

Review

Molecular modeling as a powerful technique for understanding small–large molecules interactions[☆]

Maurizio Botta *, Federico Corelli, Fabrizio Manetti, Andrea Tafi

Dipartimento Farmaco Chimico Tecnologico, Università degli Studi di Siena, Via Aldo Moro, 53100 Siena, Italy

Received 27 September 2001; accepted 13 October 2001

Abstract

In the present review we summarize recent work, aimed at a better understanding of the interactions in macromolecule–ligand complexes, performed by means of computational tools such as pseudoreceptor generation, molecular docking, conformational search and energy minimization. While the first approach has been applied when the three-dimensional structural properties of the biological target were unknown, the remaining protocols exploited the knowledge of the overall structure of the involved macromolecules and their active sites. Molecular modeling techniques were used in the cases reported to study and propose macromolecular binding sites and to predict their interactions with bioactive conformers of the ligands. © 2002 Éditions scientifiques et médicales Elsevier SAS. All rights reserved.

Keywords: Anti-HIV-1; Enzyme activity; Macromolecule–small molecule interactions; Molecular modeling; Pseudoreceptor generation

1. Introduction

Many biological processes, such as the pharmacological action of drugs, the enzymatic transformations, the signal transduction and so on, repose on the interactions of macromolecular receptors and their small-molecule ligands.

During the last few decades, multidimensional NMR spectroscopy and X-ray crystallography have led to an exponential increase in the number of solved protein structures. In addition, dramatic improvements in computational tools occurred: new computational protocols for ligand docking, conformational analysis and energy minimization have been described.

In the present paper, based on the knowledge of the three dimensional structural properties of the enzyme catalytic site and the detailed reaction mechanism of the enzyme-catalyzed reaction, molecular docking, conformational search and energy minimization protocols led to the prediction of enzyme/ligand interactions and,

thus, to the rationalization of the substrate specificity and enantioselectivity for *Candida rugosa* (CRL) and *Pseudomonas cepacia* (PCL) lipases. In particular, molecular modeling has been used to predict the ability of the enzymes to discriminate between the enantiomers of a series of substrates, and has also been applied for guidance on redesigning CRL by site-directed mutagenesis in order to test the importance of a couple of amino acids selected on the basis of the computational results.

In a different study, molecular modeling techniques have been applied to allow the identification of a possible binding model for a class of suramin-related anti-HIV-1 inhibitors (Suradistas) on the HIV-1 cellular receptor (CD4) surface, without detailed structural information about the binding site for the ligand on CD4.

In a third application, finally, in the absence of any experimental structure (NMR or X-ray) of the macromolecular biological target (*Candida albicans* lanosterol 14 α -demethylase), a model has been deduced from the ligands that bind to it, by means of pseudoreceptor modeling. This approach allowed the construction of an active-site surrogate for the structural uncharacterized macromolecule, based on the structure of known ligand molecules and required the knowledge of highly probable bioactive conformations for some ligands.

[☆] This work was presented as a lecture at the IX Meeting “Strutture Eterocicliche nella Ricerca Farmaceutica”, Palermo, Italy, May 14–17, 2000.

* Corresponding author.

E-mail address: botta@unisi.it (M. Botta).

2. Experimental procedures

All calculations and graphical manipulations were performed on Silicon Graphics workstations using the software packages MacroModel/BatchMin [1] (equipped with the AMBER* all-atom and united-atom force fields), InsightII/Discover (version 95.0) [2] and PrGen [3].

The macromolecular models used in the studies here reported were derived from the coordinates of the structures labeled 1lpm [4], 1lps [5], and 2cd4 [6,7] in the Brookhaven Protein Data Bank (PDB), which represent the complexes CRL-(1*R*)-menthyl hexylphosphonate, CRL-(1*S*)-menthyl hexylphosphonate and the N-terminal fragment of CD4, respectively. In the case of PCL, the atomic coordinates of the enzyme were derived from the X-ray structure of PCL complexed with the inhibitor DEP kindly provided by M. Cygler as a PDB file [8].

To create initial coordinates for the docking studies, all water molecules of the involved crystal structure were removed and excluded from the calculations. Hydrogen atoms were added in their idealized positions. In particular, the pH used for the kinetic measurements was simulated for CRLs by means of InsightII Biopolymer module [2].

In the present studies, the computational step corresponding to the molecular docking of ligands to enzymes was simplified, since the functional groups of the substrates interacting with the macromolecule and the amino acid residues of the catalytic machinery were well known in advance. Thus, the docking problem was essentially reduced to locate the possible conformations of the ligand within the active site of its target enzyme. Further, DEP and the phosphonyl group in the X-ray crystal structures were used to guide the building of the tetrahedral oxyanion intermediates providing a good template to adjust torsional angles of the substrate. The hydrogen atom of the catalytic serine hydroxyl group was then transferred to the sp^2 $N\epsilon_2$ nitrogen atom of the imidazole ring of the catalytic histidine, rendering it positively charged. The oxyanion generated by attack of serine was directed toward the oxyanion hole, and the oxygens of the leaving alcohol and serine side-chain were positioned so that all the catalytically essential hydrogen bonds were formed.

In the case of Suradista derivatives, complexes between CD4 and ligands **S2** were built in the following manner. One of the preferred conformers of each compound, showing a symmetrical shape with respect to the ureido carbonyl group, was interactively docked into the C'C'D strands of the CD4 molecule so that π -stacking involving naphthyl ring of the ligand and aromatic side-chain of Phe43 was created. In addition, the basic side-chain of an exposed arginine at position 59 (that lies in the sequence between C and D strands) con-

tracted some salt-bridges with sulfonate group at 3-position of the ligand. Unfavorable contacts were removed using the translation–rotation feature.

Starting from CD4-**S2** complexes, the ureido moiety of the ligands was opportunely modified and a third side-chain was added to build the complexes between CD4 and compounds **S1**.

The BatchMin Monte Carlo Multiple Minimum methodology (MCOMM command) was chosen to carry out simulations, each one involving a few thousands steps, in which the studied compounds were subjected to flexible docking. BatchMin command MOLS permitted performance of rotations and translations of the docked structures. At the same time the compounds were subjected to statistical conformational searches inside the active site of their respective targets by random changes in the torsion angles of their rotatable bonds as given by the TORS command. Values of 30° and 120° , respectively, were chosen as the minimum and maximum angular increment to be added or subtracted from the current dihedral angles at each Monte Carlo step. In the case of enzymes, the side-chain of the catalytic serine residues were rotated as well in the course of the simulations with random increments in the range of 30 – 120° . Energy minimizations of the output structures were carried. BatchMin least-square superimposition routine (COMP command) was selected to eliminate duplicate minima and the global chirality checking command (CHIG) was used to reject any structure whose chirality had been changed by the minimization.

Because of the large number of atoms in the models, to correctly optimize the complexes of the ligands and the macromolecule, the following constraints had to be imposed: (a) a subset, centered on the ligand and comprising only the substrate and a shell of residues surrounding the binding site of the receptor within a radius of 8 Å (12 Å in the case of CD4) from the ligand, was created and subjected to energy minimization. The substrate and all amino acid side-chains of the shell were unconstrained during energy minimization to allow for reorientation and thus proper hydrogen-bonding geometries and vdW contacts; (b) all the atoms external to the subset remained fixed, even if their non-bond interactions with all the relaxing atoms have been calculated.

Two intramolecular distances were monitored during the docking experiments involving CRL in order to ensure that the relative orientations of the residues of the catalytic triad remained compatible with the postulated mechanism of the hydrolytic reaction. These were the distances between the hydrogen atom bound to the $N\epsilon_2$ of the imidazole ring of the catalytic histidine and both the O_γ oxygen atom of the catalytic serine and the oxygen atom of the leaving alcoholic group.

