological elements of microsomal CYPs are similar to microbial CYPs, the spatial arrangements of these elements clearly diverge from that of microbial enzymes [17]. It is difficult to decipher structural adaptation ambiguities of microsomal CYPs based on modeling with those of structurally characterized microbial P450's. This prompted us to construct homology model of human aromatase based on CYP 2C5 crystal structure [18], which among crystallized microsomal mammalian CYPs acts on steroid substrates similar to human aromatase. Therefore, using the mammalian CYP 2C5 crystal structure, we seek to investigate the main structural and functional aspects of acidic amino acid residues present in the active site of human aromatase which play an important role in protonation of substrates. Structure-function relationships between the active site acidic residues will be assessed from docking and molecular dynamics (MD) simulations. The structural features of these residues will be used to assess the stability of the corresponding carboxylate anion which include: hydrogen bonding interactions, location in helix, fraction of side chain exposed to polar atoms ($f$), solvent accessible surface areas (SAS) and information entropy ($S$) from MD simulations. Although structural basis to determine the acidity of residues has been a subject of considerable attention in the past, the present analysis is in accord with earlier experimental studies and throws new light on the properties of the active site acidic residues [19, 20–22]. Effect of androgen and nor-androgen binding to aromatase on the acidic residues will be gauged from docking studies keeping the active site of the enzyme as well as few important regions along with the ligand flexible. Efforts are made to understand the effect of mutations on side chains of these acidic residues through modeling techniques. The results are compared with available experimental results wherever possible.

## Materials and methods

### *Homology modeling*

Sequence identity between human aromatase and CYP 2C5 (pdb code: **1NR6**) is found to be 28%.

CYP 2C5, a mammalian P450 enzyme, binds to steroids which is a property shared by aromatases and consequently, the active site is expected to have similar geometry. Moreover, the active site obtained from the corresponding model (which will be discussed later) exhibits about 50% sequence homology with T310 being conserved. It may be noted that the active site of EryF, a cytochrome P450 of known structure also binds to steroids (androstenedione) similar to that of 2C5 but does not bear noticeable resemblance. Multiple sequence analysis of available aromatase sequences obtained from Swissprot database reveals that the active site residues obtained from our atomic model are highly conserved. However, the residues share lower sequence homology with respect to the sequences of EryF implying structural variations in the active site architecture between human aromatase and EryF. Therefore, it's structure is less suitable to model especially the active site of human aromatase. In addition, the crystal structures of EryF present in the protein databank are not of mammalian origin, contrary to the human aromatase and CYP 2C5 (**1NR6**). Further validation of the model is carried out by structural validation as well as comparing it with earlier experimental studies (see later).

To obtain a reasonable model various sequence alignment protocols between human aromatase and **1NR6** were adopted [22] and subsequently the 3-dimensional models of human aromatase were built by using the program Modeller [23]. Stereo chemical quality of the models was assessed from the Ramchandran plot. The dihedral angles of the peptide bonds are also verified. The structures are further validated using Profiles-3D program with a smoothing window of size 21 in the InsightII environment. The overall self-compatibility score based on the method by Lüthy et al. [24] is also tested. It is found that the 3D-PSSM fold recognition threading technique, which uses 1-D and 3-D sequence profiles, secondary structural and solvation potential information for obtaining proper fold, gave the best model [25]. Validation of the constructed model is taken up by comparing with the reported site directed mutagenesis studies, which will be deliberated in the results section to avoid redundancy. The heme moeity was then merged into the best obtained model of the protein retaining the spatial orientation as adopted by the heme in the template CYP 2C5. The structure was then minimized using the ESFF force field [26] to achieve a satisfactory and conformationally acceptable geometry of low minimum energy. Secondary structure prediction was performed using PSIPRED [27].

The residues lining the active site cavity are identified by implementing a grid-based algorithm proposed by Exner et al. [28] wherein a point on the cartesian space is considered to be part of a cavity if it has at least two grid points in the direction of cartesian axes within the molecule. The algorithm was validated by identifying cavities in various enzymes including cytochrome P450's and hence is expected to work well on human aromatase. This methodology is implemented in MDS (Molecular Design Suite) which has been used to find the active site residues in the present work [29].

Mutations of certain active site residues are performed using the homology module in Insight II environment wherein the side chain conformations are selected based on non-bonded energy criterion. The search algorithm for the possible conformations implements an iterative approach where the side chain producing the lowest non-bond energy for each residue is found. The side chain buried area (SBA) and fraction of the side chain exposed to polar atoms of the residues involved in inhibitor binding are analyzed by using the Verify command in the Profiles-3D program [24]. The SAS values for the residues are evaluated using the Solvation module wherein Delphi based model is implemented [30]. The solvaltion energy has electrostatic and non-polar terms which is assumed to be proportional to SAS of the solute. The CFF91 parameter sets are used for the calculations [31].

### *Docking*

Calculations were performed using the program Affinity in InsightII environment. Residues in the protein within 10 Å of the geometric center of substrate were kept flexible along with the ligand. Since, progesterone hydroxylase (template for the model) also exhibits binding to steroidal substrates, the structural changes between the complexed and uncomplexed (**PDB** code: **1NR6** and **1DT6** respectively) forms were analyzed by evaluating residue-by-residue RMSD (root mean

square deviation) after optimal structure alignment [32].

$$\mathrm{RMSD} = \sqrt{\sum_{i}^{n} d_i^2/n} \quad (1)$$

$d_i$ is the distance between the $i$th atoms and $n$ is the number of such distances.

The corresponding residues in the 3D-PSSM threading alignment of human aromatase were also kept flexible during the docking calculations together with the active site residues and the ligand. The calculations used ESFF force field along with cell multipole methods to estimate non-bonded interactions of enzyme and ligands. ESFF charges were assigned for both the ligand and enzyme. Initial docking positions were generated by Monte Carlo simulation for 1000 cycles using metropolis criteria of 10 kcal $\mathrm{mol}^{-1}$ energy range 1.0 Å maximum translation and 180° maximum rotation were used to generate best possible conformations using SA (simulated annealing) docking protocol of Affinity program [33]. Further, the docking positions were energy minimized by simulated annealing method. The complex, which gave the minimum energy, was finally selected for further analysis.

