Research **The PRC-barrel: a widespread, conserved domain shared by photosynthetic reaction center subunits and proteins of RNA metabolism** Vivek Anantharaman and L Aravind

Address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

Correspondence: L Aravind. E-mail: aravind@ncbi.nlm.nih.gov

Published: 14 October 2002

Genome Biology 2002, 3(11):research0061.1-0061.9

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2002/3/11/research/0061

© 2002 Anantharaman and Aravind, licensee BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Received: 22 July 2002 Revised: 4 September 2002 Accepted: 9 September 2002

Abstract

Background

Identification of conserved protein domains that span a wide range of biological functions provide deep insights regarding the origin and evolution of complex biological systems. These versatile conserved domains often have catalytic or structural roles that can be utilized, with small variations, in different contexts. The P-loop-containing nucleotide phosphatase fold represents one such catalytic domain that is utilized in almost every conceivable biological system in all the three superkingdoms of life [1,2]. Folds such as the SH3-like barrels, the PAS-like fold, the OB fold, the double-stranded β -helix, the β -propeller and rubredoxin-like zinc ribbons are

Background: The H subunit of the purple bacterial photosynthetic reaction center (PRC-H) is important for the assembly of the photosynthetic reaction center and appears to regulate electron transfer during the reduction of the secondary quinone. It contains a distinct cytoplasmic β -barrel domain whose fold has no close structural relationship to any other well known β -barrel domain.

Results: We show that the PRC-H β -barrel domain is the prototype of a novel superfamily of protein domains, the PRC-barrels, approximately 80 residues long, which is widely represented in bacteria, archaea and plants. This domain is also present at the carboxyl terminus of the pan-bacterial protein RimM, which is involved in ribosomal maturation and processing of 16S rRNA. A family of small proteins conserved in all known euryarchaea are composed entirely of a single stand-alone copy of the domain. Versions of this domain from photosynthetic proteobacteria contain a conserved acidic residue that is thought to regulate the reduction of quinones in the light-induced electron-transfer reaction. Closely related forms containing this acidic residue are also found in several non-photosynthetic bacteria, as well as in cyanobacteria, which have reaction centers with a different organization. We also show that the domain contains several determinants that could mediate specific protein-protein interactions.

Conclusions: The PRC-barrel is a widespread, ancient domain that appears to have been recruited to a variety of biological systems, ranging from RNA processing to photosynthesis. Identification of this versatile domain in numerous proteins could aid investigation of unexplored aspects of their biology.

predominantly non-catalytic domains that are widely represented in multiple functional contexts, with roles such as small-molecule binding, nucleic-acid binding and interaction with other proteins [3-5] (see also the SCOP [6] and CATH [7] databases). Versatile globular domains appear to have emerged fairly early in evolution in various fold classes, such as the α/β or $\alpha+\beta$ mixed folds, or the all- α and all- β folds [5,8]. Comparative genomics and evolutionary studies indicate that many of these versatile folds probably emerged in contexts related to RNA binding in the ancient translation system and were subsequently re-utilized in other biological systems [9,10].

Of particular interest in this context are the small all-ß folds that assume conformations such as barrels or β-helices [3,4]. These structures have considerable potential for functional versatility, because they are able either to accommodate small molecules within cavities formed by the curved β -sheets or to interact with various larger molecules, especially nucleic acids or proteins, via the external surfaces of the sheets. A few ancient and widespread β -rich folds such as the SH3-like barrel and the OB fold appear to have colonized multiple functional niches early in evolution, although their earliest versions may have had roles related to RNA metabolism [9,11-13]. We were interested in identifying other such functionally versatile β-rich folds that could be traced back to the early stages of life's evolution. The availability of extensive genome sequence data and advances in structure determination over recent years allow the successful application of comparative genomics, sequence and structure comparisons to identify any such folds that may be somewhat less widely represented than the OB or SH3-like folds.

Here, we identify one such β -barrel fold typified by the globular domain of the H subunit of the photosynthetic reaction center (PRC-H) from purple proteobacteria such as Rhodopseudomonas viridis [14,15]. The purple bacterial photosynthetic reaction center consists of three primary subunits, of which PRC-L and PRC-M primarily bind the pigments involved in photochemistry, whereas PRC-H appears to be a key regulator of electron transfer between the quinones in photosynthetic reaction centers [16]. So far, homologs of the H subunit have only been found in photosynthetic proteobacteria [17-19] and the carboxy-terminal globular domain of PRC-H shows a distinct β-barrel fold that is structurally unrelated to other characterized β -barrels. This raises the important question of the evolutionary provenance of this unique domain. Here we use sequence-profile analysis and comparative genomics to show that the β -barrel domain of PRC-H defines a novel, widespread superfamily of β-barrel domains that is represented in several bacterial, plant and archaeal genomes. We also show that this β -barrel domain is found in the conserved protein RimM, which is involved in RNA processing and ribosomal assembly in the course of translation. Thus we

provide evidence for an unexpected evolutionary connection between RNA metabolism, translation and the redox reactions in photosynthesis in the form of a shared functionally versatile β -barrel domain.

Results and discussion Identification of the PRC-barrel domain

The PRC-H subunit is a membrane-spanning protein with a single amino-terminal transmembrane helix [14,15]. Its crystal structure reveals a cytoplasmic region comprising a largely non-globular segment followed by a β -barrel-like structure. The entire cytoplasmic region has been classified as a novel β -rich fold with no relatives in the SCOP database [6]. However, we observed that the most carboxy-terminal part of the cytoplasmic region forms a distinct folding unit in the form of a six-stranded β -barrel that could define a novel evolutionarily conserved domain (Figure 1). A DALI search



Figure I

A ribbon representation of the H (gold) and the M (gray) subunits of the photosynthetic reaction complex (PDB Ieys). The PRC-barrel is colored purple to highlight it. The two acidic residues projecting in the direction of the membrane, including the glutamate (E) involved in regulation of quinone reduction, are shown in space-filling representation. The peptide from the amino-terminal tail of the M subunit that interacts with a cleft in the PRC-barrel (PCR-M peptide) is also shown in space-filling representation.

[20] of the PDB database with this β -barrel revealed no specific structural relationships with other folds such as the OB fold or the SH₃-like barrel beyond the presence of curved β sheets, suggesting that it represents a domain with a distinct fold (Figure 2).

