A Homology Model of Penicillin Acylase from *Alcaligenes faecalis* and In Silico Evaluation of its Selectivity

Paolo Braiuca,[a] Cynthia Ebert,[a] Lutz Fischer,[b] Lucia Gardossi,*[a] and Paolo Linda[a]

A three-dimensional model of the relatively unknown penicillin acylase from Alcaligenes faecalis (PA-AF) was built up by means of homology modeling based on three different crystal structures of penicillin acylase from various sources. An in silico selectivity study was performed to compare this homology model to the structure of the Escherichia coli enzyme (PA-EC) in order to find any selectivity differences between the two enzymes. The program GRID was applied in combination with the principal component analysis technique to identify the regions of the active sites where the PAs potentially engage different interactions with ligands. These differences were further analyzed and confirmed by molecular docking simulations. The PA-AF homology model provided the structural basis for the explanation of the different enantioselectivities of the enzymes previously demonstrated experimentally and reported in the literature. Different substrate selectivities were also predicted for PA-AF compared to PA-EC. Since no crystallographic data are available for PA-AF to date, the three-dimensional homology model represents a useful and efficient tool for fully exploiting this attractive and efficient biocatalyst, particularly in enantioselective acylations of amines.

Introduction

Among industrially employed enzymes, penicillin acylase is one of the most widely studied and used, since the β-lactam antibiotics industry has replaced the traditional multistep chemical process for the production of 6-aminopenicillanic acid with an enzymatic process that only uses penicillin acylase (PA).[1] The most common source of commercially available penicillin acylase is *Escherichia coli* (PA-EC), though the same enzyme from *Alcaligenes faecalis* (PA-AF) has recently received attention in the literature,[2–4] in particular because of its high synthetic efficiency in enantioselective synthesis. Moreover, PA-AF has a clear advantage over other well-characterized PAs for industrial applications because of its higher thermostability, ascribed to the presence of two cysteine residues (absent in the *E. coli* enzyme), which form a disulfide bridge.[5] This property makes the enzyme a more attractive biocatalyst for both synthesis and modification of β-lactam antibiotics and industrial interest is demonstrated by the patent deposited in 1995.[6]

The penicillin acylase from *A. faecalis* (PA-AF) was sequenced in 1997[5] and its sequence is available in the SWISS-PROT/TrEMBL database.[7] The sequence identity between PA-AF and the penicillin acylase from *E. coli* (PA-EC) is 47%, and the residue similarity is even higher (67%). Švedas and co-workers hypothesized, on the basis of kinetic studies, that the active centres of the enzymes PA-AF and PA-EC differ in the details of their organization, especially in the subsite responsible for the interaction with the leaving group of the substrate.[4] However, no direct structural evidence has been reported so far to support this hypothesis because of the lack of a crystal structure of PA-AF. A three-dimensional model of this enzyme would also be of major importance for comparing the binding sites of the two enzymes to gain a better understanding of both the extremely effective recognition of the acyl moiety[4] and the stereoselective discrimination of the aminic part of the substrate achieved by these enzymes. We recently demonstrated that the GRID computational method, coupled with the classical docking/tetrahedral intermediate approach, can be applied successfully to the interpretation and modeling of penicillin G acylase (PGA) substrate- and enantioselectivity.[8, 9]

Herein, we present a study that leads to the first three-dimensional model of the PA-AF enzyme, obtained by means of

[a] *Dr. L. Gardossi, P. Braiuca, Prof. C. Ebert, Prof. P. Linda*
Dipartimento di Scienze Farmaceutiche
Università degli Studi di Trieste
P.le Europa 1, 34127 Trieste (Italy)
Fax: (+39) 040-52572
E-mail: gardossi@units.it

[b] *Prof. Dr. L. Fischer*
Institute of Food Technology
Department of Biochemistry
University of Hohenheim
Garbenstrasse 25 und Emil-Wolff-Strasse 14
70599 Stuttgart (Germany)

homology modeling. The obtained 3D model was used for an in silico study based on the GRID approach combined with principal component analysis (PCA) and the molecular docking approach to provide the structural basis for interpreting and predicting substrate selectivity and stereospecificity of PA-AF. The model was finally validated by comparing the theoretical results to experimental data available from the literature.

Results and Discussion

Homology model

A three-dimensional model of PA-AF was built up by means of homology modeling based on the known X-ray structures of PA from *E. coli*[10–15] and *Providentia rettgeri*,[16] available from the Protein Data Bank (PDB).[17]

At present, 18 crystallographic structures of PA-EC and one of penicillin acylase from *P. rettgeri* (PDB code: 1CP9) are available. Template structures for an investigation such as that described here should be selected by taking into account the resolution and the degree of homology of such structures. Moreover, conformational modifications induced by ligands and/or mutations must be considered. The selected structures must correspond to catalytically active proteins. In all the structures available that maintain catalytic activity, the three-dimensional structure of the protein in general, and in particular the conformation of the active site are highly conserved (for the native form and for mutants[13] or enzymes from different sources[16]).

