Research article

Open Access

The catalytic domains of thiamine triphosphatase and CyaB-like adenylyl cyclase define a novel superfamily of domains that bind organic phosphates

Lakshminarayan M Iyer and L Aravind*

Address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA E-mail: Lakshminarayan M Iyer - lakshmin@ncbi.nlm.nih.gov; L Aravind* - aravind@ncbi.nlm.nih.gov *Corresponding author

Published: 27 November 2002

BMC Genomics 2002, 3:33

Received: 19 September 2002 Accepted: 27 November 2002

This article is available from: http://www.biomedcentral.com/1471-2164/3/33

© 2002 Iyer and Aravind; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The CyaB protein from *Aeromonas hydrophila* has been shown to possess adenylyl cyclase activity. While orthologs of this enzyme have been found in some bacteria and archaea, it shows no detectable relationship to the classical nucleotide cyclases. Furthermore, the actual biological functions of these proteins are not clearly understood because they are also present in organisms in which there is no evidence for cyclic nucleotide signaling.

Results: We show that the CyaB like adenylyl cyclase and the mammalian thiamine triphosphatases define a novel superfamily of catalytic domains called the CYTH domain that is present in all three superkingdoms of life. Using multiple alignments and secondary structure predictions, we define the catalytic core of these enzymes to contain a novel $\alpha+\beta$ scaffold with 6 conserved acidic residues and 4 basic residues. Using contextual information obtained from the analysis of gene neighborhoods and domain fusions, we predict that members of this superfamily may play a central role in the interface between nucleotide and polyphosphate metabolism. Additionally, based on contextual information, we identify a novel domain (called CHAD) that is predicted to functionally interact with the CYTH domain-containing enzymes in bacteria and archaea. The CHAD is predicted to be an alpha helical domain, and contains conserved histidines that may be critical for its function.

Conclusions: The phyletic distribution of the CYTH domain suggests that it is an ancient enzymatic domain that was present in the Last Universal Common Ancestor and was involved in nucleotide or organic phosphate metabolism. Based on the conservation of catalytic residues, we predict that CYTH domains are likely to chelate two divalent cations, and exhibit a reaction mechanism that is dependent on two metal ions, analogous to nucleotide cyclases, polymerases and certain phosphoesterases. Our analysis also suggests that the experimentally characterized members of this superfamily, namely adenylyl cyclase and thiamine triphosphatase, are secondary derivatives of proteins that performed an ancient role in polyphosphate and nucleotide metabolism.

Background

Organic phosphate compounds are the central metabo-

lites of all biological systems [1,2]. Some are the basic building blocks of nucleic acids, some like ATP and GTP,

are additionally, cellular energy stores, others like cAMP or cGMP are messengers in signal transduction, and, yet others, such as FAD, NAD, thiamine phosphates and pyridoxal phosphate are cofactors for a range of enzymes [1,2]. Protein domains belonging to a relatively small set of structural folds are known to bind or catalyze reactions that utilize these organic phosphate compounds (see SCOP database: [http://scop.mrc-lmb.cam.ac.uk/scop/]) [3,4]. Several of these folds trace back to some of the earliest phases of the evolution of the protein world, and participate in a wide range of disparate biological functions in extant proteins [4,5]. Some folds, such as the P-loop fold, the Rossmann fold and the Hsp70-like fold, have been well studied, and comprise mainly of dedicated nucleotide binding or hydrolyzing proteins [6-9]. Others, such as the palm-domain, which is found in adenylyl cyclases and various nucleic acid polymerases, belong to more generalized protein folds that contain representatives with diverse biochemical activities [4,10,11]. Current availability of extensive genome sequence data, allows one to identify less numerous, nevertheless biological important organic phosphate-binding domains that may have previous eluded detection. The identification of such domain superfamilies, containing enzymes with several different activities, often throws considerable light on their evolution, structure and catalytic mechanisms [4,12].

The majority of previously known nucleotide cyclases belong to two major folds. The classical adenylyl cyclases, guanylyl cyclases and the GGDEF (diguanylate cyclase) domains share the catalytic palm domain with the family B DNA polymerases, reverse transcriptases, viral RNA dependent RNA polymerases and eukaryote-type primases [4,13,14]. The pathogenic adenylyl cyclases of several bacteria and the CyaA-like proteobacteria adenylyl cyclases are extremely divergent versions of the catalytic domain seen in the Pol- β family of nucleotidyl transferases [15] SCOP database: [http://scop.mrc-(also see lmb.cam.ac.uk/scop/]). While the catalytic domains of these two superfamilies have very different folds, they follow a similar reaction mechanism that is dependent on two Mg²⁺ ions coordinated by a cluster of acidic residues. However, the CyaB adenylyl cyclase, which was identified in the bacterium Aeromonas hydrophila is unrelated to any of these above superfamilies of enzymes [16]. Though close relatives of this enzyme exist in some bacteria, like Yersinia pestis and Borrelia burgdorferi and the archaea, its antecedents or catalytic mechanism have not been understood. Using sensitive sequence profile comparison methods, we show that the CyaB-like adenylyl cyclases are homologs of the soluble mammalian thiamine triphosphatases [17], and they define a novel superfamily of enzymes that utilize ATP and other organic phosphates. We present evidence that a representative of this domain was present in the last common ancestor of all extant life forms. The primary biological function of these proteins appears to be related to polyphosphate and nucleotide metabolism. Cyclic AMP generation and thiamine triphosphate hydrolysis appear to be secondarily acquired activities. We also identify the potential active site- and substrate interacting-residues and postulate that these enzymes are likely to catalyze a two-metal ion dependent reaction on structural scaffold that is completely different from that seen in the other two superfamilies of adenylyl cyclases.

Results and discussion Identification of the CYTH domain

In order to understand the evolutionary affinities and provenance of the CyaB adenylyl cyclase from Aeromonas hydrophila, we carried out database searches using sensitive sequence profile analysis methods. As CyaB is a small protein with no detectable low complexity regions, we used it as a seed to initiate a PSI-BLAST search [18] (run to convergence, with expect-value for inclusion in profile = .01). The search resulted in the detection of its obvious orthologs from Yersinia and various archaea at significant expect (e) values ranging from from 3×10^{-43} to 8×10^{-5} . The second iteration recovered proteins from more archaea, eukaryotes (e-value: 3×10^{-7}), Clostridium (3×10^{-7}) ⁸) and Borrelia (6×10^{-8}) . Further iterations, run to convergence, recovered the soluble mammalian thiamine triphosphatases (3×10^{-4}) , and the N-terminal region of E. coli YgiF. At convergence, several bacterial proteins, that showed a conserved EXEXK (where X is any amino acid) characteristic of this family, and a region C-terminal to a P-loop like uridine kinase domain in plants were also recovered at borderline e-values (e value ~0.01 - 0.05). Reciprocal searches initiated with the E. coli YgiF protein (residues 1-200), not only recovered its bacterial orthologs, archaeal CyaB homologs and eukaryotic proteins, but also several others such as Bacillus subtilis YjbK, Methanosarcina Ma2350, and Mesorhizobium loti Mll4592 with e-values in the range of 10-4-10-6 upon first detection. Additionally, transitive searches initiated from the region C-terminal to the uridine kinase of the plants (49D11.13 from Oryza sativa, region: 250-410) recovered archaeal CyaB homologs confirming their relationship to with the other proteins detected in these searches. Regular expression searches with the conservation pattern found in these CyaB homologs also recovered the most of the members detected in the above-mentioned profile searches, but failed to recover any new candidates.

In all these searches, the alignments more or less spanned the entire length of the CyaB protein and typically contained the same set of conserved residues. The Gibbs sampling procedure revealed the presence of seven conserved motifs, with a probability of chance occurrence less than

10⁻¹⁴, in the search space comprising of the 70 or so proteins that were identified in the above searches as having this domain. We clustered these proteins using the BLAST-CLUST program in several smaller clusters and prepared multiple alignments for the individual clusters and predicted secondary structure for these set using the PHD program. A nearly complete congruence was seen in the comparison of the secondary structures of the individual clusters. In many cases, the region of similarity to CyaB comprised the entire length of the target protein detected in these searches. However, in some cases it only comprised a part of the protein, with rest of the protein being made of other globular domains. These observations, taken together, suggested that CyaB and soluble mammalian thiamine triphosphatases define a novel superfamily of conserved domains, which may either occur by itself or in combination with other domains. We named this domain the CYTH (CyaB, thiamine triphosphatase) domain after the two experimentally characterized proteins in which it is present.

Sequence conservation, structure and biochemical activities of the CYTH domain

All complete CYTH domain sequences were aligned using the T_Coffee program, and this alignment was further refined based on the PSI-BLAST HSPs, conserved motifs detected in the Gibbs sampling procedure and predicted secondary structure for the individual groups (Fig. 1). A text copy of the alignment is being provided as an additional file (see additional file1 and additional file 2). The predicted secondary structure for this domain indicated an α + β fold, with 6 conserved β -strands and 6 conserved α -helices. Neither the predicted secondary structure, nor the pattern of the conserved residues revealed an obvious relationship with any previously recognized fold. Given that the CYTH domain is the only globular domain in the enzymes, thiamine triphosphatase and CyaB, it is predicted to be an enzymatic domain. While the reactions catalyzed by these two enzymes are distinct, their substrates, respectively thiamine triphosphate and ATP, are both organic triphosphates. The CYTH domain is also present Cterminal to the catalytic P-loop containing domain in the plant and slime mold uridine kinase homologs. In these proteins it is likely to interact with nucleoside diphosphates or triphosphates, which are substrates for these kinases. These observations suggest that the CYTH domains are likely to be domains specialized to bind nucleotides and other organic phosphates. A multiple alignment of this superfamily reveals the presence of several nearly universally conserved charged residues that are likely to form the active site of these enzymes (Fig. 1). The most prominent of these are an EXEXK motif associated with strand-1 of the domain, two basic residues in helix-2, a K at the end of strand 3, an E in strand 4, a basic residue in helix-4, a D at the end of strand 5 and two acidic residues (typically glutamates) in strand 6 (Fig. 1). The presence of around 6 conserved acidic positions in the majority of the CYTH domains suggests that it coordinates two divalent metal ions. Analogous active sites, that coordinate two metal ions, are observed in domains with similar activities, such as the classical adenylyl/guanylyl cyclases, family B DNA polymerases, pol- β fold nucleotidyl transferases, and triphosphatases or phosphoesterases of the HD and DHH superfamilies [11,15,19,20]. Consistent with these observations, both CyaB and ThTPase have been shown to require Mg²⁺ ions for their nucleotide cyclase and phosphatase activities [16,17].

The four conserved basic residues in the CYTH domain are most probably involved in the binding of acidic phosphate moieties of their substrates (Fig. 1). The conservation of these two sets of residues in the majority of CYTH domains suggests that most members of this group are likely to possess an activity dependent on two metal ions, with a preference for nucleotides or related phosphatemoiety-bearing substrates. The proposed biochemical activity, and the arrangement of predicted strands in the primary structure of the CYTH domain imply that the may adopt a barrel or sandwich-like configuration, with metal ions and the substrate bound in the central cavity. The only prominent exceptions to the basic conservation pattern of the CYTH domains are the versions found in the plant and Dictyostelium pyrimidine kinase homologs (Fig. 1, At1g26190-like). These versions lack 5 of the 6 conserved acidic residues, but retain 3 of the 4 conserved basic residues (Fig. 1). This leads to the prediction that these CYTH domains are catalytically inactive. However, as they retain the basic residues, they probably still bind the organic phosphate substrates, and function as regulatory domains that are linked to P-loop kinase domains.

