

Polymorphism on human aromatase affects protein dynamics and substrate binding: spectroscopic evidence.

Giovanna Di Nardo

Università degli Studi di Torino: Università degli Studi di Torino

Almerinda Di Venere

University of Rome Tor Vergata: Università degli Studi di Roma Tor Vergata

Chao Zhang

Università degli Studi di Torino: Università degli Studi di Torino

Eleonora Nicolai

University of Rome Tor Vergata: Università degli Studi di Roma Tor Vergata

Silvia Castrignanò

Università degli Studi di Torino: Università degli Studi di Torino

Luisa Di Paola

Campus Bio-Medico University: Università Campus Bio-Medico di Roma

Gianfranco Gilardi

Università degli Studi di Torino: Università degli Studi di Torino

Giampiero Mei (✉ mei@med.uniroma2.it)

University of Rome Tor Vergata: Università degli Studi di Roma Tor Vergata

Research

Keywords: aromatase polymorphism, ligand binding, fluorescence, molecular modeling

Posted Date: April 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-385816/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Biology Direct on April 26th, 2021. See the published version at <https://doi.org/10.1186/s13062-021-00292-9>.

Polymorphism on human aromatase affects protein dynamics and substrate binding: spectroscopic evidence.

Giovanna Di Nardo^{1,#}, Almerinda Di Venere^{2,#}, Chao Zhang¹, Eleonora Nicolai², Silvia Castrignanò¹, Luisa Di Paola³, Gianfranco Gilardi^{1,*}, Giampiero Mei^{2,*}

¹Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, Via Accademia Albertina 13, 10123 Turin, Italy

²Dipartimento di Medicina Sperimentale, Università di Roma Tor Vergata, Via Montpellier 1- 00133 Rome, Italy

³Dipartimento di Ingegneria, Unità di Fondamenti Chimico-Fisici dell'Ingegneria Chimica, Università Campus Bio-Medico di Roma, via Álvaro del Portillo 21, 00128 Rome, Italy

These authors contributed equally to this work

*Corresponding authors: mei@med.uniroma2.it, gianfranco.gilardi@unito.it

Abstract

Human aromatase is a member of the cytochrome P450 superfamily, involved in steroid hormones biosynthesis. In particular, it converts androgen into estrogens being therefore responsible for the correct sex steroids balance. Due to its capacity in producing estrogens it has also been considered as a promising target for breast cancer therapy. Two single-nucleotide polymorphisms (R264C and R264H) have been shown to alter aromatase activity and they have been associated to an increased or decreased risk for estrogen-dependent pathologies. Here, the effect of these mutations on the protein dynamics is investigated by UV/FTIR and time resolved fluorescence spectroscopy. H/D exchange rates were measured by FTIR for the three proteins in the ligand-free, substrate- and inhibitor-bound forms and the data indicate that the wild-type enzyme undergoes a conformational change leading to a more compact tertiary structure upon substrate or inhibitor binding. Indeed, the H/D exchange rates are decreased when a ligand is present. In the variants, the exchange rates in the ligand-free and –bound forms are similar, indicating that a structural change is lacking, despite the single amino acid substitution is located in the peripheral shell of the protein molecule. Moreover, the fluorescence lifetimes data show that the quenching effect on tryptophan-224 observed upon ligand binding in the wild-type, is absent in both variants. Since this residue is located in the catalytic pocket, these findings suggest that substrate entrance and/or retention in the active site is partially compromised in both mutants. A contact network analysis demonstrates that the protein structure is organized in two main clusters, whose connectivity is altered by ligand binding, especially in

correspondence of helix-G, where the amino acid substitutions occur. Our findings demonstrate that SNPs resulting in mutations on aromatase surface modify the protein flexibility that is required for substrate binding and catalysis. The cluster analysis provides a rationale for such effect, suggesting helix G as a possible target for aromatase inhibition.

Keywords: aromatase polymorphism, ligand binding, fluorescence, molecular modeling

Background

Structural flexibility is a crucial property of proteins, since it allows the molecular rearrangements required for most of their biological functions. Such rearrangements include large conformational changes (i.e. those characterizing cargo and contractile proteins), long distance displacements (for instance, those occurring in some integral membrane receptors), or even smaller, local conformational changes (as those required for the action of most enzymes). Flexibility allows substrates binding, products release and it is needed to confer allosteric properties to enzymes that control and regulate metabolic pathways and cells growth. Cytochromes P450 are a very good example of such enzymes as they are heme-containing monooxygenases involved in key metabolic pathways and xenobiotic detoxification. Moreover, they are known to undergo conformational changes to allow substrate access, catalysis and product release [1]. In particular, structural data demonstrate that some elements such as the F- and G- helices and the loop connecting them (F-G loop) are highly flexible and can open and close channels connecting the surface of the protein to the active site [2, 3].

Human cytochromes P450 are highly polymorphic and some single nucleotide polymorphisms (SNPs) are known to affect the enzyme activity by decreasing the catalytic efficiency. Although the effect of many SNPs can be rationalized by the location of the mutations in the protein structure, for others it is very difficult to predict. This is the case of two SNPs present in human aromatase, a cytochrome P450 with a key role in steroidogenesis as it catalyzes the conversion of androgens into estrogens [4]. The correct activity of this protein is crucial, because alterations in the equilibrium between androgens and estrogens, hypo- or hyper-production of estrogens are associated to a series of disorders that range from neurological diseases to breast cancer [5]. We have previously characterized the common R264C (rs700519) and the rarer R264H (rs2304462) polymorphic variants showing that the activity of these variants is modified by an alteration of the kinetic parameters [6, 7] and by a different phosphorylation propensity, as R264 is part of a consensus sequence for phosphorylation [6]. However, this residue lies on the protein surface, being part of the G-helix and

it is therefore difficult to understand how its mutation can impact the catalytic parameters only based on the location.

We previously demonstrated that aromatase requires dynamic flexibility to bind and process its substrates [8]. In this paper we have used a combination of experimental techniques (namely infrared and fluorescence spectroscopy) to evaluate the impact that a single point mutation (R264C or R264H) on the external surface of human aromatase has on its tridimensional structure and on its biological function. In parallel to such experimental approach, a contact network analysis of the protein structure has been also performed. In the last decade the application of this computational methodology to proteins has allowed to look at these complex molecules from a new perspective, since such kind of algorithms use a few descriptors to characterize the spatial inter-connections among amino acids. In this way, clusters of residues are localized within the protein tridimensional structure and long-range interactions, cooperative effects and segmental mobility can be predicted. The applications of this methodology span from protein-protein interaction [9, 10] to allosteric regulation of enzymes [11, 12], allowing, in very recent researches, to speculate on possible binding sites for drugs against degenerative diseases [13] or viruses [14]. We demonstrate that the combination of this approach with experimental data might suggest a molecular explanation to the loss of biological activity observed in R264C and R264H variants.

2. Results

2.1 H/D exchange followed by FTIR

Fourier transform infrared spectroscopy (FTIR) was applied to determine the kinetics of H/D exchange in Recombinant human aromatase (Aro) and its polymorphic variants, to detect possible differences in the kinetics of deuteration arising from protein conformational changes induced by ligand binding. Fig. 1 shows the FTIR spectra recorded at the beginning and at the end of a typical H/D exchange kinetic experiment, in the case of substrate-free Aro.

The band between 1670 cm^{-1} and 1620 cm^{-1} has been assigned to the amide I, whereas the band between 1470 cm^{-1} and 1420 cm^{-1} has been attributed to the amide II. The H/D exchange was thus followed as the increase of the absorbance of the amide II band at 1460 cm^{-1} (Fig. 1, inset), in a time range of 160 minutes.

The time dependence of the fraction of protons exchanged shows a dramatic increase in the first ten minutes, while a much slower kinetics takes place in the next 2-3 hours (Fig. 2). The data are, in fact, best fitted to a double exponential curve (cfr. Methods), which yields two distinct rate values (Table

1). The Aro wt sample is characterized by the faster initial transition ($k_1 \approx 0.4 \text{ min}^{-1}$), a process that is considerably slowed down by ligand binding (Fig. 1, panel A), the k_1 value being reduced to one half both in the presence of the substrate and the inhibitor molecules (Table 1). No significant differences are instead observed in the long-term component (k_2), indicating that the slower part of the kinetic is independent of the binding process. In the H/D exchange kinetics, unlike Aro wt, the presence of the substrate androstenedione and the inhibitor anastrozole on polymorphic variants is less relevant (Fig. 2B, C). In particular, the raw data are characterized by a slower increase in the initial part of the process ($t \approx 0\text{-}20 \text{ min}$, Fig. 2), whose fit yields a lower rate constant ($k_1 \approx 0.2 \text{ min}^{-1}$, Table 1), with respect to that of the wt protein. Again, the introduction of androstenedione or anastrozole in both mutants does not affect the second step of the kinetics (Fig. 2B, C) and, in fact, no significant changes are observed in the second rate constant value ($k_2 \approx 0.016 \text{ min}^{-1}$, Table 1).