The acylic subsite, which has been studied in detail both experimentally[8, 11] and theoretically,[8] shows high steric and electronic specificity towards phenyl rings. It has been demonstrated experimentally[11] and theoretically[8] that those substitutions of the phenyl ring that cause conformational distortions translate into a dramatic decrease of substrate affinity[11] and reaction rate.[8] These conformational modifications involve Arg A:145 and Phe A:146 (A indicates that the residues are in Chain A of the enzyme) and there are strong arguments supporting the hypothesis that these modifications are incompatible with the preservation of the catalytic activity.[12] Such conformational modifications always lead to PA molecules with strongly reduced activity; in one case, Janssen and co-workers reported on the complex between an inactive mutant of PA and penicillin G,[13] while Done and co-workers described the interaction of conformationally distorted PAs with sterically hindered analogues of phenylacetic acid.[11] Moreover, we have previously demonstrated by making use of the GRID/tetrahedral intermediate approach[8] that substitution of a bulky nitro group onto the phenyl group causes not only weak enzyme–substrate recognition as reported by Done, but also a dramatic decrease in v_0 (initial rate of acylation reaction), as confirmed by experimental data.[8]

The crystal structure with PDB code 1GM9[15] is a complex of the enzyme with a real substrate, penicillin G sulfoxide. This substrate is hydrolyzed by penicillin acylase, but very slowly, which makes it possible to get a crystal of the complex. The conformation of the active site is the same in the free enzyme as

in complexes with other substrate-analogous inhibitors reported by Done.[11] Therefore, we can reasonably conclude that there is negligible conformational freedom and adaptability of the PGA active site upon the substrate binding.

Among the X-ray structures of PA-EC, 1GK9[12] (1.30 Å) and 1PNK (1.90 Å) have the best resolution, although the structure of PGA from the Bro1 mutant strain of *P. rettgeri* (2.50 Å; PDB code: 1CP9)[16] shows the highest degree of homology with the PA-AF sequence (49% identity, 69% homology). The homology values for these structures are high enough to assure the construction of a reliable model for PA-AF. A comparison of the crystallographic structures of PA from *P. rettgeri* (1CP9) and PA from *E. coli* (1PNK), which have a sequence identity of 64% and homology of 75%, shows a root mean square deviation (RMSD) value for the positions of the α carbon atoms and the whole backbone of 0.2 Å and 0.7 Å, respectively. The homology modeling process was performed using several structures as templates, namely 1CP9 because of its high homology with PA-AF, and 1GK9 and 1PNK because of their high resolution.

The first and probably most critical step in protein homology modeling is the alignment of sequences. Alignments were performed by using the molecular operating environment (MOE)[18] implementation of the Gonnet matrix system, visually verified, and manually corrected. The correspondence and the right alignment of the catalytic site residues were verified, with particular attention to Ser B:1, Asn B:241, Ala B:127, Phe A:146, Arg A:145, and Arg B:263, which are known to be crucial to the catalytic mechanism.[10–15]

The construction of the three-dimensional model was carried out by using the MOE homology modeling module, with enzyme structure 1CP9 as a template in the first step. The α and β chains were modeled separately (the enzyme is a heterodimer) by means of calculation of 25 coarsely minimized intermediate models, which were ranked by the structure quality Z score of the MOE. One model of the α chain and one of the β chain had to be chosen from this ensemble of structures, not only on the basis of the Z score but also by taking into account the reciprocal positions of the two chains. Each single intermediate structure was then visually inspected and analyzed by means of the protein evaluation tools available with the MOE module. A careful inspection of the zone of contact between the two chains was necessary for the choice of the best model for each chain. In the best model of the β chain the positions of Cys492 and Cys525 are perfectly compatible with a disulfide bridge, so a bridge was added manually, in agreement with previous reports.[5]

Homology modeling of the two chains was repeated by following the same procedure as described above, with the 1GK9 and 1PNK structures as templates. The two further homology models that resulted showed substantial similarities to the model obtained by starting from the 1CP9 structure, since RMSD values for the backbone and sidechain carbon atoms were always below 0.7 Å for these structures. However the highest Z score was obtained by starting from 1CP9, which indicates that this homology model has greater statistical significance. This model was thus chosen for the refining process and the subsequent selectivity study.

The protein validation tools of the MOE system guided the refining process by pointing out structural errors and by evaluating the quality of the structure. This refining process was focused particularly on the zone of contact between the two chains. Atom contacts were manually corrected and the structure was relaxed by a series of consecutive local molecular dynamics simulations and energy minimizations. Refinement of dihedral angles did not require manual corrections and was performed by a series of molecular dynamics simulations that involved the movement of those residues that needed to be refined.

The quality of the model was assessed by the protein report tool of the MOE program and the results summarized in Table 1 indicate that the structure is of good quality, with more than 99% of residues in the allowed zone of the Ramachandran plot (85% in the core of the plot), and with statistically reliable χ angles.

Structural differences between *A. faecalis* and *E. coli* PAs

The protein consensus tool of the MOE program was used to identify regions of conserved structure in the two enzymes (PA-AF and PA-EC). The majority of the α-chain is highly conserved, as is the core of the protein, which includes the acyl cleft of the active site (RMSD below 0.5 Å). The residues directly involved in the catalysis (Ser B:1, Ala B:69, Asn B:241, and Arg B:263) are also strictly conserved and substitution of Ser B:67 with an Ala residue is the only significant change in the acyl cleft.

