# Structural basis for conservation in the CYP51 family

**Galina I. Lepesheva**[*] and **Michael R. Waterman**

Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-0146, USA

## Abstract

Sterol 14α-demethylases (14DM) comprise the CYP51 cytochrome P450 genome family. The 14DM reaction is essential for the biosynthesis of sterols which are necessary for production of cellular membranes. This is the most widely distributed P450, being present in all biological kingdoms. From one kingdom to another the primary amino acid sequence identity usually ranges between 30-20%. In this minireview we describe the conservation of specific amino acids and the various CYP51 orthologs and indicate the roles that they may play in the structure/function of this monooxygenase. The prediction of the roles of different amino acids in 14DM is based on high resolution tertiary structures of these enzymes which set the stage for detailed understanding of the 14α-demethylase reaction and its selective, phyla-specific inhibition which is crucial for the design of potent inhibitors for treatment of infection by pathogenic microbes.

### Keywords

cytochrome P450; sterol 14α-demethylase; CYP51; crystal structure

## 1.1 INTRODUCTION

Sterol 14α-demethylase (14DM, CYP51) is a microsomal cytochrome P450 in eukaryotes that is essential for the biosynthesis of sterols which themselves are essential for membrane formation and thus survival of a very large number of organisms in all biological kingdoms [1]. 14DM catalyzes the general reaction shown in Scheme 1: being an example of the small number of cytochrome P450s that catalyze a 3 step reaction, utilizing three molecules of molecular oxygen and three molecules of reduced pyridine nucleotide. Sterols produced after removal of the 14α-methyl group include molecules such as cholesterol in animals, ergosterol in fungi and other microbes and sitosterol in plants [1,2]. While all these compounds are necessary for membrane formation, in humans and other animals cholesterol is also essential for the synthesis of glucocorticoids, mineralocorticoids, sex hormones, bile acids and oxysterols spelling out additional important functions for CYP51 and sterol biosynthesis [3]. In yeast, it is suggested that sterols regulate gene expression [4].

CYP51 is present in all animals, plants, yeast/fungi, in many protozoa and in some bacteria. Even though it is found in all biological kingdoms, there are certain groups of organisms within the biological kingdoms that do not contain this enzyme including insects and worms.

[*]Corresponding author: Tel: 1 615 343-1373, Fax: 1 615 322-4349, galina.i.lepesheva@vanderbilt.edu.

Nevertheless, CYP51 is by far the most widely distributed of the more than 10,000 cytochromes P450 (http://drnelson.utmem.edu/CytochromeP450.html). In pathogenic microbes, 14DM is a drug target for human, animal and plant infections. The goal of this minireview is to present a comparison of the primary sequences and three-dimensional structures of CYP51 which allow an understanding of the functional basis of sterol 14α-demethylation and the potential for drug design

## 1.2 PRIMARY SEQUENCES OF 14DMs DISPLAY A LOW IDENTITY

Traditionally, sequence analysis is used to identify cytochrome P450 genes [6]. Generally primary sequence comparison below 40% identity indicates separate gene families and identity above 55% indicates members of the same subfamily. The different subfamilies arise from comparison of sequences between 40%-55% identity.

The identities of the primary sequences of more than 180 known 14DM monooxygenases range between about 22-30% when comparing one kingdom to another (e.g 21% between CYP51 from *Aspergillus nidulans* and *Trypanosoma congolense*, 26% between *Leishmania infantum* and human). The identities within kingdoms are higher but vary depending on the evolutionary diversity of the group. Thus, the CYP51s from mammals display more than 90% identity, on average (for example, the identity between CYP51 in human and chimpanzee is 98%, only 3 amino acids are different). The identity drops to about 75% in vertebrates, the sequences of the human and sea-urchin enzymes are only 64% identical, while human and *Monosiga brevicollis* (the unicellular organism which sometimes is classified as the most ancient animal [5]) share only 45% CYP51 sequence identity. In plants, the identities between the sequences vary from 98 to 41% (e.g. spruce to moth). Plant/algae CYP51s share only about 45% identity (e.g. sorghum to *Galdiera sulphuraria*). Even broader variations are seen in fungal orthologs: the pair *Candida glabrata - Aspergillus nidulans* (yeast-filamentous fungi) has 42% identity. The value increases to 65% for *Penicillum digitatum - Aspergillus fumigatus* (filamentous fungi) and to 62% for *C. albicans - Saccharomyces cerevisiae* (yeast), but drops down to only 44% for *S. cerevisiae - Schizosaccharomyces pombe* (true yeast – fission yeast).