Energy minimizations of the complexes were performed using the Polak–Ribière conjugate gradient method until the derivative convergence was 0.01 kcal/Åmol. The default non-bonded cutoff protocol employed by the program was modified. Thus, a vdW cutoff of 12.0 Å, an electrostatic cutoff of 20 Å, and a hydrogen bonding cutoff of 2.5 Å were utilized for all calculations.

In the case of PCL, the tetrahedral intermediates were built starting from the corresponding output docking geometries: one long covalent bond [3.7 Å in the case of (*R*)-**1a**] was drawn between the Ser87 O_γ oxygen atom and the carbonyl carbon atom of the different compounds while transferring one H atom from Ser87 to His286 (deletion and drawing of one bond without modifying the position of the atom). Finally the carbonyl group of the substrates and the imidazole ring of His286 were changed into charged species (C–O[−] and imidazolium ion, respectively). A direct energy minimization was performed (after having fixed in 3D space the position of the backbone of the whole PCL) with MacroModel/BatchMin in order to optimize the new structures. Structures lacking all the catalytically essential hydrogen bonds involving the imidazole ring of histidine were considered as non-productive complexes and, accordingly, were no further investigated.

Applying the pseudoreceptor modeling software PrGen, a coupling constant of 1.0 and a maximum allowed RMS of 0.3 kcal/mol for the predicted versus experimental dissociation constants of all the correlation-coupled receptor minimizations were chosen. The target RMS deviation was limited to a maximum of 0.4 kcal/mol. Both the training set and the test set structures were relaxed inside the enzyme cavity without constraints applying 100 trials of the conjugated gradient minimizer. Solvation energies were automatically calculated by the software PrGen and the essential biological data were taken from the literature [9].

3. Results and discussion

3.1. Computer simulations of enantioselective ester hydrolyses catalyzed by *P. cepacia* and *C. rugosa* lipases

Hydrolytic enzymes are widely used in organic synthesis as environmentally friendly catalysts that possess broad substrate specificities, display high stereoselectivity, are commercially available, and do not require expensive and unstable coenzyme systems. Lipases in particular are, at present, one of the most extensively used classes of hydrolases as they are not restricted to water soluble compounds [10–15]. It is evident, for a full synthetic exploitation of lipases in organic che-

mistry, the need of a sound rationalization of the activity of these enzymes able to correlate their *substrate specificity* and *enantioselectivity* to the structural features of the processed substrates.

In recent years the crystallographic data collected for several lipases [16–24] have provided the three-dimensional structures of their active sites allowing in this way the computational study of the interactions between these enzymes and their substrates at molecular level [25–28]. In conjunction with the experimental data obtained from kinetic and SAR studies, this approach can provide the chemists with theoretically reliable models, with better possibilities to rationalize and predict the activity of lipases with respect to previously developed empirical rules [29–31].

Based on the X-ray crystal structure of lipases from *P. cepacia* (PCL) [8] and *C. rugosa* (CRL) [4,5]—complexed with their inhibitors diethyl *p*-nitrophenyl phosphate (DEP), and menthyl hexylphosphonate, respectively—a computational study was performed in order to rationalize both the enantioselectivity and substrate specificity displayed by these lipases in the enantioselective hydrolysis of racemic esters (see Table 1 [32–36] for PCL, and Table 2 for CRL).


The first step of the hydrolysis reaction of the substrates was simulated by construction and energy minimization of the theoretically involved tetrahedral intermediates for both PCL and CRL. In the case of PCL, the recognition process between the enzyme and its substrates was simulated as well by the flexible docking of some of the studied compounds inside the enzymes active sites, applying a protocol previously defined by us [25].

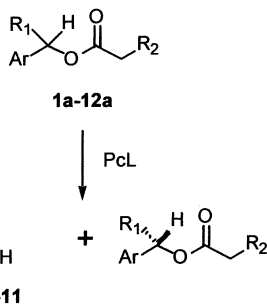
3.1.1. Calculations on PCL

The study of the molecular recognition of substrates **1a**–**12a** by PCL through docking experiments allowed the identification of different binding geometries for *R* and *S* enantiomers. Nevertheless, the docking experiments did not definitely clarify either the *enantioselectivity* nor the *substrate specificity* of the enzyme, as they also suggested that compound (*S*)-**1a** and other enantiomers with this same stereochemistry, as well as substrates not processed like compound (*R*)-**12a**, could still place their ester moiety at bond distance to SER87 and be recognized by PCL.

The finding that both *R*- and *S*- enantiomers of substrates can be recognized by a lipase has been discussed recently when the concept of transition-state model has been introduced (besides the classical concepts based on the size, shape and topology of the active site) to explain the enantiopreference of lipases in terms of the energy difference between the diastereomeric forms of the postulated transition states [37]. The stereoelectronic theory, developed by Deslongchamps [38] transforms this thermodynamic

Table 1
Hydrolysis of esters **1a–12a** catalyzed by *P. cepacia* lipase (PCL)

Substrate	Ar	R ₁	R ₂	Enantioselectivity (E)
(<i>R,S</i>)- 1a ^a	Ph	Me	H	>1000
(<i>R,S</i>)- 2a ^a	<i>p</i> -MePh	Me	H	312
(<i>R,S</i>)- 3a ^a	<i>m</i> -MePh	Me	H	>1000
(<i>R,S</i>)- 4a ^a	<i>p</i> -MeOPh	Me	H	22
(<i>R,S</i>)- 5a ^b	<i>o</i> -MePh	Me	H	not hydrolyzed
(<i>R,S</i>)- 6a ^a	Ph	CF ₃	Cl	127
(<i>R,S</i>)- 7a ^c	Ph	CCl ₃	H	not hydrolyzed
(<i>R,S</i>)- 8a ^c	Ph	CH ₂ Cl	H	145
(<i>R,S</i>)- 9a ^{d,e}	<i>o</i> -MeOPh	CN	H	45
(<i>R,S</i>)- 10a ^b		H		>1000
(<i>R,S</i>)- 11a ^a	2-naphthyl	Me	H	>1000
(<i>R,S</i>)- 12a ^b	1-naphthyl	Me	H	not hydrolyzed



^[a]Ref. [33]. ^[b]Ref. [32]. ^[c]Ref. [34]. ^[d]Ref. [35]. ^[e]The cyanohydrines formed were not isolated [35].

Table 2
Hydrolysis of Naproxen and Ketoprofen methyl esters catalyzed by CRLs

Substrates (methyl esters)	<i>C. rugosa</i> lipases	Conversion (acid mmol/protein mg) × 10 ³	Conversion (%)	e.e. _p (%) ^a	<i>E</i> ^b
Naproxen	Wild-type	7.59	32	90.91 ± 0.03	32.0
Naproxen	Phe344Val	13.70	33	90.15 ± 0.02	29.9
Naproxen	Phe344,345Val	3.79	18	73.47 ± 0.05	7.6
Naproxen	Phe345Val	0.30	9	42.19 ± 0.05	2.6
Ketoprofen	Wild-type	4.85	33	89.28 ± 0.06	27.2
Ketoprofen	Phe344Val	9.36	38	86.28 ± 0.05	23.0
Ketoprofen	Phe344,345Val	1.08	9	61.84 ± 0.01	4.5
Ketoprofen	Phe 345Val				

^a Standard deviation of six replicate injections.

^b Calculated according to Chen et al. [50].

concept into geometric constraints on the three-dimensional structure of the tetrahedral intermediates and states that, in the case of lipases, the serine attack to the substrates is effective if it generates a *gauche* conformation (with respect to the C_{alcohol}-O-C-O_{serine} dihedral angle) of the first tetrahedral intermediate [38]. Based on these concepts, the tetrahedral intermediates of esters (*R*)-**1a**, (*S*)-**1a** and (*R*)-**12a** were studied.