More significant differences between the homology model and the template molecule were observed in the aminic subsite and in general in the part of the protein that is more exposed to the solvent, although the global atom standard deviation values of the backbone and the sidechains are below 1.5 Å. The peripheral part of the protein is the most variable part and this can be ascribed to: 1) the presence of four extra residues in PA-AF between aminoacids A:57 and A:70 of PA-EC; 2) the C-terminal portion of the β chain, which contains two more residues in PA-AF than in PA-EC; 3) the deletion of two residues in the PA-AF sequence in comparison to the loop between residues B:332 and B:337 in PA-EC. These differences are shown in Figure 1.

In the aminic subsite, residue B:384 is a Thr residue in PA-EC but a Pro residue in PA-AF and this difference translates into a remarkable difference in folding in this zone of the protein. Also, the bulkier Phe B:71 of PA-EC is substituted by a Pro in PA-AF and this widens the space available in the aminic cleft (Figure 1).

Selectivity differences

In order to investigate the differences in selectivity between PA-AF and PA-EC two different strategies were used: the GRID/PCA
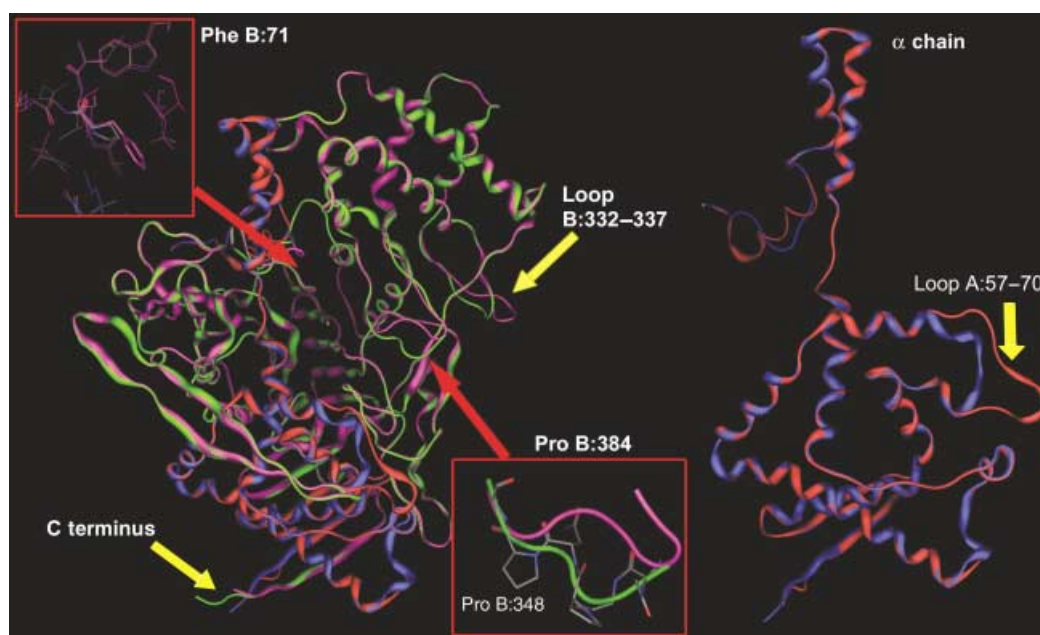
Table 1. *Protein report of the homology model.*				
Parameter	Observed		Expected	
	Mean	s.d.[a]	Mean	s.d.[a]
trans omega	171.9	7.1	180.0	5.8
C-α chirality	33.7	2.0	33.8	4.2
$\chi 1$ – gauche minus	− 59.1	17.8	− 66.7	15.0
$\chi 1$ – gauche plus	56.7	17.6	64.1	15.7
$\chi 1$ – trans	187.1	13.3	183.6	16.8
helix ϕ	− 66.1	17.7	− 65.3	11.9
helix ψ	− 35.4	19.7	− 39.4	11.3
$\chi 1$ – pooled s.d.	-	15.1	-	15.7
proline ϕ	− 60.6	18.1	− 65.4	11.2
[a] s.d. = standard deviation.				

Figure 1. *Schematic representation of the PA-AF backbone, superimposed on the PA-EC backbone. Arrows point out the zones where the two proteins differ the most. Color code: PA-AF α chain, red; PA-AF β chain, green; PA-EC α chain, blue; PA-EC β chain, magenta.*

method and molecular docking. The GRID/PCA approach is a computational strategy applicable to the study and detailed description of differences in selectivity within a family of related proteins. The method leads to the identification of those structural differences that effectively translate into relevant and significant variations in enzyme – substrate interactions.[19, 20] The GRID procedure, developed by Goodford in 1985, allows estimation of the interaction energies between small chemical groups (probes) and a target (either a protein or other small organic molecule) by establishing a regular array of grid points throughout and around the target. This computational tool detects energetically favourable binding sites on such targets of known three-dimensional structure.[21–24] The output of the calculation is a molecular interaction field (MIF), which is a three-dimensional matrix of the interaction energies between the target and the probes calculated for each grid node. This three-dimensional matrix was statistically analyzed by means of principal component analysis (PCA) with the GOLPE program.[25] PCA analysis pointed out those zones of the two structurally related proteins engaging in interactions with a defined ligand that differ in terms of nature and strength of the interaction (see the Computational Methods section and Figure 2). An adequate statistical treatment of the data set allowed irrelevant variables to be discarded, as reported in the Computational Methods section. The structure of PA-EC with the PDB code 1PNK and the PA-AF homology model were used. The probes were divided into two groups: apolar and H-bond donor/acceptor (Table 2).