Phyletic patterns, evolutionary history and potential biological functions of the CYTH domains

CYTH domains are present in most of the major lineages, for which sequence information is currently available, from the three principal superkingdomains of life (Fig. 2). We used the multiple alignment of the CYTH domain to construct phylogenetic trees using the least squares, neighbor joining and maximum likelihood methods (see Materials and Methods). The monophyletic clusters that emerged in this analysis were essentially the same as those that were derived through similarity based clustering using BLASTCLUST. The majority of archaeal and eukaryotic proteins formed a monophyletic cluster to the exclusion of most of the bacterial proteins (RELL Bootstrap support 77%) (Fig. 2). This cluster was also supported by a unique synapomorphy (a shared derived character) in the form of a conserved motif (Dh; where h is a hydrophobic residue) associated with the second strand (Fig. 1). This phylogenetic tree topology resembles that of several proteins inPredicted Sec. Str. MTH1629_Mta_15679624 MJ0240_Mja_15668415 CAC0650_Cac_1599338 CYaB_Ska_20094307 ST0100_Skt_15920492 CyaB_Ska_20094307 PH1050_FL1499570 PH1050_FL1499570 PH1050_FL1499570 PH1050_FL1499570 PH1050_FL1499570 PH1050_FL209233700 WN01296C_Mp_15790139 CyaB_Me_1020222 CyaB_Ye_16121473 PH5011.4 Ce_17531725 C014434 Em_7290766 Rsc1389_Fs01_17546108 Rh21921 Mfa_12698188 THTP_Hs_13236577 Att235260 Att_15222660 Att23527 Mfa_1589148 PH1012 Ph_14591672 PH1012 Ph_14590652 CyaB_Ye_1830144 PF0663_Ftu_18977235 PH1012 Ph_14590652 CyaB_Ye_183038948 XAC1375_Atu_15890492 SKC1279_Xca_21230736 RK2157_Xfa_1583948 XAC1375_Atu_5590492 SKC1279_Xca_21230736 RK2157_Xfa_1583948 XAC1375_Atu_5590492 SKC1279_Xca_21230736 RK2157_Xfa_1583948 XAC1310_Xax0_21242083 XC1279_Xca_21230736 RK2157_Xfa_1583948 XAC1310_Kax0_2142083 XC1279_Xca_21230736 RK2157_Kfa_15831948 XAC1310_Kax0_2142083 XC1279_Xca_2123073 RH0458_Smp_1631597 RH04158_Rm_15795049 SKC87.71c_50_Cma_21322 RH10458_Smp_1631697 RH058_Smp_1631697 Ph123_HM05_1230004 Att4644 Ana_17222186 Ch1_970_Fac_123126077 RH058_Smp_1501149 SK2867_Ac_20091185 CG1222_Cr_16127752 Ph20054_Smp_1560736 SK73381_Sc2021185 CG12231_CG1_1325002 CG12322_Cr_16127752 Ph20054_Sm1_1507948 Dc11_BPR849_19421985 MH3377_Mma_21227479 MH0351_Rm1_567344 RH058_Sm1_15675144 In0964_Lin_16603035 RH1359_At_12562342 SF1036_At_12562342 SF1036_At_12

							888888	
		· . 55	БЕБЕБ	5	нанананана		BBBBBB	
	1	MIOV	AKIHDR DEMVDRILSVG GVHVSDEEQHDLY	FNAPHRD	FARTDEALBIERS		GGRTFITTRGPKIDEES-KTRKELETEVAD\MJ0240 11ke	
	1	MIN	INVKIDDKNKVVEQLKKLGFKFIKKKFQEDIY	FNGIDRD	FRETDEALRIRDE		DGNFFVTY:GPKIDKIS-KTRE	
	1	MKOA	THIDIDVENIRTILKSVNAINIKKENOVNDI	YDFENGK	LLDNKGYARIBIV	6	KNHYFMTVEKMLSOEKF-KVMDCHEIEISD	
	1	MV	WWWIND ROVERPIRET C REVETURARDI	VOHPORD	FARTDRALBLOCS		DDEVALTY GPEV/GPE KSPV-LEV/UDD	
				A CHIE CRUD	TABIDEAU CO			
	3	HICK	INIKDISPNIBEDEKILNIKIKDENEBHQIDIT	INSPVRD	PRKIDEALMLRLV		NNEVELTTA GPALSSES-ASREETTVAIDN	
	1	MICK	VMIVVEYPSLEELRESLGKDMKLLDVEKQEDVY	FNYVYRN	FKITDEAVRININ		NKGIELTY GPKLHSSL-KARE ISITVND	
	1	GV	SEYRIECSEM EYIEAALRDAG LYLEGERLEVDIY	YRHPCRD	MLSRDEALRLEVY	1	DNGGRLTY GPRM3GGF-KARE VEIELTR	
	1	- MOV	AMEPYREGVERKVKEIARLIIEKFEHDIY	FSHPCRD	FASSDRALFIROD		VEGITLTY GPKVDVET-KSREVKLKVEN	
	1	MY OR	THOUSAN	VOUDCED	POVTOPALETETY			
				a grip on p	PORTDERMENT			
	1	MIDY	LEGIANDEIFERVRETFEFMRREIHEDIT	TOHPCRD	PSKIDEALRIRIK		HNEVPLITEGPRIDERS-KIRL IEVEIGED	
	1	MYOV	LIGYANDEIFERVREKFEFMRREIHEDIY	YQHPCRD	FSKTDEALRIRIK	4	HFEAFLTYS GPKIDERS-KTRL IEVEIDD	
	1	MIEV	VIVRADHSKIRPLLEKIGANKIGVEEQSDTY	FSAPYRD	FANTDEALRIESL		GGQAVLTY GPKLDKVS-KTRELETPVDE	
	1	MINV	VEVRADHSKIRPVLMEMGASKIGVREDSDVY	FAAPYRD	FAKTDRALSISL		GGHSVLTY GPKLDKVS-KTRV IETPVDG	
		MT T	VENDADI	POUDCED	PEATOPATHURIC		PODUTUTY CODMONCE VTDI VICEORCE	
		ni la	A INDD STANNE SUPP	PUNPCRD	PSAIDEADEVAIS		BORVIVIII OFRIGADA - KIRLAVAAQAS	
	1	MIDV	AARHIIAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	TDAPHKA	PEOLDEVARTERSA	21	DGESRVIII GPLVEAES-KIRE PETAVGD/	
	9	RFDV	PEYRLSDVDAFTCALAALNPEVMLEDNQEQDSY	FDTPEHS	LAAEGKSLVIETM	2	SGIQLWIVEGPEADRC2AVNITDA\CyaB	
	8	KY V	ELEFRVMDLTTLHEQLVAQKATAFTLNNHEKDIY	LDANGQD	LADQQISMVLEEM	2	SGIRLWIV GPGAE RCCASNIEDV/	
	25	SRNV	IKAKVONLDETVRRAVEISGKOP-TILKOHDIF	YESP	NGRLEMESV	6	HTELIWYDESDVAGPKLSNPNKFEVPSEVLD\THTP-like	
	10	ORNY	VEAR I PGGSEGFEORLILARNLS GSODAOLIEORDV	FESP	LGGRLELSYL	4	RSOLVYYDEPDVAGPKLSKPNKTVDEPEV	
	20	ADATU	TADUUDA ANI CADAAAI T DACD- PUT PODDTP	PACA	HCREWERR			
	-	ARG		PACA	I HOLDING HOLDING		NonLini Franzyadrkuszu – Frivridaru	
S	5	P10A	REPERSECTION OF THE REPORT OF	IDIPERS	LMQADHWLERRED	14	VLGHHTBIN-BUTABPTIVAQUCKVLGALGLG	
	5	LIEV	REFLPGPGTEERLQELGGTLEYRVTERDTY	YDTPELS	LMQADHWLERRED	12	VLGPHTEYE-ELTAEPTIVAQLCKVLRADGLG/	
	248	SPTY	ILKSRKEVSVDQIKAVLSDGHTETKEETYDIY	LLPPGED	PESCOSYLEMENK		DGKYSLMPE-EWVTDTPFVISPRITFEVS	
	248	NPTY	ILKSSKAVTPEQMKAALSEDFKERTEETYDIY	LLPPGED	PEACOSYLEMENR		DGKYNLMPE-EWVTDRPFIISPRITFEVS	
	242	NPMY	IL SPRTIT PTDIKVALGED HTESIRETYDI	LLPPGRD	PESCOSYLEMENR		EGKYNLMPE-EWVTDNPFIISPRITFEVS	
	201	PDUY	UCHARVENN	LYDDVVC	OADERWINTETER			
	474	MEVI	BUILDING NUTEDOU HAAPANASUMI	BODDODI'	ALL DRUGUE ARA	-		
	1	MKOL	THIIDIDVEKLRETINKLGAKNVKRENGTNDI	PDFPDRK	LLDKKGYARIETV	0	KTVYFMTTA-KMLSQDKFKVMBSNETIIED\archaeal cyaB 11)	ĸe