When aligned, it is quite clear that several of the CYP51 sequences are likely to contain sequencing errors, an example being the incorrect 20 amino acid residue C-terminal fragment in *Phanerochaete chrysosporium* [6]. Therefore, counting identical residues in the multiple sequence alignment formally results in only 15 residues that are invariant in all protein sequences including those from bacteria. The formal number of residues identical amongst eukaryotic CYP51s including protozoa, fungi, plants/alga and animals is currently 24. However, probable sequencing errors and "misaligned" residues no doubt increase this number. For example, G81 in *Trypanosomatidae* clearly plays the same role as the corresponding glycine located one residue towards the C-terminus in the CYP51s from all other kingdoms. Without question, there must be a C and not S coordinated to the heme in CYP51 from cotton. Moreover, when aligned, several sequences explicitly display incorrect fragments (233-273 in *Penicillium italicum),* etc. The 36 residues that are more than 99% conserved in eukaryotic CYP51s are listed in Table 1A, their locations in the aligned CYP51 sequences are marked with triangles in Figure 1.

Low identities in the primary sequences of sterol 14α-demethylases, especially from evolutionarily distant species, have made it difficult to comprehend what drives these enzymes to be so devoted to their single reaction and strictly specific catalytic mechanism.

## 1.3 EUKARYOTIC 14DMs HAVE HIGHLY SIMILAR THREE-DIMENSIONAL STRUCTURES

As we reported recently [7], the first eukaryotic CYP51 structures [3g1q and 3gw9] were found 1) significantly different from the structures of the water soluble CYP51 ortholog from *Mycobacterium tuberculosis* [8], 2) with no significant conformational rearrangements upon binding of a highly potent inhibitor. Contrary to the bacterial structure, 14DM from *Trypanosoma (T) brucei* does not have an opening to the upper surface: the heme is completely covered and buried inside the protein. Instead, the substrate access channel clearly goes to the left part of the distal surface. The I-helix is much less bent, helices B' F and G are located much closer to the protein core. As a result, the active site volume of *T. brucei* 14DM is much smaller than that in *M. tuberculosis* CYP51 and comparable with the volumes of the 14DM sterol substrates [7]. This might suggest that CYP51 in *M. tuberculosis* can have an additional/alternative function *in vivo*.

Within the last year the 14DM structures from four eukaryotic organisms have been determined (*T. brucei, T. cruzi, Leishmania infantum* and human), all of them being found strikingly similar: the Cα rms deviation between the human and *Trypanosomatidae* orthologs is below 1.8 Å, binding of three structurally different ligands to *T. cruzi CYP51* causes variation within 0.9-1.4 Å [3K1O, 3KHM, and 3KSW]. Thus, it appears that CYP51 resists catalytic diversification by preserving high similarity of their general fold and, through this, the rigidity of the substrate binding cavity.

## 1.4 STRUCTURAL VIEW ON THE CONSERVED RESIDUES IN THE CYP51 FAMILY

Combined information from the three structural levels (Table 1, Figures 1 and 2) uncovers the basis to understand how CYP51 enzymes might achieve their conservation. The majority of the residues conserved in the whole family are most likely to play a structural role, perhaps essential for the proper folding of the nascent polypeptide chain. Those include four groups 1) the heme binding area that presents significant conservation across the CYP families (the region from 415 to 424 in CYP51 from *T. brucei* (Figure 1)), 2) the N-terminal proline rich region (P33 area), which (similar to the more extended proline rich region in the CYP2 family [9]) makes van der Waals contacts with β1-2 (T78); 3) other conserved glycines and prolines (12 in total) that separate the major secondary structural elements thus preserving their length and location within the CYP51 sequence; and 4) five residues that provide additional heme support by forming hydrogen bonds with the porphyrin ring A and D propionates (Figure 3A). Three of these residues, Y103, R361 and H420 (probable sequence error in *Citrus reticulate*), are identical in the whole CYP51 family. Though R124 (Table 1B) is conserved only amongst protozoa and bacteria, in the CYP51s from other phyla there is always a positive charge (R or K) in the C-helix, four residues towards the C-terminus (this residue (K156) is H-bonded to ring D in the heme in human CYP51). Y116 (ηB") is conserved in all phyla except for plants, where it is replaced with F. We presume that in plant CYP51s (similar to the situation with R124) the heme support in this region can be provided by the conserved Y located one turn to the C-terminus (Y124 in potato).