Acylic subsite

The GRID/PCA analysis showed how the substitution of Ser B:67 with an Ala residue in PA-AF results in weaker interactions of the

Table 2. Probes used for the selectivity study.	
Apolar probes	
DRY	hydrophobic
C3	aliphatic carbon
C1	aromatic carbon
Polar probes	
OH	phenolic hydroxy group
ON	nitro group oxygen atom
N2:	sp^3 NH$_2$, with a lone pair
N2	neutral flat NH$_2$
N3 +	amine NH$_3$ cation
O:	carboxy oxygen atom
O	carbonyl oxygen atom

acylic subsite with H-bonding probes. Figure 3 a shows the PCA scores plot of the hydrogen-bonding probes. This plot shows how the first principal component discriminates between the enzymes and allows verification of whether a difference in substrate – enzyme interaction exists. In the PCA loading plot of the OH probe (Figure 3 b) variables responsible for this difference can be identified (circled in the figure). These variables can be translated into a three-dimensional representation of their original spatial position within the catalytic site (red cubes, Figure 3 c and 3 d). This allows immediate identification of the most important regions within the enzyme where a difference in substrate selectivity occurs. The MIF of the OH probe, represented in the figure as a blue surface, quantifies the difference. In this case, the MIF represents an isopotential surface of 2.5 kcal mol^{-1}, which corresponds to a hypothetical difference of 2.5 kcal mol^{-1} in interaction energy when a OH group binds the enzyme in this zone compared to interaction energy with the
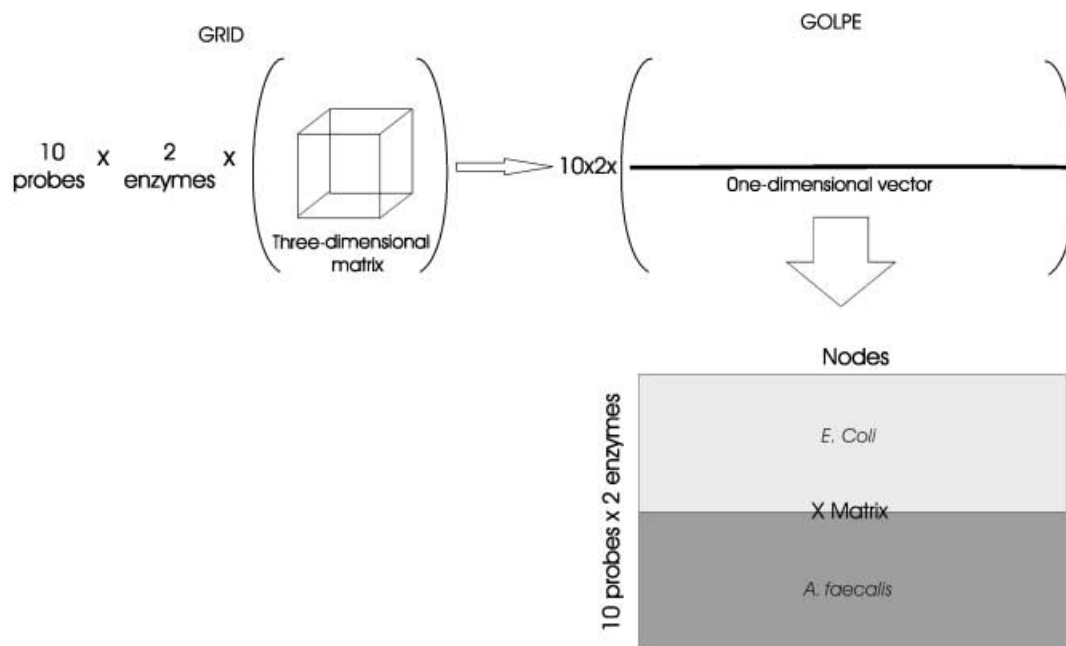