	1	-MOV	FKIKLEDFLHTLNTFNPEFVRYEEQEDVY	FEVPR	PKLLRINGV	2	LKKYYLT F -EILDENNEEFY	
	1	-M21	EVEFRVNFEDIKRKIEGLGAKFFGIEEQEDVY	FELPS	PKLLEVEKI	2	TGKSYITY - BILDKRNEBFY LEFEVQD	
	1	-MOI	IFRVSEDIKERLESLNAKFVGLEVOEDVY	PDVPL	PKLLEVERI	2	LGKSYITY - EIMDRNNE EFYEVEFEVSD/	
	1	MEDI	SKAFIPTKRLKRITKLAN-KKEKFIKRRIKTDIY	YSN	OKKITETEKI.	1	TLEKIVTPE-KKILDNNNTVEINKEIEFKIDS	
	-				grace and and the	-		
	2	AI	REFLVVGEGWRAAAHRVIPMTQGY	INDQKAL 1	KGTQKASVEVEIQ		EDAAYLNLESAAFGHTROFFEYAIPL\XF2357_Xfa	
	2	PTEI	REFLVTG DGWRDAAHAV IPMAQGY	INDOAAL 1	SGTOHASVEVEIO		GESAFLALK SREVGHTROFFEYPIPL	
	2	GINT	PERLUTA DOWPTATHAV	TNDOASL 1	SGAONASVEVELO		GDAAFLNT. SPELG	
	3			ULC.	TRUDCULULTY		ODDATE TO CARDON THE CONTRACT OF THE CONTRACT.	
	4	AL	CAP DYAG	VAS	DENKOVRVETV	-	URDATE TORSA DVRD IBISIPU	
	2	AKPI	REFLVATDGWRQHADKGIKERQAT	VVT	MEDRSVEVEIH	1	NKRAKITIKIGRSALVRNYTEYDLPM	
	2	GKQV	REFLVSSTAWRGLVEADIRILOFY	LAT	APGRTVRIRIS	1	GASAELTLE FGSGV RERD FEYPIPL	
	2	PI	EREFLIGN DNWROYADEP LLEROGY	LSV	EKERTVEVEIA		GORAWLTLKGYISELSRSEFEYEIPL	
15	159	LAM	RULLRWPRGKRGLRSCPRVITDRY	VN	GTRLELERA	1	FADGRCELE LTOKVPVLLSGAIOGLITNTYLS/	
(T.).	2	AT	PUPLVKDDPWPDLACGVLWPCCW	TAT	KDLTTVRVRTI		GNRAVITT	
472	2	CD		T.OT	DEPRIMENTS			
1113	-	or	RAF DISH- DOWROUNTO	131	DERVYRYRIA		SDRSSD11	
	2	YOOT	PREFLVNGDDWRQLAEGSATRQGT	IPS	QGATVRINVV		GNQGYLTIKGPTVNPSRSSPEYLIPL	
446	2	GL	REFLVVGEEWRALAGPALPLRQGY	LAG	-SPOGFTVRMRIM	1	MEQAWLTL APAEG IARH FEYRIPL	
	2	AI	HKFLVKDDRWRSLVQGYLXRQGY	IAT	PDLTTLEVETI		GDRAYLTIKGKNASIARLILSMKFLQW/	
	2	AO I	LEFIVNHSAV EALRDHLNTL GGEHHDPVOLLNIY	YETPDNW	LEGHDMGLEINGE		NGRYEMIMEVAGRVTGGLHORPSYNVALSEPTLDLAOLPTEVWP\vgiF Ec like	
	2	AOOT	THETWNHDAY DALENHLHTL CORNHAPSOLINTY	PETPDNW	LERHDMOLETROP		NGRYENTMETAGRYTGGL HORPEYNVALSERV	
	-		I PRIMUDAU DAL DAU DAU DELLA COMULA DOCULARIA	BORDDAN	I DDUDMOT DTOOP			
			DAPT WADAY DADAAADAT DO OBAAAPOQUDAT	EBIEDIN	LICENDIGUELEGE		NORTERIN LAOK TODE NORT TRANSFER LOUTOFASTAL	
	18	NOOT	LELELSQSGA GSLLRKNPFG SSPTILQQRSIY	FUTPGWD	LSKRSLSLRINQS		GNERIQIERAALGIAAGSFTROEWERPVAGDTPIVDDPQI	
	2	AREI	LAVPAGAH DALAGWLDAH AQPAGSVALANVY	YDTPEQA	LARNRAALEVERH		GNQWLQTL TAGDGGQGGHAGLSARH WEVPLDGDA LSVEAFAARDAA	
	20	DHII	LEFLIPPEAA EAILDVLDGEG AVROLDATY	YDTADHA	LRRAGFGLEVEDG		EGGRKOTLISAS AGGV PSRGWEETIAGPT PDRDALART	
	2	OKT	ILRVSRATL FALORHPLLKK - RNKSGWERRELYNOY	YDTPGRE	LARAKVALEMERD		GEOYIOTLESEGOSVAGLSERN, WDWYLEKNKLDLKKLDDKCWP	
	2	ETUT	FFUSPORS - TILPAKISETK VLOUSCOPECONT	PDTPDNM	LROHDIGLETEPP		DEVYVOT V TAGEVVAGL	
		111	TERTATIONAT	POTADER	LEONDICIET		DOPYDATE CONTRACT NOR AND TRED	
	-	-	THE ADDRESS OF THE AD	NOMA DOLL	I DETENDED TO MELLOT		Non-the sector of the sector o	
	2	UNEV	DELAY IPAPA DLLHQEDTNFR ILAHNTIFLANTY	DIADRY	THURHUNGTEANDE		MAGITTITATIVAPAN AND HIKLAINALTANATANA	
	6	LQUI	LAISPQIG IELPQYLAKFT ILEHQNLFLGNTY	YDYPDHF	LAKQKMGLRIRQE		DQELTLTLTINGKVVSGLHSRPSYNLPLIEKETPTNAQLRGLY-/	
	1	MKOL	LEYYAHSQIGVETICDLWVAMYDGYNLVDRANTLKHEFY	YDTEDHA	LRRKGIYARLD		VNRSLIMV - IGDSESGL PNRE ITFSTMDD FYHY-	
	18	NMOI	SEFLVPEEID FKALENLSKLASYTISEAOIOVIEDT	LDTKNKD	IMAAGYYLEVEKA	2	EKGKWVTIKSLGGFEGGTHRREYASFLPEGLSVLECPSLRIR\MA2350 Mac like	
	1	-MOT	SEFLVPERKD FKELENLSPLASYTISEVKAOLNEDS	LOTENKD	IMASGYYLEVEKS	2	EKGKWITTESLOGFKAGTHRREYVSFLPEGLSVLECPDLRIE	
		MILT	SEFLVLDRAD TONLETLSOLODYSLSECTUOUTEDT	TUTTEMU	LMSRGYPLELEVP	2	KKGOWLTTKSLGGPEAGU HKPB, VUSELDER GSULECPDIKIP	
	-			A L DUILY	LANDAUTOT DO	-	AUDITION DATE TO A DE TRANSPORTER AUDITION DE LA CONTRACTION DE LA CONTRACTICA CONTRACTICA DE LA CONTR	-
	8	HLOV	THE SYSESTQ IPQLEALAEVD MIDRTEIHQLSAVY	DTVDLR	LIRAKITLERATG		GNDAGWAIR-PFGIIGRREVQAPLIGEGATETLPPRELL-\SCP11.24_Scoe 11	ĸe
	- 4	HLOV	TAPSVSESTQ IPQLEALAEVD HIDRTEIHQUSAVY	FUTVDLR	LIRAKITLERETG		GNDAGWHII - PFGTIGRREVQAPLDGEGATETLPPRELL-	
	11	HL	BREFDVIESTVSPSFEGIAAVVRVEQSPTQQEDAVY	FDTPSHD	LARNQITLERETG		GADAGWHLE-LPAGPDKRTEMRAPLSASGDAVPAELL-	
£	5	KR T	REYEPPAGGLAGLPDLTGVGPVASVTAAGPEELDAVY	HDTVDLR	LAGSSATLERETG		GPDAGWHL - LPLTGDTRE VWAPLSDDVPAALR-/	
	2	-MOV	VELRLLTAAAHLRLTTLLTPYHLKTLHOPNTE	PDTPKND	LSLRRAVLEL FL.	11	PPRCIVSLAAKPTLANGI SRVEDDEEEIEYWI GKECVESPAKLSD	
	A	NI	SLLSL ARYDOLKKL PTHUTPUPATHU	LDSKDFF	LEKKKLALET		DESAENTLE - VEOR	
				E COLDER	I DEPENDENT I P		DEVICE TO BE	
	3	dra1	DET BORK DE INKEKDQFIGVTPVLOTNYY	DIFDER	DREALVANEINTF		EPHADULAT - Y PUS VGNPT I NUNLQUADAENY - LSKEELPUGLV -	
	3	NLOI	TLLTKNEYNRLLSQMKHVTPVTQTNYY	IDTKAFD	LKANKMSLRIGTF		VNSAELT	
	4	ELEI	FENLLTKBEYDRLIEDFRIKEDDFFEOTNFY	LDTADFG	LKERNSALRIEKL		ETQYQLTL -TPEARGLAGTTQILAADQATAI-TDGANIPVGPV-	
	4	EL	FENLLTK EEYDTLIENF RVKEDDFFEQTNYY	LDTTNFG	LKERHSALRIROL		ETQYQLTLE-TPEARGLMCTTQILGEDQASAI-ISGANIPVGPV-	
	4	EI	FINIVTEBEFHALCKSFSIEVFTKOVNHY	FETPNFS	LKEAGSALEINK		GETYTLTLE-OPAEVGLLETHOVVTENEAKMM-METNVIISGAV-	
	4	EIGT	RETLYSK ETFKRLISOL HIGEGDFKLORNHY	FETDDFO	LKKOSSALETERK		RAIFTFTLK-OPHPAGLLFTNOTLSKORAKLA-LESAHFPSGEV-	
		NH	PROMITASTYNKLORKYPROPUT PROVIDEN	TOTPDEE	LKEHRSALETBUK		DNOVENTLE TO BE VILLE YNY I UD I K DEMNILT I SNDNILDD I P	
	4	INN I	DENT LON DENT DECEMPTION DECEMPTION OF THE PROPERTY OF THE PRO	ADDENDER.	AVER AND ADDRESS OF		ANY INTER TRANSPORT OF THE TRANSPORT	
	- 4	E1 I	ABPLOBUKUL-PPSTSSI-BUTNY	ELLUND.	LINAHUSALIII IKK		CONTRACTOR OF A CONTRACT OF A	
	1.5			CONTRACTOR OF TAXABLE	T MOMPLES & THE THE DW		NORYTHATE PRAD	
	4	EIGI	FRNMLTKQEFKNIASALQLTEKDFTDQKNHI	PUIDSPA	LKOKHOULKINKK		MORIADIA - PLAD AODA INACISE ASD PROLADA	