The CYP51 I-helix signature sequence (-aGQHtS-) and E205 (the F' helix) are most likely involved in the proton delivery route (Figure 3B) that must be conserved in the whole family [7]. Mutation of the corresponding QHT and S has been experimentally proven to significantly affect the activity of rat CYP51 [10].

Most of the highly conserved residues that form the surface of the CYP51 substrate binding cavity (Y103, F105, P108, F110, G111, V114 and F116) are from the CYP51 B'-helix

signature region [1]. This explains the crucial effect of their mutagenesis in studies of the enzyme substrate binding and catalysis [11,12]. Interestingly, except for F110, these residues appear to play roles additional to the interaction with the substrate. Thus, Y103 and Y116 support the heme, and G111 breaks the helical turn. P108 (which substitution to G was found to strongly affect the enzyme activity and spectral response to the substrate without influencing the P450 expression [1] is likely to control the local topology of the substrate binding surface in this area.

## 1.5 14DM SUBSTRATE BINDING CAVITY IS HIGHLY REPRESENTED BY PHYLA-SPECIFIC RESIDUES

Seven secondary structural elements in CYP51s can be involved in substrate recognition/ binding. The substrate access channel is bordered by helices A', F" and β4 hairpin; the binding cavity is formed by five segments: the B'-helix - B'/C loop, the N-terminal portion of helix C, the N-terminal portion of helix I, K/β1-4 loop - β1-4 and β4 hairpin.

Quite surprisingly, the residues that might directly participate in the interaction with the substrate display a rather moderate degree of conservation across the kingdoms. Only three of them, Y103, F110, and Q126 could be included in Table 1A as potentially identical in the whole family. The other residues vary across the kingdoms, yet have pronounced tendency to be phyla-specific (Table 1B) Thus, M106 (*Trypanosomatidae*) is always N in plants yet T in fungi and animals. A115 is replaced by V in plants, by V or I in fungi. L127 is substituted with F in plants and with K in fungi and animals. P210 (substrate access channel) is H in animals. F290 is always M in fungi yet L in animals, M460 is conserved in all kingdoms except for filamentous fungi which contain L in this position and so forth.

Since the access channel entrance in CYP51 is highly hydrophobic [13], it is most likely that the substrate enters the channel with its hydrophobic arm, the only hydrophilic portion of the molecule (C3-OH) being positioned closest to the channel entrance. If this orientation is correct, modeling of eburicol (24-methylenedihydrolanosterol) into the structure of CYP51 from *T. cruzi* (Figure 3C) in such a way that its 14α-methyl group is positioned above the heme plane at the distance of about 5 Å from the iron and the β -methyl group of the sterol C4 atom (C28[β]) is oriented towards I105 [14] suggests that the C4 α-methyl group (C29[α]) will form hydrophobic interactions with the αK/ β1-4 loop (L356) and β4 hairpin (M461), the C3-OH will be between M358 and M360 (β1-4) and the distal part of the sterol arm will be intercalated between the helices C (L127) and I (M284). The phyla-specific residues in the binding cavity must be responsible for differences in substrate preferences and turnover rates of CYP51s across biological kingdoms (as it has been proven for Y/I105 in trypanosomes [14,15].

## 1.6. CONCLUSION

The CYP51 family members display low identity in their primary structures in different biological kingdoms yet preserve strict conservation at both secondary and tertiary structural levels. Better understanding of structure/function relations for these unusual members of the cytochrome P450 superfamily not only will provide mechanistic insights into their strictly specific three-step reaction and open an opportunity for rational development of novel highly selective antimicrobial drugs [16], but might also serve as the basis for directed P450 evolution.

## Acknowledgments

The authors congratulate our friend, Klaus Ruckpaul, on his 80th birthday. We appreciate the many important discoveries you have made, Klaus, and your dedication to training young investigators. The field of P450 research is much richer because you have been one of our leaders.

## REFERENCES

[1]. Lepesheva GI, Waterman MR. Sterol 14alpha-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. Biochim. Biophys. Acta 2007;1770:467–477. [PubMed: 16963187]

[2]. Waterman MR, Lepesheva GI. Sterol 14 alpha-demethylase, an abundant and essential mixed-function oxidase. Biochem. Biophys. Res. Commun 2005;338:418–422. [PubMed: 16153595]

[3]. Waterman, MR. Production of Sex Steroids. In: Huges, IA., editor. Frontiers in Endocrinology Sex Differentiation: Clinical and Biological Aspects, Sex Differentiation: Clinical and Biological Aspects, V20-Serono Symposia Series. Christengraf, via Anagnina, Rome, Italy: 1996. p. 61-69.