Figure 2. Scheme of the procedure for the conversion of the MIFs into the X matrix for the multivariate analysis. The three-dimensional matrices of interaction energies were converted into one-dimensional vectors, which were combined to form the X matrix.
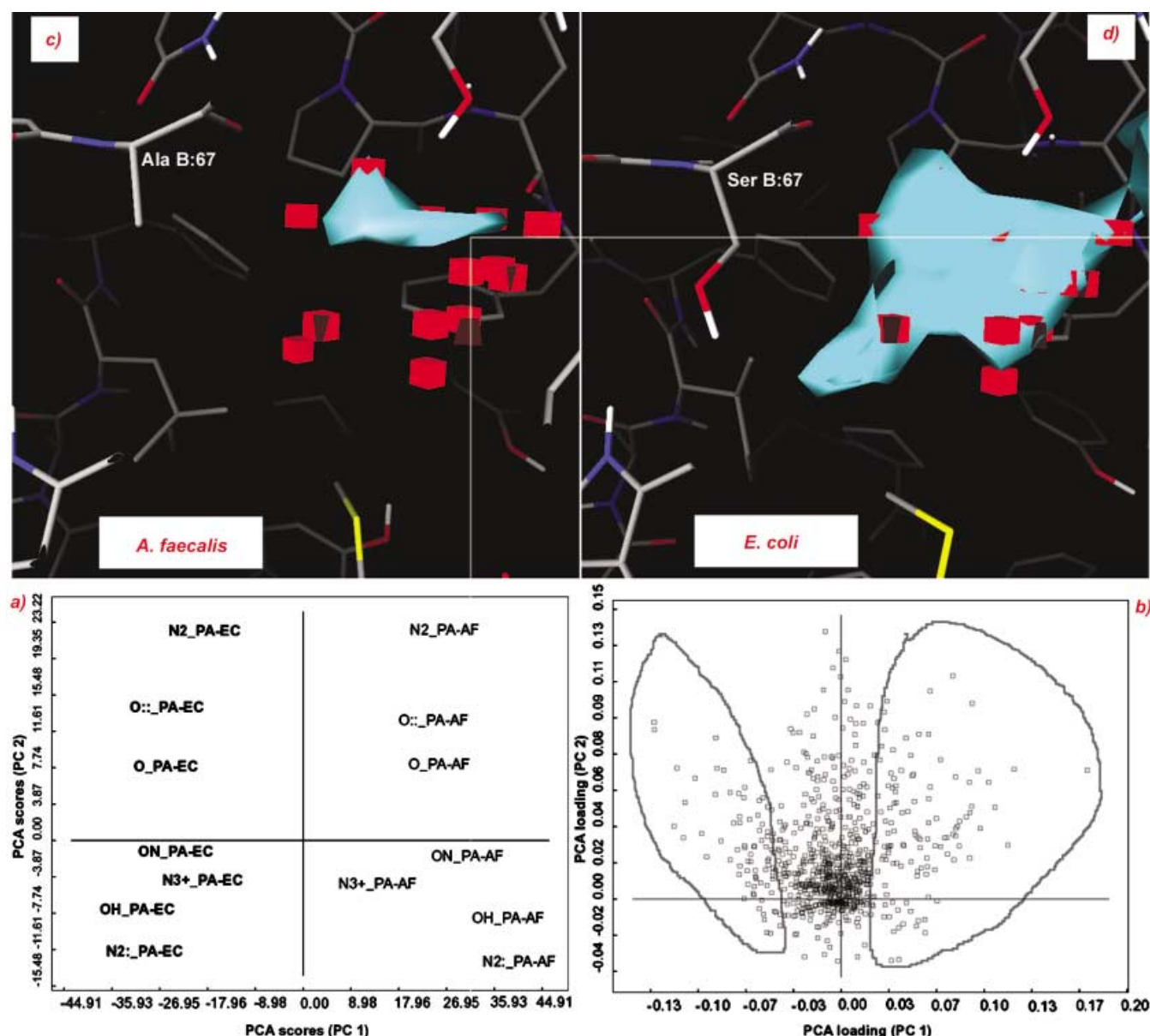
Figure 3. *Detailed description of the results of GRID/PCA analysis in the zone close to the B:67 position. The PCA scores plot of the hydrogen-bonding probes (a) shows that the first principal component discriminates between the enzymes and allows verification of whether a difference in substrate – enzyme interaction exists. Variables responsible for this difference (b, circled) can be identified in the PCA loading plot of the OH probe. These variables can be translated into a three-dimensional representation of their original spatial position within the catalytic site (red cubes; c and d). This process allows immediate identification of the most important regions within the enzyme where a difference in substrate selectivity exists between the two enzymes studied. The MIF, represented in the figure as a blue surface, quantifies the magnitude of the difference.*

template enzyme. Therefore, the different interactions with various hydrogen-bonding probes can reasonably be ascribed to the substitution of Ser B:67 with an Ala residue.

It is widely recognized and has been experimentally demonstrated that PA-EC shows its highest selectivity towards 4-hydroxyphenylacetic acid and this has been ascribed to the stabilizing electrostatic interaction that occurrs between the hydroxy substituent on the phenyl ring of the substrate and Ser B:67.[8, 9, 11] Therefore, the presence of Ala B:67 in PA-AF causes a decreased interaction energy with 4-hydroxyphenylacetic acid. This effect is expected to cause a lower selectivity of PA-AF towards this substrate. To obtain a deeper understanding of this

selectivity difference, a docking simulation was performed by using an ensemble of conformers of the 4-hydroxyphenylacetic acid and applying the GROUP program and the MOVE directive of the GRID package, which allowed us to take into account the flexibility of the side chains of the amino acids.

The calculated difference between the complexation energies of the substrate with the two enzymes is about 2.5 kcal mol^{-1}, which is perfectly consistent with the loss of an H bond. This outcome is in agreement with the expected lower selectivity of PA-AF for the 4-hydroxyphenylacetic residue, which leads to a decrease of the catalytic rate k_{cat} of the acylation/deacylation reaction, as well as to a lower inhibition constant.