To further investigate its evolutionary relationships, we used the sequence of this β -barrel unit from the PRC-H protein (gi: 132177, residues 151-257) of *Rhodopseudomonas viridis* in a PSI-BLAST [21] search of the non-redundant (NR) database at the National Center for Biotechnology Information (NCBI). This search (expect value (e) threshold for inclusion in profile = 0.01) recovered, in addition to the orthologs of the PRC-H proteins from other purple proteobacteria, several uncharacterized proteins from the cyanobacterium *Anabaena* (for example, all5315 and alr5332, iteration 2, e = 10⁻⁶-10⁻⁴), non-photosynthetic α -proteobacteria such as *Mesorhizobium, Sinorhizobium, Brucella* and *Caulobacter* (for example, SMc00885, iteration 4, e = 10⁻⁴ or CAC1676, iteration 5, $e = 10^{-6}$), several other assorted bacteria like Deinococcus, Bacillus and Streptomyces (for example, YlmC iteration 4, $e = 10^{-5}$) and several archaea with completely sequenced genomes. Interestingly, in addition to these proteins, this search also recovered the ribosome-associated RimM protein from bacteria (for example, RimM, Deinococcus radiodurans, iteration 6, $e = 10^{-4}$). To establish the validity of these relationships we collected all the true positives detected in this search and clustered them on the basis of similarity, obtained diverse representatives belonging to each cluster, and seeded PSI-BLAST searches with each of them. The majority of these searches recovered approximately the same set of proteins with statistically significant e-values. For example a search started with the archaeal protein Tao943 (gi: 10640258, whole length) recovers RimM (from *Vibrio cholerae*, iteration 4, $e = 10^{-4}$), and the PRC-H protein (*R. viridis*, iteration 8, $e = 10^{-3}$). Although some of these searches converged prematurely, they consistently recovered true positives detected in the other searches



Figure 2

A comparison of the PRC-barrel with the analogous β -barrels, namely the SH3-like barrel and the OB fold. (a) The representative of the SH3-like barrel is the dihydrofolate reductase subunit (PDB Ivie) and (b) the representative of the OB fold is the cold-shock protein SI-like RNA-binding domain (PDB Imjc). Note the difference in packing of the last strand of the OB fold with respect to the first strand in the SH3-like barrel and (c) the PRC-barrel. In the case of the OB fold, note the difference in orientation of the second strand with respect to the first as compared to the other two β -barrels.

with at least borderline statistical significance (e approximately 0.05-0.01).

We prepared separate multiple sequence alignments of this region for all the major, distinct clusters of the proteins detected in the above searches and predicted secondary structure for each of them. The predicted secondary structure [22] corresponded perfectly with that of the barrel domain of the classic PRC-H proteins. Furthermore, the smallest proteins with this region were detected in the Euryarchaea (for example, Ta0943) and their length of approximately 80 residues exactly corresponded to that of the β-barrel that forms a distinct folding unit seen in the PRC-H subunit. Consistent with this, several proteins, such as mll3685 from Mesorhizobium loti, have duplications or triplications of this region, which indicate that the boundaries of each repeat correspond perfectly to the β -barrel unit of PRC-H. These observations suggest that this region indeed defines a novel evolutionarily mobile domain of approximately 80 residues (Figure 3). We named it the PRC-barrel after the photosynthetic reaction center subunit H, in which it was first observed.

Most of the sequence conservation in the PRC-barrel is centered on the hydrophobic residues that stabilize the six strands of the domain. Additionally, there is a nearly invariant glycine (Figure 3) that corresponds to the beginning of strand 2 and is likely to stabilize the first β -hairpin in the structure. Beyond the conserved core, there is considerable variability in the residues in the loops, and these are likely to impart the specificity required for the diverse interactions of this superfamily.

Potential biological functions of the PRC-barrels

The experimentally characterized PRC-barrel-containing proteins possess diverse biological functions: the PRC-H subunits themselves are involved in photosynthesis in the purple bacteria [14], whereas RimM is a protein that associates with the 30S ribosomal subunit and is required for efficient translation and processing of 16S RNA [23-25]. Gene-disruption studies in Rhodobacter capsulatus indicate that loss of the PRC-H subunit results in disruption of the reaction center and the light-harvesting complex-1 and loss of photosynthetic growth [26,27]. Biochemical studies have pointed out that the PRC-barrel of the purple bacterial PRC-H lies on the cytosolic face of the reaction center and directly affects the redox processes during the photosynthetic reaction [16,28]. On photoactivation there is an electron-transfer chain from the primary donor - the bacteriochlorophyll molecules - to the primary quinone, and then to the secondary quinone. A glutamate residue (E173 in R. viridis PRC-H) located in the loop between strand 2 and 3 of the PRC domain is in the vicinity of the secondary quinone of the reaction center. The site-directed mutagenesis of this glutamate severely retards the first and second electron transfers from the primary quinone that successively reduce the secondary quinone to semi-quinone and quinol [16]. The crystal structure of the reaction center reveals that this acidic residue of the PRC-barrel is situated close to other acidic residues from the PRC-L subunit, which interact with the quinone [16]. Thus, the acidic residue in the loop between the PRC could act as a regulator of the electrostatic state of the reaction complex to potentiate electron transfer. The multiple alignment (Figure 3) of the PRC-barrel reveals that this glutamate, or an equivalent acidic residue, is conserved in the majority of PRC-barrels that are most closely related to the PRC-H version. In addition to the purple bacteria, such versions are seen in the cvanobacteria, α-proteobacteria such as rhizobia, Agrobacterium tumefaciens and Brucella melitensis, and Deinococcus radiodurans (Figure 3). In the case of the cyanobacteria, it is possible that some of the PRC-H-like proteins that contain an equivalent acidic residue might associate with their very distinct photosynthetic reaction centers [29-31], by analogy with their purple bacterial counterparts. One of these proteins from Anabaena cylindrica is exclusively expressed in the spore-like akinetes [32], though its actual function remains unknown. The extensive spread of this

Figure 3 (see figure on the next page)

A multiple alignment of the PRC-barrel was constructed using T-Coffee [38] and realigning the sequences by parsing high-scoring pairs from PSI-BLAST search results. The secondary structure assigned by PHD [22] is shown above the alignment, with E representing a β -strand, and H an α -helix. The 85% consensus shown below the alignment was derived using the following amino-acid classes: hydrophobic (h, ALICVMYFW, yellow shading); the aliphatic subset of the hydrophobic class (I, ALIVMC, yellow shading); small (s, ACDGNPSTV, green) and polar (p, CDEHKNQRST, blue). A 'G' denotes the conserved G of the tiny subset of the small class. Columns of residues that are peculiar to a particular category of PRC-barrels (see text) are colored red. The limits of the domains are indicated by the residue positions on each side. The numbers within the alignment are non-conserved inserts that have not been shown. The different families are shown on the right. The sequences are denoted by their gene name followed by the species abbreviation and GenBank identifier (gi). The species abbreviations are: Archaea: Af, Archaeoglobus fulgidus; Hsp, Halobacterium sp. NRC-1; Mac, Methanosarcina acetivorans; Mta, Methanobacterium thermoautotrophicum; Mj, Methanococcus jannaschii; Ph, Pyrococcus horikoshii; Tac, Thermoplasma acidophilum; Bacteria: Atu, Agrobacterium tumefaciens; Aae, Aquifex aeolicus; Ana, Anabaena sp.; Bs, Bacillus subtilis; Bb, Borrelia burgdorferi; Bmel, Brucella melitensis; Cac, Clostridium acetobutylicum; Ccr, Caulobacter crescentus; Cj, Campylobacter jejuni; Des, Desulfitobacterium hafniense; Drad, Deinococcus radiodurans; Ec, Escherichia coli; Hi, Haemophilus influenzae; Hp, Helicobacter pylori; MIo, Mesorhizobium loti; Mtu, Mycobacterium tuberculosis; Nm, Neisseria meningitidis; Pae, Pseudomonas aeruginosa; Pmar, Prochlorococcus marinus; Rcap, Rhodobacter capsulatus; Rp, Rickettsia prowazekii; Rsp, Rhodobacter sphaeroides; Rvi, Rhodopseudomonas viridis; Sli, Streptomyces lividans; Sme, Sinorhizobium meliloti; Scoe, Streptomyces coelicolor A3; Syco, Synechococcus sp.; Ssp, Synechocystis sp.; Tm, Thermotoga maritima; Tp, Treponema pallidum; Ter, Trichodesmium erythraeum; Tsyn, Thermosynechococcus elongatus; Ttep, Thermochromatium tepidum; Xf, Xylella fastidiosa; Plants: At, Arabidopsis thaliana.