The study of the tetrahedral intermediates definitively rationalized the experimental enantioselectivity of PCL in the resolution of racemic esters of secondary alcohols showing that *S* enantiomers cannot be obtained (or can be obtained much slower than *R* enantiomers), since the formation of the acyl-enzyme must follow a higher activation energy pathway involving an *anti* conformation of the tetrahedral intermediates due to unfavorable steric contacts between the substrates and the enzyme. The same principles could be used to justify the substrate specificity of PCL, especially in the case of the 1-naphthyl derivative **12a**. This result is depicted in Fig. 1, where the superimposition of the calculated tetrahedral intermediates of (*R*)-**1a** (black) and (*S*)-**1a** (gray) located at their proper respective 3D position inside the active site of PCL is shown.

In conclusion, the coupled modeling of Michaelis–Menten complexes and of the corresponding transition state analogues identified a substrate-recognition model for PCL common to one enantiomeric series of substrates **1a–12a** and suggested that the major factor

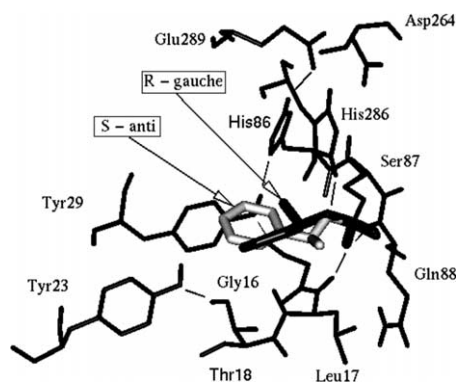


Fig. 1. Superimposition of the calculated tetrahedral intermediates (thick) of (*R*)-**1a** (black) and (*S*)-**1a** (gray) in PCL located at their proper respective 3D position inside the active site of the enzyme. Hydrogen atoms bound to heteroatoms are displayed. For sake of simplicity only residues useful for the discussion are shown. Hydrogen bonding interactions are depicted as thin lines.

controlling both the enantioselectivity and the substrate specificity of this lipase comes into play during the first step of the hydrolysis reaction leading to the formation of the tetrahedral intermediate. The stereoelectronic constraints required for an effective formation of the covalent bond between PCL and the substrates are best satisfied only by the processed compounds.

3.1.2. Calculations on CRL

Computer modeling studies of the interactions between CRL and methyl (\pm)-2-(3-benzoylphenyl)propionate (Ketoprofen methyl ester) led to the rationalization of the enzyme stereoselectivity (see Table 2) on the basis of the spatial arrangement of both the catalytic residues and substrates.

The complex between the *S* enantiomer of the substrate and the enzyme proved to be 24 kJ/mol more stable than the corresponding complex involving the *R* enantiomer. According to computational studies previously reported [39,40], the fast-reacting enantiomer appeared to be correctly predicted, although the calculated value of the energy difference did not agree quantitatively with the experimental data. In fact, the simulation indicated a much higher enantioselectivity with respect to that experimentally determined which corresponds to a difference in activation energy of about 8 kJ/mol for the enantiomers of the substrate. Considering that this kind of calculation is usually complicated by the high energy values of the enzyme–ligand complexes and their fluctuations during the simulations [41], the qualitative agreement reached by means of molecular mechanics simulations can be considered as a good result. In fact, current modeling methods are extremely useful as a qualitative tool, but quantitative prediction (for example, the degree of enantioselectivity) are still not reliable [28].

In the case of the *S* substrate, both the acyl chain and the alcoholic moiety are directed out of the active site, while the methyl group at the stereogenic center is directed toward the entrance of the hydrophobic tunnel (Fig. 2, top). On the contrary, the methyl group in the *R* enantiomer is out of the tunnel and directed toward a region defined by Gly124 and the aromatic side-chain of Phe296. As a consequence of this orientation of the methyl group at the *R* stereogenic center, a conformational rearrangement of amino acid main chains ranging from Gly122 to Gly124 occurs, which gives rise to a loss of H-bond between oxyanion and Gly123 NH, being this group 1.3 Å apart with respect to the position found in the (*S*)-enantiomer–enzyme complex. In addition, the aromatic portion of the substrate, as a result of a rotation around the C_{chiral}–C_{aromatic} bond, is forced to point toward Phe344, whose aromatic ring undergoes a rotation of about 120°. In a similar way, at the bottom of the hydrophobic pocket filled by the substrate, Phe345 aromatic ring moves toward Phe296

and Leu297 assuming a perpendicular orientation with respect to the substrate benzoyl moiety (Fig. 2, bottom) without any profitable interaction. The methyl group of the alcoholic portion of the *S* and *R* enantiomers is oriented toward Glu208 main chain, and Gly124, respectively.

A second binding mode, with the benzoyl moiety of the substrate interacting with the enzymatic hydrophobic tunnel, was also found. The *S* enantiomer–enzyme complex is 9 kJ/mol more stable than the corresponding *R* enantiomer–enzyme complex. The alcoholic moiety of the substrate is directed toward the Gly124 main chain without any interaction, and is linked by hydrogen bond to His449 N ϵ_2 .

On the basis of the molecular mechanic simulations performed, the enantiopreference for the *S* enantiomer that has emerged in this theoretical study could be attributed to the lack of a hydrogen bond between the *R* substrate and the residues of the oxyanion hole, as shown in the first binding model described.

In addition, our simulations led to the identification of two amino acids (namely, Phe344 and Phe345) mainly involved in the enzyme–substrate interactions. To test the importance of these residues to the enzymatic activity, site-directed mutagenesis of the selected amino acids has been performed accordingly, and the mutated enzymes have been evaluated for their conversion and selectivity capabilities toward different substrates. The mutations were chosen according to the following considerations: (i) A previously published paper by our group [25] highlighted that the CH group at 4-position of the Phe345 aromatic ring sterically interacts with the methyl group at the stereogenic center of the methyl (*R*)-2-(6-methoxy-2-naphthyl)propionate accounting for both the stereochemical preference and the high enantioselectivity of CRL. Substitution of Phe345 by Val removes these sterical clashes because of the shortening of the side-chain, providing favorable hydrophobic interaction between the methyl group on the chiral carbon of the substrate and the isopropyl moiety of the mutated amino acid. (ii) In the model of binding with both the acyl chain and the alcohol moiety of the substrate directed out of the active site, when the configuration of the ligand was inverted from *S* to *R*, some conformational rearrangement occurred due to unfavorable steric contacts involving both the Phe344 and Phe345 aromatic rings. From this point of view, the suggested mutations could provide some additional space useful for the *R*-substrate orientation inside the catalytic site.

In light of these considerations, both of the proposed mutations were expected to enhance the affinity of CRL toward the *R*-enantiomer, with a consequent decrease of the e.e. values.