The presence of a complex kinetics (Table 1) indicates a heterogeneity in the protons populations. Indeed, a bi-phasic behaviour of the H/D exchange process in proteins is generally explained in terms of two groups of peptide hydrogens, one more accessible to the solvent molecules (and thus subject to a higher exchanging rate) and a second one located in more buried protein segments, less accessible to water (15-18). Since the H/D exchange process at the protein surface would be too fast to be detected, the two rate constants must be assigned to partially exposed (k_1) or buried (k_2) residues protons. In such a context, a relevant meaning may be attributed to the pre-exponential factors m_i of the respective fits, as they represent the initial fractions (i.e. at time = 0) of each protons population. As shown in Table 1, the m_1 value of the wt protein is the highest (≈ 0.60) in the absence of any ligand, and the lowest (≈ 0.20) when one of the ligands is present, thus indicating that a large change, $\Delta m_1 \approx -65 \%$, occurs in the population of the partially exposed protons upon binding.

On the contrary, the decrease observed in the case of the two mutants results to be smaller ($\Delta m_1 \approx -40\%$ for R264H and $\Delta m_1 \approx -23\%$ in the case of R264C), suggesting less pronounced structural effects induced by ligand binding.

2.2 *Intrinsic fluorescence dynamics*

Ligand-induced changes in the intrinsic fluorescence of aromatase have been characterized through the phase-shift and demodulation technique and the results of data analysis are shown in Fig. 3. As previously observed [8], the protein emission decay is quite heterogeneous, being characterized by four distinct lifetime components, ranging from tenths of nanoseconds to 6-9 nanoseconds. Such a complexity is due to the presence of five tryptophan residues, located in different environments of the protein scaffolding. The main feature of the wt sample in the presence of either substrate or inhibitor is a large decrease of the fractional contribution at $\approx 5 \text{ ns}$ (Fig. 3A). Such a considerable

quenching effect has been attributed by previous fluorescence measurements to the interaction between the substrate and a specific tryptophan residue, namely W224, which is buried in the core of the protein matrix [8].

As shown in Fig. 3B and 3C, no changes were instead detected in the fractional contribution at 5 ns of both R264H and R264C upon ligands addition, indicating that in the mutants the binding of either substrate or inhibitor molecules does not produce significant effect in the environment of W224, suggesting a different local conformation with respect to that of the wt cavity.

This hypothesis has been further tested by studying the protein rotational dynamics. The presence of the long lifetime component, in fact, allows to perform anisotropy measurements using the aromatase intrinsic fluorescence. The corresponding phase shift and demodulation data are reported in Fig. 4, for Aro wt and R264C or R264H mutants. The results demonstrate that both mutated forms display slower rotational dynamics, with respect to the wt protein, the data being shifted toward a lower frequency range.

In fact, the phase and demodulation curves of each aromatase type crosses each other at ≈ 150 , 63, and 93 MHz, respectively in the case wt, R264H and R264C, indicating that R264H is the form characterized by the slowest dynamics. Non-linear fitting of the phase shift and demodulation points reported in Figure 4 yielded the results illustrated in Figure 5. Each data set required two rotational correlation times, ϕ_1 , and ϕ_2 . The first one, ϕ_1 , ranges from 21ns (wt) to 33 ns (R264H) and it is compatible with the rotational motion of whole protein, as shown in the cartoon sketched in the left side of Fig. 5. Indeed, the predicted turning on a principal axis of a spherical hydrated molecule with the same size of aromatase (≈ 58000) would be $\phi_{\text{sph}} \approx 24$ ns [19]. Longer ϕ_1 values, such those characterizing R264C and R264H, suggest a more swelled tridimensional form, since an expanded tertiary structure would produce a slower motion.

In the case of the wt sample, the entrance of the substrate (or inhibitor) in the protein binding cavity produced a considerable decrease in the ϕ_1 value ($\approx -23\%$, Fig. 5), thus implying a faster rotational dynamics. This trend was also observed in the mutated forms but with two important differences: i) the extent of the relative change in ϕ_1 is less pronounced ($\approx -11\%$ and -15% , respectively); ii) in both samples, the ϕ_1 value remains above 23 ns, independently of the ligand (substrate or inhibitor) used (Figure 5).

The second rotational lifetime, ϕ_2 , is much shorter being on the order of ≈ 400 -500 ps, in the case of wt and R264C aromatase and ≈ 800 ps, as regards R264H (Figure 5). Such values reflect a faster dynamics (with respect to ϕ_1), that can be ascribed to the average local motions of tryptophan(s) environment. The presence of the ligands in the binding pocket slows down such movements in the case of the wt and the R264C samples, the substrate molecule being the most efficient, since

ϕ_2 is almost doubled in the holo-form (Figure 5). On the contrary, the addition of the substrate or the inhibitor does not produce significant changes in the local dynamics of R264H, as shown by the similarity of ϕ_2 the values.

2.3 ANS binding

8-Anilino-1-naphthalenesulfonic acid (ANS) is a hydrophobic probe, which increases its fluorescence intensity upon binding to surface cavities and pockets of proteins [20]. For such reason, it is particularly suitable to study the conformational changes that modify the compactness of a protein tridimensional structure, such as those produced by temperature [21], pH [22], pressure [23] and substrate binding [24]. In order to characterize to what extent the presence of substrate (or inhibitor) in the active site of aromatase alter the roughness and accessibility of the protein surface, we added increasing amounts of ANS to the ligand-containing samples and compared the respective titration curves and fit in Figure 6. As shown also by the respective dissociation constants (reported in Table 2), the wt protein displays a lower affinity for ANS, its K_d value being 3 times larger than that obtained in the case of R264H and 2 times with respect to R264C (Table 2). Interestingly, in all samples the ANS binding process does not depend on the ligand used, giving the same results for the substrate and the inhibitor (Fig. 6 and Table 2).

2.4 Contact network analysis

A topological characterization of the aromatase structure has been attempted using the available crystallographic model of substrate-containing aromatase (PDB code 4kq8), and an *in silico* version of the ligand-free protein, obtained through a molecular dynamics simulation [25]. A contact network approach consists of a coarse-graining algorithm that reduces the amino acids of a protein into “nodes” of a tridimensional grid, identifying the position of each residue with that of its α -carbon. Each couple of amino acids distance is then evaluated and an (N x N) matrix built, consisting in “1” and “0” elements, corresponding to interacting or non-interacting elements of the network (according to a discriminating distance, fixed *a priori*). Eventually, “modules” are identified within the protein structure, thanks to a clustering procedure that groups nodes on the basis of the number of their mutual interactions. In the case of aromatase, such analysis allowed the identification of two distinct regions of the protein, as shown in Fig. 7 left panel, where each module of the crystal structure has been highlighted with a different color, red and green, respectively.

The clustering application does not result into an evident domain partition (such as regions with a preponderant kind of secondary structure or areas characterized by specific biological activity), nor the two clusters part in two distinct sections of the sequence. The spatial distribution of each one of

the two clusters is uniform and the border may be approximated by a plane perpendicular to the heme group. As shown in Figure 7 right panel, such an abstract plane is placed just below the channel used by the substrate to reach the catalytic site [26] indicating that the boundary between the two clusters is a critical region for the protein biological function. This feature can be more quantitatively described evaluating the so-called “participation coefficient”, P , a parameter that reflects the tendency of each node contained in one cluster to establish connections with nodes belonging to the other group [11]. Such connectivity index ranges from 0 to 1 (Fig. 8A) and the average protein value, $\langle P \rangle$, calculated on the whole sequence, results to be $\langle P \rangle \approx 0.076$ (Figure 8b). Much higher P values characterize, instead, the residues lying at the clusters border, including those that form the central section of helix G (dashed rectangle, Figure 8B). The participation coefficient has been also evaluated in the case of the substrate-bound aromatase model, P_{sb} , and the difference obtained between the holo- and apo-form, $\Delta P = P_{sb} - P$, is represented in Figure 8C.

