

Research Article

Sequence comparisons of cytochrome P450 aromatases from Australian animals predict differences in enzymatic activity and/or efficiency[†]

Anam Fatima^{1,‡}, Jessica K. Holien^{2,‡}, Chandni Tiwari^{1,‡},
Michael W. Parker^{2,3}, Raymond J. Rodgers^{4,*} and Lisandra L. Martin^{1,*}

¹School of Chemistry, Monash University, Clayton, Victoria, Australia, ²ACRF Rational Drug Discovery Centre, St Vincent's Institute of Medical Research, Fitzroy, Victoria, Australia, ³Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Victoria, Australia and ⁴Discipline of Obstetrics and Gynaecology, School of Medicine, Robinson Research Institute, The University of Adelaide, SA, Australia.

***Correspondence:** School of Chemistry, Monash University, Clayton, Victoria, Australia. Tel: +61399054514; E-mail: Lisa.Martin@monash.edu or Discipline of Obstetrics and Gynaecology, School of Medicine, Robinson Research Institute, The University of Adelaide, SA, Australia. Tel: +61883133932; E-mail: Ray.Rodgers@adelaide.edu.au.

†Grant Support: This research was supported by the National Health and Medical Research Council of Australia (NHMRC) and the Australian Research Council. This work was also supported by a capital grant from the Australian Cancer Research Foundation and infrastructure funding from the Victorian Government (Australia) Operational Infrastructure Support Scheme to St. Vincent's Institute of Medical Research. M.W.P. is a NHMRC Senior Principal Research Fellow, R.J.R. holds a Lloyd Cox Fellowship (University of Adelaide) and J.K.H. is a 5point Foundation Fellow.

‡Joint contribution as first authors.

Received 6 August 2019; Revised 1 February 2020; Accepted 25 February 2020

Abstract

Aromatase (P450arom, CYP19A1) is the terminal enzyme in the synthesis of the steroid hormone family of estrogens. Not surprisingly, this enzyme has structural similarities between the limited number of species studied thus far. This study examined the structure of aromatases from four diverse Australian species including a marsupial (tammar wallaby; *Macropus eugenii*), monotreme (platypus; *Ornithorhynchus anatinus*), ratite (emu; *Dromaius novaehollandiae*) and lizard (bearded dragon; *Pogona vitticeps*). We successfully built homology models for each species, using the only crystallographically determined structure available, human aromatase. The amino acid sequences showed high amino acid sequence identity to the human aromatase: wallaby 81%, platypus 73%, emu 75% and bearded dragon at 74%. The overall structure was highly conserved among the five species, although there were non-secondary structures (loops and bends) that were variable and flexible that may result in some differences in catalytic activity. At the N-terminal regions, there were deletions and variations that suggest that functional distinctions may be found. We found that the active sites of all these proteins were identical, except for a slight variation in the emu. The electrostatic potential across the surfaces of these aromatases highlighted likely variations to the protein-protein interactions of these enzymes with both redox partner cytochrome P450 reductase and possibly homodimerization in the case of the platypus, which has been postulated for the human aromatase enzyme. Given the high natural selection pressures on reproductive strategies,

© The Author(s) 2020. Published by Oxford University Press on behalf of Society for the Study of Reproduction.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

the relatively high degree of conservation of aromatase sequence and structure across species suggests that there is biochemically very little scope for changes to have evolved without the loss of enzyme activity.

Key words: evolution of steroidogenesis, aromatase, molecular modeling, estrogen.

Introduction

The synthesis of the steroid hormone family of estrogens is a defining step in the reproductive ability of animal species. The biosynthesis of steroid hormones begins with cholesterol, which is converted into a variety of steroid hormones by a series of enzymatically catalyzed chemical reactions. A key step in estrogen production is a final transformation of the precursors, androstenedione or testosterone, into estrogens by a three-step, six-electron process. It is catalyzed by a cytochrome P450 enzyme called aromatase (P450arom, CYP19A1) [1, 2].

Cytochrome P450 enzymes are a superfamily of heme-thiolate enzymes that achieve stereoselective insertion of an oxygen atom into an otherwise inactive carbon-hydrogen bond [3–5]. Two electrons from the reducing nicotinamide adenine dinucleotide phosphate (NADPH) are required for the catalytic cycle. Electrons are transferred to the microsomal aromatase from a membrane bound flavoprotein, NADPH cytochrome P450 reductase (CPR). This requires a specific protein-protein interaction between CPR and aromatase.

Since estrogens are essential for reproduction, which is under intense natural selection pressures, estrogen biosynthesis dictates the fitness and success of many higher order species across evolution. Additionally, estrogen has many different regulatory roles in different species [1]. This is achieved partially by the production of estrogen by different tissues, such as gonads, placenta, trophoblast and brain. To enable tissue-specific expression of the aromatase gene (CYP19A1) for estrogen production, the pig family replicated the gene twice [6–11]. In contrast, most other mammalian species replicated only the first exon, enabling the use of alternative tissue-specific promoters [1].

To date, X-ray crystallography has resolved the structural details of only one natively isolated aromatase, obtained from human placenta [12]. Since then, recombinantly modified human aromatase has been expressed and structurally characterized by Ghosh and colleagues [13, 14]. However, aromatase has not been crystallized from any other species. Furthermore, despite the essential importance of this enzyme, aromatase has only been isolated and sequenced from a few animal species, and even fewer have been characterized in terms of catalytic activity [1, 15].

Previously, we investigated the suiform family, including pigs, in which ancient gene duplication has resulted in three tissue-specific aromatase. Employing a range of biophysical (Quartz crystal microbalance with Dissipation monitoring using biomimetic lipid bilayers) and biochemical methods (FRET and enzyme kinetics assays), we previously compared the human with porcine aromatase enzymes [16]. Experimental analysis of human P450arom and porcine gonadal [6], placental [15] and the pre-implantation blastocyst [7] P450arom displays significantly different kinetics. The human aromatase has a catalytic efficiency (V_{\max}/K_m , $\text{min}^{-1}/\mu\text{mol}$) of 14.1 compared with porcine placental and gonadal isozyme, which are 19.2 and 7.8, respectively [16]. Selecting the least catalytically efficient porcine isozyme (gonadal) of aromatase to compare with the human aromatase, we undertook sequence analysis, building homological models for structural comparisons, and finally molecular dynamics simulations and calculations that

supported differences in oligomerization *in vitro*. This *in silico* analysis found that homodimerization occurs with human aromatase [17], but this appears not to occur with the porcine gonadal aromatase, and this could be the reason for the differences in enzyme kinetics between the two species [16]. These computational data were further supported by *in vitro* biophysical studies using purified recombinant and wild-type aromatase proteins with or without their redox partner CPR and also visualized by Atomic Force Microscopy (AFM) in solution at a membrane interface [17]. These data led us to propose that other species among the evolutionary tree might also have developed alternative ways to regulate the activity of aromatase.