Kinetic experiments performed using the recombinant (wild-type and mutated Phe344Val, Phe345Val,

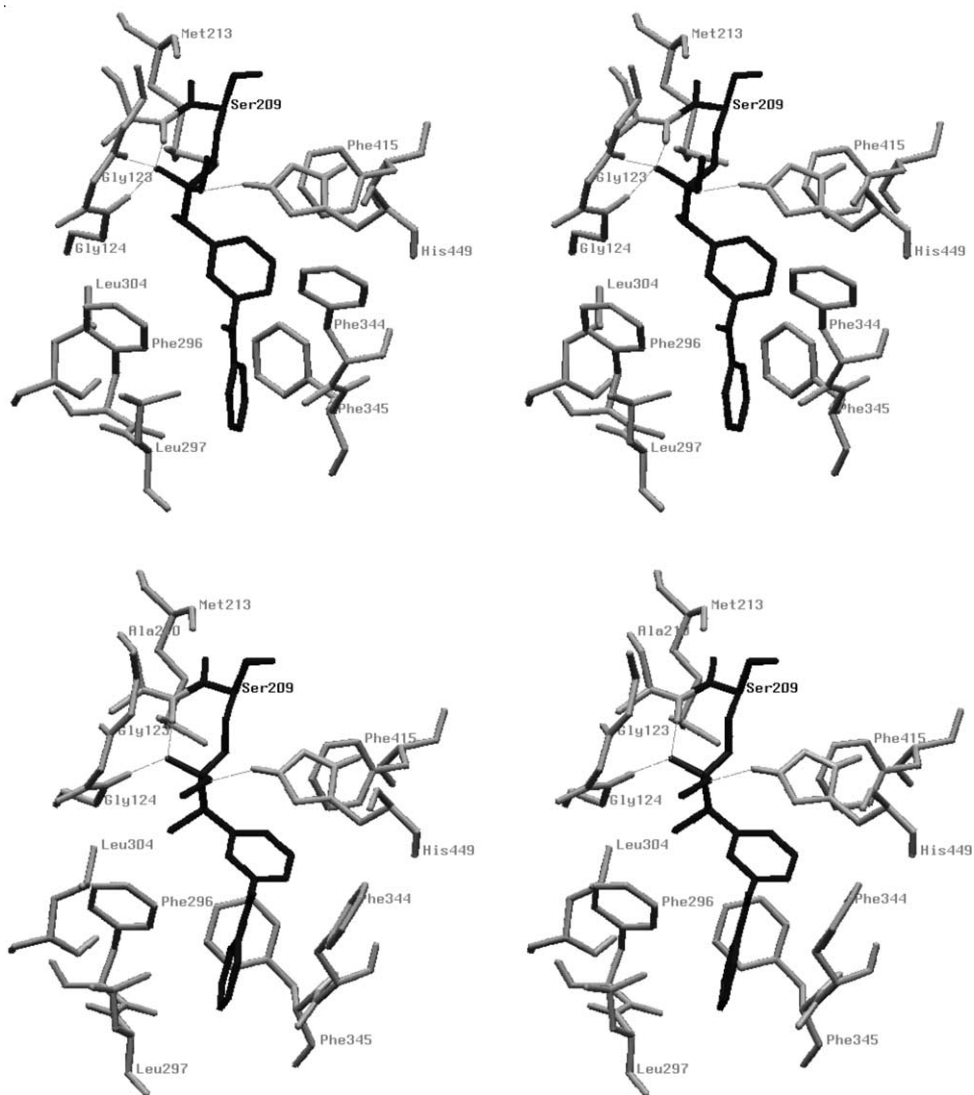


Fig. 2. Stereo representation of the CRL-substrate transition-state complexes of the fast-reacting *S*-enantiomer (top) and the slow-reacting *R*-enantiomer (bottom) of methyl 2-(3-benzoylphenyl)propionate. The tetrahedral intermediate, including Ser209, is shown in black, and the hydrogen-bonding network involving the catalytic serine and the oxyanion hole (Gly123, Gly124, and Ala210) is drawn by thin gray lines. The leaving alcoholic chain of both enantiomers points out of the active site. Top: Both the acyl and the alcoholic moieties of the *S*-enantiomer are directed out of the active site, while the methyl group at the chiral carbon atom points toward the entrance of the hydrophobic tunnel. Three hydrogen bonds between the oxyanion hole and the oxyanion were found. Bottom: The methyl group at the stereogenic center of the *R*-enantiomer is directed toward Gly124 and Phe296. Gly123 lost a hydrogen bond with the oxyanion due to a conformational rearrangement focused on the Gly122–Gly124 sequence.

and Phe344,345Val) CRLs on methyl esters of Keto-*profen* and *Naproxen* showed a decreased enantioselectivity of the hydrolysis reaction catalyzed by CRL mutants. The experimental results obtained in these biotransformation reactions indicate that Phe344 and especially Phe345 influence CRL activity, supporting their proposed interactions with substrates.

3.2. Research on anti-HIV-1 agents. Investigation on the CD4-Suradista binding mode through docking experiments

Virus infection is initiated by the attachment of a

viral particle to a specific receptor on the plasma membrane of the target cell. In the CD4-dependent pathway of entry, virus binding takes place at the cell surface and is the result of a highly specific interaction between the principal HIV cellular receptor (the CD4 molecule) and the surface envelope viral glycoprotein gp120. Subsequently, the fusion between the viral envelope and cell membranes allows the virus core to penetrate the cell cytoplasm.

The gp120 binding site on CD4 is localized on its N-terminal domain, whose structure was recently determined by X-ray crystallographic analysis [6,7].

Compounds **S1a–S8d**, structurally related to the non-toxic minor groove DNA binder distamycin (Fig. 3 and Table 3), synthesized by Pharmacia and Upjohn, have been extensively studied for their anti-HIV activity [42] and proved to be potential candidates for clinical trials.

Only compounds **S1** and **S2** exhibited antiviral activity, although monofunctional derivatives **S5–S8** displayed RT-inhibiting properties *in vitro*. Due to their polar nature, these compounds are unlikely to enter cells, and their activity has been attributed to the inhibition of viral binding. In fact, time course experiments performed by our group [42] clearly demonstrated that inhibition was at the surface of the cell and

compounds **S** were shown to interfere with attachment of virus to cells, most likely due to an effect on CD4-gp120 interactions, as well as inhibiting the fusion process between virus and cell.

In an effort to rationalize biological data of compounds **S1**, **S2**, **S5–S8**, we resorted to molecular modeling in an attempt to gain an insight into both the possible biologically active conformation(s) of these ligands and their mode of interaction with CD4 or gp120.

As a result of a conformational search on Suradistas, one of the preferred conformation of **S2b** appeared as a highly symmetrical, quasi-planar arc of circumference, as expected on the basis of the symmetrical pattern of

