RESEARCH ARTICLE



Open Access

Genome-wide analysis of putative peroxiredoxin in unicellular and filamentous cyanobacteria

Hongli Cui^{1,2†}, Yipeng Wang^{1†}, Yinchu Wang^{1,2} and Song Qin^{1*}

Abstract

Background: Cyanobacteria are photoautotrophic prokaryotes with wide variations in genome sizes and ecological habitats. Peroxiredoxin (PRX) is an important protein that plays essential roles in protecting own cells against reactive oxygen species (ROS). PRXs have been identified from mammals, fungi and higher plants. However, knowledge on cyanobacterial PRXs still remains obscure. With the availability of 37 sequenced cyanobacterial genomes, we performed a comprehensive comparative analysis of PRXs and explored their diversity, distribution, domain structure and evolution.

Results: Overall 244 putative *prx* genes were identified, which were abundant in filamentous diazotrophic cyanobacteria, *Acaryochloris marina* MBIC 11017, and unicellular cyanobacteria inhabiting freshwater and hot-springs, while poor in all *Prochlorococcus* and marine *Synechococcus* strains. Among these putative genes, 25 open reading frames (ORFs) encoding hypothetical proteins were identified as *prx* gene family members and the others were already annotated as *prx* genes. All 244 putative PRXs were classified into five major subfamilies (1-Cys, 2-Cys, BCP, PRX5_like, and PRX-like) according to their domain structures. The catalytic motifs of the cyanobacterial PRXs were similar to those of eukaryotic PRXs and highly conserved in all but the PRX-like subfamily. Classical motif (CXXC) of thioredoxin was detected in protein sequences from the PRX-like subfamily. Phylogenetic tree constructed of catalytic domains coincided well with the domain structures of PRXs and the phylogenies based on 16s *rRNA*.

Conclusions: The distribution of genes encoding PRXs in different unicellular and filamentous cyanobacteria especially those sub-families like PRX-like or 1-Cys PRX correlate with the genome size, eco-physiology, and physiological properties of the organisms. Cyanobacterial and eukaryotic PRXs share similar conserved motifs, indicating that cyanobacteria adopt similar catalytic mechanisms as eukaryotes. All cyanobacterial PRX proteins share highly similar structures, implying that these genes may originate from a common ancestor. In this study, a general framework of the sequence-structure-function connections of the PRXs was revealed, which may facilitate functional investigations of PRXs in various organisms.

Keywords: Peroxiredoxin, Structure, Phylogeny and evolution, Comparative genomics, Cyanobacteria

Background

Cyanobacteria are among the earliest organism branching groups on earth, dating back 2.5-3.5 billion years, based on the fossil evidences [1]. As a taxonomic unit characterized by the first photosynthetic organisms with an oxygenic type of photosynthesis [2,3], cyanobacteria comprise a large number of species with diverse genome sizes and

* Correspondence: sqin@yic.ac.cn

¹Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Chunhui Road, Yantai 264003, People's Republic of China ecological habitats. Specifically, the genome size varies from 1.6 Mb (*Prochlorococcus* sp. MIT9301) to 9.0 Mb (*Nostoc punctiforme* PCC 73102) and the number of genes ranges from 1,756 (*Prochlorococcus marinus* MED4) to 8,462 (*Acaryochloris marina* MBIC11017) [4-6]. The remarkable variation in genome size indicates their significance in comparative genome research [7]. Cyanobacteria may also be unicellular or filamentous and can be found in almost all the conceivable environments, including marine and freshwater habitats, soil and rocks and extreme environments [8,9]. Unicellular cyanobacteria (*Prochlorococcus* and *Synechococcus*), which can inhabit



© 2012 Cui et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[†]Equal contributors

Full list of author information is available at the end of the article

ocean and possess the smallest genome size, is responsible for significant biomass and primary production in the marine biosphere [10]. Three unicellular cyanobacteria (Thermosynechococcus elongatus BP-1, Synechococcus sp. JA-2-3B'a (2-13) and Synechococcus sp. JA-3-3Ab) were isolated from hot-springs. Other unicellular species have larger genome sizes, including water bloom forming cyanobacteria (Synechocystis sp. PCC 6803 and Microcystis aeruginosa NIES-843), a thylakoids-absence cyanobacterium (Gloeobacter sp. PCC 7421), a nitrogen-fixing cyanobacterium (Cyanothece sp. ATCC 51142), and an animal-cyanobacterial symbionsis (Acaryochloris marina MBIC11017) [11]. The diazotrophic filamentous cyanobacteria have the largest genome sizes and include strains isolated from fresh water (Anabaena PCC7120, Anabaena variabilis ATCC 29413 and Arthrospira. platensis NIES-39), from a plant-cyanobacterial symbionsis (Nostoc punctiforme PCC29133), and from tropical and subtropical oceans (Trichodesmium erythraeum IMS101). In addition, the phylogeny of sequenced cyanobacterial organisms has been reported in previous studies [7,12,13].

Similar to heterotrophic organisms, cyanobacteria need to manage the ROS generated by oxygen reduction; however, they must also regulate ROS produced during photosynthetic electron transport [14,15]. Indeed, cyanobacteria constantly produce oxygen under illumination, which makes it crucial for them to prevent electron escape from normal electron transfer pathways to oxygen and ROS production [14]. Living organisms have developed various antioxidant defense mechanisms to protect themselves against ROS damage, including enzymatic (catalases, superoxide dismutases (SOD) and peroxidases), and non-enzymatic (glutathione, peroxiredoxins, vitamin A, C, E, and carotenoids) pathway [14,16,17].

The main factors involved in the cyanobacterial ROSscavenging system are low molecular mass antioxidants (peroxiredoxins, ferredoxin, glutathione, beta-carotenoids, and tocopherol) and enzymes of the Halliwell-Asada cycle in combination with peroxisomal catalase and superoxide dismutase [15,18-20]. A catalaseperoxidase was purified and characterized from Synechococcus elongatus PCC 7942 [21]. Additionally, the katG gene (encoding bi-functional catalase-peroxidase) was cloned and characterized from Synechocystis sp. strain PCC 6803 [22-24]. Recently, several studies about the catalytic mechanisms of the bi-functional catalase KatG from Synechocystis PCC 6803 have been published (for a review, see [25]). Genome sequence analysis of 64 cyanobacterial SODs indicated that the Cu/Zn form of SOD is rare among all cyanobacteria. Specifically, the marine unicellular Prochlorococcus species only possess Ni SOD, whereas other unicellular strains possess Fe SOD and Ni SOD or Fe SOD and Mn SOD [26].

Peroxiredoxins (PRXs) comprise an important antioxidant protein family with the ability to detoxify peroxide and the prx gene has recently been identified from higher plants [27]. Members of the PRX family are thiolspecific reductases or peroxidases [28]. PRXs exist as the form of multiple isoforms and catalyze the reduction of a broad range of different peroxides, including hydrogen peroxide, alkyl hydroperoxides and peroxinitrite [29,30]. The existence of different PRX family members has already been recorded in a wide variety of organisms ranging from archaea to mammals [31]. Six different sub-classes of PRXs, PRX I-IV (2-Cys PRX), PRX V (Type II PRX) and PRX VI (1-Cys PRX), have been identified from mammalian systems [32]. However, only four PRX sub-classes (1-Cys PRX, 2-Cys PRX, Type II PRX and PRX Q) have been reported in higher plants systems [29]. Analyses of the genome sequence of Synechocystis sp. PCC 6803 revealed the presence of five genes encoding peroxiredoxins 2-Cys PRX (sll0755), 1-Cys PRX (sll1198), two PRX Q (sll0221 and slr0242) and one Type II PRX (sll1621) [19,28,33]. Analyses of the genome sequence of Synechococcus elongatus PCC 7942 led to identification of six putative prx genes including one 1-Cys PRX, one 2-Cys PRX and four PRX Q [34]. Now that with the complete and partial of genomes from several cyanobacterial species, genome-wide identification and analysis of PRXs in cyanobacteria becomes possible.

Recently, 37 genomes of unicellular and filamentous cyanobacteria became available, which has facilitated the cyanobacterial systemic analysis of carotenoid cleavage dioxygenases [35], the metacaspases family [7], fatty acid desaturases [36], serine/threonine protein kinases [12], restriction modification systems [37], and carotenoids biosynthesis [38]. Comparative genomic investigations of cyanobacterial superoxide dismutases have also been conducted [26]. In this study, we selected 11 previously characterized PRXs from *Synechocystis* sp. PCC 6803 and Synechococcus elongatus PCC 7942 to search for cyanobacterial PRXs at the genome level. A BLASTp-plus-HMMsearch-phylogeny reconstruction approach was employed to analyze PRXs, focusing on their classification, distribution, structure, phylogeny and evolution. A better understanding of cyanobacterial PRXs can help us to understand the antioxidant mechanisms of cvanobacteria.

Results

Identification of open reading frames encoding PRX proteins

A total of 37 complete and partial cyanobacterial genomes were downloaded from the JGI genome portal [39] or Cyanobase [40] and used for this analysis. The information and phylogeny of 37 sequenced cyanobacterial strains were listed in Figure 1. The BLAST (BLASTp

			Key feature	Acro-	Genome	Total	Total	Tota	1 Perce
				nym	Size(Mb)	genes	proteins	Prxs	s ntage
o4 [Gloeol	pacter violaceus PCC 7421	Unicellular marine	7421	4.66	4,484	4,430	9	(0.20)
Ĩ	Sy	nechococcus sp. JA-3-3Ab	Unicellular hot-spring	JA33	2.93	2,815	2,760	8	(0.29)
60	100 Sy	nechococcus sp. JA-2-3B'a(2-13)	Unicellular hot-spring	JA23	3.05	2,914	2,862	8	(0.28)
	Thermos	ynechococcus elongatus BP-1	Unicellular hot-spring	BP-1	2.59	2,528	-	7	-
	8d Pro	chlorococcus marinus str. MIT 9215	Unicellular marine	9215	5 1.74	2,025	1,983	5	(0.25)
	64 Pro	chlorococcus marinus str. MIT 9301	Unicellular marine	9301	1.64	1,949	1,907	4	(0.26)
	96 Pro	chlorococcus marinus str. AS9601	Unicellular marine	9601	1.67	1,964	1,921	4	(0.21)
	98 Pro	chlorococcus marinus MIT9312	Unicellular marine	9312	1.71	2,007	1,810	4	(0.22)
	Pro	chlorococcus marinus str. MIT 9515	Unicellular marine	9515	1.70	1,948	1,906	4	(0.20)
	97 Pro	chlorococcus marinus MED4	Unicellular marine	MED4	1.66	1,756	1,717	5	(0.29)
	64 Pro	chlorococcus marinus str. NATL1A	Unicellular marine	TL1A	1.86	2,236	2,193	6	(0.27)
	og Pro	chlorococcus marinus str. NATL2A	Unicellular marine	TL2A	1.84	2.207	2,163	6	(0.28)
	Proc	hlorococcus marinus SS120	Unicellular marine	S120	1.75	1,928	1,883	3	(0.16)
	56 C Proc	hlorococcus marinus str. MIT 9211	Unicellular marine	9211	1.67	1,900	1,855	4	(0.22)
	Proch	lorococcus marinus str. MIT 9303	Unicellular marine	9303	2.68	3,049	2,997	6	(0.20)
	Syne	chococcus sp. WH 7803	Unicellular marine	7803	2.37	2,586	2,533	4	(0.16)
	100 Syne	echococcus sp. CC9311	Unicellular marine	9311	2.61	2,944	2,892	6	(0.21)
	Syne	chococcus sp. WH8102	Unicellular marine	8102	2.44	2,579	2,519	5	(0.20)
	Syn	echococcus sp. CC9605	Unicellular marine	9605	2.51	2,692	2,645	5	(0.19)
98	96 Syne	chococcus sp. CC9902	Unicellular marine	9902	2.23	2,355	2,307	5	(0.30)
	L Syne	chococcus sp. RCC307	Unicellular marine	R30	7 2.22	2,583	2,535	6	(0.24)
	Syneo	hococcus elongatus PCC 6301	Unicellular freshwater	6301	2.75	2,580	2,527	7	(0.28)
	100 Syneo	hococcus elongatus PCC 7942	Unicellular freshwater	7942	2.74	2,715	2,612	7	(0.27)
	100 No	ostoc punctiforme ATCC 29133	Filamentous freshwater	29133	9.06	6,794	6,087	10	(0.16)
	Anal	baena variabilis ATCC 29413	Filamentous freshwater	29413	7.07	5,724	5,043	9	(0.18)
90	100 Anat	baena sp. PCC 7120	Filamentous soil	7120	0 7.21	6,223	5,366	9	(0.17)
	Arth	rospira platensis NIES-39	Filamentous freshwater	39	6.79	6,676	-	11	-
	91 Land Tri	chodesmium erythraeum IMS101	Filamentous marine	101	7.75	4,498	4,451	9	(0.20)
75	55 Cyar	nothece sp. ATCC 51142	Unicellular freshwater	51142	5.46	5,359	4,762	7	(015)
	64 Cyan	othece sp. PCC 8801	Unicellular freshwater	8801	4.79	4,616	4,567	7	(0.15)
[]8	<u>н</u> — с	yanothece sp. PCC 7424	Unicellular freshwater	7424	6.55	5,763	5,625	9	(016)
41	Syn	echocystis sp. PCC 6803	Unicellular freshwater	6803	3.96	3,725	3,172	6	(0.19)
64	5 Micr	ocystis aeruginosa NIES-843	Unicellular freshwater	7421	5.84	6,363	6,312	7	(0.11)
	Synec	hococcus sp. PCC 7002	Unicellular marine	7002	3.41	3,235	2.823	6	(0.21)
L	Cyanoth	ece sp. PCC 7425	Unicellular freshwater	7425	5.79	5,380	5,297	9	(0.17)
96	Acaryo	chloris marina MBIC11017	Unicellular symbiosis	11017	8.36	8,462	6,254	12	(0.19)

H 0.01

Figure 1 Phylogenetic tree of the sequenced cyanobacterial strains and *prx* **information.** A Maximum likelihood tree for 36 sequenced cyanobacteria constructed based on 16s rRNA was reconstructed as described in the Methods [35]. The General Time Reversible (GTR) substitution model was selected assuming an estimated proportion of invariant sites and four gamma-distributed rate categories to account for rate heterogeneity across sites. The reliability on internal branches was assessed using the bootstrapping method (400 bootstrap replicates). Percentages in brackets represent total PRXs as a percentage of the total proteins. *Prochlorococcus marinus* MIT 9313 are absent from this tree because they are partial genomes and have no complete *rRNA* genes.

and tBLASTn) and HMM (hmmsearch) programs were used to search for proteins similar to confirmed cyanobacterial PRXs in each cyanobacterial genome. Pfam and SMART analysis using the derived sequences were then carried out to eliminate false positives. Among the 254 investigated proteins, ten that were originally annotated as Trx (NIES39_D06120, P9215_11961, Syncc9605_1945, Syncc9902_0720, and SYNW0724), Trx like protein (9301_02651), putative SOD (Syncc9902_0982), probable BCP (BP-1_0473), HP (7421_3157) and Trx domain 2 (Syncc9902_0354) respectively, were found to lack the important catalytic domains of typical PRX upon Pfam and SMART analysis and thus excluded from further consideration. As a result, a total of 244 proteins were considered in this study and an additional table file shows this in more detail [see Additional file 1, Table S1 and S2], among which 79 were originally annotated as AhpC/TSA or AhpC/TSA family members, 66 were originally annotated as BCPs (putative BCP or BCP homolog), 25 were originally annotated as peroxidases and 25 were originally annotated as peroxiredoxins. The remaining 49 proteins were accepted as PRX family members for this study, including 12 proteins annotated by other additional domains (such as 1-Cys, 2-Cys, TSPA, glutaredoxin-family domain protein and rehydin), 25 proteins annotated as hypothetical proteins, 8 proteins annotated as redoxins, and 4 proteins annotated as twin-arginine translocation pathway proteins.

The distribution of prx genes encoding PRX proteins

Amid diverse cyanobacterial genomes, the number of prx genes varies from 3 to 12 and the percentage of PRXs in the total proteins ranges from 0.11-0.30% (Figure 1). Among all unicellular cyanobacteria, symbiont Acaryochloris marina MBIC 11017 possesses 12 prxs, which is much higher than other species. However, the percentage of PRXs within the total proteins of this organism was only 0.19%, which is not the highest among unicellular cyanobacteria. The low ratio may be a result of the large genome of Acaryochloris marina MBIC 11017. Within marine unicellular cyanobacteria, the thylakoids-lacking cyanobacterium Gloeobacter sp. PCC 7421 possesses 9 prxs, which is much higher than others. Only three prx genes were found in Prochlorococcus marinus SS120, while four to six prx genes were found in other Prochlorococcus marinus strains and all marine Synechococcus strains, including WH 7803/8102, CC 9311/9605/9902, RCC 307, and PCC 7002. The percentage of PRXs within the total proteins was approximately 0.20% in the Prochlorococcus marinus strains and marine Synechococcus strains. Three Synechococcus strains inhabiting hot springs (BP-1, JA-2-3B'a(2–3), and JA-3-3Ab) and two freshwater Synechococcus elongatus strains (PCC 6301 and PCC 7942) were found to contain eight and seven *prx* genes, respectively, and these had similar percentages of PRXs in the total proteins (0.27-0.29%). All *Cyanothece* strains were found to contain seven (ATCC 51142 and PCC 8801) or nine (PCC 7424 and PCC 7425) *prx* genes, and the percentages of PRXs within the total proteins were 0.15%-0.17% for these cyanobacteria. The water-blooming cyanobacterium *Microcystis aeruginosa* NIES-843 was found to contain seven *prx* genes and the percentage of PRXs (0.11%) was the lowest among all investigated cyanobacteria. Six *prx* genes were found in *Synechocystis* sp. PCC 6803.

Compared with unicellular cyanobacteria, filamentous diazotrophic cyanobacteria possess more *prx* genes (10 for *Nostoc punctiforme* PCC 29133, 9 for *Anabaena variabilis* ATCC 29413, 9 for *Anabaena* sp. PCC 7120, 9 for *Trichodesmium erythraeum* IMS 101, and 11 for *Arthrospira platensis* NIES-39). However, the percentages of PRXs in the total proteins of these cyanobacteria were only 0.16%-0.18%, which was lower than those from marine unicellular cyanobacteria.

The number of *prx* gene is different from various habitat niches and genome sizes (Figure 2A). Unicellular cyanobacteria habiting marine contain the minimum amount of prx than those from freshwater and hotsprings. A similar phenomenon occurred in the filamentous cyanobacteria from marine and freshwater. The number of *prx* gene is increasing along with the increasing of the genome size of different cyanobacteria (Figure 2A). However, regardless of the habitat niches and cellular morphology, the percentage of PRX in the total proteins decreased along with the increased genome sizes. It is evident from these findings that filamentous diazotrophic cyanobacteria contain more prx genes than unicellular species, whereas the number of prx genes provides insufficient representation after allowing for their larger genomes. Moreover, in order to study the relationship between gene distribution and properties of the organisms, Spearman Rank Correlation test (R) was carried out and specific results were summarized in Figure 2B. Based on the summary on Figure 2B, the correlations between different properties and gene distribution were different. The total number of prx genes and genes encoding PRX from 1-Cys and PRX-like subfamilies share close correlations with genome size and ecophysiology properties of the organisms, while the other did not.

