Amino Acids (2009) 36:235–249
DOI 10.1007/s00726-008-0057-2

ORIGINAL ARTICLE

Searching distant homologs of the regulatory ACT domain in phenylalanine hydroxylase

Jessica Siltberg-Liberles · Aurora Martinez

Received: 19 February 2008 / Accepted: 11 March 2008 / Published online: 27 March 2008

Abstract High sequence divergence, evolutionary mobility, and superfold topology characterize the ACT domain. Frequently found in multidomain proteins, these domains induce allosteric effects by binding a regulatory ligand usually to an ACT domain dimer interface. In mammalian phenylalanine hydroxylase (PAH), no contacts are formed between ACT domains, and the domain promotes an allosteric effect despite the apparent lack of ligand binding. The increased functional scenario of this abundant domain encouraged us to search for distant homologs, aiming to enhance the understanding of the ACT domain in general and the ACT domain of PAH in particular. The PDB was searched using the FATCAT server with the ACT domain of PAH as a query. The hits that were confirmed by the SSAP algorithm were divided into known ACT domains (KADs) and potential ACT domains (PADs). The FATCAT/SSAP procedure recognized most of the established KADs, as well 18 so far unrecognized non-redundant PADs with extremely low sequence identities and high divergence in functionality and oligomerization. However, analysis of the structural similarity provides remarkable clustering of the proteins according to similarities in ligand binding. Despite enormous sequence divergence and high functional variability, there is a common regulatory theme among these domains. The results reveal the close relationships of the ACT domain of PAH with amino acid binding and metallobinding ACT domains and with acylphosphatase.

Keywords ACT domain · Aromatic amino acid hydroxylases · Phenylalanine hydroxylase · Structural homology · Sequence divergence

Abbreviations

AAAH	Aromatic amino acid hydroxylase
ACYP	Acylphosphatase
AHAS	Acetoacetate synthase isozyme III small subuint
AK	Aspartokinase
ALY	Transcriptional coactivator
ASR	Ancestral sequence reconstruction
Atx1	Metallochaperone
CCS	Metallochaperone of superoxide dismutase
CE	Combinatorial extension
CutA1	Periplasmic divalent cation tolerance protein
Duf190	Domain unknown function
GCVR	Glycine cleavage transcriptional repressor
HisG_C	ATP phophoribosyl transferase
KADs	Known ACT domains
LPRA	Leucine-responsive regulatory protein A
MSCS	Small-conductance mechanosensitive channel
NikR	Nickel binding regulatory protein
PADs	Potential ACT domains
PAH	Phenylalanine hydroxylase
PheA	Prephenate dehydratase
PII A–B	P-II-like signaling protein, nitrogen regulatory protein A–B

Electronic supplementary material The online version of this article (doi:10.1007/s00726-008-0057-2) contains supplementary material, which is available to authorized users.

J. Siltberg-Liberles · A. Martinez (✉)
Department of Biomedicine, University of Bergen,
Jonas Lies vei 91, 5009 Bergen, Norway
e-mail: aurora.martinez@biomed.uib.no

J. Siltberg-Liberles
Department of Molecular Biology, University of Wyoming,
Laramie, WY 82071, USA

RMSD	Root mean square deviation
Rpia	Ribose-5-phosphate isomerase
SGUF 1–7	Structural genomics target with unknown function 1–7
SMBD	Small molecule binding domain
S6 A–B	30S ribosomal protein S6 A–B
TD	Threonine deaminase
TH	Tyrosine hydroxylase
TPH	Tryptophan hydroxylase
VAO	Vanillyl-alcohol oxidase
V-ATPase	Head domain of subunit C
Znta	Zinc, (lead, cadmium, and mercury) transporting ATPase
3-PGDH	D-3-Phosphoglycerate dehydrogenase

Introduction

The N-terminal regulatory domain of phenylalanine hydroxylase (PAH), an important enzyme which catalyzes the conversion of L-Phe to L-Tyr has been classified as an ACT domain (Aravind and Koonin 1999). The ACT domain is one of many different small molecule-binding domains (SMBDs) characterized by high sequence divergence and evolutionary mobility (Anantharaman et al. 2001). For an illustration of the ACT domain 3D structure see Fig. 1.

Originally, the functional link between the established ACT domains—found in many metabolic enzymes—appeared to be a regulatory role promoted by binding a regulatory ligand (Aravind and Koonin 1999; Chipman and Shaanan 2001; Liberles et al. 2005). The ACT domain is generally incorporated in multidomain protein contexts where allosteric regulation appears to be provided by a transmission of finely tuned conformational changes (Liberles et al. 2005). The archetypical ACT domain forms dimers (or higher order oligomers), generating a ligand-binding site at the interface (Chipman and Shaanan 2001; Liberles et al. 2005). The ligand is often an amino acid or pyrimidine. The details of the regulation provided by the ACT domain in the AAAHs are unknown. Currently, the ACT domain of PAH is the only structural representative of all regulatory domains of the aromatic amino acid hydroxylases (AAAHs), which also includes tyrosine hydroxylase (TH) and tryptophan hydroxylases 1 and 2 (TPH1 and TPH2). PAH differs from other established ACT domain containing proteins since there are no contacts between the ACT domains in tetrameric PAH (Flatmark and Stevens 1999). Moreover, mammalian PAH is allosterically regulated by its own substrate but in spite of reports about L-Phe binding both to the active site as a substrate and at the regulatory domain as an activator (Gjetting et al. 2001; Shiman 1980; Shiman et al. 1994), we have previously shown that the ACT-domain of human PAH does not bind L-Phe, which exclusively binds at the active site (Martinez et al. 1993; Thorolfsson et al. 2002). This property is most probably extended to the ACT domain of other mammalian forms of the enzyme ((Liberles et al. 2005) and references therein). To our knowledge, binding of amino acids or substrate analogs at the regulatory domain of TH or the TPHs has not been reported or suggested. This is unlike other established homologs in the ACT domain family, such as the regulatory domains of D-3-phosphoglycerate dehydrogenase (3-PGDH) and prephenate dehydratase (PheA) [or the bifunctional chorismate mutase-prephenate dehydratase (*P*-protein)], where binding of regulatory amino acids has been established (Chipman and Shaanan 2001; Pohnert et al. 1999). The ACT domain of PheA presents two amino acid motifs (GAL and I/LESRP), separated by 17 residues, which were shown to be involved in L-Phe binding and feed back inhibition in the *P*-protein from *E. coli* (Pohnert et al. 1999). These motifs are conserved in PAH and only slightly mutated in the other mammalian AAAHs, i.e. SS/AL, I/LETRP in TH; GGL and IESRK/R in the TPHs which suggest that the amino acid-binding function of these ACT domains might have adapted to another function

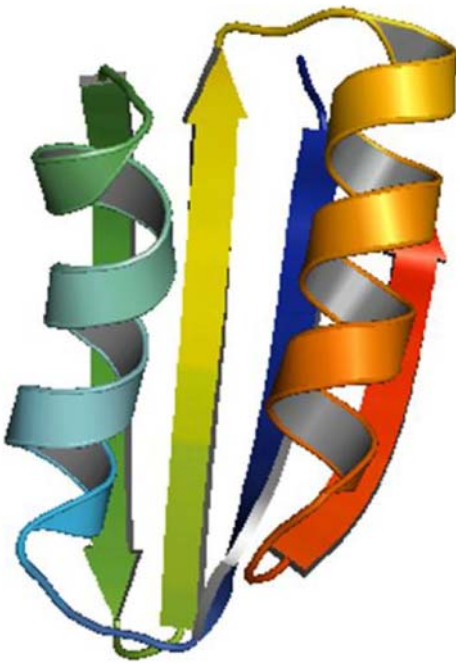

Fig. 1 A guide to the ACT domain structure. The ACT domain displays a ferredoxin-like fold, β1 (*blue*), α1 (*cyan*), β2 (*green*), β3 (*yellow*), α2 (*orange*), and β4 (*red*). Here 3-PGDH (PDB id 1YBA) is shown. ACT domain ligands frequently bind on the odd loop side (Aravind and Koonin 1999; Chipman and Shaanan 2001). The remaining figures of ACT domain structures in this article will be from a similar view, unless otherwise noted. This and remaining figures are made in PyMol (DeLano 2002)

at an early stage of the AAAH evolution. Elucidation of the evolutionary relationships of the ACT domains in PAH and in other ACT domains in general is a challenging task that may provide insights into the function of the domain itself and the regulation of the AAAHs.

The Pfam database, which classifies domain families upon sequence homology (Bateman et al. 2004), includes several ACT domain proteins originally sampled by (Aravind and Koonin 1999) using a PSI-BLAST (position-specific iterating-BLAST) sequence database search (Table 1). The database for structural classification of proteins, SCOP (Murzin et al. 1995), classifies 3-PGDH, PAH and GCVR within the ferredoxin-like fold, $\beta\alpha\beta\beta\alpha\beta$, as members of the ACT-like superfamily (Table 1). This SCOP superfamily also includes the regulatory domain of threonine deaminase (TD; an ACT-like domain by Pfam), which is involved in the metabolism of L-Ile, and the C-terminal domain of the nickel responsive regulator (NikR) (Table 1). In the other commonly used database for classification of protein domain structures, CATH (Orengo et al. 1997; Pearl et al. 2000, 2005) (Version 3.0.0), 3-PGDH is classified as a 2-layer α/β-plaits sandwich—also known as the ferredoxin-like fold—with two structural homologs, GCVR and D-ribose-5-phosphate isomerase (Rpia) (Table 1). The ferredoxin-like fold constitutes a superfold, which is a fold with at least three different homologous superfamilies that do not seem to be related, as first recognized by Orengo et al. (1994) who identified 10 superfolds when comparing folds and superfamilies in the CATH database. This particular superfold, the two-layer α/β-plaits sandwich, is shared by >70 different superfamilies according to CATH. Surprisingly, the ACT domain of PAH is not classified as one of these. Instead, the whole multidomain PAH protein has its own fold category despite having three functional, evolutionary units (Flatmark and Stevens 1999). For the remaining of this article the term ferredoxin-like fold will also implicitly designate the two-layer α/β-plaits sandwich. In addition to these ACT-domains already classified in the structural databases, coordinates for several of the proteins originally described by Aravind and Koonin have recently been deposited to the PDB database (see recent review, Grant 2006) (Table 1). Further, the leucine-responsive regulatory protein A (LRPA) and the regulatory domain of ATP-PRTASE (HisG_C) which both have been associated with the ACT domains (Ettema et al. 2002; Grant 2006), are also included in Table 1.

