J Mol Model (2010) 16:279–284
DOI 10.1007/s00894-009-0527-9

ORIGINAL PAPER

Evolutionary trace analysis of CYP51 family: implication for site-directed mutagenesis and novel antifungal drug design

Chunquan Sheng · Shuanghong Chen · Haitao Ji · Guoqiang Dong · Xiaoyin Che · Wenya Wang · Zhenyuan Miao · Jianzhong Yao · Jiaguo Lü · Wei Guo · Wannian Zhang

Received: 18 January 2009 / Accepted: 22 April 2009 / Published online: 11 July 2009

Abstract Lanosterol 14α-demethylase (CYP51) is an essential enzyme in the fungal life cycle and also an important target for the antifungal drug development. Based on the multiple sequence alignments of CYP51 family, an evolutionary tree of the CYP51 family was constructed by the evolutionary trace (ET) method. The identified trace residues could provide a reliable and rational guide to the design of CYP51 mutations and give more information about the detailed mechanism of substrate (drug) recognition and binding. The reliability of ET analysis to identify residues of functional importance was validated by the reported site-directed mutagenesis studies of CYP51s. Several residues in the active site were also validated by our mutagenesis studies. Mapping the identified trace residues onto the active site of the modeled structure of *Candida albicans* CYP51 (CACYP51) may provide useful information for the design of novel antifungal agents.

Keywords Antifungal drug design · CYP51 family · Evolutionary trace · Mutation design

Electronic supplementary material The online version of this article (doi:10.1007/s00894-009-0527-9) contains supplementary material, which is available to authorized users.

C. Sheng · G. Dong · X. Che · W. Wang · Z. Miao (✉) · J. Yao · J. Lü · W. Guo · W. Zhang (✉)
School of Pharmacy, Military Key Laboratory of Medicinal Chemistry, Second Military Medical University,
325 Guohe Road,
Shanghai 200433, People's Republic of China
e-mail: miaozhenyuan@hotmail.com
e-mail: zhangwnk@hotmail.com

S. Chen
Naval Medical Research Institute,
880 Xiangying Road,
Shanghai 200433, People's Republic of China

H. Ji
Department of Chemistry, Department of Biochemistry, Molecular Biology, and Cell Biology, and Center for Drug Discovery and Chemical Biology, Northwestern University,
Evanston, IL 60208-3113, USA

Introduction

Lanosterol 14α-demethylase (CYP51) is a member of the cytochrome P450 superfamily, which catalyzes the oxidative removal of the 14α-methyl group (C-32) of lanosterol to give a $^{\triangle 14,15}$-desaturated intermediates in ergosterol biosynthesis [1]. As an essential enzyme in the fungal life cycle, CYP51 has been a primary target for azole antifungal agents, such as fluconazole and itraconazole, which are well-established drugs for the treatment of topical and systemic mycoses [2]. The crystal structures of CYP51 from *Mycobacterium tuberculosis* (MTCYP51) for free state and ligand-bound forms have been solved [3–6]. However, the crystal structure of CYP51 from fungi has not been solved until now, because they are large membrane proteins and very difficult to be crystallized under the current conditions. In our previous studies, we have constructed the three-dimensional (3D) model of CYP51 from *Candida albicans* (CACYP51) by homology modeling and the binding modes of the natural substrate and the antifungal azoles were investigated [7–9]. Based on the 3D model, novel CACYP51 inhibitors have been designed and synthesized [8, 10]. In the present study, the evolutionary trace (ET) method [11] was used to identify important functional residues responsible for catalytic activity and find specific residues of fungal CYP51, which is valuable to design selective inhibitors targeting fungal CYP51s.

Computational methods

Protein sequences and structures

A total of 49 CYP51 members from different species were used in the ET analysis. Their amino acid sequences were taken from SWISS-PROTEIN and NCBI Data Banks. Among the reported crystal structures of MTCYP51, the MTCYP51-fluconazole complex was used in the present study, because it is very important in antifungal drug design. The crystallographic coordinates of MTCYP51 in complex with fluconazole (0.22 Å resolution, R_{cryst}=0.204) were obtained from the RCSB Protein Data Bank as entry 1EA1 [4]. The 3D structure of CACYP51 and the docking model of fluconazole were taken from our previously modeled structures [9]. All calculations were performed with the commercially available InsightII 2000 software package [12].

