Correspondence

PAS: a multifunctional domain family comes to light

Chris P. Ponting* and L. Aravind†

Light regulates gene expression by resetting molecular oscillators ('clocks') that are associated with circadian rhythmicity [1]. The mouse *Clock* gene product contains an amino-terminal putative basic helix–loop–helix DNA-binding domain followed by a large region (termed the Per, ARNT, Sim or PAS domain) that contains two imperfect repeats [2,3]; here, we shall apply the term PAS to a single repeat unit. PAS-domain-containing proteins represent an evolutionarily related family, members of which regulate circadian rhythmicity in diverse organisms [4,5]. Here, we report that PAS domains are present in several hundred proteins, including dozens of histidine kinase homologues and also ether-à-go-go-like K^+ channels (Table 1). We also reveal that similar 40–45 amino acid regions (herein referred to as PAC motifs) found carboxy-terminal to several PAS sequences are likely to contribute to the PAS structural domain (Figure 1).

Analyses of eubacterial genomes showed that PAS/PAC homologues are present in *Escherichia coli*, *Haemophilus influenzae*, *Synechocystis* sp. PCC6803, *Mycobacterium tuberculosis* and *Bacillus subtilis*, but not in *Mycoplasma genitalium*, *Mycoplasma pneumoniae* or *Helicobacter pylori*. The *Synechocystis* genome encodes the largest number — 61 — of PAS/PAC domains, perhaps reflecting a necessity for this photosynthetic bacterium to co-ordinate photoreception with cellular functions. As with eubacterial genomes, PAS/PAC sequences in archaeal genomes are either

Table 1

Representative PAS / PAC-containing proteins.

Name	Accession code[1]	Number of: PAS	PAC	Species	Function (known substrates of histidine kinases in parentheses)
Animals:					
Clock	AF000998	2	0	Mouse	Regulator of circadian clock
per	PER_DROME	2	1	*Drosophila*	Regulator of circadian clock
h-erg	U04270	1	1	Human	Voltage-activated potassium channel
Plants:					
Phytochrome	PHYA_ARATH	2	1	*A. thaliana*	Regulatory photoreceptor, phytochrome A
Protein kinase	Z30332	1	1	*S. oleracea*	Unknown; Ser/Thr kinase
Dictyosteliida:					
dokA	X96869	2	2	*D. discoideum*	Regulator of the *Dictyostelium* osmotic response system
Fungi:					
wc-1	X94300	3	2	*N. crassa*	Autoregulates *wc-1* gene expression after blue-light induction
wc-2	Y09119	1	0	*N. crassa*	Regulates expression of light-regulated genes
Yal017w	KAB7_YEAST	2	0	*S. cerevisiae*	Unknown; probable Ser/Thr kinase
Bacteria:					
air	AIR_ECOLI	1	1	*E. coli*	Putative signal transducer for aerotaxis
arcB	ARCB_ECOLI	1	1	*E. coli*	Regulator of aerobic metabolism; histidine kinase (arcA)
bvgS	BVGS_BORBR	1	1	*B. bronchiseptica*	Virulence sensor protein; histidine kinase (bvgA)
dcrA	DCRA_DESVH	1	1	*D. vulgaris*	Chemoreceptor gene A
fixL	FIXL_RHIME	1	1	*R. meliloti*	Oxygen sensor; histidine kinase (fixJ)
hoxJ	U82565	1	1	*A. hydrogenophilus*	Hydrogen sensor; histidine kinase
kinA	KINA_BACSU	3	2	*B. subtilis*	Sporulation histidine kinase (spo0A, spo0F)
kinC	KINC_BACSU	1	1	*B. subtilis*	Sporulation histidine kinase (spo0A?)
mcpA	X86707	2	0	*R. sphaeroides*	Methyl-accepting chemotaxis protein
nifL	NIFL_AZOVI	2	1	*A. vinelandii*	Inhibitor of transcriptional activation by nifA
nodV	NODV_BRAJA	3	4	*B. japonicum*	Nodulation protein V; histidine kinase (nodW)
ntrB	NTRB_AZOBR	1	0	*A. brasilense*	Nitrogen regulation protein; nitrogen-regulated histidine kinase (ntrC)
ntrY	NTRY_AZOCA	1	1	*A. caulinodans*	Nitrogen regulation protein; histidine kinase (ntrX)
nwsA	Z22637	4	4	*B. japonicum*	Suppresses nodulation defect of nodW mutant; histidine kinase
phoR	PHOR_BACSU	1	0	*B. subtilis*	Regulates alkaline phosphatase gene expression; histidine kinase (phoB)
pilS	PILS_PSEAE	1	0	*P. aeruginosa*	Regulates expression of type 4 fimbriae; histidine kinase (pilR)
pleC	PLEC_CAUCR	2	1	*C. crescentus*	Regulator of genes involved in polar organelle development (divK)
ppsR	L19596	2	0	*R. sphaeroides*	Aerobic repressor of genes involved in photopigment biosynthesis
pyp	PYP_ECTHA	1	0	*E. halophila*	Photoactive yellow protein; cytosolic photoreceptor
(Cyanobacteria):					
cyaB1	D89623	1	0	*Anabaena* sp.	Adenylate cyclase homologue; membrane-localised
plpA	D90905	4	2	*Syn.* sp. PCC6803	Probable light receptor (phytochrome)
rcaE	U59741	1	1	*F. diplosiphon*	Regulator of complementary chromatic adaptation
Archaea:					
bat	BAT_HALHA	1	1	*H. halobium*	Regulates bacteriorhodopsin gene expression
ORF	C1_585	1	1	*M. thermoautotroph.*	Unknown; contains cheY-like domain
ORF	F1_78	3	1	*M. thermoautotroph.*	Unknown; contains cheY-like domain
ORF	L_610	2	2	*A. fulgidus*	Unknown; contains periplasmic amino-acid-binding domain