To summarize, there is a lack of consensus in the classification of the ACT domain in the above databases, which perhaps can be explained by the high evolutionary mobility of the ACT domains, their high sequence divergence, its presence in multidomain proteins, and the abundance of the ferredoxin-like fold (Anantharaman et al. 2001). Additionally, multi-domain proteins often arise by fusion of a domain at the N or C terminal of e.g. an enzyme, altering enzyme specificity or regulation (Bjorklund et al. 2005). Given this and our interest to increase our understanding of the regulatory function of the ACT domain in PAH and the other AAAHs within an evolutionary context, we initiated the search for so far unrecognized putative ancient homologs of this domain. PAH has low but noticeable sequence similarity with the regulatory domains of PheA and 3-PGDH (Aravind and Koonin 1999; Gjetting et al. 2001; Liberles et al. 2005), but in general there is far too low sequence identity to attempt the investigations of evolutionary relationships on the basis of sequence alignments among ACT domain proteins. We here present the results using a combination of the FATCAT server (Ye and Godzik 2003), which is a flexible protein structure comparison algorithm, and the SSAP server (Taylor and Orengo 1989), which is based on a rigid body protein structure comparison algorithm.

Methods

Searching the PDB

A consensus approach of the FATCAT (Flexible structure AlignmenT by Chaining Aligned fragment pairs allowing Twists) bioinformatic server (Ye and Godzik 2003; Ye and Godzik 2004, 2005) and the SSAP algorithm (Taylor and Orengo 1989) used to annotate homologs in the CATH database (Orengo et al. 1997; Pearl et al. 2000, 2005) was utilized. As a first step, the ACT domain of PAH (residues 30–112 of PDB id 1PHZ) was submitted to the FATCAT server and searched against the PDB database (10 January 2007, 40% non-redundant set with 9,553 chains). A P value of $\leq$0.001 (which means that such a similarity could have happened by chance once in 1,000) or a P value of <0.05 and a FATCAT similarity score (FATCAT score) >100 were the criteria for statistically significant matches.

The FATCAT server used in the initial PDB search was originally developed to identify homologous flexible structures with different conformations (Ye and Godzik 2003), and when compared with two different rigid structure alignment programs, DALI and CE, it performed better. DALI stopped at hinge positions, where FATCAT inserted a twist, while CE continued and thereby aligned non-homologous residues (Ye and Godzik 2003). Given these properties FATCAT was selected even though we did not apply twisted alignments in this work.

Confirming the results of the PDB search

The SSAP algorithm (Taylor and Orengo 1989), used when assigning homologs in the CATH database (Orengo et al.

Table 1 Proteins with ACT and ACT-like domain proteins (known ACT domains; KADs) for which the 3D structures are presently available

Family representative	PDB id	PFAM[a]	SCOP[a]	CATH[a]
3-PGDH; D-3-phosphoglycerate dehydrogenase, regulatory C-terminal domain (*E. coli*; *Mycobacterium tuberculosis*)	1PSD, 1SC6, 1YGY, 1YBA[b]	X	X	X
AHAS; acetoacetate synthase isozyme III small subunit[c] *(E. coli)*	2F1F	X		
AK; aspartokinase[c] *(Arabidopsis thaliana)*	2CDQ			
GCVR; putative glycine cleavage system transcriptional repressor[c] (*Vibrio cholerae*)	1U8S	X	X	X
HisG_C; ATP phosphoribosyltransferase (*Mycobacterium tuberculosis*)	1NH7			
LRPA; transcriptional regulator *(Pyrococcus furiosus)*	1I1G[d]			
NikR; nickel responsive regulator, C-terminal regulatory domain (*E. coli*)	1Q5Y, 2BJ1[b]		X	
PAH; phenylalanine hydroxylase, N-terminal regulatory domain (*Rattus norvegicus*)	1PHZ, 2PHM	X	X	
PheA; putative prephenate dehydratase *(Staphylococcus aureus)*	2IQ8			
Rpia; D-ribose-5-phosphate isomerase, C-terminal lid domain (*E. coli*)	1O8B			X
SGUF1; hypothetical protein af1403[c] (*Archaeoglobus fulgidus*)	1Y7P	X		
SGUF2; structural genomics unknown function (*Streptococcus pneumoniae*)	1ZPV	X		
SGUF3; protein BT0572[c] *(Bacteroides thetaiotaomicron)*	2F06	X		
TD; allosteric threonine deaminase[c], C-terminal domains (*E. coli*)	1TDJ	X[e]	X	
YbeD *(E. coli)*	1RWU			
YkoF[c] *(Bacillus subtilis)*	1SBR			

This table includes the proteins originally described by Aravind and Koonin (1999) with coordinates at the PDB database, and/or recently recognized as such (Grant 2006). Rpia was also included since it is classified in CATH as structural homolog of 3-PGDH and GCVR

Reference (if published) and residue range used (within parenthesis) for those part of this study, for PDB: 1PSD (Schuller et al. 1995), 1SC6 (Bell et al. 2004), 1YGY (Dey et al. 2005), 1YBA (336–410) (Thompson et al. 2005), 2F1F (1–69) (Kaplun et al. 2006), 2CDQ (338–418) (Mas-Droux et al. 2006), 1U8S (3–83), 1NH7 (210–284) (Cho et al. 2003), 1I1G (64–135) (Leonard et al. 2001), 1Q5Y (Schreiter et al. 2003), 2BJ1 (53–133) (Chivers and Tahirov 2005), 1PHZ (30–112) (Kobe et al. 1999), 2PHM (Kobe et al. 1999), 2IQ8 (179–264) replaced in PDB by PDB id 2QMW, 1O8B (127–198) (Zhang et al. 2003), 1Y7P (2–82), 1ZPV (−2–83), 2F06 (4–64), 1TDJ (336–423) (Gallagher et al. 1998), 1RWU (13–87) (Kozlov et al. 2004), and 1SBR (9–86) (Devedjiev et al. 2004)

[a] The sequential database Pfam includes eight gene families for which the structure has been determined and the proteins have been classified as ACT domain (PF01842) or ACT-like domain (PF00585). SCOP has five members in the ferredoxin-like fold classified as ACT-like superfamily (55021). CATH classifies the ACT domain of 3-PGDH, GCVR and Rpia as the two-layer α/β-plaits sandwich (3.30.70.260). CATH does not handle the ACT domain of PAH as a self-standing domain, but integrates it with the catalytic domain of the enzyme in a unique fold (1.10.800.10)

[b] The actual PDB id that was detected in the FATCAT search and used in this study

[c] These proteins have ACT domain repeats. The first ACT domain was used in this study

[d] This is a structural representative of the RAM domains. It has been proposed that the ACT and the RAM domains might have a common ancestor (Ettema et al. 2002)

[e] TD is not classified as an ACT domain, but as an ACT-like domain (PF00585)

1997; Pearl et al. 2000, 2005) was used to confirm the structures identified by FATCAT as potential ACT domains, given the domain boundaries from the FATCAT search. The CATH criteria for assigning homologs at the superfamily level is SSAP score > 80, sequence identity $\geq$ 20%, and at least 60% overlap, or SSAP score $\geq$70, $\geq$60% overlap and related functions, as described in the literature or Pfam. Here, the criterion of a SSAP score $\geq$80 was used. The FATCAT results that were confirmed by the SSAP algorithm were divided into known ACT domains (KADs) and potential ACT domains (PADs). Further, pairwise SSAP comparisons—extracting scores and RMSD values—were performed for all ACT domain candidates that met both the FATCAT and SSAP criteria when compared to PAH. SSAP scores are inversely correlated to distance, assuming that high SSAP score equals a shorter evolutionary distance. Therefore, we used the negative pairwise SSAP scores as distances for building a UPGMA distance tree using Phylip (Felsenstein 1989) to show the relationships due to the SSAP comparisons.

Alignments

Multiple structurally based sequence alignments were built using Mustang (Konagurthu et al. 2006). For the analysis of the results, pairwise structural alignments were built using CE, where a z score of >3.9 equals a P value of <0.001 (Shindyalov and Bourne 1998, 2001), and visualized using PyMol (DeLano 2002), as implemented and provided by STRAP (Gille and Frommel 2001).

Results and discussion

The FATCAT/SSAP search provides putative ACT domain candidates

In the FATCAT search (Ye and Godzik 2004) with PAH as a query, 76 hits of potential ACT domains with a statistically significant P value < 0.001 and 50 hits ranging from P value > 0.001 to 0.05 and with a score of >100 were obtained. For all, the requirement of untwisted alignments (Ye and Godzik 2003) was met. These were further compared to PAH by SSAP (Taylor and Orengo 1989), using the same domain boundaries that had been identified by FATCAT in the initial search. The SSAP comparisons reduced the hit sampling to 32 structures that in addition to the FATCAT criterion also matched the criteria of SSAP score >80. However, if there were more than one hit from the same protein family with >30% sequence identity to each other, solely one entry was kept. Accordingly, five redundant hits were removed, resulting in a list of 28 structures. This list contained 10 (including the query PAH) of the already known ACT or ACT-like domains (KADs) (Table 1) and 18 potential ACT domain (PADs) candidates (Table 2). The detected KADs are 3-PGDH, AHAS, AK, GCVR, HisG_C, NikR, PheA, SGUF 1, and SGUF 2, although for AK the P value for the structural

Table 2 The ensemble of potential ACT domain (PADs) containing proteins, identified in the structural search by FATCAT and SSAP

Name; description	PDB id	P value	FATCAT score	Alignment Length (residues)	Sequence identity (%)
ACYP; acylphosphatase *(Bacillus subtilis)*[a]	2FHM[b]	4.72e-05	111.80	71	10.47
ALY; transcriptional coactivator *(Mus musculus)*	1NO8	5.11e-04	94.41	70	11.11
Atx1; metallochaperone *(Saccharomyces cerevisiae)*	1CC7	2.16e-05	114.73	71	6.17
CCS; metallochaperone of superoxide dismutase *(Saccharomyces cerevisiae)*	1JK9	5.07e-03	114.85	71	6.17
CutA1; periplasmic divalent cation tolerance protein *(Archaeoglobus fulgidus)*	1P1L	5.45e-06	139.97	75	3.26
Duf190; putative nitrogen regulatory protein/ PII-like signaling protein *(Thermotoga maritima)*	1O51	2.22e-05	115.89	76	3.37
MSCS; small-conductance mechanosensitive channel *(E. coli)*	2OAU	6.81e-03	113.94	73	4.82
PIIA; nitrogen regulatory protein P-II 2 *(E. Coli)*	2NS1	1.53e-05	138.15	76	4.90
PIIB; hypothetical protein TTHA0516 *(Thermus thermophilus)*	2CZ4	1.01e-06	142.61	76	6.12
S6A; 30S ribosomal protein S6 *(Thermus thermophilus)*	1CQM[c]	9.86e-05	113.90	76	8.60
S6B; 30S ribosomal protein S6 *(Aquifex aeolicus)*	2J5A	4.96e-05	119.95	82	9.38
SGUF4; hypothetical protein *(Thermotoga maritime)*	2NZC	4.22e-06	125.35	80	14.63
SGUF5; hypothetical protein TTHA1053 *(Thermus thermophilus)*	2CVE	2.13e-03	107.79	68	11.69
SGUF6; unknown protein[d] *(Galdieria sulphuraria)*	2NYI	2.19e-04	131.06	81	5.68
SGUF7; hypothetical protein, function unknown *(Helicobacter pylori)*	2ATZ	5.93e-04	128.14	71	5.88
V-ATPase; head domain of subunit C *(Saccharomyces cerevisiae)*	1U7L	1.14e-02	128.27	75	7.14
VAO; vanillyl-alcohol oxidase[d] *(Penicillium simplicissimum)*	1AHU	4.64e-02	129.43	77	3.26
Znta; zinc (lead, cadmium, and mercury) transporting ATPase *(E. coli)*	1MWY	1.37e-04	104.49	71	7.41