The availability of genomes for many species provides an opportunity to data-mine aromatase sequences across representative species of classes that lie at essential junctures across evolution. By doing this, we may begin to elucidate subtle evolutionary differences between these species which has led to their survival. Here we analyzed the aromatase enzyme across four Australian species that included the classes of reptilia, aves and mammalia. Specifically, bearded dragon (*Pogona vitticeps*), emu (*Dromaius novaehollandiae*), platypus (*Ornithorhynchus anatinus*) and tammar wallaby (*Macropus eugenii*) were compared with human P450arom.

Materials and methods

General methods

Analysis of the complexes was conducted using NAMD [18] and figures constructed using Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.) and Chemaxon software (ChemAxon Ltd.).

Sequences and alignment

Amino acid sequences were downloaded from NCBI for human (NP_001334185.1), bearded dragon (XP_020645700.1) and platypus (XP_003429099.2) species. The tammar wallaby aromatase used an unpublished sequence provided by Professor Marilyn Renfree (University of Melbourne) and colleagues as noted in our acknowledgements. The emu sequence (XP_025952868.1) was provided prior to publication [19] by Dr. Tim Sackton of Harvard University with assistance from Dr Chris Smith, Monash University. Alignment of sequences was undertaken using Clustal W 2.1 [20]. Species-specific amino acids are annotated by the species name, i.e., h = human, w = wallaby, p = platypus, e = emu and d = dragon.

Homology modeling

Three-dimensional models of the wallaby, platypus, emu and dragon aromatase were built using the crystal structure of human aromatase as a template (Protein Data Bank entry no. 3EQM) in Modeller v9.11 [21] and Swiss Model [22]. Model quality was determined using the Structure Assessment module in Swiss Model, which includes a combination of Ramachandran Plots, MolProbity [23], QMean [24] and visual analysis (Supplementary Figure S1). The models that scored the highest across multiple categories were utilized in further analysis. Analysis of the resultant models suggested that

Swiss Model was able to create higher quality models of the dragon species. Conversely, the Modeller program was able to create higher quality models using the wallaby, platypus and emu species. Flexible regions were removed from the N- and C-termini. The heme and androstenedione from the human structure were merged into the Australian animal structures and minimized under the Tripos Force Field for 10 000 iterations or until convergence was reached. Local protein contact potentials were generated with vacuum electrostatics in Pymol, using default settings (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC New York, NY, 2020).

Molecular dynamics simulations

The molecular dynamics program NAMD [18] was used to minimize the protein models. The proteins were initially solvated with TIP3P water using the Solvate plugin within VMD version 1.9 (VMD—Visual Molecular Dynamics, molecular graphics software, <https://www.ks.uiuc.edu/Research/vmd/>). The rotate to minimize volume was selected, the boundary reduced to 1.8 and the box padding increased to 20 Å in all directions; all other parameters were kept at default. Charges were then neutralized with NaCl using the autoionize plugin within VMD version 1.9 using default settings. Each structure was minimized and equalized for 1 nanosecond (2 femtosecond time step) under the CHARMM27 all-atom force field at 298 K. Langevin dynamics were used with group pressure, and Langevin piston was turned on. Trajectory snapshots were collected every picosecond. After approximately 200 picoseconds, the average root mean square deviation (RMSD) had leveled to approximately 2 Å and analysis was conducted after this time point. The total energy of the system was calculated during and at the end of the simulation runs and averaged.

Results and discussion

Amino acid sequence comparisons

The amino acid sequences from the four animal sequences were aligned to the human sequence (Figure 1). The sequences from the four animals showed high amino acid sequence identity to the human aromatase; wallaby 81%, platypus 73%, emu 75% and bearded dragon at 74%. The wallaby had the highest sequence similarity to human which is not surprising as evolutionarily they are the most similar of the five species. Regardless, there were still some major sequence differences between the P450arom of the different species. A closer look at these differences showed that some may impact their ability to function in the same manner as the human aromatase. The biggest difference between the aromatase of all four species was at the N-termini. Specifically, when compared to the human, the wallaby is truncated at the N-terminus, with 69 residues fewer than the human aromatase. Importantly, in the crystallized human aromatase [12], this region forms the A-helix. Thus, the lack of these amino acids may affect the overall 3D structure of the wallaby P450arom. The N-terminal region anchors the enzyme into the membrane bilayer; thus, the significance of the species variation across this region could have a role in enzyme mobility and thus function.

We also assessed the ability of the N-terminus of all species to be a part of a transmembrane domain using ΔG Predictor [25, 26]. This program uses the distribution pattern of amino acids present in a trans-membrane region (typically an alpha helix) to well defined physicochemical properties, that are employed to enable the biological machinery to send the protein to the membrane for

insertion into the endoplasmic reticulum [26]. In this program, the amino acid sequences are inputs, and the graphical outputs (Supplementary Figure S2) show the predicted ΔG_{app} values against the position in the sequence for different helix lengths (L). Notably, in all our species, a trans-membrane helix is predicted within the first 50 amino acids (Supplementary Figure S2). However, the N-terminal helix of the wallaby aromatase was the least likely to insert as a transmembrane domain, as the expected change in predicted free energy ($-\Delta G$) was slightly positive. However, the authors of the ΔG Predictor program note that “In principle, a negative value of ΔG_{app} indicates that the sequence is predicted to be recognized as a trans-membrane helix by the Sec [27] and integrated into the membrane. On the other hand, a positive value of ΔG_{app} does not necessarily imply that the sequence is not trans-membrane. Even if the predicted ΔG_{app} value is positive, that does not necessarily imply that the sequence entered cannot be trans-membrane. All it suggests is that the segment would not be inserted efficiently by itself and may need stabilizing interactions from surrounding helices” (<http://dgpred.cbr.su.se/index.php?p=instructions>). Thus, these data suggest that the wallaby aromatase sequence is less convincing as a trans-membrane helix region, especially with charged residues (Lys, Glu and Arg) present and which are not observed in the N-terminal trans-membrane helix in the other species. Further biophysical data using expressed wallaby aromatase enzyme would be required to provide a definitive functional significance for this region.

