Journal of Enzyme Inhibition and Medicinal Chemistry, 2011, 1–12, Early Online

ISSN 1475-6366 print/ISSN 1475-6374 online
DOI: 10.3109/14756366.2011.643305

informa
healthcare

ORIGINAL ARTICLE

Chemogenomics of pyridoxal 5′-phosphate dependent enzymes

Ratna Singh[1], Francesca Spyrakis[2,3], Pietro Cozzini[2,3], Alessandro Paiardini[4], Stefano Pascarella[4], and Andrea Mozzarelli[1,3]

[1]Department of Biochemistry and Molecular Biology and[2]Department of General and Inorganic Chemistry, Laboratory of Molecular Modelling, University of Parma, Parma, Italy, [3]National Institute of Biostructures and Biosystems, Rome, Italy, and [4]Department of Biochemical Sciences, "Sapienza" University of Rome, Rome, Italy

Abstract

Pyridoxal 5′-phosphate (PLP) dependent enzymes comprise a large family that plays key roles in amino acid metabolism and are acquiring an increasing interest as drug targets. For the identification of compounds inhibiting PLP-dependent enzymes, a chemogenomics-based approach has been adopted in this work. Chemogenomics exploits the information coded in sequences and three-dimensional structures to define pharmacophore models. The analysis was carried out on a dataset of 65 high-resolution PLP-dependent enzyme structures, including representative members of four-fold types. Evolutionarily conserved residues relevant to coenzyme or substrate binding were identified on the basis of sequence-structure comparisons. A dataset was obtained containing the information on conserved residues at substrate and coenzyme binding site for each representative PLP-dependent enzyme. By linking coenzyme and substrate pharmacophores, bifunctional pharmacophores were generated that will constitute the basis for future development of small inhibitors targeting specific PLP-dependent enzymes.

Keywords: Chemoprints, pharmacophore, biligands, PLP-dependent enzymes, drug targets

Introduction

Pyridoxal 5′-phosphate (PLP) is the biologically active form of vitamin B6 acting as coenzyme for enzymes that metabolize amines, amino and keto acids. The key feature of PLP-dependent enzymes is the catalytic versatility, as they are able to carry out transamination, decarboxylation, β- or γ-elimination and replacement[1]. More than 140 different enzyme activities are based on PLP as classified by the Enzyme Commission of IUMBM[2]. Despite their catalytic versatility, all structurally characterized PLP-dependent enzymes belong to five distinct structural groups, which correspond to five independent evolutionary lineages[1,3]. Fold type I is the most common and is found in a variety of aminotransferases and decarboxylases, as well as in enzymes that catalyse α-, β- or γ-eliminations. Fold type II is found mainly in enzymes that catalyse β-replacement and elimination reactions, such as tryptophan synthase. Fold type III is found in alanine racemase and in a subset of amino acid decarboxylases. Fold type IV enzymes include D-alanine aminotransferase and a few other enzymes. Fold type V includes glycogen and starch phosphorylases where PLP acts as acid-base catalyst, differently from the other fold type enzymes. The common features of the five fold types are: (i) the PLP that binds covalently through a Schiff-base linkage to the ε-amino group of the side-chain of an active-site lysine residue, and (ii) the location of the phosphate group of PLP near the *N*-terminus of an α-helix (the "anchoring" α-helix[4]). These features determine the coenzyme binding site and the adjacent substrate binding site.

As a consequence of the metabolic relevance of PLP-dependent enzymes, some of them are drug targets[5], such as DOPA decarboxylase for the treatment of Parkinson

Address for Correspondence: Andrea Mozzarelli, Department of Biochemistry and Molecular Biology, University of Parma, Viale GP Usberti 23/A, 43100 Parma, Italy. Email: andrea.mozzarelli@unipr.it; Alessandro Paiardini, Department of Biochemical Sciences, University of Rome "La Sapienza", Rome, P.le Aldo Moro, Italy. Email: alessandro.paiardini@uniroma1.it

(Received 24 October 2011; revised 18 November 2011; accepted 18 November 2011)

Abbreviations

AT, aminotransferase
CGS, cystathionine γ-synthase
CGL, cystathionine γ-lyase
MGL, methionine γ-lyase
HPAT, histidonol phosphate aminotransferase
PLP, pyridoxal 5′-phosphate
DOPA, dopamine
D-AT, D-amino acid aminotransferase
GABA, γ aminobutyric acid
MSA, multiple sequence alignment
RMSD, root mean square deviation

disease[6] and GABA aminotransferase for the treatment of epilepsy[7]. Many other PLP-dependent enzymes are considered potential drug targets, such as ornithine decarboxylase[8], serine racemase[9], kynurenine transaminase[10] and cysteine synthase[11]. Given the common reactivity of PLP with amine-containing compounds, target specificity is a key issue in order to avoid toxicity effects.

Towards the identification of tailored ligands for PLP-dependent enzymes, we have applied a cofactor-based chemogenomics approach[12,13], previously exploited for the NAD(P)-dependent enzyme family[14]. The presence of a common site where NAD(P) binds allowed to generate a library of compounds that target the coenzyme binding site within a subfamily of NAD-oxidoreductases. In a second step, the coenzyme-mimetic scaffold was coupled to substrate-mimetic scaffolds, leading to highly specific biligands[15]. In the present work, a common coenzyme pharmacophore model was generated via structure and sequence comparisons of the PLP binding sites. Subsequently, bifunctional pharmacophore models were generated to target unique PLP-dependent enzymes by coupling the coenzyme and substrate binding sites pharmacophores. Compounds that mimic either PLP or PLP-substrates are expected to compete (i) at the stage of apo-enzyme, with the coenzyme during protein folding, and (ii) at the stage of holo-enzyme, with the coenzyme derivatives that are formed during catalysis.

Materials and methods

Structure database

A database containing three-dimensional structures of PLP-dependent enzymes belonging to fold types I-IV was built. Using the classification found in several structural databases, SCOP[16], CATH[17] and MMDB[18], a total of 683 PLP-dependent crystallographic structures were retrieved from the Protein Data Bank[19]. From these structures, 65 representative members were selected on the basis of a hierarchical set of criteria: (i) engineered enzymes bearing residue mutations were discarded, (ii) in the presence of orthologous enzymes, the structure with the highest resolution was selected. Among the 65 retrieved structures, 49 belong to fold type I, 9 to fold type II, 4 to fold type III and 3 to fold type IV (Table S1).

Sequence database

Orthologous sequences were retrieved for each of the 65 structures from NCBI database. The number of orthologous sequences for each enzyme depends on their availability from different lineages. Redundant sequences were not included (identity > 90%). The number of orthologous sequences retrieved for each protein and minimum pairwise sequence identity between them is given in Table S1.

Sequence alignments

Multiple sequence alignments (MSA) were carried out between orthologous sequences using CLUSTALW[20], and manually checked to optimize the quality of the final alignment. Starting from these alignments, the program CAMPO[21] was used to assess the evolutionary conservation degree for each residue of a protein. The algorithm implemented in CAMPO assigns a score to each column of a multiple sequence alignment through the application of a user-defined mutational matrix and incorporates a weight based on the percentage of sequence identity between compared proteins.

Structure alignments

3D alignments of protein structures were carried out using EBI-SSM server[22,23], available at http://www.ebi.ac.uk/msd-srv/ssm/. Pairwise root mean deviations (RMSDs) values between structures were determined.

Pharmacophore models

Pharmacophore models were generated using the FLAP software[24] developed by Molecular Discovery Ltd (www.moldiscovery.com). The coenzyme and ligand binding sites were identified by the flapsite tool, while the GRID algorithm[25] was used to investigate the corresponding pharmacophoric space. The DRY probe was used to describe the potential hydrophobic interactions, using the sp2 carbonyl oxygen (O) and the neutral flat amino (N1) probes for the hydrogen bond donor and acceptor capacity of the target, respectively.