Structures and functions

Pfam and SMART domain analysis could not distinguish subfamilies among the cyanobacterial PRXs. Moreover, most of the proteins originally annotated as AhpC/TSA, BCP, and peroxiredoxin were not classified into distinct subfamilies. Fortunately, based on structural characteristics acquired from the CDD domain (Conserved Domain



number of *prx* in each species was determined based on the genes encoding PRX in the genomes. Percentages represent total *prx* as a percentage of total proteins. Strain names and the eco-physiological properties are as in Figure 1 and Additional file 1. Statistical analyses on the relationship between the distribution of genes encoding PRXs from different sub-families and the properties of 37 cyanobacterial organisms were performed using the Spearman Rank Correlation test (R), respectively. For the test of the distribution of *prx* and the genome size, cellular morphology, habitat and eco-physiological properties of 37 cyanobacterial organisms, the *X_i* is GS, U/E, MAR/FRE or UM/UF/UH/FM/FF and the *Y_i* is the total number of *prx* or the number of *prx* belonging to different sub-families and the in each cyanobacterial organisms. Note: *X*, the independent variable; *Y*, the dependent variable; GS, genome size (from small to large); U/F, unicellular or filamentous; MAR /FRE, marine or freshwater; UM/UF/UH/FM/FF, unicellular marine, unicellular freshwater, unicellular hot-spring, filamentous marine or filamentous freshwater. Spearman Rank test indicated that the distribution of some PRX family such as PRX-like or 1-Cys PRX correlate well with the eco-physiological properties and genome sizes of cyanobacterial ("*", *p*-value <=0.01; "**", *p*-value <=0.01).

Database) analysis, the identified cyanobacterial PRXs could be classified into five major subfamilies: 1-Cys PRX, 2-Cys PRX, PRX BCP, PRX5_like, and PRX-like (Figure 3).

Cyanobacterial PRX subfamily I (1-Cys PRX) includes 20 (8.19%) PRXs with less than 200 amino acid residues and is considered to possess the basic active sites in 26–50 residues. Genes encoding PRX proteins from this subfamily are present in five filamentous cyanobacteria (*Anabaena* sp. PCC 7120, *Anabaena variabilis* ATCC 29413, *Arthrospira platensis* NIES-39, *Nostoc punctiforme* ATCC 29133 and *Trichodesmium erythraeum* IMS101), eight unicellular cyanobacteria inhabiting freshwater (*Synechocystis* sp. PCC 6803, *Microcystis aeruginosa* NIES-843, *Synechococcus elongatus* PCC 6301/7942, *Cyanothece* sp. PCC 8801/7424/7425 and *Cyanothece* sp. ATCC 51142), and three unicellular cyanobacteria inhabiting hot-springs (*Thermosynechococcus*

elongatus BP-1, *Synechococcus* sp. JA-3-3Ab and *Synechococcus* sp. JA-2-3B'a(2–13)). It is interesting that 1-Cys PRX coding genes are a single gene in each cyanobacterial strain, whereas two genes encoding this PRX are found in *Acaryochloris marina* MBIC11017. However, genes encoding PRX from this subfamily are absent from all marine unicellular cyanobacteria except for *Gloeobacter violaceus* PCC 7421 and *Synechococcus* PCC 7002.

Subfamily II (2-Cys PRX) is the largest class of PRXs and characterized by two conserved redox-active cysteines, a peroxidatic cysteine (generally near residues 51–73) and a resolving cysteine (near residues 183–188). Subfamily II contains 37 (15.16%) proteins with less than 210 amino acid residues. Every one of all cyanobacterial organisms possess a single gene coding 2-Cys PRX respectively, suggesting that these genes are highly conserved throughout the evolutionary history.



Figure 3 Schematic representation and distribution of putative cyanobacterial PRX. Fused domains forming a single polypeptide chain are connected by a horizontal line. The red rectangles represent the Trx_like superfamily. Elliptical shadows with different colours represent different PRX subfamilies. Yellow rectangles represent the length of overlap. Strain and domain names are as in Figure 1 and Additional file 1, respectively. Figures are not drawn to scale.

Subfamily III (PRX BCP), bacterioferritin comigratory protein (BCP), was named based on its electrophoretic mobility before its function was known. BCP contains the peroxidatic cysteine and a putative resolving cysteine near the N-terminal. This subfamily was further divided into two types. Type a (PRX BCP-A) contains 85 (34.84%) proteins with less than 170 amino acid residues and was considered to possess the peroxidatic cysteinal basic structure in residues 44-61. There are several paralogous genes encoding PRXs from this type, which are widely distributed among almost all cyanobacteria except for Cyanothece sp. ATCC 51142, Prochlorococcus marinus SS120, Synechococcus PCC 7002, and Synechocystis sp. PCC 6803. Type b (PRX BCP-B) comprises 37 (15.16%) proteins with less than 200 amino acid residues and is considered to possess the peroxidatic cysteinal basic structure in residues 75-93. Compared to the paralogous genes encoding PRX BCP-A, all 37 cyanobacterial organisms possess a single gene encoding PRX BCP-B. It is apparent that the position of the peroxidatic cysteinal basic structure can be applied to distinguish these two types of PRX BCP proteins, which comprise the majority (50.00%) of cyanobacterial PRXs.

The fourth subfamily of PRX is PRX5-like, a homodimeric trx peroxidase, is widely expressed in mitochondria, peroxisomes and cytosol. This subfamily comprises 15 (6.14%) proteins with less than 190 amino acid residues and is considered to possess a peroxidatic cysteinal basic structure in residues 46–63. These 15 (6.14%) proteins are found in *Acaryochloris marina* MBIC11017, *Anabaena* sp. PCC 7120, *Cyanothece* sp. PCC 7424/ 7425, *Cyanothece* sp. ATCC 51142, *Nostoc punctiforme* ATCC 29133, *Microcystis aeruginosa* NIES-843, *Prochlorococcus marinus* 9313/9303/9311, *Synechococcus* PCC 7002, *Synechocystis* sp. PCC 6803, *Arthrospira platensis* NIES-39, and *Trichodesmium erythraeum* IMS101. *Prx* genes encoding PRX proteins from this subfamily are only detected in a few cyanobacteria, rather than all cyanbacterial strains, implying that they may exist in a species-specific fashion.

The last subfamily of PRX is PRX-like, members of which were originally annotated as hypothetical proteins. The protein sequences from this subfamily show similarity to PRXs and contain the conserved CXXC motif. We speculated that one specific cysteine in the motif corresponds to the peroxidatic cysteine of PRX. However, these proteins do not contain the other two residues of the typical catalytic triad of PRX. This subfamily was further divided into two types. Type c (PRX_like1) possesses the CXXC motif (near residues 52-65) in the N-terminal, as well as the putative typical catalytic triad of PRX in the C-terminal (near residues 134-140). The 32 (13.11%) proteins from this type were found to be distributed among all filamentous cyanobacteria and unicellular cyanobacteria living in marine (Synechococcus), freshwater (except for Synechocystis sp. PCC 6803), and hot-springs, whereas they were absent from all Prochlorococcus marinus (except for 9215 and 9301). Type d (PRX_like2) possesses the CXXC motif (near residues 64-77) in the N-terminal and contains 17 proteins (6.96%) that are distributed in all five filamentous cyanobacteria (Anabaena sp. PCC 7120, Anabaena variabilis ATCC 29413, Arthrospira

platensis NIES-39, *Nostoc punctiforme* ATCC 29133, and *Trichodesmium erythraeum* IMS101), three hot-springs inhabitant cyanobacteria (*Thermosynechococcus elongatus* BP-1, *Synechococcus* sp. JA-3-3Ab and *Synechococcus* sp. JA-2-3B'a(2–13)), and the freshwater unicellular *Cyanothece* group.

Phylogenetic analysis

To elucidate the evolutionary histories between species and cyanobacterial prx genes, the translated protein sequences of these genes and previously proven PRX proteins (Table 1) were applied to construct the phylogenetic tree. Six major clades were observed in the phylogenetic tree in general (Figure 4). PRXs from 1-Cys, 2-Cys, PRX BCP and most of the PRX-like2 subfamilies belonged to the first monophyletic (BS: 80%) group. The second monophyletic (BS: 75%) group contains members of the PRX5_like and PRX-like subfamily with all PRX-like1 and some PRX-like2, which cluster separately according to their domains, respectively. According to the results of the phylogenetic tree (Figure 4), most members (except proteins 7120_1206 and 11107_5336) of different subfamilies are consistent with the classification (Figure 3 and Additional file 1, Table S1) based on CDD domain analysis, which indicates that cyanobacterial PRXs cluster strictly according to their structural characteristics. In addition, the PRXs generally cluster within each subfamily according to the phylogeny of the species.

Several interesting results emerged from further analysis of the phylogeny of cyanobacterial PRXs. All *prx* genes

Table 1 List of organisms and PRX protein sequences analyzed in this study (except for the sequences from cyanobacterial genomes)

Species	Accession No.	Length	Protein
Arabidopsis thaliana	CAA72804.1	216	1-cys
Arabidopsis thaliana	sp Q96291.2	266	2-Cys Prx A
Arabidopsis thaliana	sp Q9C5R8.3	273	2-Cys Prx B
Arabidopsis thaliana	AEE77109.1	216	Prx Q
Arabidopsis thaliana	AAM65848.1	162	Prx -2B
Arabidopsis thaliana	AEE74337.1	201	Prx -2F
Arabidopsis thaliana	NP_176774.1	553	Prx -2A
Arabidopsis thaliana	sp 022711.2	162	Prx -2D
Arabidopsis thaliana	sp Q949U7.2	234	Prx -2E
Arabidopsis thaliana	sp Q9SRZ4.1	162	Prx -2C
Homo sapiens	AAA50464	199	Prx I (2-Cys)
Homo sapiens	AAA50465	198	Prx II (2-Cys)
Homo sapiens	BAA08389	256	Prx III (2-Cys)
Homo sapiens	AAB95175	271	Prx IV (2-Cys)
Homo sapiens	AAF03750	214	Prx V (atypical 2-Cys
Homo sapiens	BAA03496	224	Prx VI (1-Cys)

encoding PRX BCP formed three major clades and an additional figure file shows this in more detail [see Additional file 2, Figure S1]. Several paralogous genes encoding PRX BCP-A compose a monophyletic (BS: 90%) group. As expected, the PRX Q from Arabidopsis thaliana [GenBank: AEE77109.1] clusters with the PRX BCP subfamily, suggesting a cyanobacterial-origin of this gene in higher plants. Meanwhile, genes encoding PRX BCP-B proteins form a monophyletic (BS: 89%) group. Most genes encoding PRX BCP are paralogous based on their close evolutionary relationships, suggesting that they share common ancestors and may have been produced by recent gene duplication. It is obvious that PRXs BCP from Gloeobacter violaceus PCC7421 (gll_0506), Synechococcus sp. JA-2-3B'a(2-13) (CYB_1376), Synechococcus sp. JA-3-3Ab (CYA_2305), and Arthrospira platensis NIES-39 (NIES39_E02230) formed a separate cluster, respectively, indicating obvious species-specific duplication events in these strains. The 2-Cys PRX from higher plants build a monophyletic group (BS: 88%) with all the cyanobacterial 2-Cys PRXs except for 7421_3158, suggesting a common ancestor and an additional figure file shows this in more detail [see Additional file 3, Figure S2]. Surprisingly, more than one prx genes coding 2-Cys were discovered from Homo sapiens (four genes) and higher plants (two genes), indicating recent gene duplication occur in linage-specific fashion. All prx genes encoding PRX-like were clustered into two major clades and an additional figure file shows this in more detail [see Additional file 4, Figure S3]. Members belonging to PRX-like1 comprise a monophyletic (BS: 84%) group. Members from PRX-like2 build a monophyletic group (BS: 99%). It is interesting that one protein (Anabaena sp. PCC 7120: 7120_1206) belonged to Prx5_like subfamily build a monophyletic group (BS: 96%) with three prx encoding Prx-like1, suggesting that a natural recombination, a lateral gene transfer, or convergent evolution took place. In the subfamily PRX5_like and an additional figure file shows this in more detail [see Additional file 5, Figure S4], the PRX5_like subfamily also includes six type II PRXs (type 2A/2B/2C/ 2D/2E/2F) from Arabidopsis thaliana [GenBank: NP_176774.1, AAM65848.1, sp|Q9SRZ4.1, sp|O22711.2, sp[Q949U7.2, and AEE74337.1] and the typical 2-Cys PRX from Metazoa [GenBank: AAF03750]. Surprisingly, six prx genes encoding PRX from higher plants clustered with one protein from Metazoa but the cyanobacterial PRX, implying that a non-cyanobacterial origin of this gene encoding PRX typeII proteins in higher plants. Additionally, 1-Cys PRXs from Arabidopsis thaliana and Homo sapiens formed one clade and build sister group with all cyanobacterial 1-Cys PRXs, indicating a noncyanobacterial origin of 1-Cys prx genes in higher plants and an additional figure file shows this in more detail [see Additional file 6, Figure S5].



Conserved domain features

The redox-active cysteines play a crucial role in the function of all PRXs which were originally divided into two categories, 1-Cys and 2-Cys PRXs, based on the number of cysteine residues directly involved in catalysis. The guanidino group of the conserved arginine is presumed to stabilize the ionized peroxidatic cysteine [41]. We surveyed the cysteine-including motif and the number of conserved arginines from distinct protein sequences to facilitate the classification of different subfamilies (Figure 5 and Table 2). PRXs from 1-Cys and PRX5_like subfamilies contain only one cysteine-including motif (WAGDSWVVLFSHPADYTPVCTTELG) and (VVLXX-LPGAFTPTCSS) in the N-terminal, respectively. Two cysteine-including motifs (VVLFFYPLDFTFVCPTEVIA-FSD) and (DEVCPA) were found in the N-terminal and C-terminal of the 2-Cys PRXs, respectively. The results from PRX BCP are similar to those from 2-Cys PRXs, whereas the second cysteine-including motif is not conserved among some sequences. Members of PRX-like1 contain a cysteine-including motif in the N-terminal and a cysteine-including motif (AACTPDF) in the C-terminal, whereas PRX_like2 only possess the CXXC motif in the N-terminal. In addition, some arginines are conserved among all PRXs and these are primarily distributed near the C-terminal of the protein sequences.

Discussion

Peroxiredoxins (PRXs) are an important type of antioxidant protein

Photosynthetic organisms have evolved complicated mechanisms to protect themselves against ROS damage