These domains, not previously classified as ACT domains, meet the criteria of (i) either a P value ≤ 0.001 or a P value < 0.05 and a FATCAT score > 100, and (ii) a SSAP score > 80 (see Table 3 for SSAP-scores) for the domain boundaries identified in the FATCAT search in comparison to the ACT domain of PAH (PDB 1PHZ, residues 30–112)

[a] The structural similarity of another ACYP domain to the ACT domain has been noted elsewhere (Rosano et al. 2002)

[b] PDB id 2FHM was substituted with another ACYP (PDB id 1V3Z), since the latter was published at the time of writing while the former was not

[c] PDB id 1CQM (Otzen et al. 2000) corresponds to a mutated sequence, and was therefore substituted with PDB id 1G1X (Agalarov et al. 2000)

[d] This protein has ACT domain repeats. The second ACT domain was used in this study

Reference (if published) and residue range used (within parenthesis) for those part of this study, for PDB: 2FHM, 1V3Z (3–91) (Miyazono et al. 2005), 1NO8 (105–182) (Perez-Alvarado et al. 2003), 1CC7 (2–73) (Rosenzweig et al. 1999), 1JK9 (B: 3–74) (Lamb et al. 2001), 1P1L (1–87), 1O51 (−1–100) (Schwarzenbacher et al. 2004), 2OAU (180–262) (Bass et al. 2002), 2NS1 (B:0.97), 2CZ4 (1–92), 1CQM (Otzen et al. 2000), 1G1X (1–98) (Agalarov et al. 2000), 2J5A (3–108) (Olofsson et al. 2007), 2NZC (2–81), 2CVE (124–191), 2NYI (90–176), 2ATZ (71–151), 1U7L (189–263) (Drory et al. 2004), 1AHU (419–507) (Mattevi et al. 1997), and 1MWY (2–74) (Banci et al. 2002)

Table 3 Pairwise RMSD values (above the diagonal) and SSAP score (beneath the diagonal)

	1ahu	1cc7	1g1x	1i1g	1jk9	1mwy	1nh7	1no8	1o51	1o8b	1p1l	1phz	1rwu	1sbr	1tdj	1u7l	1u8s
VAO[a]	0	3.2	3.8	4.1	3.5	2.9	2.8	5.4	3.0	2.8	3.4	2.6	3.1	3.1	3.3	2.9	2.9
Atx1[a]	80.4	0	4.5	3.4	1.7	1.9	5.0	3.4	2.4	3.0	2.2	2.9	2.9	3.7	2.6	3.2	2.3
S6A[a]	78.4	72.2	0	2.8	5.3	3.00	4.2	5.4	5.5	2.8	3.3	3.7	3.0	4.2	3.3	3.0	2.6
Lrpa	75.1	82.1	81.5	0	3.2	2.9	3.1	3.9	4.0	2.5	2.7	4.0	2.9	3.4	3.5	4.1	4.1
CCS[a]	78.1	91.1	72.2	79.0	0	1.7	4.9	3.7	3.4	3.9	2.4	3.5	3.5	3.3	2.7	2.8	4.8
Znta[a]	80.3	88.8	77.8	81.6	88.5	0	2.7	3.6	3.4	3.8	2.5	3.4	2.9	4.1	2.9	3.0	2.9
HISG_C[a]	79.8	76.2	80.8	81.9	75.0	80.1	0	4.3	3.0	3.5	2.6	2.7	3.4	6.7	4.7	3.0	2.5
ALY[a]	68.2	80.7	71.5	79.2	78.8	78.9	76.1	0	3.1	4.9	4.4	3.4	4.2	4.8	4.2	3.9	3.1
Duf190[a]	81.1	82.1	77.1	76.4	78.6	79.8	81.3	80.5	0	3.5	2.6	2.7	2.9	3.5	5.9	3.0	3.2
Rpia	76.5	81.0	78.7	86.2	75.8	78.2	78.6	72.9	77.1	0	2.9	4.3	2.9	5.1	3.00	2.8	4.2
CutA1[a]	77.1	82.7	80.7	80.00	80.4	81.2	82.4	78.4	84.7	77.9	0	3.2	3.0	3.4	4.00	2.6	3.3
PAH	82.9	82.4	78.0	76.8	81.0	82.2	83.5	80.8	82.1	74.3	81.0	0	3.2	4.2	3.2	3.0	2.8
YbeD	78.7	78.9	78.6	79.1	74.0	77.7	76.5	76.5	77.5	77.9	76.4	77.7	0	4.1	3.6	2.5	3.4
YkoF	79.4	78.9	72.6	77.6	75.1	77.9	73.9	76.0	79.2	70.3	78.1	78.6	75.2	0	3.8	4.5	3.3
TD	77.6	78.8	78.3	78.3	79.3	80.1	76.7	75.0	76.8	75.1	77.9	79.8	76.9	73.1	0	3.4	4.0
VATPase[a]	81.3	83.0	77.2	76.5	81.2	81.8	85.1	74.9	78.9	83.4	82.3	82.9	78.7	74.2	77.6	0	3.3
GCVR	82.4	84.0	83.0	79.3	78.1	80.7	80.5	80.5	81.6	79.1	81.3	85.4	77.3	78.5	78.6	80.5	0
ACYP[a]	81.6	81.7	81.1	82.3	76.2	75.8	79.4	72.7	82.5	74.8	80.1	80.8	81.5	80.5	78.5	77.2	84.2
SGUF1	78.6	84.6	84.1	82.1	80.4	81.8	84.0	75.7	79.8	81.3	81.6	84.1	77.4	77.3	81.6	82.0	85.7
3-PGDH	80.6	84.9	83.1	84.1	79.3	83.4	80.2	79.2	77.5	77.7	76.8	81.3	80.3	76.3	81.0	76.3	86.0
SGUF 2	79.9	82.5	82.6	78.0	77.4	80.3	76.3	78.7	79.3	79.2	78.7	81.9	77.6	75.6	75.4	80.8	89.7
SGUF 7[a]	78.3	79.8	76.9	77.7	78.0	80.1	79.5	75.9	79.5	79.4	73.7	81.0	78.0	74.2	68.8	76.7	80.5
NikR	82.0	83.7	81.5	83.0	80.0	79.0	78.3	76.0	81.7	75.3	78.9	83.7	77.5	80.8	75.8	81.2	85.4
AK	79.8	82.0	82.4	82.2	79.0	80.9	78.9	80.9	80.0	77.9	75.0	84.1	77.5	76.8	81.4	77.6	83.7
SGUF 5[a]	77.4	85.4	77.0	81.0	83.5	82.2	85.1	74.4	81.1	81.8	82.6	83.5	75.6	80.5	77.9	87.5	83.0
PIIB[a]	82.9	79.7	79.0	77.3	77.2	78.6	82.7	79.7	87.7	73.6	84.2	82.9	79.1	78.0	77.4	80.5	84.0
SGUF 3	77.7	81.4	76.0	76.6	82.5	80.9	78.0	77.0	78.2	77.2	77.4	77.0	73.8	72.6	78.8	81.3	81.5
AHAS	79.0	85.0	82.6	83.2	81.7	83.3	83.1	79.9	80.0	80.8	80.3	83.4	79.9	74.2	82.6	82.7	86.0
PheA	79.1	81.7	81.2	80.2	80.3	81.2	79.2	77.4	79.5	74.4	77.4	83.5	81.2	76.1	81.1	81.0	84.5
S6B[a]	78.2	76.5	87.5	77.3	73.0	77.0	77.3	72.9	78.9	73.8	75.4	79.1	75.1	76.8	78.3	73.8	82.4
PIIA[a]	77.5	80.6	78.1	76.6	76.6	74.7	80.4	74.5	83.3	74.7	85.0	80.5	75.7	73.8	77.0	79.4	79.4
SGUF 6[a]	78.7	82.7	84.2	73.8	79.6	81.3	79.1	73.9	77.2	77.2	81.3	82.1	80.0	75.0	80.2	81.0	90.1
SGUF 4[a]	81.7	84.0	84.2	84.2	80.9	81.7	85.5	82.8	82.3	75.1	78.0	85.7	75.8	81.8	82.3	84.0	86.7
MSCS[a]	83.6	81.2	79.7	78.1	80.6	83.2	77.2	75.0	77.5	79.7	77.4	80.5	81.0	76.2	76.3	79.4	81.1
	1v3z	1y7p	1yba	1zpv	2atz	2bj1	2cdq	2cve	2cz4	2f06	2f1f	2iq8	2j5a	2ns1	2nyi	2nzc	2oau
VAO[a]	3.1	3.9	2.9	3.6	4.2	3.4	3.2	3.7	2.6	2.9	2.8	4.1	3.2	3.4	3.4	3.3	2.9
Atx1[a]	2.5	2.5	2.8	2.7	3.2	2.8	3.4	2.4	2.9	3.0	2.2	3.2	3.6	2.4	2.5	3.0	2.8
S6A[a]	4.5	2.2	2.4	3.0	4.2	3.4	2.5	3.2	3.7	2.8	2.2	3.0	2.5	3.2	4.0	3.2	2.5
Lrpa	2.6	3.4	3.3	4.0	3.4	3.1	3.1	2.9	4.3	3.9	3.3	3.2	3.5	3.2	5.1	3.8	3.2
CCS[a]	3.4	3.1	4.4	4.5	3.4	4.1	3.6	2.9	3.3	2.8	2.8	4.3	4.0	3.2	3.0	3.7	3.0
Znta[a]	4.6	3.3	2.8	3.3	3.0	4.0	3.6	3.1	3.3	2.7	2.5	3.5	3.2	3.8	2.9	3.7	2.1
HISG_C[a]	3.6	3.2	3.3	4.2	4.1	3.5	3.7	2.5	2.6	3.5	2.5	3.3	3.4	3.4	3.1	2.7	3.8
ALY[a]	4.9	4.6	3.7	3.3	4.3	5.2	3.1	4.2	3.5	3.1	3.5	3.8	4.2	5.3	6.6	3.5	3.9
Duf190[a]	3.7	4.3	3.2	4.2	3.4	3.0	4.0	2.6	2.5	3.0	3.6	3.3	4.2	2.5	5.2	3.0	2.9
Rpia	3.8	3.0	4.0	3.8	3.0	4.4	3.5	3.1	4.1	3.9	3.0	4.8	4.0	3.6	3.9	4.9	2.5
CutA1[a]	2.9	3.0	3.2	3.7	4.7	3.3	3.7	2.2	5.3	2.8	2.4	3.5	3.6	2.8	5.0	4.5	4.4