Results

Identification of evolutionarily conserved residues

One of the major challenges in protein science is the identification of the residues mediating function and regulation, such as residues involved in substrate and ligand binding, protein–protein interaction, catalysis, and structural roles. Various methods have been applied for the quantitative prediction of the relative importance of these residues, including scoring strict conservation, property conservation, entropy of a position or scoring conservation in related families[26]. Alignment and evolutionary

residue scoring methods assume that the relevance of a residue is reflected in its evolutionary conservation. The more important the residue, the sooner it becomes conserved in different evolutionary branches. Accordingly, the initial part of the present work was focused on aligning and ranking the residues of PLP-dependent enzymes that play the same function in different organisms. These enzymes are defined "orthologous". To implement this concept, MSAs were carried out between orthologous sequences for each enzyme contained in the structural database. Higher scores were attributed by CAMPO[21] to the more evolutionarily conserved residues. The conservation map, reported on the structure of each enzyme, allows visualizing the more conserved portions of the enzyme, suggesting structural/functional/regulatory roles. To demonstrate the procedure and information retrieved from this analysis, an example is presented from a fold type IV enzyme, *Bacillus sp. ym-1* D-amino acid aminotransferase (D-AT, EC 2.6.1.21[27]). This enzyme catalyzes the transamination between various D-amino acids and the corresponding α-keto acids. The enzyme is vital for bacteria because it provides an important route for the synthesis of the essential bacterial cell wall components D-alanine and D-glutamate, as well as other D-amino acids. Thus, D-AT is a target enzyme for the development of novel antimicrobial agents[5]. For determining the evolutionarily conserved residues of the B*acillus* D-AT, 25 orthologous sequences of D-AT were retrieved and aligned. The minimum pairwise identity that was fixed as threshold to include a sequence in the group of orthologous was 25%. The MSA was carried out between the resulting sequences and the subsequent conservation score was calculated and mapped on the D-AT structure (Figure 1). The invariant residues at PLP and substrate binding sites between orthologous sequences were found to be Tyr31, Glu32, His47, Arg50, Arg98, Arg138, Lys145, Leu149, Glu177 and Leu201 where this amino acid numbering is referred to DA-T from *Bacillus sp. ym-1*[27].

Similar analyses were carried out on the other 64 structures of PLP proteins. A dataset was prepared, containing the information on the conserved residues at substrate and coenzyme binding site of specific target available on request to the corresponding authors.

Structural analysis of coenzyme binding site similarity

In order to develop inhibitors that recognize the PLP binding site, the structural similarity of coenzyme sites and the class of residues interacting with PLP were determined for the enzymes contained in our dataset. These features define the PLP "chemoprint", and represent the functional/structural basis for PLP-based pharmacophores. We also investigated which features are in common to all PLP-dependent enzymes and which ones are fold type-dependent.

To cluster members of our dataset according to coenzyme binding site similarity, three-dimensional structures of enzymes belonging to the same fold type were superposed. Fold type I contains 49 structures that are evolutionarily too distant to be structurally compared all at once. Hence, the fold type I structures were first subdivided according to their reaction type, i.e. aminotransferases, decarboxylases and lyases. Then, their active sites were structurally compared to find coenzyme binding site similarity (Table 2). For aminotrasferases, the structures were further grouped according to Root Mean Square Deviation (RMSD) values, clustering them in three groups with RMSD values ≤3.0 Å (Table S2). We have preferred to compare PLP enzymes belonging to fold type I on the basis of reaction type and not on the basis of fold sub-type because in this way we have explored potential common structural features linked to catalytic function/mechanism.

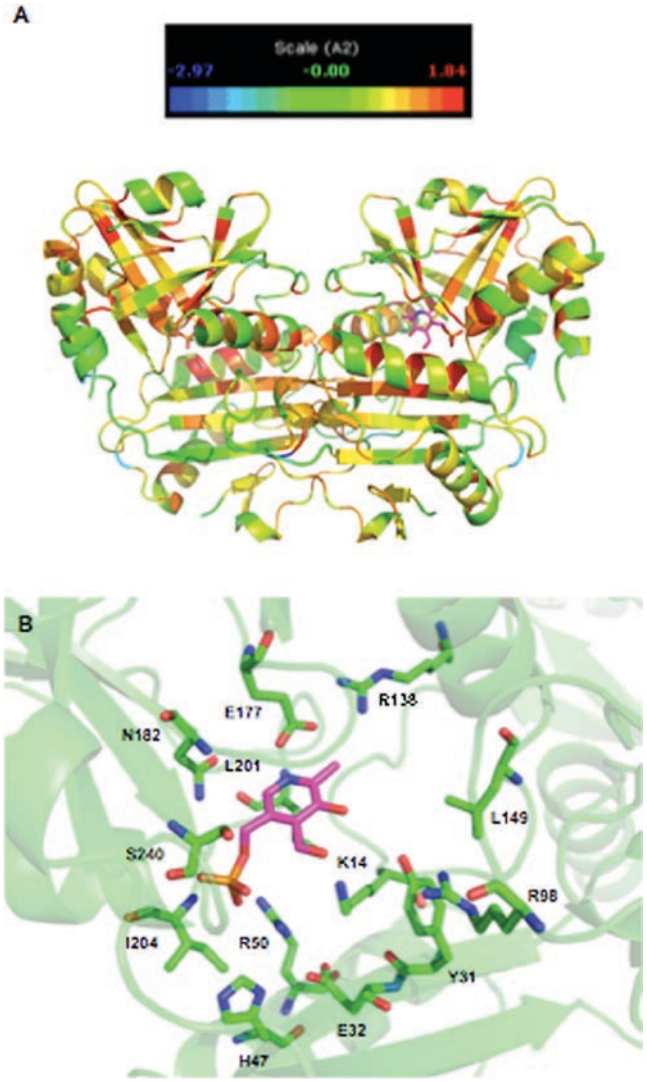

Figure 1. (A) Structure of *Bacillus sp. ym-1* D-AT (PDB code: 1DAA) with residues colored according to the conservation score, as calculated by CAMPO[21], expressed as standard deviations from the mean score. Residues in red and blue are the most and the least conserved ones, respectively. PLP is shown in pink. (B) Close-up view of D-AT active site showing residues characterized by a conservation score higher than 1 (Table 1). (See colour version of this figure online at www.informahealthcare.com/enz)

Table 1. Conservation score of active site residues in D-amino acid aminotransferase.

Active site	
Residue	Conservation score
TYR31	1.84
GLU32	1.69
VAL33	0.97
HIS47	1.84
ARG50	1.84
ARG98	1.84
HIS100	0.61
ARG138	1.84
CYS142	0.18
LYS145	1.84
LEU149	1.84
ALA152	-0.14
GLU177	1.84
GLY178	0.35
SER179	0.42
SER180	0.45
SER181	0.49
ASN182	1.84
LEU201	1.84
ILE204	1.48
THR205	0.47
ARG206	0.86
SER240	1.62
THR241	0.05
THR242	0.60

The structural comparison for enzymes belonging to fold types II, III and IV indicated that within these folds structures are more homogenous than within fold type I (Table S3).

Fold type I

Results of the structural superposition for fold type I enzymes are reported below, separated for reaction specificity:

Decarboxylases There are six decarboxylases in this cluster. L-Threonine-O3-phosphate decarboxylase and 2-2, dialkylglycine decarboxylase were not included in this cluster, since the identity of the residues at their active sites is closer to aminotransferase cluster 1 and 2 (see below), respectively, than to the decarboxylases. The analysis of the structural superposition led to the following results:

1. Two highly conserved His residues are always present at the binding site. One His is involved in a stacking interaction with PLP pyridine ring, whereas the other is at the phosphate binding site and involved in cofactor binding.
2. A D-X-A motif (Asp-variable residue-Ala) surrounds the PLP moiety. Asp interacts with the pyridine nitrogen, while Ala contributes to define the hydrophobic patch superimposed to the PLP pyridine ring, thus sandwiched between His and Ala residues.
3. Ser or Thr are evolutionarily conserved residues interacting with the 3′O of PLP.