[4]. Dahl C, Biemann HP, Dahl J. A protein kinase antigenically related to pp60v-src possibly involved in yeast cell cycle control: positive *in vivo* regulation by sterol. Proc. Natl. Acad. Sci. U. S. A 1987;84:4012–4016. [PubMed: 2438691]

[5]. King N, et al. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 2008;451:783–788. [PubMed: 18273011]

[6]. Warrilow A, Ugochukwu C, Lamb D, Kelly D, Kelly S. Expression and characterization of CYP51, the ancient sterol 14-demethylase activity for cytochromes P450 (CYP), in the white-rot fungus *Phanerochaete chrysosporium*. Lipids 2008;43:1143–1153. [PubMed: 18853217]

[7]. Lepesheva GI, Park HW, Hargrove TY, Vanhollebeke B, Wawrzak Z, Harp JM, Sundaramoorthy M, Nes WD, Pays E, Chaudhuri M, Villalta F, Waterman MR. Crystal structures of *Trypanosoma brucei* sterol 14alpha-demethylase and implications for selective treatment of human infections. J. Biol. Chem 2010;285:1773–1780. [PubMed: 19923211]

[8]. Bellamine A, Mangla AT, Nes WD, Waterman MR. Characterization and catalytic properties of the sterol 14alpha-demethylase from *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U S A 1999;96:8937–8942. [PubMed: 10430874]

[9]. Kemper B. Structural basis for the role in protein folding of conserved praline-rich regions in cytochromes P450. Toxicol. Appl. Pharmacol 2004;199:305–315. [PubMed: 15364546]

[10]. Nitahara Y, Kishimoto K, Yabusaki Y, Gotoh O, Yoshida Y, Horiuchi T, Aoyama YJ. The amino acid residues affecting the activity and azole susceptibility of rat CYP51 (sterol 14-demethylase P450). J. Biochem 2001;129:761–768. [PubMed: 11328599]

[11]. Lepesheva GI, Virus C, Waterman MR. Conservation in the CYP51 family. Role of the B' helix/BC loop and helices F and G in enzymatic function. Biochemistry 2003;42:9091–9101. [PubMed: 12885242]

[12]. Lepesheva GI, Seliskar M, Knutson CG, Stourman NV, Rozman D, Waterman MR. Conformational dynamics in the F/G segment of CYP51 from *Mycobacterium tuberculosis* monitored by FRET. Arch. Biochem. Biophys 2007;464:221–227. [PubMed: 17585868]

[13]. Lepesheva GI, Hargrove TY, Anderson S, Kleshchenko Y, Futak V, Wawrzak Z, Villalta F, Waterman MR. Structural insights into inhibition of sterol 14α-demethylase in the human pathogen *Trypanosoma cruzi*. J. Biol. Chem (under review).

[14]. Lepesheva GI, Zaitseva NG, Nes WD, Zhou W, Arase ML, Liu J, Hill GC, Waterman MR. CYP51 from *Trypanosoma cruzi*: a phyla-specific residue in the B' helix defines substrate preferences of sterol 14alpha-demethylase. J. Biol. Chem 2006;281:3577–85. [PubMed: 16321980]

[15]. Lepesheva GI, Hargrove TY, Ott RD, Nes WD, Waterman MR. Biodiversity of CYP51 in trypanosomes. Biochem. Soc. Trans 2006;34(Pt 6):1161–1164. [PubMed: 17073776]

[16]. Lepesheva GI, Ott RD, Hargrove TY, Kleshchenko YY, Schuster I, Nes WD, Hill GC, Villalta F, Waterman MR. Sterol 14alpha-demethylase as a potential target for antitrypanosomal therapy: enzyme inhibition and parasite cell growth. Chem. Biol 2007;11:1283–1293. [PubMed: 18022567]
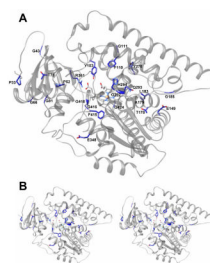
**Figure 1. Multiple sequence alignment of the CYP51 family**

The alignment was carried out using 185 CYP51 sequences and displays two protozoan (*Trypanosoma brucei* and *Monosiga brevicollis*), two animal (monkey and sea urchin), two plant (potato and pine tree) and two fungal family members (*Aspergillus nidulans* (filamentous) and *Candida glabrata* (yeast)). The residues identical in all sequenced eukaryotic CYP51s are marked with black triangles, the light grey triangles show the residues conserved in more than 99% sequences. The black circles mark the phyla-specific residues included in Table 1B.