*MD simulation protocols*

The modeled protein was placed in a box of TIP3 pre-equilibriated water molecules with the box boundaries placed at least 10.0 Å from any given protein atom. The total number of solute and solvent atoms in the periodic box was 60,882. The simulations in the present study were performed using the second generation of the AMBER force-field and the SANDER module of the AMBER6.0 package. Long-range interactions were treated using the partical mesh ewald (PME) method. The PME charge grid spacing was ~1.0 Å, and the charge grid was interpolated using a cubic B-spline of fourth order, a direct sum tolerance of 0.00001, and a 8 Å direct space cut off. Constant temperature and pressure were maintained throughout the simulations using the Berendsen scaling algorithm with a time constant of 1.0 ps [34]. All bond lengths were constrained using the SHAKE algorithm. A time step of 1 fs was used to integrate the equations of motion. The solvated system was equilibrated by gradual heating to 300 K. The first 25 ps of dynamics was not considered for analysis. Simulation was carried out for 1 ns of MD. The resulting trajectories were analyzed using the CARNAL module of AMBER6.0 [35].

*Statistical procedures*

Procedures used for analysis such as residue-by-residue RMSD [32], information entropy ($S$) [36, 37] and Swissprot database analysis of aromatase sequences, were implemented by writing fortran codes.

## Results and discussion

The acidic residues lining the active site play a significant role in the aromatization mechanism of androgens and hence it is imperative to understand the characteristics of these residues. Seven acidic residues are found to be lining the active site cavity of human aromatase namely, E129, D222, E245, E302, D309, E379 and D380. However, we include D476 also in our analysis as earlier studies demonstrate it's importance and also due to it's proximity to the active site [38]. CLUSTALW multiple sequence analysis indicates that E129, E302, D309 and D380 are conserved in all aromatase sequences irrespective of type of species whereas E245, E379 and D476 are conserved mainly in mammalian aromatases [39]. However, E379 exhibits a conserved mutation to aspartate whereas E245 and E302 remain as glutamates. The structural environment of these residues plays an important role in affecting the aromatization mechanism. Generally, residues participating in larger number of suitable hydrogen bond donors having low fraction of side chain exposed to polar negative atoms can stabilize negatively charged carboxyl groups [19–21, 40]. Increasing burial of the residues may lead to favorable interaction with buried polar groups and consequently have a moderately stabilizing effect on the carboxylate ion [40]. However, larger solvent accessible surface area of the residues can result in the residues adopting different conformations which could lead to uncertainty in determining the carboxylate ion stability. In the present study, the sampling of conformational space will be assessed by evaluating

information entropy [36, 37] of SAS during an MD simulation.

*Active site acidic residues*

MD simulations in water indicate that the average H-bond occupancy of E245–Y241 and E245–S118 interaction is highest followed by E302–R115 and D222–L227. The residue T310 also exhibits fairly good H-bond occupancy with S314 and A306. These hydrogen bonding interactions involving the carboxylate oxygens of E245, E302 and D222 help in stabilizing the corresponding carboxylate ions. Another feature that seems to contribute to the carboxylate ion stability is it's location in a helix. Forsyth et al. [40] observed that carboxylate residues located at the N-terminal end of a helix seem to exhibit higher acidity. Only E302 is found to be located at the N-terminal end of I-helix whereas D309 is positioned around the center. The remaining residues are located either in a coil or C-terminal end of a helix. The burial of an acidic residue also contributes to the stability of the corresponding carboxylate ion. However, smaller burial can lead to a larger sampling of the conformational space which makes it difficult to assess the stability of the ion. Nevertheless, the conformational space spanned can be assessed by evaluating the SAS from MD simulations and quantifying the variability in the values in time. It is found that the SAS of E302/E245 is found to be the least followed by T310, D222 and D309 (Figure 1). The rest of the residues have a large SAS and consequently, can adopt different conformations in different crystal forms [40]. The conformational flexibility of the residues can be quantified by evaluating the SAS information entropy ($S$) over the entire MD simulation. For this, the values of SAS obtained from the simulations are initially grouped into bins of size 3 $Å^2$ beginning with 0 $Å^2$. $S$ (in bits) is subsequently given by

$$S = -\sum_i P_{ij} \log_2 P_{ij} \tag{2}$$

where $P_{ij}$ is the probability that the $j$th residue has the SAS value belonging to the $i$th bin. $P_{ij}$ are evaluated from

$$P_{ij} = \frac{n_{ij}}{N} \tag{3}$$

where $n_{ij}$ are the number of values present in the $i$th bin and $N$ is the total number of values. In

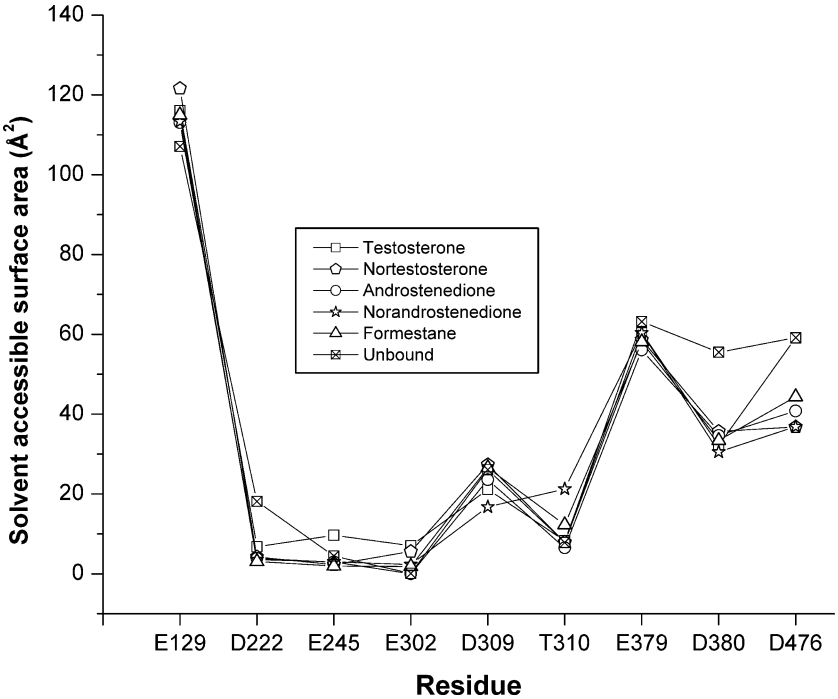


*Figure 1.* The values of SAS of the acidic residues lining the active site cavity of human aromatase.

general, a larger value of $S$ for a particular residue, indicates more conformational flexibility. It is found that conformational rearrangements are maximum in the case of E129, D476 and D380 (Figure 2), whereas D309 and D222 exhibit moderate flexibility. In contrast, E302, E245 and T310 display the least rearrangements implying less changes in the environment of the residues.