Sequence alignments of CYP51 family and evolutionary trace analysis

Multiple sequence alignments were performed using the Align 123 module of InsightII 2000 [12] software package. The Blosum-62 matrix [13] was used with a gap penalty of 10 and a gap extension penalty of 1. The alignment was then manually optimized by the knowledge-based method [7]. The aims of the optimization are: (1) to match known secondary elements in the crystal structure of MTCYP51; (2) to locate insertions and deletions in CYP51 outside the predicted α-helical and β-sheet regions of known P450 proteins; (3) to match irregular spatially conserved regions (SCRs) which are essential for protein structure and function, such as the Cys-pocket and substrate access channel identified from the crystal structure of MTCYP51.

This alignment was then used to generate an evolutionary tree using the EvolutionaryTrace program of the binding site module of InsightII 2000 software package. Based on the alignment, conserved trace residues and class-specific trace residues were identified at different percentage sequence identity cutoffs (PICs). By default, they are identified at PIC=30%, 40%, 50%, 60%, 70%, 80%, and 90%. Spatial clusters of the evolutionary trace residues were generated using a hierarchical clustering method based on the distances between the side chain atoms of the identified trace residues. Because exposed functional residues are more likely to be responsible for binding or enzymatic activity, they are less likely to mutate than residues on protein surfaces during evolution [11]. On the other hand, the buried residues are more important for maintaining the structural integrity. Therefore, the solvent accessible surface of CACYP51 was precalculated using Access_Surf command from ProStat module and only those trace residues with larger than 15% percent solvent accessible surface were mapped onto the 3D structure of CACYP51 for further analysis.

Results and discussion

Multiple sequence alignment and analysis of evolutionary tree

The sequences of 49 members of CYP51 family were aligned (see Supplementary Material) and the insertions or deletions were avoided where conserved secondary structure motifs were present. The 49 sequences in the CYP51 family vary in length from 414 to 561 residues and belong to two major kingdoms of life: bacteria (7 organisms) and eukarya (42 organisms) including fungi, plants, protozoan and animals. The sequence identities between various CYP51 proteins range from 29% to 93%. After the multiple sequence alignment, an evolutionary tree was constructed on the basis of percentage sequence identity cutoffs (PICs). The evolutionary tree shows that the CYP51 family has three major branches (Fig. 1). One of these is composed of bacteria and plant CYP51s and the fungal CYP51s form another large subgroup with sequence identities larger than 45%. The remaining subgroup is composed of animal CYP51s with sequence identities larger than 65%. It is worth noting that mammalian CYP51s (especially human CYP51) share a close relationship to fungal CYP51s with sequence identities around 45%. It was reported that the therapeutic side effects of azole antifungal agents were partly due to the interactions of azoles with human CYP51 [7]. Thus, the identification of class-specific residues in the active site of fungal subgroup might be of great value to design novel compounds with high specificity and separate their antifungal activity from toxicity.

Trace residues in substrate access channel

Trace residues are grouped into conserved residues and class-specific residues. Conserved residues are essential for maintaining protein structure and functions. Residues conserved within subgroups and different from the other subgroups are identified as class-specific residues. With the increase of the PIC level, the sequences of the protein family can be partitioned into different specific classes using hierarchical clustering method, and the identified class-specific trace residues share higher specificity. Figure 2 shows the distribution of the identified conserved trace residues on the 3D structure of CACYP51. The substrate (ligand) binding pocket was embedded in the