There are subtle variations in specific amino acid residues between species that may be associated with activity and function of the aromatase. For example, one of the residues that we have reported as capable of forming a functional homodimer, h²⁹⁰D [16], is also found in the platypus (p²⁹⁹D) and the emu (e²⁸⁹D), whereas, the wallaby and dragon have w²²¹E and d²⁹⁰E, respectively. This residue is located within the human H-helix and together with other charged residues creates an electrostatic homodimeric interface for human aromatase [16]. In this study, we compared the human with the porcine gonadal (p_g) aromatase structure using homology models and molecular dynamics and found that the porcine gonadal aromatase was more stable as a monomer. Additionally, those data were supported by dramatic differences in enzyme activities and biophysical properties [16]. Thus, the dimeric human aromatase exhibited a close salt-bridge interaction at 2.6 Å, between h²⁹⁰D...h²⁹³R, whereas the porcine gonadal aromatase showed no contact at the equivalent residues, p_g²⁸⁸E | p_g²⁹¹K suggesting it was less likely to form a stable dimer, and we proposed that it is likely to function as a monomer. The equivalent residues in the species investigated here are w²²¹E...w²²⁴G, p²⁹⁹D...p³⁰²R, e²⁸⁹D...e²⁹²A and d²⁹⁰E...d²⁹³G. This implies that the platypus is the only species we examined that is likely to exhibit activity and catalytic efficiency of aromatase similar to that found in humans. More detailed investigation, including expression and determination of the activities of the wallaby, emu and dragon P450arom is now required.

We examined the CPR-binding region, the heme-binding domain and the substrate-binding site. The residues which form the CPR binding region in humans are predominately positively charged arginine (K) and lysine (R) residues, which allow for the electrostatic interaction with the CPR [28]. In the human, this consists of eight positively charged residues: ¹⁰⁸K, ¹⁴²K, ¹⁴⁵R, ¹⁵⁰K, ³⁵²K, ³⁵⁴K, ⁴²⁰K and ⁴²⁵R; highlighted in green in Figure 1. The only difference between all four species is a single amino acid difference in the emu sequence where the human ³⁵²K is replaced by an asparagine (N). This minor change from a basic to an amidic amino acid would be expected

