Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Catalysis B: Enzymatic 23 (2003) 17–27
JOURNAL OF MOLECULAR CATALYSIS B: ENZYMATIC
www.elsevier.com/locate/molcatb

Conserved structural modules and bonding networks in isopenicillin N synthase related non-haem iron-dependent oxygenases and oxidases☆

Janet Sim, Esther Wong, Hong Soon Chin, Tiow Suan Sim*

Department of Microbiology, Faculty of Medicine, National University of Singapore, 5 Science Drive 2, Singapore 117597, Singapore

Received 5 November 2002; received in revised form 3 March 2003; accepted 4 March 2003

Abstract

Analyses of attainable primary amino acid sequences, assigned secondary structures and superimposed tertiary structures of 142 isopenicillin N synthase (IPNS) related *non-haem iron-dependent oxygenases* and *oxidases* (designated as NHIDOX) were examined in this study. Despite having low sequence homologies (∼20%), these enzymes were found to possess conserved structural domains (delineated as modules I and II) that fold into jelly-roll motifs, juxtaposed by adjacent stabilizing elements. The seven highly conserved residues, corresponding to His214, Asp216 and His270 in *Aspergillus nidulans* IPNS (IPNS_AN) for iron binding, Arg279 and Ser281 for substrate/co-substrate binding, as well as Gly40 and Gly254 with yet undetermined functions, are arrayed closely within these conserved modules. Complex hydrogen bonding interactions of these conserved residues with residues found in specific α-helices and β-strands of the conserved core motif are apparently involved in stabilizing these structures. Although the NHIDOX enzymes appear to share conserved active center architecture, differences in their hydrogen bonding networks were observed, particularly those involving the substrate/co-substrate binding ligands and the two conserved glycines. These may modulate functional versatilities, relative sizes of the jelly-roll motifs, and the specific amino acid residues involved in stabilization and folding of the active center.

Keywords: Conserved residues; Conserved structural modules; Hydrogen bonding network; Non-haem iron-dependent oxygenases and oxidases

1. Introduction

In the multi-step β-lactam biosynthetic pathway carried out by various bacteria and fungi [1], three enzymes, isopenicillin N synthase (IPNS), deacetoxycephalosporin C synthase (DAOCS) and deacetylcephalosporin C synthase (DACS), are known to bind iron with a non-haem coordinated ligand environment and also concomitantly use oxygen and ascorbate for their catalysis [2,3]. Based on these properties and their amino acid sequence relatedness, IPNS, DAOCS and DACS have been classified under a subfamily of mononuclear non-haem iron-binding oxygenases and oxidases [2]. Most enzyme members within this subfamily are involved in catalyzing the biosyntheses of plant secondary metabolites. For example, flavanone 3β-hydroxylase (FL3OH), flavonol synthase (FS), leucoanthocyanidin

☆ Supplementary data associated with this article can be found at doi: 10.1016/S1381-1177(03)00037-7.

* Corresponding author. Tel.: +65-68743280; fax: +65-67766872.
E-mail address: micsimts@nus.edu.sg (T.S. Sim).

1381-1177/03/$ – see front matter
doi:10.1016/S1381-1177(03)00037-7

dioxygenase (LDOX) and anthocyanidin synthase (ANS) are involved in the biosynthesis of flavanoids [4]; 1-aminocyclopropane-1-carboxylic acid oxidase (ACCO) catalyses the last step in the biosynthesis of ethylene [5]; hyoscyamine 6β-hydroxylase (H6H) is involved in synthesizing the plant alkanoid scopolamine [6]; gibberellin 20-oxidase ($G^{20}O$) and desacetoxyvindoline-4-hydroxylase (DVH) are involved in the biosynthesis of gibberellins [7] and vindolines [8], respectively. In this report, these enzymes are collectively named as IPNS related *non-haem iron-dependent oxygenases* and *oxidases* or designated as the NHIDOX family in this study.

Iron, oxygen and ascorbate are commonly used for catalysis although the NHIDOX enzymes carry out different oxidative reactions and exhibit different substrate selectivities. For instance, IPNS is instrumental in catalyzing a desaturative ring closure reaction whereby a tripeptide substrate L-α-aminoadipyl-L-cysteinyl-D-valine (LLD-ACV) is cyclized to form the four-membered β-lactam ring and the five-membered thiazolidine ring of isopenicillin N [1]. DAOCS, also known as expandase, catalyses the oxidative ring expansion of the five-membered thiazolidine ring of penicillin N to form the six-membered dihydrothiazine ring of deacetoxycephalosporin C (DAOC) and DACS is responsible for hydrolyzing the methyl group at the C3 position of DAOC.

Another enzyme of interest, H6H, is involved in catalyzing an epoxidation reaction in which the substrate hyoscyamine is converted to form the plant alkaloid scopolamine [6]. With the exception of IPNS and ACCO, the other members of NHIDOX family also require 2-oxoglutarate as the co-substrate for catalysis.

To date, extensive biochemical and functional studies of various enzyme members have shown conclusively the conserved residues responsible for binding iron and substrate/co-substrate [9–18]. Separate reports [2,19,20] on the high-resolution crystal structure complexes of *Aspergillus nidulans* IPNS (IPNS_AN) and *Streptomyces clavuligerus* DAOCS (DAOCS_SC) has been reported. Such information will provide clues on how an enzyme's function can be modulated by certain amino acids or strings of sequences. In this study, we have described the conserved architectural scaffolds as well as the key factors that are responsible for modulating the functional differences and substrate selectivities of enzymes in the NHIDOX family.

2. Materials and methods

2.1. Data retrieving

The analysis is based on comparison of 142 protein sequences and 2 protein structures available in the databank. All sequences were retrieved from the National Center of Biotechnology Information (NCBI)-supported sequence search program and the structure coordinates were downloaded from the Protein Data Bank (PDB). The 142 sequences include 11 IPNS, 4 DAOCS, 2 DACS, 20 FL3OH, 11 LDOX, 5 ANS, 7 FS, 19 GO, 2 H6H, 52 ACCO and 1 DVH homologous enzymes as well as proteins isolated from *Campylobacter jejuni*, fruit fly, fission yeast *Schizosaccharomyces pombe*, *Acinetobacter* and *Pseudomonas* species. The sources, protein sequence accession numbers and abbreviated names of the enzymes analyzed are described in supplementary material.