Secondary Structure		EEEEEEEEE	EEE.	EEE	EEEE	EEE	EEEEE	EE	EEEE	e	EEE	. EEF	EE.		
prcH_Rvi_13096660	143	FSIAEGDVDPRGL	P <mark>VV</mark> A	ADGVEA	GT <mark>V</mark> TD	LWVDR	SE-HYFRYL	ELSVAG	SARTAL	PLG	FCDVKKD	KIV	VTSILSEQ	222	\ \
puhA Rsp 2624659	142	FHVSAGK-NPIGI	PVRG	CDGEVA	GKVVE	IWVDI	PE-OMARFL	EVELKD	-GSTRL	PIG	MVKVOSNR-VHVN	IALS	SDLFAGIP	222	
rceh_Rcap_132175	142	MKVSAGR-DPRGM	ip <mark>v</mark> qa	GDTEV <mark>V</mark>	GK <mark>I</mark> VD	MWVDI	PE-QLVRYL	EVELNS	-GKKKL	PMT	MLKIWSDR-VRVN	ia <mark>i</mark> t	SDLFDTIP	224	1
all4050_Ana_17231542	26	YQDSFQGNDIKGI	G <mark>VY</mark> T	ERSDEK	GT <mark>V</mark> ND	VLVDD	-E-GHFRYL	VDL-GFWI-	FGKKVL	PVG	RSRIDYGAG	RVY	AIGMTRDQ	109	
ali4051_Ana_17231543 alr5332 Ana 17232824	14	INEIFGGNDIKDF	DVYA	QSGVRV	GNVDN	ILVDE	DD-GHFRYD	LIDT-GFWV-	FGKKVLI	PIG	LSQINYPER LARLNYENK	RLL	VDGLSKEQ VPGLTKEO	96 96	
all5315_Ana_17232807	14	RDTEFDNYNIKDF	d <mark>vy</mark> s	di – – dndk <mark>v</mark>	gt <mark>v</mark> kh	ILVDD	- <mark>S</mark> -GRFRYL	VDT-GFWF-	FGRQVL	PIG	RSRFNYTNR	-R <mark>V</mark> Y	ANGLTKEQ	97	
mlr9657_Mlo_13488501_A	124	APNGSLVSKIIGA	T <mark>VY</mark> N	GTDANAQT <mark>I</mark>	GKVND	VLLAK	-D-GKAQSL	IGVGGFLGI	GEKNVA	DFN	KAQWAEKDGK	·R <mark>W</mark> L	VIKTTKED	212	
mll3685 Mlo 13473176 A	253 46	RAEGNIVTNIIGE		GTGDDAENI	GKUSD	VVFDK	-D-KKVDAV	JIGVGGFLGV	GTKNVA		KLOWAEKNGD	RWL	VAOTTKDE	134	
mll3685_Mlo_13473176_B	178	RADGNLATNIIGK	T <mark>VY</mark> N	GTGD <mark>DA</mark> QN <mark>I</mark>	GS <mark>V</mark> ND	IVLSK	-E-GKAESL	/IGVGGFLGL	GAKNVT	DFN	KAQWAEKNGD	•R <mark>W</mark> L	v aqttkea	266	
mll3685_Mlo_13473176_C	331	PVGEIKASDLKGT	T <mark>VY</mark> G	ANDAK	GEIGD	VVLTP	-D-KKTDAV	IVNVGGFLGI	GTKEVA	GLD	NLKFMTDKNGK	·KYL	YTTFTKEQ	417	
BMEI0367 Bmel 17982267	151	ANEGLSAEKLIGA	MVYD	ADETKI	GTIGE	VLMSA	-R-DRWDIA -D-GQTEAF	ADVGGFLGI	NSKPVA	SIA	NLDVATGKDGK	LSV	FTOFTKKD	234 91	
BMEI0366_Bmel_17982266	46	SPGQLLTSD <mark>F</mark> IGQ	P <mark>VY</mark> N	GPSD <mark>DA</mark> DN <mark>I</mark>	GT <mark>V</mark> ND	IILGA	-D-GTPQAI	IGVGGFLGV	GEKDVA	GCN	HFSWVDKSGGT	R <mark>W</mark> L	MVDADKEQ	135	(
SMc00885_Sme_15073752	84	AENQVSANTYIGQ	SVYN	ASDESI	GEIND	LIIEK	-E-GGIAAA	VGVGGFLGI	GEKNVA	PFD	KIAVADQPDSDDI	KLT	TTETA-DS	171	
AGR L 1375 Atu 15890813	74	GENOVSANDYIGK	SVYT	GADESI	GNVTN	LIMEC	-D-GGLVAA	/IGVGGFLGI	GAKDVA	PMD	KVTMTRNTOI	GTI	RLTTTETA	159	PRC
AGR_C_3639_Atu_15889293	212	QKTDLTASQ <mark>L</mark> DGM	ID <mark>IY</mark> N	AQNEN <mark>I</mark>	ge <mark>i</mark> ed	VIG-	-DGKSVIGL	/ASVGGFLGI	DKSYVVI	DPA	SVAVHNDNGTV	ik <mark>a</mark> y	<mark>V</mark> DTTKDAL	298	1
MA0627_Mac_19914425_A	151	NPDFLSANTLKGE	KVVN	RAGEDI	GKIEE	LMIDL	QD-GRVGYA	LSFGGFLGM	SDKFFAL	PWQ.	ALQLRVHEH	AFL	LDIPKETL	93	
MA0627_Mac_19914425_B MA0627 Mac 19914425 C	469	IPDFLSANTIKTE	RVVN	TAGEDI	GRIDE	LMIDL	-DNGRVATA	LSFGGFLGM	SDKLFA	PWN	FLTYRAHEH	AFT	LDIPRDVL	554	
MA0477_Mac_19914260	15	NRTFVAATE <mark>I</mark> KGS	K <mark>IL</mark> T	VKDEE <mark>L</mark>	GT <mark>I</mark> KE	<mark>v</mark> mids	-EWGRIAYV	/FACDCFLGM	KCKFFA	PWG	ALHTSRG	•D <mark>Y</mark> I	<mark>l</mark> kvdkdaf	98	
DR1314_Drad_6459065	75	NLNDTSMYNPVGA	A <mark>AY</mark> G	VNGDKI	GT <mark>V</mark> RD	ALVEP	-ETGRIRYF	VDVGGWF	SSKEVL	PVG	YGRVDDS	•G <mark>V</mark> Y	FDSLTKDQ	156	
SCMIU.U/C_SCOE_6580633 SC10A9.21 Scoe 13122124	5	VRTDIDPRNLIGR	KAFD	GDGNKI RDGSRI	GDARH GTVDE	VFFDD	-MIGR-PEW	AIRTGLF	-SRDAF	PLE	SELVDG	-HLE -ALR	VPYGR-DQ VPFDR-AL	80	
SCF55.21c_Scoe_6448732	14	MFEAENIRD <mark>W</mark> RGL	D <mark>VV</mark> D	YDGRK <mark>I</mark>	GT <mark>L</mark> ES	I YFDT	-LTDR-PAF	ATVTIGLPT-	RRRLVF	'P <mark>V</mark> D	ATVGPS	-Y <mark>l</mark> k	VTYDK-SL	94	ĺ
AGR_C_723_Atu_15887760	29	TPSLIASDKVEGI	R <mark>VY</mark> G	ADGKHI	GSIQR	IILEK	-RGGRVAYA	LSFGGFLGI	GDDYYP	PWE	KLHYDEELD	-G <mark>Y</mark> R	IDLTKEQI	114	
SMD20522_Sme_15140376 mlr9581 Mlo 13488432	151	LGPGLRVDETVGS	EVRS	MNGENV SDDKTV	GEVRN	VVIGT	-KGGRVAYA	VASGGILGI	GKDSIV		MLDYNTDLG YIOVNEERN	-GFQ -SFY	URISS-AD	234	
mlr0376_Mlo_13470613	111	HTEAIAASRVIGT	'S <mark>VY</mark> N	TEGRS <mark>I</mark>	gs <mark>i</mark> ed	IMLDK	-TSNGIMFA	/IGFGGFLGI	GEKYHA	PWA	SLDYDENRG	-G <mark>Y</mark> V	VPFSKEQL	196	/