	10 20 30					1
			<u> </u>			
301-2361 c	STMLALGTAAPDFALPDVVSGRTVQLADFSDRAG	LLVMFICQHC	PFVKH-VEEELAAIGR	DYLDLGIVAISPN	IS LETHPODGPEQLKAQ	AERL
942-1730	STMLALGTAAPDFALPDVVSGRTVQLADFSDRAG	LLVMFICQHC	PFVKH-VEEELAAIGR	DYLDLGIVAISPN	IS LETHPODGPEQLKAQ	AERL
9-Q00510	STMLSLGTIAPDFQLPDVTSGEIISVDTFKDKKA	LLVMFICCHC	PFVKH-IENELAKLGH	DYQPLGIGIVAISAN	NS IKTHPQDGPEHLKNQ	AETV
1017-4765	STMLDLGTPAPNFQLPDVVSGQTISLETFAHRNV	LLVMFICQHC	PFVKH-VQAELAQIGH	DYSDQSLGIVAISSN	DIQSHPLDDPEHLTAM	ATSL
01-3842	STMLKLGTIAPNFELPDVVSGNIISPSTFKGKKA	LLVMFICRHC	PFVKH-IQLELAKIGK	DYVPKDVGIVAISSN	IN IESHPDDA PDKLKEM	ASEL
9413-2915	SIMVPIGTQAPDFHLPDVVSGKTISLSTFADKKA	LLVMFICRHC	PFVKH-IQDELTRIGQ	DYATRDLGIVAISAN	NDAENYPDDAPESLKAL	AIEL
801-3616	SNSLALGSVAPDFSLPDITNNRTVKLSDYRNQQP	IVLAFTRIFTEKQYC	PFCYP-HI INLNQNYE	RFIEQGIEVLMITST	TDAKQSQIV	VRDL
1142-4124	TNQLLLGSIAPDFILPDITNNRMLRLGDYFHQQP	VILAFTRIFTEKQYC	PFCYP-HIVSLNKYYE	QFVERGVDVLMITS1	TDTQQSQIV	VRDL
120-4642	SNELRLDVGT PDFQLPDITNGTLVKLSNYRGKQP	ILLAFTRIFTEKQYC	PFCFP-HIKALNENYE	QFTNRGIEVLLVTS7	TDEKQSQIV	VKDL
9133-F6083	RNQLRLGVGT PDFQLPDITNGTLVKLSDYKGKQP	VLLAFTRI FTEKQYC	PFCFP-HIKALNENYE	EFKNRGIEVLMITST	rdegosQIV	VKDL *
01-0234	INDLDNGKRTPDFELPDITNGKLVRLSNYWGVTP	VIVALTRIFTEKQYC	PFCFP-HILALNKNYE	RFLERGVEVLMITST	TDEKQSKIV	VNDL
1017-2532	KGEPAVGRLAPDFELPRIG-GDSVKLSDYRGQQP	IFLAFTRIFTEKLEC	PFCYP-HIQDLKORYS	EIVDRGAELLMISST	DQVQSEQI	VSDL
120-4404	-MALRLGDTVPNFTQASTHGDIDFYFWA	GDSWVVLFSHPADY	PVCTT-ELGTVAKLKP	EFDKRNVKAIALSVI	DVESHNGW	VGDI
9413-1358	-MALRLGDTVPNFTQASTHGDIDFYEWA	GDSWVVLFSHPADY	PVCTT-ELGTVAKLKP	EFDKRNVKAIALSVI	DVESHKGW	VGDI 4
9133-F6498	-MALRLGDTVPNFTQASTHGDIDFYCWA	GDSWVVLFSHPADF	PVCTT-ELGTVAKLKP	EFDKRNVKAIALSVI	DVESHKGW	VGDI 2
A23-0113	-MSLRLGDTAPNFTQKSTHGEIDFYSWA	GDSWVVLFSHPADY	PVCTT-ELGTVAKLMP	EFEKRGVKVIALSVI	DVDSHVGW	VKDI
1142-0135	-MALQLGDTVPDFTQETSEGTISFHEWA	GDSWVVLFSHPADY	PVCTT-ELGTVASLKP	EFDKRNVKILALSVI	DSVDSHKGW	INDI
803-1198	-MALQLGDVVPDFTQESSQGPISFHEWA	GDSWVVLFSHPADY	PVCTT-ELGTVAKLKP	EFDKRNVKVIALSVI	DVESHKGW	ICDI
A-10331	NECLRVGMQAPDFATTAVVDQEFKDITLSQYR	G-KYVVLFFYPLDF	FVCPT-EITAFSDRYS	DFSSKNTEVLGVSVI	OSKFTHLAW	IQTP C
A-0352	NECLRVGMQAPDFATTAVVDQEFKDITLSQYR	G-KYVVLFFYPLDF7	FVCPT-EITAFSDRYS	DFSSKNTEVLGVSVI	DSKFTHLAW	IQTP V
313-0754	NECLRVGQQAPDFNATAVVDQEFKEISLSQYR	G-KYVVLFFYPLDF7	FVCPT-EITAFSDRYT	DFSSRNTEVLGVSVI	DSQFSHLAW	IQTS A
803-14621	NECLRVGQQAPDFSATAVVDQEFKEISLSQYR	G-KYVVLFFYPLDF7	FVCPT-EITAFSDRYT	DESSRNTEVLGVSVI	DSQFSHLAW	IQTS Y
505-1324	TGCLRVGQLAPDFTATAVVDQEFKEVTLSQYR	G-KYVVLFFYPLDFT	FVCPT-EITAFSDRYA	DFSSKNTEVLGVSVI	DSQFSHLAW	IQTP
02-1213	TGCLRVGQQAPDFTATAVVDQEFKEISLSOVR	G-KYVVLFFYPLDFT	FVCPT-EITAFSDRYA	DESSKNTEVLGVSVI	DSQFSHLAW	IQTP O'
801-03801	MQIGDKIPEFSLLDONGVKRSNRGLKNP	LVLFFYPKDD	PGCTI-EVCGFRDKYD	LFKVIGAQVWGVSNG	STSSHLAF	ANKN C
15-03801	MQIGDKIPEFSLLDQNGVKRSNKGLKNP	LVLFFYPKDDT	PGCTI-EVCGFRDKYD	LFKVIGAQVWGVSNO	STSSHLAF	ANKN
01-03811	MQIGDKIPEFSLLDQNGVKRSNKGLKNP	LVLFFYPKDDT	PGCTI-EVCGFRDKYD	LFKVIGAEVWGVSNG	STSSHLAF	ANKN OF
12-0353	MQIGDKIPEFSLLDQNGIKRSNKGLKNP	LVLFFYPKDDT	PGCTI-EVCGFRDKYD	LFKVIGAQVWGVSNO	SSSSHLAF	ANKN
15-03821	MKIGDKVPLFSLLDONGTRRTNEGLKSP	LVLFFYPKDDT	PGCTI-EVCGFRDKYD	LFKVIGAQVWGVSNO	TSSSHLAF	ANKN
d4-0345	MQIGDKVPQFSLLDONGTKRSNNGLKTP	LVLFFYPKDD	PGCTI-EVCGFRDKYD	LFKVIGAOVWGVSNO	SSSSHLAF	ANKN A'
15-09891	DFAPEVGDQAPNFHLEGFNKNIKSKSIWELSDFO	G-KWLVMYFYPKDF	AGCHL-EAKGESELKK	DFKKYNAEIVGISAI	DNQDSHESF	CSEK O
01-09581	DFAPEVGDIAPNFQLEGFNKNIKSKNIWELNDFO	G-KWLVMYFYPKDF	AGCTL-EAKGFSELKK	DESKYNAEIVGISAI	DNQDSHESF	CSEK _
12-0897	DFAPEVGDSAPNFOLEGFNKNIKSKKIWELSDFO	G-KWLVIYFYPKDF	AGCTL-EAKGFSELKK	DFSKYNAEIIGISAI	KKDSHESE	CSEK
01-09561	DFAPEVGDYAPNFKLEGFNKNIKTKKTWELSDFE	G-OWLVMY FY PKDF	AGCTL-EAKGESELKK	DESKYNAEIVGISAI	DNEDSHESF	CSER
15-09851	DYAPEVGDIAPSFHLEGINKNIKSKKIWDSNELK	G-KWVVLY FYPKDF	AGCTL-EAKGFSKLKK	DESKYNAEIIGISAI	DKEETHESF	CSEO
d4-0903	DYAPEIGDSAPSFHLEGINKSIKSKKIWDSNELI	G-KWIVLYFYPKDF	AGCTL-EAKGESOLKN	DESKYNAEIIGISAI	DNODSHDSF	CSEK
	VPDVVFKTRVRDESVEGPNPYRWEDKTTADIF	GGKKVVLFSLPGAFT	PICSSNHLPRYEELYS	EFOAOGVDEIICLSV	VNDAFVMFKW	GKEII ST
02-A0793	UDCUUEVODUDDE CUDC DNDVDWEDVOOEATE	GGKKVVLFSLPGAFT	PTCSSNHLPRYEQLFE	EFQALGVDDIICLSV	VNDAFVMFQW	GKQI 5
02-A0793 03-1621	VESVVERINVRDESVEGENEIRWEDRITEVIE		DROCONTH DDUPPT VD	EFKALGVDEVICISV	NDAFTMEOW	GKOO AT
02-A0793 03-1621 142-3126	VPSVVFKTRVRDESVPGPNPTRWEDRTTEQTF VPSVVFKTRVRDESVPGPNPYRWQDLTSEEIF	KGKKVIVFSLPGAF	FIUSSNHLFKIELLID		ATT THE T	
02-A0793 03-1621 .142-3126 .25-1596	VPSVVFKTRVRDESVPGPNPTRWED-KTTEQIF VPSVVFKTRVRDESVPGPNPYRWQDLTSEEIF VPNVTFKTRVRDESVPGPNPYRWODLTTEDIF	KGKKVIVFSLPGAF KGKKVVVFSLPGAF	PTCSSNHLPRIEELID	EFOAOGVDOIICISV	NDAFVMFOW	GKHI
02-A0793 03-1621 142-3126 25-1596 017-3680	VPSVVFKIKVRDESVPGPNPIRMD-TITELF VPSVVFKTRVRDESVPGPNPYRWQDLTSEEIF VPNVFFKTRVRDESVPGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVPGPNPYRWOD-KTTODIF	KGKKVIVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF	PTCSSNHLPRYEELYE	EFQAQGVDQIICISV	VNDAFVMFQW	GKHI
002-A0793 803-1621 1142-3126 125-1596 1017-3680 01-0162	VPSVVFKIKV ADBSVEGPNPYRWQDLTSELF VPSVVFKIKV RDBSVEGPNPYRWQDLTSELF VPNVTFKTRVRDBSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDBSVEGPNPYRWQDKTTQDIF VPDVVFKTRVRDBSVDGPNPFRWQDKTTQEIF 110 120 130	KGKEVIVFSLPGAF KGKEVVVFSLPGAF GGKEVVLFSLPGAF GGKEVVLFSLPGAF 140	PTCSSNHLPRIEELID PTCSSNHLPRIEELID PTCSSTHLPRIEELHE PTCSSTHLPRIEELID 150 160	EFQAQGVDQIICISV EIKAQGVDEIICLSV EFKGQGIDEIICLSV 170 18	MNDAFVMFQW MNDAFVMFQW MNDAFVMFQW 80 190 1	GKHI GKHQ GKQQ
002-A0793 803-1621 1142-3126 125-1596 1017-3680 01-0162 801-2361_c 942-1730	VPSVVFKIRVRDESVEGPNPYRWQDLTSEIF VPSVVFKIRVRDESVPGPNPYRWQDLTTEDIF VPDVVFKIRVRDESVPGPNPYRWQDKTTQDIF VPDVVFKIRVRDESVDGPNPFRWQDKTTQEIF 110 120 130 	KGKKVIVFSLPGAF KGKKVVVSLPGAF GGKKVVLFSLPGAF GGKKVVLFALPGAF 140 ACTPDFHLFI ACTPDFHLFI ACTPDFHLFI	PTCSSNHLPRIBELIE PTCSSTHLPRIBELYE PTCSSTHLPRIBELYE PTCSSTHLPRIBELYE 150 160 1	EFQAQGVDQIICISV EIKAQGVDEIICLSV EFKGQGIDEIICLSV 170 16 SSDRPVTGADLRAAI SSDRPVTGADLRAAI	MDAFVMFQW MDAFVMFQW MDAFVMFQW 80 190 1 1 ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS	GKHI GKHQ GKQQ 200 1 IGCN IGCN IGCN IGCN
002-A0793 803-1621 1142-3126 125-1596 1017-3680 01-0162 001-2361_c 042-1730 -000510 017-4765	VPSVVFKIRVRDESVEGPNPYRWQDLTSEIF VPSVVFKIRVRDESVEGPNPYRWQDLTTEDIF VPDVVFKIRVRDESVEGPNPYRWQDKTTQDIF VPDVVFKIRVRDESVEGPNPYRWQDKTTQEIF 110 120 130 	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKVVLFSLPGAF GGKKVVLFALPGAF 140 1ACTPDFHLFI ACTPDFHLFI ACTPDFHLFI	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSNLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELTN EBLYN RDRLVYRGQLDDSRP ISDRQLVYRGQLDDSRP	E FQAQGVDQIICISV EIKAQGVDEIICLSV EFKGQGIDEIICLSV 170 11 SSRFVTGADLRAAI SSRFVTGADLRAAI GNNLFVTGADLRAAI	ANDAFVMFQW ANDAFVMFQW 80 190 II.S. ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS IDAILADQPINSDQKPS	CKHI CKHQ CKHQ CO CC CC CC CC CC CC CC CC CC
002-A0793 003-1621 1142-3126 125-1596 0017-3680 11-0162 001-2361_c 442-1730 -Q00510 0017-4765	VPSVVFKIRVRDSVEGPNPYRWQDLTSELF VPSVVFKIRVRDSVEGPNPYRWQDLTSELF VPDVVFKTRVRDSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDSVEGPNPRWQDKTTQLF 110 120 130	KGKKVVVESLPGAF KGKKVVVESLPGAF GGKKVVLFSLPGAF GGKKVVLFALPGAF ACTPDFHER ACTPDFHER ACTPDFHER ACTPDFHER	PTCSSNILFRIEBTD PTCSSNILFRIEBTD PTCSSNILFRIETURS PTCSSTHLPRIEELHEI PTCSSTHLPRIEELHEI PTCSSTHLPRIEELHEI PTCSSTHLPRIEELHEI PTCSSTHLPRIEELT ISDROLVYRGQLDDSRPI NEDRRLVYRGQLDDSRPI NEDRRLVYRGQLDDSRPI NEDRRLVYRGQLDDSRPI NEDRRLVYRGQLDDSRPI	EFQAQGVDQIICISV EIKAQGVDEIICLSV EFKGQGIDEIICLSV 170 16 SSDRPVTGADLRAAI SNRLPVTGADLRAAI SNNLPVTGADLRAAI SNNLPVTGADLRAAI	MDAFVMFQW MDAFVMFQW MDAFVMFQW 80 190 II.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.	CKHI SKHO GKQQ 20 IGCN IGCN IGCN IGCN VCCN
002-A0793 103-1621 1142-3126 125-1596 0013680 11-0162 001-2361_c 001-2361_c 001-23510 001-4765 11-3842 001-2355	VPSVVFKIRVRDSVSVFGPNPYRWQDLTTEDIF VPSVVFKIRVRDESVPGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVPGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVPGPNPYRWQDKTTQEIF 110 120 130 FTFPYCLDETQAIAKAYTA GFTFPYCLDETQAIAKAYTA G	KGKKVVVFSLPGAF KGKKVVVESLPGAF GGKRVVLFSLPGAF I ACTPDFHLFI ACTPDFHLFI ACTPDFFLFI ACTPDFFLFI ACTPDFFLFI	PTCSSNLEPRIBLID PTCSSNLEPRIEBLID PTCSSTLEPRIEBLHE PTCSSTLEPRIEBLHE PTCSSTLEPRIEBLHE PTCSSTLEPRIEBLY EDRLIVRGQLDDSRP NDRRLVYRGQLDDSRP NDRQLVYRGQLDDSRP NDRVLVYRGQLDDSRP NDFKLEVRGQLDDSRP	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 16 SSDRPVTGADLRAAI SNDRPVTGADLRAAI SNNLMVTGEDLRAAI SNNLMVTGEDLRAAI	MDAFVMFQW MDAFVMFQW MDAFVMFQW 100 190 11.1.1.1 ITAVLTGDAPSADQKPS IDAVLTGDAPSADQKPS IDAVLTGDAPSADQKPS IDAVLAGKNIEECOKPS	20 20 CROO
02-A0793 03-1621 142-3126 25-1596 017-3680 1-0162 001-2361 c 42-1730 -200510 017-4765 1-3842 413-2915	VPSVVFKIRVRDSVESVEGPNPYRWQDLITEDIF VPSVVFKIRVRDESVPGPNPYRWQDLITEDIF VPDVVFKIRVRDESVPGPNPYRWQDKITQDIF VPDVVFKIRVRDESVPGPNPFRWQDKITQDIF 110 120 130 FTFPYCLDETQAIAKAYTA GFTFPYCLDETQAIAKAYTA GFTFPYCLDETQEVAQAYTA GFTFPYCLDETQEVAQAYTA GFTFPYCPETQEVAKQYTA EFTFPYCPETQEVAKQYTA EFTFPYCPETQEVAKQYTA EFTFPYCPETQEVAKQYTA E	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKVVLFSLPGAF GGKKVVLFSLPGAF I 	PTCSSNLPRIBLTD PTCSSNLPRIELPRIELPRI PTCSSTLPRIELPRIELPRI PTCSSTLPRIELPRIELPRI EDGSTTLPRIELPRIELPRI EDGSTTLPRIELPRIELPRI EDGSTTLPRIELPRIELPRI EDGSTTLPRIELPRIELT EDGSTTLPRIE EDGSTTLPRIELT EDGSTTLPRIELT EDGSTTLPRIE EDGSTTL	E FQAQGVDQIICISV EIKAQGVDEIICLSV EFKGQGIDEIICLSV 170 16 SDR PVTGADLRAAI SDR PVTGADLRAAI SNNLMVTGADLRAAI SNNLMVTGBDLRAAI SNNEKPVTGADLRAAI	MDAFVMFQW MDAFVMFQW MDAFVMFQW 80 190 IIII. TTAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS IDALAEQPISANQKPS IDALAEQPISANQKPS IDAVLAGKPVIDEQKPS TDAVLAGKPVIDEQKPS	CHI GKHO CO IGCN IGCN IGCN IGCN VGCN VGCN VGCN VGCN
02-A0793 03-1621 142-3126 25-1596 017-3680 1-0162 017-3680 1-0162 017-3680 1-0162 017-3680 1-0162 017-3640 1-3842 413-2915 01-3616 142-4124	VPSVVFKIRVRDSVEGPNPYRWQDLTSEXIF VPSVVFKIRVRDSVEGPNPYRWQDLTSEXIF VPDVVFKTRVRDSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDSVEGPNPYRWQDKTTQXF 110 120 130 1 GFTFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQEVQAYTA GFSFPYCDETQEVAQAYTA G	KGKKVVVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFALPGAF I 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSNLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELT ISDRLVYRGQLDDSRP NAHRLVYRGQLDDSRP NAHRLYRGQLDDSRP NAHRLYRGQLDDSRP NAHRLYRGQLDDSRP NAHRLYRGQLDDSRP NAHRLYRGQLDDSRP	EFQAQGVDQIICISV EIKAQGVDEIICLSV EFKGQGIDEIICLSV 170 16 SSDRPVTGADLRAAI SNDLPVTGADLRAAI SNNLPVTGADLRAAI SNNLPVTGADLRAAI SNEKPVTGSDLRAAI SNEKPVTGADLRAAI	ANDAFVMFQW ANDAFVMFQW NDAFVMFQW 80 190 IIIII. ITAVLTGDAPSADQKPS IDAVLAGDAPSADQKPS IDAVLAGKNIEEEQKPS IDAVLAGKNIEEEQKPS IDAVLAGKPVIDEQKPS FDKLLKFL	SKHI SKHI SKQ 20 IGCN IGCN IGCN VGCN
02-A0793 03-1621 142-3126 25-1596 017-3680 1-0162 01-2361 c 4-20510 017-4765 1-3842 413-2915 01-3616 142-4124 20-4642	VPSVVFKIRVRDSVEGPNPYRWQDLTSELF VPSVVFKIRVRDESVPGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVPGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVDGPNPFRWQDKTTQEIF 110 120 130	KGKKVVVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF Id0 IACTPDFHLFI ACTPDFHLFI ACTPDFHLFI ACTPDFHLFI ACTPDFHLFI ACTPDFHLFI GAPLPAQFVLI GAPLPAQFVLI GAPLPAQFVLI	PTCSSNLPRIBLID PTCSSNLPRIELPRI PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELING ELLEN RDRRLVYRGQLDDSRP NDRRLVYRGQLDDSRP NDYRLVYRGQLDDSRP NDYRLVYRGQLDDSRP NDYRLYRGQLDDSRP NDYRCLUYRGQLDDSRP NTGYLYRHES	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 18 SDR PVTGADLRAAI SNDE PVTGADLRAAI SNNLMVTGEDLRGAI SNEK PVTGADLRAAI SNGK PVTGADLRAAI SNGK PVTGADLRAAI SNGK PVTGADLRAAI SNGK PVTGADLRAAI SNGK PVTGADLRAAI SNGK PVTGADLRAAI	MDAFVMFQW MDAFVMFQW MDAFVMFQW 190 11.1.1.1 ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS LDAILADQPINSDQKPS LDALAEQPISANQKPS LDAVLAGKNIEBEQKPS FDKLLKFL	ZON ZON ZON IGCN IGCN IGCN VGCN VGCN VGCN ZON ZON ZON ZON ZON ZON ZON ZO
002-A0793 003-1621 142-3126 25-1596 001-2361_c 42-1730 -000510 017-4765 1-3842 413-2915 01-3616 142-4124 20-4642 133-F6083	VPSVVFKIRVRDSVEGPNPYRWQDLTEEIF VPSVVFKIRVRDESVPGPNPYRWQDLTEEIF VPDVVFKTRVRDESVPGPNPYRWQDKTTQDIF VPDVVFKTRVRDESVDGPNPFRWQDKTTQEIF 110 120 130	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF I40 I I I I ACTPDFHLFI ACTPDFHLFI ACTPDFHLFI ACTPDFHLFI ACTPDFVFI GAPLPAQFVLI GAPLPAQFVLI GAPLPAQFVLI GAPLPAQFVLI	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSNLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE ISDRLPYRCQLDDSRP NDRRLVYRCQLDDSRP NDRLVYRCQLDDSRP NDRLVYRCQLDDSRP NDRLLYRCQLDDSRP NDRLLYRCQLDDSRP NTCRLQYKLFS	E FQAQGVDQIICISV EIKAQGVDEIICLSV EFKGQGIDEIICLSV 170 IE SSDRPVTGADLRAAI SSDRPVTGADLRAAI SNNLMVTGEDLRAAI SNKKPVTGEDLRAAI SNKKPVTGEDLRAAI SNKKPVTGEDLRAAI SNKKPVTGEDLRAAI FDHNAKV FDHNASV FDDHNASV	MDAFVMFQW MDAFVMFQW MDAFVMFQW MDAFVMFQW 1	CKHI GKHO CROO IGCN IGCN IGCN IGCN VGCN VGCN VGCN VGCN VGCN VGCN
02-A0793 03-1621 142-3126 25-1596 017-3680 1-0162 017-3680 1-0162 017-4765 1-3842 413-2915 11-3842 413-2915 11-3616 142-4124 20-4642 133-F6083 1-0234	VPSVVFKIRVRDSVEGPNPYRWQDLTSEIF VPSVVFKIRVRDSVPGPNPYRWQDLTTEDIF VPDVVFKRVRDSVPGPNPYRWQDLTTEDIF VPDVVFKRVRDSVPGPNPYRWQDKTTQIF 110 120 130	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKVVVFSLPGAF GGKVVLFSLPGAF GGKKVVLFSLPGAF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSNLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELT PTCSSTL	E FQAQGVDQIICISV EIKAQGVDEIICLSV EFKGQGIDEIICLSV 170 16 SDR PVTGADLRAAI SDR PVTGADLRAAI SNLKVYTGADLRAAI SNLKVYTGSDLRAAI SNKKPVTGSDLRAAI SNKKPVTGADLRAAI FDHNAKI FDHNASS FLDHNASS FLDHNASS	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW 80 190 II.N.I. ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS IDALAEQPISANQKPS IDALAEQPISANQKPS IDAVLAGKPVIDEQKPS FDKLLKFL VQLLLKTS IEILLEQFN IEILLEQFN	CHI GKHO CROO IGCN IGCN IGCN IGCN VGCN VGCN VGCN VGCN VGCN VGCN VGCN V
002-A0793 103-1621 1142-3126 125-1596 001-2361 001-2361 001-2361 001-2361 001-2361 001-2361 001-2361 001-2351 001-3616 1142-4124 120-4642 013-572	VPSVVFKIRVRDSVESVEGPNPYRWQDLTSELF VPSVVFKIRVRDSSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDSSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDSSVEGPNPRWQDLTTQIF 110 120 130	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFALPGAF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE ELVI ELVI ELVI RDRRLVYRGQLDDSRP AHRKLVYRGQLDDSRP AHRKLVYRGQLDDSRP MTGKIQYKLFS	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 16 SSDR FVTGADLRAAI SNDR FVTGADLRAAI SNNLMVTGEDLRGAI SNNLMVTGEDLRGAI SNEK FVTGADLRAAI SNEK FVTGADLRAAI SNEK FVTGADLRAAI 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ISO 190 II.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.	CREAT CREAT
02-A0793 03-1621 142-3126 25-1596 017-3680 1-0162 01-2361 c 42-1730 -00510 017-4765 1-3842 413-2915 01-3616 142-4124 20-4642 133-F6083 1-0234 017-2532 20-4404	VPSVVFKINVRDSVESVEGPNPYRWQDLITEDIF VPSVVFKINVRDESVEGPNPYRWQDLITEDIF VPDVVFKINVRDESVEGPNPYRWQDLITEDIF VPDVVFKINVRDESVEGPNPYRWQDLITEDIF VPDVVFKINVRDESVEGPNPYRWQDKITQEIF 110 120 130	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKVVLFSLPGAF GGKVVLFSLPGAF GGKVVLFSLPGAF GGKVVLFSLPGAF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELHE PTCSSTLPRIELT PTCSSTLPRI	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 IE SSDR PVTGADLRAAI SSDR PVTGADLRAAI SNNLMVTGEDLRAAI SNNLMVTGEDLRAAI SNKFVTGADLRAAI FFDHNASV FFDHNASV FLDRNASS FLDRNASS FLDRNASS	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFVMFQW IDAVLAGDPSADQKPS IDAVLAGQPISADQKPS IDAVLAGKPVIDEQKPS' IDAVLAGKPVIDEQKPS' IDAVLAGKPVIDEQKPS' IDAVLAGKPVIDEQKPS' IDAVLAGKPVIDEQKPS' IDAVLAGKPVIDEQKPS' IDAVLAGKPU	CKHI CKHI CKHO CROO
02-A0793 003-1621 142-3126 25-1596 017-3680 1-0162 017-3680 1-0162 017-4765 1-3842 413-2915 01-3616 142-4124 20-4642 133-F6083 1-0234 017-2532 20-4404 413-1358	VPSVVFKIRVRDSVESVEGPNPYRWQDLTSELIF VPSVVFKIRVRDESVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVEGPNPYRWQDKTTQDIF 110 120 130	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF I ACTPDFHLF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSNLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI PTCSSTLPRIELHEI ISDRUTRGQLDDSRP RDRRLVYRGQLDDSRP WDYKLVYRGQLDDSRP WDYKLVYRGQLDDSRP WDYKLVYRGQLDDSRP WTCKLQYKLFS	E FQAQGVDQIICIS\ EIKAQGVDEIICLS\ EFKGQGIDEIICLS\ 170 IE SSDRPVTGADLRAAI SSDRPVTGADLRAAI SNNLMVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI FFDHNAK FFDHNAS\ FLDRNAS3 FLDRNAS3 FLDHNAS FLDHNAS1 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW 30 190 11I. TTAVLTGDAPSADQKPS LDAILADQPINSDQKPS LDAILADQPINSDQKPS LDAILAGQPISANQKPS LDAULAGKNIEEEQKPS TDAVLAGKVIEEEQKPS TDKLLKFL ZQLLLKIS	CKHI CKHO CROO 20 20 CCN IGCN IGCN IGCN VGCN
