# The diversity of globin-coupled sensors

Tracey Allen K. Freitas<sup>a,b</sup>, Shaobin Hou<sup>a</sup>, Maqsudul Alam<sup>a,b,\*</sup>

<sup>a</sup>Department of Microbiology, Snyder Hall 207, 2538 The Mall, University of Hawaii, Honolulu, HI 96822, USA <sup>b</sup>Maui High Performance Computing Center, 550 Lipoa Parkway, Kihei, Maui, HI 96753, USA

Received 2 July 2003; revised 5 August 2003; accepted 10 August 2003

First published online 5 September 2003

Edited by Horst Feldmann

Abstract The recently discovered globin-coupled sensors (GCSs) are heme-containing two-domain transducers distinct from the PAS domain superfamily. We have identified an additional 22 GCSs with varying multi-domain C-terminal transmitters through a search of the complete and incomplete microbial genome datasets. The GCS superfamily is composed of two major subfamilies: the aerotactic and gene regulators. We postulate the existence of protoglobin in Archaea as the predecessor to the chimeric GCS.

© 2003 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.

*Key words:* Globin; Archaeon; Heme-based sensor; Globin-coupled sensor

#### 1. Introduction

Homo- and heteromeric heme-based sensors are mediators of cellular responses to metabolic and environmental stimuli such as NO, CO and  $O_2$  [1]. Changes in intracellular gas concentrations are sensed by a heme moiety and result in either aerotaxis or gene regulation. Presently, there are six known types of heme sensors: CooA, NPAS2, sGC, Dos and *Ax*PDEA1, FixL, and HemAT. The HemATs, by homology, are the only aerotactic heme sensors combining globin and MCP signaling domains, whereas the remaining function in gene regulation, either by binding DNA directly, modulating a small metabolite 2nd messenger (cyclic mono- and di-NMPs), or directly interacting with a transcription factor or regulator.

CooA is a CO sensor that controls the transcription of COutilizing genes. Binding of CO to the heme domain of CooA homodimers modulates the DNA-binding C-terminal domain [2]. Neuronal PAS domain protein 2 (NPAS2) is expressed in mammalian brain tissue [3] and regulates transcription as a heterodimer with BMAL1 [4–6]. Dissociation of the NPAS2:BMAL1 heterodimer occurs upon CO binding to the NPAS2 monomer, effectively removing its DNA-binding, and hence, transcription capability [3]. The soluble guanylate cyclase (sGC) contains a heme-binding and guanylate cyclase domain. Binding of NO to the sGC heterodimer produces cGMP from GTP [7], whereby gene regulation ensues by the cGMP 2nd messenger. The direct oxygen sensor (Dos), first described in *Escherichia coli* [8], functions as a tetrameric phosphodiesterase (PDE) by converting cAMP to 5'-AMP while in the ferrous form, and is strongly inhibited by CO and NO ligands [9]. A1 from *Acetobacter xylinum* (*AxP*-DEA1) also functions as a PDE by linearizing cyclic bis( $3' \rightarrow 5'$ )diguanylate, an allosteric activator of the bacterial cellulose synthase, to the ineffectual pGpG [10,11]. Both Dos and *AxPDEA1* possess similar heme-binding PAS domains fused to the PDE C-terminus, consisting of a GGDEF and EAL domain. Histidine kinase FixL binds heme at an N-terminal PAS domain and controls transcription of oxygen-sensitive genes by its response regulator, FixJ [13,14]. Phosphorylated FixJ acts as the transcriptional activator and permits transcription of the *fix* genes [15,16].

Heme-based aerotaxis transducers, the HemATs, possess a heme-binding globin domain and a signaling domain typical of methyl-accepting chemotaxis proteins (MCP) [17]. Hem-ATs, originally discovered in the archaeon *Halobacterium salinarum* and the Firmicutes *Bacillus subtilis*, are members of the family of globin-coupled sensors (GCSs) [18,19]. Variance in the C-terminal transmitter domain indicates that not all GCSs are involved in aerotaxis. In this report, we further identify the diversity of these GCSs resulting from exhaustive searches of completed and in-progress microbial genomes. We also report their putative functions and categorize them in relation to other non-globin heme-based sensors and propose two possible evolutionary models of the GCS and globin.

#### 2. Materials and methods

#### 2.1. Genome and protein sequences

The following preliminary sequence data was obtained from the Institute for Genomic Research website: Acidithiobacillus ferrooxidans, Bacillus anthracis, Bacillus cereus, Carboxydothermus hydrogenoformans, and Geobacter sulfurreducens; DOE Joint Genome Institute: Azotobacter vinelandii, Burkholderia fungorum, Geobacter metallireducens, Magnetococcus, Magnetospirillum magnetotacticum, Rhodobacter sphaeroides, Rhodospirillum rubrum, and Novosphingobium aromaticivorans; National Center for Biotechnology Information: Escherichia coli O157 H7, Halobacterium salinarum, Agrobacterium tumefaciens, Caulobacter crescentus, Bacillus halodurans, Bacillus subtilis, Vibrio vulnificus, and Shigella flexneri; the Bordetella pertussis, Bordetella parapertussis, and Bordetella bronchiseptica sequence data was produced by the Bordetella pertussis Sequencing Group at the Sanger Institute and can

<sup>\*</sup>Corresponding author. Fax: (1)-808-956 5339.

E-mail address: alam@hawaii.edu (M. Alam).

be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/bp/. At present, five genomes are incompletely sequenced and therefore accession numbers are not available for those proteins (see Table 1 for details).

#### 2.2. Multiple alignments and secondary structure

All sequences were aligned in a two-stage process. Multiple alignments in ClustalX v1.8 [20] were followed by manual adjustment in DNAStar's MegAlign. At this stage, globin crystal structures (*E. coli* HMP, PDB ID: 1GVH; *Vitreoscilla stercoraria* Hb, PDB ID: 1VHB; *Ralstonia eutropha* FHb, PDB ID: 1CQX; *Chlamydomonas eugametos* trHb, PDB ID: 1DLY; *Paramecium caudatum* trHb, PDB ID: 1DLW; HemAT-*Bs*, PDB ID: 1OR6) and Jnet [21] secondary structure predictions were used as guides to produce the finished alignments in Fig. 1A.