[1]Accession codes use SwissProt or EMBL nomenclature. *A. fulgidus* and *M. thermoautotrophicum* ORFs labelled according to genome annotation.

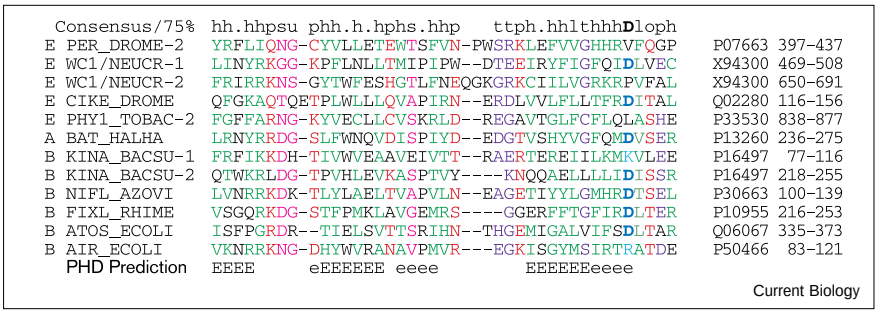
apparently absent (*Methanococcus janaschii*) or abundant (*Archeoglobus fulgidus* and *Methanobacterium thermoautrophicum*).

Approximately half of all eubacterial or archaeal PAS/PAC-containing proteins are soluble or membrane-bound histidine-kinase sensor molecules. Without exception, these PAS and PAC sequences occur in regions amino-terminal to their histidine-kinase domains. These amino-terminal regions represent regulatory domains that sense input signals [6], suggesting that PAS sequences have sensory function(s). PAS and PAC sequences occur also in eubacterial proteins that contain other types of signalling domains (see Table 1 and Supplementary material published with this paper on the internet).

In addition to previously reported eukaryotic PAS-like sequences [4,5], we have identified PAS/PAC sequences in human and plant serine/threonine kinases, two additional PAS motifs and two PAC motifs in *Neurospora crassa* wc-1 and, unexpectedly, a single PAS+PAC domain in members of the ether-à-go-go family of K^+ channels [7].