2.2. Primary sequence analysis

The sequences were aligned and the mean pair-wise comparison between and within various enzyme groups were calculated using CLUSTALW [21]. The calculated percentage sequence identity is defined as $(n_{\text{identical}}/L_{\text{min}})100$ where $n_{\text{identical}}$ is the number of positions in the alignment that has the same amino acid and L_{min} the number of residues in the shorter of the two sequences being compared.

2.3. Secondary structure analysis

Two prediction programs, the self-optimized prediction method (SOPM) [22] and GOR IV secondary structure prediction method (GOR IV) [23] were used to assign structure information for enzymes with undetermined protein structures. To improve the reliability of this analysis, three homologous protein sequences from each enzyme group were selected based on disparate protein sequence

identity for structure prediction analysis. For example, FS_PH and FS_MI were selected because they share the lowest and the highest average sequence identity values with other FS enzymes, at 23 and 53%, respectively. And the third homologous enzyme, FS_AT, was selected because it shares intermediate sequence identity value (∼28%) with other FS enzymes. For the analysis, the protein sequences of selected enzymes (IPNS_AN, DAOCS_SC, DACS_SC, DACS_NL, FL3OH_PA, FL3OH_IB, FL3OH_PH, LDOX_DC, LDOX_IP, LDOX_OS, ANS_CC, ANS_FI, ANS_IB, FS_AT, FS_PH, FS_MI, GO_SD, GO_PS, GO_CM, H6H_HN, H6H_AB, ACCO_HA, ACCO1_PX, ACCO_SL, DVH_CR, OXIDOR_CJ, T41002_SP, SDPP_SP, unknown_Acine, OXIDOR1_PA, OXIDOR2_PA, CG5346_DM and CG5340_DM) (supplementary material) were initially arranged to have aligned positions identical to that in the extensive multiple sequence alignment of the entire lot of 142 protein sequences. Thereafter, the SOPM and GOR IV predicted secondary structures of these enzymes were encoded by color (yellow for strands, red for helices and gray for loops) and manually highlighted onto their corresponding aligned sequence positions. Since both SOPM and GOR IV assigned secondary structure-based analysis produced synonymous results, we have chosen to present only the structure analysis data based on SOPM predictions (Fig. 1). Moreover, since the secondary structures assigned to the selected homologous protein sequences within each enzyme group bear strong resemblance to one another, only the results obtained for one of the homologous protein sequences is presented.

2.4. Tertiary structure analysis

The protein structures of IPNS_AN and DAOCS_SC (PDB accession codes 1BLZ and 1RXG, respectively) were viewed and manipulated using Swiss PDB Viewer v3.7b2 (SPdbV). All modeled structures were generated using the SWISS-MODEL program [24]. Aligned structure analysis, superimpositions, residue distance determinations and hydrogen bonding network evaluations were carried out using tools provided by the software program based on default parameters. Hydrogen bonds were detected if a hydrogen atom is located within 1.2–2.76 Å range of a "compatible" donor atom [25].

3. Results and discussion

3.1. Comparative primary sequence analysis

An extensive sequence analysis of NHIDOX enzymes was carried out 5 years ago [9]. This analysis included 52 non-redundant enzyme members, consisting only of those involved in making β-lactams in certain prokaryotic and eukaryotic lineages and those involved in synthesizing secondary signaling products in various plant species. Unexpectedly, our sequence searches and analyses revealed that several proteins isolated from more diverse origins, such as the fruit fly, fission yeast, *C. jejuni*, *Pseudomonas* and *Acinetobacter* species, are also closely related. This suggests that the NHIDOX family is more diverse than previously thought [2,9] and that new members unveiled through the growing number of completed genome sequence projects [26–28] will constitute enzymes from wider spectrum of mechanistic functions and metabolic pathways than is currently known. These enzymes have also been described as iron/ascorbate oxidoreductase family under the pfam accession number PF00671 [29].

The sequences of the 142 enzymes compared in this study are mostly 300–400 amino acids in lengths with the exception of CG5340_DM fruit fly protein, which is 542 amino acids in length. The other related fruit fly protein, CG5346_DM is only 388 amino acids in length. Clustal analysis showed that only the 200–542 amino acid sequence region of CG5340_DM aligned well with other NHIDOX enzymes (Fig. 1). Interestingly, further scrutinizing of the sequence revealed that the N-terminal 46–197 amino acid region of CG_5340 have high homologies with the 354–507 region of the same sequence. This region constitutes part of a conserved domain (module II in Fig. 1) elucidated in the latter analysis in this study. Hence, this suggests the possible existence of a duplicated structure in CG5340_DM and it will be interesting to understand more about these structures through detailed analysis of the conserved module II domain.

The enzymes compared have very low sequence identities, ranging from an average of 20% to as