#### 2.3. Protein domain detection and analyses

Protein sequences were analyzed with the Pfam (http:// pfam.wustl.edu/), SMART (http://smart.embl-heidelberg.de/), and SCOP (http://scop.berkeley.edu/) datasets and domain descriptions were taken from the InterPro database (http://

		Zhelix Ahelix Bhelix Chelix Ehelix	
HemAT-Bs Structure			-
		70 80 90 100 110 120 130 140 15	i0
B. subtilis	30	KHADVKKQLKMVRLGDAELYVLEQLQPLIQENIVNTVDA KNLDHESSLADIIN-DHSSVDRLGDTLKRIIQE-MFACVI	D - C
B. halodurans	33	AESELSAQLRMIHLTILDDLKRMKALQPLVEENMEVLADA ASNIIKQPNLNEIIETHSSVERLZCTLKQIILE-MENTEI	D -
B. anthracis	31	TNSELKVQMDMLHISKEDLQIVKVLQPFIYEEIDWITEK ANITKQPNLITIIERYSSIPKLQUTLKTIKE-LFSED	9H -
B. cereus	31	TNSELKIQMDMLHISKEDLQIVKVLQPFIYEEIDWITEK SNITKQPNLITIIERYSSIPKLKUTIKE-LFS:DM	0H -
C. hydrogenoio mans	8	SRYQETLSFLNLTAEDLOLMAFFRELFIOKAQEFVNK OHLTKFPYLQELIKKHSTVEKLSKTQAE FIS-LTSEK	D -
H. Salinarum	29	DEAE I AVALSET GIDDD I MAALA AEQULFEE AT AD ALVID O DILESSERTUDLFAN-SIKTVE ULST QAE LLG-LGRESS	- C
Magnetococcus1	23	DHIEUMKRFVGFTEKU ASILKKLERV AAKHATAVNIT TILSUFAHLEKI IGG-AGSSVERLATIGEE LVU-LFUGE	- U
M. magnetotacticum2	94		- U
C craccaptuci	12	EPTRECIDENT DEST DU DU DET CAN DU CANDATERERE DE DE ASCOLO AUCTORISTICA CON	- v
A timefacienc	13	ERINE MURANESSERVICE TO THE OTHER AND THE REFS - STATUTE TO STATUTE TO THE S	- W
C crescentus?	50		G-
R. sphaeroides	9	FRIRELATE ADD ATRALING AVERALDR FRMRITS ABG-FFA-D ATHM SA SERVICE WAR-LAS FEL	D -
W aromaticino rane	12	DKL AFFWIDHKD FEDROWI AKUF FWA PRALDKI, VOOTATTUETASTERSSDAMDHA DKKOTE WAG MES PR	- D
R Tubran	8		r
V. vulnificus	9	KVED ADELLKI, HDL/TEAD LALTRIKE GOT MVPKL DEVKH - DVI. RNTPEYEOYEG D AOKLORVODS OVR. WKT-FED RI	D -
A. ferrooxidans	4	DNSGTLPAFLGLODSDFOVTORYRDALDKEASALAHA DYLLSHPATAAVFRDFSSARLDALIOKOTE AKG-LLV-RL	D -
S. flexneri	9	KDEWTGLVEGAD PLIRAKAAE IALAHAHYLSI B RIVRID PHAEEFLSNEGVER OL SSAMER I IN-VLSTOV	
E. coli	9	KDEWTGLVEOAD PPTRAKAAEIAVAHAHYLSIE RIVRIDPHAEEFLSNEOVEROLKSAMER IIN-VLSTOV	
A. vinelandii	18	AREVILL GOF PAPVVAO TRELATINO SELPGY COMLODE CAMLELTHEOVKSRLHETLRO IVS-VESMSE	DD
B. pertussis	64	ALRYKDT CAHYS PHEW AARNVYTANK AALADY ECMLADPNAAFFLSDOLVKTKLHASMODILES-VY ATAF	-T
B. parapertussis	47	ALRIKOT CAHYS PHEW AARNYY TANK AALAD Y CALADPNAAFFLSD OLVKTKLHASMOD LES-YY A AF	-T
B. bronchiseptica	64	ALRIKOT CAHYS PHEWA BARNVVTANK AALAD Y CEMLAD PNAAFFLSD QLVKTKL HASMOD LES-VY A AF	P-T
Magnetococcus2	9	EQRLKDIYLGVD AEKVNFIGDLIKDRLNQTVER TELLEVE SARFFLDSALVKERLHGSLIE LQM-LFSHKD	D-D
G. metallireducens	6	EIKAHYLFGDEDAETLKSLLSIAQANRELMIED WYLLGIPETAAFLQDDTVLQRLKLSHGGYFVN-LFREVY	
G. sulfurreducens	6	EIKAHYRFTDED AELLGSLFPL AETNKERLAD 0 DYLLGIPETAEFLKEDLVLOKLKOTHOD FVS-LF & SY	
B. fungorum	20	GSHL <mark>Y</mark> SQARASALTSLTEVLRINAVE <mark>T</mark> VKR DGLTRLPKSKHTLAALSEHELQHLKTQQIQNLYALASPDLT	A-
		Fhelix Ghelix Hhelix	
HemAT-Bs Structure		F helix G helix H helix	
Hem&T-Bs Structure	111	F helix  G helix  H helix    160  170  180  190  200  210  220  230    160  170  180  190  200  210  220  230	Tos
HemāT-Bs Structure B. subtilis B. baladurans	111	F helix G helix H helix 160 170 180 190 200 210 220 230 	QS
HemāT-Bs Structure B. subtilis B. halodurans B. antbracis	111 114 112	F helix G helix H helix 160 170 180 190 200 210 220 230 DEFICKENET SI LOT IL PKYMIG FOELLISHIDIYE ASITN	QS EN ES
HemāT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus	111 114 112 112	F helix G helix H helix 160 170 180 190 200 210 220 230 DFFL KRNRTSIL TI LEYKWAG FOELLISHIDIYE ASITMQUELLKAIKAT TÄTLHLEQULVLE OFFL KRNRTSIL GI UKWAYS FOOLTOSL JULE QSSDYSA REIVLATRSIL KILHLEQULVLE ENFFL ORWAT KR VOIL UHRVAYTA YUELFASIMKILK TUVE	QS EN ES
HemùT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans	111 114 112 112 87	F helix G helix H helix 160 170 180 190 200 210 220 230 	QS EN ES IA
HemMT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans M. salimarum	111 114 112 112 87 111	F helix  G helix  H helix    160  170  180  190  200  210  220  230   DEFICKENEL'SI LCT LL PKWMG FOELLISMIDIYE ASITN  OELLKAIKAT TÄULNLEQULUER  OELLKAIKAT TÄULNLEQULUER   DEFICKENEL'SI LCT LL PKWMG FOELLISMIDIYE ASITN  OELLKAIKAT TÄULNLEQULUER  OELLKAIKAT TÄULNLEQULUER   DEFICKENEL'SI LCT LL PKWMG FOELLISMIDIYE ASITN  OELLEVIA  OELLEVIA   DEFIC DEVKTIKE VOIL HAKKMT A VOELFESIKKILKTOVE  SDFSYSINVIN SILLEDELVIA   ENFIE DEVKTIKE VOIL HAKKMT A VOELFESIKKILKTSSD YSA AEFSYSINVIN SILLEDELVIA	QS EN ES IA ID
HemàT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans N. salinarum Magnetococcusi	111 114 112 112 87 111 103	F helix  G helix  H helix    160  170  180  190  200  210  230   DEFTEKRRETST LET LLPKYMG FUELLSMEDTYERSTIN  OELLSMEDTYERSTIN  OELLKRIKATTELALEDUVLE   DEFTEKRRETST LET LUPKYMG FUELLSMEDTYERSTIN  OELLKRIKATTELALEDUVLE   DEFTEKRRETST LET LUPKYMG FUELLSMEDTYERSTIN  OELLKRIKATTELALEDUVLE   DEFTEKRRETST LET LUPKYMG FUELFRSINKIKTSSD  SDFSYSIRVINS LFILDEUVLA   DEFTEKRRETST VUI LUPKYMG FUELFRSINKIKTSSD	QS EN ES IA ID