Aminic subsite

Most of the structural differences between the two PAs are located in the aminic subsite. The GRID/PCA procedure revealed a larger MIF for PA-AF than for PA-EC with both hydrophobic and polar probes. This difference is mainly due to the exchange of the Phe B:71 found in PA-EC for a Pro residue in PA-AF. This variation does not mutate the chemical nature of this zone in the active site but enables the aminic subsite to accommodate bulkier nucleophiles. The GRID/PCA analysis shows how the presence of a proline residue in position B:384 in PA-AF leads to a different backbone conformation from that in PA-EC and causes a slight shrinkage of the cleft. This result is particularly important since this zone is directly involved in the enantiospecific recognition of aminoesters, which are *N*-acylated by PA.[9] Arg B:263 and Ser B:386 interact with the carbonyl oxygen atoms of esters of L-aromatic aminoacids, whereas D-substrates cannot achieve a conformation suitable to establish this type of stabilizing interaction.[9] The different conformation of the backbone in this part of the subsite results in a reduction of the space available for substrate accommodation in PA-AF and this characteristic can strongly influence the selectivity of the enzyme towards chiral nucleophiles. This structural evidence could provide an explanation for the enantioselectivity of PA-AF towards chiral amines lacking the acid/ester group, as demonstrated experimentally by Švedas and co-workers.[2, 3] In order to test and further validate the theoretical model, a docking simulation was performed with five amides accepted by both PA-AF and PA-EC, although with a different enantioselectivity.[2, 3]

The docking simulation was carried out with the MOVE directive of the GRID package, which allows the flexibility of the side chains of amino acids to be taken into account. All the resulting conformations were visually inspected in order to discard those with a geometry incompatible with catalysis. Fewer than five structures resulted from each simulation carried out and most of these structures could be discharged simply on the basis that the position of the substrate in the active site was clearly incompatible with catalysis, either because the ligand was outside the active site or because of the position of the carbonyl group with respect to the catalytic Ser residue. The remaining complexes were subjected to energy minimization, during which all the protein residues included in a sphere of 10 Å from the substrate were allowed to move. We observed, however, that the protein did not undergoe any conformational modification during minimization. In all the complexes, the substrates have the carbonyl carbon atom of their amide group at a distance of 2.45 – 2.65 Å from the hydroxy oxygen atom of Ser B:1, and their carbonyl oxygen atom points towards the oxyanion hole in an orientation suitable for nucleophilic attack by the catalytic serine residue.

The docking results (Table 3) are in a good agreement with experimental data for both the substrates. The different shapes of the aminic subsites of the two enzymes that result from the change of Phe B:71 into Pro in PA-AF, and the different backbone conformations of the enzymes caused by the change of Thr B:384 into Pro in PA-AF were shown to have a major effect on enzyme – substrate interactions.

The two enantiomers of each of the compounds **1**, **3**, and **5** have quite similar conformations in PA-EC and thus have slightly different interaction energies, and this difference translates into low enantioselectivity. The enantiomers of the same substrates in PA-AF show noticeable interaction energy differences as a result of the changed shape of the cleft caused by the shrinkage of the space available in the zone close to Pro B:384 in comparison to that in PA-EC. This change in shape results in an energy penalty for the interaction of the enzyme with (*S*)-**1**, (*S*)-**3** and (*R*)-**5**. The enantiopreference of both the enzymes towards the (*S*)-enantiomer of **2** is ascribable to the loss of the polar interactions between the hydroxy group and both Ser B:386 and Arg B:263 in the case of the (*R*)-enantiomer. Again, the narrower cleft of PA-AF enhances the enantioselectivity.

The lack of enantiospecific recognition of **4** by both the enzymes can be abscribed to the higher planarity of the indanyl moiety compared the structures of the other substrates. The indanyl group occupies almost the same space for both enantiomers, which results in a very low interaction energy difference between the two enantiomers and translates into a lack of enantioselectivity.
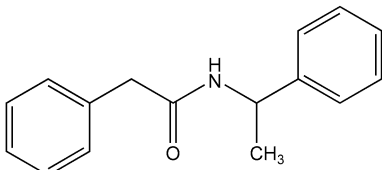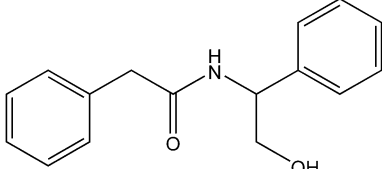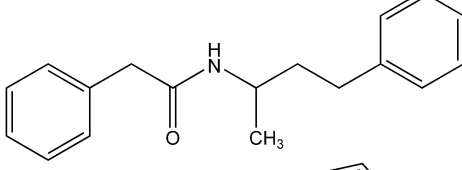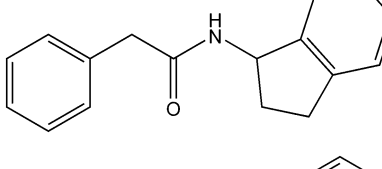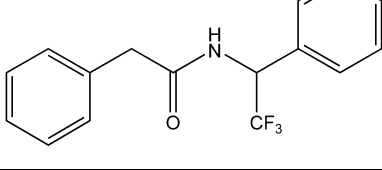
The value of $\Delta\Delta E_{EC\text{-}AF}$ gives an estimate of the relative stability of the complexes formed by the two enzymes with the same substrate. The large values calculated for the substrates are ascribable to the different space available for accommodating these molecules in the enzymes. Therefore, the $\Delta\Delta E_{EC\text{-}AF}$ value is an index of substrate selectivity and the results are in agreement with the experimentally verified higher selectivity found for PA-AF than for PA-EC.