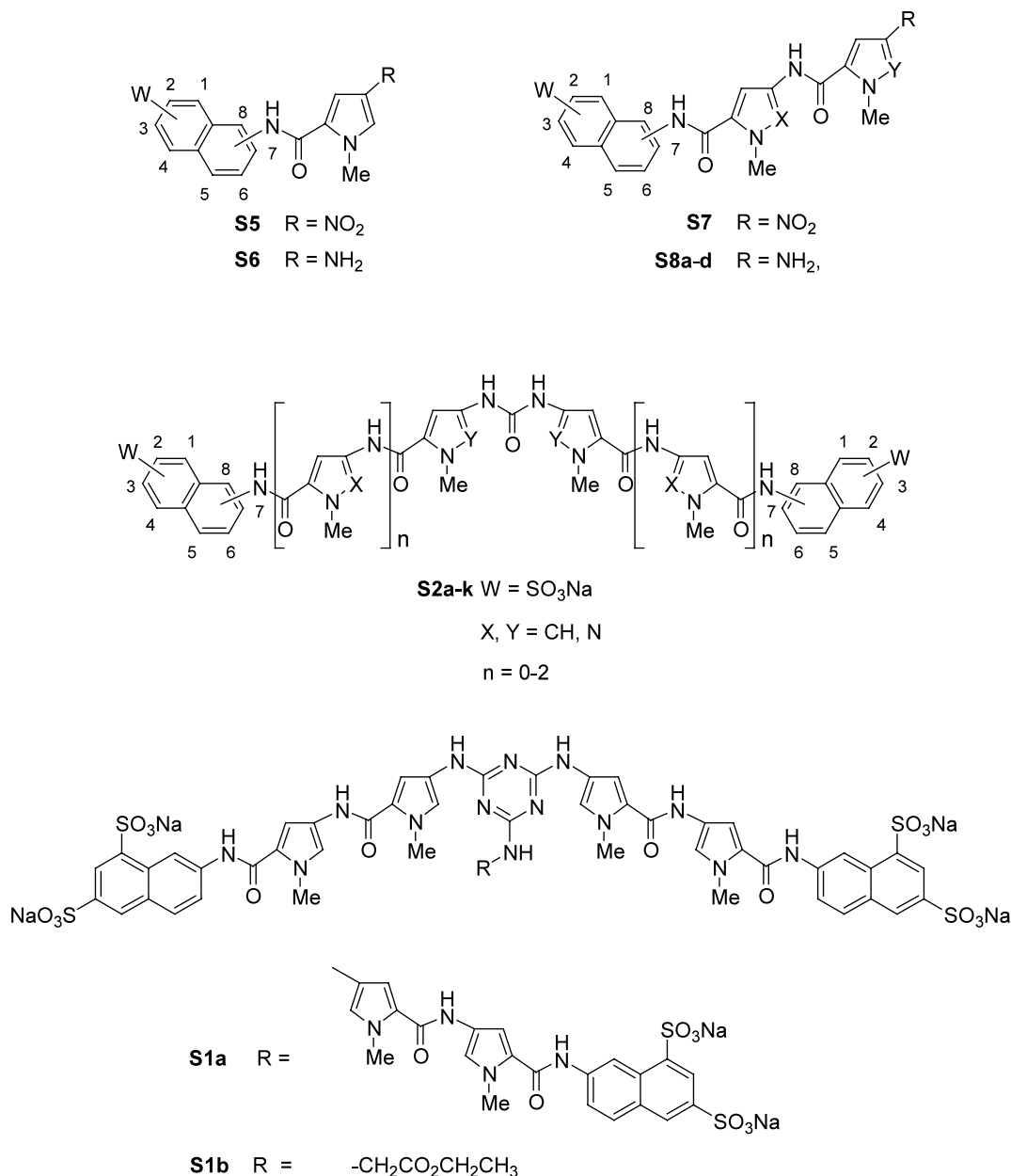


Fig. 3. Structure of compounds **S1**, **S2**, **S5–S8**.

Table 3
Chemical data and anti-HIV activities of compounds **S1**, **S2** and **S5–S8**

Comp.	NH pos.	X	Y	W	<i>n</i>	EC ₅₀ (μM) ^a	IC ₅₀ (μM) ^b
S1a						0.06	>200
S1b						3.9	>200
S2a	7	CH	CH	1,3-SO ₃ Na	1	29.6	>200
S2b	7	CH	CH	1,3,5-SO ₃ Na	1	4.6	>200
S2c	7	CH	CH	1,3-SO ₃ Na	2	1.0	86
S2d	8	CH	CH	1,3,5-SO ₃ Na	1	31.6	>200
S2e	7	CH	CH	1,3-SO ₃ Na	0	80.7	>200
S2f	8	CH	CH	3,5-SO ₃ Na	1	12.3	>200
S2g	7	CH	CH	4,8-SO ₃ Na	1	23	153
S2h	7	N	N	1,3-SO ₃ Na	1	7.5	>200
S2i	8	CH	CH	1,5-SO ₃ Na	1	>200	>200
S2j	7	CH	N	1,3-SO ₃ Na	1	7.9	>200
S2k	7	N	CH	1,3-SO ₃ Na	1	6.5	>200
S5	8			3,5-SO ₃ Na		>200	>500
S6	8			3,5-SO ₃ Na		>200	>300
S7	7	CH	CH	1,3,5-SO ₃ Na		256	>200
S8a	7	CH	CH	1,3,5-SO ₃ Na		155.7	>200
S8b	8	CH	CH	3,5-SO ₃ Na		34.9	313
S8c	7	N	CH	1,3-SO ₃ Na		19.5	>200
S8d	7	CH	N	1,3-SO ₃ Na		>200	>200

^{a, b} All compounds were evaluated from a high test concentration of 200 μM for their 50% effective antiviral concentration, EC₅₀ (reported as the concentration of drug required to inhibit 50% of virus induced cell killing), and their concentration that caused 50% cell death, IC₅₀ (reported as the concentration of drug required to reduce cell viability by 50%), utilizing the XTT cytoprotection assay for each congener.

its NMR spectrum [43,44]. In this conformer, the distance between the sulfonate groups at the position 3 of the two naphthalene units is 24–26 Å, which is in full agreement with the 25 Å extension of the ridge containing the C'' strand [6]. These observations highlighted a possible good shape complementarity between conformations of the ligands and the ridge on CD4 macromolecule leading us to choose the symmetrical, planar structure as the putative bioactive conformation of compound **S2b**.

When compound **S2b** in its symmetrical conformation was docked into the X-ray structure of CD4 (Fig. 4), very good interactions between the naphthalenesulfonate groups of the ligand and two distinct regions (hereafter referred to as subsite 1 and subsite 2, respectively) on CD4 surface were found. Interestingly, all the amino acid residues involved in the CD4–ligand interactions lie in the sequence 38–59, which is thought to be a fundamental cluster for CD4–gp120 interactions [6,7].

In order to further test this binding model and obtain an insight into the possible relationship between the structural features of CD4–**S1** and –**S2** complexes and the biological activity of these compounds, the same computational study was performed starting from the other ligands.

Examination of the complexes obtained from each of compound **S2a**, **S2c** and **S2e** quite well accounted for their different biological activity.

In Fig. 5 is depicted the case of complexes between CD4 and compound **S1a**. Two of the three side-chains

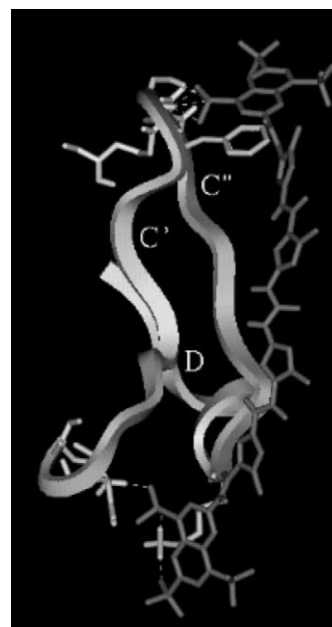


Fig. 4. CD4–**S2b** complex. Ribbon representation of sequence 30–57 including C', C'' and D strands of CD4, and ligand **S2b**. Side-chains of Ser42, Phe43, and Arg59 are depicted at the top of the figure. Pi-interactions between naphthalene ring of the ligand and Phe43 side-chain are evident. At the bottom, side-chains of Asn30, Asn32, and Lys50 are shown. Dashed lines represent electrostatic interactions (hydrogen bonds and salt bridges) between CD4 amino acids and sulfonic moieties of **S2b**. The direction of the view is from the GFCC' face [6].



Fig. 5. CD4–**S1a** complex. Two side-chains of compound **S1a** projecting its naphthalene moieties toward the top and the bottom of the figure, are nearly parallel to C'' strand (on the right side of the figure). The third side-chain of the ligand binds in a large groove whose walls are represented by FG (on the left side of the figure) and CC' (on the middle of the figure) corners of CD4. The most important amino acid side-chains of the three binding sites on CD4 are also shown. Triazine moiety of the ligand overlaps C'' strand. Dashed lines represent hydrogen bonds and salt bridges between CD4 and the ligand. Ribbon diagram is extended to the complete D1 domain of CD4. The direction of the view is from the GFCC' C'' face of CD4 [6].

of the ligand assume a conformational disposition very similar to that of compounds **S2**. Both compounds **S1** and **S2** are nearly parallel to the C'' strand, while the third side-chain of compound **S1a** binds in a C-shaped portion of CD4 surface, very similar to a large groove. Both CC' and FG corners (residues 29–35 and 87–90, respectively) as well as short amino acid sequences of C, C' , F, and G strands constitute the wall and the floor of the groove, respectively. Strands C and C' are near to the triazine moiety of compound **S1a**, while strands F and G are far from it. The most significant interactions between CD4 and compound **S1a** are contacts between the above described groove and the third side-chain bound to the triazine moiety. The naphthyl ring is directed toward a hydrophilic region of CD4 with some positively charged amino acids (subsite 3).