HemāT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans H. salinarum Magnetococusi M. magnetococusi	111 114 112 112 87 111 103 83	F helix G helix H helix 160 170 180 190 200 210 220 230 	QS EN ES ES IA D E
HemAT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans M. salinarum Magnetococcus1 M. magnetotacticum2 M. magnetotacticum3	111 114 112 112 87 111 103 83 105	F helix G helix H helix 160 170 180 190 200 210 220 230 DEFIEKRNRI SI LEI LLPKWMG FOELLISMIDIYE ASIIN	QS EN ES IA D ES IA ES IA ES IA
HemàT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans H. salinarum Magnetococcus M. magnetotacticum M. magnetotacticum C. crescentus	111 114 112 112 87 111 103 83 105 88	F helix G helix H helix 160 170 180 190 200 210 220 230 OFFIC KURKIST LET LIL PKYMAG FUELLISHIDIYE ASITMQUELKAIKAT TSTILLIEDULVLE OFFIC KURKIST LET LIL PKYMAG FUELLISHIDIYE ASITM	QS EN ES IA ID ID IE IE LE
HemàT-Bs Structure B. subtilis B. halodarans B. anthracis B. cereus C. hydrogenoformans N. salinarum Nagnetococcus1 M. magnetotacticum1 C. crescentus1 A. tumefaciens	111 114 112 112 87 111 103 83 105 88 88	F helix G helix H helix 160 170 180 190 200 210 220 230 DEFIE KERRETST LETTIL PENDIG FORLILSMIDITE ASITNQUELLKATKAT TÄLLALEOQUVLE DEFIE OKWAT GK VOIL URGWAG FORLILSMIDITE ASITNQUELLKATKAT TÄLLALEOQUVLE DEFIE OKWAT GK VOIL URGWAG FORLISSING VICTURESSING	QS EN ES ES IA D E E E L E
HemAT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans M. salinarum Magnetotacticum M. magnetotacticum M. magnetotacticum M. magnetotacticum A. tumefaciens C. crescentus A. tumefaciens	111 114 112 112 87 111 103 83 105 88 88 88 125	F helix  G helix  H helix    160  170  180  190  200  210  220  230   DEFIEKRNRI SI LEILEKNYMG FUELLSMIDIYERSIN  OUELLKAIKAT TÄILMLEQULLER  OUELLKAIKAT TÄILMLEQULLER   DEFIEKRNRI SI LEILEKNYMG FUELLSMIDIYERSIN  OUELLKAIKAT TÄILMLEQULLER  SDFSYSINVIN SILLULEUSSD  SDFSYSINVIN SILLULEUSSD   DEFIE QRVATIKR VOI LHRKWYTR YUELFRSIMKIIKUVE  SDFSYSINVIN SILLULEUSSD  SDFSYSINVIN SILLULEUSVIN   DEFIE QRVATIKR VOI LHRKWYTR YUELFRSIMKIIKUSSD SDFSYSINVIN SILLUEUSVIN  SDFSYSINVIN SILLUEUSVIN   ENFIE QRVATIKR VOI LHRKWYTR YUELFRSIMKIKKISSD	QS EN ES ES IA D D E E E L E D D E E ID D
HemàT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans H. salinarum Magnetococcus M. magnetotacticum M. magnetotacticum C. crescentus A. tumefaciens C. crescentus R. sphaeroides	111 114 112 112 87 111 103 83 105 88 88 125 82	F helix G helix H helix 160 170 180 190 200 210 220 230 OFFIC KURKITST LTL IL FKNMAG FOELLISHIDTVE ASITMOUELKAIKATTSTILHLEOOLVLE OFFIC KURKITST LTL TKINAG FOELLISHIDTVE ASITMOUELKAIKATTSTILHLEOOLVLE DFFIC GKVAT GA VII LUTKYNYT A YOLFFSILKILK TSD	QS EN ES ES IA DD EE EE LD DD FD
HemàT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans Magnetococcus M. magnetotacticum C. crescentus A. tumefaciens C. crescentus R. sphaeroides N. aromaticivorans	111 114 112 112 111 103 83 105 88 88 88 125 82 87	F helix G helix H helix 160 170 180 190 200 210 220 230 DEFIEKRNRITSI LITIL PKYRMG FUELLISMIDIYE RSITNQUELLKAIKATTKILALEQUIVLE DEFIEKRNRITSI LITIL PKYRMG FUELLISMIDIYE RSITNQUELLKAIKATTKILALEQUIVLE DEFIE QKWICK K VOI LURKKYTR YUELFNSINKILK VIVESDFSYSINVIN LITLEQUIVLE 	QS EN ES ES IA DD EE LE LD DF FR
HemàT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans M. salinarum Magnetococcusi M. magnetotacticum M. magnetotacticum M. magnetotacticum C. crescentusi A. tunnefaciens C. crescentus R. sphaeroides M. aromaticivorans M. aromaticivorans M. aromaticivorans	111 114 112 87 111 103 87 103 105 88 88 125 88 88 125 87 81	F helix G helix H helix 160 170 180 190 200 210 220 230 DEFIEKRNRICSI LET LLPKWMG FOELLLSMIDIYERSITNQUELLKAIKATTSILMLEQUVLER QAFLOKKLOUT OR WIT LOTKNYVS FOQLIDSLIDLLEQSD	QS EN ES ES IA D E E ES IA D E E E F F F F A
HemàT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenofoimans M. salinarum Magnetococcusi M. magnetotacticum2 M. magnetotacticum2 M. magnetotacticum3 A. tumefaciens C. crescentus1 A. tumefaciens C. crescentus2 R. sphaeroides M. aromaticivorans R. rubrum V. vulnificus	111 114 112 112 87 111 103 87 103 88 88 125 88 88 125 82 87 81 90	F helix G helix H helix 160 170 180 190 200 210 220 230 	QS EN ES ES IA D D IE IE LD D FD FR FA SR
Hemàl-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans H. salinarum Magnetotacticum I. magnetotacticum C. crescentus A. tumefaciens C. crescentus R. sphaeroides N. aromaticivorans R. rubrum V. vulnificus A. ferrooxidans	1111 114 112 112 87 1111 103 83 105 88 88 125 82 87 81 90 85	F helix G helix H helix 160 170 180 190 200 210 220 230 DEFIEKRNRITSI LITIL PKYRMG FUELLISMIDIYERSITNQUELLKAIKATTKILMLEQUVLE DEFIEKRNRITSI LITIL PKYRMG FUELLISMIDIYERSITNQUELLKAIKATTKILMLEQUVLE DEFIE QKWITCK VOI LURKKYTR YUELFSINKILK VVE 	QS EN ES ES IA DD DE EE LD DD FD FR FA SR
HemàT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans N. salinarum Magnetococcus M. magnetotacticum M. magnetotacticum M. magnetotacticum C. crescentus A. tunnefaciens C. crescentus R. sphaeroides N. aromaticivorans R. rubrum V. vulnificus A. ferrooxidans S. flexmeri	1111 114 112 112 87 111 103 83 105 88 88 83 125 82 87 81 90 85 83	F helix G helix H helix 160 170 180 190 200 210 220 230 	QS EN ES ES IA DD DE EE LD DD FD FR FA SR AH