(Figure I - Continued on next page)

Predicted Sec. Str.			EEEEEEE	
MTH1629 Mta 15679624	PGTAAEILESLGFRMVREVVKERRIYSVGEFTVSITVMG	LG	TYLEICRDLPDGSDY-AGALREIFELYRKL	GIEDGFERRSYLELLEL 174\ MJ0240 like
MJ0240 Mjan 15668415	KEKMRQIFKKLGFKEVPPIRKIREIYKKEDIEASI DVEG	LG	LFLELEKSISDINEK-DKVLEEMMEILKAL	NISKDNIIRKSYLELRGL 175
CAC0650 Cac 15893938	PSEGEEIFKALGLVKKQSIKKYRESYKYKNSLIEIINDK	-DFC	-PFPYICIETAFENELEDIVKLLGYSMED	TSSETIYEILNE 175
CyaB_Mka_20094307	FETTDAILRHLGFEPLEHAEVKKLRTVYTLEVNGEKVVAAL	LG	TFLECKADDESEV-DEKEKLLVSILEEL	RVEGKRVRHSYLEMLLD 178
ST0300 Sat 15920492	LDSMDLILQRLGFIKVLKIEKIRKNYKINNFTVSLERVFD	LG	DFVEIGIDITDEELINFVNNFLKEY	NIEGEKTLKSYLELLVD 174
cyaB_Sso_15897197	LNKTIELLRKIGFYPVITIKKTRINYLDKSFIVSLLVKD	LG	QFICIDAINEISDQEIQNYTSNFIRKY	KIKGKLTTKSYLELLID 173
APE1761 Ap 14601609	VEEARLLLARLGFEETLVVRKRRKYYRGEGVLVTLENVES	LG	CFV AED PGGIEMLESL	KLEDRRVDETYAEMALK 168
AF1988 Af 11499570	FEAAKQILEKLGFRAVAEVKKLRRIYGLCEAIICL	LG	KFVEIEVEADNIDAKEKVFSIAEQL	GYSRNESIRDSYLELILO 167
PAB2098_Pab_14520591	TEKYFQLLQNLGFREVLRIVNTREKYYVD-KGITITLEVEG	LG	KFIETLVKEKEEI-PRTVEKLEGILRNL	GVERFERKSYLELLLE 175
PH1819 Ph 14591570	VDKYFELLDRLGFKEVLKVVKTREKYVVE-KGVTITLEVEG	LG	KFI TLVKEKDEI-PEAVEKLEKILREL	GVEKFERRSYLELLLE 176
PF1859 Pfu 18978231	VDKYSQLLELLGFREILTITTTREKYYVE-KGVTITL	LG	KFVEIETLVKNEEEI-PQAVEKLEAILREI	GVEQFERRSYLELLVE 175
cyaB_Mac_20092837	ATTAKIFRALGFSEAGAVREKREIFRAGEVIVCLOAVEG	LG	EFLEVELDVEDEKDL-ESSREKLFEFLSQF	GVDEKDSIRTSYLEMVLG 173
MM0868_Mma_21226970	TATAKIFHSLGFLEAGAVRKKRDIFRAGEIIVCLAVEG	LG	EFLOVEDKKDL-ESSRAELFKFLSQF	GLSEKDSIRTSYLEMVLE 173
PAE0832_Pyae_18312214	EVVELLNRLGFKAVAVIKKRREYYRNGDVLLSL	LG	EFVELEKMVEEESQI-ASAIEEIRRLASTL	GLAEEVRETYLELYFN 170
VNG1296C_Hsp_15790339	DAQMAAIIEQLGFESAAVVRKLRTKHHFEGFTVLLAVED	VG	BYVEISTTVEAEHAV-SGAREDAYAVLRRL	GLDPTDQIRTSYLGLRTS 201/
cyaB_Aehy_3093292	DKAASMLRTLGYRQVLAISKRRSIYFVGPFHYTRCHLEG	IG	DFALAIMTDDEALL-PDYRQQLQDLATRL	GLSSAQLETRSYRTLCEQ 180\CyaB
cyaB_Ype_16121473	SKVQSMLATLGYHPAFTIEKQRSIYFVGKFHITVCHLTG	LG	DFASIAIMTDDATEL-DKLKAECRDFANTF	GLQVDQQEPRSYRQLLGF 179/
F35D11.4_Ce_17533725	ALKLSLQSSMGVKGEVKKTRTLVLHGQTRIHICRVDG	LG	DFMOLOVCLSPEETP-EHGEKIAHEIRELL	AVPETDLLTGAYMDMLKA 197\THTP-like
CG14434_Dm_7290769	LEKILSQSNGVLGVLAKRRHLFLCGQTRIHLEVKD	LG	YFMELEVCLTEDQTL-EEGQAIAEKLSREL	GIQEADLMIGSYFDALRK 184
RSc1389_Rsol_17546108	ALRETLTAALGATGRVINTRRLYIAGQTRIHVAVQG	LG	DFVOLOVMLRDDQSE-ADGTRIAHAMMTSL	GIAEADLLDVAYIDLLAA 169
BAB21921_Mfas_12698188	AGDVAAVLDPLGLQEVASFVTKRSAWKLVLL-GTDEEEPQLKVLLGTA	-DFG	YAVG VALVHEEAEV-PAALEKIHRLSSML	GVPAQETAPAKLIVYLQRF 201
THTP_Hs_13236577	AGDVAAVLGPLGLQEVASFVTKRSAWKLVLL-GADEEEPQLRVLLDTA	-DFG	YAVGOVDALVHEEAEV-PTALEKIHRLSSML	GVPAQETAPAKLIVYLQRF 201/
At1g26190_At_15222660	VRLLGGLMALGYTIATILKENSHVFAT-DKVFVKIWLEQ	LN	RHYMQVQGKDRQLVQSTAEQL-	GLEGSFIPRTYIEQIQL 410\At1g26190-like
At1g73980_At_15221102	VRLLGGLMALGYTIATILKEKSHIFDD-DKVIVKTWLEQ	LN	RTYVOVQGKDRTFVKNVADQL-	GLEGSYVPHTYIEQIQL 410
49D11.13_Osa_19387247	VRLLGGLMALGYTIAAILKRKSRVFSD-GKATVKIDWLEQ	LN	RNYIQVQGRDRNHVKFVAEKL-	GLDGSYIPRTYIEQIQL 404
AAG32531_Ddi_11245938	VKTLGGLLSLGYQIGAILNETVEVWYDKNGVVITKEYIKE	LE	KHFIQIKGHSRREVLDSAEKL-	KITGNHVPQTPLYLYFK 460/
CPE1352_Cpe_18310334	GAMGKNIFKSLGLELFETIKKYRESYKYKDSLIEIGINDP	-SFC	PFPYICIETSSEEKLNEILIDL	GYSMEDTTSKTIYEILKD 175
PF0863_Pfu_18977235	FEKAVEVFKRLGFKIQATIKKRWVYKL-NGVTLEVNRVEG	IG	DFVEIDVISDSPEEAKEKIWEVAKML	GLKEEDVEPRLYLELINE 165\archaeal cyaB like
PH1012_Ph_14590852	PEGAIELFKRLGFKVQGVVKKRRWIYKL-NNVTFELNRVEK	AG	DFLIPVITSNPEEGKKIIWDVARRL	GLKEEDVEPKLYIELING 165
cyaB_Pab_14521172	FNLAIEVFKRLGFDVKASIKKERLIYKL-GDVTFELNKIPG	LG	NFLAIDVISDDPEEAKRKIWEVAEKL	GLKKKDVEPRLYIELVNE 164
BB0723_Bb_15595068	INNFLTLIKELKFKKLYKKIKKSLIYQT-NNLNVEINEIKN	LG	FFLEICKIINNQNDI-DLAKKEIDNIINQF	GLKENIETRPYSELLSL 171/
YP2357 YFa 15838048	CONDUCTION	DNAG		CURVED DARYYCLALAC 155 YR2257 YFa
XAC1330 Xavo 21242083	ADAPALLALCU	DNAG	-I.TVANTOLEPPD	GTEVTD DVEVVILALAS 155
XCC1220 Xca 21230236	DDARNI L NI CV CCI TDERRUI VEV - NC	DNAG	- LINIA WIT FUAD	CARPERD DARY WI ALAS 155/
ACRE 752 Atu 15890492	KDARELMASAD	OVRG	LVVA WNDEN	CREVIC
CMc03164 Cmc 15050452	NDADDLINGAT GTVTDEDDDDTDU V/2	AL PC	- LTUA W MYDET	CPRITC DREVENOALAT 149
m114592 MIA 13473857	ADALEMI DEAT	ALAG	LVVA PPPPV	CREVIC PORTANGLAL 149
NMA2128 Nm 15295049		DMAD		CIPETTA IVINETNAVI CD 140
SCOP7 23c Scop 21226075	DARVINULACI.D	PLOG	- LVI CAN PTTOP EV ONEVEDARC	VARVIDDARRACOCLUO 202/
elr0698 Sen 16331931	VDANOTI TRI CS DDI TRI VRVCI DV. OC	DNOG	LILAN VISOAD	CKEVTD DARY NUNLTO 148 all4694 Ana like
Mmc1 p 2134 Mcep 23000473	DEALLINELCH ODITERUEVOTEN.CO	ENAC	- LITA VICAPD	COEVER DURYANACING 148
all4694 Ana 17232186	ADAORMI, DULCO	VNOG	- LUAN UTDEA	GTEVING DNEVENSYLVE 147
Pmit p 1750 Pmar 23132446	IDARALWGLAPDRLI-STRVRLSL-KGGDWVVCCFRG	ANAP	LVLASVOLVSAR	WORUTGASEWNNAALAR 149
as11045 Sap 16331697	KPIKFILTRLCSPPLIEKYRYCLDY-NGKTWRVEFLG	DNOG	LILAUVULTYTGRKISLLPWI	GREVTDDARYYNVNLAO 148/
Vai F Fc 15803596		KAGE	FARPIONULFILS COTRAVIKLANOLVSOT	GLEOGSL-SKAAPGYHLAOGN 202\vgiF Ec like
vaiF_StLT2_16766502	DGNLPAGLASSVOPLE-STDFYERKWCLDV-DGSRIETALLGDV	KAGE	FARPIOLELELE - GDTRA VLKLAKOLLSOT	GLROGSL-SKAARGYHLAOGN 202
STY3381 Sen 16761976	DGNLPDGLASSVOPLF-STDFYREKWCLDV-DGSRIETALLGDV	KAGE	FARPIONLELRGDTRAVLKLAKOLLSOT	GLROGSL-SKAARGYHLAOGN 202
ml19114 Mlo 13488075	RDLRAGAGPELAPLF-EVHVKEHRWNVD-GDATLEVTLLGKV	VAAD	REAPLO LEKKA GSPTA LFALARKUDLIT	PAHLOVL-SKAERGYRLLGSA 211
RSc2886 Rsol 17547605	RAADEVRPHADRLVPLE-RTDETERLNHTAA-DGGEIRTAL	2 POTD	AHRAIDELEVEWEPAAGNTLD ESALAREHAWT	GLAPLDA-SKALRGYRLHATA 218
CC3522 Ccr 16127752	PVGEMLARAALAPVF-TARVERTIRMVAA-GETLIEVALORGEL	SAGE	ROATVOLLELET GEPOA LEDLARTLARHA	PLRLSLI-SKAERGYGLAAGD 210
PA5209 Pae 15600402	AALKDLDKKOLKPIF-STDFVRORAEIAWGRGKARVVVEAAL	VAGD	NOEEIOLELELROGDAAALLELAAELAADL	PLMPCDI-SKAERGYRLFDPN 207
VC2440 Vch 15642436	OGKELTOLOAELMPLF-STNFTREOWLISMADGSOVEVAFIOGLV	VAGD	ROEPICEVELELKS GOTDA LFTLAROLCEHG	GMRLONL-SKAARGYRLAANY 203
YP00652 Ype 16120977	EGWAVDLLQAELOPLF-RTDFTREKWVITY-GESEIELALIOGSI	SADT	LSEPLS LELKO GTOGD LLALAAELAKMG	GLROSNL-SKAARGYHLAOGN 202
PM0251 Pmu 15602116	LTLPNYAQWELKPVF-STDFERESWLIECGNGTHIEVAFLOGKI	VAGE	KOTPIC VEFELKS GLPID LLRFVOTLTLEN	EIRLSSA-SKAKRGYLLANPT 201
HI1598 Hi 16273488	PFEQLPSSTLOPIF-STDFN#TFWLVEF-QQSKIEVAFLOGKI	IAGE	YEOPISCIEFELKS GNVOD LFDFVETLPFER	DIYFSSA-SKAKRGYLLGSKO 203/
Dch1 BPRB49 19421985	VDSLGVNSDLIVPLL-QVMTYRKSLTIDL-EGAIVENCKTFYY	3GE	1 MGDAHYELEFELKS GDEKA L-DIIREMMLEY	DLSPSEI-SKAERGFOLMTKE 197
MA0084 Mac 20088983	MIFDFTSGLDLTPLI-TLKOKRIIROLNL-EKKLVAEIYLLRVNL	KSES	RKKFYNEFEVELKS-EGTVKDLEAVRDFLLDNY	NLEESRF-SKFERAYLFRENL 224\MA2350 Mac like
MM1377 Mma 21227479	MIFEFTSGLDLFPLI-TLKQKRMIRQLKL-EEKPIADLYLLKVNL	KSES	RKKIYS POVELKS-GGTAEDLETIRDFLLSHY	NLVENPF-SKFERAYMFRENL 206
MA2350 Mac 20091185	RIFELSSGFDLIPVM-KLKQKRLVRQVKL-GETQVAELSL	KSET	KEKLYSELEIELKA - EGTLQD LQAITEYLLENY	NLGENPF-SKFERAIFFKNNF 206/
Cg12231 Cg1_21325002	HIRALIQGRELTPIA-QVDNERHMSYLADEDGAVIAEPCDHVST	3 LPGG	VRKOWR FELAD GTLAEEAISVLLOSAOSVL	TAAGAFV-SNSPSKLVSALDE 210\SCF11.24_Scoe like
Cg12182 Cg1 19553432	HIRALIQGRELTPIA-QVDNERHMSYLADEDGAVIAEFCDHVST	3 LPGG	VRKOWR FELAD GTLAEEAISVLLOSAOSVL	TAAGAFV-SNSPSKLVSALDE 206
Rv2226_Mtu_15609363	VVLAIVRDQPVQPVA-RISTHRESQILYGAGGDALAEPCNDVTA	7 AAGA	5 AEQQWR MLELVTTDGTADTKLLDRLANRL	LDAGAAP-AGHGSKLARVLGA 218
SCF11,24_Scoe_21219079	LVLSRTRGAPLRPVV-RIRSTRAVRRLHDAEGGVLAELSL	AAGG	GRGEWGELVELAEGVHAGLLDAVEKKL	IVRSDSP-SKLARALRDTGVG 203/
At2g11890_At_18397064	IGSRVKRVKEEYGFNDFLGFVCLGGFENVRNVYEWRGVKLEV	-DFG	NCYCIECETEEPERVKTMIEEFLTEE	KIEFSNSDMTKFAVFRSGK 206
ydgF_Lla_15672342	LSECELLTARDINLEEITLIGSLTTINYEQHLPIGLAALCKNDY	-LGH	TDY LEVDDSKQGKKDFFD LDKN	RVEYRFSKSKVVRFLDCLRHL 195\yjbk_Bs like
SP1096_Spn_15900964	LDELAKHGIQSKNWQVLGCLTTL YEMKTAIGLMAL	-FDI	TDYCLEVENHEQGKQDFRQFLEKN	QISYQKAPSKLVRFVKSMKNS 189
SPy1124_Spy_15675104	LDIIISKGIKPSALVTFGNLTTVKRETVIPIGKLAL	-ANT	KDYCLEVSDALQGKIDFDSFLSEY	HITFKYAKSKVARCINTLKKF 189
lin0964_Lin_16800033	RDTLKEIGINHEDLQVFGSLKTIRAEKDYKKGLLVFIKNFY	-GSI	SDFLEYEVSDYDKGKEIFDKLKEY	QITNHPAENKVARFYNHVYKN 192
lmo0965_Lmo_16803005	RDTLKELGINHEDLQVFGSLKTIRAEKDYKKGLLVFIKNFY	-GSI	SDF LEYEVSDYDKGKEVFDKLLKEY	QITNQPAENKVARFYHHVYEN 192
BA1745_Ban_21399120	MNQLCKLQIPVSALTYMGSLTTERAETLFEGGTLVFUHSFY	-YNH	DDY FEVQDEETGKAAFIHLKQH	NIPIRHTNNKVKRFFLAKQNK 190
BH2851_Bha_15615414	MDALRDLSIPISQLKHIGTLSTSRAEIS <mark>Y</mark> EQGI <mark>LCLO</mark> HSSY	-LGI	EDY FEGTS EEH ATVTFQE LKTF	SISQVPTENKIQRPFSKKEKN 192
SAV1004_Sa_15923994	QIIVEQFGVKDTTLSILGALTTYNQETK <mark>Y</mark> KGDL <mark>LVL</mark> KSEY	-FDT	ADYOLOFEVKDYNQGLQKFQSLINEL	NLEHHQPLNKVQRFFKKKETL 192
OB1219_Oih_23098674	SKQLSEMNISPAELKYVGYLTTNEMEIEFNGTQLVLYSNY	-LGT	EDFELELEAST KEH GKKI FYELLNKY	NIPIRQTPSKIERFFQQAQRS 191
yjbK_Bs_16078223	KDQLHKLQIDTDAIQYFGSLATNRAEKETEKGLIVLTHSRY	-LNK	EDY FEAAD WHE GRQAFEKLLQQF	SIPQRETKNKILRFYEEKRKS 189/
concensus/90\$	t h hD		-hF h	n h