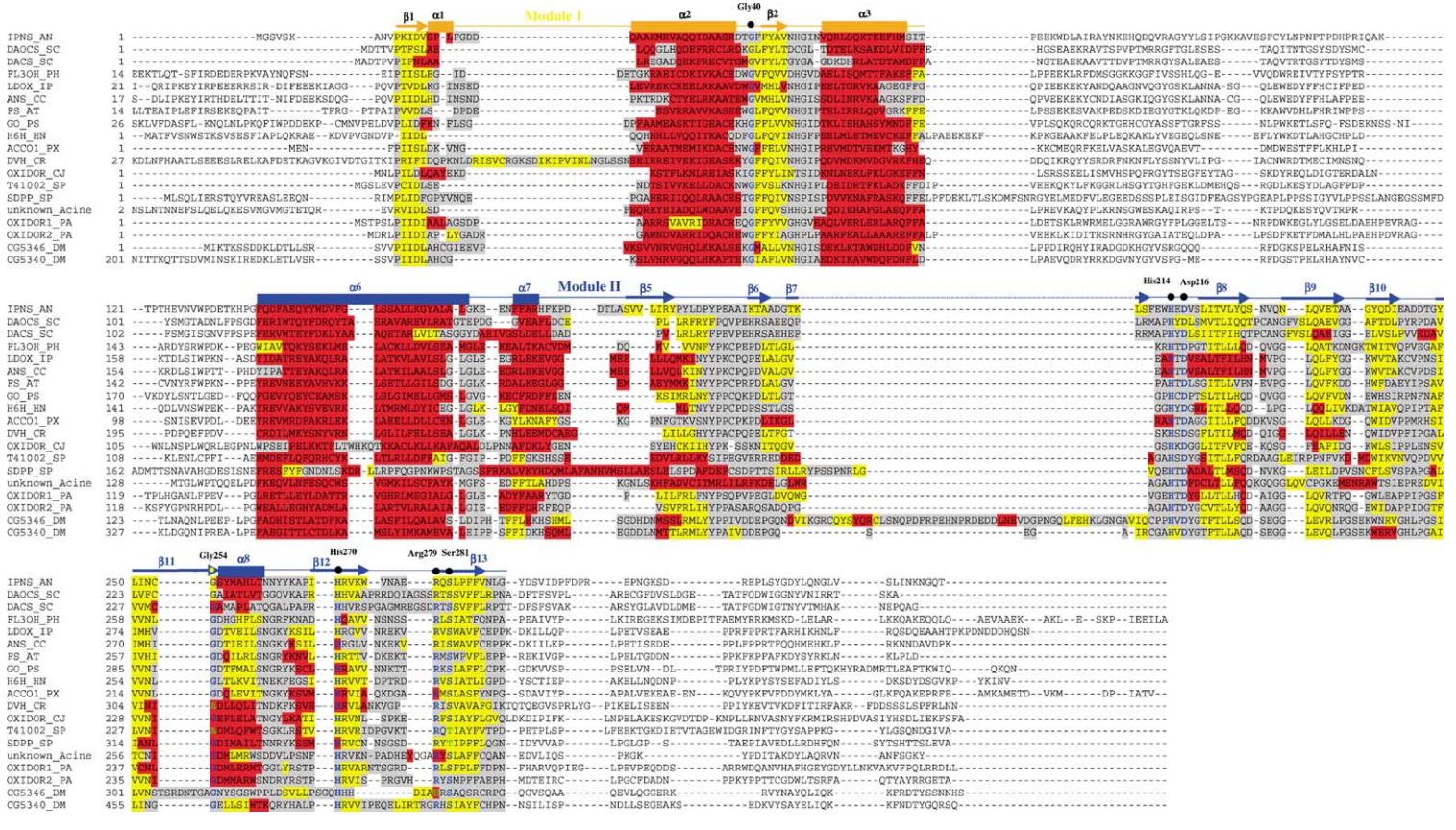

Fig. 1. Sequence-based comparative analysis of SOPM-predicted secondary structures of NHIDOX enzymes. The protein sequences of IPNS_AN, DAOCS_SC, DACS_SC, FL3OH_PH, LDOX_IP, ANS_CC, FS_AT, GO_PS, H6H_HN, ACCO1_PX, DVH_CR, OXIDOR_CJ, T41002_SP, SDPP_SP, unknown_Acine, OXIDOR1_PA, OXIDOR2_PA, CG5346_DM and CG5340_DM were arranged according to their respective aligned positions in the multiple sequence alignment of 142 NHIDOX enzymes. The SOPM-predicted secondary structures (yellow, β-strand; red, α-helix; gray, loop) of these enzymes are highly conserved in two delineated regions, modules I and II. Seven highly conserved residues (bold, blue) are found to juxtapose within the conserved modules. DVH_CR and CG5346_DM contain substantially long sequence stretches that intervene at the specific regions of modules I and II, respectively.

low as 10% (data not shown) and they share seven highly conserved residues position (Fig. 1). The motif comprising of His214, Asp216 and His270 of IPNS_AN, shown to be important for binding iron in IPNS, DAOCS, ACCO and FL3OH [9–13,17] are strictly conserved in all the enzymes compared. Two other conserved residues, corresponding to Arg279 and Ser281 of IPNS_AN, have been shown to bind the ACV substrate in IPNS [14,15] and the 2-oxoglutarate co-substrate in FL3OH [16]. The remaining conserved residues are glycines (Gly40 and Gly254, IPNS_AN numbering). Their specific roles are unknown although functional studies suggest that these residues when mutated affect protein folding properties in IPNS [30]. Additionally, certain residues such as Pro10, Ala35 and His47 (IPNS_AN numbering) appear to be relatively conserved, although, the reasons for their conservedness are not known.

After tracking through the corresponding amino acid residue positions in 142 enzymes, the three (His214, Asp216 and His270) proposed to bind iron are extremely restricted in terms of the type of residues found at these locations. This observation is in agreement with studies demonstrating the constraints in the spatial organization and residue preference of the iron-binding ligands in IPNS [31]. In contrast, the proposed substrate/co-substrate binding residues are not so strictly conserved. For instance, the residue position in the CG5346_DM fruit fly protein corresponding to Arg279 is occupied by a histidine (Fig. 1). The residue positions in *S. pombe* T41002_SP and SDPP_SP proteins corresponding to Ser281 are both occupied by threonine instead of serine. Hence, it seems that this particular conserved position can be occupied by either serine or threonine, both of which have hydroxyl side groups. Between the two conserved glycines, only the residue positions corresponding to Gly40 are strictly conserved whereas those corresponding to Gly254 are not. For example, in DVH_CR and T41002_SP, they are occupied by alanines. These results, which were not observed in previous sequence analyses of enzymes in this family [9,11] suggest that there are variable residue type restrictions in these seven conserved sites and our latter structural analyses provide support that this may have evolved according to their specific functional roles and distinct structural involvements.

3.2. Sequence-based comparative secondary structure analysis

High-resolution structural data of IPNS and DAOCS [2,19,20] were analyzed in this study. In a number of enzyme families [32–36], detailed sequence analyses and three-dimensional structure comparisons have revealed the sequence patterns, unique architectural features and how key residues are involved in stabilizing the common folds. Similar studies are not possible for enzymes in the NHIDOX family due to the limited number of available enzyme structures. However, using comparative biocomputational methods such as sequence-based analysis of secondary structures and superimposed tertiary structure analyses, the otherwise cryptic structural relationships indigenous to this family of enzymes are unveiled.