HemàT-Bs Structure B. subtilis B. halodarans B. anthracis B. cereus C. hydrogenofoimans H. salinarum Magnetococcus M. magnetotacticum M. magnetotacticum C. crescentus A. tumefaciens C. crescentus A. tumefaciens C. crescentus A. tumefaciens C. sphaeroides M. aromaticivorans R. rubrum V. vulnificus A. ferrooxidans S. flexmeri E. coli	1111 114 112 112 83 103 83 105 88 88 88 125 82 87 81 90 85 83 83 83	F helix 6 helix H helix 160 170 180 190 200 210 220 230 	QS ENSES IADDE IE LE LDDF FRASRAH TF
HemàT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans Magnetococcus1 M. magnetotacticum2 M. magnetotacticum2 C. crescentus1 A. tunefaciens C. crescentus2 R. sphaeroides M. aromaticivorans R. rubrum V. vulnificus A. fernoxidans S. flexmeri E. coli A. vinelandii	1111 114 112 87 1111 103 88 88 88 88 82 87 81 25 87 81 90 85 83 99	F helix G helix H helix 160 170 180 190 200 210 220 230 DEFIEKRNRIT SI LET IL PKNPMG FUELLISMIDIYE RSITNQUELLKAIKATTKILMLEQUVLES DEFIEKRNRIT SI LET IL PKNPMG FUELLISMIDIYE RSITN	QS EN ES ES IA D D E E E LE LD D F P F F A H TF TF SD
HemàT-Bs Structure B. subtilis B. halodurans B. anthracis B. cercus C. hydrogenoformans M. salinarum Magnetococcus M. magnetotacticum M. magnetotacticum C. crescentus A. tunefaciens C. crescentus R. sphæroides M. aromaticivorans R. rubrum V. vulnificus A. ferrooxidans S. flexmeri E. coli A. vinelandii B. pertussis	1111 114 112 87 111 103 83 88 88 88 81 25 82 87 81 82 87 81 85 83 83 83 83 83 83 83 83 83 83 83 83 83	F helix G helix H helix 160 170 180 190 200 210 220 230 	OS EN ES ES ES ES ES ES ES ES ES ES ES ES ES
HemAT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans Magnetococcus1 M. magnetotacticum2 M. magnetotacticum2 C. crescentus1 A. tumefaciens C. crescentus2 R. sphaeroides M. aromaticivorans R. rubrum V. vulnificus A. ferrooxidans S. flexmeri E. coli A. vinelandii B. paragertussis B. paragertussis	1111 114 112 87 1111 103 88 88 82 87 81 90 85 82 87 81 90 83 83 94 9122	F helix 6 helix H helix 160 170 180 190 200 210 220 230 DEFTE KRNRTIST LTT 1L PKNMAG FOULLISHTOTE ASITMQUELKATKAT TÄTLALEOQUVLES DEFTE GRWAT GR VÖL LOTKMYS FOOLTDSL JULE QSSD	QS EN ES ES IA DD E E E LD DD E FRA SRA FF F SO SV SV
HemàT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans H. salinarum Magnetococcus I. magnetotacticum M. magnetotacticum I. magnetotacticum C. crescentus I. crescentus C. crescentus R. sphaeroides M. aromaticivorans R. rubrum V. vulnificus A. fernoxidans S. flexmeri E. coli A. vinelandii B. partussis B. parcussis B. parchiseptica	1111 114 112 87 1111 103 88 88 88 81 85 82 87 81 94 139 94 139 94 122 139	F helix G helix H helix 160 170 180 190 200 210 220 230 	OS EN ES ES A D D E E E D D E E FA SR A F F SO SV
HemilT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans H. salinarum Magnetococcus1 M. magnetotacticum2 M. magnetotacticum2 C. crescentus1 A. tumefaciens C. crescentus2 R. sphaeroides H. aromaticivorans R. rubrum V. vulnificus A. ferrooxidans S. flexmeri E. coli A. vinelandii B. pertussis B. paragertussis B. prochiseptica Magnetococcus2	1111 114 112 87 1111 103 88 88 88 81 255 82 87 81 90 90 85 83 83 83 83 83 83 83 94 139 122 84 23 98 24 23 94 23 98 24 23 98 24 23 98 24 23 24 24 24 24 24 24 24 24 24 24 24 24 24	F helix G helix H helix 160 170 180 190 200 210 220 230 	OS ENSES AD D EE EE LD D EP FRA SRAH FF FS SV SV VR
HemAT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans M. salinarum Magnetotacticum M. magnetotacticum M. magnetotacticum M. magnetotacticum M. magnetotacticum M. magnetotacticum M. magnetotacticum M. subartum C. crescentus A. tumefaciens C. crescentus R. sphaeroides M. aromaticivorans R. rubrum V. vulnificus A. ferrooxidans S. flexmeri E. coli A. vinelandii B. paragertussis B. bronchiseptica Magnetococcus G. metallireducens	1111 114 112 112 113 87 113 103 88 88 81 90 85 82 87 83 83 90 85 83 83 91 22 139 9122 139 84 88 80 85 83 83 83 83 83 83 83 83 83 83 83 83 83	F helix G helix H helix 160 170 180 190 200 210 220 230 DEFIE KRNRITSI LITIL PKYRMG FUELLISMIDTYE RSITNQUELLKAIKAT TÄTLALEOULVLE DEFIE KRNRITSI LITIL PKYRMG FUELLISMIDTYE RSITN	OS ENESSIA DO DE DE LO DO PRASA TE TESOS VIVE DE
HemäT-Bs Structure B. subtilis B. halodurans B. anthracis B. cereus C. hydrogenoformans Magnetococcus1 M. magnetotacticum2 M. magnetotacticum2 M. magnetotacticum2 M. magnetotacticum2 M. magnetotacticum2 R. sphaeroides M. aromaticivorans R. rubrum V. vulnificus A. ferrooxidans S. flexneri E. coli A. vinelandii B. parapertussis B. bronchiseptica Magnetococcus? G. metallireducens G. sulturreducens	111 114 112 87 111 103 88 88 81 25 82 87 81 90 85 83 83 83 83 83 83 83 83 83 83 83 83 83	F helix G helix H helix 160 170 180 190 200 210 220 230 	OS ENSESSION DE LE LO DE FRANCE ANTE SO SV SV REE ES