**Figure 2. Ribbon diagram of CYP51 from *T. brucei***
A**.** The residues identical in all eukaryotic CYP51s (black triangles in Figure 1) are labeled, the side chains are presented as stick models. B**.** Stereo view.
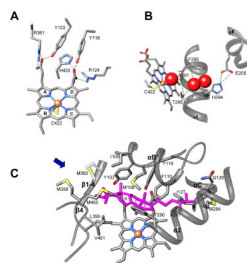
**Figure 3. Three groups of conservation in sterol 14α-demethylase**
A. Heme support. B. Hypothetical proton delivery route: ligand-free *T. brucei* CYP51 [3g1q]. The heme, iron coordinating C422 and the side chains of the residues that form H-bonds (green dashed lines) are shown as stick models. Three water molecules are presented as red spheres. C. Substrate binding cavity: *T. cruzi* CYP51 [3k10] with the modeled preferred substrate eburicol (magenta). Side chains of the family conserved and phyla-specific residues that are located within wan der Waals contacts (<4.5A) with eburicol are labeled. The backbone of the secondary structural elements (αB'/B'C loop, N-terminal parts of αC and αI, β1-4 with the preceding K'/β1-4 loop and the β4 hairpin) which form the binding cavity is displayed as ribbons. The substrate access channel location is marked with the blue arrow. Sterol rings are numbered.
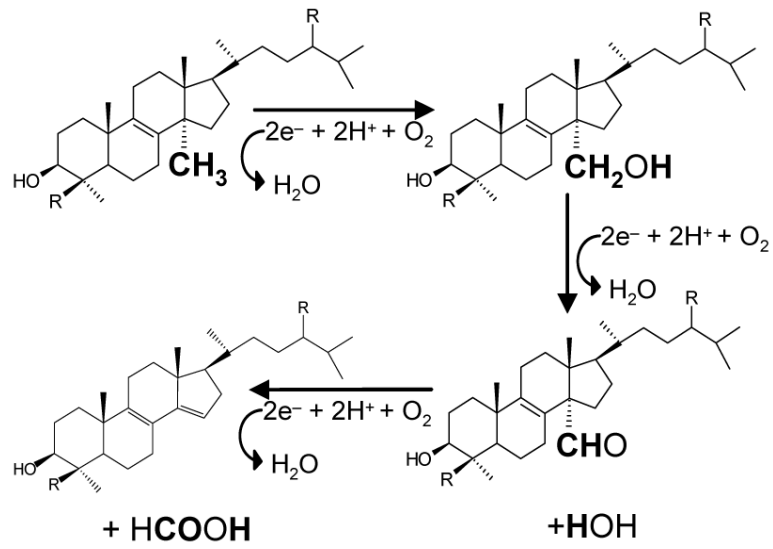
**Scheme 1.**
Sterol 14α-demethylase reaction

**Table 1**

Conserved residues in the CYP51 family. Location and possible role (*T. brucei* numbering)