Conclusion

Herein, we report the first three-dimensional model of penicillin acylase from *A. faecalis*, built up by means of homology modeling. The theoretical homology model was compared to the *E. coli* enzyme structure, and an in silico selectivity study based on the innovative GRID/PCA technique and on a molecular docking approach enabled identification of the regions of the active sites in which the two PAs engage in different interactions with ligands. PCA proved to be a very effective tool for the comparison of the two structures on a rational basis, and also sped up the computational analysis. A comparison of two 3D structures carried out only on an empirical basis (simply by observation of differences) would be both more time consuming and less reliable than the GRID/PCA method since this computational approach is able to point out differences between the enzymes and also to evaluate the nature and extent of these differences once the "noise" of insignificant differences has been cleared away. As a result, only significant variations in chemical and structural terms were investigated in detail and these variations turned out to be in agreement with experimental data on selectivity.

Important differences in both the acylic and the aminic subsites were revealed and shed light on the structural basis of the different substrate enantiorecognition of the two enzymes, which was previously demonstrated experimentally by Švedas and co-workers. Furthermore, a different selectivity towards

Table 3. *Comparison of experimental* E[2, 3] *values with docking results.*[a]

Substrate	Enzyme	E (preferred enantiomer)	$\Delta\Delta E$ [kcal mol^{-1}]
1	E. coli	10 (R)	− 2.3 (R)
	A. faecalis	310 (R)	− 16.8 (R)
2	E. coli	95 (S)	− 36.1 (S)
	A. faecalis	670 (S)	
3	E. coli	2 (R)	− 0.9 (S)
	A. faecalis	110 (R)	− 11.1 (R)
4	E. coli	12 (R)	− 1.9 (R)
	A. faecalis	2.8 (R)	− 0.4 (R)
5	E. coli	2 (S)	− 1.8 (S)
	A. faecalis	100 (S)	− 9.8 (S)

[a] *E*, interaction energy; $\Delta\Delta E$, difference between the interaction energies of the two enantiomers of each substrate with the enzyme, calculated by molecular modeling.

4-hydroxyphenylacetic acid was predicted on the basis of a change in the amino acid sequence at position B:67.

This study provides a theoretical structural model that can be used to fully exploit the stability and selectivity of PA-AF in the development of novel asymmetric syntheses and for the modification of β-lactam antibiotics.

Computational methods

All molecular modeling calculations were performed on an i686 Linux workstation, multivariate analyses were executed by the GOLPE program on an SGI O2 R10000 workstation.

Homology modeling, molecular dynamics simulations, and energy minimizations were calculated by using the molecular operating environment (MOE) version 2001.01 software. For GRID MIF calculation the GRID 20 Linux edition package was used. The docking simulations were made by using the GROUP program of the GRID 20 package. Energy minimizations were calculated with the Sybyl 6.8 program [26] by using the Amber force field.[27]

The chains alignment was performed with the align algorithm of the MOE program by using the Gonnet matrix with a penalty gap of 3 and a penalty for extending a gap of 1. The construction of the three-dimensional model was carried out with the MOE homology modeling module by means of calculation of 25 intermediate models that were coarsely minimized by using the 1CP9 structure as a template. The calculations were repeated with structures 1PNK and 1GK9. The resulting models were compared by means of the "protein consensus" tool of the MOE package for identification of regions of conserved structure within a set of superposed 3D protein structures. This comparative analysis indicated that the best models for both α and β chains were those calculated with structure 1CP9 as a template. The α and β chains models were joined together and the final structure was refined by manual corrections, local dynamics simulations, and energy minimizations, during which only residues within 5 Å of the substrate were allowed to move. The molecular dynamics simulations were set to 5 ps of heating to 600 K, 10 ps of equilibrium at 300 K, and 5 ps of cooling to 0 K. Energy minimizations were made in 100 steps of steepest descent with an RMS gradient test of 1000, 100 steps of conjugated gradient with an RMS gradient test of 100, and 500 truncated Newton steps with an RMS gradient test of 0.01. The Amber force field was used.

Water molecules were removed from the crystallographic structure 1PNK and then the two structures were superimposed by using the protein modeling tools of the MOE package. To perform an efficient GRID mapping of the enzymes, their global charges had to be neutralized. This task was performed by adding potassium ions to the proteins in positions calculated on a K$^+$ GRID MIF by using the MINIM and FILMAP programs. The GRID calculations were then based on the neutral structures. The GRID cage was built to include the active sites of the enzymes. The calculation was performed taking into account the conformational mobility of amino acid side chains by setting the MOVE directive to 1.

The three-dimensional matrixes of interaction energies (MIFs) obtained from the GRID calculations were converted into one-dimensional vectors and combined by using the GOLPE program to form a bidimensional X matrix. The first dimension of the X matrix represents the probe – enzyme interactions while the second dimension contains the values of the interaction energy calculated for each node for the multivariate analysis. Pretreatment of the data set was necessary to focus attention only on relevant variables, therefore all the variables with an absolute value below 0.01 or with a standard deviation of less than 0.03 were set to zero. Positive interaction energy values were also set to zero by using the "maximum positive cut-off" tool of the GOLPE software. The use of only the negative, favorable interaction energies has the advantage of removing the information related to interactions that are sterically

unfavourable because of closeness of the probe to the protein. The "region cut-out" tool of the GOLPE software was applied to discharge all the points of the grid lying outside the active site from the model and thus define the region of binding.[25] The data pretreatment allowed reduction of the number of active variables to about 1% of the whole data set.

Since hydrophobic interactions are usually much lower in absolute value than polar ones, two different submodels were built up for the apolar and for the polar probes, respectively. This process allowed the risk of underestimating hydrophobic interactions to be avoided and circumvented the need to use an autoscaling process, which can introduce significant noise into the model.