alr1129_Ana_17228624_A	3	SEQIIRRSD <mark>I</mark> LNI	'Q <mark>VI</mark> T	RDNGKRL	GI <mark>V</mark> SQ	VWVDI	-D-QREVVA	GLRDSLI	SISG	PRY	MYLSSIHQYG	D <mark>V</mark> I	LVDNEDVI	80	\
alr1129_Ana_17228624_B Terv p 2245 Ter 23041561 A	82	DIEVEALSNLING	EVIT	ETGEVL		FRFEA	-ETGKLNTI	IASLGVP	QIPDQFI AVAG	STY	EISI-EEVVSTGPN MFLKDVCEIG		VFEGAEER	169	
Tery_p_2245_Ter_23041561_B	83	DIDTEAYSGLINS	EVLT	ENGDFL	GR <mark>V</mark> RD	FKCDV	-RDGKVLSL	IASIGIP	QIPDQI	STY	EMPI-DEIVASGPN	RLI	VFEGSEEK	170	
Pmar_p_57_Pmar_23121939_A	18	SERLWLRSE <mark>L</mark> MGI	Q <mark>VI</mark> T	TDTGRRL	GV <mark>V</mark> GE	VVVDI	-D-RREVVA	GLRDNPLTR	FLPG	PKW	MPLESIKQVG	D <mark>V</mark> I	LVDSLDSL	97	Plant-
Pmar_p_57_Pmar_23121939_B	100	GFSPERYGKVINC	QVIT FVIT	ESGQLL	GRVLG	FSFDI	-ETGDLISL	MGAVGVP	LLGEGVI	STW	EIPV-DEIVSSGTD		VYEGAEEK	182	cyano-
slr1220_Ssp_1652582_B	82	LVEVDLYSPLVNS	EVVT	ETGEPL	GRVRD	FQFDL	-ATGKVSSI	IASLGLP	QIPDQL	STY	ELSI-DEVVSSGPN	RLI	VFEGAEER	169	group
tll0853_Tsyn_22298396_A	31	STQLLQRAD	'Q <mark>VI</mark> T	RDTGRK <mark>L</mark>	GV <mark>I</mark> NQ	VWVDT	'- <mark>D</mark> -QRQVVA	GVRNTLF	TG <mark>I</mark>	QRY	ILLDSIRQIG	- D <mark>V</mark> I	LVEHDDDV	106	
t110853_Tsyn_22298396_B	109	ALNTYNYSTLIDS		ETGELL	GRURG	FKFDP	-DTGDITDL	LASIGLP	WIPAQL	STY	ELPI-EEIVSTGPD	RII	VFEGAETR	191	
At2g38570_At_15224902_A At2g38570_At_15224902_B	196	TYILVYIKISVGY	R <mark>VV</mark> T	PGGRNI	GKVST	YRLDV	ISWEVEV	VDVRF3 VL	QDIV	VQ-	EDAASRKQ	RLT	KGLWDAQF	266	
Tery_p_1588_Ter_23040898_A	4	IPEVIKHSQ <mark>L</mark> LKR	L <mark>VL</mark> D	IETVEDQ	GC <mark>I</mark> EE	LCLNI	QSHQVIG	FICKSGFF	GRQI	(K <mark>Y</mark> F	TWKQVETIGV	- D <mark>A</mark> T	<mark>L</mark> VNGK SQP	80	į
Tery_p_1588_Ter_23040898_B	83	LENYNNIIHIIGN	EVWT	DTGEK <mark>V</mark>	GFIVE	YLLNI	-KTGEIVNY	FKYNGL	-QGALNS	IYL	LNPEAISSVGNK	-R <mark>V</mark> I	VVNNYLDN	162	
orf288 Syco 974616 B	83	PEKVDRLESLOH	E <mark>VW</mark> T	DAGLH <mark>V</mark>	GKIDR	CRFDR	PAHRVLG 2-OTGYISAY	LVSPHGW	-RGVAG	ILWD	LDPQLVLSYGQA	-R <mark>V</mark> I	VAELDPET	162	1
yrrD_Bs_2635192_A	18	DHTLRTCHEVEGF	P <mark>VY</mark> S	ERTSSY <mark>L</mark>	GT <mark>I</mark> SD	ICFSL	KGDCLGF	ILAQKRFL	- HHHHAL <mark>I</mark>	RAC	DISSIFDD	-R <mark>I</mark> L	<mark>a</mark> svsseql	95	N.
yrrD_Bs_2635192_B	99	PKSCFTYEQMKMK		QEGDIL	GMLED	VYFCL	DRGIIVA	YELSDGFFSD	1 AGSKRQ		DSLVEVRKD	-EIV	LNG	174	
Desu p 7167 Des 23121582 A Desu p 7167 Des 23121582 B	96	LSLIKEKLTIIGI	RVIT	0SGKTL	GIVEE GVVEE	YYVDP	QAKALAA -DTGKITOM	SISGGKI		GKA	ILEA-DYVVTIGOD		AOKGCETS	175	YrdD-
DR0894_Drad_6458620_A	1	MIKGKELLGF	.P <mark>VV</mark> A	VSNGAQ <mark>V</mark>	dn <mark>v</mark> hd	VVFDH	IQGNRVLA	LVEEGGW	LRA	K <mark>A</mark> V	PFGRVQSIGE	-K <mark>A</mark> V	MIATPEDV	74	like
DR0894_Drad_6458620_B	84	KEALAGKTNLIGM	TLLT DUVC	TDGQTL	GKITD	VYFDE	RSGQVEG	EASGGLFAD	1 TNGRTF	PAP	QSIQIGKDSA	· I <mark>V</mark> P	AGVAA-AM	169	group
CAC1676 Cac 15894953_A CAC1676 Cac 15894953 B	72	KGSFIRFNSIKNM	DV15 DV15	YDGTIL	GAVDL GMVEE	IMFDE	SSGFVLG	IVSRGFFAN	2 SGKKII	EGN	YMLGKK	-NLF	CFSNDKNL	155	
SC5H1.16_Scoe_7479623_A	1	MDELMAARSLTTE	.P <mark>VV</mark> T	lGGDA <mark>V</mark>	ah <mark>v</mark> kd	TVCDA	AAERITG	TLTGRGLLS	GPLKEG	PWS	AVYALGH	-d <mark>a</mark> v	MIRDRRGL	83	i
SC5H1.16_Scoe_7479623_B	86	AHQQNLRARLRGA	RVLT	EAGTEI	GTVLD	V VVEG	G-TGGRVVG	RVAASRRFV	5 HRRRVY	PRG	RTLTVSG	-R <mark>A</mark> L	VLPDGAVR	174	/
AF0359_AI_2650274 AF1532 Af 2649046	1	MYVPARSLARK	RVFT	DEGRYV	GRVED	VILDO	KTGTIVN	ATSDYNKAL	4 EEGLYI	PFE	VVRSLKD	TTT	VDRRRVR-	86	
VNG1995C_Hsp_10581428	1	-MADILAENLSGK	A <mark>VM</mark> G	SDGTEL	GML YN	ITMDL	KSGTLEH	LVEQTEESV	7 DDGRYR	PVS	RVQAVKD	Y <mark>I</mark> V	vrr	79	
VNG0982C_Hsp_10580536	2	DADADEITSLVGR	E <mark>VY</mark> S	NNGVF <mark>V</mark>	GEVED	IQLNL	NAEAVTG	LALTELNADL	6 GQKGVI	PYR	WVRAVGD	· I I L	INDVVERY	90	
MA3528_Mac_1991/583 MA0818 Mac_19914631	1	MRAELTSLEGI		NNGVYV	GTLSN	LVVEL	KGGHIID.	AVSDINEL	3 SSRGII	PFD	SVSAIKD	· ¥ 🖬 ⊥ - T T T	TRDVTORY	84	
MJ1118_Mj_1591757	1	-MEKMPAKLLFER	SIIG	NKGSVI	GK <mark>V</mark> KD	IVFDE	KVGRLVS	EVEPAEHSP	4 EGRNVL	PYK	LVVAIKD	vvv	IDETNLNR	86	
MJ0916_Mj_2826349	1	MAIRVSD <mark>I</mark> LDK	PIYT	TTAIY <mark>V</mark>	GK <mark>V</mark> YD	V MLDL	NKGVISG	IVSDIQNGC	7 PSKKVVI	PFN	LITAIGN	· I I L	V KPPADSG	88	
MTH974_Mth_2622073 MTH815_Mth_2621906	10	GGKMVELSSLYGL	ETAL	SRGKYV		VVLNI	KKGRVST	KVRPMRHDK	Z6 EEGIID	NYD.	RVQAVGD		LATAADEA	80	Archaeal