The PAS family may be divided into those that have the PAS motif followed in sequence by a PAC motif, and those that do not; PAC sequences always appear to be preceded by PAS. Evidence that PAS+PAC, and not PAS alone, represents the structural domain comes from the known tertiary structure of *Ectothiorhodospira halophila* photoactive yellow protein (PYP) [8]; this protein appears to contain a PAS sequence without an identifiable PAC sequence ([4,5]; this work). The PAS-like region of PYP encompasses the majority of the

Figure 1



Multiple alignments of representative PAC sequences. Predicted [9] secondary structures are shown beneath the alignment: H/h denotes an α-helix and E/e denotes a β strand predicted with an expected accuracy higher than 82% (upper case) or 72% (lower case). Consensus sequences are shown above the alignment [threshold = 75%; c = charged (DEHKR), h = hydrophobic (ACFGHIKLMRTVWY), p = polar (CDEHKNQRST), o = alcohol (ST), t = turn-like (ACDEGHKNQRST), and s = small (ACDGNPSTV)]; polar residues are shown in red, charged in cyan, turn-like in blue, tiny or small in magenta, and hydrophobic in green. EMBL or SwissProt accession codes, and residue limits, are shown to the right of the alignments. Entries are subdivided into eukaryotic (E), archaeal (A) or eubacterial (B) sequences. Insertions or deletions are represented by hyphens. Species: AZOVI, *Azotobacter vinelandii*; BACSU, *Bacillus subtilis*; BORBR, *Bordetella bronchiseptica*; DROME, *Drosophila melanogaster*; ECOLI, *Escherichia coli*; HALHA, *Halobacterium halobium*; NEUCR, *Neurospora crassa*; RHIME, *Rhizobium meliloti*; TOBAC, *Nicotiana tobacum*; PAS and PAC units in *B. subtilis* kinA were initially identified and their significance assessed using MACAW [19] and two BLAST programs [20]. For example, a BLAST1 search with the non-transmitter region of kinA yielded apparently significant ($p < 10^{-4}$) similarities to human, mouse and *Drosophila* Sim homologues; three kinA repeats could be aligned using MACAW with $p = 2.0 \times 10^{-4}$ (calculated using a maximal searchspace). Iterative methods, including SWise, HMMer, pattern-matching methods and a recently developed method combining BLAST with profile analysis, PSI-BLAST [20–22], were used subsequently to identify additional PAS/PAC sequences. Extensive alignments of PAS and PAC sequences, and a larger version of Table 1 are available as Supplementary material, published with this paper on the internet.

PYP α/β fold, yet excludes three carboxy-terminal β strands. As the predicted PAC secondary structure comprises three β strands [9], we suggest that PAS+PAC represents the structural domain.

PAS+PAC domain sequences are relatively divergent and are likely to have evolved different functions as a result of contrasting selective pressures on organisms. Animal PAS domains have protein-binding and dimerisation functions [10–12], which are expected to be retained by other homologues. Indeed, oat phytochrome A PAS domains are known to form homotypic and heterotypic dimers [13].

We speculate, however, that PAS/PAC is a versatile ligand-binding domain: PYP covalently binds hydroxycinnamic acid [8], nifL contains flavin adenine dinucleotide (FAD) [14], and the fixL PAS+PAC domain contains haem and can bind oxygen [15]. Although mammalian hypoxia inducible factor HIF1α and bacterial oxygen-sensor fixL possess similar functions and common domains, HIF-1α is unlikely to be a mammalian oxygen-sensor haemoprotein [16] because a haem-binding histidine residue in fixL [15] is not conserved in either of the two HIF-1α PAS+PAC domains.

Other evidence suggests that many PAS+PAC domains may act as light sensors resulting from FAD-binding functions. *Azotobacter vinelandii* nifL and *E. coli* Aer are known and proposed flavoproteins, respectively [14,17]; however, precise FAD-binding sites remain unknown. Although PAS+PAC domains contain neither a βαβ FAD-binding motif [18] nor a conserved residue involved in covalent attachment to FAD, the presence of FAD in *N. crassa* wc-1 and wc-2 might account for their participation in blue-light signal transduction [5], because this flavin specifically absorbs blue light.

In conclusion, two motifs termed PAS and PAC are represented in each of the three major branches of the tree of life. The presence of PAS/PAC homologues in circadian-cycle regulatory proteins from diverse organisms may reflect the ability of some of these domains to sense cyclical variations in blue light and redox status in cellular environments.

Acknowledgements

We would like to thank Eugene Koonin for encouragement, suggestions and the PSI-BLAST analysis, and Peer Bork and Annalisa Pastore for valuable discussions. Pre-release *M. thermoautotrophicum* data were kindly provided by the Genome Therapeutics Corporation, USA. C.P.P. is a Wellcome Trust Career Development Research Fellow and a member of the Oxford Centre for Molecular Sciences. L.A. was supported by the Regent's Fellowship.