Comparative sequence-based structure analysis revealed that these enzymes share highly conserved secondary structures in two broad regions (Fig. 1) computed to be highly divergent in sequence. These conserved regions are marked as module I which comprises of three α-helices and two β-strands and module II which comprises of nine β-strands and three α-helices. Some of these secondary structure elements are more conserved and in certain enzyme members, stretches of sequence structures intervene the conserved modules. The most prominent one is that in CG5346_DM which is composed of ∼60 amino acids made up of essentially loops that intersperse around the β7 region. However, this is absent in CG5340_DM. By gene locus, CG5346_DM is found at ∼2 kbp upstream from CG5340_DM in chromosome 3R. Hence, the sequence and structure differences imply that these two related *Drosophila* proteins are most likely functionally distinct. Alternatively, the apparent intervening sequence in CG5346_DM may be an artifact introduced during gene annotations in the *Drosophila* genome project, as it is not uncommon to have errors in demarcating the exon–intron regions. Close examination of the genome sequence for this region revealed two highly potential splice consensus sequences at about nucleotide positions 99265 and 99105. Hence, the intervening sequence will be removed if the splicing of the mRNA occurs at these sites. Besides the intervening sequences, the N- and C-terminal regions as well as regions connecting modules I and II of the enzymes com-

pared are not conserved in terms of their secondary structures.

Also in Fig. 1, it is clearly illustrated that all the seven conserved residues are juxtaposed within the conserved structural modules. Module I carries only one conserved residue, Gly40, and the remaining six conserved residues are all located in module II. The secondary structure elements in which the conserved residues are located differed in the various enzymes. Although the iron-binding ligands have preference for certain residues, the various ligand-binding sites are not restricted to specific secondary structures. Residue positions corresponding to Asp216 in the enzymes analyzed are found in the loops. However, corresponding positions to His214 are found in loops at 75% frequency while for His270, corresponding positions are found in the helical regions at 40% occurrence. Residue positions corresponding to the substrate/co-substrate binding ligands, Arg279 and Ser281, are mostly found in loops (53%) and β-strands (61%), respectively. The two conserved glycines are both found in loops, with about 75–77% occurrence. Interestingly, five of the seven conserved residues are located more frequently in loop regions.

3.3. Computational superimposed tertiary structure comparison

Scrutiny of the conserved secondary structure domains aligned against IPNS_AN protein structure revealed their tertiary orientations as well as their unique structural features. The contiguous β-strands in module II were found to correspond to the eight antiparallel running β-strands in IPNS_AN responsible for forming the jelly-roll active center (highlighted blue in Fig. 2A). The longest conserved α-helix, labeled as α6, aligned precisely to the α-helix proposed to stabilize the IPNS_AN active center. The corresponding conserved secondary structures β1 and α1 of module I in IPNS_AN (highlighted yellow in Fig. 2A) were observed to connect via a short loop to α2 (possibly with the exception of DVH_CR). Another short loop that connects α2 to β2 was observed to return β2 alongside β1 to form an antiparallel β-sheet that continues with β11, β8, β13 and β5 to form a larger extended sheet. Superimposition analysis carried out with DAOCS_SC also revealed similar observations. Hence, the significant conservation of secondary structures in modules I and II suggest that all NHIDOX enzymes may adopt common scaffolds and generate a conserved architecture for their active centers as observed for IPNS_AN and DAOCS_SC.

A topology diagram illustrating the arrangements of the secondary structures in the conserved modules, particularly that of the connections and orders between the β-strands and loops making up the jelly-roll motif, is shown in Fig. 2B. Jelly-roll motifs are commonly observed in plant lectin concanavalin A and the coat proteins of most spherical viruses such as the hemagglutunin protein from influenza virus [37]. However, it is very clear from this analysis Fig. 2A and B) as well as previous observations [2] that the jelly-roll structures in IPNS_AN consists of an incompletely closed and distorted structural unit. Across the enzymes (Fig. 1), some of the β-strands that made up the core motif, such as those corresponding to β2, β8, β11 and β13, are more conserved than others. Particularly interesting is that in some enzymes, they have "missing" β-strands around the β6 and β7 regions. These β-strands are arranged antiparallel to each other and make no connections to those in the core motif. Moreover, the same region in CG5346_DM is interspersed by a broad region of loops. Hence, whether these or other enzymes also form a distorted or perfectly closed jelly-roll motif and whether the differences existing within this important core have implications on biological functions, can only be made clear with high-resolution structure elucidations.

3.4. Hydrogen bonding networks within conserved residue clusters

Clearly, all the seven highly conserved residues are brought into close proximity around the active center (Fig. 2A). The spatial orientation of Gly40 is particularly interesting. This residue is brought in close proximity to the jelly-roll motif from the N-terminus region via the folding and arrangement of secondary structures that bring the residues in β2 to form bondings with β11 (Fig. 2B). Based on their spatial distance relationships, the seven conserved residues can be subgrouped into two clusters. The iron-binding ligands and the two conserved glycine residues are located within an averaged distance of 10 Å from one another. However, the substrate/co-substrate binding residues are located further away, within an averaged