002-A0793 103-1621 1142-3126 125-1596 001-2361 c 11-0162 001-2361 c 11-0162 001-2361 c 11-0162 001-2361 c 11-0162 11-0162 11-3842 11-3842 11-3842 11-3842 11-3842 11-2532 12-2552 12-2552 12-2552 12-	VPSVVFKIRVRDSVFGPNPYRWQDLTSEIF VPSVVFKIRVRDSVPGPNPYRWQDLTTEDIF VPDVVFKTRVRDSVPGPNPYRWQDLTTEDIF VPDVVFKTRVRDSVPGPNPYRWQDKTTQIF 110 120 130 GFFFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCDETQEVQAYTA GFSFPYCDETQEVQAYTA GFSFPYCDETQEVQAYATA GFSFPYCDETQEVAKSYIA G	KGKKVVVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELTD STLPRIELTD STLPRIELTD STLPRIELTD RDRRLVYRGQLDDSR RDRRLVYRGQLDDSR NDRLVYRGQLDDSR NDRLVYRGQLDDSR NDRLVYRGQLDDSR NDRLVYRGQLDDSR NDRLVYRGQLDDSR NDRLVYRGQLDDSR NDRLVYRGQLDDSR NDRLVYRGQLDDSR NDRLVYRGQLDSR NDRLVYRGLSS STLP ST	E FQAQGVDQI ICISV E IKAQGVDEI ICLSV E IKAQGVDEI ICLSV 170 16 SSDR PVTGAD LRAAI SNDL PVTGAD LRAAI SNNL PVTGAD LRAAI SNNL PVTGAD LRAAI SNNL PVTGAD LRAAI SNNL PVTGAD LRAAI SNEK PVTGAD LRAAI SNEK PVTGAD LRAAI SNEK PVTGAD LRAAI SNEK	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS LDAILADQFINSDQKPS LDAILADQFINSDQKPS LDAILAEQPISANQKPS LDALAEQPISANQKPS LDALLKKFD	SKHI GKHI GKUQ 20 1 GKUQ 1 GCN IGCN IGCN IGCN VGCN
02-A0793 03-1621 142-3126 25-1596 017-3680 1-0162 017-3680 1-0162 017-465 1-3842 413-2915 01-3616 142-4124 20-4642 133-F6083 1-0224 017-2522 20-4404 413-1358 23-6498 23-0113	VPSVVFKIRVRDSVEGPNPYRWQDLTSELF VPSVVFKIRVRDSSVEGPNPYRWQDLTTELF VPDVVFKTRVRDSSVEGPNPYRWQDLTTELF VPDVVFKTRVRDSSVEGPNPYRWQDLTTQLF 110 120 130	KGKKVVVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF I40 IACTPDFHLF ACTPDFHLF ACTPDFHLF ACTPDFHLF ACTPDFHLF GAPLPAQFVLL GAPLPAQFVLL GAPLPAQFVLL GAPLPAQFVLL GAPLPAQFVLL GAPLSQFVLL GAPLSQFVLL GAPLSQFVLL 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELTE	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 18 SDR PVTGADLRAAI SNDR PVTGADLRAAI SNNLMVTGEDLRAAI SNNLMVTGEDLRAAI SNNLMVTGEDLRAAI SNKFPTGADLRAAI FFDHNASS FFDHNASS FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	CKHI CKHI CKHO CROO
$\begin{array}{c} 002-A0793\\ 003-1621\\ 142-3126\\ 25-1596\\ 017-3680\\ 1-0162\\ 017-3680\\ 1-0162\\ 017-4765\\ 1-3842\\ 413-2915\\ 01-3616\\ 142-4124\\ 20-4642\\ 133-F6083\\ 11-0234\\ 017-2532\\ 20-4404\\ 413-1358\\ 1132-F6498\\ 132-F6498\\ 132-F6498$	VPSVVFR.IV.VDBSVEGPNPYR.WQDLTSELIF VPSVVFR.IV.VDBSVEGPNPYR.WQDLTTEDIF VPDVVFKT.VV.DESVPGPNPYR.WQDLTTEDIF VPDVVFKT.VV.DESVPGPNPYR.WQDKTTQDIF 110 120 130	KGKKVVVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELTD PTCSSTLPRIELTD INDRRLVYRGQLDDSRP RDRRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDYKLYRGQLDDSRP NDYKLFS	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 IE SSDR FVTGADLRAAI SSDR FVTGADLRAAI SNNLMVTGADLRAAI SNNLMVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKKPVTGADLRAAI SNKTAUSSAN SNKTAUS	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFV	CALL CALL
002-A0793 103-1621 1142-3126 125-1596 017-3680 11-0162 017-3680 11-0162 017-3680 11-0162 017-3680 11-0162 017-3680 11-0162 017-3616 113-3616 113-2532 1142-4124 120-4642 113-56083 113-56083 113-56498 123-56658 123-56688 123-56688 123-56688 123-56688 123-56688	VPSVVFKIRVRDSVESVEGPNPYRWQDLTSEIF VPSVVFKIRVRDSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDSVEGPNPYRWQDLTTQIF 110 120 130	KGKKVVVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKPVVFSVFI 	PTCSSNLPR/BLITD PTCSSNLPR/BLITD PTCSSTLPR/BLITD PTCSSTLPR/BLITD PTCSSTLPR/BLITD STALPR/BLITD STALPR/BLITD <td< th=""><th>E FQAQGVDQI ICISV E IKAQGVDEI ICLSV E IKAQGVDEI ICLSV 170 16 SSDR PVTGAD LRAAI SNL PVTGAD LRAAI SNL PVTGAD LRAAI SNL PVTGAD LRAAI SNL PVTGAD LRAAI SNL PVTGAD LRAAI SNL PVTGAD LRAAI SNEK PVTGAD</th><th>NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS LDAILADQFINSDQKPS LDAILADQFINSDQKPS LDATUAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQCPS IDA</th><th>SKHI GKHI GKUQ 20 IGCN IGCN IGCN IGCN VGCN</th></td<>	E FQAQGVDQI ICISV E IKAQGVDEI ICLSV E IKAQGVDEI ICLSV 170 16 SSDR PVTGAD LRAAI SNL PVTGAD LRAAI SNL PVTGAD LRAAI SNL PVTGAD LRAAI SNL PVTGAD LRAAI SNL PVTGAD LRAAI SNL PVTGAD LRAAI SNEK PVTGAD	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS LDAILADQFINSDQKPS LDAILADQFINSDQKPS LDATUAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKPVIDEQCPS IDA	SKHI GKHI GKUQ 20 IGCN IGCN IGCN IGCN VGCN
002-A0793 103-1621 1142-3126 125-1596 001-2361 017-3680 11-0162 001-2361 042-1730 -Q00510 0017-4765 11-3842 1413-2915 101-3616 1142-4124 120-4642 123-F6083 11-0234 011-2532 220-4404 1133-F6498 123-0113 1142-0135 1133-F6498 123-0113 1142-0135 1142-1198 1-10331	VPSVVFRIKVRDESVEGPNPYRWQDLTSELF VPSVVFRIKVRDESVEGPNPYRWQDLTSELF VPDVVFKTRVRDESVEGPNPYRWQDLTTELF VPDVVFKTRVRDESVEGPNPYRWQDLTTELF VPDVVFKTRVRDESVEGPNPYRWQDLTTQLF 110 120 130	KGKKVVVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF I40 I	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELTD STLPRIELTRIELTD RDRRLVYRGQLDDSRP RDRRLVYRGQLDDSRP NDRRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDVKLVYRGQLDDSRP NDVKLVYRGQLDDSRP NDVKLVYRGQLDDSRP NDVKLVYRGQLDDSRP NDVKLVYRGQLDDSRP NDVKLVYRGQLDDSRP NDVKLVYRGQLDSRP NDVKLVRGQLDSRP NDVKLVYRGQLDSRP NDVFTRFN NDGRLYKHLFS NCGRLYKHLFS NCGRLYKHLFS NDCGRLYKHLFS NDCGRLYKHLFS NDCGRLYKHLFS NDCGRLYKHLFS NDCGRLYKHLFS NDCGRLYKHLFS NDCGRLYKHLFS NDCGRLYKHLFS NDCGRLYKHFFS NDCGRLYKHFS NDCGRLYKHFS NDCGRLYKHFS NDCGRLYKHFS NDCGRLYKHFS NDCGRLYKHFS NDCGRLYKHFS NDCGRLYKHFS NDCGRLYKHFS NDCGRLYKHFS NDCGRLYKHFS NDCGRLYK	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 16 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFV	CREAT CR
002-A0793 003-1621 142-3126 125-1596 017-3680 01-0162 017-3680 01-0162 017-3680 01-0162 017-4765 11-3842 017-4765 11-3842 017-3616 1122-4124 120-4642 0133-F6683 11-0234 017-2532 20-4404 1132-F6498 113	VPSVVFRIKTVRDESVEGPNPYRWQDLTSELIF VPSVVFRIKTVRDESVEGPNPYRWQDLTTEDIF VPDVVFRTRVRDESVEGPNPYRWQDLTTEDIF VPDVVFRTRVRDESVEGPNPYRWQDKTTQDIF VPDVVFRTRVRDESVEGPNPFRWQDKTTQDIF 110 120 130	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKRVVLFSLPGAF GGKRVVLFSLPGAF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELTD PTCSSTLPRIELTD PTCSSTLPRIELTD PTCSSTLPRIELTD PTCSSTLPRIELTD PTCSTLPRIELTD PTCSSTLPRIELTD	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 IE SSDR PVTGADLRAAI SSDR PVTGADLRAAI SNDK PVTGADLRAAI SNNL PVTGADLRAAI SNNL PVTGADLRAAI SNNL PVTGADLRAAI SNNL PVTGADLRAAI SNNL PVTGADLRAAI SNGK PVTGADLRAAI S	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFV	CALL CALL CALL CALL
002-A0793 003-1621 1142-3126 125-1596 017-3680 11-0162 017-3680 11-0162 017-3680 11-0162 042-1730 0-00510 0.017-4765 11-3842 013-616 113-2915 01-3616 01-2915 01-29	VPSVVFKIRVRDSVESVEGPNPYRWQDLTSEXIF VPSVVFKIRVRDSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDSVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDSVEGPNPYRWQDLTTQIF 110 120 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA G	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKVVLFSLPGAF GGKPVLFSVFVF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELTD PTCSSTLPRIELTD ELTTSSTLPR	EFQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 IE SSDRPVTGADLRAAI SSDRPVTGADLRAAI SNLKVTGADLRAAI SNLKVTGADLRAAI SNLKVTGADLRAAI SNKKPVTGALKAAI SNKKPVTGADLRAAII SNKKPVTGADLRAAII SNKKPVTGA	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS DATLADQFINSDQKPS DATLADQFINSDQKPS DATLADQFINSDQKPS DATLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKPVIDEQKPS IDAVLAGKP	SKHI GKHI GKUQ 20 IGCN IGCN IGCN IGCN VGCN
002-A0793 003-1621 1142-3126 125-1596 001-2361 017-3680 11-0162 001-2361 017-3680 11-0162 001-2361 01-3842 01-3842 01-3842 01-3842 01-3842 01-3842 01-3842 01-3842 01-234 01-234 01-0235 01-358 0-	VPSVVFRIKVRDESVEGPNPYRWQDLTSELF VPSVVFRIKVRDESVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVEGPNPYRWQDKTTQEIF 110 120 130	KGKMVUFSLPGAF KGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVLFSLPGAF GGKWVLFSLPGAF GGKWVLFSLPGAF GGKWVLFSLPGAF GGKWVLFSLPGAF GGKPULFSLP 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELTD STLPRIELTD STLPRIELTD RDRRLVYRGQLDDSRP NDRRLVYRGQLDDSRP NDRRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRUYRGQLDDSRP NDRUYRGQLDDSRP NDRUYRGQLDDSRP NDRUYRGQLDDSRP NDRUYRGQLDDSRP NDRUYRGQLDDSRP NDRUYRGQLDDSRP NDRUYRGQLDDSRP NDRUYRGQLDSRP NDRUYRGQLDSRP NDRUKLYRGQLDSRP NDRUKLISS NDGRLYKHFS NDRUKLRSFY PSTG NDRUKLRSFY PSTG NDRUKLRLSFY PSTG NDRUKLRLFY PSTG NDRUKLR PSTG	E FQAQGVDQIICISV E IKAQGVDEIICLSV E IKAQGVDEIICLSV 170 16 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS TTAVLTGDAPSADQKPS LDAILADQPINSDQKPS LDAILADQPINSDQKPS LDAILADQPINSDQKPS LDAILAGPVIDEQKPS' FDKLLKFL TDQLLLKTS TDQLLLKTS IEILLEQIPOST FDKLLKFD IEILLEQIPOST TDNYS-VATPADWKDG LTDNYS-VATPADWKDG LTDNYS-VATPADWKDG LTDNYS-VATPADWKDG LTDNYS-VATPADWKDG LTDNYS-VATPADWKDG LTDNYS-VATPADWKDG LTDYQ-VATPAWKDG LTDYQ-VATPAWKDG CVQSHPDEVCPAWTPG VQSHPDEVCPAWTPG	CREAT CR
002-A0793 003-1621 1142-3126 125-1596 001-2361 c 142-1730 001-3680 01-0162 001-3680 01-0162 001-3680 01-0162 001-3680 01-3616 1142-4124 120-4642 0133-F6683 1132-F6498 0133-F6498 0133-F6498 0133-F6498 0131-F6498 0131-F6498 0131-198 0	VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPDVVFRTEVRDESVEGPNPYRWQDLTTEDIF VPDVVFRTEVRDESVEGPNPFRWQDKTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCPETQEVAQAYTA GFSFPYCPETQEVAQAYTA GFSFPYCPETQEVAQAYTA G	KGKKVVVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKRVVLFSLPGAF GGKRVVLFSLPGAF GGKRVVLFSLPGAF 	PTCSSNLEPRIELTD PTCSSNLEPRIELTD PTCSST	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 IE SDR PVTGADLRAAI SSDR PVTGADLRAAI SNDK PVTGADLRAAI SNNLMVTGEDLRAAI SNNLMVTGEDLRAAI SNNLMVTGEDLRAAI SNNLMVTGEDLRAAI SNGKPVTGADLRAAI FFDHNASV FFDHNASV FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI FLDRNASI RNFDELLRVIDSLQI RNFDELLRVIDSLQI RNFDELLRVIDSLQI RNFDELLRVIDSLQI RNFDELLRVIDSLQI RNFDELLRVIDSLQI RNFDELLRVIDSLQI RNFDELLRVIDSLQI RNTDETLRVLQAFQI RNVDETLRVLQAFQI RNVDETLRVLQAFQI	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFV MFQW INDAFV	CKHI CKHI CKHI CKHI CKHI CKHI CKHI IGCN IGCN IGCN Free Free CKHI IGCN Free IGCN Free IGCN Free IGCN Free IGCN Free IGCN IGCN Free IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN IGCN IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN Free IGCN IGCN IGCN Free IGCN IGCN IGCN IGCN IGCN IGCN Free IGCN IGCN IGCN IGCN IGCN Free IGCN I
002-A0793 003-1621 1142-3126 125-1596 001-2361 c 142-1730 0017-3680 01-0162 012-361 c 142-1730 0-000510 0017-4765 101-3616 1132-4705 1132-46083 113-66083 113-66083 113-66083 113-66083 113-66532 113-6683 113-67855 113-67855	VPSVVFRIKTVRDESVEGPNPYRWQDLTSEIF VPSVVFRIKTVRDESVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVEGPNPYRWQDLTTEDIF VPDVVFKTRVRDESVEGPNPYRWQDKTTQIF 110 120 130	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKVVLFSLPGAF GGKVVLFSLPGAF GGKVVLFSLPGAF GGKVVLFSLPGF 	PTCSSNLPRYELLT PTCSSNLPRYELT PTCSSTLPRYELT PTCSSTLPRYELT ELT STLPTSSTLPRYELT ELT RDRRLVYRGQLDDSR RDRRLVYRGQLDDSR NDRRLVYRGQLDDSR NDRLVYRGQLDDSR NDRCLVRGQLDDSR NDRCLVRGQLDDSR NDRKLYRGQLDDSR NDRKLYRGQLDDSR NDRKLYRGQLDDSR NDRKLF STL STL STL STL STL STL STL STL	E FQAQGVDQIICISV E IKAQGVDEIICLSV E IKAQGVDEIICLSV ITAQGVDEIICLSV ITA SSDR PVTGADLRAAI SSDR PVTGADLRAAI SNLKVTGADLRAAI SNLKVTGADLRAAI SNKKPVTGADLRAAI SNKTGAU SNKTTANI SNKTANI SNKTANI SNKTTANI SNKTA	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW INDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS IDAVLAGKDPINBQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIE IDAVLAGKNIE IDAVLAGKAPVIDGCKPS IDA	SKHI GKHI GKUQ 20 IGCN IGCN IGCN IGCN VGCN
002-A0793 103-1621 1142-3126 125-1596 001-2361 11-0162 001-2361 017-3680 11-0162 001-2361 017-2680 11-0162 11-0162 11-0162 11-0162 11-0164 112-2915 101-3616 1142-4124 120-4642 1132-F6083 11-0234 1132-F6098 123-6098 123-6098 1132-F6098 123-6	VPSVVFRIGVRDSVESVEGPNPYRWQDLTSEIF VPSVVFRIGVRDSVEGPNPYRWQDLTSEIF VPDVVFKTRVRDSSVEGPNPYRWQDLTTQIF VPDVVFKTRVRDSSVEGPNPRWQDLTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCDETQAIAKAYTA GFSFPYCDETQEVAKYTA GFSFPYCDETQEVAKYTA GFSFPYCDETQEVAKYTA GFSFPYCDETQEVAKYTA G	KGKKVVVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKVVLFSLPGAF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELTD STCSTLPRIELTD STCSTLPRIELTD ELITISSTLPRIELTD INDRRLVYRGQLDDSRP NDRRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDDSRP NDRLVYRGQLDSRP NDRLVYRGQLDSRP NDRLVYRGP	E FQAQGVDQIICISV E IKAQGVDEIICLSV E IKAQGVDEIICLSV 170 16 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS IDAVLAGKPVIDEQKPS' IDALADQPINSDQKPS IDALADQPISDQKPS IDAVLAGKPVIDEQKPS' FDKLLKFL	CREAT CR
002-A0793 003-1621 1142-3126 125-1596 017-3680 01-0162 017-3680 01-0162 017-3680 01-0162 017-4765 11-3842 142-4174	VPSVVFRIGVRDESVEGPNPYRWQDLTSEIF VPSVVFRIGVRDESVEGPNPYRWQDLTSEIF VPDVVFRGVRDESVEGPNPYRWQDLTTEDIF VPDVVFRGVRDESVEGPNPFRWQDKTTQIF 110 120 130 GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA G	KGKMVIVFSLPGAF KGKMVIVFSLPGAF GGKRVVLFSLPGAF GGKRVVLFSLPGAF GGKRVVLFSLPGAF GGKRVVLFSLPGAF GGKRVVLFSLPGAF GGKRVVLFSLPGAF GGKPVLFSLPGAF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELTD	E FQAQGVDQIICISV E IKAQGVDEIICLSV E IKAQGVDEIICLSV 170 18 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFV	CKHI CHI CHI CHI CHI CHI CHI CHI C
002-A0793 003-1621 142-3126 25-1596 017-3680 1-0162 017-3680 1-0162 017-3680 1-0162 017-4765 1-3842 413-2915 01-3616 142-4124 20-4642 133-F6083 11-0234 017-2532 20-4404 413-1358 133-F6498 23-0113 142-0135 03-1198 1-0352 13-0754 03-14621 03-14621 03-14621 03-124 02-1233 01-03801 15-03801 01-03811 12-0353 01-03811 01-03	VPSVVFRIEVERUBSVEGPNPYRWQDLTSERIF VPSVVFRIEVERVBSVPGPNPYRWQDLTTEDIF VPDVVFRTEVRDSVPGPNPYRWQDLTTEDIF VPDVVFRTEVRDSVPGPNPYRWQDLTTEDIF VPDVVFRTEVRDSVPGPNPYRWQDKTTQDIF 110 120 130	KGKKVVVFSLPGAF KGKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKVVLFSLPGAF GGKVVLFSLPGAF 	PTCSSNLPRYELLT PTCSSNLPRYELT PTCSSNLPRYELT PTCSSTLPRYELT PTCSS	E FQAQGVDQI ICISV E IKAQGVDEI ICLSV E IKAQGVDEI ICLSV 170 IF SSDR PVTGAD LRAAI SSDR PVTGAD LRAAI GNNL PVTGAD LRAAI GNNL PVTGAD LRAAI SNNK PVTGAD LRAAI RNFDELL RVI DSLQI RNFDELL RVI DSLQI RNFDELL RVI DSLQI RNFDELL RVI DSLQI RNFDELL RVI DSLQI RNFDEL RVI DSLQI RNTDETL RVLQAFQY RNVDETL RVLQAFQY RNVDETL RVLQAFQY EHIKEAI RVLKEI QY EHIKEAI RVLKEI QY	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKVIEQKPS IDAVLAGKANOK	SKHI GKHI GKUQ 20 IGCN IGCN IGCN VGC
002-A0793 003-1621 142-3126 25-1596 017-3680 1-0162 017-3680 1-0162 017-3680 1-0162 017-4765 1-3842 413-2915 01-3616 142-4124 20-4642 113-F6083 102-3618 113-F6083 113-F6083 113-F6498 123-F	VPSVVFRIKVRDESVEGPNPYRWQDLTSEIF VPSVVFRIKVRDESVEGPNPYRWQDLTSEIF VPDVVFKTRVRDESVEGPNPYRWQDLTTQIF VPDVVFKTRVRDESVEGPNPRWQDKTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCDETQAIAKAYTA GFSFPYCDETQEVAKYTA G	KGKKVVVFSLPGAF KGKKVVVFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKKVVLFSLPGAF GGKVVLFSLPGAF 	PTCSSNLPRYELLT PTCSSNLPRYELT PTCSSTLPRYELT STTLPRYELT STTLPRYELT PTCSSTLPRYELT STTLP	E FQAQGVDQIICISV E IKAQGVDEIICLSV E IKAQGVDEIICLSV 170 16 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS LDALLADQPINSDQKPS LDALLADQPINSDQKPS LDALLACPISS	CKHI GKHI GKUQ CO CO CO CC CC CC CC CC CC CC
002-A0793 003-1621 142-3126 25-1596 017-3680 1-0162 017-3680 1-0162 017-3680 1-0162 017-4765 1-3842 413-2915 01-3616 142-4124 20-4642 413-3F6083 1-0234 01-3258 133-F6498 133-F6498 133-F6498 133-F6498 133-F6498 133-F6498 133-F6498 133-F6498 133-F6498 133-F6498 133-F6498 133-754 003-14621 105-1324 003-14621 105-3281 115-03801 112-0353 15-03821 12-0353 15-03821 15-03	VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPDVVFRTEVRDESVEGPNPYRWQDLTTEDIF VPDVVFRTEVRDESVEGPNPFRWQDKTTQIF 110 120 130 GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA G	KGKMVIVESLEGAF KGKMVIVESLEGAF GGKRVVLESLEGAF GGKRVVLESLEGAF GGKRVVLESLEGAF GGKRVVLESLEGAF GGKRVVLESLEGAF GGKRVVLSSLEGAF GGKRVVLSSLEGAF GGKRVVESUES ACTPDFHEN 	PTCSSNLPRYELT PTCSSNLPRYELT PTCSSTLP PTCSSTLPYELT PTCSSTLP PTCSSTLPYELT PTCSSTLP PTCS	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 1E 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS LDALLADQPINSDQKPS LDALLAQPISDQKPS LDALLAQPISDQKPS IDALLAKEVIDEQKPS FDKLLKFL	CKHI CHI CHI CKHI CHI CHI CHI CHI CHI CHI CHI C