References

1. Dunlap JC: Genetic and molecular analysis of circadian rhythms. *Annu Rev Genet* 1996, 30:579-601.
2. Nambu JR, Lewis JO, Wharton KA Jr, Crews ST: The ***Drosophila*** single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* 1991, 67:1157-1167.
3. Hoffman EC, Reyes H, Chu FF, Sander F, Conley LH, Brooks BA, Hankinson O: Cloning of a factor required for activity of the Ah (dioxin) receptor. *Science* 1991, 252:954-958.
4. Lagarias DM, Wu SH, Lagarias JC: Atypical phytochrome gene structure in the green alga ***Mesotaenium caldariorum***. *Plant Mol Biol* 1995, 29:1127-1142.
5. Crosthwaite SK, Dunlap JC, Loros JJ: ***Neurospora*** wc-1 and wc-2: transcription, photoresponses, and the origins of circadian rhythmicity. *Science* 1997, 276:763-769.
6. Parkinson JS: Signal transduction schemes of bacteria. *Cell* 1993, 73:857-871.
7. Warmke JW, Ganetzky B: A family of potassium channel genes related to eag in ***Drosophila*** and mammals. *Proc Natl Acad Sci USA* 1994, 91:3438-3442.
8. Borgstahl GEO, Williams DR, Getzoff ED: 1.4 Å structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry* 1995, 34:6278-6287.

9. Rost B, Sander C: Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 1994, 19:55-72.
10. Gekakis N, Saez L, Delahaye-Brown AM, Myers MP, Sehgal A, Young MW, Weitz CJ: Isolation of timeless by PER protein-interaction: defective interaction between timeless protein and long-period mutant PERL. *Science* 1995, 270:811-815.
11. Huang ZJ, Edery I, Rosbash M: PAS is a dimerization domain common to ***Drosophila Period*** and several transcription factors. *Nature* 1993, 364:259-262.
12. Lindebro MC, Poellinger L, Whitelaw ML: Protein–protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor–Arnt transcription factor complex. *EMBO J* 1995, 14:3528-3539.
13. Edgerton MD, Jones AM: Subunit interactions in the carboxy-terminal domain of phytochrome. *Biochemistry* 1993, 32:8239-8245.
14. Hill S, Austin S, Eydmann T, Jones T, Dixon R: ***Azotobacter vinelandii*** NIFL is a flavoprotein that modulates transcriptional activation of nitrogen-fixation genes via a redox-sensitive switch. *Proc Natl Acad Sci USA* 1996, 93:2143-2148.
15. Monson EK, Weinstein M, Ditta GS, Helinski DR: The FixL protein of ***Rhizobium meliloti*** can be separated into a heme-binding oxygen-sensing domain and a functional carboxyl-terminal kinase domain. *Proc Natl Acad Sci USA* 1992, 89:4280-4284.
16. Guillemin K, Krasnow MA: The hypoxic response: Huffing and HIFing. *Cell* 1997, 89:9-12.
17. Bibikov SI, Biran R, Rudd KE, Parkinson JS: A signal transducer for aerotaxis in ***Escherichia coli***. *J Bacteriol* 1997, 179:4075-4079.
18. Baker PJ, Britton KL, Rice DW, Rob A, Stillman TJ: Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold — Implications for nucleotide specificity. *J Mol Biol* 1992, 228:662-671.
19. Schuler GD, Altschul SF, Lipman DJ: A workbench for multiple alignment construction and analysis. *Proteins* 1991, 9:180-190.
20. Altschul SF, Gish W: Local alignment statistics. *Methods Enzymol* 1996, 266:460-480.
21. Birney E, Thompson JD, Gibson TJ: PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. *Nucleic Acids Res* 1996, 24:2730-2739.
22. Eddy SR: Hidden Markov models. *Curr Opin Struct Biol* 1996, 6:361-365.

Addresses: *University of Oxford, The Old Observatory, Fibrinolysis Research Unit, South Parks Road, Oxford OX1 3RH, UK. †Department of Biology - BSBW, Texas A&M University, College Station, Texas 77843, USA.
Correspondence: Chris P. Ponting
E-mail: Ponting@bioch.ox.ac.uk