The polarity of the environment of each residue contributes in a major way in stabilizing the carboxylate ions of the acidic residues. Large number of polar negative atoms surrounding the side chain of the residues do not favor the carboxylate ion formation. We therefore analyze the environment of the acidic residues by evaluating the fraction of side chain exposed to polar atoms ($f$) which includes oxygen and nitrogen atoms. It should be noted that even the protonated nitrogens of basic amino acid residues which carry a positive charge are included in evaluating $f$. However, it is found that hardly any protonated nitrogens are located near the residues and even if present are located at a fairly large distance from the carboxylate oxygens. Therefore, evaluation of $f$ for these residues can be considered to involve only negative polar atoms. In general, higher $f$ values destabilize the anion which renders deprotonation unfavorable. MD simulations suggest that the average values of $f$ are low for E302, E245, D309 and D222 (Figure 3) and lowest for T310. To assess the possibility of a residue to attain variable values of $f$ which also indicates conformational flexibility, we again make use of Equation (2). However, $P_{ij}$ is now defined as the probability that the $j$th residue has the value of $f$ belonging to the $i$th bin. For this, the values of $f$ are classified into bins of size 0.01 starting from the value 0.0 and subsequently $P_{ij}$ are evaluated. It is found that D309 has the maximum entropy followed by D380. E302 and E245 are observed to exhibit the least entropy in $f$ (Figure 2). Maximum entropy for D309 implies more variation in $f$ in the entire MD simulation implying larger changes in the residue environment. Hence, although D309 has low negative polar atmosphere surrounding the side chain, it exhibits conformational flexibility.

*Effect of complexation to steroidal substrates*

The foregoing analysis is based on $f$, SAS, H-bond interaction and structural criteria reveal that E302

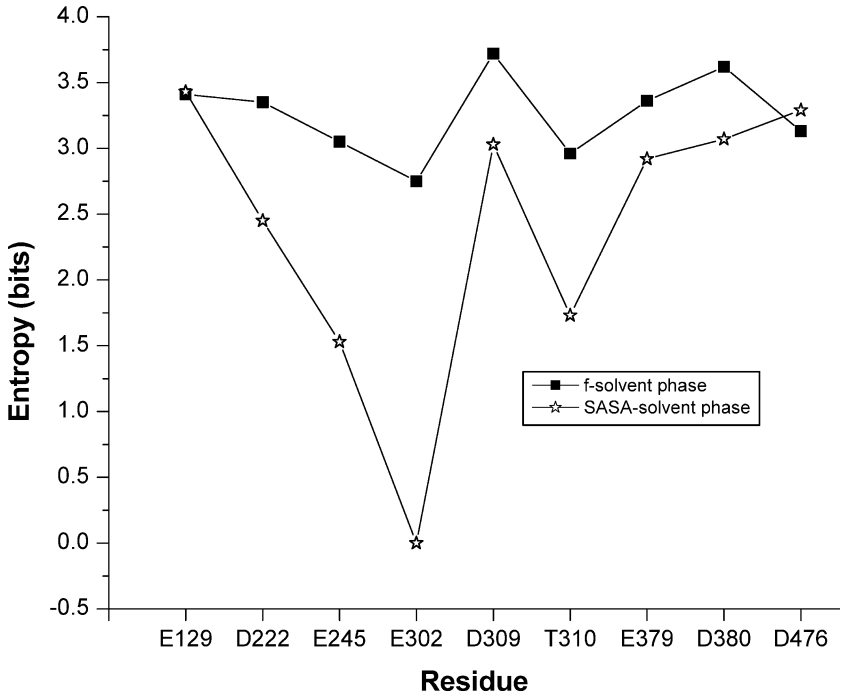


*Figure 2.* Information entropy over the SAS and values of $f$ obtained from MD simulations.

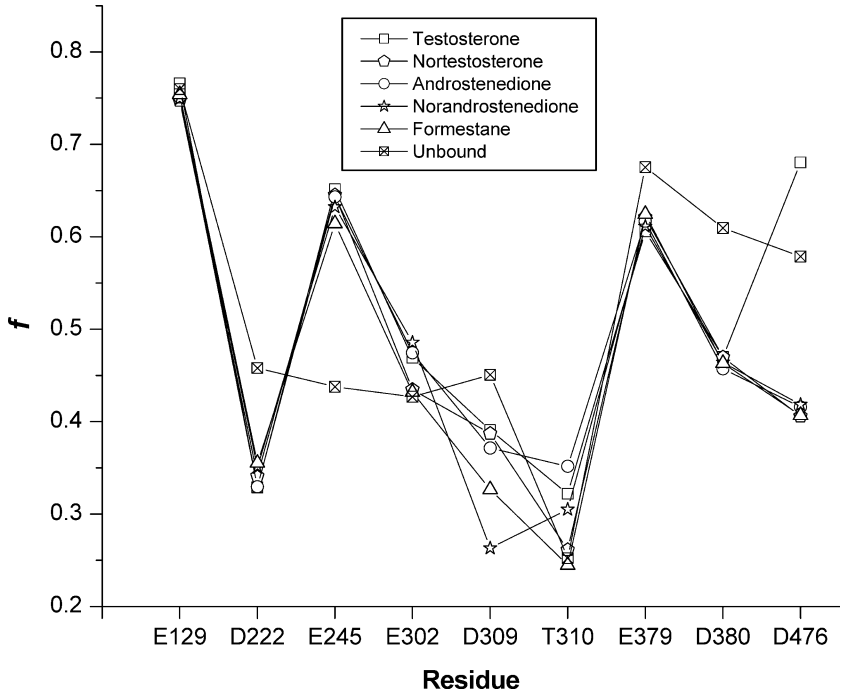


*Figure 3*. Fraction of side chain exposed to polar negative atoms (*f*) of acidic residues in the active site of human aromatase.