consensus/85%

... pthht ... t1 p. hcls1t. phh. tpht. h.p. 1ht. 1. tp...... tp..... tp..... th. c. h. .s. ph. .ts

Fig. 1. Diversity of GCSs. The structural alignment (A) and the phylogenetic tree (B) of the GCS globin domain. A: The structural alignment of the globin domains from 27 GCSs was created in ClustalX and MegAlign and includes the 2D structure of the recent HemAT-*Bs* crystal structure (PDB ID: 10R6) (personal communication) as a reference. The traditional helical assignments are maintained as helices A through H, with an additional Z helix at the N-terminus. The asterisk (\*) indicates the conserved proximal histidine. Amino acid conservation has been based on an 85% consensus sequence and colors are assigned to amino acid groups as follows: charged (c, DEHKR) in white on blue back-ground; polar (p, KRHEDQNST) in red; turn-like (t, ACDEGHKNPQRST) in green; bulky hydrophobic (h, ACLIVMHYFW) and aliphatic (l, LIVM) in yellow; aromatic (a, FHWY) in white on pink background; small (s, ACDGNPSTV) in purple; and tiny (u, AGS) in white on purple background. B: The phylogenetic tree is based on the alignment of part A of this figure with branches grouping according to transmitter type. Branches supported with bootstrap values > 5000 are indicated. Taxonomic listings for the GCS-containing organisms are listed with the organisms' names colored according to the type of transmitter domain. Pink, GAF:EAL; orange, unclassified; blue, ERERQR:GGDEF; purple, GGDEF:EAL; green, STAS; red, MCP or HAMP:MCP.

www.ebi.ac.uk/interpro). Various BLAST and PSI-BLAST searches were performed against the non-redundant database and the microbial database at the National Center for Bio-technology Information (http://www.ncbi.nih.gov/BLAST/). Transmembrane regions were identified by the algorithms TMHMM2 and DAS (http://www.cbs.dtu.dk/services/TMHMM-2.0/ and http://www.sbc.su.se/~miklos/DAS/).

#### 2.4. Phylogenetic analyses

The distance tree was created using the neighbor-joining (ClustalX) method. Bootstraps (10000 replicates) were calculated directly in ClustalX. Trees were generated in TreeView and NJPlot (distributed with the ClustalX package) and further refined in Adobe Illustrator 10.

#### 3. Results and discussion

An exhaustive heuristic search of the non-redundant protein database and (un)finished microbial genome database at NCBI yielded 27 GCSs. Criteria for identifying a putative GCS included a primary match with the globin domain followed by an accompanying transmitter domain(s). In addition, the length of the globin domain was taken into account as well as the presence of a proximal histidine. In almost all cases, a hydrophobic aromatic residue pair at the end of the B helix (usually Phe-Tyr) was also present. Secondary structure predicting algorithms and the 3D-PSSM fold-recognition server were used to support their inclusion into the family. Using (PSI)BLAST as the primary search algorithm, once a GCS was identified, it was added to the seed alignment. Since the GCS globin domains are highly divergent, each GCS sequence



Fig. 1 (Continued).

Neither the SMART database nor the manually curated Pfam-A dataset recognizes the GCS globin domain yet, though the automatically generated Pfam-B family 7730 has an incomplete and partially incorrect (on the basis of the above criteria) GCS globin domain dataset. Fig. 1A represents the alignment of the globin domain of all 27 GCSs. The resulting Neighbor-joining phylogenetic tree was created based on this alignment and is presented in Fig. 1B.