Table 3 continued

	1v3z	1y7p	1yba	1zpv	2atz	2bj1	2cdq	2cve	2cz4	2f06	2f1f	2iq8	2j5a	2ns1	2nyi	2nzc	2oau
PAH	3.1	3.4	3.0	3.5	3.7	2.7	2.9	2.3	2.6	2.4	2.3	3.1	3.4	2.7	3.1	2.6	3.4
YbeD	2.5	3.2	3.0	3.5	3.9	3.2	3.9	2.7	4.2	3.8	2.9	4.0	3.7	3.1	3.5	4.5	3.1
YkoF	3.1	3.4	3.0	4.2	6.6	2.9	4.9	2.6	3.7	4.0	6.7	3.4	3.1	3.2	3.5	3.2	5.1
TD	5.3	3.0	2.8	3.8	5.0	4.6	2.8	2.8	3.9	2.1	2.1	3.4	3.4	3.2	2.9	2.9	5.4
VATPase[a]	3.9	2.9	3.5	3.4	4.0	3.3	3.5	2.1	2.7	2.8	2.6	3.4	3.5	2.9	2.9	3.3	3.5
GCVR	2.6	2.5	2.1	2.0	3.8	2.5	3.2	2.3	2.8	2.1	1.7	3.0	2.1	3.7	1.3	2.5	3.5
ACYP[a]	0	3.5	2.3	4.2	3.8	3.9	3.1	3.0	3.7	3.1	2.7	2.5	3.4	3.1	3.4	2.8	2.6
SGUF1	81.1	0	1.7	3.1	3.9	3.3	2.5	2.6	2.9	2.3	1.5	2.6	3.8	3.4	3.7	3.0	2.8
3-PGDH	82.2	88.6	0	2.6	3.1	2.9	1.9	3.4	3.5	2.1	1.7	2.4	2.5	2.8	2.1	2.3	3.2
SGUF 2	80.1	83.2	83.6	0	3.7	3.8	2.8	3.5	3.5	3.3	2.2	3.0	3.1	4.4	1.9	3.1	3.0
SGUF 7[a]	78.8	78.6	82.1	80.4	0	4.3	4.2	3.2	4.2	3.2	3.2	3.6	3.4	4.6	3.9	4.2	2.7
NikR	79.0	84.0	79.4	79.2	76.3	0	4.1	2.4	2.9	3.1	2.2	3.1	4.1	3.2	3.0	1.9	4.0
AK	81.5	83.9	86.5	84.5	78.8	79.5	0	3.6	3.7	2.5	2.6	3.1	3.2	3.7	3.2	3.2	3.5
SGUF 5[a]	78.5	81.0	77.2	78.1	80.4	82.7	82.1	0	2.9	3.0	2.5	3.7	3.1	2.1	2.9	2.3	4.0
PIIB[a]	80.8	80.9	78.5	78.6	75.4	80.4	78.7	81.9	0	2.8	3.5	4.6	3.9	2.0	4.1	3.2	3.8
SGUF 3	74.4	81.9	82.4	79.0	75.8	76.7	81.4	82.0	76.3	0	2.2	2.2	3.0	2.9	2.5	2.9	3.2
AHAS	81.0	88.7	89.4	83.0	79.0	84.7	83.8	83.0	77.2	86.4	0	2.1	2.4	3.0	1.9	3.0	3.0
PheA	82.0	85.0	85.3	83.3	80.4	82.0	85.9	79.6	75.6	80.6	85.5	0	3.5	3.1	3.3	3.1	4.1
S6B[a]	79.7	76.6	79.7	79.4	75.2	77.4	78.7	76.6	74.7	75.0	78.6	77.7	0	3.5	4.0	3.5	2.7
PIIA[a]	76.4	77.6	78.0	75.1	73.4	78.5	75.2	81.8	84.9	75.9	74.5	77.5	75.2	0	4.0	2.9	3.0
SGUF 6[a]	78.9	82.8	85.4	89.2	78.5	79.6	84.3	82.2	81.1	79.0	84.0	82.1	79.2	79.5	0	2.6	3.5
SGUF 4[a]	80.3	87.3	87.2	84.0	77.7	88.6	85.5	85.7	81.4	78.4	87.3	85.7	80.4	80.9	82.2	0	4.1
MSCS[a]	80.5	80.9	82.0	81.1	81.5	79.2	81.9	78.9	79.8	75.9	82.5	76.5	78.5	75.1	80.8	80.6	0

The *first column* gives the abbreviation and the *first row* gives the corresponding PDB id

[a] Potential ACT domains (PADs)

similarity detected by FATCAT to the ACT domain of PAH is at the limit of significance (P value < 0.05). For 3-PGDH and NikR the FATCAT search selected a different PDB id than in the previous annotations (1YBA and 2BJ1, respectively). LRPA, TD, and YkoF were all selected by the FATCAT search procedure, but did not pass the selection criterion of SSAP score >80 (i.e., the values were 76.78, 79.78, and 78.60 for LRPA, TD, and YkoF, respectively), while Rpia, SGUF 3, and YbeD were not present in the FATCAT hit sampling. Except for SGUF 3, these proteins are in fact known for having structures which do not conform to the archetypical ACT-domain architecture (see references in Table 1). Moreover, LPRA is a structural representative of the RAM domain, which is found to be fused to a DNA-binding domain of prokaryotic transcriptional regulator proteins. Structural and functional associations between the ACT and RAM domains have been previously highlighted (Ettema et al. 2002), but the firm identification of LPRA, and also of HisG_C, as ACT domains have not been performed yet.

The sequence identities are often very low (Table 2), and these potential ACT domain homologs would be difficult to identify without a structural approach. Pairwise SSAP calculations were performed for all KADs (Table 1) and all PADs selected by the FATCAT/SSAP procedure (Table 2) and the resulting SSAP score and RMSD outputs are shown in Table 3. For a complete list of pairwise sequence identity and length of alignments see Table S1 in Supplementary material. An UPGMA distance tree (Phylip software, Felsenstein 1989) based on the negative SSAP score as distance is shown in Fig. 2. The second column in the tree shows the Pfam classification and the third column shows the ligand if known. The tree shows that the negative SSAP scores appear as good distance measurements since it is noteworthy that the protein relations correlate well with the Pfam classification of the proteins. The amino acid-binding domains form one cluster, while the domains containing metallobinding motifs form another cluster, which also includes SGUF3. PAH, which has lost the ACT domain dimerization and amino acid-binding function, and NikR, which is a metallobinding domain that uses a different metallobinding motif than Atx1, CCS, and ZntA, form a cluster together with SGUF4. This cluster is located between the metallobinding domains and the amino

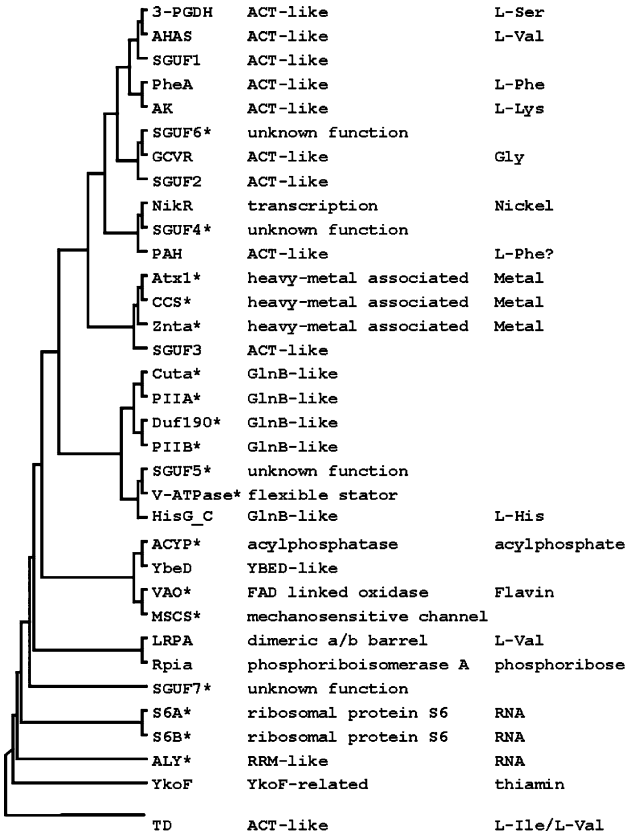

Fig. 2 Distance tree generated with the SSAP-scores as distances. Phylip software (Felsenstein 1989) (UPGMA) was utilized to reveal the relationships between the proteins from SSAP-scores in Table 3. The tree was visualized using TreeView (Page 1996). The *first column* is the abbreviated name, the *second column* refers to their Pfam domain family classification (TD and the rest of the ACT classified as ACT-like), and the third column gives the ligand (if known)

acid-binding domains (Fig. 2), suggesting a bridge between the two clusters. The homology between the regulatory domains of 3-PGDH, AHAS, AK and PheA (Aravind and Koonin 1999; Chipman and Shaanan 2001), all allosterically regulated by an amino acid, was confirmed by this approach. Also the GlnB-like and the RNA-binding proteins form separate clusters. With respect to the SGUF proteins, the grouping pattern suggests amino acid-binding capabilities for SGUF 1, SGUF 2, and SGUF 6, while metallobinding seems plausible for SGUF 3 and SGUF 4.