While the majority of the protein residues is not affected by the ligand binding process, the topological analysis suggest that a significant increase in the participation coefficient occurs in the case of those residues that form the binding channel (Figure 8C). The central section of helix G displays, instead, an opposite effect characterized by the largest decrease in the P value, suggesting that the presence of the ligand reduces the connectivity role that that region has in the open, ligand-free protein form.

3. Discussion

Polymorphism of human proteins is a well-known phenomenon [27] that has recently raised a large interest in the scientific community, due to the possibility that cancer predictive studies (based on enzyme variants analyses) have in precision medicine [28]. The systematic study of proteins' variants databases through new algorithms [29-30] has considerably improved the accuracy of cancer survival predictions [28, 31-33], giving insights on drugs-induced damages [34] and allowing the identification of new, promising biomarkers [35]. In this general framework, the structural and functional characterization of aromatase polymorphisms is particularly important, as this enzyme is considered as a potential target for breast cancer therapy [36].

The crystallographic structures of placental [37] and recombinant [38] aromatase have demonstrated that the protein has a compact, globular fold and that R264 is localized on the protein surface, far away from the substrate binding site. Despite such a peripheral location, the conversion of ARG to HIS or CYS is known to influence the protein functional properties, reducing its ability to bind the substrate and decreasing its catalytic efficiency [6]. Circular dichroism spectroscopy showed that the two polymorphic variants retain the overall wt secondary structure, but are characterized by a lower

thermal stability suggesting that the mutation might change the motility of the helix G to which R264 belongs [6]. Structural studies and molecular dynamics simulations of bacterial P450 have revealed that such alpha helix is involved, through the adjacent helix F, in the opening of the channel which provides the access of substrate to the active site [26, 39, 40].

The contact network analysis reported in the present study has identified two main clusters (Figure 7), which are “communicating” through residues mainly dislocated along the channel that allows the entrance of the substrate in the active site. Such topological analysis thus suggests that large conformational changes characterize the substrate binding process of aromatase, involving long-range interactions between the two modules. Helices F and G are no exception: their central sections contain a relevant number of crucial nodes connecting the two clusters (Figure 8A and 8B) and undergo the largest (negative) change in their participation coefficient, P , as the protein binds the substrate molecule (Figure 8C). The kinetics of H/D exchange in Aro wt (Figure 2A and Table 1) indicates that these conformational changes produce a large decrease in the fraction of the more exposed protons in the presence of both androstenedione and anastrozole. Trivial explanations (such as a screening effect exerted by the ligands) seem to be excluded by fluorescence anisotropy measurements, that demonstrate how in the ligand-bound form the protein assumes a more compact tertiary structure through a conformational change that confers the molecule a faster rotational dynamics (Figure 5). The results obtained for the two polymorphic variants demonstrate that these samples undergo much smaller changes in the H/D kinetics (Table 1) and in the long rotational correlation time (Figure 5). More importantly, in the presence of both ligands, they retain a higher ANS binding capability, with respect to that of Aro wt (Figure and Table 2), confirming that they lack that compactness which characterizes ligand-bound Aro wt. As a consequence, the correct entrance of the substrate (or inhibitor) molecule to the protein active site is compromised, as demonstrated by dynamic fluorescence measurements. Indeed, in a previous study [8], we have demonstrated that the lifetime component at 4.5-5.0 ns can be attributed to W224, since it is not present in mutants such as W224F [8]. Since this tryptophan resides in the proximity of aromatase binding site and it is dramatically quenched by the entrance of androstenedione or anastrozole (Figure 3A), it can be used to monitor ligand-induced structural changes in the specific region of the active site. The absence of any significant difference of this specific fluorescence lifetime component in the two polymorphic variants (Figure 3B,C) demonstrates that the local perturbation introduced at the extremity of helix G upon the substitution of R264 impairs the correct placement of the ligand in the active site, thus providing a structural rationale for the decreased activity of the mutants [6].

The normal mode analysis of aromatase [25] has demonstrated that the protein displays a very rigid core (corresponding approximately to the heme group), which is surrounded by concentric shells of

progressively more flexible regions. The helices F and G and their inter-connecting loop are one of such external mobile structural areas [38]. Indeed, long-range effects connected to the flexibility of helix G has been already reported in the case of bacterial cytochromes P450 [41]. In particular, it was demonstrated that the breaking of the salt bridge which involves ASP 251 has crucial consequences on ligands recognition and binding.

According to the contact network analysis reported in this study, the protein could be figured as a sort of sandwich, the central slice corresponding to the substrate channel (Figure 7 and 8). In this picture, the packing of the sandwich would be controlled by the clamping effected exerted by helices F and G, whose hinge mechanism resides in their highly inter-connected central elements (Figure 8). The substitution of R264 with residues characterized by a different geometry (HIS and CYS), length and charge (CYS) must severely affect the salt bridge interaction with ASP 251, altering the “pivot” mechanism by which the flexible F-G group controls the interconnection between the two clusters.

3. Conclusion

In conclusion, the data reported in this study demonstrate by two independent spectroscopic techniques that human aromatase assumes a more compact tridimensional conformation upon ligand binding. On the contrary, mutants R264C and R264H (that impair the correct binding of both substrate and inhibitor molecules) undergo a less efficient packing process, suggesting a minor flexibility of their mobile segments. A rationale to such behavior could be envisaged thanks to a contact network analysis of the protein structure, which has revealed the connecting role of helices F and G between the two clusters that form the substrate channel to the active site. Thus, based on *in vitro* measurements, such an explanation shed new light on molecular impact that the two polymorphisms have in living beings.

4. Methods

4.1 Materials

Recombinant human aromatase (Aro) was expressed in *E. coli* and purified as previously described [7, 8]. Both wild-type and mutants were purified with the same type of buffer, i.e. 100 mM phosphate buffer pH 7.4 containing 20% glycerol, 1 mM β -mercaptoethanol, 0.1% Tween-20

All chemicals used for protein purification were purchased from Sigma–Aldrich (St. Louis, MO USA) and were analytical grade.

Proteins (wt and mutants) concentration for all fluorescence experiments was about 6.5 μM , while if present, androstenedione and anastrozole concentrations were 20 μM and 10 μM , respectively.

4.2 H/D exchange kinetics experiments by ATR-FTIR

Kinetics of H/D exchange was followed by FTIR as reported by Di Nardo *et al.* [8]. Experiments were performed at room temperature using an infrared spectrophotometer Bruker Model Tensor 27 (Bruker Instruments, USA) coupled with an attenuated total reflectance (ATR) sampling tool (Harrick Scientific Products, USA). The working parameters were set as follows: scan velocity 10 kHz; resolution 4 cm^{-1} ; spectra acquisition frequency limits 4000 and 800 cm^{-1} . Protein sample was analysed by depositing a thin protein film (30 μL , 50 μM) directly on ATR germanium crystal. In particular, Aro was analysed in ligand free form and in ligand bound form, obtained by incubating the enzyme with androstenedione as substrate or anastrozole as inhibitor. The correct binding of both substrate and inhibitor to the active site of Aro was examined by UV/vis spectroscopy by monitoring a shift of the maximum absorbance from 418 nm to 394 nm for Androstenedione and from 418 nm to 422 nm for anastrozole. During acquisition, the spectrophotometer sample chamber was continuously purged with D_2O enriched nitrogen. Spectra were collected at intervals of 1 minute, for the first 10 minutes, and every 8 minutes for the following 160 minutes. For each time point 60 scans were collected and averaged. Spectra were collected at least in quadruplicate and averaged for each time point, corrected by the contribution of control sample, represented by protein storage buffer, and normalized. The H/D exchange kinetics was studied by following the absorbance increase at about 1460 cm^{-1} , corresponding to the shift of the Amide II band upon deuteration. The relative absorbance values were plotted as function of time, fitted to a double exponential function and the deuteration rates were calculated and compared by ANOVA statistical analysis using SigmaPlot software.

4.3 Fluorescence measurements

Fluorescence spectra were collected using a K2-ISS (ISS, Inc., Champaign, IL, USA) photon-counting fluorimeter thermo-stated at 20 $^{\circ}\text{C}$ using an external bath circulator. ANS binding was

studied measuring the fluorescence emission spectra from 450 to 550 nm of the probe using an excitation wavelength of 350 nm. All spectra were corrected by blank subtraction

8-Anilino-1-naphthalenesulfonic acid (ANS) binding curves were analyzed assuming a simple binding equilibrium, namely, $\text{ANS} + \text{P} \leftrightarrow \text{ANS-P}$, and fitting the total fluorescence, F , at increasing concentration of ANS, $[\text{ANS}]$, according to:

$$F = F_{\infty} \left(([\text{ANS}] + P_0 + K_d) - \sqrt{([\text{ANS}] + P_0 + K_d)^2 - 4[\text{ANS}]P_0} \right) / (2 P_0)$$

where P_0 is the total aromatase concentration and K_d the dissociation constant of the process.