Figure I

Multiple alignment of CYTH domains Proteins are represented by their corresponding gene names, followed by a species abbreviation, followed by the Genbank gi number. The coloring reflects the amino acid conservation at 90% consensus. The consensus abbreviations and coloring scheme are as follows: h: hydrophobic residues (L,I,Y,F,M,W,A,C,V), l: aliphatic (L,I,A,V) and a: aromatic (F,Y,W,H) residues shaded yellow, c: charged (K,E,R,D,H) residues, and p: polar (S,T,E,C,D,R,K,H,N,Q) residues colored purple; +: basic (K,R,H) residues shaded blue, -: acidic (D,E) residues shaded red, s: small (S,A,C,G,D,N,P,V,T) and u:tiny (G,A,S) residues, colored green; b: big (L,I,F,M,W,Y,E,R,K,Q) residues shaded gray. Secondary structure assignments: H: Helix, E: Extended (Strand).

volved in core cellular functions such as the DNA recombinase RecA, aminoacyl tRNA synthetases, RNA polymerase and other RNA metabolism proteins [21–23]. This suggests that a CYTH domain was present in the last universal common ancestor of all extant life forms and the extant forms are in part vertically inherited from this ancestral form.

However, there are certain anomalies to this pattern. The CYTH domains are entirely absent from the small genomes of pathogenic bacteria such as *Rickettsia* and *Chlamydia*

as well as some of the large genomes such as *Deinococcus radiodurans*. At least, a single copy of the CYTH domain is seen in most archaeal and eukaryotic genomes sampled to date, with the exception of *Thermoplasma* and the yeasts, where it is absent. This implies that the CYTH domain has been lost independently on a number of occasions in evolution. CyaB homologs from *Aeromonas*, *Clostridium*, *Borrelia*, and *Ralstonia* lie firmly (RELL Bootstrap >= 70%) within the archaeal and eukaryotic clusters, rather than with their bacterial counterparts (Fig. 2). These bacterial forms also share the unique sequence signature in the sec-



Figure 2

Maximum-likelihood phylogenetic tree of CYTH domains. RELL bootstrap values are shown below the branches and branches with bootstrap values <50% are collapsed. The protein nomenclature follows the convention given in the legend to Fig. 1.



Figure 3

Domain architecture, predicted operons and contextual information map for CYTH domains Proteins are represented by their gene names, species abbreviations and gi as in Fig. 1. Operons are shown with genes represented as boxarrows. The contextual map shows different types of associations between the domains. Unidirectional black arrows represent domains co-occuring in the same protein. Bidirectional red arrows represent domains co-occuring in operons, the dotted red arrow represents adjacent gene transcribed in opposite directions, and the green arrow represents an experimentally derived functional association. Domain abbreviations: GuK, Guanylate kinase, NuK, Nucleotide kinase, TK, Thymidylate kinase.

ond strand with this group suggesting that they have been derived through horizontal transfer from different archaeal and eukaryotic sources (Fig. 1). In particular the CyaB homolog from Ralstonia groups very strongly with the animal versions and appears to be a recent horizontal transfer in this bacterium from the latter clade. The possibility of lateral transfer of Aeromonas CyaB from an archaeal source has been previously suggested, and is consistent with the enzyme being optimally functional under high temperature [16]. There are 3 distinct CYTH domains in the euryarchaeon Methanosarcina acetivorans, in addition to the version which groups with the CYTH domains that are found, in single or duplicate copies, in other archaea and eukaryotes. These former versions, strongly group with CYTH domains from actinobacteria (Rell Boostrap 77%) to the exclusion of other lineage (Fig. 2). Furthermore, they share a fusion to a novel conserved domain with characteristic histidines (see below) with the actinobacterial versions. Thus they appear to have been transferred laterally from the actinobacteria into the *Methanosarcina* lineage followed by a small lineage specific expansion in the latter. Bacteriophages, like RB49, that contain a solo CYTH domain, which is closer to the version seen in its proteobacterial hosts, could have served as conduits for the lateral distribution of this domain.

The phyletic pattern of the CYTH domain is not very typical of signaling enzymes like nucleotide cyclases. Classic adenylyl and guanylyl cyclases show a far more sporadic distribution, and are often present in multiple copies fused to a variety of signaling domains such as the cyclic nucleotide binding domains [24]. Cyclic nucleotide generating activity is not known to exist in a subset of the archaea [25], though most of them contain a well-conserved copy of the CYTH domain. Experimental analysis on the Methanococcus CyaB homolog revealed no adenylyl cyclase activity comparable to that seen in Aeromonas CyaB [16]. Likewise, there is no evidence for a widespread presence of the ThTPase activity in the organisms that contain CYTH domain proteins [17]. Hence, the principal biological function of the CYTH domains may be different from those of the experimentally characterized members of this family. General conservation of this domain across a range of organisms, and its presence in the LUCA suggests the possibility of an important general role in cellular process of free-living organisms. In order to decipher the potential biological roles of this domain we used different forms of contextual information regarding these domains, in the form of domain architectures, gene neighborhoods (predicted operons) and experimental evidence for interactions between proteins. Both domain architecture and operon analysis have been used extensively to make functional predictions of poorly characterized domains or genes [26–29]. Moreover, the presence of evolutionarily conserved operons often correlates with the involvement of the component genes in a sequential pathway or physically interacting complex [30,31]. We summarize the different forms of contextual information that we extracted from the CYTH domains in the form a network diagram (Fig. 3).

The CYTH domain shows a small array set of fusions to other conserved domains (Fig. 3). One of the most prevalent fusions is to an uncharacterized domain, with a characteristic pattern of conserved histidines and other charged residues. This domain is predicted to adopt an α helical fold and is according referred to as CHAD (CHAD: conserved <u>h</u>istidine $\underline{\alpha}$ -helical <u>d</u>omain; Fig. 4). The sequence conservation pattern (Fig. 4) suggests that this domain is likely to contain two repeat units, with at least 4 helices each, at its core. While no clear functional prediction can be made for the CHAD, the conserved charged residues could form a strongly polar surface that could participate, either in metal chelation, or act as phosphoacceptors. Another notable fusion is with a specific version of the HD hydrolase domain [19], which is also found fused to the C-terminus of the HSP70-fold domain in the exopolyphosphatase (PPX) (Fig. 3). HD domains typically possess phosphoesterase activity, and are fused to catalytic domains that possess nucleotide kinase, nucleotidyltransferase, nucleotide cyclase or diguanylate cyclase activity [19,29]. This fits well with the observed cyclase activity seen in CyaB, but is also consistent with phosphotransferase or nucleotidyl transferase for the CYTH domain. Finally, catalytically inactive versions of the CYTH domain are found fused to the nucleotide kinase domain

in uridine kinase homologs and may serve as an allosteric nucleotide-binding site in these enzymes (Fig. 3).

In terms of conserved gene neighborhoods, CYTH-domain-encoding genes, like mll4592 from α-proteobacteria, are frequently found in the neighborhood of genes encoding solo CHADs (Fig. 3). Additionally, in Methanosarcina the CYTH-encoding genes are found in predicted operons or in the neighborhood of genes encoding exopolyphosphatase (PPX) and polyphosphate kinase (PPK). Furthermore, in Sulfolobus, a gene for a CHAD protein is in a predicted operon along with genes for thymidylate kinase and PPX. CHAD- and CTYH-encoding genes are also found in the neighborhood of PPK and PPX in Chlorobium tepidum and Geobacter metallireducens, respectively (Fig. 3). Genes for CYTH domains also co-occur in predicted operons with genes for another polyphosphate utilizing enzyme, the polyphosphate-dependent NAD kinase (PPNK), in certain cyanobacteria (eg. Prochlorococcus marinus) and Gram positive bacteria like Oceanobacillus iheyensis. Other potential connections are furnished by the co-occurrence of genes for CYTH-domain proteins with genes involved in with nucleoside polyphosphate metabolism. These include co-occurrence with the gene for adenosine tetraphosphatase (APAH; eg. in Magnetococcus sp.) and with genes encoding the YjbM-domain in Gram-positive bacteria. The YjbM domain, most often, occurs fused to pentaphosphate guanosine-3'-pyrophosphohydrolase (SpoT) and GTP pyrophosphokinase (RelA), suggesting a role for it in the metabolism of the stringent-response nucleoside polyphosphate.

These contextual connections are consistent with the participation of CYTH domains in organo-phosphate biochemistry, and circumstantially associate it with the metabolic network related to polyphosphates and nucleoside polyphospates (Fig. 3) [32]. Specifically, PPK and PPX have been shown to, respectively, lengthen or shorten polyphosphate polymers [32]. These two enzymes also appear to interact with the nucleotide metabolism of the cell. In particular PPK and PPNK can utilize Poly(P) to synthesize nucleoside polyphosphates, while PPK along with adenylate kinase can carry out polyphosphate-dependent phosphorylation AMP [33-35]. Hence, it is likely that the CYTH domain proteins participate directly in this biochemical network along with these proteins. One possibility is that the CYTH domains utilize polyphosphates to synthesize different organo-phosphate derivatives including nucleotides. Alternatively, they could also function as phosphoesterases that hydrolyze particular nucleoside polyphosphates. These leads could aid further experimental investigations on the CYTH domain that might help in uncovering ancient, as-yet-unexplored links between polyphosphate and nucleotide metabolism.