MTH1151_Mth_2622257	8	KGEEKYWSE <mark>I</mark> KGY	Q <mark>VA</mark> T	NNARIL	GELEE	LIIND	KTGKITD	VIKVDSGRN	8 KGDYLL	PFG	KVEKVGE	FII	ALAADEA	88	
MTH1859_Mth_2622995	1	MRIVEE <mark>M</mark> VGK	E <mark>VL</mark> D	SSAKV <mark>I</mark>	GK <mark>V</mark> KD	<mark>v</mark> evdi	ESQAIES	LVLGKGGISE	4 SKGETI	PYE	MVKKIGD	-K <mark>I</mark> L	LKGPEE	77	
PH0209a_Ph_14591739	1	MVMELSKMYGK	DUYT	TRGKYI		IVIDI	KEGEGKV		2 NGDRIS	PYE	KVTAVGD	·IIL 	VQAPERK-	70	
Ta0850_Tac_10640127	46	MEFYSRVNQK	PVMS	SSGEII	GY <mark>V</mark> KS	FMADE	KWNLTTM	AKPSGGRGR	6 DDGNYL	PMS	NVNVGND	AIV	VMNIESIR	131	
ylmC_Bs_2633909	1	MISISE <mark>F</mark> QVK	D <mark>VV</mark> N	VSNGKK	<mark>GSI</mark> GD	IDINV	TTGKIQA	LILGGNGKVL	4 KEEELV	PWR	NIVKIGED	V <mark>I</mark> L	VRLSEPHA	81	
ymxH_BS_16078735 CAC1697 Cac 15894974	2	DLPSHSINAMKSM	E <mark>IV</mark> D	IKRAERL	GVLGQ	TDLEI	N-EQDGQIT.	LTT BGS-KAG	4 QGHDIR	PWH DWT	HIQKIGSD	·MLL VTL	UNCONLEV	85	1
ag_1057_Aae_2983543	94	GEEEYYAYELVGM	EVET	DKGKKL	GK <mark>V</mark> ER	VQDM-	GPYDALV	D	-KENLL	PFV	SDIVLKVDKENF	KII.	VKEELLPV	166	\
ylqE_Bs_2633974	97	NEGEFYFHE <mark>I</mark> IGC	EVFT	EEGEL <mark>I</mark>	GK <mark>V</mark> KE	ILTP-	GANDVWV	IGRK	GKKDAL	PYI	ESVVKHIDVREF	K <mark>I</mark> E	IELMEGLI	174	
BB0697_Bb_2688631 BMET0148 Bmel 17982028	93	KEGEYYLGKLIGY	AIVN FAVD	NNKKL		FFEY-	LNSVFLE	VRV	GIKFFF	PFL	SIYIGDINTQEH	TIE	VEDHAACL.	166	
CAC1757_Cac_15895034	89	KEGQYFIKDIIGC	NVFD	ENDKDL	GEVYD	VIST-	KNNDVYC	IRKE	GQQDIL	PAL	KDIVLKIEIENH		IKAVEEWL	166	
CAC0683_Cac_15893971	4	LLNNFFLSE <mark>V</mark> LYR	K <mark>VY</mark> D	EYGEY <mark>V</mark>	GK <mark>L</mark> WD	IYVTA	DESYPRA	IGYKIKKGGE	2 NYEFKS	HFY	REDESRKI	•Y <mark>M</mark> Q	<mark>v</mark> kavkdti	89	į
CC3651_Ccr_16127881	96	DDDEFYLTDLVGL	TVRH	IQTDQLL	GRVKS	VQNF-	GAGDILE	TPDLGG	PTWYI	PFT	RAAVPEVRVSEC		ADPPALVG	178	
DR2010_Drad_6459796	94 97	EEGTFYYHDLRGL	EVPE	AGGERL	GT <mark>V</mark> SD	VMDA-	-GASILFE GHQDLLV	DYGGGT	I IFRIF.	PY-	VEVPLAGGKI	GAV	HLTADAPA	177	
yfjA_Ec_1788960	105	EEGDYYWKDLMGC	Q <mark>VV</mark> T	TEGYD <mark>L</mark>	GK <mark>V</mark> VD	MMET-	GSNDVLV	KANLKDAF -	GIKERL	PFL	DGQVIKKVDLTTF	2S <mark>I</mark> E	VDWDPGF-	185	[
HP1149_Hp_2314304	100	KEGEFFYCDLVGL	SVVE	ENEIL	GK <mark>V</mark> IE	IQRI-	SQTDYFL	VETTLNLVEK	1 LAKIFL	PYR	DFYIQEILLQDF	KIT.	THNAKTLL	184	
mll4288_Mlo 13473626	98 94	EEGEFYHADLVGI	T <mark>VVN</mark> E <mark>VRD</mark>	LEGYTM DTGTAL	GK <mark>V</mark> VA	VHNF-	GSNDVLV GGGDILD	TLAG	-RKGVI	PFL	15QVVKRVDLTTH QAAVPDVSVAEC	SF <mark>V</mark> R	VDWDAGF -	⊥/8 174	RimM
rimM_Mtu_1403428	99	EPDTYYDHQLVGL	M <mark>V</mark> QT.	ATGEG <mark>V</mark>	GV <mark>V</mark> TE	VVHT-	AAGELLA	/KR	DSDEVL	PFV	RAIVTSVSLDDO	31 <mark>V</mark> E	IDPPHGLL	176	
alr2812_Ana_17230304	158	GEDEYHVVDLIGM	E <mark>VF</mark> L	QASGDL <mark>V</mark>	GA <mark>V</mark> VD	VIP	AGNDLLE	SLHEPVTSD	1 KPKTVL	PFV	KAIAPVVDLQTH	RIE	ITPPPGLL	245	
nmmo591_Nm_/225821 rimM Pae 9949912	96 97	EEGEYYWHOLECT	KV1D	NDHTVL 0GROUT	GVIDH	LET-	GANDVLM	KPCAGSI	DDRERU	PFV PYT	SQYIETVDTGSP GQCVLSIDIAAG	SEMB	VDWDADF-	175 175	
RP348_Rp_3860908	93	EEDEFYIADL	Q <mark>IL</mark> D	QSNKEI	GK <mark>I</mark> KN	ILNF-	GAGDIIE	EFL	DKTTEL	PFN	KEFFPIITKD	-Y <mark>V</mark> I	LNYQTKV-	165	
ROS_Sme_15076269	95	DEDEFFYADLEGL	E <mark>AV</mark> D	RTGKSY	CS <mark>V</mark> TG	VFDF-	GAGDLLE	KGP	GLRPVL	PFT	EWSVLEIDLEAC	KLV	IDPTAAGL	174	
009681_S11_14423884 slr0808 Ssp 1653194	92 107	EPGEFHVTDL CI		EDGTEV	GIVTD	TSHL-	G-NDLLG	TLDKN	PDKEVI	PFV	SEIVIGIDLDAQ EAIVPTVELARO	RLF	INTIPGLI	185 1	
TM1568_Tm_4982137	/	CEDEVVEVET DC	DVEV	ESGENV		TTET-	GSNDVLV	7PK	KKKETL	PMT	KDCIVEIDKGAK	KII	AKEMEWI-	176	
TD0007 TD 222220	103	SEDETIFIE	DALT	D 00011			COMPANY,	VICIC			CIVEIDICO A				
1P0907_1p_3323220	103	SAGEFYLADLCRC	ELVF	EGSAV	GV <mark>V</mark> LS	VVEG-	GGSSLLE	QRT	HGGVCY	PFH	RTFIGDVDVGRH	CK <mark>I</mark> E	LQRWILE	174	
XF0108_Xfa_9104888 BAB10810 At 10177579	103 100 97 176	SAGEFYLADLCRC HPDEYYWVDLEGL DEGEFYSRDLVCM	ELVF QVHT RVL	EGSAV LEGVVL KETCOLV	GV <mark>V</mark> LS GS <mark>V</mark> SH GTVAM	VVEG- LFSN- IFDN-	GGSSLLE GANDVIV	VQRT IH	HGGVCY GERERLI	PFH PFV PFV	RTFIGDVDVGR QPDYVKSVDFEA DAIVPDVDIEP	KIE RIV EMV	LLQRWILE VDWDPEF-	174 170 268	 /