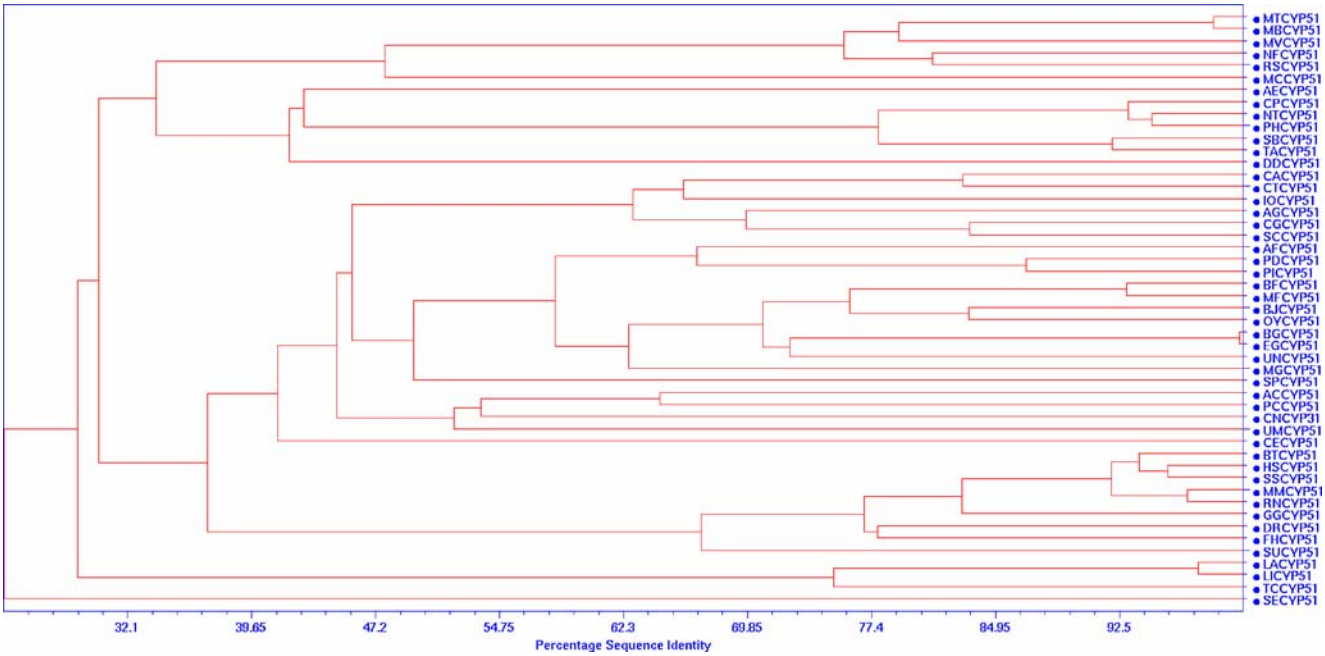

Fig. 1 Sequence identity dendrogram of 49 sequences of CYP51 family

interior of the P450 proteins. The active site was accessible through a long channel. The crystal structure of MTCYP51 [4] and our modeled structure of CACYP51 [9] show two putative substrate access channels. Channel 1 (B' helix / BC loop) is parallel to the heme, while channel 2 (C-terminal of the F-helix and N-terminus of the G-helix, the FG loop) is perpendicular to the heme. The identified trace residues in the substrate access channels are listed in Table 1. To assess