Structure-based multiple sequence alignment: detecting similarities at extreme functional and sequence divergence

The total sampling of PADs and KADs were joined in a multiple structure-based sequence alignment prepared by Mustang (Konagurthu et al. 2006) (Fig. 3). Known ligand-binding positions are highlighted in the alignment (Fig. 3). There are no strictly conserved positions but there are semi-conserved positions, mostly of a hydrophobic character. Most ACT domain proteins for which the ligand bound structures are known use loop 1 for ligand binding, i.e. 3-PGDH (Schuller et al. 1995), YkoF where tiamin binds to first of the two tandem ACT domains (Devedjiev et al. 2004), and AK (Mas-Droux et al. 2006) (Fig. 3). Also, the metallobinding proteins Znta, CCS, and Atx1, with the motif CXXC (Banci et al. 2002; Lamb et al. 2001; Rosenzweig et al. 1999) and ACYP (Rosano et al. 2002) use loop 1 for ligand binding (Fig. 3). Further, Asn41, located in the third loop, is critical for the enzymatic activity of ACYP (Stefani et al. 1997). The ACT domain of NikR, which also is a metallobinding protein, does not use the same metallobinding motif in loop 1, but instead the end of the second and beginning of the third β-strands, plus the end of the first α-helix from an adjacent chain (Chivers and Tahirov 2005). The GAL motif in PAH, which is part of the original ACT domain consensus, is located in the same area as the metallobinding motif and the G is present in Atx1 (Figs. 3, 4a, c). This loop is highly similar in PAH and Atx1, and this entire alignment has a CE z score of 4.2, and a RMSD of 2.64 Å over 70 residues. Further, the second α-helix in these metallobinding domains is rather peculiar looking. Despite that, five residues in this helix are conserved between PAH and CCS (Fig. 4b, d). The PAH motif IESRP (in the second β-strand) seems to be specific to the ACT in the AAAHs in addition to PheA, and is not detected in the ACT-domain candidates. There is however a motif that resembles IESR in GCVR, namely, IDSR, although shifted two residues towards the N-terminus in the alignment (Fig. 3). The N-terminals of most sequences, and several of the semi-conserved positions in the loop regions (Fig. 3), are relatively Gly rich. Most structures have at least one Gly or one Pro in most loops; these residues are frequently involved in forming specific loop structures. Interestingly, in ACYP several conserved Gly residues have been found to simply prevent aggregation, which also is an important driving force for the evolution of proteins (Parrini et al. 2005).

The metallobinding domains provide regulation, which is a common theme among the ACT-domains. Together, the RMSD and the SSAP-scores (Table 3, Fig. 2) suggest that the metallobinding domains are strong ACT domain candidates and distant homologs of the amino acid-binding ACT domains. Reinforcing this idea is the similarities in the ligand-binding sites, which also point to ACYP as a distant homolog. ACYP is a small enzyme—it basically consists of a ferredoxin-like fold with one extra β-strand—that hydrolyses the carboxyl-phosphate bond of acylphosphates (Stefani et al. 1997), the only enzyme among the PADs (Table 2). The substrate, e.g. 1,3-bisphosphoglycerate, carbamoylphosphate, succinylphosphate, acetylphosphate,

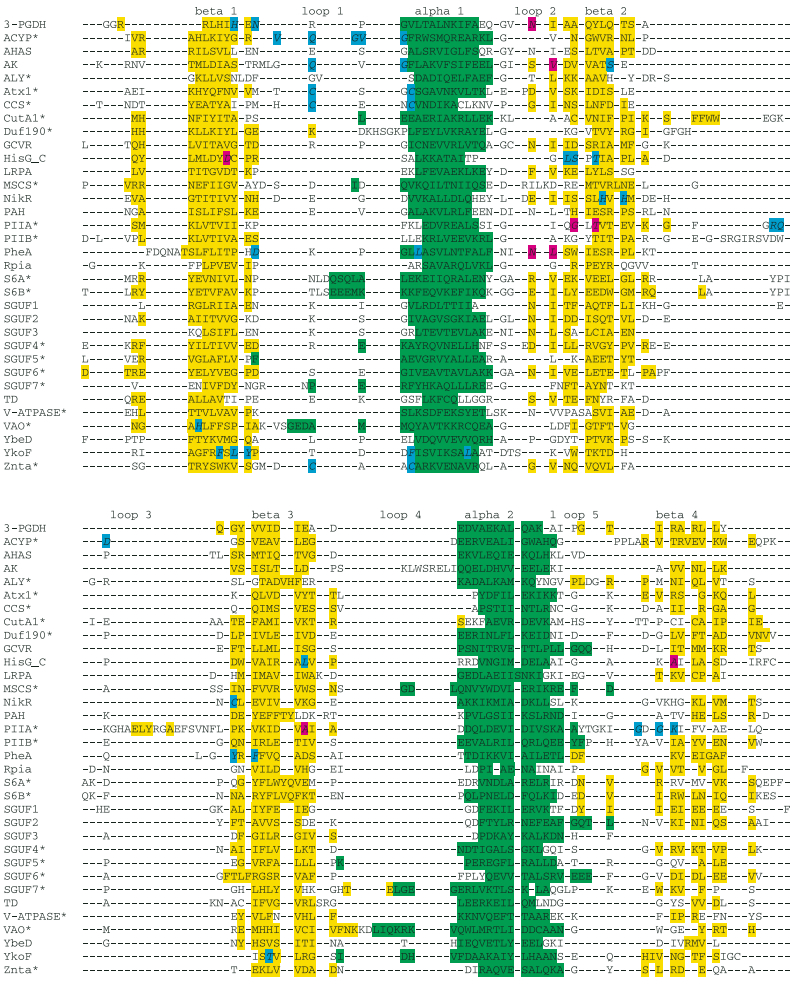

Fig. 3 Sequence alignment. **a** A multiple structural/sequence alignment was prepared for the known ACT domains (KADs) and selected potential ACT domains (PADs)*. The alignment was generated using Mustang (Konagurthu et al. 2006). The secondary structure elements, i.e. β-strands (*yellow*) and α-helices (*green*), are shown for each corresponding PDB structure (Tables 1, 2). Ligand-binding residues are shown in turquoise, and in the case of two ACT domains contributing to one binding site, the ligand-binding residues from the other chain are marked in *purple* (Banci et al. 2002; Chivers and Tahirov 2005; Devedjiev et al. 2004; Lamb et al. 2001; Mas-Droux et al. 2006; Mattevi et al. 1997; Rosano et al. 2002; Rosenzweig et al. 1999; Schuller et al. 1995; Stefani et al. 1997). The ligand-binding residues for PheA are the equivalent residues that bind L-Phe in PheA2 (PDB id 2QMX)

or β-aspartylphosphate (Stefani et al. 1997), binds at the end of the first β-strand and the following loop 1 (Fig. 3) (Corazza et al. 2006). The structural similarity of HypF ACYP-like domain with both the metallobinding domain Atx1 and 3-PGDH has been noted previously by Rosano et al. (2002). ACP also shows relatively high sequence identity (19% SSAP id, see Supplementary material) with the ribosomal protein S6 1. S6 does not bind small ligands, but RNA and other proteins. Analysis of the residue composition for all proteins in Table 3, comparatively with the folding nucleus of ribosomal proteins S6 1 and S6 2 (V6, I8, I26, L30, V65, V72, L75, and L79) (Lindberg et al. 2006; Olofsson et al. 2007) shows a high degree of conservation (Fig. 3). The residues in the folding nucleus (the foldons) are believed to be evolutionary conserved (Mirny and Shakhnovich 2001) even at extreme sequence divergence, further supporting that these domains might share a common ancestry.

Another interesting cluster with high structural similarity is the one formed by the PAD GlnB-like proteins CutA1, Duf190, PIIA (also known as GlnK), and PIIB; HisG_C is also close (Fig. 2). Duf190 is a domain with unknown function but CutA1 provides divalent cation tolerance in microorganisms, and binding Cu(II) has been shown to have an impact on its multimerization (Tanaka et al. 2004) and eukaryote CutA1 has been assigned a potential role in signal transduction (Arnesano et al. 2003). The PII proteins are encoded by the *glnB* gene, hence the name of this superfamily. The PII proteins, which basically consist of a ferredoxin-like fold with an extended loop 3 (Fig. 3), interact with many different proteins in the nitrogen regulatory pathway. PII proteins bind ATP and 2-ketoglutarate (Kamberov et al. 1995). They are uridylated by GlnD to increase L-Gln synthesis (Kamberov et al. 1994). GlnD was part of the original ACT domain sampling by Aravind and Koonin (1999), but no structures of the GlnD domain are available yet. HisG is a multidomain enzyme that synthesizes L-His in some bacteria, plants, and fungi (Bond and Francklyn 2000). The regulatory domain of HisG, HisG_C, controls the feed-back regulation of HisG by switching the conformation from the active dimer to the inactive hexamer upon L-His binding (Cho et al. 2003).

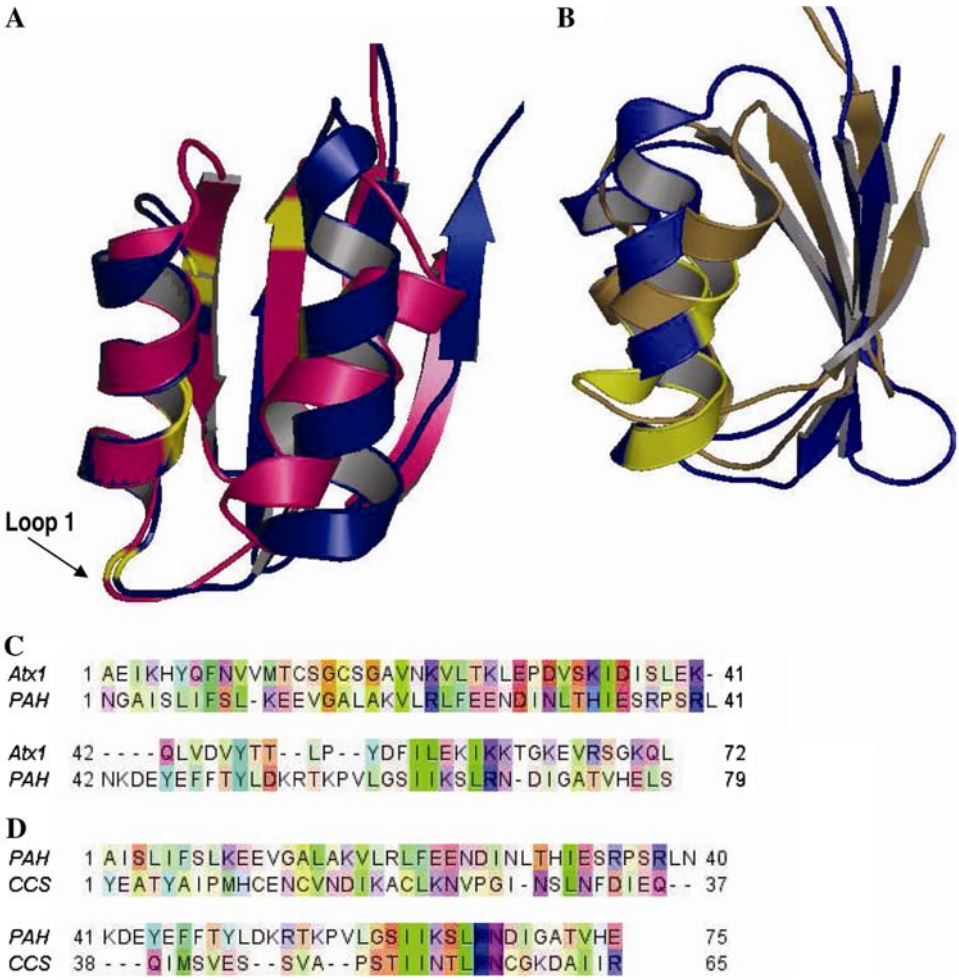

Fig. 4 Representative superimpositions with metallobinding domains. (**a**, **b**) Structural and (**c**, **d**) corresponding sequence alignments of Atx1 and PAH (**a**, **c**), and CCS and PAH (**b**, **d**). PAH (*blue*, PDB id 1PHZ), Atx1 (*pink*, PDB id 1CC7), and CCS (*sand*, PDB id 1JK9). Atx1 (*pink*) in (**a**) has the G in the GAL motif conserved, inside the metallobinding motif, located in loop 1. CCS (*sand*) in (**b**) has five residues in the second α-helix in common with PAH (*blue*), shown as a side view since these conserved residues are more or less buried. (**a**, **c**) Atx1 (PDB id 1CC7) and PAH (PDB id 1PHZ); CE *z* score 4.2, RMSD 2.64 Å over 70 residues. (**b**, **d**) CCS (PDB id 1JK9) and PAH (PDB id 1PHZ); CE *z* score 4.2, RMSD 2.65 Å over 65 residues. All conserved residues are shown in yellow in the structural alignments (**a**, **b**)