Predicted Sec. Str.	
MM1377_Mma_21227479	541 EDPMAFV <mark>A</mark> HRIFAYQ <mark>F</mark> SQMLAHEKGTRKGEDIEDLH <mark>DMR</mark> VAVR <mark>RMR</mark> AA <mark>A</mark> IVFDEYIESEKVEPHIKGLKRTLGTLGGVRD-LDVFREKAED-YLKTLPPGREHDLDPLFLILSEEREKAREEMLDYLDSEKYSRFKKD <mark>F</mark> SE
MA0084 Mac 20088983	571 EDSMAFVAHRIFAYOFSOMLAHEKGTRKGEDIEELHDMRVAVRRTRAAATVFNKYLESEKLEPHLKGLKRTLGALGGVRD-LDVFREKAET-YLETLPTGHEHDLDLLFVTLTEEREKARENMLEYLDSEKYSRIKKDFSE
MA2350 Mac 20091185	537 ENSMAMUAOKVPSOOFARMLAHEKGTRKGEDIEELHOMRVSIRMPAAAKVFEAYLOSKKLGPHLKGLKSTLGALGDVRD-LDVFREKAEG-YLKKLPPEKEHDLDPLFAVLAEEREKSRKNMLIYLESEKYSSFKKEFSE
Gmet0782 Gme 23053876	
alloE67 Pro 17228062	
all0567_Ana_17228063	9 VKTEGNIAIEALQKHEKKTEKWERAVKEDEDPEALH, MKVEMKKERTALSKEDIALMESKPA-SDKNIGKIAKEEGNEKD-LDVLKGTLETLIQPHLPEKEQKTEQKAPDALTKQKVSVLDKTQETLKDESIKSLKHSLEE
s1r1444_Ssp_16330148	4 THSFGDR <mark>A</mark> VFAFGKHTGK <mark>I</mark> FKYAPRVLKDQDPEDLH, LRVGVRRLRTAVIGFMPAVQLPQGI-TERK <mark>I</mark> GKIGQQL©VQRD-NDVLQEILLHKYLPHLPTSEQGCLEKVLKRLQKERKQAFKATYKLLTHDKFTQFKQSWEI
CT0884_Chte_21673712	218 HASIHEN <mark>V</mark> RRLLQFTTSI <mark>M</mark> EANEEGIRKDIDSEFLHD <mark>F</mark> RVAIRRS <mark>R</mark> SILRLLNGVFDPEKTAWMLAG <mark>L</mark> RELGKRTNDLRD-SDVYLLRREE-YTSLLPPSLRPALDPFFSDLEADKRLHHRQFCRYLTGREYSGFMTS <mark>L</mark> KE
SCF11.24 Scoe 21219079	216 AGSSGSSYVLSYLREQVGVLVGLDPAVR-RELPDAVHRMRVTCRRLRSCLRSYRSVLDRRVTDPVAAELRWLAGELGAARD-QEVLRERIGA-ALDGLPDELV-LGPVAARLRVWDVSRDEDVRSRTRMALGSPRHLRLLDALD
Rv2226 Mtu 15609363	228 POPPADPVHRAVSEOVEOLLUWDRAVR-ADAYDAVHOMRVTTRKIRSLIDSQESFGLKESAWVIDELRELADVLGVARD-AEVLGDRYQR-ELDALAPELV-RGRVRERLVDGARRRYQTGLRRSLIALRSQRYFRLLDALDA
Ca12231 Cal 21325002	226 KNDPARGVLAAIAANASKTAEYDPRVR-ADEYDSVH MRVATRELSHLOTEGULGGEDVLNLEKELKVLANILGRARD-AEVVEERLSN-LINTEVGDSI-EEETKKELLEDLGAEVREHERVVRALDNDRYTDLIOALEN
Cal2182 Cal 19553422	
D32972 Dec 15509069	
PA28/2 Pae 13398088	
m114591_M10_134/3856	8 REPETDEVKILAEEIGKALQHEDAAR-SRPEQGEHCCRKREKSAKALDREVHSGDETFCTT-ENQCTRNVAALLAGPRE-ATALIETVDRLAAAFPRES-ADGGEDAVRDREVARQHDEHEGAGEEAAIGAATAACADG
SMc03155_Sme_15966639	8 KRPFTED <mark>F</mark> RAVGGEQ <mark>IERA</mark> IAMLEVQP-EGVHEAIHD <mark>A</mark> RKGFK <mark>RL</mark> RSLYRL <mark>V</mark> AADAPLFQRQ-ENAR <mark>F</mark> RAMARSLSTFRD-AAALVENAGYLRRHAASEE-QQLALDKICSILASRRDRIAEDEGDLDRKIEETIVNCRK
Atu4492_Atu_17938181	8 KKPFGDE <mark>I</mark> RRAGLEL <mark>I</mark> DD <mark>A</mark> VTILRDQP-SGSHEAV <mark>HDAR</mark> KRFK <mark>RL</mark> RALYRLIRKVAPDFARE-ENTR <mark>F</mark> RDIARSLAFARD-ATALVETAEYLEAFAASDT-QGEALRSVTVMLRERRDDALDHEAGLDEAIAAAIAG <mark>C</mark> EK
RSc3043 Rsol 17547762	9 ILPVFAG <mark>I</mark> VLEHVNVLREMLTKLAAPAPTDEDLH <mark>OFR</mark> VSLRRLRSAMVTFAPVLPDVFMEVWKPRLRDLATSTGPVRB-WDVLLLDWLPAARAALDPADMRALNWLDKTAARARAARARARAKVLHAELVSPGMASMLDALKD
DR2614 Dr 15807595	
SS01190 SE0 15898044	
CC3522 Cor 16127752	
m119114_M10_13488075	225 EMSARTAFVRIATACLEVZERLINERVLSWSRDAEALH, ARVSLRKLKSLCSICKSLFDDSRFDHMREEELTWLASEFGDARG-IDVMIDRASSEALSSRLQDAREDAIAAVEASLSSARARALMIDAAE
VC2440_Vch_15642436	219 NDTAESCFIRALEHALAHWHYHEQIYTERENVAALHIIRHAVSYLSQLLSVYGGIIPRRASAILRQELKWLEQELQWLKS-FEYLESLQED-KGYALRKLDARKFLVTALKTLQESLPQ-REDTLRLLSSARYTGLLLDLSR
YP00652_Ype_16120977	218 KSTVEQGMAGGLESILEHWQYHEELWL-RGEPAAKTMIIDALAMVRQSLAIFGGLVPRKASTELRALLIALEPQLEPKSVDAALLCYSVDY
STY3381 Sen 16761976	218 KATVEQGLEASLDLALSQWQYHEELWL-RGDESAKEHVLDAMGLVRHALMLFGGIVPRKASAHLRDLLTQAEATMTSAVS-AVTAVYSTQT
vgiF StLT2 16766502	218 KATVEOGLEASLDLALSOWOYHEELWL-RGDKSAKEHVLDAMGLVRHALMLFGGIVPRKASAHLRDLLTOAEATMISAVS-AVTAVYSTOT
vgiF Ec 16130950	
consensus/90%	
consensus/ 50%	
Predicted Sec. Str.	NH
Predicted Sec. Str. MM1377 Mma 21227479	ин
Predicted Sec. Str. MM1377 Mma_21227479 MA0084 Mac_20088983	нн
Predicted Sec. Str. MM1377_Mma_21227479 MA0084_Mac_20088983 MA2350 Mac 20091185	HR NHHH-NHHHHHHHHHH FLDFPESWVL 13 VXDVLPSILYARLANISAY 9 SVERLHRL LAA GLAVILEFENVLGKE VESLIKOFKALQOHL GERDAVAAEMLGTYIKTGANK 19 IEAVLAYKEAELLTLLDFPESWEKVRT 844 FLEFPETWAF 13 VXDVLPSILYARLADISAY 9 SVERLHRL LAA GLAVILEFESVLGKE VESLIKOFKALQOHL GLAFDAVATDMLGFYLKTGTNN 18 VEAVLAYKEELTLLILDFFPESWEKVRT 834 DLADYERAL 13 IKDVLPSILYARLADISAY 9 SVERLHRL LAA GLAVILEFESVLGKE VESLIKOFKALQOHL GLAFDAVATDMLGFYLKTGTNN 18 VEAVLAYKEELTLLILDFFPESWEKVR 833
Predicted Sec. Str. MM1377 Mma 21227479 MA0084_Mac_20088983 MA2350_Mac_20091185 Gmet0782 Gme 23053876	HH
Predicted Sec. Str. MM1377_Mma_21227479 MA0084_Mac_20088983 MA2350_Mac_20091185 Gmet0782_Gme_23053876 pl0657_png_7232063	HH
Predicted Sec. Str. MM1377 Mma_21227479 MA0084 Mac_20088983 MA2350 Mac_20091185 Gmet0782 Gme_23053876 all0567 Ana_17228063 all0567 Jana 17228063	HH
Predicted Sec. Str. MM1377 Mma 21227479 MA0084 Mac 20088983 MA2350 Mac 20091185 Gmet0782 Gme 23053876 al10567 Jana 17228063 slr1444_Ssp_16330148	HH
Predicted Sec. Str. MM1377 Mma_21227479 MA0084 Mmc_20088983 MA2350 Mac_20091185 Gmet0782 Gme_23053876 al10567_Ana_17220063 slr1444_ssp_16330148 GT0884_GT0842_GT3712	HH HHHHHHHHHHHH HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Predicted Sec. Str. M41377 Mma 21227479 MA0084 Mac_2008983 Ma2350 Mac_20091185 Gmet0782 Gme_23053876 all0567 Jma I7228063 slr1444_Ssp_16330148 C70884 Chte_21673712 Scrl1.24_Ecce_21219079	HH HHH HH H
Predicted Sec. Str. NM1377 <u>Mma_21227479</u> NA0084 <u>Mac_20088983</u> Ma2350 <u>Mac_2008185</u> Gmet0782 <u>Gme_2008185</u> Global <u>Alama</u> <u>17228063</u> slr1444 <u>Ssp_16330148</u> C70884 <u>Chte_21673712</u> Scr11.24 <u>Scoe_21219079</u> Sr211.24 <u>L5609363</u>	HH HHH HH
Predicted Sec. Str. MM1377_Mma_21227479 MA0084_Mac_20088983 MA2350_Mac_20091185 Gmet0782_Gme_23053876 all0567_Ana_17228063 slr1444_Ssp_16330148 cr0844_chte_21673712 Scr11.74_scoe_21219079 Rv2226_Mtu_15609363 cal231_col_21235002	HH HHH HH <th< td=""></th<>
Predicted Sec. Str. MM1377_Mma_21227479 MA0364_Mac_2008993 MA2350_Mac_20091185 Gmet0782_Gme_23053876 all0567_Ama_17228063 all1644_58p_16330142 cc011.Chcc_21213009 cc011.Chcc_21213009 cc01231_Cc1_21325002 cc12231_Cc1_19553432	HH HHH HH H Interpolation H
Predicted Sec. Str. MM1377 Mma_21227479 MA0034 Mac_20088983 MA2350 Mac_2008185 Gmet0782 come_23053876 all0567 Jna_17228063 s1r1444 Sep_16330148 cr0848 Chte_21673712 Scr11.74 Scoe_21219079 Rv2226 Mtu_15609363 cg12231_cg1_21325002 cg122182_cg1_19553432 D32872 Pac_555432	HH HHH HH In <
Predicted Sec. Str. MM1377_Mma_21227479 MA0084_Mac_20089983 MA2350_Mac_20091185 Gmet0782_Gme_23053876 all0667_Ana_T7228063 s1r1444_Ssp_16330148 Cr0884_Chte_21673712 SCP11.24_Scce_21219079 Rv2226_Mtu_1550363 Cg12213_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_2553432 PA1872_Pac_15559066_f	HH HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Predicted Sec. Str. MM1377 <u>Mma_21227479</u> MA0364 <u>Mac_20088983</u> MA2350 <u>Mac_2008185</u> Gmet0782 <u>cme_23053876</u> all0567 <u>Ama_17228063</u> slr1444 <u>Ssp_16330148</u> <u>CT0844 Chte_21673712</u> Scrl1.74 <u>Scoe_21219079</u> Nv2226 <u>Mtu_15509463</u> Cg12231 <u>Cg1_2135004</u> Cg12231 <u>Cg1_15531432</u> PA2872 <u>Pae_15598066</u> ml1459 <u>J</u> ML0_13473855.	HH HHH HH I I