and E245 carboxylates would be stabilized more than the other residues. This is indeed surprising since many of the previous studies have indicated the important role of D309 in the proton transfer process during the aromatization of androgens which results in the formation of D309 carboxylate ion [13, 14]. This indicates that D309 could undergo structural changes essentially due to substrate binding which favor carboxylate ion formation. Recently, Pochapsky co-workers [41] have reported conformational changes in the active site residues upon ligand binding in an analogous cytochrome $P450_{cam}$ enzyme (CYP101). Therefore the properties of the acidic residues have been evaluated in the presence of different substrates i.e., androstenedione, testosterone, nor-androstenedione, nor-testosterone and the inhibitor formestane (Figure 4) by performing docking calculations. Care was taken to analyze frames obtained from docking calculations, having energies close to the lowest energy conformation, before obtaining a consensus. Two methods are implemented to calculate the properties which involving (i) removal of the ligand from the docked structure by retaining the protein conformation (RL) and (ii) keeping the ligand in the complex (KL). The analysis was found to be essentially in RL and KL for all the active site acidic residues except for D309 which will be deliberated in the following discussion. Therefore, barring analysis involving D309, the method adopted will not be mentioned in the following discussion. Similar values obtained by both the RL and KL procedures imply that the observations are chiefly due to changes in protein structure accompanying ligand binding rather than the ligand itself.

Generally, it is observed that the carboxylate oxygens of the residues participate in larger number of H-bond interactions following substrate binding. Interestingly, it is found that D309 forms H-bonds with T310 following binding of the enzyme with androgen or nor-androgen substrates. However, this interaction is not present in the lowest energy formestane (inhibitor) complex obtained from docking calculations [8]. Another interesting feature observed is that D222, D309, D380 and D476 exhibit decrease in the polarity of the atmosphere surrounding the side chain whereas E245 shows a substantial increase. The values of *f* for D309 (also T310) obtained by KL (Figure 5) are much lower than

those obtained by RL implying that in addition to structural changes due to complexation, the substrate itself directly influences the polar atmosphere surrounding this residue. In particular, the values of $f$ are now found to be least for D309 followed by D222, D476 and E302

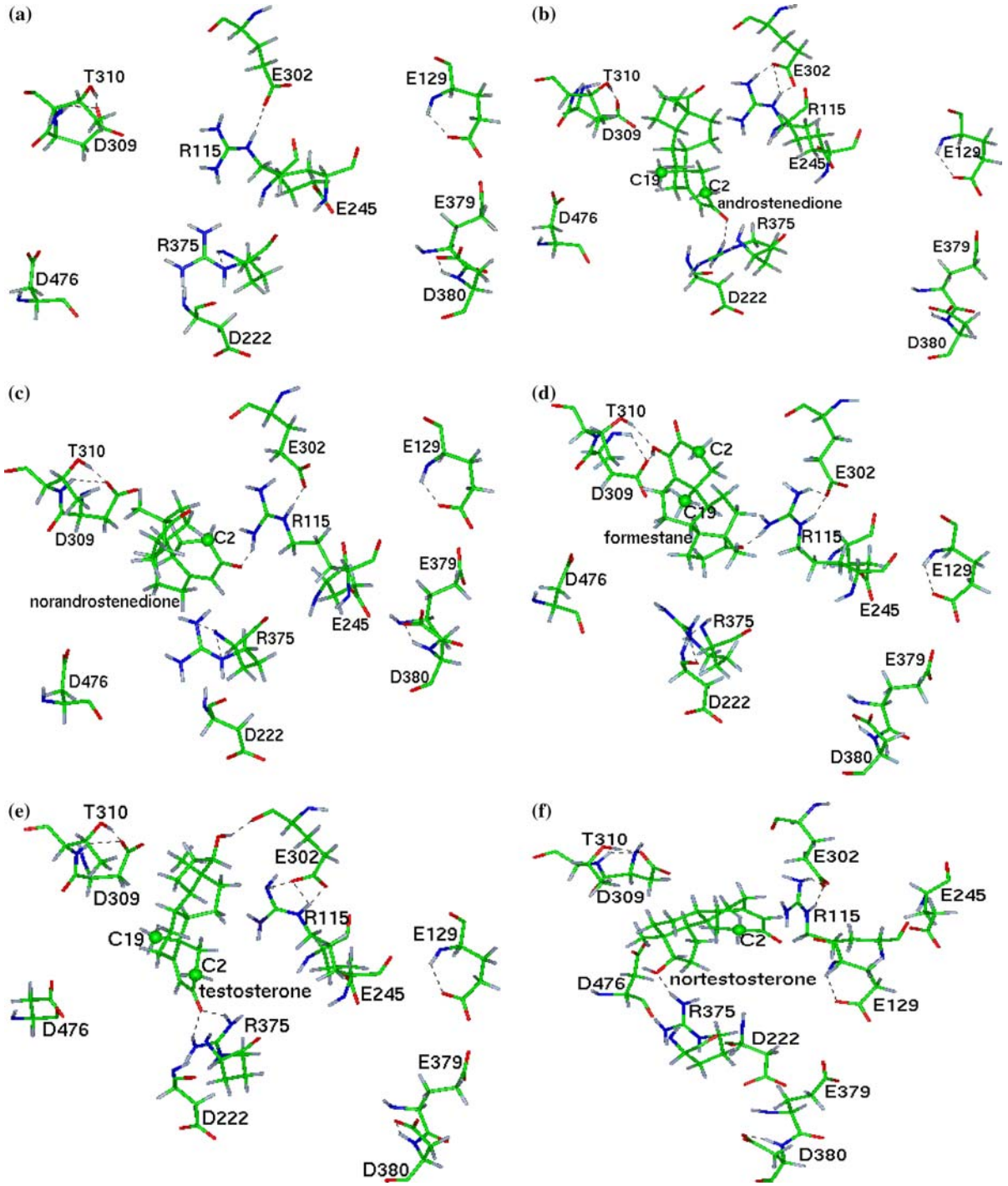


*Figure 4*. The active site acidic residues when complexed to ligands. (A) is the active site prior to substrate binding.