Human	MVLEMLNPIHYNITSIVPEAMPAATMP-----VLLLTGLFLLVWNYEGTSSIPGPG	51
Wallaby	-----	0
Platypus	MLLDGLRWAQLNLAGLLEAGPRTAAATLTLLLLLLLLLAGTPLALWSGRDAAPGPG	60
Dragon	MVAEALNPAHHNMSSILSETMPAATVP-----LLVTAFLLMWSYEETSSIPGPG	51
Emu	MVLETLPNMH-YITSLVPDVPVATVP-----ILLICFLFLIWNQEETSSIPGPG	50
	A' helix A helix β1-1 β1-2 B helix	
Human	YCMGIGPLISHGRFLWMGIGSACNYNRYVGEFMRVWISGEETLIISKSSSMFHIMKHHH	111
Wallaby	-----MYESV-DICKIMF-VKENFL--IIEIFYIFSLFRSSSMFHIMKHSH	42
Platypus	YCLGLGPLVSHGRFLWMGIGSACNYNRYGEFMRVWIGGEETLIISKSSSMFHIMKHSH	120
Dragon	YCMGIGPLLSHGRFLWMGVGSACNYNRYGEFMRVWINGEETLIISKSSSMFHVMKHGH	111
Emu	YCMGIGPLLSHGRFLWMGVGNACNYNRYGEFVRVWISGEETFIISKSSSVFHVMMKHH	110
	:: :: . *: *: * : *	
	B' helix C helix D helix	
Human	YSSRFSGSKLGLQCIGMHEKGIIFNNNPWLKTTTPFFMKALSGPGLVVRMVTVCESLKH	171
Wallaby	YISRFSGSKLGLQCIGMHENGVIIFNNNPALWKVIRPFTKALSGPGLVVRMVAVCVNSTVIH	102
Platypus	YSCRFGSKLGLQCIGMHEKGIIFNNNPWLKAIKIRPFTKALSGPGLVHMVAVCVQSTITH	180
Dragon	YISRFSGSKLGLQCIGMHENGVIIFNNNPALWKEIRPFTKALSGPGLVGMVIMCIESTTEH	171
Emu	YVSRFGSKLGLQCIGMYENGIIFNNNPALWKEIRPFTKALSGPGLVVRMIAICVESTTDH	170
	* .***** *: *	
	E helix F helix	
Human	LDRLEEVNTESGYVDVLTLLRRVMDTSNTLFLRIPLDESAIVVKIQGYFDAWQALLIKP	231
Wallaby	LDKLEEDVTNELGYVNLNLLRRVMDTSNVLFGLIPLDENAIIVVKIQGYFDAWQALLIKP	162
Platypus	LDKLEEDVTNELGYISVLTLMRRIMLNTSNALFLGIPMDESEMVMKIQGYFNAWQALLIKP	240
Dragon	LNRLLEEDVTEDGYIDCLNLMRRIMLDTSNRFLGLIPLDENAIIVKIQNYFDAWQALLIKP	231
Emu	LDKLEEDVTMGNIIVNLNLMRRIMLDTSNKFLGLIPLDENAFVVKIQNYFDAWQALLIKP	230
	* : : : : : * * : . * : *	
	G helix helix1 H helix	
Human	DIFFKISWLYKKYKSVKDLKDAIEVLIAEKRRRISTEEKLEECMDFAELILAEKRGDL	291
Wallaby	DIFFKISWLYKKYKSVKDLKDAIEVLIAEKRRRLSIDDKLEDCMDFAELILAEKRGDL	222
Platypus	DIFFKISWLYKKYKSAIKELQDAVETLLDEKRLRLSAAEKLEDHVFASQLIFAQDRGDL	300
Dragon	DIFFKISWLYKKYKSAIKELQDAVETLLDEKRLRLSAAEKLEDHVFASQLIFAQDRGDL	291
Emu	DIFFKISWLYKKYKSAIKELQDAVETLLDEKRLRLSAAEKLEDHVFASQLIFAQDRGDL	290
	***** : : : : : * : : : : : * : : : : : * : : : : : * : : : : : * : : : : *	
	I helix J helix J' helix	
Human	TRENVNQCIEMLELIAAPDTSVSLFFMLFLIAKHPNVEEAIKEIQTVIGERDIKIDDIQ	351
Wallaby	TRENVNQCIEMLELIAAPDTSVSVYFMLLLIAKHPNVEEAVINEIQTIIIGERDITIGDIQ	282
Platypus	TRENVNQCIEMLELIAAPDTSVTIFFMLLLIAKHPNVEQAIVEEIRTEIGRDIEMEDMQ	360
Dragon	TRENVNQCIEMLELIAAPDTSVTLFFMLFLIAEHQVEDAMMKEIQAQIGNREIQGEDMP	351
Emu	TAENVNQCIEMLELIAAPDTSVTLFFMLFLIAEHQVEDAMMKEIQAQIGNREIQGEDMP	350
	* ***** : : : : : * : : : : : * : : : : : * : : : : : * : : : : *	
	K helix β1-4 β2-1 β2-2 β1-3 helix2	
Human	KLKVMENFIYESMRYQPVVDLVMRKALDDVIDGYPVKKGNTNIIILNIGRMHREFFPKPN	411
Wallaby	KLKVIESFIYESMRYQPVVDLVMRKALDDVIDGYPVKKGNTNIIILNIGRMHREFFPKPN	342
Platypus	KLKVIESFIYESMRYQPVVDLVMRKALDDVIDGYPVKKGNTNIIILNIGRMHREFFPKPN	420
Dragon	KLKVVENFIYESMRYQPVVDLVMRKALDDVIDGYPVKKGNTNIIILNIGRMHREFFPKPN	411
Emu	NLKIVENFIYESMRYQPVVDLVMRKALDDVIDGYPVKKGNTNIIILNIGRMHREFFPKPN	410
	* : : : * * ***** : : : : : * : : : : : * : : : : : * : : : : *	
	helix3 haem L helix β3-3	
Human	EFTLENFAKNVPYRYFQPFGRGPRGKAGYIAMVMMKAILVTLRRFHVKTLLQGQCVESI	471
Wallaby	EFTLENFAKNVPYRYFQPFGRGPRGKAGYIAMVMMKAILVTLRRFHVKTLLTGCCLERI	402
Platypus	EFTLENFAKNVPYRYFQPFGRGPRGKAGYIAMVMMKAILVTLRRFHVKTLLTGCCLERI	479
Dragon	EFTLENFAKNVPYRYFQPFGRGPRGKAGYIAMVMMKAILVTLRRFHVKTLLTGCCLERI	469
Emu	EFTLENFAKNVPYRYFQPFGRGPRGKAGYIAMVMMKAILVTLRRFHVKTLLTGCCLERI	470
	* : : : : * * ***** : : : : : * : : : : : * : : : : : * : : : : *	
	β4-1 β4-2 β3-2	
Human	QKIHDLSLHPDETCKNMLEMIFTPRNSDRCLH--	503
Wallaby	QKNDLSLHPDETCKNMLEMIFTPRTEKCRE---	433
Platypus	HKRNDLSLHPDETCKNMLEMIFTPRTEGGLRQP--	511
Dragon	QKNDLSLHPDETCKNMLEMIFTPRINRDERCNEE	503
Emu	QKNDLSLHPDETCKNMLEMIFTPRSTNKNQDV--	503
	* : : : : * * ***** : : : : : * : : : : : * : : : : *	

Figure 1. Amino acid sequence of P450arom from wallaby (unpublished), platypus (XP_003429099.2), emu (XP_025952868.1) and dragon (XP_020645700.1) aligned with the human (NP_001334185.1) sequence. Highlighted in gray are the long helices, in blue are the beta strands and in yellow are the small helices. Amino acids marked in green represent those predicted to be involved in CPR binding. Below the aligned sequences is a code whereby, * indicates amino acids that are fully conserved; : indicates amino acids with properties that are strongly conserved and . indicates amino acids with properties that are weakly conserved across species.

to have only a slight influence on the binding affinity of the CPR. In comparison with the porcine aromatases, the positively charged amino acids required for CPR binding were identical to the human except for the porcine blastocyst isozyme that had ³⁵²E [16]. Thus, it is difficult to gauge whether this single change to the CPR binding

in the emu will impact the strength of the protein-protein interaction or affect the nature of the electron transfer needed between CPR and P450arom. Site-directed mutagenesis and enzyme activity studies on expressed aromatase would be required and were beyond the scope of this study.

Table 1. Total energy for each species obtained by molecular dynamics minimizations of P450arom structures created using Swiss Model and Modeller.

Total Energy/(kcal/mol)	wallaby	platypus	emu	dragon
Minimum	−154 367	−167 206	−161 000	−161 009
Maximum	−154 344	−167 189	−159 480	−160 993
Mean	−154 355	−167 200	−160 160	−161 001
S.D.	11.86	9.136	772.5	11.24
S.E.M.	6.850	5.275	446.0	7.948

In the heme-binding domain, there are more amino acid differences between the human (⁴³⁰FGFGPRGCA) and the other four species. This is surprising as all species are expected to bind heme. The wallaby sequence, also being a mammal, is predicted to have identical heme-binding amino acids to the human. The next most similar sequence to the human is the emu, which has a single minor conservative residue change (⁴³⁸A → V), which we expect to have no significant impact on the binding of the heme. The platypus also has a single difference from the human sequence; however, this is a non-polar to polar mutation (⁴³⁶G → S), which, depending on the location, may reduce hydrophobic interactions and thus the binding affinity for the heme. Interestingly, the bearded dragon sequence had both the conservative platypus change (⁴³⁸A → V) and the less conservative emu change (⁴³⁶G → S), plus another non-polar to polar substitution (⁴³²F → S). Since heme is essential for the transfer of electrons, these changes suggest that they do not interfere with binding to a heme group in a major way.