Figure 3 (see legend on the previous page)

version of the domain in non-photosynthetic proteobacteria such as the rhizobia, suggest that a similar mechanism of regulating electron transfers may, perhaps, be used in regulating non-photosynthetic electron-transfer reactions.

Many PRC-barrels, including those of the RimM family, lack the acidic residue typical of those related to the PRC-H subunit and required for redox regulation. The crystal structure shows that the PRC-barrel domain mediates a contact with the amino-terminal cytoplasmic peptide of the PRC-M subunit and also makes contacts with the other structurally less ordered regions of the PRC-H polypeptide [14,15] (Figure 1). This suggests that, in addition to the specific electrostatic regulatory function, the PRC-barrel mediates specific interactions with other molecules through multiple surfaces. These interactions are consistent with the 'foundational role' postulated for the PRC-H protein in the assembly of the reaction center [33]. Such a protein-protein interaction role in the assembly of complexes could be a potential function of the PRC-barrels that do not possess the features suggestive of redox regulation. Given that RimM specifically associates with the 30S ribosomal subunit rather than the fully assembled ribosome, and participates in the maturation of the 16S rRNA [23-25], it is possible that the carboxyterminal PRC-barrel could be used to bind RNA or proteins. An acidic residue at the beginning of strand 3 and a patch of large and acidic residues at the extreme amino terminus of the PRC-barrel that is specific to the RimM proteins are of particular interest in this regard (Figure 3). On the basis of the structure of the PRC-barrel it can be predicted that in RimM these residues are likely to line a cleft that may accommodate a peptide from an interacting protein (Figures 1,3). Other surfaces of the PRC-barrel in the RimM protein could interact with other proteins, or, alternatively, interact with RNA. The original report on the crystal structure of the photosynthetic reaction center suggested that the region corresponding to the PRC-barrel of the PRC-H subunit could bind a small-molecule ligand [14]. While the structure of the PRC-barrels with a central aperture (Figures 1,2) makes this a tempting possibility, currently there is no evidence to support the possibility that these domains bind small-molecule ligands.

Evolutionary history and diversification of the **PRC**-barrels

The phyletic patterns and relationships of the PRC-barrels have a number of important implications, including some for the evolution of the photosynthetic reaction center in bacteria. Phylogenetic analysis and similarity-based clustering show that these domains essentially form three large and distinct groups (Figure 4) and at least two smaller clusters. The PRC-H subunits and their close relatives from cyanobacteria, non-photosynthetic α -proteobacteria and *D. radiodurans* form the first of these groups. They are represented in multiple copies in the proteomes of most α -proteobacteria. In addition, they are also seen in multiple copies in the cyanobacterium Anabaena, the actinomycete Streptomyces coelicolor, D. radiodurans and the euryarchaeon Methanosarcina acetivorans. In this group, the PRC-barrel occurs either linked to amino-terminal transmembrane helices or as tandem repeats, with up to three copies, or linked to another α -helical repetitive domain also found in the Bacillus protein YsnF (Figure 4). Most members of this group contain a negatively charged residue in the loop between strand 2 and 3 (Figure 3) that could be implicated in electrostatic regulatory processes, as in the case of PRC-H proper. The second major group is represented in all the euryarchaeal genomes available to date and is additionally found in a few Gram-positive bacteria. All proteins belonging to this class are the stand-alone minimal version of the PRC-barrel. The third large group of these proteins comprises the RimM orthologs. These are the most prevalent of all the PRC-barrel proteins and are present in a single copy in all bacterial proteomes available to date, and also in plants in the form of a chloroplast-derived version. The RimM proteins have an additional specific amino-terminal β-strand-rich domain with no detectable relationship to other domains.