#### 3.1. Biological heme-sensor classification

Using the identified functions of CooA, NPAS2, sGC, Dos, *Ax*PDEA1, FixL, and HemATs, all currently identified biological heme-based sensors can be classified as either aerotactic or gene regulating. Gene regulation is observed to occur via one of three different pathways: via protein–DNA interaction [2–6], via modulation of small-metabolite 2nd messengers [7–12], or by protein–protein interaction as in a transcription factor or regulator [13–16]. The resulting organization schema is illustrated in Fig. 2. GCSs are found in organisms with various physiological and metabolic systems: Gram-positive and Gram-negative, aerobic and anaerobic, oxic and anoxic phototrophs, and even a nitrogen fixer (*A. vinelandii*).

3.1.1. Aerotactic. HemATs are the only known hemebased aerotaxis sensors [17,18] and approximately half of the predicted GCSs are HemATs. Each possess an N-terminal globin domain and a C-terminal MCP-like domain. The original HemAT signaling domain was classified as an  $\sim$  MCP [17]; however, additional HemATs exhibit a  $\sim$  HAMP:MCP module. Such a combination is typical of transmitter regions of methyl-accepting chemotaxis proteins such as the *E. coli* serine receptor, Tsr, and hence these proteins may mediate aerotaxis as well. All HemATs are soluble proteins.

The aerotactic subfamily is predominantly Gram-negative  $\alpha$ -Proteobacteria (nine proteins), but also includes the Firmicutes (five proteins) and one Archaea. In particular, the magnetotactic proteobacterium *M. magnetotacticum* possesses two aerotactic transducers, whereas *Magnetococcus* MC-1 cells possess only one. Magnetotaxis has been shown to work in conjunction with aerotaxis [22]. Though only a single Archaeal transducer has been found, this is not surprising since at least half of the sequenced Archaeal genomes do not contain recognizable taxis genes. Moreover, the representative sample size of the Archaeal genomes (one GCS out of 18 genomes  $\sim 6\%$ ) is miniscule compared to that of the bacterial genomes (26 GCSs out of 228 genomes  $\sim 11\%$ ).

3.1.2. Modulation of a 2nd messenger. Proteins possessing the GGDEF domain have been implicated in c-diGMP modulation [23] and eight such proteins were identified in this group, incorporating either the GGDEF domain or a GGDE-F:EAL domain pair. Closer inspection of these proteins reveals another highly conserved domain centered between the N-terminal globin sensor and the C-terminal GGDEF domain. This new domain has been designated as ERERQR, after a conserved patch of residues ( $\geq 85\%$  of five acidic, seven basic, 32 polar and 25 hydrophobic sites in a primarily alpha helical and coiled structure, data not shown). Af-GReg2M has the exact C-terminal domain organization as EcDos and AxPDEA1 (~GGDEF:EAL), PDEs that inactivate the 2nd messengers cAMP and c-diGMP, respectively. The GCS from B. fungorum (BfGReg) possesses a C-terminal

Table 1				
Source information	and	classification	of	GCSs

No.	Organism	Name	NCBI accession	Classification	SMART	Pfam	Taxonomy	Protein	Other name
			no.					length	
1	Bacillus anthracis	HemAT-Ba	NP_653892	Aerotactic	MA	МСР	Firmicutes	434	BA_0532
2	Bacillus halodurans	HemAT-Bh	NP_241371	Aerotactic	MA	MCP	Firmicutes	441	BH505
3	Bacillus subtilis	HemAT-Bs	NP 388919	Aerotactic	MA	MCP	Firmicutes	433	YhfV
4	Bacillus cereus	HemAT-Bc	NP 835085	Aerotactic	MA	MCP	Firmicutes	434	_
5	Carboxvdothermus	HemAT-Ch	TIGR 129958	Aerotactic	MA	MCP	Firmicutes	251	_
	hvdrogenoformans								
6	Halobacterium sp. NRC-1	HemAT-Hs	NP_280321	Aerotactic	MA	МСР	Archaea	490	HtrX, HtB, Htr10
7	Magnetospirillum	HemAT-MmA	ZP_00054774	Aerotactic	MA	МСР	α-Proteobacteria	444	Magn7582
0	Magnetotacticum	Ham AT Man D	<b>7D</b> 00054075	A	МА	MCD	or Ducto chartonia	722	Mara (9/7
8	magnetospirilium magnetotacticum	HemA1-MmB	ZP_00054075	Aerotactic	MA	МСР	α-Proteobacteria	/32	Magn6867
9	Rhodobacter sphaeroides	HemAT-Rs	ZP_00006252	Aerotactic	MA	MCP	α-Proteobacteria	371	Rsph2166
10	Rhodospirillum rubrum	HemAT-Rr	ZP_00014161	Aerotactic	MA	MCP	α-Proteobacteria	442	Rrub1164
11	Agrobacterium tumefaciens	HemAT-At	NP_354049	Aerotactic	HAMP-MA	HAMP-MCP	α-Proteobacteria	500	AGR_C_1888
12	Caulobacter crescentus	McpB	NP_419247	Aerotactic	HAMP-MA	HAMP-MCP	α-Proteobacteria	538	McpB
13	Caulobacter crescentus	McpM	NP_421120	Aerotactic	HAMP-MA	HAMP-MCP	α-Proteobacteria	556	McpM
14	Novosphingobium	HemAT-Na	ZP_00095064	Aerotactic	HAMP-MA	HAMP-MCP	α-Proteobacteria	482	Saro2089
	aromaticivorans								
15	Magnetococcus sp. MC-1	HemAT-Mg	ZP_00043038	Aerotactic	HAMP-MA	HAMP-MCP	α-Proteobaceria	519	Mmc10749
16	Magnetococcus sp. MC-1	MgGReg	ZP_00042662	Gene regulator	EREROR:DUF1	EREROR:GGDEF	α-Proteobaceria	467	Mmc10355
	0 1	0 0	-	(2nd messenger)					
17	Bordetella bronchisentica	<i>Bb</i> GReg	n/a	Gene regulator	EREROR:DUF2	EREROR:GGDEF	β-Proteobacteria	531	_
	I IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII			(2nd messenger)			F		
18	Bordetella parapertussis	<i>Bna</i> GReg	n/a	Gene regulator	EREROR: DUF3	EREROR:GGDEF	β-Proteobacteria	514	_
		-1.01118		(2nd messenger)			P		
19	Bordetella pertussis	<b>BneGReg</b>	n/a	Gene regulator	EREROR DUF4	EREROR	<b>B-Proteobacteria</b>	531	_
.,	Der derende per tabbib	Dreoneg	11,00	(2nd messenger)	Littlinging of t	Lingingeobli	priotecoueteria	001	
20	Escherichia coli	EcGReg	NP 287665	Gene regulator	EREROR DUE5	FREROR	v-Proteobacteria	460	YddV
		Bronds	111_207000	(2nd messenger)	Literique or c	Linginoophi	, i roteooueteria		1 44 1
21	Azotobacter vinelandii	AvGReg	ZP 00090857	Gene regulator	EREROR DUF6	FREROR	v-Proteobacteria	472	Avin2552
21	nizorooderer vinendindin	moneg	21_0000000007	(2nd messenger)	Enterique of a	Enteriqueoobei	Trocoodecena	172	1101112002
22	Shigella flexneri 2a str 301	SfGReg	NP 707605	Gene regulator	EREROR DUE7	FREROR	v-Proteobacteria	381	YddV
22	Shigena Jexheri 2a sti.501	Sjones	111_707005	(2nd messenger)	ERERQR.DOI /	ERERQR:00DEI	/ 1 lotoobacteria	501	1 dd V
23	Acidithiobacillus ferrooxidans	AfGReg	n/a	Gene regulator	DUF1-DUF2	GGDFR FAI	v-Proteobacteria	880	_
23	Actanniobactinas Jerrooxiaans	AJORCE	11/a	(2nd messenger)	D011-D012	GODER.EAL	<i>γ</i> -110te00aete11a	000	
24	Burkholderia fungorum	<b>RfGR</b> eg	<b>7P</b> 00030046	Gene regulator	GAE.DUE2	GAE-EAI	<b>B</b> -Proteobacteria	724	Bcen2859
24	Burkholderid Jungorum	bjokeg	ZI_00030040	(2nd messenger or	UAL DULZ	GAI-EAL	p-110teobacteria	724	Deep2859
				(2nd messenger of Trascon Pag)					
25	Vibria multifaug CMCD6	<b>W</b> CD ag	ND 762050	Cana ragulatar	STAS	ST A S	v Drataahaataria	206	VV20072
23	v who vumpicus CMCP0	v vOReg	111/02039	(and messanger)	51A5	51A5	y-rioteobacteria	300	v v 20073
20	Cook and an and former have a	C-CC8	a la	(2nd messenger)			S Desta sha sta	200	
20	Geodacter sulfurreaucens	GUUS	n/a	Unclassfilea	_	_	o-proteobacteria	300	C
27	Geodacter metallireducens	GmGCS	ZP_00082251.1	Unclassfiled	—	—	o-Proteobacteria	300	Gmet3020