Predicted Sec. Str. MM1377_Mma_21227479 MA0084_Mac_2008993 MA2350_Mac_20091185 Gmet0782_Gme_23053876 all0667_Ana_17228063 slr1444_Ssp_16330148 Cr0884_Chte_21673712 Scr11.74_Scco_21219079 Rv2226_Mtu_15609363 Cg12231_Cg1_21352002 Cg12231_Cg1_21553432 PA2872_Pac_155594068 ml14591_ML0_1373856 Shc03155_Smc_155966639	HH HHH HH
Predicted Sec. Str. MM1377 <u>Mma_21227479</u> MA0364 <u>Mac_20088983</u> MA2350 <u>Mac_2008185</u> Gmet0782 <u>cme_23053876</u> all0567 <u>Ama_17228063</u> slr1444 <u>Ssp_16330148</u> CT0844 <u>Chte_21673712</u> Scrl1.74 <u>Ssco</u> 21219079 Rv2226 <u>Mtu 15509463</u> Cg12231 <u>Cg1_21353432</u> PA2872 <u>Pae_15598066</u> m114591 <u>Mto_13473856</u> SMc03155 <u>Sme_15966639</u> Aut4492 <u>Atu 1</u> 7338181	HH HHH HH
Predicted Sec. Str. MM1377_Mma_21227479 MA084_Mac_20088983 MA2350_Mac_2008983 sln444_Seme_32053876 all0657_Ama_17228063 sln444_Sem_16303148 CT0884_Chte_21673712 SCF11.74_Scoe_21219079 Rv2226_Mtu_156039363 Cg12231_Cg1_21325002 Cg122182_Cg1_21525002 Cg122182_Cg1_21525002 Cg122182_Cg1_21525002 Smc03155_Smc_15596866 ml14597_M10_13473856 Smc03155_Smc_15766639 Atu4492_Atu_17938181	HH HHH HH HH<
Predicted Sec. Str. MM1377_Mma_21227479 MA0364_Mac_2008993 MA2350_Mac_20091185 Gmet0782_Gme_23053876 all0567_Ama_17228063 slr1444_Spm_16330148 C70884_Chte_21617312 Sfm_2128063 cfg12231_Cg1_21125002 cfg12231_Cg1_21550459 PA2872_Paa_15598068 m11459_H10_13473856 SMc03155_Sme_15966639 SMc03155_Sme_15966639 SMc03155_Sme_15966539 SMc03155_Sme_15966539 SMc03155_Sme_15966539 SMc03155_Sme_15966539 SMc03155_Sme_15966539 SMc03155_Sme_15966539 SMc03155_Sme_15966539 SMc03155_Sme_15966539 SMc03155_Sme_15966539 SMc03155_Sme_15966539 SMc03155_Sme_1596535 SMc03155_Sme_1596755 SMc03155_Sme_1596755 SMc03155_Sme_1596755 SMc03155_Sme_1596755 SMc03155_Sme_1596755 SMc03155_Sme_1596755 SMc03155_Sme_1596755 SMc03155_Sme_159555 SMc03155_Sme_159555 SMc03155_Sme_159555	HH NHHH-HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Predicted Sec. Str. MM1377_Mma_21227479 MA0384_Mac_20088983 MA2350_Mac_2008983 allower_2008185 Gmet0782_dme_23053876 all0567_Ama_17228063 all1444_Sep_16330148 CT0884_Chte_21673712 SCF11.74_Sco_21219079 Rv2226_Mtw_15607363 Cg12231_Cg1_21325002 Cg122142_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Rv2225_Bm_1558056639 Atu4492_Atw_17938181 Rvc3043_Rsc0_115547762 DR2614_Dr_15807595 SS01190_Sso_15889044	HH HHH HH
Predicted Sec. Str. MM1377_Mma_21227479 MA0084_Mac_20089933 MA2350_Mac_2009185 Gmet0782_0me_23053876 all0667_Ana_I7228063 slr1444_Ssp_16330148 Cr0884_Chte_21673712 SCP11.24_Scoc_21219079 Rv2226_Mtu_1560363 Cg12231_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Sg1105_Scot_35986619 Mt4922_Atu_179381811 Rsc3043_Rsc1_17547762 Ss01190_Sso_15898044 Cs1322_Ccr_16127752	HH
Predicted Sec. Str. MM1377_Mma_21227479 MA0364_Mac_20088983 MA2350_Mac_20088983 BA2350_Mac_20091185 Gmet0782_Gme_23053876 all0567_Ama_17228063 slr1444_Spi_16330148 CT0844_Chte_21673712 SCF11_74_Sco_21219079 Rv2226_Mtu_15603963 Cg12231_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Rv2126_Mtu_15603966 ml14591_M10_1347856 Shc03155_Sso_155896049 Atu4492_Atu_17938181 Rsc3043_Rsc_115547762 DR2614_Dr_15807595	HH
Predicted Sec. Str. MM1377_Mma_21227479 MA0084_Mac_20089983 MA2350_Mac_20091185 Gmet0782_Gme_23053876 all0667_Ana_17228063 slr1444_Ssp_16330148 CT0884_Chte_21673712 SCP11.24_Scoc_21219079 Rv2226_Mtu_1560363 Gg12231_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Gg122182_Cg1_21325002 Gg122182_Cg1_21325002 M14591_Mto_15138056 Atu4492_Atu_17938181 Rc3043_Rss1_17547762 DR2644_Dr_15807595 SS01190_Sso_15389044 CC3522_Ccr_16127752 m13114_10_13488055	HH HHH HH HH<
Predicted Sec. Str. MM1377 <u>Mma 21227479</u> MA0364 <u>Mac 20088983</u> MA2350 <u>Mac 20088983</u> Ma2350 <u>Mac 20091185</u> Gmet0782 <u>cme 23053876</u> all0567 <u>Jma 17228063</u> slr1444 <u>Spc 21637312</u> Scr11.24 <u>Spc 2123079</u> Rv2226 <u>Mtu 1560363</u> cg12231 <u>cg 21325002</u> cg122182 <u>cg1 21325002</u> sl2182 <u>cg1 21325002</u> sl2182 <u>cg1 21325002</u> sl2182 <u>cg1 21325002</u> sl2182 <u>cg1 21325002</u> cg12182 <u>cg1 21325002</u> sl2182 <u>cg1 21355002</u> sl2182 <u>cg1 21355002</u> sl218	HH
Predicted Sec. Str. MM1377_Mma_21227479 MA0084_Mac_2008993 MA2350_Mac_20091185 Gmet0782_Gme_23053876 all0667_Ama_17228063 slr1444_Spc_16330148 Cr0884_Chte_21673712 Scr11.74_Scc_21219079 Rv2226_Mtu_1560363 Gj2231_Gj_21232002 Cg122182_Gj_21325002 Cg1221752 m1449_Atu_1560359 Sc0148_Sc014885 Sc0148_Sc01485 Sc0148_Sc014885 Sc0148_Sc014885 Sc0148_Sc014885 Sc0148_Sc014885 Sc0148_Sc014885 Sc0148_Sc01485 Sc0148_Sc01485 Sc0148_Sc01485 Sc0148_Sc01485 Sc0148_Sc01485 Sc0148_Sc01485 Sc0148_Sc01485 Sc	HH
Predicted Sec. Str. MM1377_Mma_21227479 MA0364_Mac_2008993 MA2350_Mac_20091185 Gmet0782_Gme_23053876 all0667_Ama_17228063 slr1444_Ssp_16330148 CT0884_Chte_21617112 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc550245 Str1124_Bc552 Str1124_Bc562 Str1124_	HH
Predicted Sec. Str. MM1377_Mma_21227479 MA084_Mac_20088983 MA2350_Mac_2008983 sl1244_5pm_203876 sl12464_5pm_103876 Strll.24k_5pm_1038048 CT0884_Chte_21673712 Strll.24k_5co_21219079 Rv2226_Mtu_15609363 Ggl2231_Ggl_21325002 Ggl22182_Ggl_21325002 Ggl221640_Ggl_21325002 Ggl221640_Ggl_21325002 Ggl221640_Ggl_21325002 Ggl221640_Ggl_21325002 Ggl221640_Ggl_21325002 Ggl221640_Ggl_21325002 Ggl221640_Ggl_21325002 Ggl22166502 Ggl2216	HH
Predicted Sec. Str. MM1377 Mma_21227479 MA0084 Mac_20089933 MA2350 Mac_2009185 Gmet0782 Cme_23053876 all0667 Ama_T228063 s1r1444 Sep_16330148 CT0884 Chte_21673712 SCF11.24 Scco_21219079 Rv2226 Mtu_1560363 Cg12211 Cg1_21325002 Cg122182 Cg1_21325002 Cg122182 Cg1_21325002 Cg122182 Cg1_21325002 Cg122182 Cg1_21325002 Gg12182 Cg1_21325002 Sg12182 Cg1_21325002 Sg12182 Cg1_21325002 Sg12182 Cg1_21325002 Sg12182 Cg1_21325002 Sg12182 Cg1_21325002 Sg1190 Sg1255 Sg01190 Sg155 Sg01190 Sg155 Sg1100 Sg155 Sg1555 Sg1100 Sg155 Sg1555 Sg1555 Sg1100 Sg1555 Sg1100 Sg1555 Sg1100 Sg1555 Sg1100 Sg1555 Sg1100 Sg1555 Sg1100 Sg1555 Sg1100 Sg1555 Sg1100 Sg1555 Sg15555 Sg1100 Sg15555 Sg1100 Sg15555 Sg1100 Sg15555 Sg1100 Sg15555 S	HH
Predicted Sec. Str. MM1377_Mma_21227479 MA0084_Mac_20088983 MA2350_Mac_2008983 Sinter_20088983 Sinter_20091185 Gmet0782_Gme_23053876 all0567_Ana_17228063 Sint444_Sep_16330148 CT0884_Chte_21673712 SCF11.74_Scc_21219079 Rv22226_Mtu_1560363 Gj12231_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Cg122182_Cg1_21325002 Sinter_2132502 Si	HH