In the substrate-binding site, there is a requirement to bind the hydrophobic androstenedione and allow for its conversion to estrone. Therefore, the active site of aromatase needs to consist of predominately hydrophobic, non-polar residues. According to Ghosh et al. [12], in the human crystal structure, the residues which form direct van der Waals interactions with androstenedione are ¹¹⁵R, ¹³³I, ¹³⁴F, ²²¹F, ²²⁴W, ³⁰⁶A, ³¹⁰T, ³⁷⁰V, ³⁷³V, ³⁷⁴M and ⁴⁷⁷L. However, in order to orient the substrate correctly and to stabilize intermediates, there needs to be polar interactions with the two distal oxygens of the androstenedione. In the human crystal structure, this is achieved via amino acids ³⁰⁹D and ³⁷⁴M. Specifically, the ³⁰⁹D interacts with O1 (oxygen 1) on the androstenedione via its side-chain and the backbone of ³⁷⁴M interacts with O2 (oxygen 2) on of androstenedione [12]. All of the aromatase residues directly interacting with androstenedione mentioned above are common between all five species with the exception of the emu that has a conservative residue substitution at h³⁷³V becoming a similarly non-polar amino acid isoleucine (I) in the emu sequence (e³⁷²I).

Homology models were successfully created for each species

Using both Swiss Model and Modeller, we were able to successfully create high quality models of the wallaby, platypus, emu and dragon aromatase proteins. For further analysis, we used the Swiss Model dragon models and the models created by Modeller for the wallaby, platypus and emu. Amino acids computationally predicted to be unstructured were deleted from both the N- and C- termini. Molecular dynamics minimizations were then undertaken on models of each species. The total energies of each species were found to be similar (Table 1); therefore, visual analysis was required to identify specific differences between the structures of each animal species.

The homology models were aligned in Pymol (Figure 2), and the ligand (androstenedione) and heme were manually docked in and re-minimized. Based on the reasonably high sequence homology, it was not unexpected that all species aligned relatively well. The root mean square deviation (RMSD) was calculated by aligning the alpha carbons of the other species to the human structure (Supplementary Table S1 and Supplementary Figure S1). This gave RMSDs of 0.12, 0.19, 0.23 and 0.12 Å for wallaby, platypus, dragon and emu, respectively. Although the wallaby had the closest RMSD to the human, as anticipated from the sequence alignments, the wallaby is missing structure at the N-terminus when compared to the human crystal structure. Specifically, this consists of two alpha helices and approximately one and a half beta strands from the first beta sheet. Although not unexpected, the extent of the missing structural components was surprising, as from the sequence, we only expected a single alpha helix not to be present.

Visually, the backbone alignment of all the other species looks structurally similar to the human crystal structure. At approximately 0.1 Å worse than the other species RMSD, the dragon only had very minor differences in overall structural features. Specifically, this was just some very minor differences at the C-terminus, perhaps due to the two amino acids truncation in the sequence when compared to the human.

Local protein contact potentials using vacuum electrostatics were successfully generated for each animal model using Pymol. The surface charges (Figure 3) represent the surface electrostatic morphology with regions of positive (blue), negative (red) and neutral (white) surface potential. The human aromatase (Figure 3A) illustrates the positively charged (R/K) residues, shown in blue spheres that are needed to bind the redox partner protein CPR. These residues, discussed earlier, are all conserved across species (see Figure 1) except for the emu sequence in which the human ³⁵²K is replaced by an asparagine (N). In Figure 3A, this substitution reduces the positive charge (Supplementary Figure S4); however, the reduction from eight to seven positive charges is not expected to have a significant impact on the overall strength of CPR binding, although possibly it could impact the coupling efficiency of the electron transfer from CPR of aromatase in the emu.

Another feature of interest is the hypothesized human aromatase dimer interface [16]. The 90° rotation of Figure 3A shows the homodimerization interface (Figure 3B). As discussed above, analysis of the amino acid residues that form a close salt bridge in the postulated human aromatase dimer interface suggests that the platypus aromatase is the only one of these five species likely to exhibit activity and catalytic efficiency of aromatase similar to that found in the human. In Figures 3C–F, the animal species are oriented similarly to the human to show the equivalent interface used by the human aromatase for homodimerization. Interestingly, the charge distribution across the wallaby, platypus, emu and dragon has

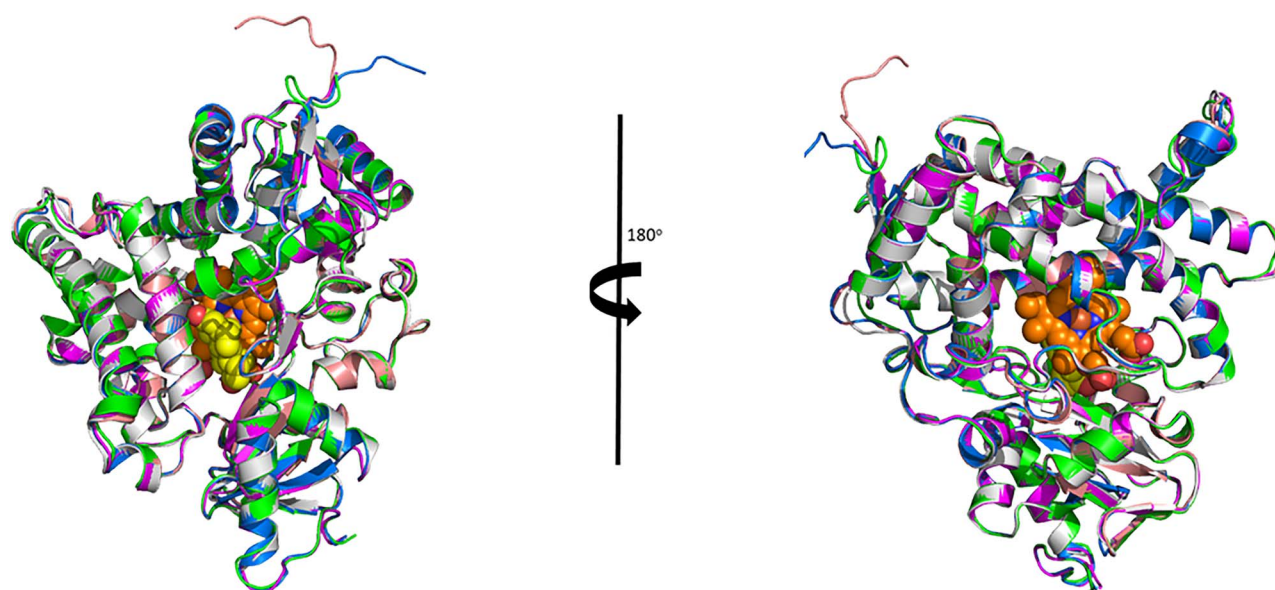


Figure 2. Alignment of the Australian animal homology models with the human P450arom crystal structure. Shown is the human structure (magenta) with the dragon (green), emu (blue), platypus (gray) and wallaby (wheat) using an alpha carbon alignment. Also shown is the position of the heme (orange spheres) and the androstenedione (yellow spheres) as crystallized in the human crystal structure.