(Figure 3) and the largest values are found for E129 and E245. Analysis based on SAS shows marked difference between the RL and KL procedures for the residue D309, while the results are almost similar for the remaining residues. D309 exhibited considerable decrease in SAS obtained from KL procedure. The SAS obtained from RL method remained almost identical to the uncomplexed form for this residue. This implies that the solvent accessibility is largely defined by the presence of a substrate for this residue. In contrast, the SAS values of D476 and E379 do not change significantly on complexation (Figure 1). Moreover, the SAS of E302 shows a slight increase, although still on the lower side, whereas the SAS of D222 decreased. In general, E302, D309, D222, E245 and T310 exhibit lowest SAS values on complexation.

The present analysis indicates that complexation with steroidal substrates results in structural changes rendering the environment of D309 favorable for carboxylate ion formation. This is in agreement with earlier site directed mutagenesis studies which showed decrease in enzymatic activity with androgens on mutating this residue [6, 42]. However, the SAS values of D309 show significant decrease only in the presence of steroidal substrates implying that the increasing burial of this residue (D309) due to the steroids would lead to favorable interactions between the charged carboxyl groups and buried polar groups in the enzyme. Although, the structural changes on complexation, surrounding E302 are minimal as seen from the values of $f$ and SAS, it exhibits increase in number of H-bonds making it favorable for carboxylate ion formation. The involvement of this residue which plays different roles in androgen and nor-androgen aromatization processes has been deliberated in previous experimental studies [10, 43]. The larger stabilization of D309 anion is in accord with the mechanisms proposed by Auvray et al. [10] according to which E302 reprotonates in one step lesser than D309 by assisting in hydroxylation of the C1 position during aromatization of nor-androgens. The residue D476 exhibits similar trends as D309 on complexation. However, the values of $f$ for D476 are more than 1.3 times that of neutral D309

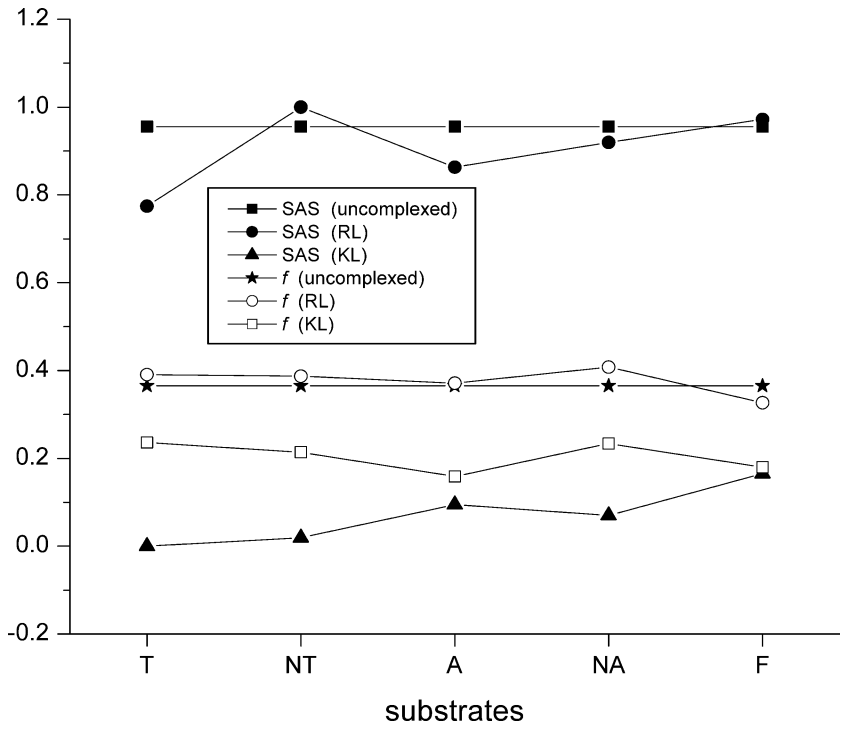


*Figure 5*. The SAS (in Å$^2$; scaled by the highest SAS value obtained for the residue) and $f$ values obtained from RL and KL methods for the residue D309. T – Testosterone, NT – nortetsosterone, A – androstenedione, NA – norandrostenedione, F – formestane.

indicating repulsion between D309 anion and D476 which is a factor facilitating aromatization as proposed by Auvray et al. [10]. Similar to the favorable interaction between D476 and D309, E302 might have such an interaction with E245 since the *f* values show similar trend. It is seen that E245 has *f* value similar to E302 in the uncomplexed state which is increased to more than 1.3 times upon complexation. Moreover, the carboxylate oxygens of the two residues are placed at analogous distance and are oriented towards each other. In addition, CLUSTALW multiple sequence analysis demonstrates that the residues E245 and D476 are conserved only in mammalian aromatases implying similar functionality of the two residues [39]. The residues D380 and D222 exhibit decrease in the polar atmosphere as well as SAS on steroid binding. In contrast, E379 and E129 are not affected on complexation with substrates.

*Proximity of substrate with the acidic residues*

The other important factor which determines the proton transfer process is the proximity of these residues to the key sites of the androgen and nor-androgen substrates. In the initial stages of the enzymatic mechanism, D309 (in the case of androgens) and E302 (in the case of nor-androgens) seem to lose a proton resulting in the formation of carboxylate anion [14]. This was substantiated from earlier mutational studies wherein the D309A mutant induced activity loss with androstenedione as well as testosterone. Nevertheless, the mutant exhibited activity with nor-testosterone contrary to the mutant E302A which displayed activity loss with all the three substrates [10]. Therefore, we examine the proximity between the closest carboxylate oxygens of these acidic residues and C2/C19 of the ligands which are considered to be involved in proton transfer process during the catalytic mechanism. It should be noted that C19 is absent in nor-androgens. Analysis of androgen complexes indicates that C19 is closest to the carboxylate oxygen of D309 (Figure 6) followed by E302 and D476 which have similar values. Placed at slightly larger distances are D222 and E245. C2 which is present in all the substrates exhibits shortest distances in nor-androgen complexes with E302 (Figure 7). However, in androgen complexes C2 exhibits proximity to both E302 as well as D309. Nonetheless, C19 being closer to D309 than C2 in androgen complexes, the initial protonation takes place at C19. The above analysis corroborates with the mechanisms proposed by Auvray et al. [10] wherein C2 was considered to be involved in the proton transfer process in the case of aromatization of nor-androgens and C19 in the case of androgens. However, D476 is located further away from C2 in all the complexes but is situated closer to C19 in the androgen complexes. This indicates that during androgen aromatization, besides D309, D476 plays a major role as suggested by earlier site directed mutagenesis studies [10].