Lifetimes and dynamic fluorescence anisotropies were performed using the phase shift and demodulation technique on a KOALA-ISS fluorimeter. The excitation source was a 300-nm laser diode, and the emission was collected through a WG320-nm cutoff filter to avoid scattering. Anisotropy decays were collected through Glan-Thompson polarizers taking into account the G-factor correction. All measurements were repeated in triplicate to obtain a good statistic, using a set of at least 30 frequencies. The data have been analyzed with the software provided by ISS.

4.4 Contact network analysis

Protein contact networks (PCNs) rely on a minimalist perspective on protein structures, seen as networks of active contacts between residues (network nodes are the protein residues, and the active contacts between pair of residues are the network links). The definition of active contacts is crucial in defining the properties emerging from the network analysis. In this work, we define “active contact” any contact corresponding to a distance between residues comprised between 4 and 8 Å. This choice includes only significant noncovalent bonds, sensible to the environmental cues. The distance between residues is computed starting from the coordinates of the residues’ α -carbons. The mathematical description of PCNs is provided by the adjacency matrix, defined as:

$$A_{ij} = \begin{cases} 1 & \text{if } 4 \text{ \AA} < d_{ij} < 8 \text{ \AA} \\ 0 & \text{otherwise} \end{cases}$$

Once the network has been built up, it is possible to derive several network descriptor. The most important and simple is the node degree k_i , defined as:

$$k_i = \sum_j A_{ij}$$

Based on the node degree, we applied a spectral clustering algorithm to partition PCN into clusters. We applied the method to the network Laplacian, defined as follows:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}$$

\mathbf{A} being the adjacency matrix and \mathbf{D} the degree matrix, i.e. a diagonal matrix whose diagonal is the degree vector. The eigenvalue decomposition is applied to the laplacian \mathbf{L} : the eigenvector

corresponding to the second minor of eigenvalue v_2 is of interest for the clustering partition. Considering the partition in two clusters, for instance, nodes are divided into the two clusters according to the sign of the corresponding components of the vector v_2 .

The partition is binary and hierarchical; the number of cluster (power of 2) is an input value.

In this work, we partitioned the structures into two clusters, so we applied just once the spectral clustering algorithm to the PCNs.

Once the network clustering has been applied, it is possible to derive a clustering descriptor, the participation coefficient P_i , for the i -th residue (node) defined as:

$$P_i = 1 - \left(\frac{k_{si}}{k_i} \right)^2$$

k_{si} is the i node degree including only links with nodes belonging to the same cluster s and k_i the node i degree. Thus, P addresses the role of nodes in signal transmission between different clusters (domains in PCNs).

List of Abbreviations

SNPs: single nucleotide polymorphisms
Aro: Recombinant human aromatase
FTIR: Fourier transform infrared spectroscopy
ATR: attenuated total reflectance
ANS: 8-Anilino-1-naphthalenesulfonic acid
PCNs: Protein contact networks

Declarations section

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of supporting data

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Authors' contributions

G.D.N. and A.D.V.: designed the experiments, interpreted the data and contributed to manuscript writing; C.Z.: contributed to engineering the mutants, purified and characterized the wild type and the mutants; E.N.: designed, performed and interpreted fluorescence experiments; C.S.: designed, performed and interpreted the FT/IR experiments; L.D.P.: designed, performed and interpreted Contact network analysis and contributed to manuscript writing; G.G and G.M.: designed the research project, designed the experiments, interpreted the data and contributed to manuscript writing.

Acknowledgements

Not applicable

Authors' information

Not applicable

References

1. Cojocaru V, Peter JW, Wade RC. The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta* 2007; 1770: 90-401.
2. Poulos TL. Cytochrome P450 flexibility. *PNAS* 2003; 100:13121-13122.
3. Pochapsky TC, Kazanis S, Dang M. Conformational plasticity and structure/function relationships in cytochromes P450. *Antioxid Redox Signal.* 2010; 13:1273-1296.
4. Di Nardo G, Gilardi G. Human aromatase: perspectives in biochemistry and biotechnology. *Biotechnology and Applied Biochemistry* 2013; 60:92-101.
5. Patel S. Disruption of aromatase homeostasis as the cause of a multiplicity of ailments: A comprehensive review. *J Steroid Biochem Mol Biol.* 2017; 168:19-25.
6. Baravalle R, Di Nardo G, Bandino A, Barone I, Catalano S, Andò S, Gilardi G. Impact of R264C and R264H polymorphisms in human aromatase function. *J. Steroid Biochem. Mol. Biol.* 2017; 167: 23-32.

7. Parween S, Di Nardo G, Baj F, Zhang C, Gilardi G, Pandey AV. Differential effects of variations in human P450 oxidoreductase on the aromatase activity of CYP19A1 polymorphisms R264C and R264H. *J Steroid Biochem Mol Biol.* 2020; 196:105507.
8. Di Nardo G, Breitner M, Sadeghi SJ, Castrignanò S, Mei G, Di Venere A, Nicolai E, Allegra P, Gilardi G. Dynamics and flexibility of human aromatase probed by FTIR and time resolved fluorescence spectroscopy. *Plos One* 2013; 8:e82118.
9. Di Paola L, Mei G, Di Venere A, Giuliani A. Exploring the stability of dimers through protein structure topology. *Curr. Protein Pept. Sci.* 2016; 17:30-36.
10. Minicozzi V, Di Venere A, Nicolai E, Giuliani A, Caccuri AM, Di Paola L, Mei G. Non-symmetrical structural behavior of a symmetric protein: the case of homo-trimeric TRAF2 (tumor necrosis factor-receptor associated factor 2). *J. Biomol. Struct. Dyn.* 2020; 13:1-11.
11. Di Paola L, Giuliani A. Protein contact network topology: a natural language for allostery. *Curr. Opin. Struct. Biol.* 2015; 31:43-48.
12. Di Paola L, Mei G, Di Venere A, Giuliani A. Disclosing allostery through protein contact networks. *Methods. Mol. Biol.* 2021; 2253:7-20.
13. Platania CB, Di Paola L, Leggio GM, Romano GL, Drago F, Salomone S, Bucolo C. Molecular features of interaction between VEGFA and anti-angiogenic drugs used in retinal diseases: A computational approach. *Front. Pharmacol.* 2015; 6:248.
14. Di Paola L, Hadi-Alijanvand H, Song X, Hu G, Giuliani A. The discovery of a putative allosteric site in SARS-CoV-2 spike protein by an integrated structural/dynamic approach. *J Proteosome Res.* 2020; 19:4576-4586.
15. Yu S, Fan F, Flores FC, Mei F, Cheng X. Dissecting the mechanism of Epac activation via hydrogen-deuterium exchange FT-IR and structural modelling. *Biochemistry* 2006; 45:15318–15326.
16. Kim KS, Fuchs JA, Woodward CK. Hydrogen exchange identifies native-state motional domains important in protein folding. *Biochemistry* 1993; 32:9600–9608.
17. de Jongh HH, Goormaghtigh E, Ruyschaert JM. Tertiary stability of native and methionine-80 modified cytochrome c detected by proton-deuterium exchange using on-line Fourier transform infrared spectroscopy. *Biochemistry* 1995; 34:172–179.
18. Li J, Cheng X, Lee JC. Structure and dynamics of the modular halves of Escherichia coli cyclic AMP receptor protein. *Biochemistry* 2002; 41:14771–14778.
19. Cantor CH, Schimmel PR, *Biophysical Chemistry Vol. II.* W.H. Freeman and Company, San Francisco (1980).