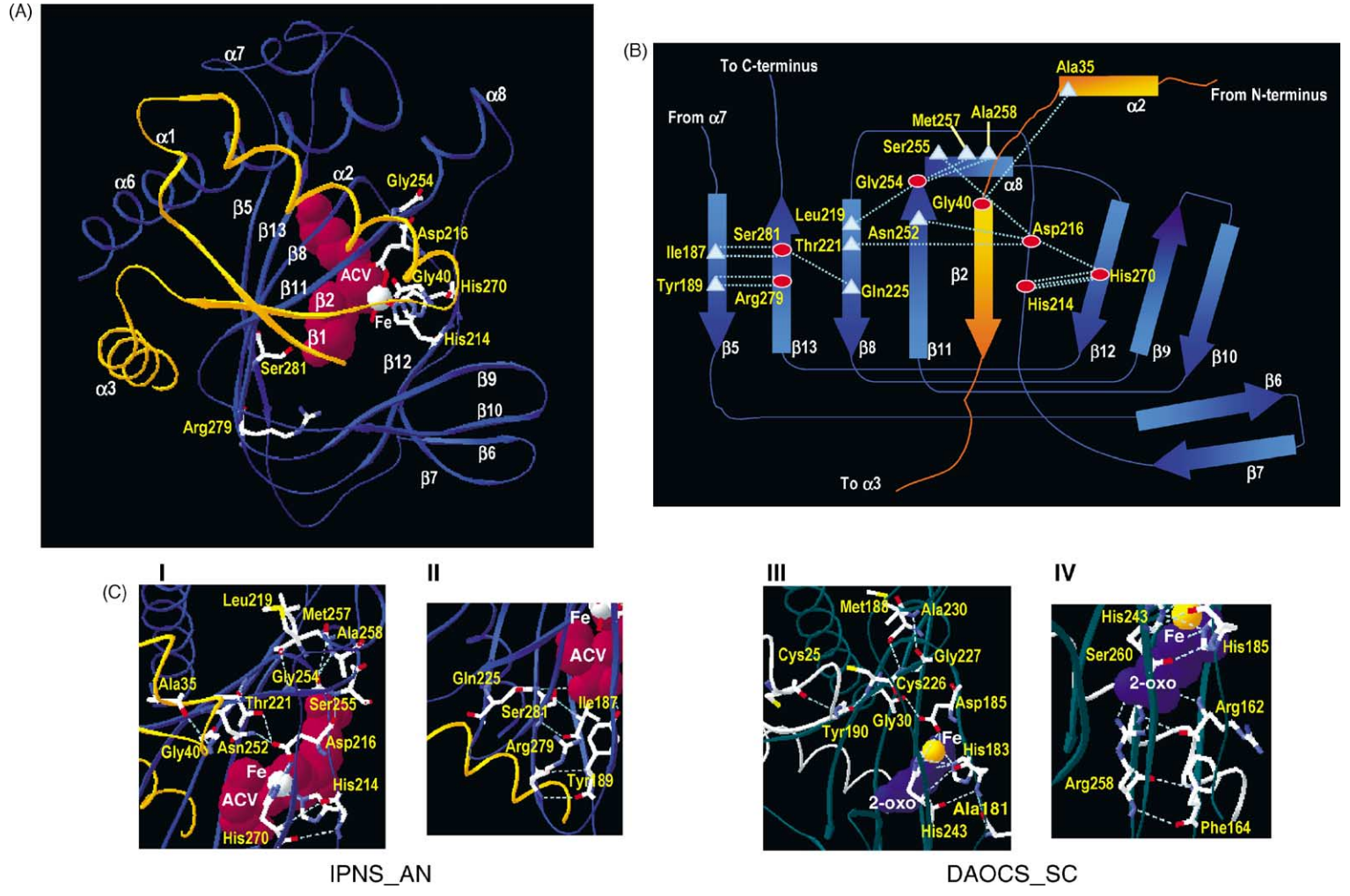

Fig. 2. Spatial orientations of the conserved structural modules and the bonding networks of conserved residue clusters. (A) As deciphered from IPNS_AN tertiary structure, the conserved secondary structures in module I (yellow) are brought into close proximity around the jelly-roll motif structure regions in module II (blue). The spatial positions of the conserved residues (CPK colors), iron (gray) and ACV substrate (maroon) are also indicated. (B) Topology diagram illustrating the connectivity within the jelly-roll barrel. Extensive hydrogen bonding interactions are observed to connect between the seven conserved residues (red circles) and the various partners (blue triangles) to form the core motif structure. β6 and β7 form an antiparallel β-sheet which did not appear to make any bonding connections with the other two β-sheets in this conserved region. (C(I, II)) The bonding network within the iron-binding ligand and the substrate/co-substrate binding ligand clusters of IPNS_AN are shown, respectively. (C(III, IV)) The hydrogen bonding networks clustering around the corresponding conserved residues in DAOCS_SC are shown. The conserved secondary structures in modules I and II of DAOCS_SC are shown in white and turquoise ribbons, respectively, and the spatial locations of iron (yellow) and co-substrate 2-oxoglutarate (2-OXO) (purple) are also indicated. Hydrogen bonding connections in all figures are represented by dotted lines. CPK color coding, white: carbon and hydrogen, blue: nitrogen, red: oxygen and yellow: sulfur.

distance of 20 Å from the earlier cluster. Most strikingly, structure analysis revealed extensive hydrogen bonding networks within both conserved residue clusters. As shown in Fig. 2C(I), the iron-binding residues, His214 and His270 of IPNS_AN, are connected via three interactions to each other, consisting of two main-chain:main-chain and one side-chain:side-chain hydrogen bonding interactions. Moreover, the axial nitrogen atom in the imidazole side chain of His270 also participates in a fourth hydrogen bonding interaction with the carbonyl group in the side chain of Asp216. Also clearly revealed from the analysis is that Asp216 is held in its spatial orientation via three additional hydrogen bonding interactions, Asn252 (β11), Ser255 (α8) and Thr221 (β8). In close proximity, Gly254 intercalates with three backbone interactions with the corresponding main chains of Leu219 (β8), Met257 (α8) and Ala258 (α8). Gly40, located in the turn region of the loop connecting α2 and β2, is observed to form a main-chain:main-chain hydrogen bonding interaction with Ala35, which is located at the C-terminus cap of the preceding helix (α2). On the opposite end of the core motif, the substrate binding residues are also extensively connected (Fig. 2C(II)). For instance, Arg279 forms a bifurcate backbone interaction with the carbonyl oxygen and amide nitrogen of Tyr189 (β5) and Ser281 makes three hydrogen bonding interactions, a bifurcate backbone interaction with the carbonyl oxygen and amide nitrogen of Ile187 (β5) and a main-chain:side-chain interaction with Gln225 (β8). The locations of the conserved residues and their bonding partners as well as the extensive network interactions are illustrated in Fig. 2B. Interestingly, majority of the bonding partners are clustered in β5, β8 and α8. Intricate networks of hydrogen bondings involving key residues have been proven, both through theories and experiments, to be important for protein folding and stability [38–41]. Similarly, the extensive network structures observed for the conserved residues in IPNS_AN suggest that apart from their known functional roles (except Gly40 and Gly254), they also participate in juxtaposing the respective conserved strands and loops to form the core motif structure in IPNS_AN via bonding interactions.