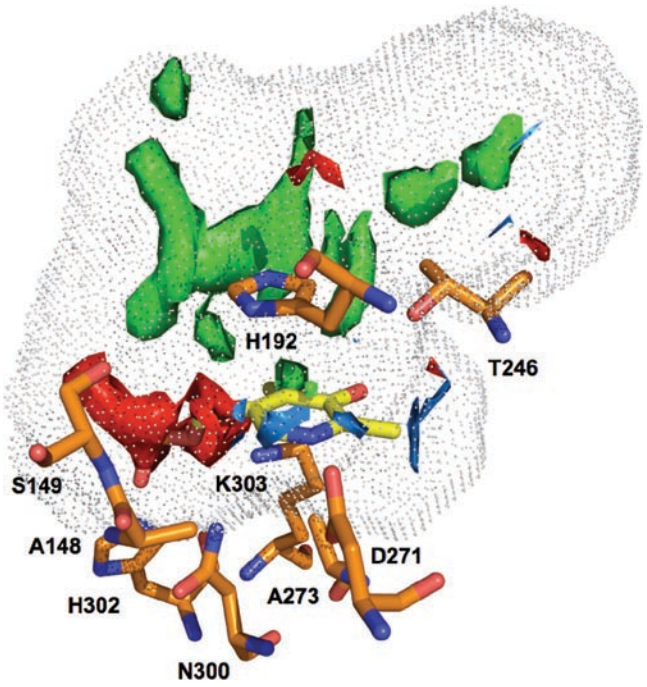

Figure 2. Decarboxylase pharmacophoric map derived from DOPA decarboxylase structure (PDB code: 1JS3), generated with the GRID algorithm[24]. Green, red and blue contours identify regions sterically and energetically favourable for hydrophobic, H-bond donors and H-bond acceptors groups, respectively. The dotted surface defines the accessible volume of the pocket. (See colour version of this figure online at www.informahealthcare.com/enz)

Applying the GRID algorithm[25], a representative pharmacophoric map for decarboxylases was generated exploring the active and PLP sites of DOPA decarboxylase (Figure 2).

Aminotransferases Nineteen aminotransferases belonging to fold type I are present in the database. Aminotransferase structures were divided into three clusters on the basis of the RMSD (Table S2) of the superposed structures. The structure of phosphoserine aminotransferase[28] was discarded since it exhibits an RMSD greater than 3.0 Å, compared to all the other structures of the dataset. However, a visual comparison indicates that its active site displays some similarity with the structures belonging to cluster "3" (see below).

Based on the structural alignment, similarities between clustered structures at the cofactor binding site have been identified (Table 2). These are summarized as follows:

1. An aromatic residue (Tyr, Trp or Phe) is located above the PLP pyridine ring in all of the aminotransferase structures. In cluster "1" and "3" (Figures 3A and 3C, respectively) the aromatic residue is involved in a stacking interaction with the pyridine ring, whereas in cluster "2" (Figure 3B) the aromatic residue is tilted of ~90 degrees, thus excluding any stacking interaction.

Table 2. Chemoprints for cofactor recognition in fold type I enzymes.

Structure	PDB	N	3′O	Pyridine	O1P, O2P, O3P
Aminotransferase cluster1					
Aspartate aminotransferase	1MAP	D222	N194, Y225	W140, A224	Y70, R266, S107, G108, T109, S255.
N-succinyldiaminopimelate aminotransferase	2OOR	D198	N170, Y201	Y119, V200	Y57, K240, G93, A94, T95, S229
Histidinol-phosphate aminotransferase	3CQ5	D197	N172, Y200	Y123, A199	Y63, R236, G97, S98, N99, T225, S227
LL-diaminopimelate aminotransferase	3EI6	D237	N209, Y240	Y152, A237	Y94, R278, G127, A128, L129, S267
Aromatic amino acid aminotransferase	2AY8	D222	N194, Y225	W140, A224	Y70, R266, G107, G108, T109, S255, S257
Tyrosine aminotransferase	1BW0	D216	N188, Y219	F138, I218	Y71, R261, G112, G113, S114
Kynurenine aminotransferase-I	1W7M	D213	N185, Y216	F125, V215	Y63, L225, G99, G100, Y101, S244
Kynurenine aminotransferase-II	2R2N	D230	N202, Y233	Y142, P232	Y74, R270, G116, S117, Q118, S260, S262
Alanine aminotransferase	1XI9	D205	N177, Y208	Y127, I207	Y66, R245, A101, T103, V102, S236
Glutamine aminotransferase	1V2F	D191	N163, Y194	F112, V193	Y57, R230, G86, A87, T88, S219
α-aminodipate aminotransferase	2Z1Y	D202	N194, Y205	Y125, A204	Y70, R245, G99, S100, Q101, S235, S237
L-Threonine-O-3-phosphate decarboxylase	1LC8	D185	N157	F108, A187	Y56, R224, E85, T86, S213, T215
Chemoprints*		**Acc**	**Don, Don/Acc**	**Ar, Hyd**	**Don/Acc, Don, Don, Don, Don/Acc**
Aminotransferase cluster2					
4-aminobutyrate aminotransferase	1OHW	D298	Q301	F189, V300	T353, C135, G136, S137
Acetyl ornithine aminotransferase	1VEF	D225	Q228,E197	F140, I227	T283, S112, G113, T114
Ornithine aminotransferase	2OAT	D263	Q266	F177, I265	T322, T141, G142, V143
Lysine aminotransferase	2CJH	D271	Q274	F167, V273	T330, G128, A129, T330
7, 8-Diaminopelargonic acid synthase	1MLY	D245	A217	Y144, I247	T309, S111, G112, S113
2,2-Dialkylglycine decarboxylase	1D7U	D243	Q246	W138, A245	T303,G111, A112
Chemoprints*		**Acc**	**Don**	**Ar, Hyd**	**Don/Acc, Don, Don, Don**
Aminotransferase cluster3					
Alanine glyoxylate aminotransferase	1H0C	D183	S158	W108, V185	T263, Y260, S81, G82, H83
3-Hydroxykynurenine aminotransferase	2CH1	D179	S154	W104, V181	T259, Y256, S77, A78, H79
Aminoethyphosphonate transaminase	1M32	D167	S170, T142	Y91, M169	T242, S64, G65, S66
Phosphoserine aminotransferase	1BJO	D174	T153	W102	T240, N239, Q197, A76, R77
Chemoprints*		**Acc**	**Don/Acc**	**Ar, Hyd**	**Don/Acc, Don, Don, Don**
Decarboxylase					
DOPA decarboxylase	1JS3	D271	T246	H192, A273	H302, G354, A148, S149, N300
Cysteine sulfinic acid decarboxylase	2JIS	D273	T248	H191, A275	H304, G357, G152, S153, N302
Glutamate decarboxylase	2OKJ	D373	T348	H291, A375	H404, G456, G252, A253, N402
L-Tyrosine decarboxylase	3F9T	D206	T181	H132, A208	H244, T285, G94, T95, D242
Glycine decarboxylase	1WYU	D236	S73	H166,A241	H265, T321, G132, A133, N263
Ornithine decarboxylase	1ORD	D316	W319	H283, A318	H354, S352, S198, S199, S396
Chemoprints*		**Acc**	**Don/Acc**	**Ar, Hyd**	**Don, Don, Don, Don, Don/Acc**
Lyase					
NifS CsdB	1JF9	D200	Q203	H123, A202	H225,THR278,T94,T95,S223
cysteine desulfurase	1T3I	D205	Q208	H128, C207	H230,T282,A99,T100,S228
cystine C-S lyase	1ELU	D197	Q200	H114, A199	H222, T276, V88, T89, T220
Chemoprints* (similarity with decarboxylase)		**Acc**	**Don**	**Ar, Hyd**	**Don, Don, Don, Don, Don/Acc**
Cystathionine γ lyase	3COG	D187, T189	N161	Y114, T189, S209	Y60, R62, G90, L91, T211
Methionine γ lyase	1E5F	D184, T186	N158	Y111, T186, S206	Y56, R58, G86, M87, T208
O-Acetyl homoserine sulfhydrylase	2CB1	D176, T178	N151	F105, T178, S199	Y50, R52, G80, G81, S199
Cystathionine γ synthase	1CS1	D173, T175	T175	Y101, T175, S179	Y46, R48, G76, M77, T197
Cystathionine β lyase	1CL2	D185, T187	W340, T187	Y111, T87, A207	Y56, R58, G86, A87, T209
Tyrosine phenol lyase	2VLF	D214, E103	N185, R217	F123, S254	Y71, R100, Q98, G99, S254
Tyrptophanase	1AX4	D223, E104	N194, R226	F132, A225	Y72, R101, Q99, G100, S263