20. Lakowicz JR, Principles of fluorescence spectroscopy. Kluwer Academic/Plenum Publishers, New York (1999)
21. Kim KH, Yun S, Lee EK. Thermodynamic analysis of ANS binding to partially unfolded α -lactalbumin: correlation of endothermic to exothermic changeover with formation of authentic molten globules. *J. Mol. Recognit.* 2016; 29:446-451.
22. Varriale A, Marabotti A, Mei G, Staiano M, D'Auria S. Correlation spectroscopy and molecular dynamics simulations to study the structural features of proteins, *Plos One* 2013; 8:e64840.
23. Ceccarelli A, Di Venere A, Nicolai E, De Luca A, Minicozzi V, Rosato N, Caccuri AM, Mei G. TNFR-associated factor-2 (TRAF2): not only a trimer. *Biochemistry* 2015; 54:6153-6161.
24. Benítez-Rangel E, Rodríguez-Hernández A, Velasco-García R. The substrate of the glucose-6-phosphate dehydrogenase of *Pseudomonas aeruginosa* provides structural stability. *Biochim Biophys Acta Proteins Proteom.* 2020; 1868:140331.
25. Di Nardo G, Cimicata G, Baravalle R, Dell'Angelo V, Ciaramella A, Catucci G, Ugliengo P, Gilardi G. Working at the membrane interface: Ligand-induced changes in dynamic conformation and oligomeric structure in human aromatase. *Biotechnol. Appl. Biochem.* 2018; 65:46-53.
26. Sgrignani J, Magistrato A. Influence of the Membrane Lipophilic Environment on the Structure and on the Substrate Access/Egress Routes of the Human Aromatase Enzyme. A Computational Study. *J. Chem. Inf. Model.* 2012; 52:1595-1606.
27. Lewis, W. Polymorphism of Human Enzyme Proteins. *Nature* 1971; 230: 215–218.
28. Amelio I, Bertolo R, Bove P, Candi E, Chiocchi M, Cipriani C, Di Daniele N, Ganini C, Juhl H, Mauriello A, Marani C, Marshall J, Montanaro M, Palmieri G, Piacentini M, Sica G, Tesaro M, Rovella V, Tisone G, Shi Y, Wang Y, Melino G. Cancer predictive studies. *Biol Direct.* 2020; 15:18.
29. Mihaylov I, Kańduła M, Krachunov M, Vassilev D. A novel framework for horizontal and vertical data integration in cancer studies with application to survival time prediction models. *Biol Direct.* 2019; 14:22.
30. Werner J, Géron A, Kerssemakers J, Matallana-Surget S. mPies: a novel metaproteomics tool for the creation of relevant protein databases and automatized protein annotation. *Biol Direct.* 2019; 14:21.
31. Han Y, Ye X, Wang C, Liu Y, Zhang S, Feng W, Huang K, Zhang J. Integration of molecular features with clinical information for predicting outcomes for neuroblastoma patients. *Biol Direct.* 2019; 14:16.
32. Han Y, Ye X, Cheng J, Zhang S, Feng W, Han Z, Zhang J, Huang K. Integrative analysis based on survival associated co-expression gene modules for predicting Neuroblastoma patients' survival time. *Biol Direct.* 2019; 14:4.
33. Kim SY, Jeong HH, Kim J, Moon JH, Sohn KA. Robust pathway-based multi-omics data integration using directed random walks for survival prediction in multiple cancer studies. *Biol Direct.* 2019; 14:8.

34. Chierici M, Francescato M, Bussola N, Jurman G, Furlanello C. Predictability of drug-induced liver injury by machine learning. *Biol Direct.* 2020; 15:3.
35. Liu L, Wang G, Wang L, Yu C, Li M, Song S, Hao L, Ma L, Zhang Z. Computational identification and characterization of glioma candidate biomarkers through multi-omics integrative profiling. *Biol Direct.* 2020; 15:10.
36. Adhikari N, Amin SA, Saha A, Jha T. Combating breast cancer with non-steroidal aromatase inhibitors (NSAIs): Understanding the chemico-biological interactions through comparative SAR/QSAR study. *Eur J Med Chem.* 2017; 137:365-438.
37. Ghosh D, Griswold J, Erman M, Pangborn W. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. *Nature* 2009; 457:219-223.
38. Lo J, Di Nardo G, Griswold J, Egbuta C, Jiang W, Gilardi G, Ghosh D. Structural Basis for the Functional Roles of Critical Residues in Human Cytochrome P450 Aromatase. *Biochemistry* 2013; 52:5821-5829.
39. Jiang W, Ghosh D. Motion and flexibility in human cytochrome p450 aromatase. *Plos One* 2013; 7:e32565.
40. Park J, Czapl L, Amaro RE. Molecular simulations of aromatase reveal new insights into the mechanism of ligand binding. *J. Chem. Inf. Model.* 2013; 53:2047-2056.
41. Di Nardo G, Dell'Angelo V, Catucci G, Sadeghi SJ, Gilardi G. Subtle structural changes in the Asp251Gly/Gln307His P450 BM3 mutant responsible for new activity toward diclofenac, tolbutamide and ibuprofen. *Arch. Biochem. Biophys.* 2016; 602:106-115.

Figure legend

Figure 1: FTIR spectra of apo-rArom at the beginning (blue) and at the end (red) of a deuteration kinetics. In the inset (right side) the increase of the absorption intensity upon deuteration, at $\lambda = 1460 \text{ cm}^{-1}$ is shown (red arrow) after normalization of the spectra at 1650 cm^{-1} .

Figure 2: Kinetics of H/D exchange followed by FTIR for wt-rArom (panel A, blue symbols) and the two polymorphic variants (namely R264H, panel B, and R264C, panel C, blue symbols). The kinetics in the presence of the substrate androstenedione (green symbols) and the inhibitor anastrozole (red symbols) are also reported. The solid lines correspond to the best double exponential fitting curves, whose parameters are reported in Table 1.

Figure 3: lifetime components of wt-rArom fluorescence decay (panel A), R264H (panel B) and R264C (panel C) in the absence (dashed bars) and in the presence of substrate (green) or inhibitor (red).

Figure 4: phase shift (black symbols) and demodulation data (purple symbols) of wt-rArom (circles), R264H (squares) and R264C (triangles). The solid lines correspond to the best fit obtained using two rotational correlation times.

Figure 5: long (ϕ_1) and short (ϕ_2) fluorescence correlation lifetimes of apo-wt and apo-mutants (blue) aromatase. The values of ϕ_1 and ϕ_2 upon the addition of substrate or inhibitor are reported in green and red, respectively. In the cartoons, a ribbon model of aromatase is reported in which the positions of R264 and of the five tryptophans are highlighted in green and pink VDW spheres.

Figure 6 and Table II:ANS-binding curves of wt-rArom (circles) and mutants R264H (squares) or R264C (triangles) in the presence of saturating concentration of substrate (green) or inhibitor (red). The concentration of all the proteins is 6.4 μ M. The solid lines correspond to the best fit obtained for each sample by a chi-squared minimization procedure (see materials and methods).

Figure 7: in the left side of the figure the two clusters (red and green) identified on the basis of the contact network analysis applied to the 4kq8 pdb file are reported (the heme group is also shown in blue). The G-helix has been highlighted by a red dot cloud. On the right side, the same structure is reported together with residue R264 (blue, VDW) and those aminoacids (purple, VDW) which form the access channel for the substrate androstenedione (according to Sgrignani and Magistrato, 2012) and the heme group (in red). The black line in both pictures marks the frontier of the two clusters.

Figure 8: in panel A the participation coefficient, P , of each amino acid of wt-rArom is shown, according to the colors code whose numerical value is reported in the left scale. In panel B the distribution of the P values is presented as a function of the sequence position (X-axis) and of the belonging to one of the two clusters (red or green) identified in Figure 7. $\langle P \rangle$ is the protein average participation coefficient. The area of the dashed rectangle corresponds to the G-helix residues. In panel C the difference $\Delta P = P_{sb} - P$ between the P values of the holo and apo forms is reported. The green color has been attributed to those residues whose P value does not change ($\Delta P=0$) upon ligand binding.