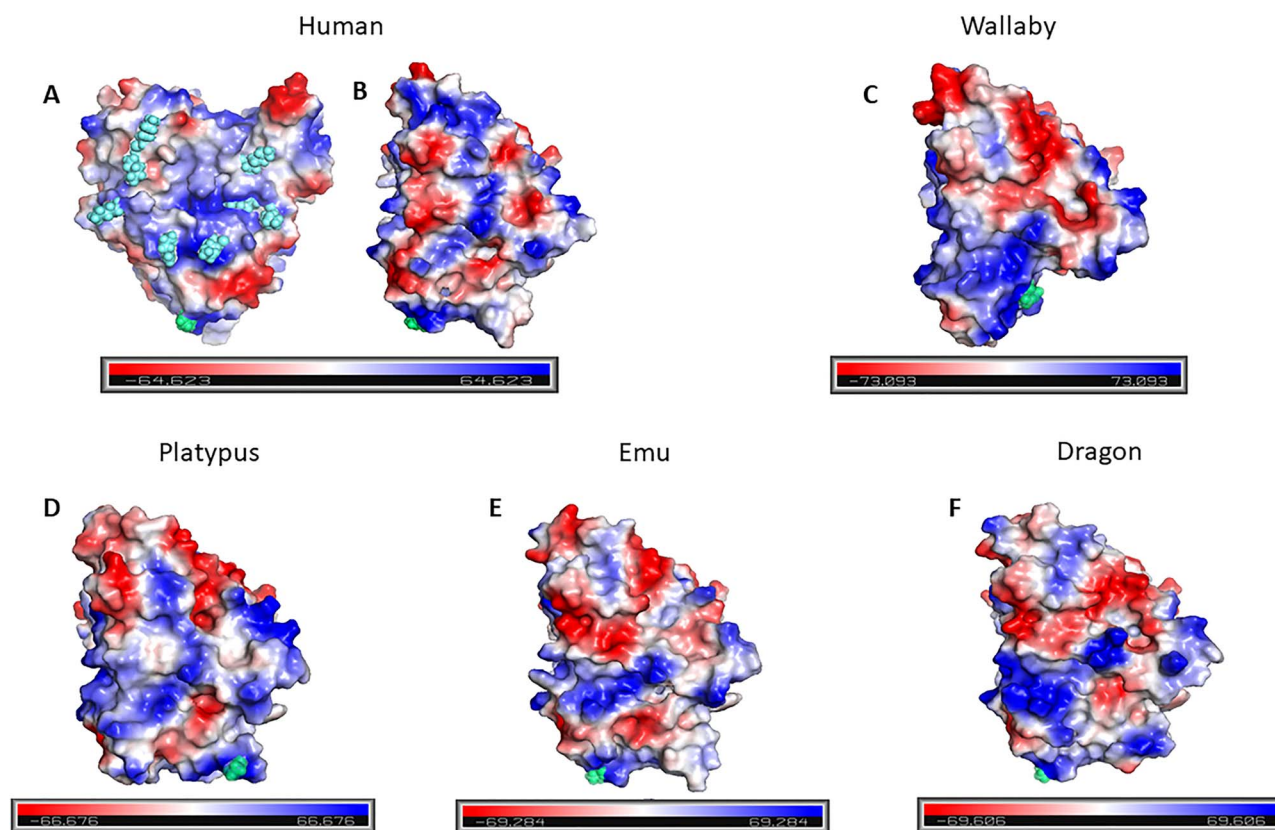


Figure 3. Surface electrostatic potential for all species. The human aromatase showing (A) the CPR binding site with positively charged residues (R/K) highlighted as light blue spheres, (B) rotation of (A) by 90° to show the face for protein-protein homodimerization [16] interface. (C–F) The equivalent homodimerization face for the wallaby, platypus, emu and dragon, respectively. All structures are with the N-termini orientated at the bottom (lime green) that represents the membrane anchor.