Accompanying each GCS is the source organism, suggested naming convention along with any previous names, NCBI accession numbers (available except for those with genome sequencing in-progress), classification according to Fig. 2, domain topology as identified by SMART and Pfam, taxonomy and sequence length. Naming conventions for the GCSs are as follows: HemAT = heme-based aerotactic transducers; GReg = gene regulating.



## **Biological Heme-based Sensors**

Fig. 2. Functional classification scheme of biological heme-based sensors. Heme-based sensors CooA, NPAS2, sGC, AxPDEA1, Dos, FixL, and HemAT can be grouped according to their primary functions described in the literature. The GCSs are tentatively categorized according to this schema. *Bf*GReg is believed to be a gene regulator of either the 2nd messenger or transcription regulator class. No function could be assigned to the two membrane-bound *Geobacter* GCSs, *Gs*GCS and *Gm*GCS. Domains with an asterisks (\*) indicate new domains not presently a part of the SMART database. See text and Table 1 for details.

GAF:EAL together with an additional PAS domain. Proteins possessing the GAF domain regulate small molecules like cAMP and cGMP and function in transcription [23–25].

3.1.3. Protein–protein interactions. VvGReg from V. vulnificus possesses a C-terminal STAS (sulfate transporter and anti- $\sigma$  factor antagonist) domain recognized by Pfam as an anti-anti- $\sigma$  factor. Spore formation in *B. subtilis* is an example of such a regulated process utilizing  $\sigma F$  ( $\sigma$  factor initiating prespore development), its antagonist SpoIIAB, and the antianti- $\sigma$  factor, SpoIIAA. To our knowledge, VvGReg is the first example of a globin domain with a transcriptional regulator. GCSs predicted to be involved in DNA binding have yet to be identified.

3.1.4. Unclassified GCS. Two GCSs identified in the strict anaerobic  $\delta$ -Proteobacteria may be involved in sulfate/sulfur reduction. GsGCS from G. sulfurreducens and GmGCS from G. metallireducens exhibit a bundle of four transmembrane helices at C-terminal resemble either glutathione S-transferase (GST) or ferritin-like proteins. These are generally soluble proteins; however, a distinct microsomal membrane-bound GST family has been identified [26,27]. Both proteins are involved in cellular protection from toxicity of reactive oxygen species [28].

#### 3.2. Phylogenetics of the GCSs

The phylogenetic tree (Fig. 1B) results in two interpreta-

tions: (1) there is a predisposition of bacterial lineages for particular signal-transducing elements, or (2) the globin domains are customized to function in concert with particular signal-transducing elements.

In the case of the GCS, a more evolved and ordered protein is built up from the less ordered components; namely, the ancestor globin, or *protoglobin*, and the signaling domains. This higher ordered protein imparts a new function(s) to the host organism that allow descendants to thrive in environments that may not have been able to survive before. Rapid response to toxic oxygen or other highly reactive species that otherwise might quickly kill a microbe is a significant pressure to retain such a fusion protein. Within the tenet of the biological evolution, as atmospheric oxygen levels rose and eukaryotic cells evolved, the need for oxygen taxis may have diminished, resulting in the absence of such chimeric systems in the upper eukaryotes. There are three organisms that possess two GCSs: C. crescentus, M. magnetotacticum, and Magnetococcus. All four proteins in C. crescentus and M. magnetotacticum are HemATs and therefore it seems likely that they arose from gene duplication, i.e. they are paralogs. In contrast, the two GCSs from Magnetococcus perform different functions. One is a HemAT and the other, a predicted gene regulator. This indicates that each globin evolved independently with its particular signaling domain to reflect the observed diversity (Fig. 1B) and predicts the existence of the protoglobin in more primitive organisms like the Archaea or the deeply branching photosynthetic bacteria.