When the bonding partners of the seven conserved residues in three other IPNS isozymes, IPNS_CA, IPNS_SJ and IPNS_SC, were evaluated using modeled protein structures, similar extensive hydrogen bonding networks were observed (Table 1). Although the bonding partners for Gly40, His214 and His270 appear to be conserved, differences in terms of partner residue identities and the exact number of bonding partners were observed for those of Gly254, Asp216, Arg279 and Ser281. When similar analyses were carried out for DAOCS isozymes, specific differences in the bonding relationships were also observed for the conserved residues in these enzymes and those of IPNS, in particular, the respective bonding partners to Arg279 and Ser281. Thus, instead of making interactions with residues corresponding to Tyr189/Ile187, the corresponding conserved Arg279/Ser281 co-substrate binding site in various DAOCS isozymes interact with a Phe/Arg partner pair (though the number of bondings observed in DAOCS_LL is different) which corresponds to Leu192/Pro190 in IPNS_AN. This results in a lateral shift of two amino acid positions in the partnering patterns of the respective β-strands in DAOCS and IPNS.

Although we could not carry out similar bonding analyses for the remaining NHIDOX enzymes due to the lack of available protein structures, we have evaluated the residue types found at positions corresponding to all bonding partners identified in IPNS and DAOCS (Table 1). Those of ACCO and FL3OH are specifically enumerated since experimental evidence for the essentiality of their conserved residues have been obtained. Interestingly, most of these positions are also highly conserved, albeit the residue identities at specific bonding sites differ from those found in IPNS and DAOCS. When the analyses were extended to include all 142 NHIDOX enzymes, the corresponding positions which aligned to the bonding partners of the three iron-binding ligands, His214, Asp216 and His270, were found to be most conserved. This suggests that the corresponding conserved iron-binding ligands in the NHIDOX enzymes utilize a similar repertoire of bonding partners to generate the same organized structure. This network is structurally important as can be seen in Fig. 2B, the bonding interactions between His214 and His270, appear to be essential for anchoring Asp216 such that the latter can function to hold the respective conserved strands and loops in proper conformation. This finding, together with our observations that these ligands are most restricted in terms of residue identities, suggest that functional

Table 1

Hydrogen bonding network of conserved residue clusters in NHIDOX enzymes. *Top panel*: The identities of the residues that form bondings with the corresponding conserved residues in IPNS_AN, IPNS_CA, IPNS_SJ, IPNS_SC, DAOCS_SC, DAOCS/DACS_CA, DAOCS_LL, DAOCS_NL were determined using the respective crystal and modeled structures. Their corresponding positions are marked with respect to IPNS_AN amino acid numbering. If more than one hydrogen bond is formed between the respective residues, the bonding numbers are indicated in superscripts. *Middle panel*: the residue types found at the corresponding positions of the respective bonding partners were evaluated in 52 ACCO and 20 FL3OH isozymes and those enumerated from all 142 NHIDOX enzymes are shown in the *bottom panel*. The numbers in parentheses indicate the occurrence frequency of the residue types

Identities of residues that are bonded to the corresponding conserved residues evaluated in various IPNS and DAOCS isozymes																					
	Gly40[a]		**Gly254**			**His214**		**Asp216**				**His270**		**Arg279**				**Ser281**			
	Ala35[b]	Cys253	Leu219	Met257	A258	Glu212	His270	Thr221	Asn252	Ser255	His270	His214	Asp216	Tyr189	Val275	Ala277	Leu192	Ile187	Gln225	Pro190	
IPNS_AN	Ala		Leu	Met	Ala		His[3]	Thr	Asn	Ser	His	His[2]	Asp	Tyr[2]				Ile[2]	Gln		
IPNS_CA	Ala		Leu	Met			His[2]		Asn	Ser	His	His[2]	Asp	Tyr[2]	Val	Glu		Ile[3]			
IPNS_SJ	Ala		Met	Met			His[3]		Asn	Thr[2]	His	His[3]	Asp	Tyr[2]	Ala			Ile[3]			
IPNS_SC	Ala		Met	Met			His[3]		Asn	Thr[2]	His	His[3]	Asp	Tyr[2]				Ile[2]			
DAOCS_SC	Cys	Cys	Met	Ala		Ala	His[3]	Thr		Ala	His	His[3]	Asp				Phe[2]				Arg[2]
DAOCS/DACS_CA	Cys	Cys	Thr	Gly			His[2]			Ala	His	His[3]	Asp				Phe[2]				Arg[2]
DAOCS_LL	Cys	Cys	Met	Ala			His[2]	Thr		Ala	His	His[3]	Asp				Phe				Arg[2]
DAOCS_NL	Cys	Cys	Ile	Ala			His[2]			Ala	His	His[3]	Asp				Phe[c]				Arg[c]
Residue types found at the corresponding positions of ACCO and FL3OH																					
ACCO (52 isozymes)	Ala (94)[c] Val (2)	Leu (85) Ile (15)	Gly (100)	Leu (90) Ile (8) Val (2)	Glu (100)	Arg (98) Ser (2)	His (100)	Ile (98) Val (2)	Asn (100)	Asp (100)	His (100)	His (100)	Asp (100)	Gly (100)	Gln (92) Arg (2) His (2)	Gly (90) Thr (4) Glu (2)	Val (100)	Thr (54) Asn (44) Pro (2)	Gln (98) Glu (2)	Thr (96) Ser (2) Ala (2)		
FL3OH (20 isozymes)	Ala (100)	Leu (100)	Thr (95) Leu (5)	Gly (95) Ala (5)	His (100)	Lys (100)	His (100)	Thr (100)	Asn (100)	Asp (100)	His (100)	His (100)	Asp (100)	-[d]	Asn (100)	Asn (40) Gln (30) Ser (25)	Val (90) Leu (10)	-	Gln (100)	Val (50) Ile (30) Leu (15)		
Residue types found at the corresponding positions of NHIDOX enzymes																					
All 142 NHIDOX enzymes	Ala (94) Cys (2) Ile (1) Pro (1) Val (1) Trp (1) Ser (1)	Leu (49) Ile (35) Cys (10) Val (4) Ser (1) Gly (1)	Gly (39) Thr (17) Ser (14) Ala (13) Leu (7) Met (4) Ile (4) Cys (1) Tyr (1) Phe (1)	Leu (47) Gly (14) Phe (14) Ile (9) Met (7) Val (4) Ala (4) Asp (1)	Glu (54) His (14) Met (14) Ala (5) Thr (3) Gly (2) Pro (2) Gln (1) Lys (1) Leu (1) Asn (1) Ser (1)	Arg (37) Gly (20) Glu (17) Lys (14) Ala (4) Val (4) Pro (1) Ser (1) Gln (1)	His (100)	Thr (60) Ile (37) Ser (3) Val (1)	Asn (80) His (15) Phe (3) Leu (1) Met (1)	Asp (86) Ala (4) Thr (4) Ser (3) Glu (1) Leu (1)	His (100)	His (100)	Asp (100)	Gly (43) Ile (16) Gln (12) Leu (7) Tyr (7) Glu (5) Val (4) Met (2) His (1) Arg (1) Asn (1) Ser (1) Thr (1)	Asn (45) Gln (35) Pro (7) Val (4) Asp (4) Ile (2) Ser (1) His (1) Arg (1) Thr (1)	Gly (36) Lys (17) Ser (11) Thr (10) Ala (6) Arg (5) Ile (4) Glu (4) His (3) Cys (3) Tyr (1) Asn (1) Phe (1) Val (1)	Val (51) Leu (25) Ile (17) Met (2) Thr (1) Phe (1) Lys (1) Tyr (1) Gly (1)	Thr (31) Asn (26) Leu (15) Ile (12) Val (4) Glu (3) Met (2) His (1) Ser (1) Tyr (1) Trp (1) Pro (1)	Glu (83) His (11) Pro (4) Lys (1) Thr (1) Gln (1)	Thr (36) Met (22) Leu (17) Val (7) Pro (6) Ile (6) Phe (2) Lys (1) Cys (1) Ser (1) Ala (1)