This conformational flexibility and regulation upon amino acid binding is typical for the ACT domains, and this domain is indeed very similar to the ACT domain of PAH. A pairwise structural alignment made using CE (Shindyalov and Bourne 1998, 2001) has a *z* score of 4.4 and a RMSD of 2.16 Å over 66 aligned positions (Fig. 5a, c). Further, as both CutA1 and HisG_C are part of the highly diverse GlnB-superfamily, a pairwise CE structural alignment of HisG_C and CutA1 (Fig. 5b, d) has a *z* score of 4.1, and a RMSD of 2.54 Å over 69 residues, which indeed is comparable to the alignment of HisG_C and PAH (Fig. 5a, c), which might indicate an ancient link between the GlnB and the typical ACT domains. Nevertheless, the lack of functional binding similarities between the amino acid binding ACT domains and the GlnB- and the RNA-binding proteins renders the identification of these proteins as ancient homologs of the ACT domain of PAH as very speculative.

The FATCAT/SSAP approach thus contributes to the identification proteins with relevant structural (and partially functional) similarity to the ACT domains and which would not have been detected with basic sequence searches, as also observed by Ettema et al. (2002) on their comparative studies of the RAM- and ACT-domains. Still, this structural superimposition approach failed to detect other proteins which have been associated with the ACT and ACT-like domains (LRPA, TD, YkoF, Rpia, SGUF 3, and YbeD). This is most probably an effect of the choice of query structure and strict cut-offs. The use of several other ACT domain proteins through a structural consensus as query structure, as well as less stringent selection criteria would have aided to capture other ACT-like domain proteins. However, with the use of less strict selection criteria the number of non-homologous proteins would increase even further beyond functional significance.

Implications for the regulation of PAH

Emerging from the clustering around PAH in the distance tree (Fig. 2) is an ancestral-binding functionality for the ACT domain in this enzyme, i.e. a role as a metallobinding

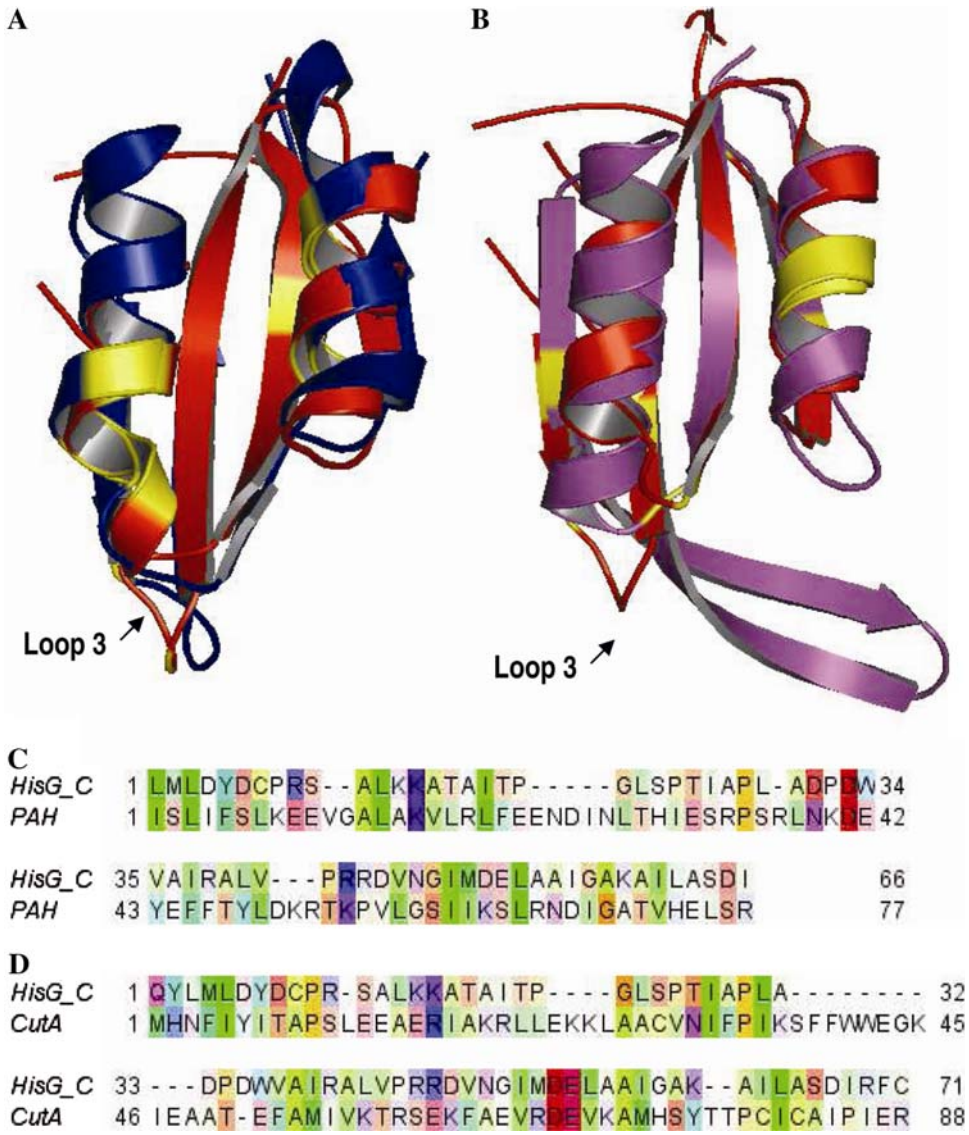

Fig. 5 Representative superimpositions with GlnB-like domains. (**a**, **b**) Structural and (**c**, **d**) corresponding sequence alignments of PAH and HisG_C (**a**, **c**), and CutA1 and HisG_C (**b**, **d**). HisG_C (*red*) in (**a**, **c**) and CutA1 (magenta) in (**b**, **d**) have the P in the IESRP motif in PAH conserved, located at the end of the third βstrand, adjacent to loop 3. Loop 3 in CutA1 (*magenta*) in (**b**) is extended as in many other GlnB-like domains, although not in HisG_C (red). (A,C) PAH (blue, PDB id 1PHZ) and HisG_C (PDB id 1NH7); CE z score 4.4, RMSD 2.16 Å over 66 residues. (**b**, **d**) HisG_C (PDB id 1NH7) and CutA1 (PDB id 1P1L); CE z score 4.1, RMSD 2.54 Å over 69 residues. All conserved residues are marked in yellow in the structural alignments (**a**, **b**)

domain (see above). For PAH and the rest of the AAAHs this premise leads to the interesting hypothesis that an ancient function of their ACT domains might have been to store and provide iron to their catalytic domains, which require non-heme iron for catalysis (Fitzpatrick 2003). The structural alignment of PAH and metallochaperone Atx1 (Fig. 4a, c) reveals a stunningly similar loop 1, where the metallobinding site is located.

In the light of this work it is also interesting to discuss functional and structural differences between the ACT domain in PAH and in the other amino acid-binding proteins. PAH clusters with amino acid-binding proteins in the distance tree (Fig. 2) and the strong evolutionary relationships between PAH and PheA (Pohnert et al. 1999) support an original L-Phe-binding function for the ACT domain of PAH (as found for PheA). Despite this, PAH appears to have derived enough to miss the perfect clustering obtained for the other amino acid-binding proteins (Fig. 2), a finding that it is maybe associated with the fact that the ACT domains in PAH do not dimerize (Heil et al. 2002; Kobe et al. 1999). In the active PAH tetramer there would be no contact between ACT domains which interact with the catalytic domain of the adjacent subunit in the dimer (Flatmark and Stevens 1999). On the other hand, in the structures of ACTs complexed with amino acids the binding site is often generated at the dimer interface (Chipman and Shaanan 2001; Liberles et al. 2005; Schuller et al. 1995). Moreover, for human PAH it has been shown that only 1 mol L-Phe binds per mol subunit at the active site in the catalytic domain (Thorolfsson et al. 2002). Discussions regarding a putative ability of PAH to bind L-Phe in the regulatory domain have been centered on the GAL-IESRP motifs (Gjetting et al. 2001) which are also involved in the binding of L-Phe in PheA and *P*-Protein (Pohnert et al.