Principal component analysis was applied to the X matrix to extract the relevant information. The multivariate characterization of the catalytic sites produced a mathematical hyperspace, where each node corresponds to one dimension and each probe is represented by a point. The closer the positions of two points, the more strongly these points are correlated. The original data were projected onto a few principal components, or "latent variables". These principal components are linear combinations of all the original node values. The plot of the principal components gives information on the overall structure of the data set.

The docking simulations were performed by using the GROUP program of GRID 20 package. An ensemble of conformers of substrates was generated by a systematic search as implemented in the MOE package by rotating each single bond in 10 degrees steps. The conformers were used as input for the GROUP programs by means of the AutoGroup utility; the MOVE directive was set to 1. The resulting complexes were visually inspected in order to discard the unacceptable ones. All the complexes were finally minimized by using the Amber force field in its Sybyl implementation, with the Powell algorithm: the substrate and all the residues within a sphere of radius of 10 Å centred on the substrate were allowed to move. The $\Delta\Delta E$ values were calculated from the differences between E values calculated for each enantiomer – enzyme complex.

[1] M. A. Wegman, M. H. A. Janssen, F. van Rantwijk, R. A. Sheldon, *Adv. Synth. Catal.* **1995**, *343*, 559 – 576.

[2] L. M. van Langen, N. H. P. Oosthoek, D. T. Guranda, F. van Rantwijk, V. K. Švedas, R. A. Sheldon, *Tetrahedron: Asymmetry* **2000**, *11*, 4593 – 4600.

[3] D. T. Guranda, L. M. van Langen, F. van Rantwijk, R. A. Sheldon, V. K. Švedas, *Tetrahedron: Asymmetry* **2001**, *12*, 1645 – 1650.

[4] V. K. Švedas, D. T. Guranda, L. M. van Langen, F. van Rantwijk, R. A. Sheldon, *FEBS Lett.* **1997**, *417*, 414 – 418.

[5] R. M. D. Verhaert, A. M. Riemens, J. van der Laan, J. Duin, W. J. Quax, *Appl. Environ. Microbiol.* **1997**, *63*, 3412 – 3418.

[6] E. De Vroom, M. van der Mey, **1995**, 0638649 A2.

[7] A. Bairoch, R. Apweiler, *Nucleic Acids Res.* **2000**, *28*, 45 – 48.

[8] A. Basso, P. Braiuca, S. Clementi, C. Ebert, L. Gardossi, P. Linda, *J. Mol. Catal. B: Enzym.* **2002**, *19 – 20*, 423 – 430.

[9] A. Basso, P. Braiuca, C. Ebert, L. Gardossi, P. Linda, *Biochim. Biophys. Acta* **2002**, *1601*, 85 – 92.

[10] H. J. Duggleby, S. P. Tolley, C. P. Hill, E. J. Dodson, G. Dodson, P. C. E. Moody, *Nature* **1995**, *373*, 264 – 268.

[11] S. H. Done, J. A. Brannigan, P. C. E. Moody, R. E. Hubbard, *J. Mol. Biol.* **1998**, *284*, 463 – 475.

[12] J. A. Brannigan, G. G. Dodson, S. H. Done, L. Hewitt, C. E. McVey, K. S. Wilson, *Appl. Biochem. Biotechnol.* **2000**, *88*, 313 – 319.

[13] W. B. L. Alkema, C. M. H. Hensgens, E. H. Kroezinga, E. de Vries, R. Floris, J. van der Laan, B. W. Dijkstra, D. B. Janssen, *Protein Eng.* **2000**, *13*, 857 – 863.

[14] W. B. L. Alkema, A. Dijkhuis, E. de Vries, D. B. Janssen, *Eur. J. Biochem.* **2002**, *269*, 2093 – 2100.

[15] C. E. McVey, M. A. Walsh, G. G. Dodson, K. S. Wilson, J. A. Brannigan, *J. Mol. Biol.* **2001**, *313*, 139 – 150.

[16] M. A. McDonough, H. E. Kley, J. A. Kelly, *Protein Sci.* **1999**, *8*, 1971 – 1981.

[17] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235 – 242.

[18] Molecular operating environment (MOE) version 2001.01 distributed by Chemical Computing Group Inc., Montreal, Canada.

[19] G. Cruciani, K. Watson, *J. Med. Chem.* **1994**, *37*, 2589 – 2601.

[20] M. A. Kastenholz, M. Pastor, G. Cruciani, E. E. J. Haaksma, T. Fox, *J. Med. Chem.* **2000**, *43*, 3033 – 3044.

[21] P. J. Goodford, *J. Med. Chem.* **1985**, *28, 849 – 857.*

[22] D. N. A. Boobbyer, P. J. Goodford, P. M. McWhinnie, R. C. Wade, *J. Med. Chem.* **1989**, *32*, 1083 – 1094.

[23] R. Wade, K. J. Clerk, P. J. Goodford, *J. Med. Chem.* **1993**, *36*, 140 – 147.

[24] GRID 19 distributed by Molecular Discovery Ltd., London, UK.

[25] GOLPE distributed by MIA srl, Perugia, Italy.

[26] SYBYL 6.8 distributed by Tripos Inc., St. Louis, MO, U.S.A.

[27] S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, *J. Comput. Chem.* **1986**, *7*, 230 – 252.