002-A0793 003-1621 142-3126 25-1596 017-3680 1-0162 017-3680 1-0162 017-3680 1-0162 017-4765 1-3842 413-2915 01-3616 142-4124 20-4642 133-F6083 11-0234 017-2532 20-4404 413-158 133-F6498 23-0113 142-0135 03-14621 103-14621 105-1324 001-03801 15-03801 15-03821 145-03821 15-03821 15-03821	VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPDVVFRTEVRDESVEGPNPYRWQDLTTEDIF VPDVVFRTEVRDESVEGPNPFRWQDKTTQIF 110 120 130 GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA G	KGKMVUFSLDGAF KGKWVUFSLDGAF GGKWVUFSLDGAF GGKWVUFSLDGAF GGKRVUFSLDGAF GGKRVUFSLDGAF GGKRVUFSLDGAF GGKRVFSLDGAF GGKRUFSLDGAF GGKDF GGKDF GGKDF GGKDF GGKDF GGKDF GGKDF GGKDF GF GGKDF GT GT GT GT GT GT GT GT GT GT GAD FA G	PTCSSNLPR/BLITD PTCSSNLPR/BLITD PTCSSNLPR/BLIPS PTCSSTLPR/BLIPS PTCSSTLPR/BLIPS BTCSTLPR/	E FQAQGVDQIICISV E IKAQGVDEIICLSV E IKAQGVDEIICLSV 170 If SSDR PVTGADLRAAI SSDR PVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI SNULWTGADLRAAI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDSLQI RNEDELRVIDALQA SNUDETLRVLQAFQY RNUDETLRVLQAFQY EHIKEAIRVLKEIQN EHIKEAIRVLKEIQN EHIKEAIRVLKEIQN EHIKEAIRVLKEIQN EHIKEAIRVLKEIQN EHIKEAIRVLKEIQN	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFVMFQW IDATLADQFINSDCKPS UDAVLAGKNIEEQKPS UDAVLAGKNIEEQKPS UDAVLAGKNIEEQKPS UDAVLAGKNIEEQKPS UDAVLAGKNIEEQKPS UDAVLAGKNIEEQKPS UDAVLAGKNIEEQKPS UDAVLAGKNIEEQKPS UDAVLAGKNIEEQKPS UDAVLAGKNIEEQKPS UDAVLAGKNIEEQKPS UDAVLAGKPVIDGCKPS UDAVLAGKPVIDGCKPS UDAVLAGKPVIDGCKPS UDAVLAGKPVIDGCKPS UTDKY-VATPADWKDG UTDNYS-VATPADWKDG UTDNYS-VATPADWKDG UTDNYS-VATPADWKDG UTDNYS-VATPADWKDG UTDYQ-VATPANWQDG (VESHPDEVCPAWTPG VQSNPDEVCPAWTPG VQSNPDEVCPAWTPG Q	SKHI GKHI GKUQ 20 IGCN IGCN Frt Wech IGCN Frt Wech CO CO CO CO CO CO CO CO CO CO
002-A0793 142-3126 25-1596 017-3680 1-0162 017-3680 1-0162 017-3680 1-0162 017-3680 1-0162 017-4765 1-3842 413-2915 01-3616 142-4124 20-4642 113-F6083 113-F6083 113-F6083 113-F6498 123-6498 123-0754 03-1198 10331 0352 13-0754 03-1198 10331 0352 13-0754 03-1198 10331 0352 13-0754 03-1198 10331 0352 13-0754 03-1198 10331 0352 13-0754 03-1198 10331 0352 13-0754 03-1198 10331 0352 13-0754 03-1198 10331 0352 13-0754 03-1198 10331 0352 13-0754 03-1198 10331 0352 15-03801 0352	VPSVVFRIKVRDESVEGPNPYRWQDLTSEIF VPSVVFRIKVRDESVEGPNPYRWQDLTSEIF VPDVVFKTRVRDESVEGPNPYRWQDLTTQIF VPDVVFKTRVRDESVEGPNPYRWQDKTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCDETQEVAKYTA GFSFPYCDETQEVAKYTA G	KGKMVUFSLPGAF KGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVESLPGAF GGKPUFSLP 	PTCSSNLPRYELLT PTCSSNLPRYELT PTCSSTLPRYELT STTLPRYELT STTLPRYELT PTCSSTLPRYELT ELHE PTCSSTLPRYELT STTLPRYELT STTLPRYELT STTLP RDRRLVYRGQLDDSRP AHRKLVYRGQLDDSRP AHRKLVYRGQLDDSRP AHRKLVYRGQLDDSRP WDYKLVYRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDDSRP WTGKIQVRGQLDSRP WTGKIQVRHFS STTLP PARKLRLFYP PSGIMMSTINLPVG PDGVIMHSTINNLPVG PDGVIMSTINNLPVG PDGVIMSTINNLPVG PDGVIMSTINCPVG PDGVIMSTINCPVG PDGVIMSTINCPVG PDGVZ	E FQAQGVDQIICISV E IKAQGVDEIICLSV E IKAQGVDEIICLSV 170 16 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW B0 190 II.I.I.I. ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS LDAILADQPINSDQKPS LDAILADQPINSDQKPS LDAILAQPISDQKPS IDVLAGKNIEBEQKPS IDVLAGKPVIDEQKPS IDVLLKIS DQLLLKIS DQLLLKIS DQLLLKIS IEDLLGILDNYIFE IEDLLGILDNYIFE IEDLLGILDNYIFE IEDLLGILDNYIFE UTDNYS-VATPADWKDG LTDNYS-VATPADWKDG LTDNYS-VATPADWKDG LTDYQ-VATPADWKDG LTDYQ-VATPADWKDG LTDYQ-VATPADWKDG LTDYQ-VATPADWKDG LTDYQ-VATPADWKDG LTDYQ-VATPADWKDG VVQSHPDEVCPA\WTPG VVQSHPDEVCPA\WTPG VVQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPDEVCPA\WTPG VQSNPCPA\WTPG VQSNPCPA\WTPG VQSNPCPA\VCPA\WTPG VQSNPCPA\VCPA\WTPG VQSNPCPA\VCPA\WTPG VQSNPCVPA\WTPG VQSNPCPA\VCPA\WTPG VQSNPCVPA\WTPG VQSNPCPA\VCPA\WTPG VQSNPCVPA\WTPG VQSNPCPA\VCPA\WTPG VQSNPCVPA\WTPG VQSN	SKHI GKHI GKUQ 20 20 20 1GCN IGCN IGCN IGCN VGCN
002-A0793 003-1621 142-3126 25-1596 017-3680 1-0162 017-3680 1-0162 017-3680 1-0162 017-3680 1-0162 017-4765 1-3842 413-2915 01-3616 142-4124 20-4642 413-2542 113-6683 1-0234 01-3616 142-4124 20-4404 413-158 113-6498 113-6498 113-0754 103-14621 105-1324 003-14621 115-03801 112-0353 15-03821 15-03821 01-0581 15-03821 01-0583 15-03821 012-0583 15-03821 012-0583 15-03821 012-0583 15-03821 012-0583 15-03821 012-0583 15-03821 012-0583 012-0585 00	VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPDVVFRTEVRDESVEGPNPYRWQDLTTEDIF VPDVVFRTEVRDESVEGPNPFRWQDKTTQEIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA G	KGKMVIVESLEGAF KGKMVIVESLEGAF GGKMVIESLEGAF GGKMVIESLEGAF GGKMVIESLEGAF GGKMVIESLEGAF GGKMVIESLEGAF GGKMVIESLEGAF GGKMVIESLEGAF ACTPDFHEN 	PTCSSNLPRYELLT PTCSSNLPRYELT PTCSSNLPRYELT PTCSSTLPRYELT PTCSTL	E FQAQGVDQIICISV EIKAQGVDEIICLSV EIKAQGVDEIICLSV 170 1E 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS LDAILADQPINSDQKPS LDAILADQPINSDQKPS LDAILADQPINSDQKPS LDAILAGPISANQKPS IDAULAGKPVIDEQKPS FDKLLKFL	CKHI CHI CHI CHI CHI CHI CHI CHI C
002-A0793 003-1621 1142-3126 125-1596 001-2361 c 142-3126 125-1596 001-360 1-0162 01017-3680 1-0162 0-00510 0-00510 0-00510 0-00510 017-4765 11-3842 142-4124 142-4124 142-4124 142-4124 142-4124 142-4124 142-4124 142-4124 142-4124 142-4124 142-4135 113-76498 113-76498 113-76498 113-754 103-14621 105-1324 105-1324 101-03811 112-0381 112-03	VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPDVVFRTEVRDESVEGPNPYRWQDLTTQIF VPDVVFRTEVRDESVEGPNPFRWQDKTTQIF 110 120 130 GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA GFFPYCLDETQAIAKAYTA G	KGKMVUFSLDGAF KGKWVUFSLDGAF GGKWVUFSLDGAF GGKWVLFSLDGAF GGKWVLFSLDGAF GGKWVLFSLDGAF GGKWVLFSLDGAF GGKWVLFSLDGAF GGKWVLFSLDGAF GFKWVFSUF 	PTCSSNLPRYELPI PTCSSNLPRYELPI PTCSSNLPRYELPI STCSTLPRYELPI ELPI STCSTLPRYELPI ELPI STCSTLPRYELPI ELPI STCSTLPRYELPI ELPI STCSTLPRYELPI STCSTLPRYELPI STCSTLPRYELPI STCSTLP STCSTCSTLP STCSTLP STCSTLP STCSTLP STCSTLP	E FQAQGVDQIICISV E FKAQGVDEIICLSV E FKAQGVDEIICLSV 170 IF SDR PVTGADLRAAI SNDR PVTGADLRAAI SNL PVTGADLRAAI SNL PVTGADLRAAI SNL PVTGADLRAAI SNL PVTGADLRAAI SNL PVTGADLRAAI SNL PVTGADLRAAI SNL PVTGADLRAAI SNL PVTGADLRAAI SNL PVTGADLRAAI SNE VTGADLRAAI SNE VTGADLR	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS DATLADQPINSDQKPS DATLADQPINSDQKPS DATLADQPINSDQKPS DATLADQPINSDQKPS DATLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKVIE IDAVLAGKPVIDEQKPS IDAVLAGKAPVIDEQKPS IDAVLAGKAPVIDEQKPS IDAVLAGKAPVIDEQKPS IDAVLAGKAPVIDEQKPS IDAVLAGKAPVIDEQKPS IDAVLAGKAPVIDEQKPS IDAVLAGKAPVIDEQKPS IDAVLAGKAPVIDEQKPS IDAVLAGKAPVIDEQKPS IDAVLAGKAPVIDEQKPS IDAVLAGKAPVID IDAVLAGKAPV	SKHI GKHI GKUQ 20 IGCN IGCN IGCN VGCN
002-A0793 003-1621 1142-3126 125-1596 017-3680 11-0162 017-3680 11-0162 0-200510 0.017-4765 11-3842 0413-2915 101-3616 112-0234 113-F6083 113-F6083 113-F6498 123-F6498 1	VPSVVFRIKVRDESVEGPNPYRWQDLTSEIF VPSVVFRIKVRDESVEGPNPYRWQDLTSEIF VPDVVFKTRVRDESVEGPNPYRWQDLTTQIF VPDVVFKTRVRDESVEGPNPYRWQDKTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCDETQEVAKAYTA G	KGKMVUFSLPGAF KGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF 	PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSNLPRIELTD PTCSSTLPRIELTD STLPRIELTD STLPRIELTD RDRRLVYRGQLDDSRP RDRRLVYRGQLDDSRP SDRQLVYRGQLDDSRP SDRQLVYRGQLDDSRP MTGKIQVKGQLDDSRP WDYKLVYRGQLDDSRP WTYKGQLDDSRP WTGKIQVKHCFS KEGKLSYKHFS KEGKLSYKHFS KEGKLSYKHFS KEGKLSYKHFS PHKKLRLFTYPSTG PHKKLRLSFYPSTG PHKKLRLSFYPSTG PHKKLRLFTYPSTG PHKKLRLSFYPSTG PHKKLRLFTYPSTG PHKKLRLFTYPSTG PHKKLRLFTYPSTG PHKKLRLLFYPSTG PHKKLRLFTYPSTG PHKKLRLLFYPSTG PHKKLRLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKKLRLLFYPSTG PHKLRLSFYPSTG PHKLRLSFYPSTG PHKLRLSFYPSTG PHKLRLSFYPSTG PHKLRLSFYPSTG PHKLRLSFYPSTG PHKLRLSFYPSTG PHKLRLFYPSTG PHKLRLSFYPSTG PHKLRLSFYPSTG PHKLRFY PHKLRSFYPSTG PHKLRSFYPSTG PHKLRFYPSTG PHKLRSFYPSTG PHKLRFYPSTG PHKLRSFYPSTG PHKLRFYPSTG PHKLRFYPSTG PHKLRFYPSTG PHKLRFYPSTG PHKLRFYPSTG PHKLRFYPSTG PHKLRFYPSTG PHKLRFYPSTG PHKLRFYPSTG PHKLRFYPSTG PHKLFYPSTG P	E FQAQGVDQIICISV E IKAQGVDEIICLSV E FKGQGIDEIICLSV 170 16 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW 30 190 II.I.I.I. ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS LDAILADQPINSDQKPS LDAILADQPINSDQKPS LDALLAGPISANQKPS IDVLAGKNIEBEQKPS' FDKLLKFL	SKHI GKHI GKUQ 20 20 20 1GCN IGCN IGCN VGCN
002-A0793 003-1621 1142-3126 125-1596 001-2361 c 142-1730 0017-3680 01-0162 01017-3680 01-0162 011-3842 1424-1730 001-3616 1142-4124 120-4642 0133-F6683 1132-6498 1132-64	VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPDVVFRTEVRDESVEGPNPYRWQDLTTEDIF VPDVVFRTEVRDESVEGPNPFRWQDKTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCDETQEVAQAYTA GFSFPYCDETQEVAQAYTA G	KGKMVUFSLBGAF KGKWUVFSLBGAF GGKWVUFSLBGAF GGKWVLFSLBGAF GGKRVUFSLBGAF GGKRVUFSLBGAF GGKRVUFSLBGAF GGKRVUFSLBGAF GGKRUFSSC GFKUVFSSCH GGKPUFSSCNTFLS WIPPSSCNTFLS WIPPSSCNTFLS WIPPSSCNTFLS	PTCSSNLPRIEDID PTCSSNLPRIEDID PTCSSNLPRIENTEDID PTCSSTLPRIENTEDID PTCSSTLPRIEDID PTCSSTLPRIENTEDID PTCSSTLPRIENTEDID PTCSSTLPRIENTEDID PTCSSTLPRIEND	E FØAQGVDQI ICISV E IKAQGVDEI ICLSV E FKGQGIDEI ICLSV 170 IF SDR PVTGADLRAAI SNDR PVTGADLRAAI SNDR PVTGADLRAAI SNDR PVTGADLRAAI SNUR PVTGADLRAAI RNEDELRVIDSLOI RNEDELRV	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW ITAVLTGDAPSADQKPS TTAVLTGDAPSADQKPS LDALABQPISADQKPS LDALABQPISADQKPS LDALABQPISADQKPS LDALAGKVIDEQKPS FDKLLKFL	SKHI GKHI GKUQ 20 IGCN IGCN VGCN
002-A0793 003-1621 1142-3126 125-1596 01-0162 01-0162 01-0162 01-0162 01-0162 0-000510 0-000510 0-000510 0-000510 017-4765 113-3842 011-3842 011-3842 011-2344 017-2532 120-4642 013-F6083 113-66083 113-66083 113-66083 113-66083 113-66083 113-66083 113-66083 113-66083 113-66083 113-66083 113-66083 113-66083 113-66083 113-6754 113-6754 113-0754 113-0754 113-0353 113-0353 113-0353 112-0353 112-0353 112-0353 112-0353 112-0353 112-0353 112-0353 112-03551 112-03551 113-09851 115-	VPSVVFRIGVRDSVESVEGPNPYRWQDLTSEIF VPSVVFRIGVRDSVEGPNPYRWQDLTSEIF VPDVVFRTRVRDSVEGPNPYRWQDLTTQIF UPDVVFRTRVRDSVEGPNPFRWQDKTTQIF 110 120 130 GFFFPYCLDETQAIAKAYTA GFFFPYCLDETQAIAKAYTA GFFFPYCLDETQAIAKAYTA GFFFPYCLDETQAIAKAYTA GFFFPYCLDETQAIAKAYTA GFFFPYCLDETQAIAKAYTA G	KGKMVUFSLDGAF KGKWUVFSLDGAF GGKWVUESLDGAF GGKWVLFSLDGAF GGKWVLFSLDGAF GGKWVLFSLDGAF GGKWVLFSLDGAF GGKWVLFSLDGAF GGKWVLFSLDGAF GFKWVESUFS 	PTCSSNLPRIEDID PTCSSNLPRIEDID PTCSSNLPRIEDID PTCSSTLPRIEDID ELLEN E	E FQAQGVDQIICISV E FKAQGVDEIICLSV E FKAQGVDEIICLSV 170 IE SSDR PVTGADLRAAI SSDR PVTGADLRAAI SNLKPVTGADLRAAI SNLKPVTGADLRAAI SNLKPVTGADLRAAI SNLKPVTGADLRAAI SNLKPVTGADLRAAI SNLKPVTGADLRAAI SNLKPVTGADLRAAI SNEKTGAD SNEKTGAD	ANDAFVMFQW ANDAFVMFQW ANDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFV-DQFINSDQKPS LDAVLAGKDQFINSDQKPS LDAVLAGKNIEEQCKPS LDAVLAGKNIEEQCKPS IDAVLAGKPVIDSCKPS LDAVLAGKPVIDSCKPS IDAVLAGKPS IDAVLAGKPVIDSCKPS IDAVLAGKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGKPVIDSCKPS IDAVLAGK	SKHI GKHI GKUQ 20 10 10 10 10 10 10 10 10 10 1
002-A0793 003-1621 1142-3126 125-1596 001-2361 c 142-1730 01017-3680 01-0162 01017-3680 01-0162 01017-4765 11-3842 142-4765 11-3842 142-4765 11-3842	VPSVVFRIEVROESVEGPNPYRWQDLTSEIF VPSVVFRIEVROESVEGPNPYRWQDLTSEIF VPDVVFKTRVRDESVEGPNPYRWQDLTTQIF VPDVVFKTRVRDESVEGPNPYRWQDLTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCDETQEVAKAYTA GFSFPYCDETQEVAKAYTA G	KGKMVUFSLPGAF KGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF 	PTCSSNLPRYELLPS PTCSSNLPRYELPS PTCSSTLPRYELPS PTCSSTLPRYELPS ELLPS STLPRYELPS ELLPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTILPS INTERSTINE INTERSTI	E FQAQGVDQIICISV E IKAQGVDEIICLSV E IKAQGVDEIICLSV 170 16 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFV	CREEV SAVE
002-A0793 002-A0793 003-1621 1142-3126 125-1596 001-0162 0017-3680 01-0162 0017-3680 01-0162 0017-3680 01-0162 0017-4765 11-3842 0442-1730 0-00510 0017-4765 11-3842 0413-2915 001-3616 1142-4124 120-4042 0133-F6083 11-0234 011-0234 011-0234 011-0234 011-0234 011-0234 011-0234 011-0234 011-0234 011-0234 011-0234 011-0234 011-0234 011-0234 011-0234 011-0234 0133-F6498 0133-F6498 0133-F6498 0133-F6498 0133-F6498 0133-F6498 0133-F6498 0133-F6498 0133-F6498 0133-1621 015-03801 015-03801 015-03801 012-0353 015-03851 012-0353 012-0353 012-0351 012-035	VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPDVVFRTEVRDESVEGPNPYRWQDLTTEDIF VPDVVFRTEVRDESVEGPNPFRWQDKTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA G	KGKMVUFSLBGAF KGKWUVFSLBGAF GGKWVUFSLBGAF GGKWVLFSLBGAF GGKWVLFSLBGAF GGKWVLFSLBGAF GGKWVLFSLBGAF GGKWVLFSLBGAF GGKWVLFSLBGAF GGKWVLFSLBGAF 	PTCSSNLPR/BLID PTCSSNLPR/BLID PTCSSNLPR/BLIP PTCSSTLPR/BLIP ETCSSTLPR/BLIP	E FQAQGVDQIICISV E IKAQGVDEIICLSV E FKGQGIDEIICLSV 170 If SSDR PVTGADLRAAI SSDR PVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI RNEDELRVIDSLOI RNEDEL	ANDAFVMFQW ANDAFVMFQW ANDAFVMFQW ANDAFVMFQW INDAFVMFQW ISON ISON ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS IDAVLAGKNIEEEQKPS IDAVLAGKNIEEEQKPS IDAVLAGKNIEEEQKPS IDAVLAGKNIEEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKNIEEQKPS IDAVLAGKIEQTS I	SKHI GKHI GKUQ 20 IGCN IGCN VGCN
002-A0793 803-1621 1142-3126 125-1596 001-0162 9-000510 1017-3680 01-0162 9-000510 1017-4765 101-3842 9413-2915 301-3616 1142-4124 120-4642 9133-F6083 101-0234 1017-2532 120-4404 9133-F6083 9133-F6083 9133-F6083 9133-F6083 9133-F6083 9133-F6083 9133-F6083 9133-F6083 9133-F6083 9133-F6083 1142-0135 803-1198 1142-0135 803-1198 1142-0135 803-1198 115-03801 125-03801 125-03801 125-03801 125-03801 125-03801 125-03801 125-03801 125-03801 125-03801 125-03801 125-03801 125-03801 125-03801 125-03801 125-03851 120-0351 1312-0353 1312-0355 1312-	VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPDVVFRTEVRDESVEGPNPYRWQDLTTQIF UPDVVFRTEVRDESVEGPNPYRWQDKTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA G	KGKMVUFSLBGAF KGKWUVFSLBGAF GGKWVUFSLBGAF GGKWVUFSLBGAF GGKWULFALBGAF GGKWULFALBGAF GGKWULFALBGAF GGKWULFALBGAF GGKWULFALBGAF GGKPULFALBGAF 	PTCSSNLPRYELPI PTCSSNLPRYELPI PTCSSNLPRYELPI STCSTLPRYELPI ELPI STCSTLPRYELPI ELPI STCSTLPRYELPI ELPI STCSTLPRYELPI ELPI STCSTLPRYELPI RORRLVYRGQLDDSRP RORRLVYRGQLDDSRP WDYKLVYRGQLDDSRP WDYKLVYRGQLDDSRP WDYKLVYRGQLDDSRP WDYKLVYRGQLDDSRP WDYKLFS SCDDRP WTGKIQYKHFS STCST WDGLPKHFS STCST ST	E FQAQGVDQIICISV E FKAQGVDEIICLSV E FKAQGVDEIICLSV 170 14 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW 190 ITAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS IDAVLAGKNIEEQKPS LDALLADQPINSDQKPS LDALAEQPISANQKPS LDALLACQPISDQKPS LDALLAKFD	SKHI GKHI GKUQ 20 20 20 1GCN IGCN IGCN IGCN IGCN VGCN
002-A0793 803-1621 1142-3126 425-1596 1017-3680 01-0162 942-1730 9-00510 1017-4765 01-3842 9433-2915 801-3616 1142-4124 120-4642 9413-2915 801-3616 1142-4124 120-4642 9133-F6083 01-0234 1107-2532 1120-4404 9413-1358 9133-F6083 01-0234 1102-4649 803-1198 A-10331 A-0352 313-0754 303-14621 605-1324 102-1213 301-03801 215-03801 215-03801 601-09581 312-0353 515-09851 603-09581 515-09851 603-032 1142-3126 425-1596	VPSVVFRIEVROESVEGPNPYRWQDLTSEIF VPSVVFRIEVROESVEGPNPYRWQDLTSEIF VPDVVFKTRVRDESVEGPNPYRWQDLTTQIF VPDVVFKTRVRDESVEGPNPYRWQDLTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCDETQEVAKAYTA G	KGKMVUFSLPGAF KGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVUFSLPGAF GGKWVIFSLPGAF GGKWVIFSLPGAF 	PTCSSNILFRIELTD PTCSSNILFRIELTD PTCSSNILFRIELTD PTCSSTLPRIELTD STCSTHLPRIELTD STCSTHLPRIELTD PTCSSTLPRIELTD RDRLVYRGQLDDSRP RDRRLVYRGQLDDSRP RDRLVYRGQLDDSRP RDRQLVYRGQLDDSRP RDRQLVYRGQLDDSRP WDYKLVYRGQLDDSRP WTYGLQYRGQLDDSRP WTYGLQYRGQLDDSRP WTYGLQYRGQLDDSRP WTYGLQYRGQLDDSRP WTYGLQYRGQLDSRP WTYGLQYRGQLDSRP WTYGLQYRGQLDSRP WTYGLQYRGQLDSRP WTYGLQYRGQLDSRP WTYGLQYRGQLDSRP WTYGLQYRGQLDSRP WTYGLQYRGQLDSRP WTYGLYRGQLDSRP WTYGLYRGQLDSRP WTYGLYRGQLDSRP WTYGQLMYRGQLDSRP WTYGQLPYRGQLDSRP WTYGQLPYRGQLDSRP WTYGQLYRGQLDSRP WTYGQLYRHFS PTSSI PTSSI PTSSI WTYGQL WTYGQL WTYGQLDSRP WTYGQL WTYG	E FQAQGVDQIICISV E IKAQGVDEIICLSV E IKAQGVDEIICLSV 170 16 	NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW NIDAFVMFQW INDAFVMFQW INDAFVMFQW ISTAVLTGDAPSADQKPS ITAVLTGDAPSADQKPS IDAVLAGKPVIDEQKPS' IDAVLAGKPVIDEQKPS' IDAVLAGKPVIDEQKPS' IDAVLAGKPVIDEQKPS' IDAVLAGKPVIDEQKPS' IDAVLAGKFVIDEQKFVIDEQKPS' IDAVLAGKFVIDEQKFVIDEQKFVIDEQKFVIDEQKFVIDEQKFVIDEQKFVIDEQKFVIDEQ	CREEV CRAFY
002-A0793 803-1621 1142-3126 425-1596 1017-3680 01-0162 942-1730 9-Q00510 1017-4765 01-3842 9413-2915 801-3616 1142-4124 120-4642 9133-F6083 01-0234 1017-2532 120-4642 9133-F6083 01-0234 1017-2532 120-4642 9133-F6083 01-0234 1017-2532 120-4642 9133-F6083 01-0352 313-754 9133-F6498 9133-F6498 9133-F6498 9133-F6498 9133-F6498 9133-F63801 215-03801 215-03801 215-03891 501-0351 312-0353 515-03821 912-0897 301-09561 515-09851 312-0897 301-09561 515-09851 302-A0793 303-1621 1142-3126 125-1596 101-0360	VPSVVFRIEVRDESVEGPNPYRWQDLTSEIF VPSVVFRIEVRDESVEGPNPYRWQDLTTEDIF VPDVVFRTEVRDESVEGPNPYRWQDLTTEDIF VPDVVFRTEVRDESVEGPNPFRWQDKTTQIF 110 120 130 GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA GFSFPYCLDETQAIAKAYTA G	KGKMV UVESLEGAF KGKW UVESLEGAF GGKRVULESLEGAF GGKRVULESLEGAF GGKRVULESLEGAF GGKRVULESLEGAF GGKRVULESLEGAF GGKRULEALEGAF GGKRULEALEGAF GGKRULEALEGAF GGKRULEALEGAF 	PTCSSNLPR/EBLID PTCSSNLPR/EBLID PTCSSNLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBLID PTCSSTLPR/EBSID PTCSSTLPS/EBSID PTCSST	E FQAQGVDQIICISV E IKAQGVDEIICLSV E FKGQGIDEIICLSV 170 If SSDR PVTGADLRAAI SSDR PVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI SNULPVTGADLRAAI RNEDELRVIDSLOI RNEDELRVISSON RNEDEL	ANDAFVMFQW ANDAFVMFQW ANDAFVMFQW ANDAFVMFQW INDAFVMFQW INDAFVMFQW INDAFV	SKHI GKHI GKUQ 20 IGCN IGCN V