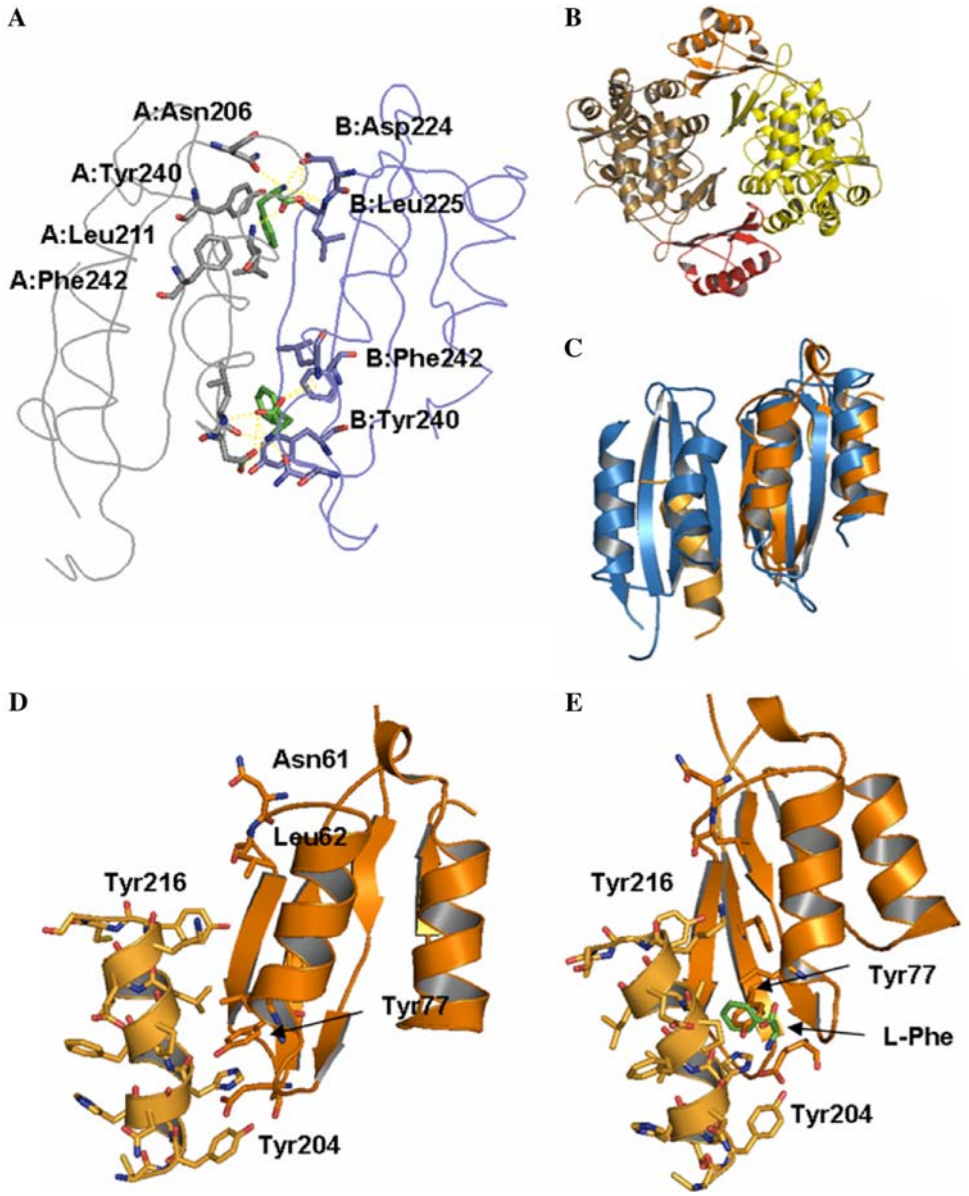

Fig. 6 Interdomain interactions in rat PAH compared to the L-Phe binding site in PheA2. A comparison of the L-Phe binding site in PheA (PDB id 2QMX) and the equivalent region of PAH (PDB id 2PHM, Kobe et al. 1999). **a** The dimeric structure of PheA2 shown as ribbon. The ACT domains are shown in *grey* and *blue*, respectively. The dimer interface provides two identical binding sites for L-Phe (*green*). The residues involved in coordination of L-Phe are shown in *stick*. **b** The truncated dimeric structure of PAH from rat (PDB id 1PHM) is shown as *cartoon*. The catalytic domain of chain A (*brown*) interacts with the regulatory domain of chain B (*orange*), while the catalytic domain of chain B (*yellow*) interacts with the regulatory domain of chain A (*red*). No contacts are formed between the two regulatory ACT domains (**a** red, **b** orange). **c** A close-up view of one ACT domain (*orange*) superimposed with PheA2 (*blue*) and the interaction to the adjacent catalytic domain (*sand*) in PAH. **d** A close-up view of one ACT domain (*orange*) and the interaction to the adjacent catalytic domain (*sand*) in PAH. The residues corresponding to those involved in L-Phe binding in PheA2, and other residues from the adjacent α-helix involved in important interactions in this region, are shown in *stick*. **e** Same as in **d**, with L-Phe (*green*) the superimposed PheA2 structure (not shown)

1999). In PAH, the motif is at the interface between the interacting regulatory and catalytic domains. A recent structure of PheA2 from *Chlorobium tepidum TLS* shows L-Phe bound to its regulatory ACT domain (PDB id 2QMX). L-Phe is coordinated by Asn206 (Asp in PheA), Leu211, Tyr240, and Phe242, plus Asp224 (Asn in PheA) and Leu225 from the adjacent ACT domain (Fig. 3). Only Leu211 is part of the GAL-IESRP motif that has earlier been involved in L-Phe binding, but the GAL-IESRP motif is in the ligand-binding region. The conservation of these ligand-binding residues among PheA, PheA2, and PAH is high (Fig. 3).

The ligand-binding region of PheA2 is located at the dimer interface between two ACT regulatory domains (Fig. 6a), and the amino acid-binding site is similar to that in 3-PGDH. There are no contacts between the different ACT domains in PAH (Fig. 6b), but a superimposition of PAH and PheA2 reveals that a α-helix from the catalytic domain in the adjacent chain in PAH is located in a similar position to that in the second ACT domain in PheA2 (Fig. 6c). However, the α-helix from the adjacent catalytic domain that interacts in PAH with the homologous region to the ligand-binding site in PheA2, only spans halfway up the ligand-binding region as seen in PheA2 (Fig. 6c). Therefore, the chance of L-Phe being able to bind to the equivalent of the upper binding site (Asn61 and Leu62 in PAH, equivalent to Asp224 and Leu225 in PheA2) (Fig. 6a) is excluded simply because the adjacent chain is missing (Fig. 6d). Binding L-Phe to an Asp and Leu solely is not likely. For the lower binding site (Fig. 6a), the equivalent of Asp224 and Leu225 are missing from PAH. The α-helix from the adjacent domain is flanked by Tyr204 and Tyr216 which occupy the amino acid-binding region preventing access for L-Phe (Fig. 6d). The interface is rich in aromatic stacking interactions as well as hydrogen bonds. Tyr77, equivalent to Tyr240 that stacks L-Phe in the PheA2-binding site, forms a hydrogen bond with Ile209 of the adjacent catalytic domain, which probably causes it to reach further into the ligand-binding site. In fact, a superimposition of PheA2 with L-Phe and the ACT domain of PAH shows how the hydroxyl group on Tyr77 reaches straight through the aromatic ring in L-Phe (Fig. 6e). Thus, it seems plausible that the amino acid-binding capacity in this region of PAH might have been lost along evolution. In mammalian PAH changes would have occurred in concert with the interacting catalytic domain, instead of with another ACT domain as is the case in PheA.

Conclusion

There are as expected additional distant homologs of the ACT domain within the ferredoxin-like superfold and our work has highlighted some putative potential candidates. Studies of other superfolds have found evidence for homologous relationships between several different superfamilies (Kiel and Serrano 2006; Qi and Grishin 2005). Energetic and structural genomics considerations have shown that the superfolds can support a much broader repertoire of sequences than other folds (Orengo and Thornton 2005; Shakhnovich et al. 2003). The potential distant ACT domain homologs detected here show a conserved regulatory theme among them. Analysis of these additional domains led us to an improvement of the hypotheses regarding the ancestral function of the ACT domain to include metallobinding or chaperone-like properties.

There is a limited set of superfolds, although their total contribution to the domain pool is very large (Orengo and Thornton 2005); the methodological approach presented here may contribute to improve functional annotation and to understand the evolutionary relationships of superfamilies and superfolds.

Acknowledgments We are grateful to Prof. Randy Lewis, University of Wyoming, for providing a good research environment for JSL. This research was supported by The Research Council of Norway and Helse-Vest.

References

Agalarov SC, Sridhar Prasad G, Funke PM, Stout CD, Williamson JR (2000) Structure of the S15,S6,S18-rRNA complex: assembly of the 30S ribosome central domain. Science 288:107–113

Anantharaman V, Koonin EV, Aravind L (2001) Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol 307:1271–1292

Aravind L, Koonin EV (1999) Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol 287:1023–1040

Arnesano F, Banci L, Benvenuti M, Bertini I, Calderone V, Mangani S, Viezzoli MS (2003) The evolutionarily conserved trimeric structure of CutA1 proteins suggests a role in signal transduction. J Biol Chem 278:45999–46006

Banci L, Bertini I, Ciofi-Baffoni S, Finney LA, Outten CE, O'Halloran TV (2002) A new zinc-protein coordination site in intracellular metal trafficking: solution structure of the Apo and Zn(II) forms of ZntA(46–118). J Mol Biol 323:883–897

Bass RB, Strop P, Barclay M, Rees DC (2002) Crystal structure of Escherichia coli MscS, a voltage-modulated and mechanosensitive channel. Science 298:1582–1587

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR (2004) The Pfam protein families database. Nucleic Acids Res 32: D138–D141

Bell JK, Grant GA, Banaszak LJ (2004) Multiconformational states in phosphoglycerate dehydrogenase. Biochemistry 43:3450–3458

Bjorklund AK, Ekman D, Light S, Frey-Skott J, Elofsson A (2005) Domain rearrangements in protein evolution. J Mol Biol 353:911–923

Bond JP, Francklyn C (2000) Proteobacterial histidine-biosynthetic pathways are paraphyletic. J Mol Evol 50:339–347

Chipman DM, Shaanan B (2001) The ACT domain family. Curr Opin Struct Biol 11:694–700

Chivers PT, Tahirov TH (2005) Structure of Pyrococcus horikoshii NikR: nickel sensing and implications for the regulation of DNA recognition. J Mol Biol 348:597–607

Cho Y, Sharma V, Sacchettini JC (2003) Crystal structure of ATP phosphoribosyltransferase from *Mycobacterium tuberculosis*. J Biol Chem 278:8333–8339

Corazza A, Rosano C, Pagano K, Alverdi V, Esposito G, Capanni C, Bemporad F, Plakoutsi G, Stefani M, Chiti F, Zuccotti S, Bolognesi M, Viglino P (2006) Structure, conformational stability, and enzymatic properties of acylphosphatase from the hyperthermophile Sulfolobus solfataricus. Proteins 62:64–79

DeLano WL (2002) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto. http://www.pymol.org

Devedjiev Y, Surendranath Y, Derewenda U, Gabrys A, Cooper DR, Zhang RG, Lezondra L, Joachimiak A, Derewenda ZS (2004) The structure and ligand binding properties of the B. subtilis YkoF gene product, a member of a novel family of thiamin/HMP-binding proteins. J Mol Biol 343:395–406

Dey S, Grant GA, Sacchettini JC (2005) Crystal structure of *Mycobacterium tuberculosis* D-3-phosphoglycerate dehydrogenase: extreme asymmetry in a tetramer of identical subunits. J Biol Chem 280:14892–14899

Drory O, Frolow F, Nelson N (2004) Crystal structure of yeast V-ATPase subunit C reveals its stator function. EMBO Rep 5:1148–1152

Ettema TJ, Brinkman AB, Tani TH, Rafferty JB, Van Der Oost J (2002) A novel ligand-binding domain involved in regulation of amino acid metabolism in prokaryotes. J Biol Chem 277:37464–37468

Felsenstein J (1989) PHYLIP: phylogeny inference package (Version 3.2). Cladistics 5:164–166

Fitzpatrick PF (2003) Mechanism of aromatic amino acid hydroxylation. Biochemistry 42:14083–14091