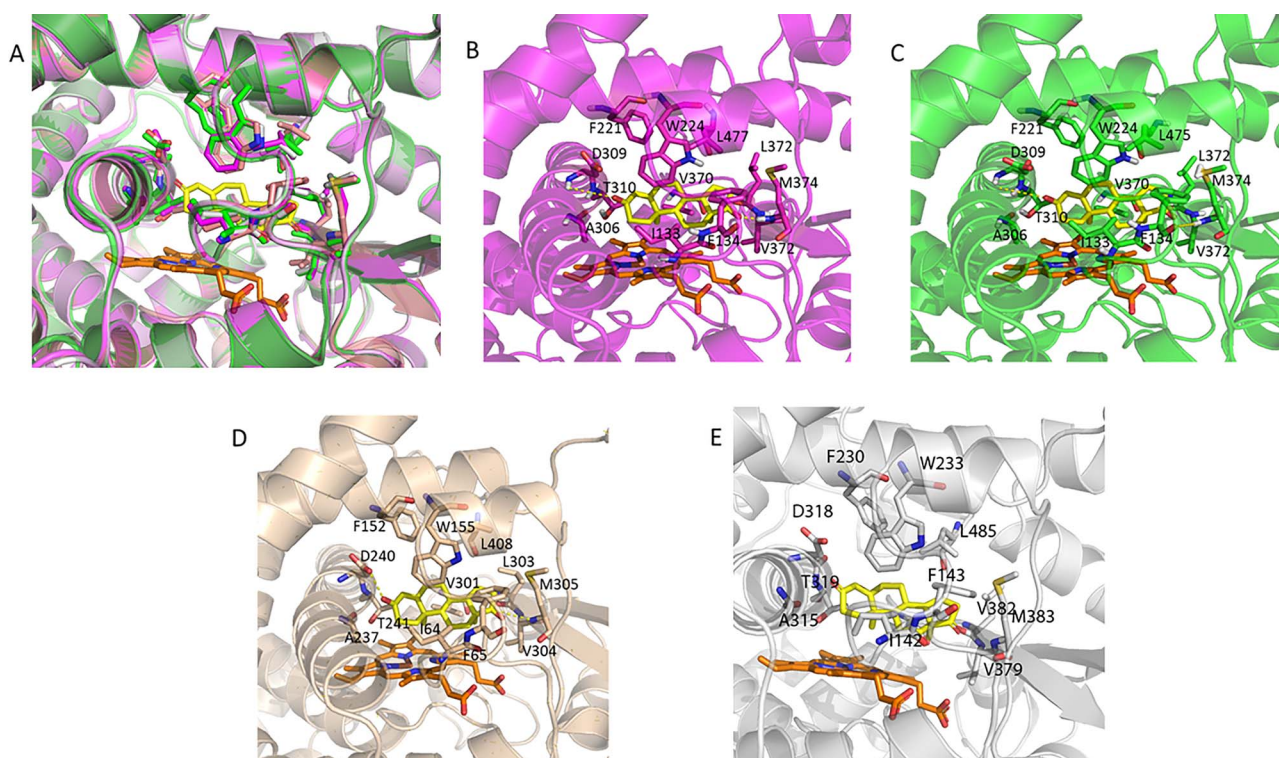


Figure 4. Human, dragon and wallaby have almost identical substrate binding sites. (A) Alignment of the human (magenta), dragon (green) and wallaby (wheat) and platypus (gray) models highlights how similar these sites are. The amino acids which interact with the androstenedione (yellow) are shown as sticks and labeled in the individual figures of (B) human only, (C) dragon only and (D) wallaby only and (E) platypus only. The heme is shown in orange.

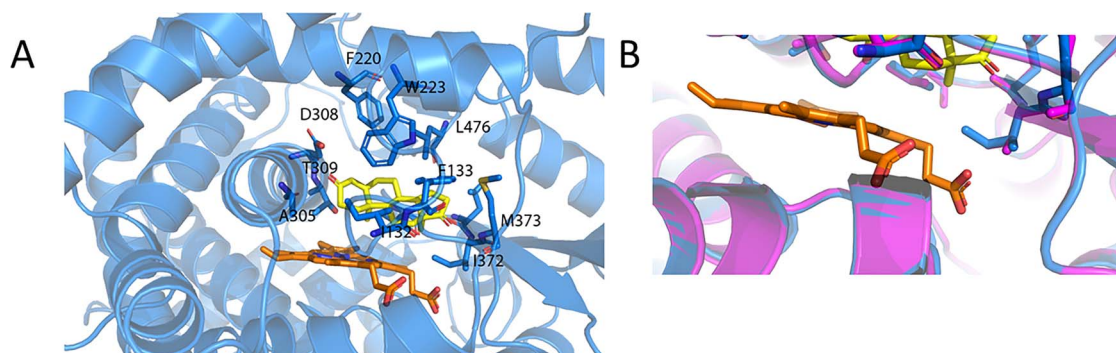


Figure 5. (A) The emu amino acids (blue) predicted to interact with androstenedione (yellow). (B) A close-up view of the h³⁷³V (magenta) to e⁴⁰⁷I (blue) amino acid difference in the active site. The heme is shown in orange.

broadly similar patterns, but are different to the human aromatase, that has a “stripe” of positive charge flanked by negative charges along the dimerization interface. These models are also shown in [Supplementary Figure S3](#), where each aromatase structure is rotated through 90° stages to provide a 3D visualization of the electrostatic surfaces. [Supplementary Figure S4](#) provides a qualitative but relative comparison between the species, and the subtle changes observed could in turn translate into altered activity of the respective aromatases.

Most species have identical substrate binding sites

Computational and visual analysis of the models showed that the active site residues of the human, dragon, platypus and wallaby

are very similar. This is interesting, as the dragon not only has less sequence identity to the human than the wallaby, but the structural model of the dragon also has a higher RMSD than the wallaby, when compared to the human ([Figure 4](#)).

The human structure has 12 hydrophobic interactions with androstenedione (h¹³³I, h¹³⁴F, h²²¹F, h²²⁴W, h³⁰⁶A, h³⁰⁹D, h³¹⁰T, h³⁷⁰V, h³⁷²L, h³⁷³V, h³⁷⁴M and h⁴⁷⁷L). These equivalent residues are found in the dragon, platypus and the wallaby and thus as expected, these same hydrophobic interactions are observed. As expected from the sequence analysis, the oxygens of the androstenedione are orientated into place by hydrogen bonds between the side-chain of ³⁰⁹D and O1 and ³⁷⁴M and O2 for human. Evolutionary, this might be interesting, as subtle variation in these structures could alter the sequential

hydroxylation reactions that result in aromatization of the steroid A-ring.

There are small differences between human and emu

There is one very small difference in the amino acid sequences of human and emu, which was predicted to have a very minor effect on androstenedione binding. Namely, h³⁷³V is an e⁴⁰⁷I in emu. Visual analysis of this residue shows that they align perfectly (Figure 5) with the extra carbon of e⁴⁰⁷I predicted to give a slight increase in hydrophobic interactions with the heme and the androstenedione.

Conclusion

The homology models built for wallaby, platypus, emu and dragon represent a wide and diverse variation in species and yet, these show good 3D structural similarities. The amino acid sequences from the four Australian animals showed high amino acid sequence identity to the human aromatase; wallaby 81%, platypus 73%, emu 75% and bearded dragon at 74%. In the 3D structures of all the enzymes, with regard to the active sites, we found that the dragon, platypus and wallaby had identical substrate binding sites to the human aromatase. Comparisons between the emu and the human showed only small differences with variations due to conserved amino acid substitutions. Among the five species, the aromatase structure was highly conserved, although there were non-secondary structures (loops and bends) that were variable and flexible, and these may result in some differences in the catalytic activity. Certainly, the electrostatic potential across the surfaces of these aromatases highlighted likely variations to the protein-protein interactions of these enzymes with CPR and possibly, like the porcine gonadal isoform, an inability to homodimerize in the case of the wallaby, emu and dragon. Given the high natural selection pressures on reproductive strategies, the relatively high degree of conservation of aromatase sequence and structure across species suggests that there is biochemically very little scope for changes to have evolved whilst still retaining aromatase activity.