Figure 4

Multiple alignment of the CHAD domain The coloring scheme, secondary structure abbreviations and species abbreviations are as in Fig. 1. The coloring reflects the consensus at 90% conservation.

Finally, at least in some lineages, the CYTH domain proteins may have been secondarily recruited for other functions. The CyaB protein may represent one such case where after the original transfer from an archaeon into the proteobacterial lineage it may have acquire the novel function as an adenylyl cyclase. However, it is entirely possible that even in this case the adenylyl cyclase activity is secondary to some other uncharacterized metabolic activity. The vertebrate soluble thiamine triphosphatase has undergone accelerated divergence, as it is present on a long branch in the phylogenetic tree (Fig. 2). ThTPase is also unusual in lacking the 3rd conserved acidic domain of the CYTH domain. Hence, it may represent a case of relatively recent acquisition of a new catalytic activity.

Conclusions

We show that *Aeromonas* adenylyl cyclase CyaB and thiamine triphosphatase define a novel superfamily of catalytic domains that act on nucleotides and organo-phosphate substrates. These domains are widely distributed in all the 3 superkingdoms of life and can be traced back to the last ancestor of all life forms. We identify 6 conserved acidic residues, that are likely to form the active site of these en-

zymes, and 4 conserved basic residues, that may participate in interactions with phosphate-moiety-containing substrates. We postulate that these enzymes are likely to chelate 2 divalent cations and are likely follow a bimetal reaction mechanism similar to what has been proposed for nucleotide cyclases, nucleic acid polymerases, or certain phosphoesterases such as those of the HD and DHH superfamilies. A version of the CYTH domain, which is fused to the catalytic domain of nucleotide kinases, lacks the predicted catalytic residues, and probably function as an allosteric regulatory domain. Additionally, we detected a novel domain, termed CHAD, which occurs fused to the CYTH domain or is encoded by genes occurring in the same operon as those encoding CYTH domains. CHAD contains conserved histidines that are predicted to either chelate metals or serve as phosphoacceptors. Based on phyletic distribution and contextual information, we conclude that these enzymes may play a critical role in the interface between nucleotide and polyphosphate metabolism.

Methods

The non-redundant (NR) database of protein sequences (National Center for Biotechnology Information, NIH, Bethesda) was searched using the BLASTP and PSI-BLAST programs [18]. Profile searches using the PSI-BLAST program were conducted either with a single sequence or an alignment used as the query, with a profile inclusion expectation (E) value threshold of 0.01 and were iterated until convergence. Additionally, hidden Markov model based searches using a multiple alignment of known members were run using the HMMER2 package [36]. The Gibbs sampling procedure, as implemented in the MA-CAW program was used to detect and evaluate statistically significant conserved motifs [37]. Multiple alignments were constructed using the T_Coffee program [38], followed by manual correction based on the PSI-BLAST results. Protein secondary structure was predicted using a multiple alignment as the input for the JPRED and PHD programs [39,40]. Preliminary clustering of proteins was done using the BLASTCLUST program with empirically determined length and score threshold cut off values (For documentation see [ftp://ftp.ncbi.nih.gov/blast/documents/README.bcl]). Phylogenetic analysis was performed using the neighbor joining or least square method followed by local rearrangements using the maximum likelihood algorithm to predict the most likely tree. The robustness of tree topology was assessed with 10,000 Resampling of Estimated Log Likelihoods (RELL) bootstrap replicates. The MOLPHY and Phylip software packages were used for phylogenetic analyses [41,42].

The species abbreviations used in the alignments are: Aehy: Aeromonas hydrophila, Af: Archaeoglobus fulgidus, Aga: Anopheles gambiae, Ana: Anabaena sp. PCC 7120, Ap: Aeropyrum pernix, At: Arabidopsis thaliana, Atu: Agrobacterium tumefaciens, Ban: Bacillus anthracis, Bb: Borrelia burgdorferi, Bha: Bacillus halodurans, BPRB49: Bacteriophage RB49, Bs: Bacillus subtilis, Cac: Clostridium acetobutylicum, Cau: Chloroflexus aurantiacus, Ccr: Caulobacter crescentus, Ce: Caenorhabditis elegans, Chte: Chlorobium tepidum, Cgl: Corynebacterium glutamicum, Cpe: Clostridium perfringens, Ddi: Dictyostelium discoideum, Dm: Drosophila melanogaster, Ec: Escherichia coli, Gme: Geobacter metallireducens, Hi: Haemophilus influenzae, Hs: Homo sapiens, Hsp: Halobacterium sp., Lin: Listeria innocua, Lla: Lactococcus lactis, Lmo: Listeria monocytogenes, Mac: Methanosarcina acetivorans, Mcsp: Magnetococcus sp. Mfas: Macaca fascicularis, Mjan: Methanococcus jannaschii, Mka: Methanopyrus kandleri, Mlo: Mesorhizobium loti, Mma: Methanosarcina mazei, Mta: Methanothermobacter thermautotrophicus, Mtu: Mycobacterium tuberculosis, Nm: Neisseria meningitidis, Oih: Oceanobacillus iheyensis, Osa: Oryza sativa, Pa: Pyrococcus abyssi, Pae: Pseudomonas aeruginosa, Pfu: Pyrococcus furiosus, Ph: Pyrococcus horikoshii, Pmar: Prochlorococcus marinus, Pmu: Pasteurella multocida, Pyae: Pyrobaculum aerophilum,

Rsol: Ralstonia solanacearum, Sa: Staphylococcus aureus, Scoe: Streptomyces coelicolor, Sen: Salmonella enterica, Sme: Sinorhizobium meliloti, Spn: Streptococcus pneumoniae, Spy: Streptococcus pyogenes, Sso: Sulfolobus solfataricus, Ssp: Synechocystis sp. PCC 6803, Sst: Sulfolobus tokodaii, StLT2: Salmonella typhimurium LT2, Vch: Vibrio cholerae, Xaxo: Xanthomonas axonopodis, Xca: Xanthomonas campestris, Xfa: Xylella fastidiosa, Ype: Yersinia pestis.

Authors' contributions

Author 1 (LMI) contributed to the discovery process, preparation of the multiple sequence alignments and figures, Author 2 (LA) contributed to the discovery process, preparation of the figures and manuscript and conceived the study. All authors read and approved the final manuscript.

Additional material

Additional file 1

A text copy of the alignments of the CYTH domain and the CHAD are provided in simple alignment and Clustal aln formats Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2164-3-33-S1.txt]

Additional file 2

Alignments presented in the article in PDF format. Click here for file [http://www.biomedcentral.com/content/supplementary/1471-2164-3-33-S2.pdf]

References

- 1. Stryer L: Biochemistry Newyork, NY: W H Freeman and Co 1995
- 2. Nelson DL, Cox MM: Lehninger Principles of Biochemistry Worth Publishers Inc 2000
- 3. Murzin AG, Brenner SE, Hubbard T, Chothia C: **SCOP:** a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995, **247**:536-540
- Åravind L, Mazumder R, Vasudevan S, Koonin EV: Trends in protein evolution inferred from sequence and structure analysis. Curr Opin Struct Biol 2002, 12:392-399
- Mushegian AR, Koonin EV: A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci U S A 1996, 93:10268-10273
- Saraste M, Sibbald PR, Wittinghofer A: The P-loop a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci 1990, 15:430-434
- Gorbalenya AE, Koonin EV: Superfamily of UvrA-related NTPbinding proteins. Implications for rational classification of recombination/repair systems. J Mol Biol 1990, 213:583-591
- 8. Vetter IR, Wittinghofer A: Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer. Q Rev Biophys 1999, 32:1-56
- Bork P, Sander C, Valencia A: An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. Proc Natl Acad Sci U S A 1992, 89:7290-7294
- Artymiuk PJ, Poirrette AR, Rice DW, Willett P: A polymerase I palm in adenylyl cyclase? Nature 1997, 388:33-34
- 11. Murzin AG: How far divergent evolution goes in proteins. Curr Opin Struct Biol 1998, 8:380-387

- Leipe DD, Wolf YI, Koonin EV, Aravind L: Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 2002, 317:41-72
- 13. Pei J, Grishin NV: **GGDEF** domain is homologous to adenylyl cyclase. *Proteins* 2001, **42:**210-216
- Koonin EV, Wolf YI, Kondrashov AS, Aravind L: Bacterial homologs of the small subunit of eukaryotic DNA primase. J Mol Microbiol Biotechnol 2000, 2:509-512
- Aravind L, Koonin EV: DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. Nucleic Acids Res 1999, 27:1609-1618
- 16. Sismeiro O, Trotot P, Biville F, Vivares C, Danchin A: Aeromonas hydrophila adenylyl cyclase 2: a new class of adenylyl cyclases with thermophilic properties and sequence similarities to proteins from hyperthermophilic archaebacteria. J Bacteriol 1998, 180:3339-3344
- 17. Lakaye B, Makarchikov AF, Antunes AF, Zorzi W, Coumans B, De Pauw E, Wins P, Grisar T, Bettendorff L: **Molecular characteriza**tion of a specific thiamine triphosphatase widely expressed in mammalian tissues. J Biol Chem 2002, **277**:13771-13777
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389-402
- Aravind L, Koonin EV: The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem Sci 1998, 23:469-472
- 20. Aravind L, Koonin EV: A novel family of predicted phosphoesterases includes Drosophila prune protein and bacterial RecJ exonuclease. Trends Biochem Sci 1998, 23:17-19
- Leipe DD, Aravind L, Grishin NV, Koonin EV: The bacterial replicative helicase DnaB evolved from a RecA duplication. Genome Res 2000, 10:5-16
- 22. Wolf YI, Aravind L, Grishin NV, Koonin EV: Evolution of aminoacyl-tRNA synthetases – analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res* 1999, **9**:689-710
- 23. Anantharaman V, Koonin EV, Aravind L: Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids* Res 2002
- Ponting CP, Aravind L, Schultz J, Bork P, Koonin EV: Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. J Mol Biol 1999, 289:729-745
- 25. Schultz JE, Klumpp S: Cyclic GMP in lower forms. Adv Pharmacol 1994, 26:285-303
- Huynen M, Snel B, Lathe W 3rd, Bork P: Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. Genome Res 2000, 10:1204-1210
- 27. lyer LM, Koonin EV, Aravind L: Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* 2002, 3:8
- Aravind L: Guilt by association: contextual information in genome analysis. Genome Res 2000, 10:1074-1077
- Makarova KŠ, Aravind L, Grishin NV, Rogozin IB, Koonin EV: A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res 2002, 30:482-496
- Dandekar T, Snel B, Huynen M, Bork P: Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem Sci 1998, 23:324-328
- Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV: Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome* Res 2001, 11:356-372
- Kornberg A, Rao NN, Ault-Riche D: Inorganic polyphosphate: a molecule of many functions. Annu Rev Biochem 1999, 68:89-125
- Shiba T, Tsutsumi K, Ishige K, Noguchi T: Inorganic polyphosphate and polyphosphate kinase: their novel biological functions and applications. *Biochemistry (Mosc)* 2000, 65:315-323
- Ishige K, Noguchi T: Polyphosphate:AMP phosphotransferase and polyphosphate:ADP phosphotransferase activities of Pseudomonas aeruginosa. Biochem Biophys Res Commun 2001, 281:821-826

- 35. Ishige K, Noguchi T: Inorganic polyphosphate kinase and adenylate kinase participate in the polyphosphate:AMP phosphotransferase activity of Escherichia coli. Proc Natl Acad Sci U S A 2000, 97:14168-71
- 36. Eddy SR: Profile hidden Markov models. Bioinformatics 1998, 14:755-63
- Schuler GD, Altschul SF, Lipman DJ: A workbench for multiple alignment construction and analysis. Proteins 1991, 9:180-190
- Notredame C, Higgins DG, Heringa J: T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000, 302:205-17
- 39. Rost B, Sander C: Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 1993, 232:584-99
- Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ: JPred: a consensus secondary structure prediction server. Bioinformatics 1998, 14:892-893
- Felsenstein J: Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol* 1996, 266:418-27
- Hasegawa M, Kishino H, Saitou N: On the maximum likelihood method in molecular phylogenetics. J Mol Evol 1991, 32:443-5



"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime." Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours you keep the copyright

Submit your manuscript here: http://www.biomedcentral.com/info/publishing_adv.asp **BioMed**central