a. The seven highly conserved residues in IPNS-related NHIDOX enzymes.
b. All labelings are stated with respect to IPNS_AN.
c. The figures in brackets represent percentage occurrences rounded off to the nearest whole number.
d. No residues are aligned in this position.
If more than one hydrogen bond is formed between the respective residues, the bonding numbers are indicated in superscripts.

constraints are imposed on the evolution of these ligands as well as their hydrogen bonding partners. Flexibility around the iron-binding ligand cluster region appears to be modulated by the neighboring Gly40 and Gly254 bonding networks. In comparing Fig. 2C(I, III), the extra bonding between Gly30 (corresponding to Gly40 of IPNS_AN) and Cys226 of DAOCS_SC, appears to result in a tight arrangement of the conserved structures around its jelly-roll motif. This has been confirmed through superimposition analyses of IPNS_AN and DAOCS_SC protein structures. Moreover, the number and types of bonding interactions that Gly254 is involved in may be critical for controlling the flexibility of the loop region connecting β11 and β12 and between β8 and β11. It is interesting to observe that the corresponding bonding partners of Gly40 and Gly254 are less conserved in the NHIDOX enzymes compared to the iron-binding ligands, suggesting that different network organizations may exist. However, it is important to note that the corresponding conserved Gly40 and Gly254 in these enzymes may either be bonded to the respective aligned residues identified in Table 1, to adjacent or other neighboring residues, or make no bonding interaction at all.

Most strikingly, the analyses revealed that the hydrogen bonding partners of Arg279 and Ser281 are most diverse amongst the conserved residues. This is specifically evident from the analyses of IPNS, DAOCS, ACCO as well as FL3OH isozymes. Hence, although the various NHIDOX enzymes have evolved a conserved ArgXSer residue pair for binding substrate/co-substrate, control over their specificities and selectivities is modulated via the bonding network organizations. Particularly interesting is that a similar bonding involving the main-chain:side-chain bonding interaction of Ser281 and Gln225 which connects β13 and β8 in IPNS_AN is absent in various IPNS and DAOCS isozymes. However, the corresponding residue positions to Gln225 are rather conserved in other NHIDOX enzymes and it will be interesting to pursue whether the missing bondage observed is substituted by other connections.

4. Conclusions

Comparison of the protein sequences of NHIDOX enzymes essentially reveals their linear relationships of having seven highly conserved residues. However, analysis of their two- and three-dimensional structure relationships showed that these enzymes have evolved common topological scaffolds, comprising of highly conserved antiparallel running β-strands and bondings with certain amino acid residues, to form the jelly-roll motif structure for their active centers. The remarkable conservation of core motifs throughout NHIDOX enzymes can be interpreted in terms of their stringent requirement of an inimitable chemical and structural environment for dioxygen binding and generation of highly oxidized Fe(IV)-oxo moieties to mediate appropriate substrate oxidations. Interestingly, we have identified a conserved hydrophobic cleft made up of 11 hydrophobic residues, appropriately configured in the active center of IPNS [42]. This cleft is possibly important for isolation of reactive intermediates generated during the oxidative reactions from destruction of enzyme function. It is suggested that structurally similar hydrophobic patches exist in the active centers of NHIDOX enzymes since residues corresponding to the proposed hydrophobic cleft are also highly conserved or conservatively substituted. Although these enzymes have evolved a stable architectural framework for iron oxidation, this investigation reveals their esoteric structural differences. It is quite apparent that extensive networks of hydrogen bonding interactions involving the conserved residues are colocalized within the conserved jelly-roll motifs of these enzymes. However, the different network organizations observed in various NHIDOX enzymes are reminiscent of differences in catalytic prowess and substrate specificities. Further structural studies are eminent, especially those designed to compare the fine structural differences between their core motifs and product formation. The importance for elucidating fundamental folding pathways to stabilize the jelly-roll core motifs and the involvement of the seven conserved residues and their respective interaction partners also cannot be overemphasized.

In principle, the results clearly illustrate that in addition to sequence and invariant residue relationships, the conserved structural modules and common features observed in the NHIDOX enzymes can be manipulated as useful emblems for more accurate annotation and identification of sequence and structural relatives of this family of catalytically diverse enzymes. Although the NHIDOX enzymes possess

functional residues that are highly conserved, the interactions that these residues are involved in are the key modulators for fine tuning structural features allowing the diverse enzymes to achieve exquisite reactions.