Although androstenedione is similar in structure to formestane except for a hydroxyl group at C4, the later inhibits the aromatase activity. It is found that in the formestane complexed structure, C2 and C19 are closest to D309 followed by E302. However, D476 is located fairly away from these atoms. Since D476 is displaced away when formestane is bound which is an inhibitor of the enzyme, the above analysis further substantiates the role of D476 besides D309 in androgen aromatization. Earlier experimental observations indicated that the presence of the C4-hydoxyl group in formestane changes the binding orientation of the inhibitor by displacing D476 which was expected to bring the steroid closer to D309 and T310 [8]. This aspect is observed in our model wherein, docking studies reveal that C19 of this inhibitor is at a larger distance from D309 oxygen compared to androstenedione (and testosterone) complex. Similarly, C2 is placed at a larger distance from E302 compared to C2 of androstenedione. However, a feature observed in our model is that the hydroxyl oxygen and hydrogen of formestane form a critical hydrogen bonding interactions with the hydroxyl hydrogen of T310 and carboxylate oxygen of D309 respectively (Figure 4). This hydrogen bond inhibits participation of the hydroxyl group of threonine in the heme ferroxy radical formation.

*Mutational analysis*

Earlier site directed mutagenesis studies revealed that the D309A mutant exhibited higher activity loss than the D309N mutant, with androgens as substrates [8]. However, it was observed that D309A mutant showed similar activity as the wild type with nor-androgen substrates. Calculations

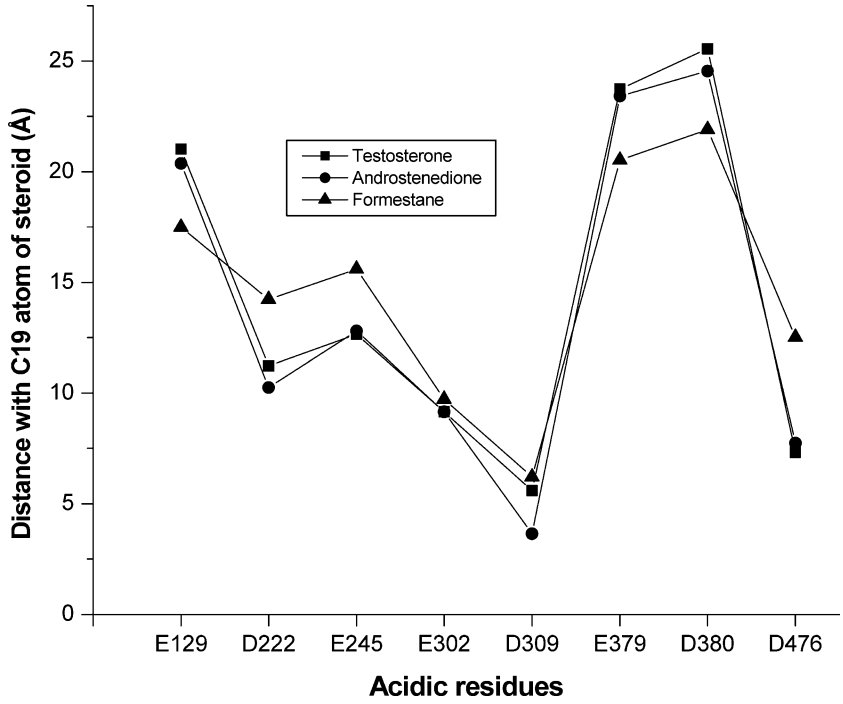


*Figure 6*. Distance between the carboxylate oxygen of active site acidic residues and C19 of C19-androgens as well as the inhibitor formestane.

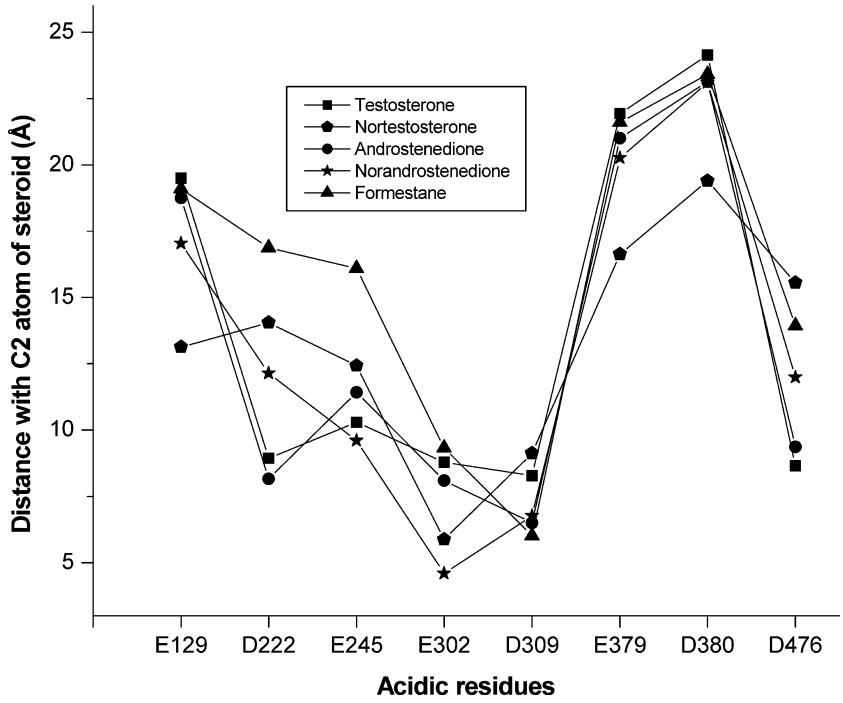


*Figure 7*. Distance between the carboxylate oxygen of active site acidic residues and C2 of steroids and formestane.