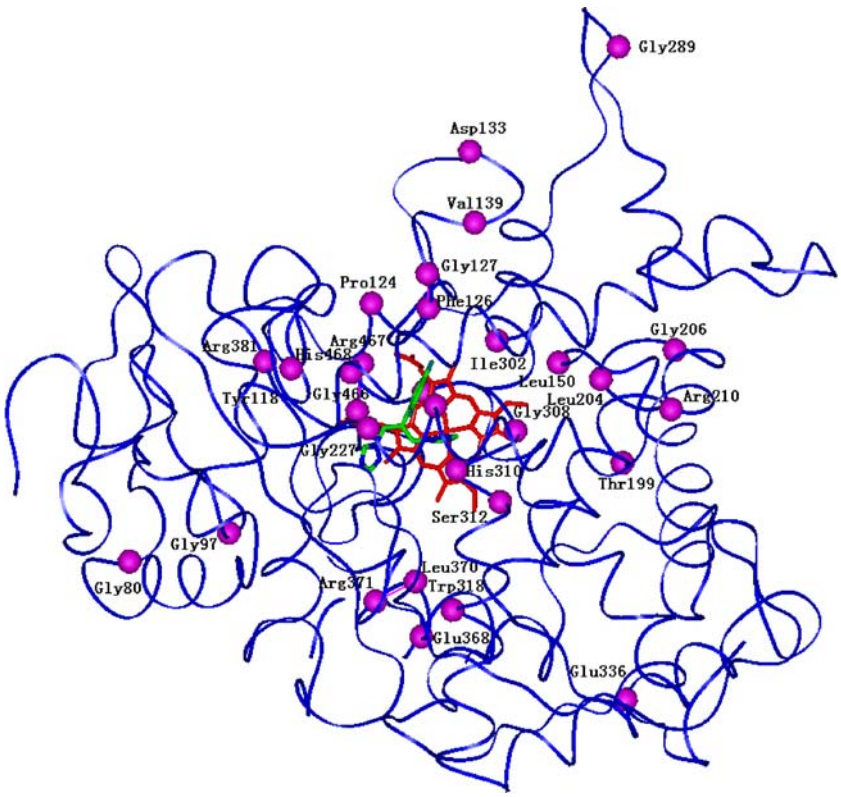

Fig. 2 The identified conserved residues (Cα atoms) mapped onto the three-dimensional structure of CACYP51. The conserved residues were depicted as a purple ball-and-stick model and CACYP51 as a blue ribbon model

Table 1 The identified trace residues in the channel 1, channel 2 and active site of CACYP51[a]

PIC	Conserved residues in the Channel 1	Conserved residues in the Channel 2	Conserved residues in the active site	Class-specific residues in the active site
30%	Tyr118 (Tyr76), Pro12 (Pro81), Phe126 (Phe83), Gly127 (Gly84), Val130 (Val87), Asp133 (Asp90)	Leu224 (Leu172), Gly227 (Gly175)	Tyr118 (Tyr76), Phe83 (Phe126), Leu204 (Leu152), Ile302 (Ile251), Gly308 (Gly257), His310 (His259), Ser312 (Ser261), Arg381(Arg326)	Leu121 (Phe78), Ala114 (Gln72)
40%[b]	-	-	Gln142(Gap), Leu305(Met254)	Thr122 (Met79), Thr123 (Thr80), Val125 (Ile82), Lys143 (Gap), Lys144 (Met99), Phe233 (Tyr181), Met306 (Phe255), His377 (Ile322), Ser378 (Ile333)
50%	-		Gln309(His258)	Gly307(Ala256), Thr311(Thr260)
60%	-	-	-	Leu139(Arg96), Gly303(Ser252)
70%	-	Leu220 (Leu168), Arg246 (Arg194), Asp247 (Asp195), Ala249 (Ala197), Arg264 (Arg212)	-	-
80%	-	-	Asp225(Glu73)	-
90%	-	-	Pro375(Pro320), Pro406(Pro349)	-

[a] The important functional residues that are confirmed by site-directed mutagenesis are underlined. The residues in the brackets mean corresponding residues in MTCYP51

[b] For the PIC values 40% to 90%, only newly identified trace residues are shown. The symbol "-" means no new trace residue is identified at this PIC level