In addition to these three major groups, there is one small group that has representatives only in the cyanobacteria and plants (Figures 2,4) and appears to contain more distant relatives of the PRC-H-like PRC-barrels. All members of this group contain two copies of the PRC-barrel, and, in most cases, one of these copies contains an acidic residue in the loop between strand 2 and 3 that might be equivalent to the classic PRC-H-like regulatory acidic residue. The presence of a predicted amino-terminal transit peptide in the plant proteins suggests that it probably functions in the chloroplast, and that these proteins may, in part, have a regulatory function analogous to the PRC subunit. The other small group, typified by the Bacillus subtilis protein YrrD, has so far been found only in the actinomycetes, Gram-positive bacteria and D. radiodurans (Figures 3,4). This group is most closely related to the above-discussed plant-cyanobacterial group, and is ultimately a distant relative of the larger PRC-H like group. Most members of this group have duplicate copies of the PRC-barrel domain and typically lack the acidic residue in the loop between strands 2 and 3.

The phyletic pattern of RimM is reminiscent of proteins that for part of the ribosome or participate in RNA metabolism [9]. Likewise, the presence of the euryarchaeal-type solo PRC-barrels in every euryarchaeal genome, despite the metabolic diversity of these euryarchaea, suggests that these proteins probably have a core cellular function. One likely possibility is that they function in RNA metabolism, perhaps as the archaeal equivalents of RimM. In contrast to these two groups of PRC-barrels, the PRC-H-like group and their more distant relatives have a phyletic distribution, mainly limited to bacteria or archaea with large genomes and complex metabolism. This suggests that they were, perhaps, derived later in bacteria evolution from a version involved in



Figure 4

Phylogenetic relationships of the PRC-barrel-containing proteins along with the domain architectures. The phyletic pattern of each family is shown, along with the number of proteins (if there is more than one). Species abbreviations are as in Figure 3. The RELL bootstrap values for the major branches are shown at their base. The thickness of a given branch is approximately proportional to the number of proteins contained within it. Ysnf, a repeat domain typified by the *Bacillus subtilis* YsnF protein; N-term, the specific amino-terminal domain of the RimM proteins.

RNA metabolism. The extensive presence of multiple copies of the PRC-H-like versions in diverse α-proteobacteria suggests that this particular form, with the characteristic acidic residue, was derived in their common ancestor. A corollary to this is that these proteins probably functioned, initially, as regulators of electron-transfer chains in non-photosynthetic energy metabolism systems in the ancestral α -proteobacterium. They were subsequently utilized in the photosynthetic reaction center after a subset of these bacteria acquired photosynthesis. The phylogenetic tree supports a close relationship between some of the cyanobacterial PRC-barrels and the PRC-H proteins of the photosynthetic α -proteobacteria (Figure 4). This implies that the cyanobacteria probably acquired these forms of the PRC-barrel through lateral transfer from the purple bacteria and may have incorporated them as regulatory subunits into the organizationally distinct cyanobacterial photosystems [29-31] or into other uncharacterized electron-transfer chains. This is consistent with the previously observed case of horizontal transfer of α -proteobacterial reaction-center genes, including that for PRC-H, into the β -proteobacteria [34].

Conclusions

We show that the carboxy-terminal β -barrel domain of the H subunit of the photosynthetic reaction center defines a novel all- β -sheet fold, representatives of which are widespread throughout the prokaryotic world. Homologs of PRC-H, with a conserved acidic residue that has been shown to have a role in regulating electron transfer in the reduction of the secondary quinone of reaction centers, are also found in non-photosynthetic α -proteobacteria, cyanobacteria, *D. radiodurans* and the archaeon *M. acetivorans*. It appears likely that the PRC-H-like version of the PRC-barrels was first derived in the

ancestral *a*-proteobacteria, followed by dissemination into other lineages. Probably, these proteins originally functioned in non-photosynthetic electron-transfer chains and were subsequently incorporated into the photosynthetic apparatus after its emergence in the α -proteobacteria. In addition, this domain also appears to mediate specific protein-protein interactions. This is likely to be the principal role of versions of this domain present in the pan-bacterial RimM proteins and other proteins widely distributed in bacteria and archaea. A protein comprising only a stand-alone copy of the PRC-barrel is conserved in all the euryarchaeal proteomes available to date, and, by analogy with the RimM protein, is predicted to function in RNA metabolism. It seems possible that PRC-barrels with a sporadic distribution and a regulatory function in energy metabolism or photosynthesis are likely to have been derived from more conserved and ancient versions that were probably involved in RNA metabolism. The identification of this domain may help in the exploration of hitherto unexplored facets of diverse biological processes such as photosynthesis, energy metabolism and RNA metabolism.

Materials and methods

The non-redundant (NR) database of protein sequences at NCBI was searched using the BLASTP program [21]. Profile searches were conducted using the PSI-BLAST program with either a single sequence or an alignment used as the query, with a profile-inclusion expectation (E) value threshold of e = 0.01, and were iterated until convergence [21,35]. Before use in PSI-BLAST searches, the PRC domain was evaluated for compositional bias using the SEG program [36]. No such bias that could skew the statistics of sequence relationships in searches of the NR database was detected. Accordingly, to achieve maximum sensitivity, all searches were run with the compositional-bias-based statistics turned off [37]. Multiple alignments were constructed using the T_Coffee program [38], followed by manual correction based on the PSI-BLAST results.

Structural manipulations were carried out using the Swiss-PDB viewer program [39] and the ribbon diagrams were constructed with MOLSCRIPT [40]. Searches of the PDB database with query structures was conducted using the program DALI [20]. Protein secondary structure was predicted using a multiple alignment as the input for the program PHD [22]. Signal peptides were predicted using SIGNALP [41,42] and the transmembrane regions were predicted using TOPRED [43].

Phylogenetic analysis was carried out using the maximumlikelihood, neighbor-joining and least squares methods [44,45]. Briefly, this involved the construction of a leastsquares tree using the FITCH program or a neighbor-joining tree using the NEIGHBOR program (both from the PHYLIP package) [46], followed by local rearrangement using the Protml program of the Molphy package [42] to arrive at the maximum likelihood (ML) tree. The statistical significance of various nodes of this ML tree was assessed using the relative estimate of logarithmic likelihood bootstrap (Protml RELL-BP) with 10,000 replicates.