(*Continued*)

Table 2. Continued.

Chemoprints* (similarity with aminotransferase)		**Acc**	**Don**	**Ar, Hyd**	**Don/Acc, Don, Don, Don, Don/Acc**
Others					
L-Threonine aldolase	1LW5	D168	N139, R171	H83, A170	R231, G58, T59, S207
Phenylserine aldolase	1V72	D179	S181, R182	H93, S181	Y39, K237, G68, T69, Q219
2-Amino-3-ketobutyrate-CoA ligase	1FC4	D210, S212	H213, S185	H136	N275, S274, S111, F112, T241
3-Amino-5-hydroxybenzoic acid synthase	1B9I	D159, S91	H162	F88, A161	R236, N234, G62, T63, S183
5-aminolevulinate synthase	2BWP	D214	H217, S189	H142, V216	S277, T278, A115, Y116, T245
Kynureninase	3E9K	D178, D250	H253	F165, L137	Y275, N333, S332, T138
8-amino-7-oxononanoate synthase	1DJ9	D204	H207, S179	H133, A206	S265, T266, G108, F109, T233
Glutamate-1-semialdehyde	2E7U	D238	N211	Y144, V240	T298, G117, T118
Serine hydroxymethyltransferase	1BJ4	D238	H231, S203	H148, A230	H256, GLY120, S121, S119, T254
Allinase	2HOR	D225	N207, Y228	Y165, V227	Y92, R259, V132, T133, T248, SER250
ACC synthase	1M7Y	D230	T233	Y145, I282	R281, G119, A120, T121, S272, S270

*Hyd, hydrophobic; Ar,aromatic; Don, donor; Acc, Acceptor; Don/Acc-donor/acceptor.

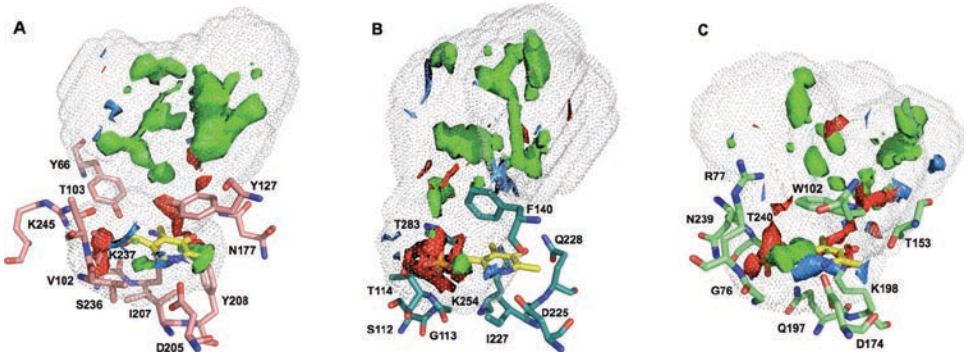

Figure 3. Pharmacophoric maps of the active and PLP binding sites of three representative members of aminotransferase clusters "1", "2" and "3". (A). Alanine aminotransferase (PDB code: 1XI9); (B). Acetyl ornithine aminotransferase (PDB code: 1VEF); (C). Phosphoserine aminotransferase (PDB code: 1BJO). The contours and the cavity profile were generated by GRID. Green, red and blue contours identify regions sterically and energetically favourable for hydrophobic, H-bond donors and H-bond acceptors groups, respectively. The dotted surface defines the accessible volume of the pocket. (See colour version of this figure online at www.informahealthcare.com/enz)

2. A D-X-[A,V,I] motif is present at the pyridine ring side. Asp is a conserved residue interacting with the pyridine nitrogen, whereas the apolar residues interact with the *si* side of the PLP ring, as defined on the basis of the prochirality of C4′, thus sandwiched between the conserved aromatic and apolar residues side chains, as displayed in Figure 3. The only exceptions to this sequence pattern are represented by kynurenine aminotransferase and 2-aminoethylphosphonate aminotransferase, where D-X-P and D-X-M motifs were found, respectively.
3. The 3-O′ of PLP interacts with Asn and Tyr in cluster "1", Gln in cluster "2", with the exception of 7,8-diaminopelargonic aminotransferase (Ala or Glu), and Ser or Thr in cluster "3". Tyr plays a dual role because it is involved in cofactor binding, as well as in catalysis in some of the enzymes[29]. In cluster "2" and "3" Tyr is replaced by Thr. In cluster "1", at the phosphate binding site, Arg or Lys are conserved residues.

Applying the GRID algorithm[25], representative pharmacophoric maps for the three aminotransferase clusters were generated exploring the active and PLP sites of alanine aminotransferase (cluster "1") (Figure 3A), acetyl ornithine aminotransferase (cluster "2") (Figure 3B) and phosphoserine aminotransferase (cluster "3") (Figure 3C).

Lyases The superposition of the PLP binding site of ten lyase structures revealed that in some of them the cofactor binding site resembles that of decarboxylases, while in some others resemble that of aminotransferases. For example, in cysteine desulfurase[30], selenocysteine lyase[31] and C-S lyase[32] structures, apart from the interaction of Asp with PLP, the neighbouring residues are conserved and display a rearrangement analogous to the one typical of decarboxylase structures: two His residues at the active site of lyases are localized in the same position of His residues of decarboxylase coenzyme binding sites and, similarly, one His is involved in a stacking interaction with the pyridine ring and another His residue interacts with the phosphate moiety of PLP. Again, the PLP pyridine ring is allocated between a histidine and an alanine or cysteine residue. On the contrary, other structures of lyase i.e. cystathionine β lyase[33], cystathionine γ synthase[34], cystathionine γ lyase[35], methionine γ lyase[36], *O*-acetylhomoserine sulfhydrylase[37], tyrosine phenol lyase[38] and tryptophanase[39], were found to be structurally

Table 3. Chemoprints for cofactor recognition fold type in II, III, IV enzymes.