(See figure on previous page.)

Figure 5 Conserved cysteine-including domains and arginine in cyanobacterial PRX from different subfamilies. A total of 42 sequences were used in the alignment of cyanobacterial PRX. Six samples were extracted to represent their own subfamilies. Shades with different colours indicate the distinct PRX subfamilies. Typical catalytic traits of PRX are indicated by the red box. The conserved arginines are indicated by yellow shading.

(for a review, see [14,42]). These include enzymatic methods (superoxide dismutases, peroxidases and catalases) that can be used to sequentially detoxify superoxide and hydrogen peroxide [43], and non-enzymatic mechanisms (glutathione, vitamin A, C, E, carotenoids, etc.) [14]. Peroxiredoxins (PRXs) are an important type of antioxidant proteins that are also known as the thioredoxin peroxidases or alkyl-hydroperoxide-reductase-C22 proteins [44,45]. PRXs have been identified from plants [27] and have received considerable attention in recent years. PRXs exert their protective antioxidant role in host cells through their peroxidase activity, which leads to the reduction and detoxification of hydrogen peroxide, peroxynitrite and a wide range of organic hydroperoxides (ROOH) [46-48]. The catalytic efficiency (~ $10^5 \text{ M}^{-1} \text{ s}^{-1}$) of PRXs is lower than that of better known glutathione peroxidases (~ $10^8 \text{ M}^{-1} \text{ s}^{-1}$) [46] and catalases (~ $10^6 \text{ M}^{-1} \text{ s}^{-1}$) [49], which makes their importance as other peroxidases questionable.