### 4. Summary

The diversity of heme-based sensors in prokaryotes is predominantly globin based. The family of GCSs can be grouped into two subfamilies, the aerotactic and the gene regulating. Though approximately half of the GCSs fall into the generegulating subfamily, the HemATs are the only known hemebased sensors involved in aerotaxis. The GCSs form a family of proteins (Fig. 2) that, thus far, populate all but the direct DNA-binding sensors. Considering the diversity of the GCSs and that the flavohemoglobins are similar to the GCSs, we propose that this form of globin was particularly suited for forming multi-domain chimeric proteins with novel functions. We postulate that protoglobin was the predecessor to the chimeric GCS and should therefore be found in more ancient organisms, like the Archaea.

Acknowledgements: We thank Sergei Vinogradov and an unknown reviewer for critical comments on the manuscript. This investigation was supported by the National Science Foundation grant no. MCB 0080125 and by the University of Hawaii intramural bioinformatic grant awarded to M.A.

#### References

- [1] Chan, M.K. (2001) Curr. Opin. Chem. Biol. 5, 216-222.
- [2] Lanzilotta, W.N., Schuller, D.J., Thorsteinsson, M.V., Kerby, R.L., Roberts, G.P. and Poulos, T.L. (2000) Nat. Struct. Biol. 7, 876–880.
- [3] Dioum, E.M., Rutter, J., Tuckerman, J.R., Gonzalez, G., Gilles-Gonzalez, M.A. and McKnight, S.L. (2002) Science 298, 2385– 2387.
- [4] Gekakis, N., Staknis, D., Nguyen, H.B., Davis, F.C., Wilsbacher, L.D., King, D.P., Takahashi, J.S. and Weitz, C.J. (1998) Science 280, 1564–1569.
- [5] Hogenesch, J.B., Gu, Y.Z., Jain, S. and Bradfield, C.A. (1998) Proc. Natl. Acad. Sci. USA 95, 5474–5479.
- [6] Reick, M., Garcia, J.A., Dudley, C. and McKnight, S.L. (2001) Science 293, 506–509.
- [7] Zhao, Y., Brandish, P.E., Ballou, D.P. and Marletta, M.A. (1999) Proc. Natl. Acad. Sci. USA 96, 14753–14758.

- [8] Delgado-Nixon, V.M., Gonzalez, G. and Gilles-Gonzalez, M.A. (2000) Biochemistry 39, 2685–2691.
- [9] Sasakura, Y., Hirata, S., Sugiyama, S., Suzuki, S., Taguchi, S., Watanabe, M., Matsui, T., Sagami, I. and Shimizu, T. (2002) J. Biol. Chem. 277, 23821–23827.
- [10] Tal, R., Wong, H.C., Calhoon, R., Gelfand, D., Fear, A.L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H., Cohen, A., Sapir, S., Ohana, P. and Benziman, M. (1998) J. Bacteriol. 180, 4416–4425.
- [11] Chang, A.L., Tuckerman, J.R., Gonzalez, G., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M. and Gilles-Gonzalez, M.A. (2001) Biochemistry 40, 3420–3426.
- [12] Weinhouse, H., Sapir, S., Amikam, D., Shilo, Y., Volman, G., Ohana, P. and Benziman, M. (1997) FEBS Lett. 416, 207–211.
- [13] Gilles-Gonzalez, M.A., Ditta, G.S. and Helinski, D.R. (1991) Nature 350, 170–172.
- [14] Gilles-Gonzalez, M.A. and Gonzalez, G. (1993) J. Biol. Chem. 268, 16293–16297.
- [15] David, M., Daveran, M.L., Batut, J., Dedieu, A., Domergue, O., Ghai, J., Hertig, C., Boistard, P. and Kahn, D. (1988) Cell 54, 671–683.
- [16] Virts, E.L., Stanfield, S.W., Helinski, D.R. and Ditta, G.S. (1988) Proc. Natl. Acad. Sci. USA 85, 3062–3065.
- [17] Hou, S., Larsen, R.W., Boudko, D., Riley, C.W., Karatan, E., Zimmer, M., Ordal, G.W. and Alam, M. (2000) Nature 403, 540– 544.
- [18] Hou, S., Belisle, C., Lam, S., Piatibratov, M., Sivozhelezov, V., Takami, H. and Alam, M. (2001) Extremophiles 5, 351–354.
- [19] Hou, S., Freitas, T., Larsen, R.W., Piatibratov, M., Sivozhelezov, V., Yamamoto, A., Meleshkevitch, E.A., Zimmer, M., Ordal, G.W. and Alam, M. (2001) Proc. Natl. Acad. Sci. USA 98, 9353–9358.
- [20] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) Nucleic Acids Res. 24, 4876–4882.
- [21] Cuff, J.A. and Barton, G.J. (1999) Proteins 40, 502-511.
- [22] Frankel, R.B., Bazylinski, D.A., Johnson, M.S. and Taylor, B.L. (1997) Biophys. J. 73, 994–1000.
- [23] Galperin, M.Y., Nikolskaya, A.N. and Koonin, E.V. (2001) FEMS Microbiol. Lett. 203, 11–21.
- [24] Kanacher, T., Schultz, A., Linder, J.U. and Schultz, J.E. (2002) EMBO J. 21, 3672–3680.
- [25] Hurley, J.H. (2003) Sci. STKE 164, PE1, Review.
- [26] Prabhu, K.S., Reddy, P.V., Gumpricht, E., Hildenbrandt, G.R., Scholz, R.W., Sordillo, L.M. and Reddy, C. (2001) Biochem. J. 360, 345–354.
- [27] Morgenstern, R., Guthenberg, C. and Depierre, J.W. (1982) Eur. J. Biochem. 128, 243–248.
- [28] Raza, H., Robin, M.A., Fang, J.K. and Avadhani, N.G. (2002) Biochem. J. 366, 45–55.