Figures

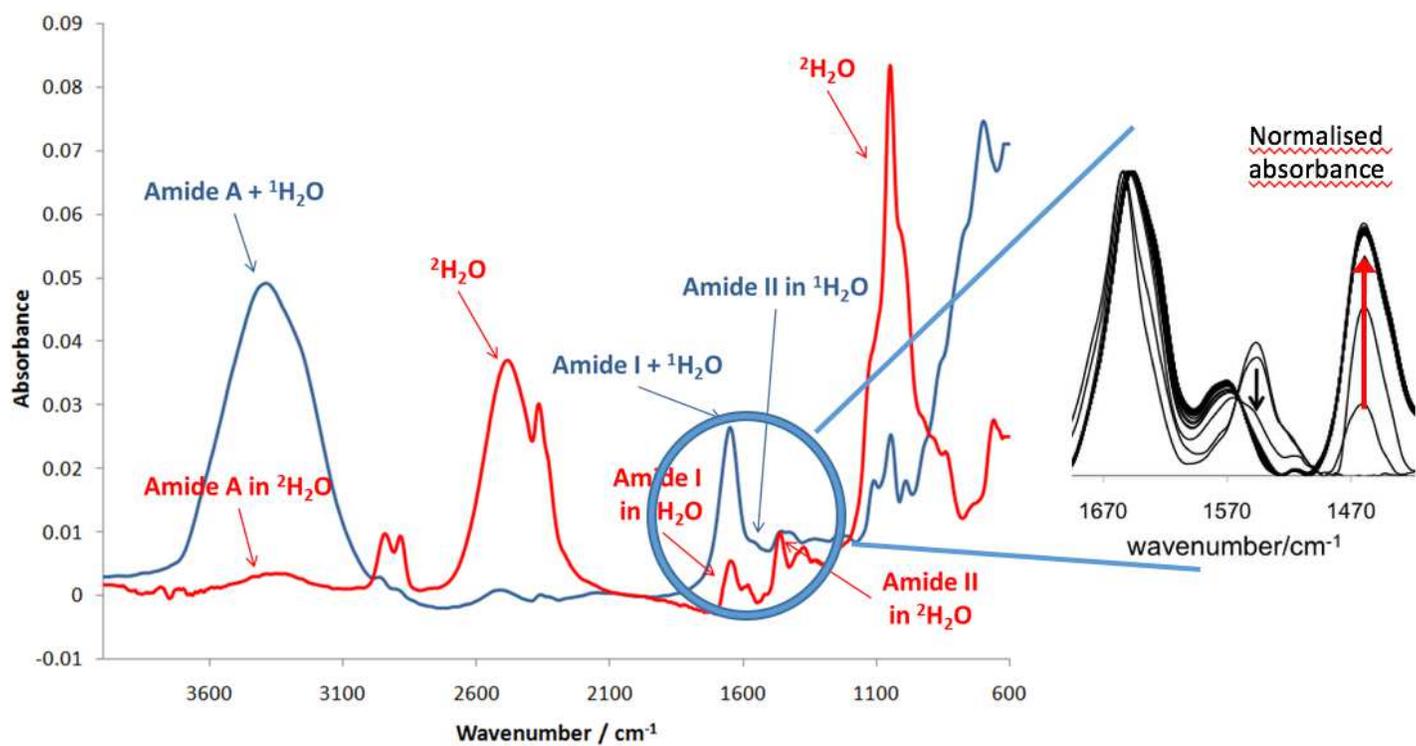


Figure 1

FTIR spectra of apo-rArom at the beginning (blue) and at the end (red) of a deuteriation kinetics. In the inset (right side) the increase of the absorption intensity upon deuteriation, at $\lambda = 1460 \text{ cm}^{-1}$ is shown (red arrow) after normalization of the spectra at 1650 cm^{-1} .

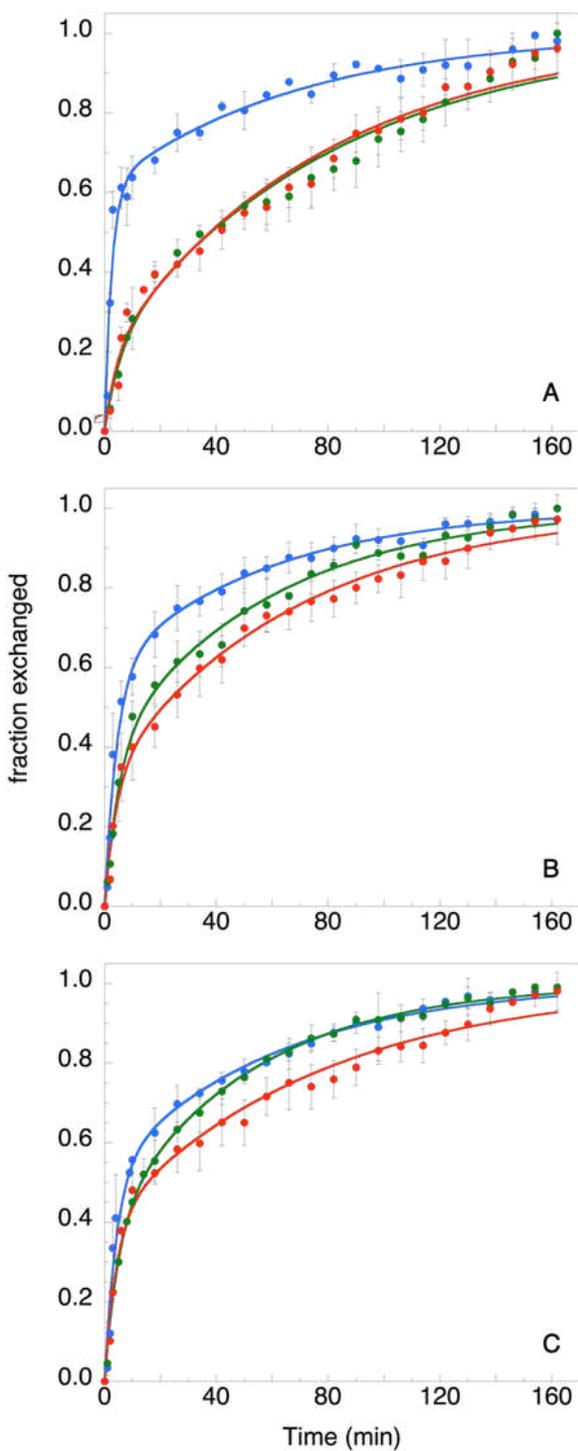


Figure 2

Kinetics of H/D exchange followed by FTIR for wt-rArom (panel A, blue symbols) and the two polymorphic variants (namely R264H, panel B, and R264C, panel C, blue symbols). The kinetics in the presence of the substrate androstenedione (green symbols) and the inhibitor anastrozole (red symbols) are also reported. The solid lines correspond to the best double exponential fitting curves, whose parameters are reported in Table 1.

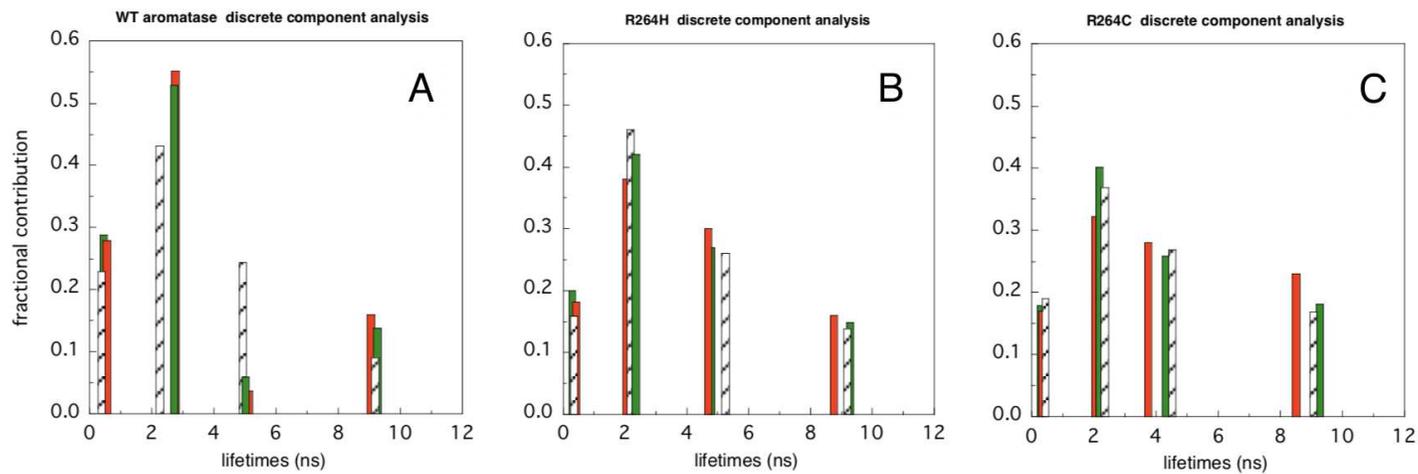


Figure 3

lifetime components of wt-rArom fluorescence decay (panel A), R264H (panel B) and R264C (panel C) in the absence (dashed bars) and in the presence of substrate (green) or inhibitor (red).