What makes PRXs so important and interesting in cyanobacteria? The multi-isoforms and the high abundance of PRXs in a wide range of cells may be the first reason [41,50,51]. Additionally, a recent study revealed that a bacterial PRX (alkyl hydroperoxide reductase C22 (AhpC)), rather than catalase, is responsible for the reduction of endogenously generated hydrogen peroxide

[52]. Finally, based on the evaluation of 37 cyanobacterial genomes in this study, it could be found that all *Prochlorococcus marinus* strains and most of the other cyanobacteria do not possess gene(s) with homology to catalase, but possess several genes with homology to PRXs (according to our unpublished results and [53]). Taken together, these characteristics indicate that PRX

Peroxiredoxins (PRXs): a structural conserved enzyme

ide in cyanobacterial and other living cells.

may actually be important to the detoxification of perox-

Six different sub-classes of PRXs, PRX I–IV (2-Cys PRX), PRX V (Type II PRX) and PRX VI (1-Cys PRX), have been identified from mammalian systems [32]. Among these, only four have been reported in plant systems, namely, 1-Cys PRX, 2-Cys PRX, Type II PRX and PRX Q [29]. According to our results, cyanobacterial PRXs were classified into five major subfamilies (1-Cys, 2-Cys, BCP, PRX5_like, and PRX-like) according to their domain structures. Based on the crystal structures of six PRXs that has been published to date, including four typ-ical 2-Cys PRXs (PRXI, PRXII, TryP and AhpC [54-56], one atypical 2-Cys PRX (PRXV [57]) and one 1-Cys PRX (PRXVI [58]). All PRXs share a similar structure, with each containing a thioredoxin fold and a few additional secondary-structural elements present as insertions. In

Subfamily	Cys-including motif 1	Cys-including motif 2	Position (Arg)
Prx 1-Cys	WAGDSWWLFSHPADYTPV <u>C</u> TTELG		5 (62, 126, 138,
	(26–50)		149 and 156)
Prx 2-Cys	VVLFFYPLDFTFV <u>C</u> PTEVIAFSD	DEVCPA	7 (9, 46, 74, 104,
	(51–73)	(183–188)	140, 163 and 170)
Prx BCP-A		C	2 (64 and 132)
	(44–57)	61	
Prx BCP-B	WWVLYFYPQDFTPG <u>C</u> TLEA	<u>C</u>	1 (160)
	(75–93)	125	
Prx_like1	LLVMFI <u>C</u> QH <u>C</u> PFVK	AACTPDF	3 (152, 164
	(52–65)	(134–140)	and 178)
Prx_like2	IFTEXXYCPFCXPH		4 (12, 47, 63
	(64–77)		and 130)
Prx5_like	WLXXLPGAFTPT <u>C</u> SS		3 (18, 138
	(46–61)		and 141)

Table 2 Conserved cysteine-including motifs and arginines of PRXs in cyanobacteria

The typical catalytic domains and CXXC motif of PRXs was represented by red and blue colour, respectively.

addition, the structure and sequences of the peroxidatic active site are highly conserved in the protein sequences from all the PRX subfamilies [41]. According to previous study [41], the peroxidatic cysteine in the reduced (SH) form is in a narrow, solvent-accessible pocket formed by a loop-helix structural motif. The cysteine is located in the first turn of the helix and is surrounded by three residues conserved among all classes-Pro44, Thr48 and Arg127 (PRX II numbering) [41]. Our results indicated that the typical catalytic triad of PRXs is found in the N-terminal of those proteins from the 1-Cys, 2-Cys, PRX BCP, and PRX5_like subfamilies (Figure 5 and Table 2). The resolving cysteine near the C-terminal was detected in the proteins from the 2-Cys PRXs subfamily. It is interesting to note that another cysteine was identified in the C-terminal of PRX BCP. This result is not consistent with the results of previous studies, which showed PRX BCP contains the peroxidatic cysteine but without a resolving cysteine [41,59]. However, the role of the second cysteine is still unknown. Members of PRX-like (1 and 2) contain a CXXC motif near the N-terminal that is similar to the classic redox active CXXC motif of Trx [60]. Schultz et al. (1999) claimed that the second cysteine in this motif corresponds to the peroxidatic cysteine of PRXs. However, these proteins do not contain the other two residues of the catalytic triad of PRXs [61]. All PRXs share a highly conserved active-site arginine, which would lower the pKa of the peroxidatic cysteine somewhat by stabilizing its thiolate form (see review [41]). As expected, at least one conserved active-site arginine was detected in all cyanobacterial PRXs (Figure 5 and Table 2). Therefore, we speculated that the mechanisms of PRXs of 1-Cys, 2-Cys, PRX BCP, and PRX5_like are similar [41,62], whereas the mechanisms of the PRX_like subfamily are different. According to the definition of the Thioredoxin_like Superfamily [CDD: cl00388], we inferred that PRX_like members do not function as protein disulfide oxidoreductases, even though they containing a Trx-fold domain. However, the catalytic triad of PRXs was discovered in C-terminal sequences from the PRX-like1 subfamily, which exceeded our expectations. Additional experimental results are needed to determine whether this predicted catalytic triad of PRXs is active in PRX-like1. However, such an analysis is beyond the scope of this paper.

The distribution of PRXs is related to genome sizes and habitat niches

Although the number of *prx* genes and their transcriptional regulation under stress in some cyanobacteria have been reported in previous studies, modification and supplementation is needed with the complete and partial sequencing genomes of several cyanobacterial species. Five genes encoding peroxiredoxin 2-Cys PRX (*sll0755*),

1-Cys PRX (sll1198), two PRX Q (sll0221 and slr0242) and one Type II PRX (sll1621) were reported in Synechocystis sp. PCC 6803 [28,33,34], whereas another gene (ID: sll1159, annotation: probable BCP) was detected and classified into the PRX-like2 subfamily. Analysis of the genome of Synechococcus elongatus PCC 7942 led to the identification of six putative prx genes [34] with one 1-Cys PRX, one 2-Cys PRX and four PRX Q, while a gene (ID: 7942_1730, annotation: hypothetical protein) was found and classified into the PRX-like1 subfamily. The computational method and the quality of the genome data may be responsible for these different results. Moreover, multi-isoforms (3-12) of genes encoding PRXs were present in all cyanobacteria investigated in the present study. However, the reason for the existence of multiple prx genes in these cyanobacteria is still unclear [33,34].

The distribution of putative PRX encoding open reading frames (ORFs) from some sub-families like PRX-like or 1-Cys PRX in different cyanobacteria correlate with the genome size, eco-physiology, and physiological properties of the organisms. Although the number (8-11) of prx genes in filamentous cyanobacteria (with large genome size) is higher than those (3-6) from marine unicellular cyanobacteria (with small genome size), the percentage (0.16-0.18%) of PRXs among the total proteins from the former is lower than the latter (0.20-0.30%). Moreover, most of the cyanobacteria possess disproportionate numbers of putative prx genes with different genome sizes, indicating that not a basic set is amplified to achieve a larger genome, but that additional functions may be encoded by larger genomes. This result is not consistent with the previous studies who found that not only the number of Serine/threonine kinases and metacaspase genes in filamentous cyanobacteria is higher than those from marine unicellular cyanobacteria, but also the percentage of Serine/threonine kinases and metacaspase genes in the total proteins is higher [7,12]. The reason for this phenomenon may be that PRXs are not the only protein to protect against ROS. For example, other proteins such as catalase, SOD and ferredoxin have been detected in cyanobacteria and the number of genes encoding SODs in filamentous cyanobacteria (with large genome size) is much higher than other cyanobacteria with small genome size [22,26,63]. However, two unicellular cyanobacterial strains inhabiting freshwater (Synechococcus elongatus PCC 7942 and Synechococcus elongatus PCC 6301) and three unicellular cyanobacterial strains living in hot-springs (Thermosynechococcus elongatus BP-1, Synechococcus sp. JA-2-3B'a (2-13), and Synechococcus sp. JA-3-3Ab) maintain more prx genes (7-8) than unicellular cyanobacteria from marine. Considering that unicellular cyanobacterial strains from different habitats share similar genome sizes, various environmental selective pressures may be responsible for the number of *prx* genes in these organisms. The distribution of a small numbers of *prx* genes in cyanobacteria from the ocean is consistent with Serine/threonine kinases and metacaspase genes in cyanobacteria, which are remarkably reduced in marine species [7,12]. Gene loss has been shown to facilitate the acclimatization of these cyanobacteria to the oligotrophic environment of the sea. The major force driving this phenomenon was reportedly a selective process favoring the adaptation of these cyanobacteria, which has been discussed in detail by Alexis Dufresne et al. [64].

The evolution of PRXs

The protein sequences from the 1-Cys PRX subfamily contains a single conserved catalytic cysteine and is thus denoted 1-Cys PRX [65-67]. Our results revealed that the 1-Cys PRX subfamily was absent from all marine unicellular cyanobacteria except for Gloeobacter violaceus PCC 7421 and Synechococcus PCC 7002. The phylogenic relationship among 1-Cys PRXs from cyanobacteria, higher plants, and Metazoa strongly supports a non-cyanobacterial origin of these proteins in higher plants, indicating that genes encoding 1-Cys PRX are not unique for cyanobacteria and the higher plants do not acquire this gene by endosymbiosis event. Immunochemical study revealed that the 1-Cys PRXs from higher plants are preferentially localized in the nucleus and within the nucleolus [17,65,68]. In addition, the 1-Cys PRXs have been widely recorded in mammalian systems [69]. The 2-Cys PRXs (classical or typical) functioned as a homodimer in a head-to-tail arrangement in which the sulfenic acid derivative of the peroxidatic cysteine of one subunit interacts with the resolving cysteine of the other subunit during the catalytic cycle [70,71]. The 2-Cys PRX subfamily includes chloroplastic 2-Cys PRX, mammalian PRX I-IV and yeast thiol-specific antioxidant (TSA) [17]. Meanwhile, this subfamily is highly conserved among all cyanobacteria. The phylogenetic tree for 2-Cys PRXs revealed that cyanobacteria and higher plants share a common ancestor, which is consistent with the previous studies [70] and the sub-cellular localization (chloroplast) of this protein in A. thaliana [17]. PRX BCP subfamily constitutes the largest group of prx in cyanobacteria. The prxq genes cloned from higher plants are homologous to the bacterioferritin comigratory protein (BCP) from Escherichia coli [72] and cluster into the cyanobacterial PRX BCP group. Thus many prx genes were originally annotated BCP (PRX Q) in cyanobacteria. PRX Q is the only one that has not been isolated from an animal system [72]. Type II PRXs (A/B/C/D/E/F) from higher plants build a monophyletic group with members from PRX5_like as a sister group, implying that the higher plants acquire this gene via photoautotrophic endosymbiosis. In addition to the above subfamilies, a novel subfamily (PRX-like1 and PRX-like2) was firstly identified from cyanobacteria in this study. Most members of this subfamily are noted as hypothetical proteins that show sequence similarity with PRXs. The structure and mechanism of members of this subfamily are currently unclear.

Conclusions

Comparative analysis based on the availability of cyanobacterial genome sequences becomes a powerful tool for systematic studies of gene families. Peroxiredoxins comprise one of the most important proteins that play key roles in protecting own cells from the damage of ROS. In this study, 244 putative prx genes were identified from 37 species of cyanobacteria using BLASTp, tBLASTn, HMMsearch and SMART domains analysis. Among these putative PRXs, 25 prx genes originally annotated as hypothetical proteins were accepted as PRXs firstly in this study. The quantity of prx genes in unicellular and filamentous cyanobacteria depends on the genome size, eco-physiology, and ecological habitats. According to the results of CDD domain and phylogenetic analysis, the 244 PRXs were divided into five major groups (1-Cys, 2-Cys, PRX BCP, PRX5_like, and PRXlike). The 2-Cys, PRX BCP, and PRX-like subfamilies are conserved and widely distributed among cyanobacteria. However, PRXs from other subfamilies have only been detected in a few cyanobacterial strains, indicating that they are species or habitat-specific. The typical catalytic trait of PRXs was identified in all PRXs except those from the PRX-like2 subfamily. The proteins from the PRX-like2 subfamily share the classical redox active CXXC motif of thioredoxin. Phylogenetic trees based on the catalytic domains of PRXs from each subfamily coincide well with the phylogenies based on the16s rRNA.

Methods

Identification of *prx* genes encoding PRX proteins

A total of 37 species of cyanobacteria, including *Prochlorococcus, Synechococcus, Synechocystis, Gloeobacter, Cyanothece, Microcystis, Trichodesmium, Acaryochloris, Anabaena* and *Nostoc* were used in this analysis. These cyanobacterial genomes were downloaded from the JGI genome portal [39] or Cyanobase [40]. Ten photosynthetic eukaryotic PRX proteins from *Arabidopsis thaliana* and six eukaryotic PRX proteins from *Homo sapiens* were also downloaded from NCBI Genbank [73].

To identify genes encoding peroxiredoxins, eleven previously characterized PRXs from freshwater cyanobacteria *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942 [34] and ten PRXs from *Arabidopsis thaliana* were used to construct a query protein set. BLASTp [74-76] and tBLASTn [77] programs were conducted locally to search for all prx genes from all 37 cyanobacterial genomes using a threshold e-value of 1e-10. Briefly, the prx genes encoding PRX proteins used in this study were first identified by local BLASTp and tBLASTn program rather than from the COG database in IMG. Following, we manually checked the extracted proteins by SMART and Pfam analyses to avoid falsepositive hits that commonly arise during large-scale automated analyses. PRXs found by this method were added to the query set for another round of BLASTp searches. This procedure was continued until no new proteins were found. Moreover, in order to check for false negatives, two hmm models [Pfam: PF00578] and [Pfam: PF08534] derived from the known PRX proteins were applied to search for genes encoding PRX on all proteins encoded in the 37 cyanobacterial genomes [78,79]. All translated protein sequences of genes encoding PRXs used in this paper were listed in more detail [see Additional file 7].

Multiple sequence alignment and structure analysis

Proteins identified by the BLAST searches were aligned using ClustalW [80,81] with a gap opening penalty of 10, a gap extension penalty of 0.2, and Gonnet as the weight matrix. The SMART [82] and Pfam 26.0 [78] databases were applied to delete false positives. The alignment was then examined by inspection of the PRX_1cys, PRX_Typ2cys, PRX_BCP, PRX5_like, PRX_like1, and PRX like2 domains [CDD: cd03016, cd03015, cd03017, cd03013, cd02969, and cd02970] in the NCBI Conserved Domain Database [83]. A protein was accepted as PRX if it was possible to recognize any domain above or known to participate in the function of PRXs. Structural analysis of the obtained PRXs was performed using the SMART (Simple Modular Architecture Research Tool) [82] and the CDD (Conserved Domains Database) [83], methods, relying on hidden Markov models and Reverse Position-Specific BLAST, separately.

Phylogenetic analysis

Maximum likelihood trees of 16s *rRNA* and PRX proteins were constructed using PhyML [84]. For the 16S *rRNA* tree, the General Time Reversible (GTR) substitution model was selected to assume an estimated proportion of invariant sites and four gamma-distributed rate categories to account for rate heterogeneity across sites [85]. The reliability of internal branches was assessed using the bootstrapping method (400 bootstrap replicates). The Le and Gascuel evolutionary model [86] was selected for analysis of the protein phylogenies assuming an estimated proportion of invariant sites and a gamma correction (four categories). Bootstrap values (BS) were inferred from 400 replicates. Graphical representation and edition of the phylogenetic tree were performed with TreeDyn (v198.3) [87].

Statistical analyses

Statistical analyses on the relationship between the distribution of genes encoding PRXs and properties of 37 cyanobacterial organisms were performed using the Spearman Rank Correlation test (R). For all of the data analyses, a *p*-value <0.01 was considered statistically significant [88].

Additional files

Additional file 1: Table S1. Cyanobacterial genes encoding peroxiredoxin were predicted by BLAST program (BLASTp and tBLASTn). Note: "-" stands for those do not belong to PRX family. The "complete" and "in complete" present for the complete or partial of the genomes (data collected at 1 Jan, 2012). Table S2 Cyanobacterial genes encoding peroxiredoxin were predicted by hiddenMarkov model (hmmsearch). Note: "*" stands for additional genes encoding PRXs predicted basedon hmmsearch analysis. The "complete" and "in complete" present for the complete or partial of thegenomes (data collected at 1 Jan, 2012).

Additional file 2: Figure S1. Maximum likelihood tree of PRXs from Prx-BCP subfamily.

Additional file 3: Figure S2. Maximum likelihood tree of PRXs from 2-Cys subfamily.

Additional file 4: Figure S3. Maximum likelihood tree of PRXs from Prx-like subfamily.