the functional importance of these trace residues, we reviewed the published records of their mutations. The available mutagenesis data of MTCYP51 have shown that there are four important functional residues (Tyr76, Phe83, Gly84 and Asp90) [14] lining channel 1 and four important functional residues (Leu172, Gly175, Arg194 and Asp195) [14] lining channel 2. Among those residues, Phe83, Asp90, Leu172 and Arg194 may form direct contacts with the substrate [14]. Tyr76 and Asp195 might maintain proper conformation of heme and Arg194, respectively [14]. The two glycines (Gly84 and Gly175) could control conformational flexibility of the secondary structural elements upon formation of the enzyme/substrate complex [14]. At PIC level of 30%, Tyr76, Phe83, Gly84, Asp90, Leu172 and Gly175 were successfully identified as conserved trace residues. Arg194 and Asp195 were also identified as conserved trace residues at the PIC level of 70%. These interesting results demonstrate the powerfulness of the ET method in identifying the important functional residues of protein. After mapping the conserved trace residues onto the structure of MTCYP51 and CACYP51, we can find that Try76, Phe83, Asp90, Leu172 and Arg194 of MTCYP51 (corresponding to Try118, Phe126, Asp133, Leu224 and Arg246 in CACYP51) also become a portion of ceiling of the active site and they might form a direct interaction with the substrate or inhibitor.

Trace residues in cysteine pocket

In the crystal structure of MTCYP51, the cysteine pocket lies on the opposite side of the heme and cannot participate directly in inhibitor binding. Cys394 of MTCYP51 (corresponding to Cys470 in CACYP51) provides the axial thiolate ligand for the heme iron atom. At the PIC level of 30%, Gly464, Gly466, Arg467, His468 and Cys470 in the cysteine pocket of CACYP51 were identified as conserved trace residues. The site-directed mutagenesis of CACYP51 revealed that the substitutions G464S [15] and R467K [16] were associated with the resistance of fluconazole to *Candida albicans*. Although these functional residues could not form direct interaction with the substrate or azole inhibitors, they might be involved in interdomain conformational changes upon the inhibitor or substrate binding and alter the heme environment.

Trace residues in the active site

On the basis of the docking result of fluconazole with CACYP51 [8], the active site was defined as all the residues within the 8 Å around fluconazole. The identified trace residues in the active site of CACYP51 at different PIC levels were listed in Table 1. Although mutants that confer resistance to fluconazole in *Candida albicans*, such as F105L, E266D, K287R, G450E, and V488I, have been

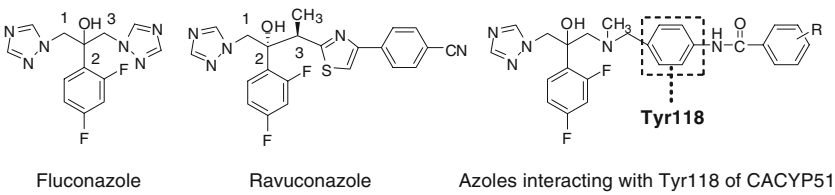

Fig. 3 Chemical structure of fluconazole, ravuconazole and our synthesized azoles

reported [17], none of them is located in the active site and could not form direct interaction with the substrate or inhibitor. In the active site, only mutations of Phe145 [18] and Thr315 [19] were reported to reduce the affinity of fluconazole with CACYP51. However, Phe145 and Thr315 are located at the edge of the active site and they cannot form direct interaction with fluconazole. It was proposed that Phe145 and Thr315 could maintain the proper conformation of the heme environment [18, 19]. Thr315 was identified as a class-specific residue at the PIC level of 80% in the present ET analysis. The identified trace residues in the active site can provide useful information to design CYP51 mutations and explore the substrate or drug binding.