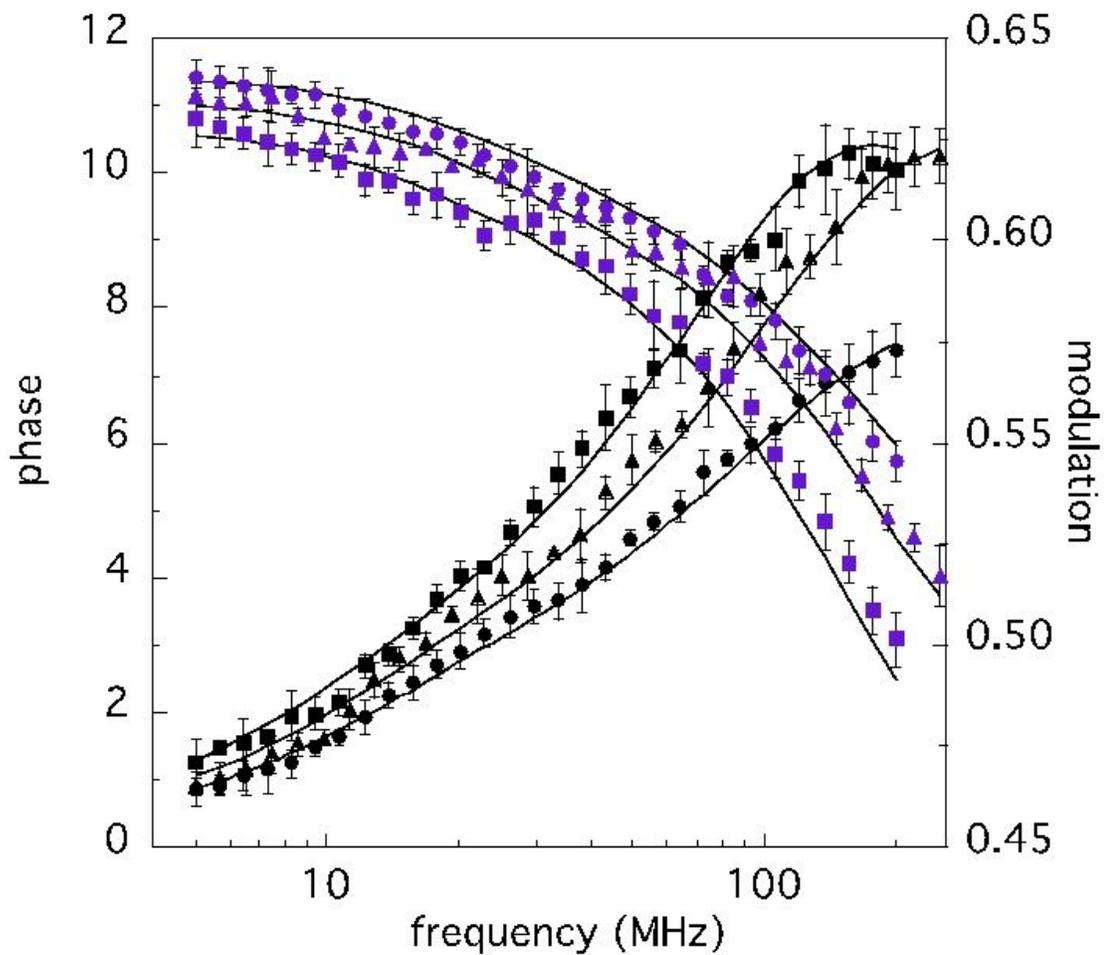


Figure 4

phase shift (black symbols) and demodulation data (purple symbols) of wt-rArom (circles), R264H (squares) and R264C (triangles). The solid lines correspond to the best fit obtained using two rotational correlation times.

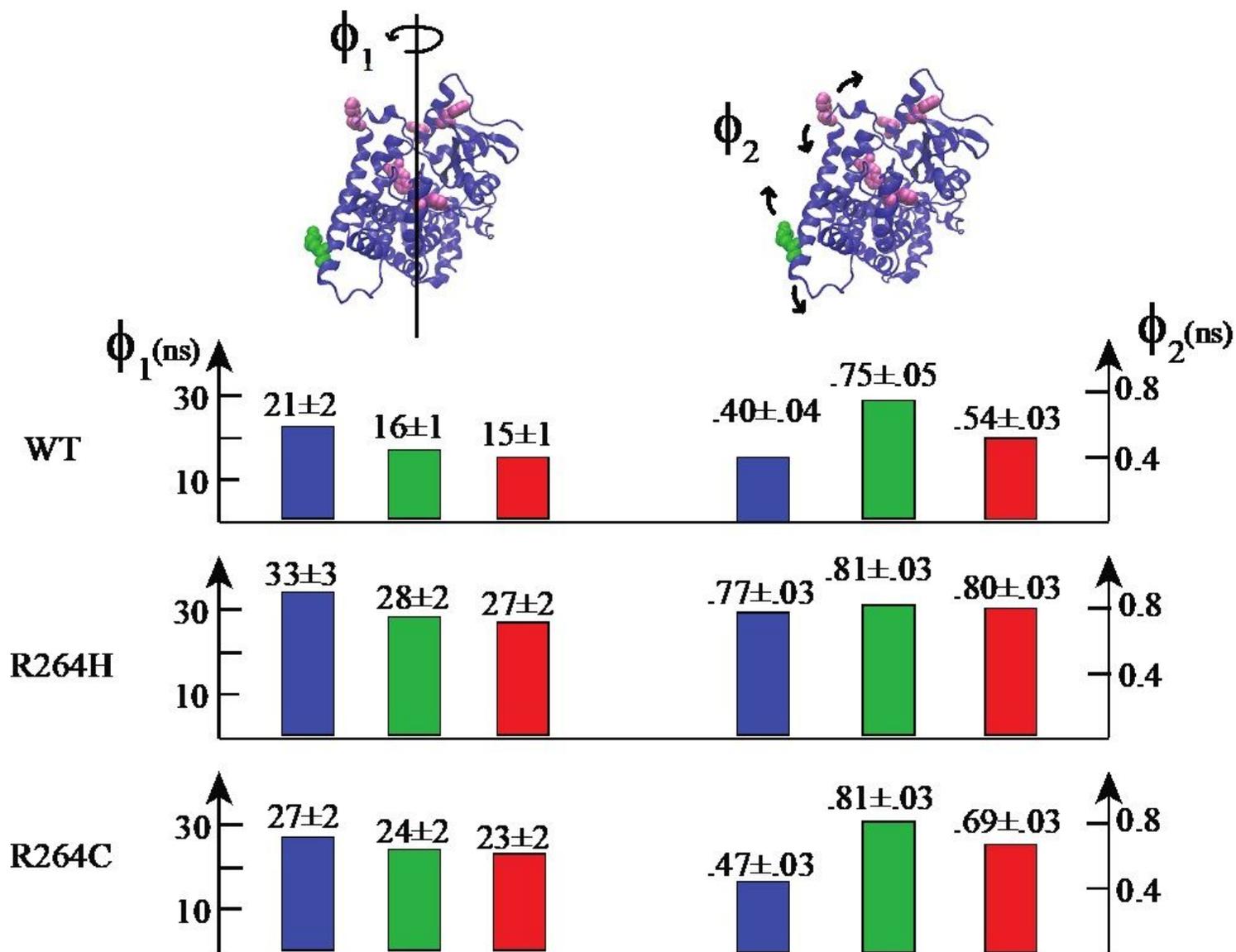


Figure 5

long (ϕ_1) and short (ϕ_2) fluorescence correlation lifetimes of apo-wt and apo-mutants (blue) aromatase. The values of ϕ_1 and ϕ_2 upon the addition of substrate or inhibitor are reported in green and red, respectively. In the cartoons, a ribbon model of aromatase is reported in which the positions of R264 and of the five tryptophans are highlighted in green and pink VDW spheres.