Structure	PDB	N	3′O	Pyridine	O1P, O2P, O3P
Fold type II					
Tryptophan synthase	1TJP	S377	Q114, E350	H86, G303, L304	G232, G233, G234, S235, N236
Threonine synthase	1V7C	T317	N87	F60, A240, I241	G187, N188, A189, G190, N191
Serine racemase	1V71	S308	N84	F56, G236, A237	G183, G184, G185, G186, L187
Threonine deaminase	1VE5	S303, T279	N75	F50, G232, V233	G178, G179, G180, G181, L182
Serine dehydratase	1PWH	S166	N67	F40, A222, L223	G168, G169, G170, G171, L172
Cystheine synthase-B	2BHS	S255	N71	V40, G208, I209	G174, T175, T176, G177, T178
Cysteine synthase-A	1Y71	S272	N72	V41, P299, G228, I229	G177, T178, G179, G180, S181
Cystathionine β-synthase	1JBQ	S349	N149	V118, P375, G305, I306	G256, T257, G258, G259, T260
ACC deaminase	1J0B	T308	N82	Y282, T308	V201, T202, G203, S204, T205
Chemoprints*		**Don/Acc**	**Don/Acc**	**Hyd, Hyd**	**Don, Don, Don, Don, Don**
Fold type III					
Ornithine decarboxylase	1D7K	E274,H197,D88	C360,R154	H197,A67	Y389, R277, G276, G237
Arginine decarboxylase	2NVA	E252,D67	C324,R133	H176,A46	Y353, R255, G216, G180, S179
Diaminopimelate decarboxylase	1KO0	E268,H191	C342,R142	H191,A52	Y378,R271,G270, G227
Alanine racemase	1SFT	R219,HIS166	R136	H166,V37	Y43, Y374, I222, G221, S204
Chemoprints*		**Don/Acc**	**Don/Acc, Don**	**Ar, Hyd**	**Don/Acc, Don, Don, Don, Don**
Fold Type IV					
Aminodeoxychorismate lyase	1I2K	E173, A177	Y92	V197, A176	R45, I200, M201, A237
Branched-chain amino acid aminotransferase	1IYE	E193, E197	Y164	V217, G196	R59, I220, T221, T257
DAA aminotransferase	1DAA	E177, S181	Y21, S179	L201, S180	R50, I204, T205, T241
Chemoprints*		**Acc**	**Don/Acc**	**Hyd, Hyd**	**Don, Don, Don, Don/Acc**

*Hyd, hydrophobic; Ar, aromatic; Don, donor; Acc, Acceptor; Don/Acc, donor/acceptor.

more similar to cluster "1" aminotransferase, containing identical residues for cofactor recognition (Table 2). The pharmacophoric map was generated using the GRID algorithm applied to the substrate and PLP sites of cysteine desulfurase (Figure 4).

Inter- and intra-clusters comparisons of the coenzyme binding site led to define different chemoprints for the cofactor binding (Table 2). Since the structures of fold type I enzymes are highly divergent and the spatial position of the identified common features especially around phosphate binding group are not identical among the different structures, cluster-specific pharmacophoric models were defined in order to represent the conserved feature of the cofactor recognition and approximated position of the sites around PLP. By comparing the distinct pharmacophores of fold type I enzymes, the common pharmacophore model includes one acceptor, one donor/acceptor, one hydrophobic site, one aromatic site and three-four donors matching the chemical features of PLP.

Fold type II, III, IV

Nine enzymes belonging to fold type II, four to fold type III and three to fold type IV are present in our dataset.

Fold type II

Residues involved in cofactor recognition for fold type II structures are listed in Table 3. The interesting feature, that is peculiar of fold type II, is the conservation of a motif [ST]-X3-G-[NQ] in eight structures, excluding 1-aminocyclopropane-1-carboxylate deaminase. The residues from this motif are involved in the interaction with 3′O of PLP. This generates a specific cofactor binding site model for fold type II enzymes. As a representative structure of fold type II, cysteine synthase-A was used to generate the pharmacophore model (Figure 5). The proposed model is composed of three donor/acceptor sites, two hydrophobic and six to seven donor sites, as reported in Table 3.

Fold type III

In fold type III, three decarboxylases and a racemase were structurally aligned. Ornithine decarboxylase was selected as the representative structure to generate a common pharmacophore model, which includes one hydrophobic, one aromatic, three donor/acceptor and six donors sites (Figure 6, Table 3).

Fold type IV

In fold type IV, three structures were superimposed. D-amino acid aminotransferase was selected as the representative structure for model generation. The pharmacophoric model (Figure 7) includes an acceptor, two hydrophobic residue, two donor/acceptor and four donor sites (Table 3).

Pharmacophores models to target specificity: the biligand approach The comparison of the pharmacophores among different fold types allows defining the few features representing the conserved chemoprint for PLP recognition. This is composed of two donors/acceptors, and two hydrophobic sites sandwiching the pyridine ring, three donors around phosphate group of PLP, and

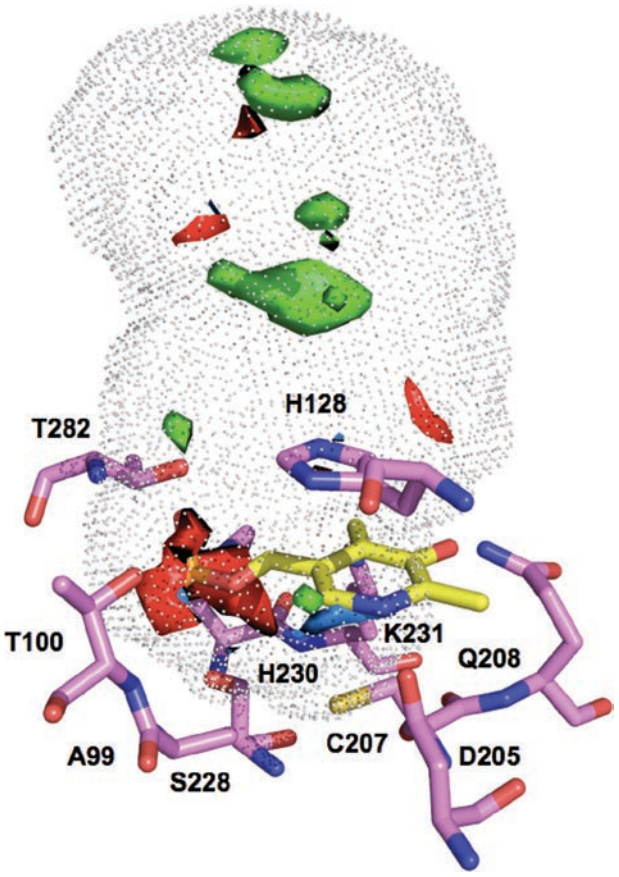

Figure 4. Lyase pharmacophoric map derived from the cysteine desulfurase structure (PDB code: 1T3I) and generated with the GRID algorithm. Green, red and blue contours identify regions sterically and energetically favourable for hydrophobic, H-bond donors and H-bond acceptors groups, respectively. The dotted surface defines the accessible volume of the pocket. (See colour version of this figure online at www.informahealthcare.com/enz)

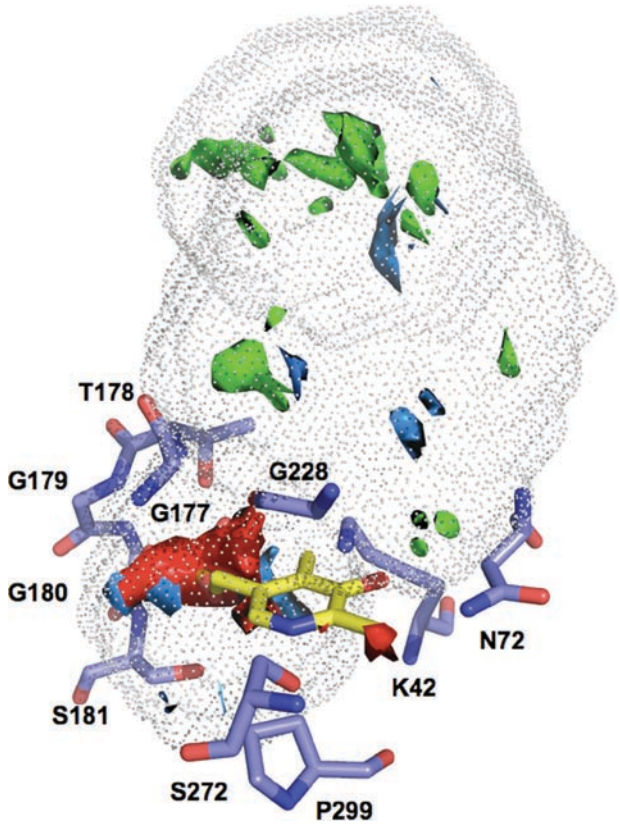

Figure 5. Fold type II pharmacophoric map derived from the cysteine synthase-A structure (PDB code: 1Y7L). Green, red and blue contours identify regions sterically and energetically favourable for hydrophobic, H-bond donors and H-bond acceptors groups, respectively. The dotted surface defines the accessible volume of the pocket. (See colour version of this figure online at www.informahealthcare.com/enz)