Validation of the model by site-directed mutagenesis

In order to validate the present ET model, Tyr118, His310 and Ser378 were chosen for site-directed mutagenesis study because they were postulated to interact with the azole inhibitors in our previous docking studies [8, 10]. Tyr118 and His310 are identified as conserved trace residues. From the 3D model of CACYP51, they might not be involved directly in the catalytic reaction of CYP51, but they are important for the maintenance of the architecture of the active site. On the other hand, Ser378 is a specific trace residue of fungal CYP51 and is less important for the structure and function of CACYP51. It is often used as a hydrogen-bonding site in inhibitor design. Wild type and mutants of Tyr118, His310 and Ser378 from CACYP51 were constructed and heterologously expressed in *Saccharomyces cerevisiae* with deletion of the CYP51 gene [20]. The catalytic activities of wild type and its mutants were assayed by determining the ability of the enzymes to convert lanosterol to its C-14α demethylated product, 4,4-dimethyl-5α-cholesta-8,14,24-trien -3β-ol. The catalytic activity of wild type CACYP51 is 11.6 nmol/min/mg. Introduction of Y118A, Y118F, and Y118T substitution into CACYP51 considerably reduced the catalytic activity to 6.25, 6.54, and 5.82 nmol/min/mg protein, respectively. For His310 mutants, the catalytic activity of H310A and H310R is also reduced to 6.74 and 6.20 nmol/min/mg protein, respectively. As compared to conserved trace residues Tyr118 and His310, the influence of class-specific residue mutants S378A and S378T on the catalytic activity is not so obvious. The catalytic activity of H310A and H310R is 8.22 and 8.05 nmol/min/mg protein, respectively. These results are consistent with the predictions from the ET model. Moreover, the mutation changes also decreased azole susceptibility. For example, the catalytic activity of Tyr118 mutants was no more than one-half of that of the wild type [20]. Moreover, Tyr118 mutants show different affinities with different azoles. The IC_{50} values of Fluconazole to Tyr118 mutations are about 0.17 μM to 0.20 μM, while the IC_{50} values of our azole derivatives to mutation type enzymes were increased to 2.4 μM~3.4 μM. Comparing the IC_{50} values of our azole derivatives [8] and fluconazole to mutation enzyme, our azole derivatives exhibited negative inhibitory activity with a 15–20 fold decrease. These results further supported our modeling results [8] that conserved trace residue Tyr118 could form a π-π interaction with the phenyl group of our synthesized compounds (Fig. 3), while Tyr118 forms no direct interaction with fluconazole.

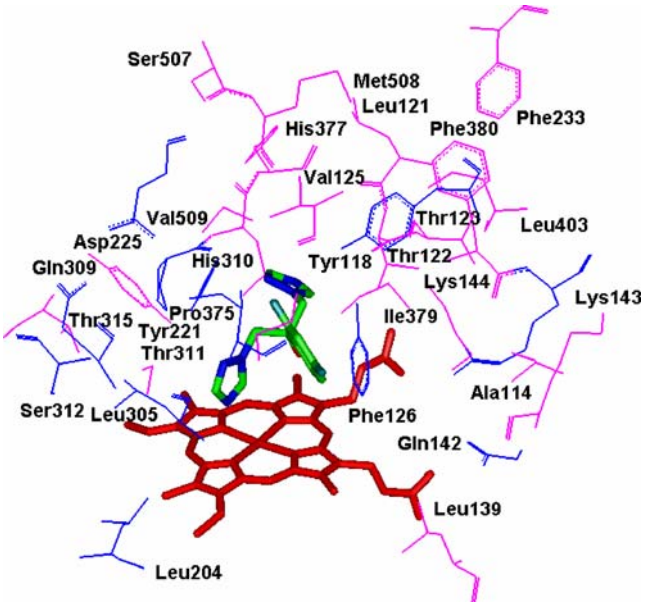

Fig. 4 Binding mode of fluconazole in the active site of CACYP51. Conserved residues (blue) and class-specific residues (purple) within 8Å in the active site of CACYP51 are displayed

Conclusions

In summary, multiple sequence alignments were performed on the CYP51 family and an evolutionary tree of CYP51 family was constructed. The ability of ET analysis to identify residues of functional importance is validated by published site-directed mutagenesis results of CYP51 members. Thus, this study may provide useful information to the design of CYP51 mutations, especially those in the active site of fungal CYP51s. In order to validate the trace residues interacting with the azole inhibitors, Tyr118, His310, and Ser378 were chosen for site-directed mutagenesis and reduced catalytic activity of these mutants were observed. On the other hand, the identified trace residues in the active site of CACYP51 should be especially useful for the design of novel antifungal agents. Mapping the identified trace residues onto the active site of CACYP51 reveals three potential binding sites for the inhibitor design (Fig. 4). First, a hydrophobic pocket defined by Phe126, Leu204, Leu305, Ile302, Leu139, and Met306 could form strong hydrophobic interactions with the inhibitors. Second, Gln309, His310, Ser312, and Thr311 form a hydrophilic hydrogen-bonding pocket for the inhibitor design. Third, most of the class-specific residues form a pocket, which was found to be important for the high selectivity and affinity of azole antifungal agents. In this pocket, the only conserved residue Tyr118 may be a potential site for the π-π interaction with the inhibitor. Class-specific residues, such as Phe233, Phe380 and Met508, could form hydrophobic and *van der Waals* interactions with the inhibitors. The above observation may provide useful information for the design of highly potent inhibitors.