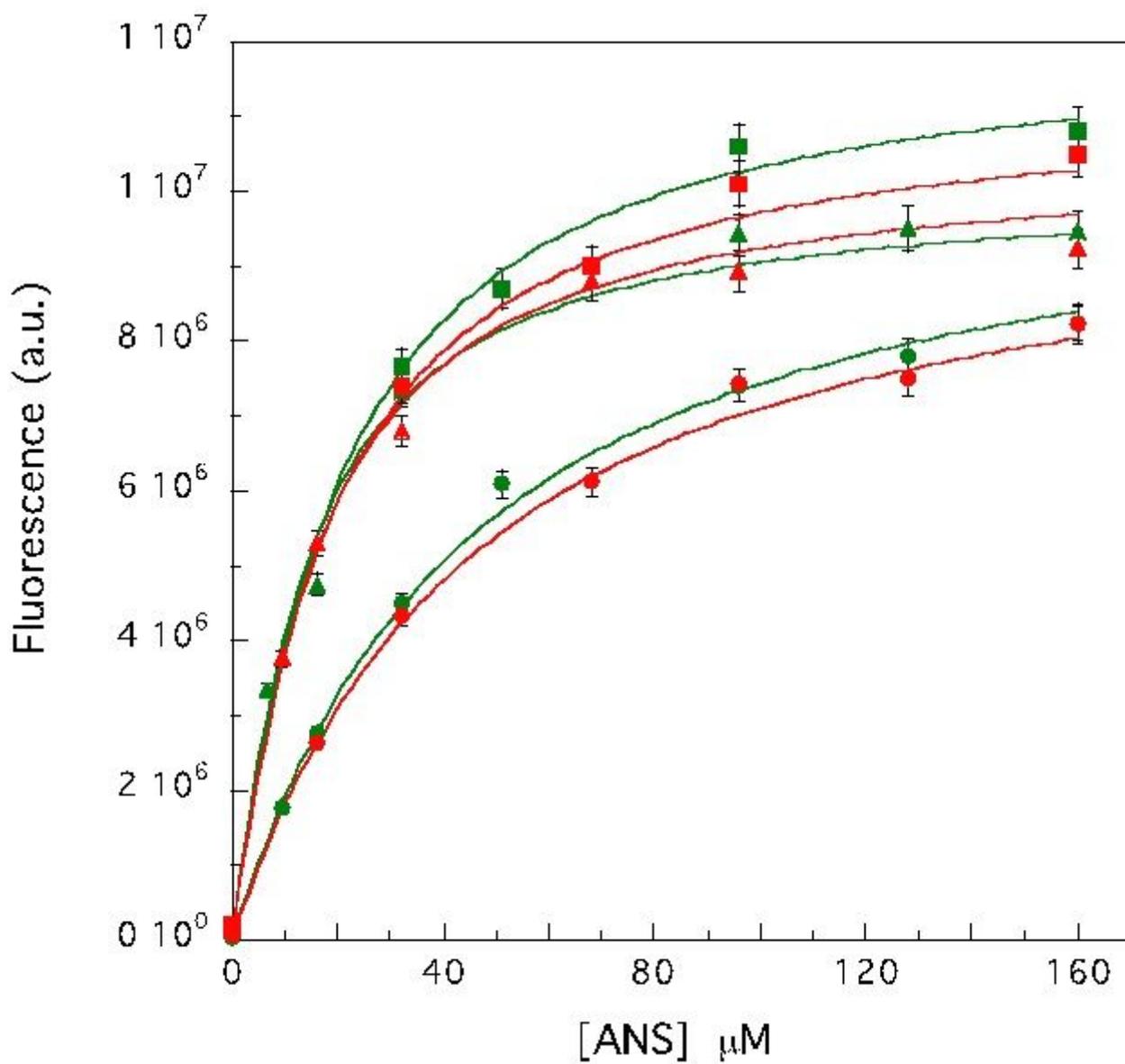


Figure 6

and Table II: ANS-binding curves of wt-rArom (circles) and mutants R264H (squares) or R264C (triangles) in the presence of saturating concentration of substrate (green) or inhibitor (red). The concentration of all the proteins is 6.4 μM. The solid lines correspond to the best fit obtained for each sample by a chi-squared minimization procedure (see materials and methods).

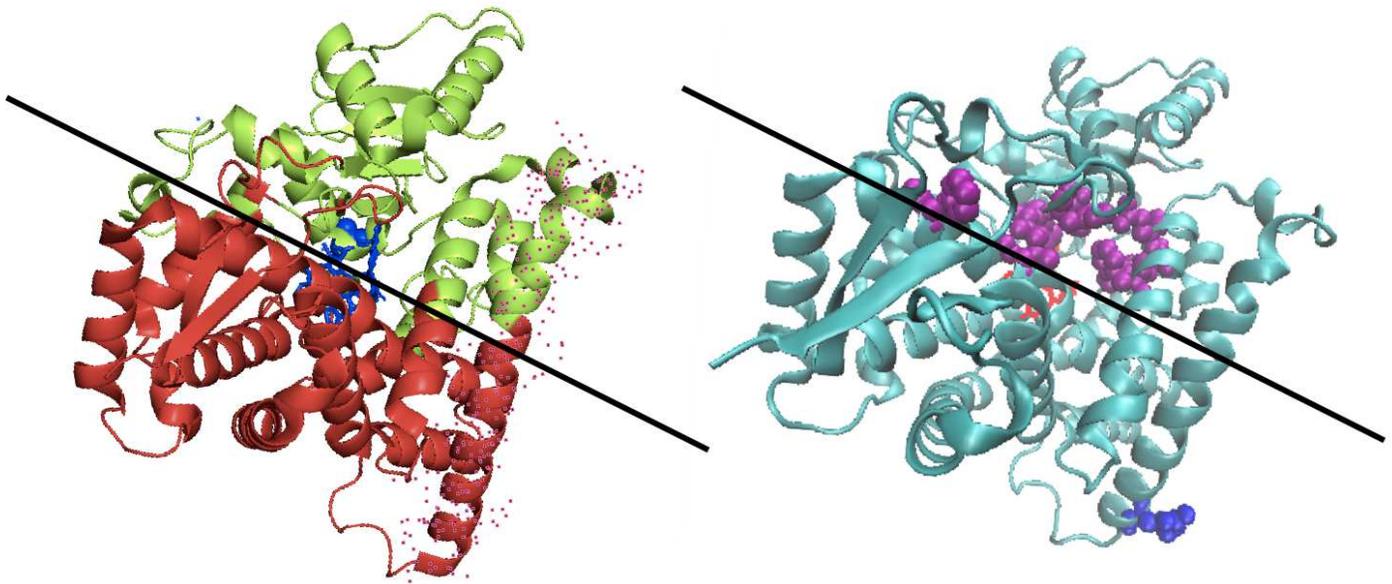


Figure 7

in the left side of the figure the two clusters (red and green) identified on the basis of the contact network analysis applied to the 4kq8 pdb file are reported (the heme group is also shown in blue). The G-helix has been highlighted by a red dot cloud. On the right side, the same structure is reported together with residue R264 (blue, VDW) and those aminoacids (purple, VDW) which form the access channel for the substrate androstenedione (according to Sgrignani and Magistrato, 2012) and the heme group (in red). The black line in both pictures marks the frontier of the two clusters.

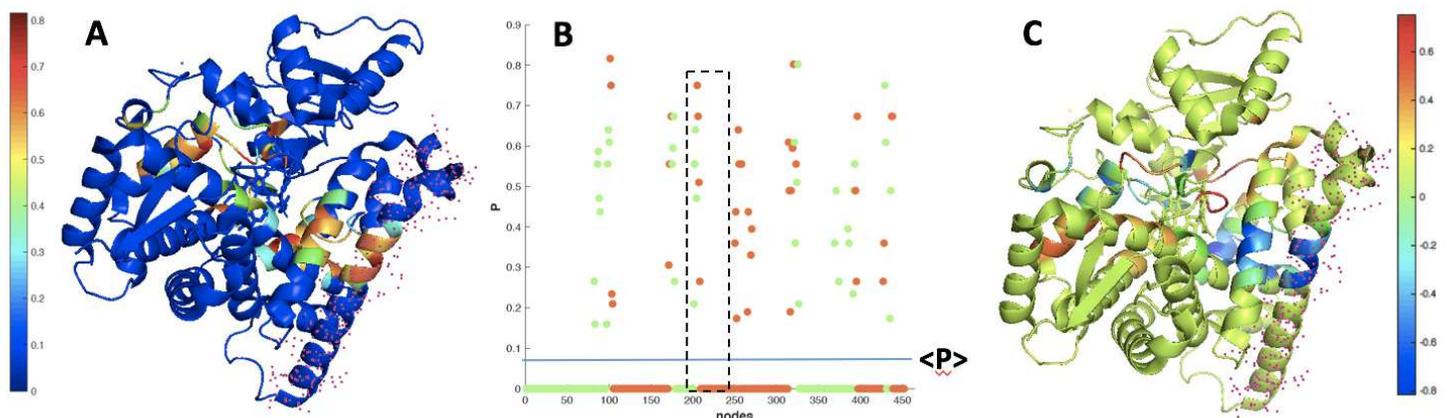


Figure 8

in panel A the participation coefficient, P , of each amino acid of wt-rArom is shown, according to the colors code whose numerical value is reported in the left scale. In panel B the distribution of the P values is presented as a function of the sequence position (X-axis) and of the belonging to one of the two clusters (red or green) identified in Figure 7. $\langle P \rangle$ is the protein average participation coefficient. The area

of the dashed rectangle corresponds to the G-helix residues. In panel C the difference $\Delta P = P_{sb} - P$ between the P values of the holo and apo forms is reported. The green color has been attributed to those residues whose P value does not change ($\Delta P=0$) upon ligand binding.