The findings reported above may allow some conclusions to be drawn. Relationships between structural features of compounds **S1**–**S2** and binding affinity to CD4 can be gathered from the analysis of CD4–ligand complexes. The proposed model is able to rationalize the effect of the distance between the naphthalene rings on the observed activity for compounds **S2**. Particularly, compound **S2a** seems to be the optimal ligand for CD4, showing the proper distance of 24–26 Å between

sulfonate groups on the two naphthalene rings. Due to the higher distance between the naphthalene moieties, **S2c** would be at least as good as **S2a** as a CD4 ligand. This idea is in agreement with the experimental evidence. Finally, on the basis of the proposed binding mode, compound **S2e** seems to have structural features avoiding a profitable interaction with CD4 surface.

The marked increase in the activity of **S1a** can be explained taking into consideration the interaction of the third naphthylsulfonate moiety with a third accessory binding site (subsite 3).

In light of the biological and computational results reported above, we infer that compounds **1**, **2** might interact with CD4 by the attachment of their naphthylsulfonic moieties to three different hypothetical binding sites. In spite of the lack of activity of ligands **S5**–**S8** possessing only one naphthylsulfonic moiety, to justify the biological data of compounds **S8b** and **S8c**, we cannot exclude a model of interaction which implies monodentate binding of these ligands to the CD4 surface, though we are not able to identify the most probable binding site.

Very recently, after the completion of this work, the structure of an HIV gp120 envelope glycoprotein complexed with the CD4 receptor and a neutralizing human antibody (ternary complex) was reported in the literature [45]. The ternary structure was solved by X-ray crystallography at 2.5 Å resolution and shows that residues of CD4 in contact with gp120 are concentrated in the span from 25 to 64 with Phe43 and Arg59 of CD4 establishing multiple contacts. Moreover, three CD4 lysine residues are implicated in binding (residues 29, 35 and 46). Accordingly, the amino acids of CD4 involved in the CD4–Suradista and ternary complexes are basically the same. Both the experimentally and computationally obtained complexes agree in highlighting the fundamental role played by Phe43 in interacting with compounds **S1**, **S2** and gp120 sequences, respectively.

3.3. Pseudoreceptor modeling of azole antifungal agents active against *Candida albicans* lanosterol 14 α -demethylase

The knowledge of the three-dimensional structure of the active site of enzymes can primarily help to discover structure–activity relationships and to get some information about the binding mode of active inhibitors by docking experiments (receptor fitting). A pure indirect approach also, which focuses mainly on structural and physico-chemical properties of active and inactive compounds, can identify structure–activity relationships helpful to design new, more specific drug candidates and eventually to understand the topography of uncharacterized enzyme active sites. This approach is typically termed receptor mapping.

The pseudoreceptor modeling technique is a new approach to computer aided drug design based on comparative analysis of active and inactive compounds [3]. It is aimed at generating a hypothetical three-dimensional peptide model of the binding site of a structurally unknown target protein, which is expected to predict binding affinities and geometries of ligands. It combines the two philosophies of receptor mapping and fitting as it provides atomistic models (receptor surrogates) of active sites that might help an easier design of new drug candidates, also giving an insight into key intermolecular interactions. Results from experimental and even theoretical structural investigations of the target protein can be used in this technique to tailor the pseudoreceptor generation protocol.

Cytochrome P-450-dependent lanosterol 14 α -demethylase (P-450_{DM}) is an enzyme that catalyzes the first step of the conversion of lanosterol to cholesterol (mammals) or ergosterol (fungi), by causing the removal of the 14 α -methyl group of lanosterol to give the Δ 14-15-desaturated sterol [9]. Under normal conditions, azole (N-substituted imidazole and triazole) antibiotics selectively inhibit the fungal P-450_{DM}, preventing the binding of the natural substrate lanosterol to P-450_{DM} by coordination of the ring nitrogen atom (N3 of imidazole and N4 of triazole) to the sixth coordination position of the iron atom of the enzyme protoporphyrin system. The accumulation of 14 α -methylated sterols in azole-treated fungal cells affects membrane structure and functions resulting in an inhibition of the growth of fungi [9].

The search for antifungal agents that show increased specificity towards pathogenic fungal enzymes remains a primary target in medicinal chemistry research. To

this end the azole antifungal agents **t1a–t4m** (Chart 1) structurally related to bifonazole (**t1a**), already investigated by using CoMFA (a 3D-QSAR computational technique) to develop a model able to explain and predict the activity of bifonazole-like inhibitors [9], were considered for a new structure-based drug design study. Promising results were already obtained with the CoMFA approach, which successfully explained structure–activity relationships. Moreover, the alignment method utilized to derive the CoMFA model gave an insight into a possible binding mode between the inhibitors and the protoporphyrin system of the target enzyme. The CoMFA model, however, has been recently found not to accurately predict the biological activity of azole active antifungal agents not used in deriving the model itself [46].

The pseudoreceptor modeling technique has been applied in this study, aimed at the construction of a peptide binding-site model for the structurally uncharacterized P-450_{DM}. The primary goal of this research was to bring up to date the previously developed CoMFA model based on bifonazole, by predicting the activity of new azole *Candida albicans* P-450_{DM} inhibitors [46].

The methodology of pseudoreceptor modeling is well established and has been recently reviewed in a detailed manner. It consists of five distinct steps: generation of ligand alignment, identification of receptor nucleation sites (potential anchor points for receptor residues in space), construction of the pseudoreceptor, energetic equilibration, and validation–pseudoreceptor analysis [47].

The three-dimensional coordinates of the training set ligand molecules, superimposed in their bioactive

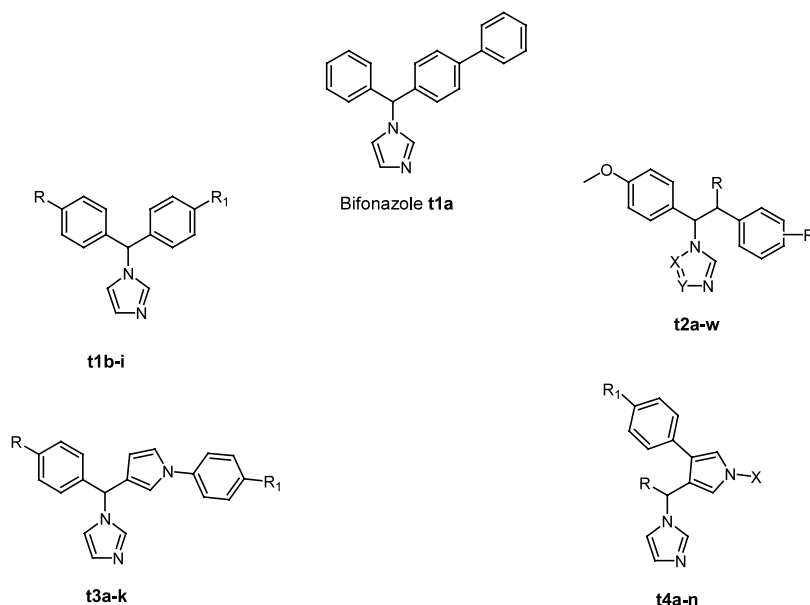


Chart 1. Azole antifungal agents investigated (see Table 4).