Acknowledgements

The assistance of Francis Ng Hoong Kee in the computational calculations is greatly appreciated. This research was supported by National University of Singapore (Singapore) Research Grant Number R-182-000-019-112.

References

[1] J.E. Baldwin, E.P. Abraham, Nat. Prod. Rep. 5 (1998) 129–145.
[2] P.L. Roach, I.J. Clifton, V. Fütöp, K. Harlos, G.J. Barton, J. Hadju, I. Andersson, C.J. Schofield, J.E. Baldwin, Nature 375 (1995) 700–704.
[3] C.J. Schofield, J.E. Baldwin, M.F. Byford, I. Clifton, J. Hajdu, C. Hensgens, P.L. Roach, Curr. Opin. Struct. Biol. 7 (1997) 857–864.
[4] W. Heller, G. Forkmann, in: J.B. Harborne (Ed.), The Flavanoids, Advances in Research Since 1986, Chapman & Hall, London, 1993, pp. 399–425.
[5] D.O. Adams, S.F. Yang, Proc. Natl. Acad. Sci. U.S.A. 70 (1979) 170–174.
[6] T. Hashimoto, J. Matsuda, Y. Yamada, FEBS Lett. 329 (1993) 35–39.
[7] T. Lange, P. Hedden, J.E. Graebe, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 8552–8556.
[8] F. Vazquez-Flota, E. De Carolis, A.M. Alarco, V. De luca, Plant Mol. Biol. 34 (1997) 35–948.
[9] I. Borovok, O. Landman, R. Kreisberg-Zakarin, Y. Aharonowitz, G. Cohen, Biochemistry 35 (1996) 1981–1987.
[10] V.J. Lay, A.G. Prescott, P.G. Thomas, P. John, Eur. J. Biochem. 242 (1996) 228–234.
[11] D.S.H. Tan, T.S. Sim, J. Biol. Chem. 271 (1996) 889–894.
[12] P. Loke, J. Sim, T.S. Sim, FEMS Microbiol. Lett. 157 (1997) 137–140.
[13] R. Lukačin, L. Britsch, Eur. J. Biochem. 249 (1997) 748–757.
[14] P. Loke, T.S. Sim, FEMS Microbiol. Lett. 164 (1998) 107–110.
[15] P. Loke, T.S. Sim, FEMS Microbiol. Lett. 165 (1998) 353–356.
[16] R. Lukačin, I. Groning, U. Pieper, U. Matern, Eur. J. Biochem. 267 (1999) 853–860.
[17] J. Sim, T.S. Sim, Biosci. Biotechnol. Biochem. 64 (2000) 828–832.
[18] H.J. Lee, M.D. Lloyd, I.J. Clifton, K. Harlos, A. Dubus, J.E. Baldwin, J.M. Frere, C.J. Schofield, J. Biol. Chem. 276 (2001) 18290–18295.
[19] P.L. Roach, I.J. Clifton, C.M.H. Hensgens, N. Shibata, C.J. Schofield, J. Hadju, J.E. Baldwin, Nature 387 (1997) 827–830.
[20] K. Valegård, A.C. Terwisscha van Scheltinga, M.D. Lloyd, T. Hara, S. Ramaswamy, A. Perrakis, A. Thompson, H.J. Lee, J.E. Baldwin, C.J. Schofield, J. Hajdu, I. Andersson, Nature 394 (1998) 805–809.
[21] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673–4680.
[22] C. Geourjon, G. Deléage, Protein Eng. 7 (1994) 157–164.
[23] J. Garnier, J.F. Gibrat, B. Robson, Methods Enzymol. 266 (1996) 540–553.
[24] N. Guex, A. Diemand, M.C. Peitsch, TIBS 24 (1999) 364–367.
[25] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P.K. Weiner, J. Am. Chem. Soc. 106 (1984) 765–784.
[26] D.R. Bentley, Med. Res. Rev. 20 (2000) 189–196.
[27] P.S. Lee, K.H. Lee, Curr. Opin. Biotechnol. 11 (2000) 171–175.
[28] A. Pandey, M. Mann, Nature 405 (2000) 837–846.
[29] A. Bateman, E. Birney, R. Durbin, S.R. Eddy, R.D. Finn, E.L.L. Sonnhammer, Nucleic Acids Res. 27 (1999) 260–262.
[30] P. Loke, T.S. Sim, Can. J. Microbiol. 47 (2001) 961–964.
[31] R. Kreisberg-Zakarin, I. Borovok, M. Yanko, Y. Aharonowitz, G. Cohen, Antonie van Leeuvenhock 75 (1999) 33–39.
[32] L. Banei, I. Bertini, A. Rosato, G. Varani, J. Biol. Inorg. Chem. 4 (1999) 824–837.
[33] B. Taneja, S.C. Mande, Protein Eng. 12 (1999) 815–818.
[34] A.E. Todd, C.A. Orengo, J.M. Thornton, Protein Eng. 12 (1999) 375–379.
[35] K. Ginalski, C. Venclovas, B. Lesyng, K. Fidelis, FEBS Lett. 482 (2000) 119–124.
[36] G. Pujadas, J. Palau, Mol. Biol. E 18 (2001) 38–54.
[37] C. Branden, J. Tooze, Beta structures, Introduction to Protein Structure, second ed., Garland, New York, 1998, pp. 67–88.
[38] D. Bordo, P. Argos, J. Mol. Biol. 243 (1994) 504–519.
[39] M.P. Byrne, R.L. Manuel, L.G. Lowe, W.E. Stites, Biochemistry 34 (1995) 13949–13960.
[40] G. Vogt, S. Woell, P. Argos, J. Mol. Biol. 269 (1997) 631–643.
[41] E.J. Herbert, A. Giletto, J. Sevcik, L. Urbanikova, K.S. Wilson, Z. Dauter, C.N. Pace, Biochemistry 37 (1998) 16192–16200.
[42] E. Wong, J. Sim, T.S. Sim, Biochem. Biophys. Res. Commun. 283 (2001) 621–626.