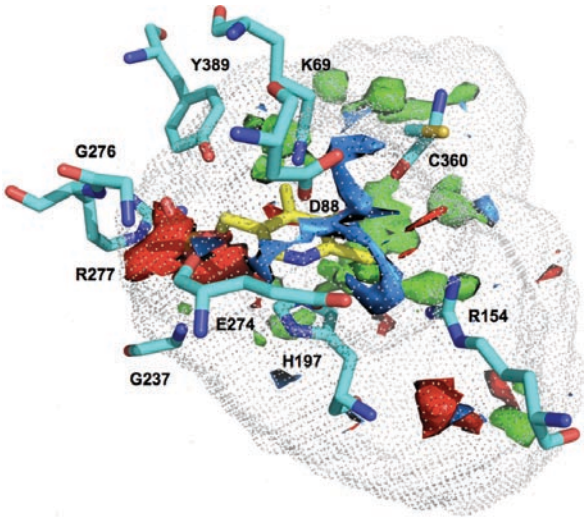

Figure 6. Fold type III pharmacophoric map derived from the ornithine decarboxylase structure (PDB code:1D7K). Green, red and blue contours identify regions sterically and energetically favourable for hydrophobic, H-bond donors and H-bond acceptors groups, respectively. The dotted surface defines the accessible volume of the pocket. (See colour version of this figure online at www.informahealthcare.com/enz)

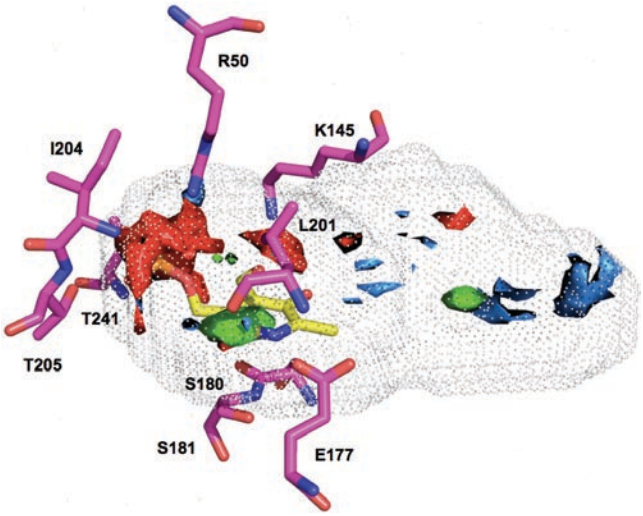

Figure 7. Fold type IV pharmacophoric map derived from the D-amino acid aminotransferase structure (PDB code:1DAA). Green, red and blue contours identify regions sterically and energetically favourable for hydrophobic, H-bond donors and H-bond acceptors groups, respectively. The dotted surface defines the accessible volume of the pocket. (See colour version of this figure online at www.informahealthcare.com/enz)

Table 4. Unique conserved residues mapped at substrate binding site.

PLP proteins	PDB	Unique conserved residues
Histidinol-phosphate aminotransferase	3CQ5	Y21, N35, N99, P256, R333
LL-Diaminopimelate aminotransferase	3EI6	Y37, K129, N309, T390,
7,8-Ddiamino-pelargonic-acid aminotransferase	1MLY	W53,D147,P308,M376,P392,F393
2-Aminoethylphosphonate transaminase	1M32	R240, Y328, P329
3-Hydroxykynurenine transaminase	2CH2	S43, N44,F45, Q344,
L-tyrosine decarboxylase	3F9T	C38, E182,S362,R371
Tryptophan synthase	1TJP	E109,A112,D305,F306,
Cysteine synthase B	2BHS	R99,R210,
O-Acetylserine sulfhydrylase	1Y7L	R100,H224,Q227s
Diaminopimelate decarboxylase	1KO0	H164,S344,Y386,
Alanine racemase	1SFT	Y265,M312,R136
D-amino acid aminotransferase	1DAA	Y31,V33,H100,T242

one donor from the conserved Lys residue that binds the coenzyme. This common pharmacophoric model identifies all the necessary chemical features required for PLP binding, and, hence, it defines a potential scaffold able to bind to the cofactor site.

After having defined a common cofactor pharmacophoric model to all fold-types enzymes, we have implemented target specificity by (i) identifying the unique conserved residues at the substrate binding site of desired targets, (ii) generating pharmacophores that are target-specific, and (iii) linking these target-specific pharmacophores to the common cofactor pharmacophore. Two conditions have been used to select the target-specific residues at substrate binding site: (a) residues should be conserved in orthologous proteins; (b) residues should not be shared nor present at the same position in substrate binding site of the nearest homologous proteins.

To identify specific residues between the receptors catalyzing similar reaction or exhibiting similarity at substrate binding site, conserved residues at substrate binding site of given structures, identified from orthologous MSA analysis (Figure 1), were mapped onto the structural alignment of paralogous PLP enzymes. By the subtraction of common residues between homologous sequences, the unique conserved residues between orthologous sequences of a particular target are mapped at substrate binding site. On the basis of this orthologous and paralogous sequences-structures comparison, target specific residues were mapped for a selection of PLP-dependent enzymes, which are absent in humans (Table 4). Examples of this approach is provided and discussed below.

Discussion

The workflow of the present project is based on two independent sequence-structure analyses: (i) for each of the 65 representative structures of PLP-dependent enzymes, orthologous sequence alignments were carried out to identify conserved residues, and the degree of evolutionary conservation of each residue was scored and mapped on every structure composing the dataset; (ii) for each fold type, coenzyme and substrate binding sites were structurally compared and clustered. For fold type I, containing a significant number of structures, a sub-clustering approach was carried out for enzymes catalyzing the same reaction, i.e. decarboxylases, lyases and aminotransferases. The two sets of information were then coupled to generate pharmacophore models, bearing either both cofactor and substrate specificity aimed at targeting individual enzymes, or common features to the analyzed fold types. The resulting pharmacophore models can be exploited for chemogenomics-based virtual screening or *de novo* drug design. Identified or designed compounds mimicking PLP can lead, *in vivo*, to the replacement of the coenzyme during enzyme folding or displacement of the coenzyme upon enzyme folding. The net result is the loss of the function associated to the target enzyme(s).

The results obtained in the present study lead to two possible scenarios: (i) targeting specific fold types by using the unique pharmacophore model identified for a fold type; (ii) targeting all PLP enzymes by using the identified common pharmacophore model. Common scaffolds for PLP-dependent enzymes are a powerful strategy when targeting pathogenic microbes, simultaneously inhibiting different PLP-dependent enzymes. The major drawback of this approach is potential toxicity. Thus, to increase the selectivity of the scaffolds and, consequently, of the targeted enzyme, the variable substrate binding site adjacent to PLP binding site was exploited. Coupling a common PLP scaffold with a specificity scaffold, which is derived from the substrate active site pharmacophore, produces specific biligand pharmacophores. This design strategy has been already successfully applied to NAD+ dependent oxidoreductases enzymes[14,15]. The rationale behind this strategy is that a moiety of the biligand is common among a group of enzymes, which provide baseline affinity across a pharmaco-family and the other moiety of the biligand depends on the variable substrate binding site, providing target specificity. The latter can be considered at two levels: (1) targeting an enzyme that is present both in humans and pathogens, discriminating between them; (2) targeting an enzyme that is present either in humans or pathogens.