Supplementary data

Supplementary data are available at *BIOLRE* online.

Acknowledgements

Access to the revised tammar wallaby genome 3.4, was kindly provided by Thomas Heider, Asao Fujiyama, Geoff Shaw, Marilyn Renfree, Andrew Pask and Rachel O'Neill. We also thank Dr's Craig Smith and Tim Sackton for access to the emu aromatase sequence prior to publication. The emu sequence is now published on NCBI and can be found at https://www.ncbi.nlm.nih.gov/genome/123?genome_assembly_id=391045. We thank Dr John Bruning, School of Biological Sciences, University of Adelaide for invaluable discussions on aromatase structure.

Conflict of interest

All the authors declare they have no financial or other potential conflicts of interest.

References

- Conley AJ, Hinshelwood M. Mammalian aromatases. *Reproduction* (Cambridge, U K) 2001; 121:685–695.
- Lange IG, Hartel A, Meyer HHD. Evolution of oestrogen functions in vertebrates. *J Steroid Biochem Mol Biol* 2003; 83:219–226.
- Akhtar M, Wright JN. Acyl-carbon bond cleaving cytochrome P450 enzymes: CYP17A1, CYP19A1 and CYP51A1. *Adv Exp Med Biol* 2015; 851:107–130.
- Guengerich FP, Macdonald TL. Mechanisms of cytochrome P-450 catalysis. *FASEB J* 1990; 4:2453–2459.
- Ortiz de Montellano PR. *Cytochrome P450: Structure, Mechanism, and Biochemistry*, 3rd ed. New York: Kluwer Academic/Plenum Publishers; 2005.
- Corbin CJ, Berger T, Ford JJ, Roselli CE, Sienkiewicz W, Trainor BC et al. Porcine hypothalamic aromatase cytochrome P450: Isoform characterization, sex-dependent activity, regional expression, and regulation by enzyme inhibition in neonatal boars. *Biol Reprod* 2009; 81:388–395.
- Corbin CJ, Hughes AL, Heffelfinger JR, Berger T, Waltzek TB, Roser JF et al. Evolution of Suiform aromatases: Ancestral duplication with conservation of tissue-specific expression in the collared peccary (*Pecari tayassu*). *J Mol Evol* 2007; 65:403–412.
- Corbin CJ, Mapes SM, Lee YM, Conley AJ. Structural and functional differences among purified recombinant mammalian aromatases: Glycosylation, N-terminal sequence and kinetic analysis of human, bovine and the porcine placental and gonadal isozymes. *Mol Cell Endocrinol* 2003; 206:147–157.
- Corbin CJ, Mapes SM, Marcos J, Shackleton CH, Morrow D, Safe S et al. Paralogues of porcine aromatase cytochrome P450: A novel hydroxylase activity is associated with the survival of a duplicated gene. *Endocrinology* 2004; 145:2157–2164.
- Corbin CJ, Moran FM, Vidal JD, Ford JJ, Wise T, Mapes SM et al. Biochemical assessment of limits to estrogen synthesis in porcine follicles. *Biol Reprod* 2003; 69:390–397.
- Corbin CJ, Trant JM, Conley AJ. Porcine gonadal and placental isozymes of aromatase cytochrome P450: Sub-cellular distribution and support by NADPH-cytochrome P450 reductase. *Mol Cell Endocrinol* 2001; 172:115–124.
- Ghosh D, Griswold J, Erman M, Pangborn W. Structural basis for androgen specificity and oestrogen synthesis in human aromatase. *Nature* 2009; 457:219–223.
- Ghosh D, Griswold J, Erman M, Pangborn W. X-ray structure of human aromatase reveals an androgen-specific active site. *J Steroid Biochem Mol Biol* 2010; 118:197–202.
- Lo J, Di Nardo G, Griswold J, Egbuta C, Wenhua J, Gilardi G et al. Structural basis for the functional roles of critical residues in human cytochrome P450 aromatase. *Biochemistry* 2013; 52:5821–5829.
- Conley AJ, Corbin CJ, Hughes AL. Adaptive evolution of mammalian aromatases: Lessons from suiformes. *J Exp Zool, Part A* 2009; 311A:346–357.
- Martin LL, Holien JK, Mizrahi D, Corbin CJ, Conley AJ, Parker MW et al. Evolutionary comparisons predict that dimerization of human cytochrome P450 aromatase increases its enzymatic activity and efficiency. *J Steroid Biochem Mol Biol* 2015; 154:294–301.
- Praporski S, Ng SM, Nguyen AD, Corbin CJ, Mechler A, Zheng J et al. Organization of cytochrome P450 enzymes involved in sex steroid synthesis: Protein-protein interactions in lipid membranes. *J Biol Chem* 2009; 284:33224–33232.
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E et al. Scalable molecular dynamics with NAMD. *J Comput Chem* 2005; 26:1781–1802.
- Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE et al. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* 2019; 364:74.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; 22:4673–4680.

21. Šali A, Pottterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. *Proteins Struct Funct Bioinf* 1995; **23**:318–326.
22. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018; **46**:W296–W303.
23. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 2010; **66**: 12–21.
24. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 2011; **27**:343–350.
25. Hessa T, Kim H, Bihlmaier K, Lundin C, Boeckl J, Andersson H et al. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 2005; **433**:377–381.
26. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 2007; **450**:1026–1030.
27. Denks K, Vogt A, Sachelaru I, Petriman N-A, Kudva R, Koch H-G. The sec translocon mediated protein transport in prokaryotes and eukaryotes. *Mol Membr Biol* 2014; **31**:58–84.
28. Hamdane D, Xia C, Im SC, Zhang H, Kim JJ, Waskell L. Structure and function of an NADPH-cytochrome P450 oxidoreductase in an open conformation capable of reducing cytochrome P450. *J Biol Chem* 2009; **284**:11374–11384.