| Residue | Remark on conservation (if not 100% identical) | Predicted role | Location/details |
|---------|-----------------------------------------------|----------------|------------------|
| **A). Residues identical in eukaryotic CYP51s (Residues shown in bold are 100% identical in bacterial CYP51s as well)** | | | |
| P33 | | structural/folding | Follows the N-terminal membrane anchor, proline rich region, essential for P450 expression (nucleus for initial folding?) |
| **G43**\* | | structural | Starts αA' |
| **P52** | | structural | αA/ separates αA from αA' by providing a sharp turn between them |
| G66 | | structural | αA - β1-2/ Separates helix A and β1-2 strand allowing a sharp turn |
| T78 | | unclear (folding?) | β1-2/ forms van der Waals contacts with the proline rich region |
| G81 | Misaligned in *Trypanosomatidae* (left shifted) | structural | β1-2 – αB/ Separates β1-2 and helix B |
| **Y103** | | heme support/ substrate binding | αB' signature/H-bond with the ring A propionate, forms the surface of the substrate binding cavity |
| P108 | Y122 in *Cunninghamella elegans* (seq. error?) | structural/(functional?) | αB' signature /Turn within the B' helix |
| **F110** | | substrate binding | αB' signature/ forms the surface of the substrate binding cavity (possibly interacts with the sterol C16/the arm area) |
| **G111** | | structural | αB' signature/ brakes αB' |
| V114 | L122 in *Aphanomyces euteiches* | unclear | αB'- αB" loop |
| Q126 | H in *Schizosaccharomyces pombe* (seq.error?) | substrate binding | αC/ forms the surface of the substrate binding cavity, (sterol arm) |
| E149 | | unclear | αD/proximal surface |
| T178 | | unclear | αE/ proximal surface |
| A179 | | unclear | αE/ proximal surface |
| **L183** | | unclear | αE/ proximal surface |
| **G185** | | structural | αF'/ Forms turn separating helices E and F' |
| R247 | T255 in *Penicillum italicum* (seq. error?) | unclear | αG |
| Y270 | | unclear | HI loop |
| **G292** | | proton delivery route | αI /I-helix signature |
| Q293 | | proton delivery route | αI /I-helix signature |
| **H294** | | proton delivery route | αI /I-helix signature/ H-bond with E205 |
| T/S295 | S in filamentous fungi | proton delivery route | αI /I-helix signature/ "conserved threonine", forms the surface of the substrate binding cavity |
| S296 | T261 in *Citrus reticulate* | unclear | αI//I-helix signature |
| W302 | G267 in *Citrus reticulate* (seq. error?) | unclear | αI |
| **E348** | | structural | αK/ ExxR salt bridge |
| I350 | I/L | unclear | αK |
| R351 | C372 in *Mycosphaerella* | structural | αK/ ExxR salt bridge |

| Residue | Remark on conservation (if not 100% identical) | Predicted role | Location/details |
|---|---|---|---|
| | *graminicola* (seq. error?) | | |
| **R361** | | heme support | β1-4 /H-bond with the ring A propionate |
| R404 | P404 in rice | unclear | αK' - ηK" loop (meander) |
| F415 | | heme binding area | αK' - ηK" loop |
| **G416** | | heme binding area | αK' - ηK" loop |
| **G418** | | heme binding area | ηK" |
| H420 | R394 in *Citrus reticulate* (seq. error?) | heme binding area | ηK'/H-bond with the ring D propionate |
| C422 | S439 in cotton (seq. error) | heme binding area | ηK"–αL loop/ iron coordinating Cys (the fifth axial ligand) |
| **G424** | | structural | ηK"–αL loop/ starts helix L |
| **B). Examples of conserved phyla-specific residues with predicted function** | | | |
| F105 | L (animals, fungi), I (*T. cruzi, Monosiga brevicollis*) | substrate binding | αB' signature/ forms the surface of the substrate binding cavity (sterol C4 atom area) |
| Y116 | F (plants) | heme support/ substrate binding | αB"/H-bond with the ring D propionate (the same role in plants might play a Y located 4 residues (one turn) to the C-terminus (Y124 in potato sequence)), forms the surface of the substrate binding cavity |
| R124 | Conserved in *Trypanosomatidae* and bacteria | heme support | αC/H-bond with the ring D propionate; the same role in other phyla must be played by the residue located one turn to the C-terminus: R (plants), K (fungi, animals: K156 in human) |
| L127 | F (plants), K (fungi, animals) | substrate binding | αC/ forms the surface of the substrate binding cavity (sterol arm) |
| P210 | H in vertebrates | substrate recognition? | αF"/forms the surface of the access channel entrance |
| V213 | F (fungi), W (animals) | substrate recognition? | αF"/forms the surface of the access channel entrance |
| M284 | L (plants), M/I/L (Fungi), M (animals) | substrate binding | αI/ forms the surface of the substrate binding cavity (sterol arm) |
| L356 | I/L (fungi), I (animals) | substrate binding | αK-β1-4 loop/forms the surface of the substrate binding cavity (sterol C29[α]) |
| M360 | L/M/F (fungi) | substrate binding | β1-4/forms the surface of the substrate binding cavity (sterol C3-OH) |
| M460 | M/L (fungi) | substrate binding | β4 hairpin/forms the surface of the substrate binding cavity (sterol C3-OH) |
| V461 | F/V (fungi), I (animals) | substrate binding | β4 hairpin/forms the surface of the substrate binding cavity (sterol C29[α]) |