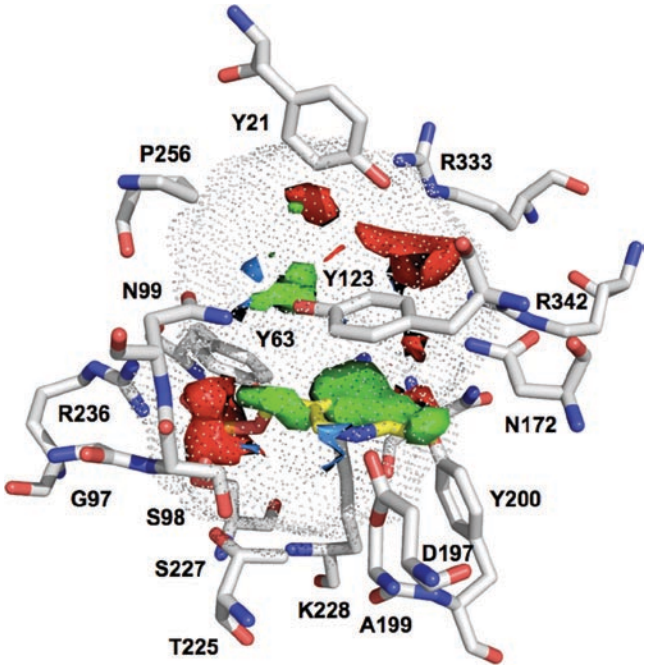

Figure 8. Cofactor and ligand binding sites corresponding to fold type I histidinol-phosphate aminotransferase (PDB code: 3CQ5). Residues lining the two cavities and interacting with the PLP cofactor are displayed in capped sticks. Green, red and blue contours identify regions sterically and energetically favourable for hydrophobic, H-bond donors and H-bond acceptors groups, respectively. The dotted surface defines the accessible volume of the pocket. (See colour version of this figure online at www.informahealthcare.com/enz)

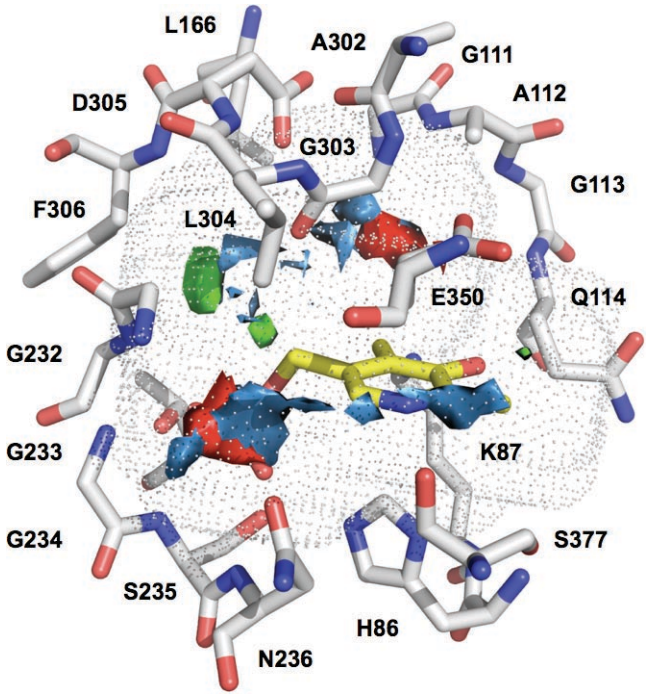

Figure 9. Cofactor and ligand binding sites corresponding to fold type II typtophan synthase (PDB code: 1TJP). Residues lining the two cavities and interacting with the PLP cofactor are displayed in capped sticks. Green, red and blue contours identify regions sterically and energetically favourable for hydrophobic, H-bond donors and H-bond acceptors groups, respectively. The dotted surface defines the accessible volume of the pocket. (See colour version of this figure online at www.informahealthcare.com/enz)

To achieve the first level of specificity, a pharmacophore was generated exploiting the non-conserved residues at the substrate binding site for a defined target. To achieve the second level of target specificity, only enzymes that are either absent in humans or in pathogens were selected for the generation of the pharmacophore.

This procedure is shown for two representative cases, histidinol aminotransferase from fold type I and tryptophan synthase from fold type II.

Histidinol-phosphate aminotransferase (HPAT)

HPAT is a PLP-dependent enzyme belonging to fold type I, catalyzing the reversible transamination between histidinol phosphate (His-P) and 2-oxoglutarate[40]. HPAT is only present in microbes, making it a potential target for a novel class of antibiotics. However, in spite of the absence of HPAT in human, the enzyme is homologous to other fold type I aminotransferases present in humans. To identify conserved residues in HPAT, 34 orthologous sequences were used, with sequence identity varying among them from 28% to 75%. Conserved residues identified at the substrate binding site from the MSA alignment of orthologous sequences are mapped on the structural alignment of homologous aminotransferase enzymes. The residues that are common in the substrate binding site are Tyr63, Gln172, Tyr200, Lys228, Arg236, and Arg342. The common residues between homologous structures were excluded and then residues uniquely conserved between orthologous sequences are identified and mapped in the enzyme active site. Tyr21, Gln35, Gln99, Pro256 and Arg333 are identified as residues that provide target specificity for this enzyme in any pathogen microbes. Consequently, a pharmacophoric map was generated (Figure 8).

Tryptophan synthase

Tryptophan synthase catalyzes the final two steps in the biosynthesis of L-tryptophan. The enzyme is found in a variety of bacteria, fungi, and plants and is absent in human[41,42]. For the identification of conserved residues between orthologous sequences of tryptophan synthase, 38 sequences were used with sequence identity varying among them from 28% to 82%. Conserved residues identified from orthologous sequences were mapped on the structural alignment of fold type II structures. By subtracting the common residues at substrate binding site of the other fold type II enzymes Glu109, Ala112, Phe280, Asp305, and Phe306 are identified as unique residues of tryptophan synthase. The corresponding pharmacophoric map is shown in Figure 9.

Conclusions

Identification of conserved residues is an important task in structure-based drug design, as well as the

identification of residues that are uniquely present in the active site of the target. The chemogenomics approach, previously applied to NAD-dependent enzymes and now tailored for PLP-dependent enzymes, is aimed at the generation of biligand target-specific scaffolds, based on the similarity at the cofactor binding sites and the diversity at the substrate binding sites. To this goal, common and diverse residues at binding sites were identified leading to chemoprints. Common three-dimensional pharmacophore models were generated from it for multiple targets, by exploiting common conserved residues at cofactor binding site of PLP proteins. Specific residues were identified in the substrate binding sites that account for target specificity.

An *in silico* screening of compound libraries using the pharmacophoric scaffolds of cofactor binding site might identify inhibitors designed to target either multiple or unique PLP enzymes, and is currently under way.