Table 4
Substitution pattern and antimycotic activities of azole antifungal agents investigated

Comp.	R	R ₁	X	Y	$-\log \frac{\text{MIC}_{90}(\text{comp})}{\text{MIC}_{90}(\text{bifon})}$
<i>Training set</i>					
t1a					0.00
t1b	H	1-pyrrolyl			0.60
t1c	H	NO ₂			−0.30
t1d	CH ₃	1-pyrrolyl			0.90
t1f	Cl	NO ₂			0.00
t1g	CH ₃	NO ₂			0.30
t2a	H	4-Cl	CH	CH	−0.60
t2b	H	2,4-Cl ₂	CH	CH	−0.60
t2c	H	4-NH ₂	CH	CH	−0.30
t2d	H	3-(1-pyrrolyl)	CH	CH	−1.80
t2e	H	4-Cl	CH	N	−1.80
t2f	H	2-Cl	CH	CH	−1.80
t2g	H	3-F	CH	CH	−1.80
t2h	H	2-NO ₂	CH	CH	−2.10
t2i	H	4-OCH ₃	CH	N	−1.80
t2j	H	2-(1-pyrrolyl)	CH	CH	−0.90
t2k	H	4-NO ₂	CH	CH	−1.50
t2l	H	4-OCH ₃	CH	CH	−0.90
t2m	H	4-Cl	N	CH	−1.50
t2n	C ₂ H ₅	4-NO ₂	CH	CH	−0.90
t2o	2,4-Cl ₂ -Bn	4-OCH ₃	CH	CH	−1.50
t2p	2,4-Cl ₂ -Bn	4-Cl	N	CH	−1.80
t3a	H	H			−0.60
t3b	CH ₃	H			0.00
t3c	Cl	Cl			−0.60
t3d	F	Cl			−0.30
t3e	H	Cl			0.30
t3f	CH ₃	F			−0.30
t3g	Cl	F			−0.30
t3h	H	F			−0.60
t4a	Ph	Cl	H		0.30
t4b	Ph	Cl	CH ₃		0.00
t4c	Ph	F	H		0.00
t4d	Ph	OCH ₃	2,4-Cl ₂ -Bn		−1.80
t4e	Ph	H	2,4-Cl ₂ -Bn		−1.50
t4f	Ph	Cl	2,4-Cl ₂ -Bn		−1.80
t4g	1-naphthyl	H	H		−0.90
t4h	2-naphthyl	H	H		−0.90
t4i	2-naphthyl	H	CH ₃		−2.40
<i>Test set</i>					
t1h	Cl	1-pyrrolyl			0.30
t1i	H	NH ₂			−0.30
t2q	H	3-Cl	CH	CH	−1.50
t2s	H	4-F	CH	CH	−1.20
t2t	C ₂ H ₅	4-Cl	N	CH	−1.50
t2u	C ₂ H ₅	4-Cl	CH	N	−1.80
t2v	2,4-Cl ₂ -Bn	4-Cl	CH	CH	−1.20
t3i	F	H			−0.30
t3j	CH ₃	Cl			0.00
t3k	F	F			0.00
t4j	Ph	CH ₃	CH ₃		−0.60
t4k	Ph	H	CH ₃		−0.60
t4m	Ph	F	CH ₃		−0.60

conformation according to our previous CoMFA study results (comprising 39 azole fungicides, see Chart 1 and Table 4) [9], were used in this study as the ligand alignment or pharmacophore.

The two steps of the methodology corresponding to identification of the receptor nucleation sites and construction of the pseudoreceptor (active site of the enzyme in our case) were accomplished defining the

receptor site from a recently published model of the active site of P-450_{DM} built by homology modeling [48], by deleting all residues further than 3.5 Å from any of the bound ligands, following in this way a methodology already described in the literature [49]. The binding pocket included 13 residues, namely Gly89, Phe380, Phe105, Arg381, Phe126, His468, Gly307, Val510, Thr311, Leu511, Pro375, Leu376 and Ile379 plus the heme moiety. The 39 azole ligands were then docked into the active site cavity and the whole system was fully minimized [3].

Thereafter, the pharmacophore-active site complex was subjected to the energetic equilibration protocol to

derive an energetically relaxed model with a high correlation between calculated and experimental free energy of ligand binding. An RMS value of 0.45 kcal/mol was reached yielding a refined model with a correlation coefficient of 0.917 after three cycles of energetic equilibration (Fig. 6).

To validate the model a test set of 13 fungicides (see Chart 1) was added to the receptor surrogate and minimized within the fixed receptor cavity, yielding a RMS error in predictions of 1.76 kcal/mol (Fig. 7). Within the test set, the largest deviation of predicted and experimental activities was obtained for compounds **t4**.

The promising results obtained with the pseudoreceptor modeling approach reported here (Fig. 7) demonstrate its validity in bridging structure-based receptor fitting and property-based receptor mapping methodologies to study and propose macromolecular binding sites and predict their interactions with bioactive conformers of the ligands.

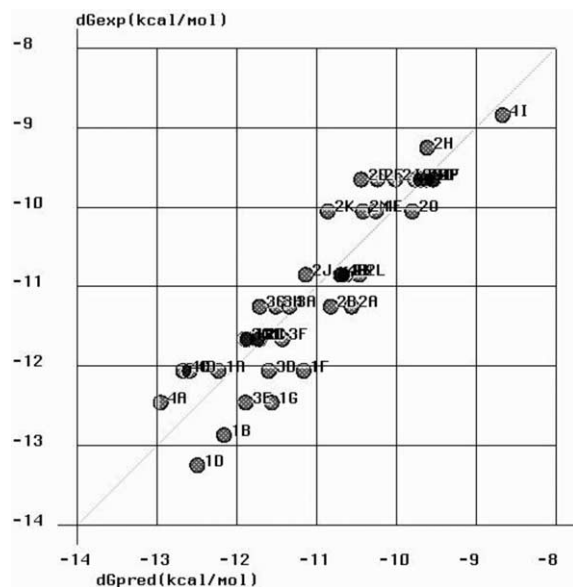


Fig. 6. Fitted predictions vs. measured activities of the 39 lanosterol 14 α -demethylase **t** inhibitors.

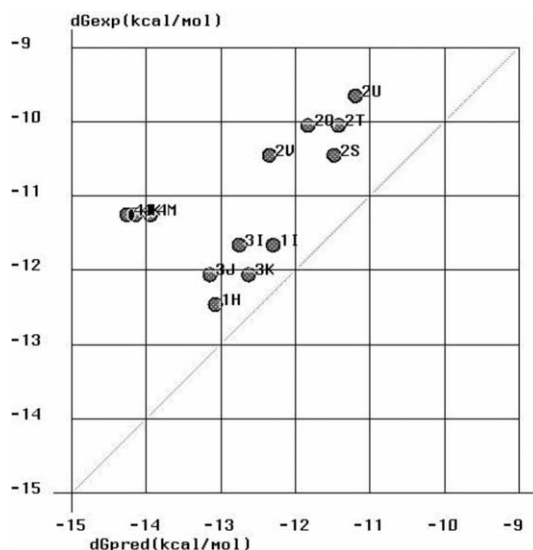


Fig. 7. Predicted vs. measured activities for the test set **t**-compounds.