on the modeled mutant D309A showed decrease in *f* and SBA in the androgen complexes whereas D309N exhibited marginal decrease in *f* value and increase in SBA. In addition, the androgen complexes of D309N mutants exhibited H-bond interaction between D309N side chain oxygen and T310. Very similar values of *f* for the wild type and the mutant enzymes along with the H-bond interaction between N309 and T310 in the mutant enzyme complexed to androgens, accounts for the D309N mutant to exhibit 3–5 times the activity of D309A as seen from earlier site directed mutagenesis studies [8, 10]. Similar observation is also seen in nor-androgen complexes. However in the nor-androgen complexes, the D309A mutant showed H-bond interaction between alanine and T310 but with reasonable decrease in *f* which could result in activity loss. Further, the proximity criterion discussed earlier could also be a major factor for the D309A mutant to show similar activity as the wild type enzyme in the case of nor-androgen binding since E302 is expected to play a major role in nor-androgen aromatization process. Earlier site directed mutagenesis studies revealed that E302D mutant exhibited 30% activity of wild type enzyme, whereas, E302Q had minimal activity [42, 10]. Mutating E302 with D exhibited decrease in SBA and increase in *f* in all the androgen/nor-androgen complexes except in the nor-androstenedione complex wherein the value was similar to the wild type enzyme. In contrast, SBA of E302Q remained similar to the wild type whereas the values of *f* decreased. Therefore, the analysis indicates the central role of *f* and SBA in controlling the enzyme activity at the position 302 in both androgen and nor-androgen complexes. Single mutations of D476 with A, K, L, N and E leads to increase in the polar atmosphere with simultaneous lowering of SBA compared to the wild type enzyme. Earlier, mutagenesis studies had shown decrease in enzymatic activity in all the mutants except D476E which had activity similar to wild type enzyme [10]. Hence, the presence of an acidic residue along with a highly polar atmosphere surrounding this position, seems to favor enzyme activity. The increase in *f* for D476E indicates the importance of repulsive interaction with D309.

Earlier studies by Amarneh et al. [6] implied that the residue R365 in the K-helix is buried internally in a hydrophobic region. This residue is found to be highly buried in the model with moderately polar environment, which may be essentially due to the presence of a conserved glutamate [6]. This highly buried nature of the residue seems to play a critical role in enzyme catalysis since earlier site directed mutagenesis showed that this residue cannot be replaced even with lysine which may be traced to a marked decrease in the SBA values [13]. Earlier mutational studies of Q225 with N and A exhibited increased affinity to the substrate which was attributed to easier substrate access to the heme pocket [6]. Interestingly, the modeled mutant Q225N exhibits decreased buried side chain area, which is more prominent in the case of the alanine mutant. The large decrease in the SBA of Q225A exhibited by our model is in tune with increase in $V_{max}$ as observed by the authors indicating easier ejection of the product out of the channel [6]. Similar observations have been made with respect to the L228A mutant. A slight decrease in SBA is observed in the L228E mutant for the glu residue. Another interesting observation from the atomic model is that a cation-$\pi$ interaction is observed between K130 and F134 which could explain the steric constraints posed by this entity affecting enzymatic activity with respect to testosterone but not with nor-testosterone as studied in previous mutagenesis analysis [10]. The polar area formed by the residues H128, Q218, Q225 and H475 is evident by high values of *f* evaluated from the model [10]. Mutation of V288 to glutamate resulted in complete loss of activity as observed by Amarneh et al. [6] This residue was considered to form a hydrophobic glue that holds the main body of the $\beta$-sheet domain. This residue is found to be highly buried with low *f* and probably plays a major role in imparting stability to the hydrophobic pocket.

## Conclusions

The present study delineates the role of eight acidic residues (E129, D222, E245, E302, D309, E379, D380) lining the active site cavity of human aromatase, including D476 which lies in proximity to the cavity. Mammalian CYP 2C5 crystal structure was taken as a template for building the 3-D model of aromatse, which is followed by docking, molecular dynamics and statistical anal-

ysis. The structural environment of these acidic residues is explored through H-bond, SAS and *f* to assess the stability of the corresponding carboxylate anions following which we infer that the environment of the residues E245, E302 and D222 stabilizes the carboxylate ion formation in the uncomplexed form. It is interesting to note that the stability of D309, D222 and D476 anions increases upon complexation to steroidal substrates. In particular, the interaction between D309 and T310, which assists proton transfer, is found to be formed following androgen/nor-androgen complexation. While substrate binding leads to substantial changes in the protein side chains, the bound substrate is observed to be selectively clamping the aspartate residue D309. Triggering the burial of D309 upon steroidal binding may facilitate the stabilizing electrostatic interactions of the carboxylate of D309 with the polar environment in the enzyme. The structural changes consequent to steroidal complexation, surrounding E302 are minimal as seen from the values of *f* and SAS, but the number of H-bonds are enhanced facilitating the carboxylate ion formation. Information entropic analysis indicates that the residues D309, D222 and D476 have more conformational flexibility compared to E302 and E245 prior to substrate binding. Interaction similar to that between D476 and D309, which is expected to assist androgen aromatization, is proposed between E302 and E245. Interestingly, D476, E245 and E379 are conserved only in mammalian aromatase sequences implying similar functionality of the two residues [36]. The variations in the structural parameters of the glutamate residues E379 and E129 are insignificant. The inhibition of aromatase activity by 4-hydroxy androstenedione (formestane) is attributed to a critical hydrogen bond formation between the hydroxy moiety and T310/D309 and the large distance from D476. Structurally, the distance between side chain oxygen of D309 and C19 is found to be very close in the androgen complexes as compared to the distance D309 and C2. However, in the nor-androgen complexes wherein C19 is absent, C2 is found to be very close to side chain oxygen of E302. Hence, D309 and E302 are expected to play a major role in the proton transfer processes of androgen and nor-androgen aromatization respectively. Our observations corroborate well with the mechanisms proposed by Auvray et al. [10] and suggest the importance of the role played by E245, especially in aromatization of nor-androgens. In addition, the model is supported by many of the earlier site directed mutagenesis studies [38, 39]. Therefore, the presence of steroids triggers suitable changes in the environment of the active site acidic residues in such a way that carboxylate ion formation is favored and consequently assists the mechanism of aromatization.

## Acknowledgements

JNM. and NM. acknowledge financial assistance from UGC and CSIR respectively. GMS thanks DST for financial support.