Acknowledgements

This work is partially supported by grants from the Italian Ministry of University and Research for (COFIN projects). Dr. G. Cruciani (Molecular Discovery Ltd, www.moldiscovery.com) has kindly provided the FLAP software.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

1. Christen P, Mehta PK. From cofactor to enzymes. The molecular evolution of pyridoxal-5′-phosphate-dependent enzymes. Chem Rec 2001;1:436–447.
2. Percudani R, Peracchi A. A genomic overview of pyridoxal-phosphate-dependent enzymes. EMBO Rep 2003;4:850–854.
3. Grishin NV, Phillips MA, Goldsmith EJ. Modeling of the spatial structure of eukaryotic ornithine decarboxylases. Protein Sci 1995;4:1291–1304.
4. Denesyuk AI, Denessiouk KA, Korpela T, Johnson MS. Functional attributes of the phosphate group binding cup of pyridoxal phosphate-dependent enzymes. J Mol Biol 2002;316:155–172.
5. Amadasi A, Bertoldi M, Contestabile R, Bettati S, Cellini B, di Salvo ML et al. Pyridoxal 5′-phosphate enzymes as targets for therapeutic agents. Curr Med Chem 2007;14:1291–1324.
6. Burkhard P, Dominici P, Borri-Voltattorni C, Jansonius JN, Malashkevich VN. Structural insight into Parkinson's disease treatment from drug-inhibited DOPA decarboxylase. Nat Struct Biol 2001;8:963–967.
7. Storici P, De Biase D, Bossa F, Bruno S, Mozzarelli A, Peneff C et al. Structures of γ-aminobutyric acid (GABA) aminotransferase, a pyridoxal 5′-phosphate, and [2Fe-2S] cluster-containing enzyme, complexed with γ-ethynyl-GABA and with the antiepilepsy drug vigabatrin. J Biol Chem 2004;279:363–373.
8. Heby O, Persson L, Rentala M. Targeting the polyamine biosynthetic enzymes: A promising approach to therapy of African sleeping sickness, Chagas' disease, and leishmaniasis. Amino Acids 2007;33:359–366.
9. Smith MA, Mack V, Ebneth A, Moraes I, Felicetti B, Wood M et al. The structure of mammalian serine racemase: Evidence for conformational changes upon inhibitor binding. J Biol Chem 2010;285:12873–12881.
10. Passera E, Campanini B, Rossi F, Casazza V, Rizzi M, Pellicciari R et al. Human kynurenine aminotransferase II–reactivity with substrates and inhibitors. FEBS J 2011;278:1882–1900.
11. Salsi E, Bayden AS, Spyrakis F, Amadasi A, Campanini B, Bettati S et al. Design of O-acetylserine sulfhydrylase inhibitors by mimicking nature. J Med Chem 2010;53:345–356.
12. Jacoby E, Mozzarelli A. Chemogenomic strategies to expand the bioactive chemical space. Curr Med Chem 2009;16:4374–4381.
13. Singh R, Mozzarelli A. Cofactor chemogenomics. Methods Mol Biol 2009;575:93–122.
14. Sem DS, Bertolaet B, Baker B, Chang E, Costache AD, Coutts S et al. Systems-based design of bi-ligand inhibitors of oxidoreductases: Filling the chemical proteomic toolbox. Chem Biol 2004;11:185–194.
15. Ge X, Wakim B, Sem DS. Chemical proteomics-based drug design: Target and antitarget fishing with a catechol-rhodanine privileged scaffold for NAD(P)(H) binding proteins. J Med Chem 2008;51:4571–4580.
16. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
17. Greene LH, Lewis TE, Addou S, Cuff A, Dallman T, Dibley M et al. The CATH domain structure database: New protocols and classification levels give a more comprehensive resource for exploring evolution. Nucleic Acids Res 2007;35:D291–D297.
18. Chen J, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S et al. MMDB: Entrez's 3D-structure database. Nucleic Acids Res 2003;31:474–477.
19. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
20. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–2948.
21. Paiardini A, Bossa F, Pascarella S. CAMPO, SCR_FIND and CHC_FIND: A suite of web tools for computational structural biology. Nucleic Acids Res 2005;33:W50–W55.
22. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr 2004;60:2256–2268.
23. Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. Bioinformatics 1999;15:327–332.
24. Cross S, Baroni M, Carosati E, Benedetti P, Clementi S. FLAP: GRID molecular interaction fields in virtual screening. validation using the DUD data set. J Chem Inf Model 2010;50:1442–1450.
25. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 1985;28:849–857.
26. Valdar WS. Scoring residue conservation. Proteins 2002;48:227–241.
27. Sugio S, Petsko GA, Manning JM, Soda K, Ringe D. Crystal structure of a D-amino acid aminotransferase: How the protein controls stereoselectivity. Biochemistry 1995;34:9661–9669.
28. Hester G, Stark W, Moser M, Kallen J, Markovic-Housley Z, Jansonius JN. Crystal structure of phosphoserine aminotransferase from Escherichia coli at 2.3 A resolution: Comparison of the unligated enzyme and a complex with α-methyl-l-glutamate. J Mol Biol 1999;286:829–850.
29. Inoue K, Kuramitsu S, Okamoto A, Hirotsu K, Higuchi T, Kagamiyama H. Site-directed mutagenesis of Escherichia coli aspartate aminotransferase: Role of Tyr70 in the catalytic processes. Biochemistry 1991;30:7796–7801.
30. Tirupati B, Vey JL, Drennan CL, Bollinger JM Jr. Kinetic and structural characterization of Slr0077/SufS, the essential cysteine desulfurase from Synechocystis sp. PCC 6803. Biochemistry 2004;43:12210–12219.

31. Lima CD. Analysis of the *E. coli* NifS CsdB protein at 2.0 Å reveals the structural basis for perselenide and persulfide intermediate formation. J Mol Biol 2002;315:1199–1208.
32. Clausen T, Kaiser JT, Steegborn C, Huber R, Kessler D. Crystal structure of the cystine C-S lyase from Synechocystis: Stabilization of cysteine persulfide for FeS cluster biosynthesis. Proc Natl Acad Sci USA 2000;97:3856–3861.
33. Clausen T, Huber R, Messerschmidt A, Pohlenz HD, Laber B. Slow-binding inhibition of Escherichia coli cystathionine β-lyase by L-aminoethoxyvinylglycine: A kinetic and X-ray study. Biochemistry 1997;36:12633–12643.
34. Clausen T, Huber R, Prade L, Wahl MC, Messerschmidt A. Crystal structure of Escherichia coli cystathionine γ-synthase at 1.5 Å resolution. EMBO J 1998;17:6827–6838.
35. Sun Q, Collins R, Huang S, Holmberg-Schiavone L, Anand GS, Tan CH et al. Structural basis for the inhibition mechanism of human cystathionine γ-lyase, an enzyme responsible for the production of H(2)S. J Biol Chem 2009;284:3076–3085.
36. Ronda L, Bazhulina NP, Morozova EA, Revtovich SV, Chekhov VO, Nikulin AD et al. Exploring methionine Y-lyase structure-function relationship via microspectrophotometry and X-ray crystallography. Biochim Biophys Acta 2011;1814:834–842.
37. Tran TH, Krishnamoorthy K, Begley TP, Ealick SE. A novel mechanism of sulfur transfer catalyzed by O-acetylhomoserine sulfhydrylase in the methionine-biosynthetic pathway of Wolinella succinogenes. Acta Crystallogr D Biol Crystallogr 2011;67:831–838.
38. Milic D, Demidkina TV, Faleev NG, Matkovic-Calogovic D, Antson AA. Insights into the catalytic mechanism of tyrosine phenol-lyase from X-ray structures of quinonoid intermediates. J Biol Chem 2008;283:29206–29214.
39. Isupov MN, Antson AA, Dodson EJ, Dodson GG, Dementieva IS, Zakomirdina LN et al. Crystal structure of tryptophanase. J Mol Biol 1998;276:603–623.
40. Haruyama K, Nakai T, Miyahara I, Hirotsu K, Mizuguchi H, Hayashi H et al. Structures of Escherichia coli histidinol-phosphate aminotransferase and its complexes with histidinol-phosphate and N-(5′-phosphopyridoxyl)-L-glutamate: Double substrate recognition of the enzyme. Biochemistry 2001;40:4633–4644.
41. Raboni S, Bettati S, Mozzarelli A. Tryptophan synthase: A mine for enzymologists. Cell Mol Life Sci 2009;66:2391–2403.
42. Dunn MF, Niks D, Ngo H, Barends TR, Schlichting I. Tryptophan synthase: The workings of a channeling nanomachine. Trends Biochem Sci 2008;33:254–264.