References

- [1] Programs MacroModel and BatchMin, version 4.5, Columbia University, New York, NY.
- [2] InsightII software, version 95.0, Molecular Simulation, Inc., San Diego, CA.
- [3] P. Zbinden, M. Dobler, G. Folkers, A. Vedani, *Quant. Struct.-Act. Relat.* 17 (1998) 122.
- [4] P. Grochulski, F. Bouthillier, R.J. Kazlauskas, A.N. Serreqi, J.D. Schrag, E. Ziomek, M. Cygler, *Biochemistry* 33 (1994) 3494–3500.
- [5] M. Cygler, P. Grochulski, R.J. Kazlauskas, J.D. Schrag, F. Bouthillier, B. Rubin, A.N. Serreqi, A.K. Gupta, *J. Am. Chem. Soc.* 116 (1994) 3180–3186.
- [6] J. Wang, Y. Yan, T.P.J. Garrett, J. Liu, D.W. Rodgers, R.L. Garlick, G.E. Tarr, Y. Husain, E.L. Reinherz, S.C. Harrison, *Nature* 348 (1990) 411.
- [7] S.-E. Ryu, P.D. Kwong, A. Truneh, T.G. Porter, J. Arthos, M. Rosenberg, X. Dai, N. Xuong, R. Axel, R. Sweet, W. Hendrickson, *Nature* 348 (1990) 419.
- [8] P. Grochulski, Y. Li, M. Cygler, personal communication, 1996.
- [9] A. Tafi, J. Anastassopoulou, T. Theophanides, M. Botta, F. Corelli, M. Artico, R. Costi, R. Di Santo, S. Massa, R. Ragno, *J. Med. Chem.* 39 (1996) 1227–1235.
- [10] M.P. Schneider (Ed.), *Enzymes as Catalysts in Organic Synthesis*, Reidel, Dordrecht, The Netherlands, 1986.
- [11] L. Poppe, L. Novák, *Selective Biocatalysis*, VCH, Weinheim, 1992.
- [12] K. Faber, *Biotransformations in Organic Chemistry*, Springer, Berlin, Germany, 1992.
- [13] S. Servi (Ed.), *Microbial Reagents in Organic Synthesis*; NATO ASI Series, Vol. 381, Kluwer Academic, Dordrecht, The Netherlands, 1992.
- [14] C.-H. Wong, G.M. Whitesides, *Enzymes in Synthetic Organic Chemistry*, Elsevier Science Ltd, Oxford, England, 1994.
- [15] K. Drauz, H. Waldmann (Eds.), *Enzyme Catalysis in Organic Synthesis*, VCH, Weinheim, 1995.
- [16] Z.S. Derewenda, U. Derewenda, G.G. Dodson, *J. Mol. Biol.* 227 (1992) 818.

- [17] U. Derewenda, A.M. Brzozowski, D.M. Lawson, Z.S. Derewenda, *Biochemistry* 31 (1992) 1532.
- [18] D. Blow, *Nature* 351 (1991) 444.
- [19] A.M. Brzozowski, U. Derewenda, Z.S. Derewenda, G.G. Dodson, D.M. Lawson, J.P. Turkenburg, F. Bjorkling, B. Hugen-Jensen, S.A. Patkar, L. Thim, *Nature* 351 (1991) 491.
- [20] L. Brady, A.M. Brzozowski, Z.S. Derewenda, E. Dodson, G.G. Dodson, S. Tolley, J.P. Turkenburg, L. Christiansen, B. Hugen-Jensen, L. Norskov, L. Thim, U. Menge, *Nature* 343 (1990) 767.
- [21] P. Grochulski, Y. Li, J.D. Schrag, F. Bouthillier, P. Smith, D. Harrison, B. Rubin, M. Cygler, *J. Biol. Chem.* 268 (1993) 12843.
- [22] H. van Tilbeurgh, L. Sarda, R. Verger, C. Cambillau, *Nature* 359 (1992) 159.
- [23] F.K. Winkler, A. D'Arcy, W. Hunziker, *Nature* 343 (1990) 771.
- [24] M.M.G.M. Thunnissen, E. Ab, K.H. Kalk, J. Drenth, B.W. Dijkstra, O.P. Kuipers, R. Dijkman, G.H. de Haas, H.M. Verheij, *Nature* 347 (1990) 689.
- [25] M. Botta, E. Cernia, F. Corelli, F. Manetti, S. Soro, *Biochim. Biophys. Acta* 1337 (1997) 302.
- [26] M. Cyglier, J.D. Schrag, *Biochim. Biophys. Acta* 1441 (1999) 205–214.
- [27] F. Haefner, T. Norin, *Chem. Pharm. Bull.* 47 (1999) 591–600.
- [28] R.J. Kazlauskas, *Curr. Opin. Chem. Biol.* 4 (2000) 81–88.
- [29] R.J. Kazlauskas, A.N.E. Weissfloch, A.T. Rappaport, L.A. Cucchia, *J. Org. Chem.* 56 (1991) 2656.
- [30] S.M. Roberts, *Phil. Trans. R. Soc. London B* 324 (1989) 557.
- [31] T. Umemura, H. Hirohara, in: J.R. Whitaker, P.E. Sonnet (Eds.), *Biocatalysis in Agricultural Biotechnology*, ACS, Washington, DC, 1989 Chapter 26.
- [32] K. Laumen, Ph.D. Thesis, BUGH-Wuppertal, 1987.
- [33] K. Laumen, M.P. Schneider, *J. Chem. Soc., Chem. Commun.* (1988) 598.
- [34] D. Breitgoff, Ph.D. Thesis, BUGH-Wuppertal, 1989.
- [35] A. van Almsick, J. Buddrus, P. Hoenicke-Schmidt, K. Laumen, M.P. Schneider, *J. Chem. Soc., Chem. Commun.* (1989) 1391.
- [36] A. van Almsick, Ph.D. Thesis, BUGH-Wuppertal, 1990.
- [37] T. Ema, J. Kobayashi, S. Maeno, T. Sakai, T. Utaka, *Bull. Chem. Soc. Jpn.* 71 (1998) 443.
- [38] H. Dugas, *Bioorganic Chemistry. A Chemical Approach to Enzyme Action*, 2nd ed., Springer, New York, 1989 Chapter 4.
- [39] M. Holmquist, F. Hæffner, T. Norin, K. Hult, *Protein Sci.* 5 (1996) 83–88.
- [40] M. Norin, K. Hult, A. Mattson, T. Norin, *Biocatalysis* 7 (1993) 131–147.
- [41] F. Hæffner, T. Norin, K. Hult, *Biophys. J.* 74 (1998) 1251–1262.
- [42] D.J. Clanton, R.W. Buckheit, S.J. Terpening, R. Kiser, N. Mongelli, A.M. Borgia, R. Schultz, V. Narayanan, J.P. Bader, W.G. Rice, *Antiviral Res.* 27 (1995) 335.
- [43] G. Biasoli, M. Botta, M. Ciomei, F. Corelli, M. Grandi, F. Manetti, N. Mongelli, A. Paio, *Med. Chem. Res.* 4 (1993) 202.
- [44] F. Manetti, V. Cappello, M. Botta, F. Corelli, N. Mongelli, G. Biasoli, A. Lombardi Borgia, M. Ciomei, *Bioorg. Med. Chem.* 6 (1998) 947.
- [45] P.D. Kwong, R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski, A. Hendrickson, *Nature* 393 (1998) 648.
- [46] Unpublished results.
- [47] M. Gurrath, G. Müller, H.D. Höltje, *Perspectives in Drug Discovery and Design*, 12/13/14: 135–157, 1998. KLUWER/ES-COM.
- [48] H.-D. Hoeltje, C. Fattorusso, *Pharm. Acta Helv.* 72 (1998) 271–277.
- [49] J.M. Jansen, K.F. Koehler, M.H. Hedberg, A.M. Johansson, U. Hacksell, G. Nordvall, J.P. Snyder, *J. Chem. Inf. Comput. Sci.* 37 (1997) 812.
- [50] C.S. Chen, Y. Fujimoto, G. Girdankas, C.J. Sih, *J. Am. Chem. Soc.* 104 (1982) 7294–7299.