Additional file 5: Figure S4. Maximum likelihood tree of PRXs from Prx5_like subfamily.

Additional file 6: Figure S5. Maximum likelihood tree of PRXs from 1-Cys subfamily.

Additional file 7: The translated protein sequences of 244 putative PRXs in 37 cyanobacteria.

Abbreviations

PRX: Peroxiredoxin; Cys: Cysteine; 1-Cys PRX: 1-cysteine peroxiredoxin; 2-Cys PRX: 2-cysteine peroxiredoxin; PRX Q: Peroxiredoxin Q; PRX II: peroxiredoxin type II; ROS: Reactive oxygen species; Trx: Thioredoxin; AhpC/TSA: Alkyl hydroperoxide reductase/thiol specific antioxid; BCP: Bacterioferritin comigratory protein; RDP: Redoxin domain protein; TSAP: Thiol specific antioxidant protein; HP: Hypothetical protein.

Competing interests

The authors have declared that no competing interests exist.

Authors' contributions

H C conceived of the study, participated in the sequence alignment and phylogeny analysis and drafted the manuscript. Y W participated in the domain analysis and statistical analyses. Y W and S Q participated in drafting the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (40876082); International Innovation Partnership Program: Typical Environmental Process and Effects on Resources in Coastal Zone Area; Public Science and Technology Research Funds Projects of the Ocean (200905021–3); and Outstanding Young Scholars Fellowship of Shandong Province (Molecular Phycology JQ200914).

Author details

¹Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Chunhui Road, Yantai 264003, People's Republic of China. ²University of Chinese Academy of Sciences, Yuquan Road, Beijing 100049, People's Republic of China.

Received: 3 March 2012 Accepted: 25 October 2012 Published: 16 November 2012

References

- Stanier RY, Cohen-Bazire G: Phototrophic prokaryotes: the cyanobacteria. Annu Rev Microbiol 1977, 31:225–274.
- Blankenship RE, Hartman H: The origin and evolution of oxygenic photosynthesis. Trends Biochem Sci 1998, 23(3):94–97.
- Whitton BA, Potts M: In The ecology of cyanobacteria: their diversity in time and space. Edited by Whitton BA, Potts M.; 2000:1–11.
- Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA, Arellano A, Coleman M, Hauser L, Hess WR, et al: Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 2003, 424(6952):1042–1047.
- Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, Barbe V, Duprat S, Galperin MY, Koonin EV, Le Gall F, *et al*: Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci U S A* 2003, 100(17):10020–10025.
- Meeks JC, Elhai J, Thiel T, Potts M, Larimer F, Lamerdin J, Predki P, Atlas R: An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. *Photosynth Res* 2001, 70(1):85–106.
- Jiang Q, Qin S, Wu QY: Genome-wide comparative analysis of metacaspases in unicellular and filamentous cyanobacteria. BMC Genomics 2010, 11:198.
- Chi XY, Yang QL, Zhao FQ, Qin S, Yang Y, Shen JJ, Lin HZ: Comparative analysis of fatty acid desaturases in cyanobacterial genomes. *Comp Funct Genomics* 2008, 16:1099–1108.
- Scanlan D: Cyanobacteria: ecology, niche adaptation and genomics. Microbiol Today 2001, 28:128–130.
- Partensky F, Hess WR, Vaulot D: Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* 1999, 63(1):106–127.
- Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, *et al*: Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. DNA Res 2003, 10(4):137–145.
- Zhang XW, Zhao FQ, Guan XG, Yang Y, Liang CW, Qin S: Genome-wide survey of putative serine/threonine protein kinases in cyanobacteria. BMC Genomics 2007, 8:395.
- Wang H, Fewer DP, Sivonen K: Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. *PLoS One* 2011, 6(7):e22384.
- 14. Latifi A, Ruiz M, Zhang CC: Oxidative stress in cyanobacteria. FEMS Microbiol Rev 2009, 33(2):258–278.
- 15. Dietz KJ: The dual function of plant peroxiredoxins in antioxidant defence and redox signaling. *Peroxiredoxin Syst* 2007, 44:267–294.
- Benzie IFF: Evolution of antioxidant defence mechanisms. Eur J Nutr 2000, 39(2):53–61.
- Tripathi BN, Bhatt I, Dietz KJ: Peroxiredoxins: a less studied component of hydrogen peroxide detoxification in photosynthetic organisms. *Protoplasma* 2009, 235(1–4):3–15.
- Dietz KJ, Jacob S, Oelze ML, Laxa M, Tognetti V, de Miranda SMN, Baier M, Finkemeier I: The function of peroxiredoxins in plant organelle redox metabolism. J Exp Bot 2006, 57(8):1697–1709.
- Dietz KJ, Stork T, Finkemeier I, Lamkemeyer P, Li WX, El-Tayeb MA, Michel KP, Pistorius E, Baier M: The role of peroxiredoxins in oxygenic photosynthesis of cyanobacteria and higher plants: peroxide detoxification or redox sensing? *Photoprotection, Photoinhibition, Gene Regulation, and Environ* 2006, 21:303–319.
- Regelsberger G, Jakopitsch C, Plasser L, Schwaiger H, Furtmuller PG, Peschek GA, Zamocky M, Obinger C: Occurrence and biochemistry of hydroperoxidases in oxygenic phototrophic prokaryotes (cyanobacteria). *Plant Physiol Biochem* 2002, 40(6–8):479–490.
- Mutsuda M, Ishikawa T, Takeda T, Shigeoka S: The catalase-peroxidase of Synechococcus PCC 7942: purification, nucleotide sequence analysis and expression in Escherichia coli. Biochem J 1996, 316(Pt 1):251–257.

- Tichy M, Vermaas W: In vivo role of catalase-peroxidase in Synechocystis sp. Strain PCC 6803. J Bacteriol 1999, 181(6):1875–1882.
- Jakopitsch C, Rüker F, Regelsberger G, Dockal M, Peschek GA, Obinger C: Catalase-peroxidase from the cyanobacterium Synechocystis PCC 6803: cloning, overexpression in Escherichia coli, and kinetic characterization. Biol Chem 1999, 380(9):1087–1096.
- Obinger C, Regelsberger G, Furtmuller PG, Jakopitsch C, Ruker F, Pircher A, Peschek G: Catalase-peroxidases in cyanobacteria similarities and differences to ascorbate peroxidases. *Free Radic Res* 1999, 31(Suppl 1):243–249.
- Smulevich G, Jakopitsch C, Droghetti E, Obinger C: Probing the structure and bifunctionality of catalase-peroxidase (KatG). J Inorg Biochem 2006, 100(4):568–585.
- Priya B, Premanandh J, Dhanalakshmi RT, Seethalakshmi T, Uma L, Prabaharan D, Subramanian G: Comparative analysis of cyanobacterial superoxide dismutases to discriminate canonical forms. *BMC Genomics* 2007, 8:435.
- Baier M, Dietz KJ: Primary structure and expression of plant homologues of animal and fungal thioredoxin-dependent peroxide reductases and bacterial alkyl hydroperoxide reductases. *Plant Mol Biol* 1996, 31(3):553–564.
- Kobayashi M, Ishizuka T, Katayama M, Kanehisa M, Bhattacharyya-Pakrasi M, Pakrasi HB, Ikeuchi M: Response to oxidative stress involves a novel peroxiredoxin gene in the unicellular cyanobacterium *Synechocystis* sp PCC 6803. *Plant Cell Physiol* 2004, 45(3):290–299.
- Dietz KJ, Horling F, Konig J, Baier M: The function of the chloroplast 2-cysteine peroxiredoxin in peroxide detoxification and its regulation. *J Exp Bot* 2002, 53(372):1321–1329.
- Rouhier N, Jacquot JP: Plant peroxiredoxins: alternative hydroperoxide scavenging enzymes. *Photosynth Res* 2002, 74(3):259–268.
- Brehelin C, Meyer EH, de Souris JP, Bonnard G, Meyer Y: Resemblance and dissemblance of *Arabidopsis* type II peroxiredoxins: Similar sequences for divergent gene expression, protein localization, and activity. *Plant Physiol* 2003, 132(4):2045–2057.
- Fujii J, Ikeda Y: Advances in our understanding of peroxiredoxin, a multifunctional, mammalian redox protein. *Redox Rep* 2002, 7(3):123–130.
- Hosoya-Matsuda N, Motohashi K, Yoshimura H, Nozaki A, Inoue K, Ohmori M, Hisabori T: Anti-oxidative stress system in cyanobacteria - Significance of type II peroxiredoxin and the role of 1-Cys peroxiredoxin in Synechocystis sp strain PCC 6803. J Biol Chem 2005, 280(1):840–846.
- 34. Stork T, Michel KP, Pistorius EK, Dietz KJ: Bioinformatic analysis of the genomes of the cyanobacteria Synechocystis sp. PCC 6803 and Synechococcus elongatus PCC 7942 for the presence of peroxiredoxins and their transcript regulation under stress. J Exp Bot 2005, 56(422):3193.
- Cui H, Wang Y, Qin S: Genome-wide analysis of carotenoid cleavage dioxygenases in unicellular and filamentous cyanobacteria. Comp Funct Genomics 2012, in press.
- Chi XY, Yang QL, Zhao FQ, Qin S, Yang Y, Shen JJ, Lin HZ: Comparative analysis of fatty acid desaturases in cyanobacterial genomes. *Comp Funct Genomics* 2008, 2008:284508.
- Zhao FQ, Zhang XW, Liang CW, Wu JY, Bao QY, Qin S: Genome-wide analysis of restriction-modification system in unicellular and filamentous cyanobacteria. *Physiol Genomics* 2006, 24(3):181–190.
- Liang C, Zhao F, Wei W, Wen Z, Qin S: Carotenoid biosynthesis in cyanobacteria: structural and evolutionary scenarios based on comparative genomics. Int J Biol Sci 2006, 2(4):197–207.
- 39. JGI: http://genome.jgi.doe.gov/.
- 40. Cyanobase: http://genome.kazusa.or.jp/cyanobase.
- Wood ZA, Schroder E, Harris JR, Poole LB: Structure, mechanism and regulation of peroxiredoxins. *Trends Biochem Sci* 2003, 28(1):32–40.
- Imlay JA: Pathways of oxidative damage. Annu Rev Microbiol 2003, 57:395–418.
- Noctor G, Veljovic-Jovanovic S, Foyer CH: Peroxide processing in photosynthesis: antioxidant coupling and redox signalling. *Philos Trans R* Soc Lond B Biol Sci 2000, 355(1402):1465–1475.
- 44. Chae HZ, Robison K, Poole LB, Church G, Storz G, Rhee SG: Cloning and sequencing of thiol-specific antioxidant from mammalian brain - alkyl hydroperoxide reductase and thiol-specific antioxidant define a large family of antioxidant enzymes. *Proc Natl Acad Sci U S A* 1994, 91(15):7017–7021.

- 45. Chae HZ, Chung SJ, Rhee SG: Thioredoxin-dependent peroxide reductase from Yeast. J Biol Chem 1994, 269(44):27670–27678.
- Hofmann B, Hecht HJ, Flohe L: Peroxiredoxins. *Biol Chem* 2002, 383(3–4):347–364.
- 47. Bryk R, Griffin P, Nathan C: **Peroxynitrite reductase activity of bacterial peroxiredoxins.** *Nature* 2000, **407**(6801):211–215.
- Peshenko IV, Shichi H: Oxidation of active center cysteine of bovine 1-Cys peroxiredoxin to the cysteine sulfenic acid form by peroxide and peroxynitrite. *Free Radic Biol Med* 2001, 31(3):292–303.
- Hillar A, Peters B, Pauls R, Loboda A, Zhang H, Mauk AG, Loewen PC: Modulation of the activities of catalase-peroxidase HPI of *Escherichia coli* by site-directed mutagenesis. *Biochemistry* 2000, 39(19):5868–5875.
- Link AJ, Robison K, Church GM: Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis* 1997, 18(8):1259–1313.
- Chae HZ, Kim HJ, Kang SW, Rhee SG: Characterization of three isoforms of mammalian peroxiredoxin that reduce peroxides in the presence of thioredoxin. *Diabetes Res Clin Pract* 1999, 45(2–3):101–112.
- Seaver LC, Imlay JA: Alkyl hydroperoxide reductase is the primary scavenger of endogenous hydrogen peroxide in *Escherichia coli*. *J Bacteriol* 2001, 183(24):7173.
- Perelman A, Uzan A, Hacohen D, Schwarz R: Oxidative stress in Synechococcus sp. strain PCC 7942: various mechanisms for H₂O₂ detoxification with different physiological roles. J Bacteriol 2003, 185(12):3654–3660.
- Hirotsu S, Abe Y, Okada K, Nagahara N, Hori H, Nishino T, Hakoshima T: Crystal structure of a multifunctional 2-Cys peroxiredoxin heme-binding protein 23 kDa/proliferation-associated gene product. Proc Natl Acad Sci U S A 1999, 96(22):12333–12338.
- Alphey MS, Bond CS, Tetaud E, Fairlamb AH, Hunter WN: The structure of reduced tryparedoxin peroxidase reveals a decamer and insight into reactivity of 2Cys-peroxiredoxins. J Mol Biol 2000, 300(4):903–916.
- Wood ZA, Poole LB, Hantgan RR, Karplus PA: Dimers to doughnuts: Redox-sensitive oligomerization of 2-cysteine peroxiredoxins. *Biochemistry* 2002, 41(17):5493–5504.
- Declercq JP, Evrard C, Clippe A, Vander Stricht D, Bernard A, Knoops B: Crystal structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 angstrom resolution. J Mol Biol 2001, 311(4):751–759.
- Choi HJ, Kang SW, Yang CH, Rhee SG, Ryu SE: Crystal structure of a novel human peroxidase enzyme at 2.0 Å resolution. *Nat Struct Mol Biol* 1998, 5(5):400–406.
- Flohe L, Budde H, Bruns K, Castro H, Clos J, Hofmann B, Kansal-Kalavar S, Krumme D, Menge U, Plank-Schumacher K, et al: Tryparedoxin peroxidase of *Leishmania donovani*: molecular cloning, heterologous expression, specificity, and catalytic mechanism. *Arch Biochem Biophys* 2002, 397(2):324–335.
- Fomenko DE, Gladyshev VN: Identity and functions of CxxC-derived motifs. *Biochemistry* 2003, 42(38):11214–11225.
- Schultz LW, Chivers PT, Raines RT: The CXXC motif: crystal structure of an active-site variant of *Escherichia coli* thioredoxin. *Acta Crystallogr D: Biol Crystallogr* 1999, 55(Pt 9):1533–1538.
- Karplus PA, Hall A: Structural survey of the peroxiredoxins. Peroxiredoxin Syst 2007, 44:41–60.
- Mazouni K, Domain F, Chauvat F, Cassier-Chauvat C: Expression and regulation of the crucial plant-like ferredoxin of cyanobacteria. *Mol Microbiol* 2003, 49(4):1019–1029.
- Dufresne A, Garczarek L, Partensky F: Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol* 2005, 6(2):R14.
- 65. Stacy RAP, Munthe E, Steinum T, Sharma B, Aalen RB: A peroxiredoxin antioxidant is encoded by a dormancy-related gene, Per1, expressed during late development in the aleurone and embryo of barley grains. *Plant Mol Biol* 1996, **31**(6):1205–1216.
- Haslekas C, Stacy RAP, Nygaard V, Culianez-Macia FA, Aalen RB: The expression of a peroxiredoxin antioxidant gene, AtPer1, in Arabidopsis thaliana is seed-specific and related to dormancy. *Plant Mol Biol* 1998, 36(6):833–845.
- 67. Mowla SB, Thomson JA, Farrant JM, Mundree SG: A novel stress-inducible antioxidant enzyme identified from the resurrection plant *Xerophyta viscosa Baker. Planta* 2002, **215**(5):716–726.

- Stacy RAP, Nordeng TW, Culiáňez-Macià FA, Aalen RB: The dormancy-related peroxiredoxin anti-oxidant, PER1, is localized to the nucleus of barley embryo and aleurone cells. *Plant J* 1999, 19(1):1–8.
- Rhee SG, Chae HZ, Kim K: Peroxiredoxins: A historical overview and speculative preview of novel mechanisms and emerging concepts in cell signaling. *Free Radic Biol Med* 2005, 38(12):1543–1552.
- Baier M, Dietz KJ: The plant 2-Cys peroxiredoxin BAS1 is a nuclear-encoded chloroplast protein: its expressional regulation, phylogenetic origin, and implications for its specific physiological function in plants. *Plant J* 1997, 12(1):179–190.
- Jang HH, Chi YH, Park SK, Lee SS, Lee JR, Park JH, Moon JC, Lee YM, Kim SY, Lee KO, et al: Structural and functional regulation of eukaryotic 2-Cys peroxiredoxins including the plant ones in cellular defense-signaling mechanisms against oxidative stress. *Physiol Plant* 2006, **126**(4):549–559.
- Kong W, Shiota S, Shi YX, Nakayama H, Nakayama K: A novel peroxiredoxin of the plant Sedum lineare is a homologue of *Escherichia coli* bacterioferritin co-migratory protein (Bcp). *Biochem J* 2000, 351:107–114.
- 73. NCBI: http://www.ncbi.nlm.nih.gov/sites/gquery.
- 74. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990, 215(3):403–410.
- Mount DW: Using the Basic Local Alignment Search Tool (BLAST). CSH Protoc 2007, 2007:pdb top17.
- 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(17):3389–3402.
- 77. Gertz EM, Yu YK, Agarwala R, Schaffer AA, Altschul SF: Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol* 2006, **4**:41.
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J: The Pfam protein families database. Nucleic Acids Res 2004, 32(suppl 1):D138–D141.
- 79. Johnson LS, Eddy SR, Portugaly E: Hidden Markov model speed heuristic and iterative HMM search. *BMC Bioinforma* 2010, 11:431.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al: Clustal W and Clustal X version 2.0. Bioinformatics 2007, 23(21):2947–2948.
- Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994, 22(22):4673–4680.
- Schultz J, Milpetz F, Bork P, Ponting CP: SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 1998, 95(11):5857–5864.
- Marchler-Bauer A, Bryant SH: CD-Search: protein domain annotations on the fly. Nucleic Acids Res 2004, 32(Web Server issue):W327–W331.
- Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2003, 52(5):696–704.
- Kern R, Bauwe H, Hagemann M: Evolution of enzymes involved in the photorespiratory 2-phosphoglycolate cycle from cyanobacteria via algae toward plants. *Photosynth Res* 2011, 109(1–3):103–114.
- Le SQ, Gascuel O: An improved general amino acid replacement matrix. Mol Biol Evol 2008, 25(7):1307–1320.
- Chevenet F, Brun C, Banuls AL, Jacq B, Christen R: TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinforma* 2006, 7:439–445.
- Team R: R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2010.

doi:10.1186/1471-2148-12-220

Cite this article as: Cui *et al.*: Genome-wide analysis of putative peroxiredoxin in unicellular and filamentous cyanobacteria. *BMC Evolutionary Biology* 2012 **12**:220.