Acknowledgments This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 30430750), the Shanghai "Phosphor" Science Foundation (Grant Nos. 09QA1407000) and Shanghai Leading Academic Discipline Project (Project Nos. B906).

References

1. Aoyama Y, Yoshida Y, Sato R (1984) J Biol Chem 259:1661–1666
2. Sheehan DJ, Hitchcock CA, Sibley CM (1999) Clin Microbiol Rev 12:40–79
3. Eddine AN, von Kries JP, Podust MV, Warrier T, Kaufmann SH, Podust LM (2008) J Biol Chem 283:15152–15159
4. Podust LM, Poulos TL, Waterman MR (2001) Proc Natl Acad Sci USA 98:3068–3073
5. Podust LM, von Kries JP, Eddine AN, Kim Y, Yermalitskaya LV, Kuehne R, Ouellet H, Warrier T, Altekoster M, Lee JS, Rademann J, Oschkinat H, Kaufmann SH, Waterman MR (2007) Antimicrob Agents Chemother 51:3915–3923
6. Podust LM, Yermalitskaya LV, Lepesheva GI, Podust VN, Dalmasso EA, Waterman MR (2004) Structure 12:1937–1945
7. Ji H, Zhang W, Zhou Y, Zhang M, Zhu J, Song Y, Lu J (2000) J Med Chem 43:2493–2505
8. Sheng C, Zhang W, Ji H, Zhang M, Song Y, Xu H, Zhu J, Miao Z, Jiang Q, Yao J, Zhou Y, Lu J (2006) J Med Chem 49:2512–2525
9. Sheng C, Zhang W, Zhang M, Song Y, Ji H, Zhu J, Yao J, Yu J, Yang S, Zhou Y, Lu J (2004) J Biomol Struct Dyn 22:91–99
10. Ji H, Zhang W, Zhang M, Kudo M, Aoyama Y, Yoshida Y, Sheng C, Song Y, Yang S, Zhou Y, Lu J, Zhu J (2003) J Med Chem 46:474–485
11. Lichtarge O, Bourne HR, Cohen FE (1996) J Mol Biol 257:342–358
12. Insight II 2000: Accelrys Inc, 10188 Telesis Court, Suite 100, San Diego, CA 92121. Phone: (858) 799-5000. Fax: (858) 799-5100. Website:http://www.accelrys.com/
13. Henikoff S, Henikoff JG (1992) Proc Natl Acad Sci USA 89:10915–10919
14. Lepesheva GI, Virus C, Waterman MR (2003) Biochemistry 42:9091–9101
15. Kelly SL, Lamb DC, Loeffler J, Einsele H, Kelly DE (1999) Biochem Biophys Res Commun 262:174–179
16. Lamb DC, Kelly DE, White TC, Kelly SL (2000) Antimicrob Agents Chemother 44:63–67
17. Loffler J, Kelly SL, Hebart H, Schumacher U, Lass-Florl C, Einsele H (1997) FEMS Microbiol Lett 151:263–268
18. Bellamine A, Lepesheva GI, Waterman MR (2004) J Lipid Res 45:2000–2007
19. Lamb DC, Kelly DE, Schunck WH, Shyadehi AZ, Akhtar M, Lowe DJ, Baldwin BC, Kelly SL (1997) J Biol Chem 272:5682–5688
20. Chen SH, Sheng CQ, Xu XH, Jiang YY, Zhang WN, He C (2007) Biol Pharm Bull 30:1246–1253