Flatmark T, Stevens RC (1999) Structural insight into the aromatic amino acid hydroxylases and their disease-related mutant forms. Chem Rev 99:2137–2160

Gallagher DT, Gilliland GL, Xiao G, Zondlo J, Fisher KE, Chinchilla D, Eisenstein E (1998) Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase. Structure 6:465–475

Gille C, Frommel C (2001) STRAP: editor for STRuctural Alignments of Proteins. Bioinformatics 17:377–378

Gjetting T, Petersen M, Guldberg P, Guttler F (2001) Missense mutations in the N-terminal domain of human phenylalanine hydroxylase interfere with binding of regulatory phenylalanine. Am J Hum Genet 68:1353–1360

Grant GA (2006) The ACT domain: a small molecule binding domain and its role as a common regulatory element. J Biol Chem 281:33825–33829

Heil G, Stauffer LT, Stauffer GV (2002) Glycine binds the transcriptional accessory protein GcvR to disrupt a GcvA/GcvR interaction and allow GcvA-mediated activation of the Escherichia coli gcvTHP operon. Microbiology 148:2203–2214

Kamberov ES, Atkinson MR, Feng J, Chandran P, Ninfa AJ (1994) Sensory components controlling bacterial nitrogen assimilation. Cell Mol Biol Res 40:175–191

Kamberov ES, Atkinson MR, Ninfa AJ (1995) The Escherichia coli PII signal transduction protein is activated upon binding 2-ketoglutarate and ATP. J Biol Chem 270:17797–17807

Kaplun A, Vyazmensky M, Zherdev Y, Belenky I, Slutzker A, Mendel S, Barak Z, Chipman DM, Shaanan B (2006) Structure of the regulatory subunit of acetohydroxyacid synthase isozyme III from *Escherichia coli*. J Mol Biol 357:951–963

Kiel C, Serrano L (2006) The ubiquitin domain superfold: structure-based sequence alignments and characterization of binding epitopes. J Mol Biol 355:821–844

Kobe B, Jennings IG, House CM, Michell BJ, Goodwill KE, Santarsiero BD, Stevens RC, Cotton RG, Kemp BE (1999) Structural basis of autoregulation of phenylalanine hydroxylase. Nat Struct Biol 6:442–448

Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: a multiple structural alignment algorithm. Proteins 64:559–574

Kozlov G, Elias D, Semesi A, Yee A, Cygler M, Gehring K (2004) Structural similarity of YbeD protein from Escherichia coli to allosteric regulatory domains. J Bacteriol 186:8083–8088

Lamb AL, Torres AS, O'Halloran TV, Rosenzweig AC (2001) Heterodimeric structure of superoxide dismutase in complex with its metallochaperone. Nat Struct Biol 8:751–755

Leonard PM, Smits SH, Sedelnikova SE, Brinkman AB, de Vos WM, van der Oost J, Rice DW, Rafferty JB (2001) Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. Embo J 20:990–997

Liberles JS, Thorolfsson M, Martinez A (2005) Allosteric mechanisms in ACT domain containing enzymes involved in amino acid metabolism. Amino Acids 28:1–12

Lindberg MO, Haglund E, Hubner IA, Shakhnovich EI, Oliveberg M (2006) Identification of the minimal protein-folding nucleus through loop-entropy perturbations. Proc Natl Acad Sci USA 103:4083–4088

Martinez A, Olafsdottir S, Flatmark T (1993) The cooperative binding of phenylalanine to phenylalanine 4-monooxygenase studied by 1H-NMR paramagnetic relaxation: changes in water accessibility to the iron at the active site upon substrate binding. Eur J Biochem 211:259–266

Mas-Droux C, Curien G, Robert-Genthon M, Laurencin M, Ferrer JL, Dumas R (2006) A novel organization of ACT domains in allosteric enzymes revealed by the crystal structure of Arabidopsis aspartate kinase. Plant Cell 18:1681–1692

Mattevi A, Fraaije MW, Mozzarelli A, Olivi L, Coda A, van Berkel WJ (1997) Crystal structures and inhibitor binding in the octameric flavoenzyme vanillyl-alcohol oxidase: the shape of the active-site cavity controls substrate specificity. Structure 5:907–920

Mirny L, Shakhnovich E (2001) Evolutionary conservation of the folding nucleus. J Mol Biol 308:123–129

Miyazono K, Sawano Y, Tanokura M (2005) Crystal structure and structural stability of acylphosphatase from hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. Proteins 61:196–205

Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247:536–540

Olofsson M, Hansson S, Hedberg L, Logan DT, Oliveberg M (2007) Folding of S6 structures with divergent amino acid composition: pathway flexibility within partly overlapping foldons. J Mol Biol 365:237–248

Orengo CA, Thornton JM (2005) Protein families and their evolution: a structural perspective. Annu Rev Biochem 74:867–900

Orengo CA, Jones DT, Thornton JM (1994) Protein superfamilies and domain superfolds. Nature 372:631–634

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM (1997) CATH: a hierarchic classification of protein domain structures. Structure 5:1093–1108

Otzen DE, Kristensen O, Oliveberg M (2000) Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: a structural clue to amyloid assembly. Proc Natl Acad Sci USA 97:9907–9912

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Parrini C, Taddei N, Ramazzotti M, Degl'Innocenti D, Ramponi G, Dobson CM, Chiti F (2005) Glycine residues appear to be evolutionarily conserved for their ability to inhibit aggregation. Structure (Camb) 13:1143–1151

Pearl FM, Lee D, Bray JE, Sillitoe I, Todd AE, Harrison AP, Thornton JM, Orengo CA (2000) Assigning genomic sequences to CATH. Nucleic Acids Res 28:277–282

Pearl F, Todd A, Sillitoe I, Dibley M, Redfern O, Lewis T, Bennett C, Marsden R, Grant A, Lee D et al (2005) The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. Nucleic Acids Res 33:D247–251

Perez-Alvarado GC, Martinez-Yamout M, Allen MM, Grosschedl R, Dyson HJ, Wright PE (2003) Structure of the nuclear factor ALY: insights into post-transcriptional regulatory and mRNA nuclear export processes. Biochemistry 42:7348–7357

Pohnert G, Zhang S, Husain A, Wilson DB, Ganem B (1999) Regulation of phenylalanine biosynthesis: studies on the mechanism of phenylalanine binding and feedback inhibition in the *Escherichia coli* P-protein. Biochemistry 38:12212–12217

Qi Y, Grishin NV (2005) Structural classification of thioredoxin-like fold proteins. Proteins 58:376–388

Rosano C, Zuccotti S, Bucciantini M, Stefani M, Ramponi G, Bolognesi M (2002) Crystal structure and anion binding in the prokaryotic hydrogenase maturation factor HypF acylphosphatase-like domain. J Mol Biol 321:785–796

Rosenzweig AC, Huffman DL, Hou MY, Wernimont AK, Pufahl RA, O'Halloran TV (1999) Crystal structure of the Atx1 metallochaperone protein at 1.02 A resolution. Structure 7:605–617

Schreiter ER, Sintchak MD, Guo Y, Chivers PT, Sauer RT, Drennan CL (2003) Crystal structure of the nickel-responsive transcription factor NikR. Nat Struct Biol 10:794–799

Schuller DJ, Grant GA, Banaszak LJ (1995) The allosteric ligand site in the Vmax-type cooperative enzyme phosphoglycerate dehydrogenase. Nat Struct Biol 2:69–76

Schwarzenbacher R, von Delft F, Abdubek P, Ambing E, Biorac T, Brinen LS, Canaves JM, Cambell J, Chiu HJ, Dai X, Deacon AM, DiDonato M, Elsliger MA, Eshagi S, Floyd R, Godzik A, Grittini C, Grzechnik SK, Hampton E, Jaroszewski L, Karlak C, Klock HE, Koesema E, Kovarik JS, Kreusch A, Kuhn P, Lesley SA, Levin I, McMullan D, McPhillips TM, Miller MD, Morse A, Moy K, Ouyang J, Page R, Quijano K, Robb A, Spraggon G, Stevens RC, van den Bedem H, Velasquez J, Vincent J, Wang X, West B, Wolf G, Xu Q, Hodgson KO, Wooley J, Wilson IA (2004) Crystal structure of a putative PII-like signaling protein (TM0021) from *Thermotoga maritima* at 2.5 A resolution. Proteins 54:810–813

Shakhnovich BE, Dokholyan NV, DeLisi C, Shakhnovich EI (2003) Functional fingerprints of folds: evidence for correlated structure-function evolution. J Mol Biol 326:1–9

Shiman R (1980) Relationship between the substrate activation site and catalytic site of phenylalanine hydroxylase. J Biol Chem 255:10029–10032

Shiman R, Xia T, Hill MA, Gray DW (1994) Regulation of rat liver phenylalanine hydroxylase. II: Substrate binding and the role of activation in the control of enzymatic activity. J Biol Chem 269:24647–24656

Shindyalov IN, Bourne PE (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng 11:739–747

Shindyalov IN, Bourne PE (2001) A database and tools for 3-D protein structure comparison and alignment using the combinatorial extension (CE) algorithm. Nucleic Acids Res 29:228–229

Stefani M, Taddei N, Ramponi G (1997) Insights into acylphosphatase structure and catalytic mechanism. Cell Mol Life Sci 53:141–151

Tanaka Y, Tsumoto K, Nakanishi T, Yasutake Y, Sakai N, Yao M, Tanaka I, Kumagai I (2004) Structural implications for heavy metal-induced reversible assembly and aggregation of a protein: the case of *Pyrococcus horikoshii* CutA. FEBS Lett 556:167–174

Taylor WR, Orengo CA (1989) Protein structure alignment. J Mol Biol 208:1–22

Thompson JR, Bell JK, Bratt J, Grant GA, Banaszak LJ (2005) Vmax regulation through domain and subunit changes: the active form of phosphoglycerate dehydrogenase. Biochemistry 44:5763–5773

Thorolfsson M, Ibarra-Molero B, Fojan P, Petersen SB, Sanchez-Ruiz JM, Martinez A (2002) L-Phenylalanine binding and domain organization in human phenylalanine hydroxylase: a differential scanning calorimetry study. Biochemistry 41:7573–7585

Ye Y, Godzik A (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics 19(Suppl 2):II246–II255

Ye Y, Godzik A (2004) FATCAT: a web server for flexible structure comparison and structure similarity searching. Nucleic Acids Res 32:W582–W585

Ye Y, Godzik A (2005) Multiple flexible structure alignment using partial order graphs. Bioinformatics 21:2362–2369

Zhang R, Andersson CE, Savchenko A, Skarina T, Evdokimova E, Beasley S, Arrowsmith CH, Edwards AM, Joachimiak A, Mowbray SL (2003) Structure of *Escherichia coli* ribose-5-phosphate isomerase: a ubiquitous enzyme of the pentose phosphate pathway and the Calvin cycle. Structure 11:31–42