References

- Leipe DD, Wolf YI, Koonin EV, Aravind L: Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 2002, 317:41-72.
- Neuwald AF, Aravind L, Spouge JL, Koonin EV: AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* 1999, 9:27-43.
- Lo Conte L, Brenner SE, Hubbard TJ, Chothia C, Murzin AG: SCOP database in 2002: refinements accommodate structural genomics. Nucleic Acids Res 2002, 30:264-267.
- Orengo CA, Bray JE, Buchan DW, Harrison A, Lee D, Pearl FM, Sillitoe I, Todd AE, Thornton JM: The CATH protein family database: a resource for structural and functional annotation of genomes. Proteomics 2002, 2:11-21.
- Todd AE, Orengo CA, Thornton JM: Evolution of function in protein superfamilies, from a structural perspective. J Mol Biol 2001, 307:1113-1143.
- SCOP: structural classification of proteins [http://scop.mrclmb.cam.ac.uk/scop/]
- 7. CATH: protein structure classification
- [http://www.biochem.ucl.ac.uk/bsm/cath_new/index.html]
- Koonin EV, Wolf YI, Aravind L: Protein fold recognition using sequence profiles and its application in structural genomics. Adv Protein Chem 2000, 54:245-275.
- Anantharaman V, Koonin EV, Aravind L: Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res 2002, 30:1427-1464.
- Aravind L, Anantharaman V, Koonin EV: Monophyly of class I aminoacyl tRNA synthetase, USPA, ETFP, photolyase, and PP-ATPase nucleotide-binding domains: implications for protein evolution in the RNA world. Proteins 2002, 48:1-14.
- Anantharaman V, Koonin EV, Aravind L: Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol 2001, 307:1271-1292.
- Agrawal V, Kishan RK: Functional evolution of two subtly different (similar) folds. BMC Struct Biol 2001, 1:5.
- Nakagawa A, Nakashima T, Taniguchi M, Hosaka H, Kimura M, Tanaka I: The three-dimensional structure of the RNAbinding domain of ribosomal protein L2: a protein at the peptidyl transferase center of the ribosome. EMBO J 1999, 18:1459-1467.
- Deisenhofer J, Michel H: Nobel lecture. The photosynthetic reaction centre from the purple bacterium Rhodopseudomonas viridis. EMBO J 1989, 8:2149-2170.
- Allen JP, Feher G, Yeates TO, Rees DC, Deisenhofer J, Michel H, Huber R: Structural homology of reaction centers from Rhodopseudomonas sphaeroides and Rhodopseudomonas viridis as determined by x-ray diffraction. Proc Natl Acad Sci USA 1986, 83:8589-8593.
- Takahashi E, Wraight CA: Potentiation of proton transfer function by electrostatic interactions in photosynthetic reaction centers from Rhodobacter sphaeroides: first results from site-directed mutation of the H subunit. Proc Natl Acad Sci USA 1996, 93:2640-2645.
- Berard J, Gingras G: The puh structural gene coding for the H subunit of the Rhodospirillum rubrum photoreaction center. Biochem Cell Biol 1991, 69:122-131.
- Fathir I, Mori T, Nogi T, Kobayashi M, Miki K, Nozawa T: Structure of the H subunit of the photosynthetic reaction center from the thermophilic purple sulfur bacterium, Thermochromatium tepidum. Implications for the specific binding of the lipid molecule to the membrane protein complex. Eur J Biochem 2001, 268:2652-2657.
- Beja O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, Hamada T, Eisen JA, Fraser CM, DeLong EF: Unsuspected diversity

commer

among marine aerobic anoxygenic phototrophs. *Nature* 2002, **415:**630-633.

- 20. Holm L, Sander C: Protein structure comparison by alignment of distance matrices. *J Mol Biol* 1993, 233:123-138.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389-3402.
- 22. Rost B, Sander C: Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 1993, 232:584-599.
- Bylund GO, Persson BC, Lundberg LA, Wikstrom PM: A novel ribosome-associated protein is important for efficient translation in Escherichia coli. J Bacteriol 1997, 179:4567-4574.
- Bylund GO, Wipemo LC, Lundberg LA, Wikstrom PM: RimM and RbfA are essential for efficient processing of 16S rRNA in Escherichia coli. J Bacteriol 1998, 180:73-82.
- Bylund GO, Lovgren JM, Wikstrom PM: Characterization of mutations in the metY-nusA-infB operon that suppress the slow growth of a DeltarimM mutant. J Bacteriol 2001, 183:6095-6106.
- Wong DK, Collins WJ, Harmer A, Lilburn TG, Beatty JT: Directed mutagenesis of the Rhodobacter capsulatus puhA gene and orf 214: pleiotropic effects on photosynthetic reaction center and light-harvesting I complexes. J Bacteriol 1996, 178:2334-2342.
- Sockett RE, Donohue TJ, Varga AR, Kaplan S: Control of photosynthetic membrane assembly in *Rhodobacter sphaeroides* mediated by *puhA* and flanking sequences. J Bacteriol 1989, 171:436-446.
- Knox PP, Churbanova IY, Zakharova NI, Krasil'nikov PM, Lukashev EP, Rubin AB, Shaitan KV: Effects of extraction of the H-subunit from Rhodobacter sphaeroides reaction centers on relaxation processes associated with charge separation. Biochemistry (Mosc) 2001, 66:91-95.
- Scheller HV, Jensen PE, Haldrup A, Lunde C, Knoetzel J: Role of subunits in eukaryotic photosystem I. Biochim Biophys Acta 2001, 1507:41-60.
- Xu W, Tang H, Wang Y, Chitnis PR: Proteins of the cyanobacterial photosystem I. Biochim Biophys Acta 2001, 1507:32-40.
- 31. Huber R: Nobel lecture. A structural basis of light energy and electron transfer in biology. *EMBO J* 1989, 8:2125-2147.
- Zhou R,Wolk CP: Identification of an akinete marker gene in Anabaena variabilis. J Bacteriol 2002, 184:2529-2532.
- Cheng YS, Brantner CA, Tsapin A, Collins ML: Role of the H protein in assembly of the photochemical reaction center and intracytoplasmic membrane in *Rhodospirillum rubrum*. *J Bacteriol* 2000, 182:1200-1207.
- Igarashi N, Harada J, Nagashima S, Matsuura K, Shimada K, Nagashima KV: Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *Mol Evol* 2001, 52:333-341.
- 35. Aravind L, Koonin EV: Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol 1999, 287:1023-1040.
- Wootton JC: Non-globular domains in protein sequences: automated segmentation using complexity measures. Comput Chem 1994, 18:269-285.
- 37. Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF: Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res 2001, 29:2994-3005.
- Notredame C, Higgins DG, Heringa J: T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000, 302:205-217.
- 39. Guex N, Peitsch MC: SWISS-MODEL and the Swiss-Pdb-Viewer: an environment for comparative protein modeling. *Electrophoresis* 1997, 18:2714-2723.
- 40. Kraulis PJ: Molscript. J Appl Crystallogr 1991, 24:946-950.
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G: A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int J Neural Syst 1997, 8:581-599.
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G: Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 1997, 10:1-6.

- von Heijne G: Membrane protein structure prediction: hydrophobicity analysis and the 'positive inside' rule. J Mol Biol 1992, 225:487-494.
- Hasegawa M, Kishino H, Saitou N: On the maximum likelihood method in molecular phylogenetics. J Mol Evol 1991, 32:443-445.
- Felsenstein J: Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol* 1996, 266:418-427.
- Felsenstein J: PHYLIP Phylogeny Inference Package (Version 3.2). Cladistics 1989, 5:164-166.