## References

1. Thijssen, H.H., Maturitas, 49 (2004) 25.
2. Yue, W., Santen, R.J., Wang, J.P., Li, Y., Verderame, M.F., Bocchinfuso, W.P., Korach, K.S., Devanesan, P., Todorovic, R., Rogan, E.G. and Cavalieri, E.L., J. Steroid Biochem. Mol. Biol., 86 (2003) 477.
3. Meunier, B., de Visser, S.P. and Shaik, S., Chem. Rev., 104 (2004) 3947.
4. Akthar, M., Robichund, P.L., Akthar, M.E. and Wright, J.N., J. Steroid Biochem. Mol. Biol., 61 (1997) 127.
5. Murthy, J.N., Rao, A.R.R. and Sastry, G.N., Curr. Med. Chem-Anti-Cancer Agent, 4 (2004) 523.
6. Amarneh, B., Corbin, C.J., Peterson, J.A., Simpson, E.R. and Grahm-Lorence, S., Mol. Endocrinol., 7 (1993) 1617.
7. Chen, S., Zhou, D., Swiderek, K.M., Kodohama, N., Osawa, Y. and Hall, P.F., J. Steroid. Biochem. Mol. Biol., 44 (1993) 347.
8. Kao, Y.C., Cam, L.L., Laughton, C.A., Zhou, D. and Chen, S., Cancer Res., 56 (1996) 3451.
9. Kodohama, N., Yorborough, C., Zhou, D., Chen, S., Osawa, Y. and Hall, P.F., J. Steroid. Biochem. Mol. Biol., 43 (1992) 693.
10. Auvray, P., Nativelle, C., Bureau, R., Dallemagne, P., Seralini, G.E. and Sourdaine, P., Eur. J. Biochem., 269 (2002) 1393.
11. Laughton, C.A., Zvelebil, M.J. and Needle, S., J. Steroid Biochem. Mol. Biol., 44 (1993) 399.
12. Blobaum, A.L., Harris, D.L. and Hollenberg, P.F., Biochemistry, 44 (2005) 3831.
13. Chen, S. and Zhou, D., J. Biol. Chem., 267 (1992) 22587.
14. Grahm-Lorence, S., Amarneh, B., White, R.E., Peterson, J.A. and Simpson, E.R., Protein Sci., 4 (1995) 1065.
15. Cavalli, A., Greco, G., Noellino, E. and Recanatini, M., Bioorg. Med. Chem., 8 (2000) 2771.
16. de Graaf, C., Vermeulen, N.P.E. and Feenstra, K.N., J. Med. Chem., 48 (2005) 2725.
17. Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F. and McRee, D.E., J. Inorg. Biochem., 81 (2000) 183.

18. Wester, M.R., Johnson, E.F., Marques-Soares, C., Dijols, S., Dansette, P.M., Mansuy, D. and Stout, C.D., Biochemistry, 42 (2003) 9335.
19. Matthew, J.B., Gurd, F.R.N., Garcia-Moreno, E.B., Flanagan, M.A., March, K.L. and Shire, S.J., CRC Crit. Rev. Biochem., 18 (1985) 91.
20. Schutz, C.N. and Warshel, A., Proteins, 44 (2001) 400.
21. Warshel, A. and Russel, S., Q Rev. Biophys., 17 (1984) 283.
22. Martin, A.C., MacArthur, M.W. and Thortan, J.M., Proteins (Suppl), (1997) 14.
23. Sali, A. and Blundell, T.L., J. Mol. Biol., 234 (1993) 779.
24. Luthy, R., Bowie, J.U. and Eisenberg, D., Nature, 356 (1992) 83.
25. Kelley, L.A., Mac Callum, R.M. and Sternberg, M.J.E., J. Mol. Biol., 299 (2000) 501.
26. Shi, S., Yan, L., Yang, Y. and Fisher, J., ESFF Force Field Project Report II, Accelrys, San Diego.
27. Jones, D.T., J. Mol. Biol., 292 (1999) 195.
28. Exner, T., Keil, M., Moeckel, G. and Brickmann, J., J. Mol. Model., 4 (1998) 340.
29. MDS 1.0: Molecular Design Suite, Vlife Sciences Technologies Pvt. Ltd., Pune, India, 2003.
30. Sharp, K.A., Nicholls, A., Friedman, R. and Honig, B., Biochemistry, 30 (1991) 9686.
31. Schmidt, A.B. and Fine, R.M., Mol. Simulat., 13 (1994) 347.
32. Bindu, P.H., Sastry, G.M., Murty, U.S.N. and Sastry, G.N., Biochem. Biophys. Res. Commun., 319 (2004) 312.
33. Luthy, B.A., Wasserman, Z.R., Stouten, P.F.W., Hodge, C.N., Zacharias, M. and McCammon, J.A., J. Comp. Chem., 16 (1995) 454.
34. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W., DiNola, A. and Haak, J.R., J. Chem. Phys., 81 (1984) 3684.
35. Pearlman, D.A., Case, D.A., Cadwell, J.W., Ross, W.S., Cheatham, T.E.I., DeBolt, S. and Ferguson, D., Comput. Phys. Commun., 2 (1995) 287.
36. Shannon, C.E., Bell Syst. Tech. J., 27 (1948) 379.
37. Gadre, S.R., Phys. Rev A., 30 (1984) 620.
38. Auvray, P., Moslemi, S., Sourdaine, P., Seralini, G.E., Enguehard, C., Dallemagne, P., Sonnet, P., Bureu, P. and Rault, S., Eur. J. Med. Chem., 33 (1998) 451.
39. Thompson, J.D., Higgins, D.G. and Gibson, T.J., Nucleic Acids Res., 22 (1994) 4673.
40. Forsyth, W.R., Antosiewicz, J.M. and Robertson, A.D., Proteins, 48 (2002) 388.
41. Wei, J.Y., Pochapsky, T.C. and Pochapsky, S.S., J. Am. Chem. Soc., 127 (2005) 6974.
42. Zhou, D., Korzekwa, K.R., Poulos, T. and Chen, S., J. Biol. Chem., 267 (1992) 762.
43. Kao, Y.C., Korzekwa, K.R., Loughton, C.A. and Chen, S., Eur. J. Biochem., 268 (2001) 243.