Microbiology (2004), **150**, 4189–4197 DOI 10.1099/mic.0.27480-0

Evolution of the PPM-family protein phosphatases in *Streptomyces*: duplication of catalytic domain and lateral recruitment of additional sensory domains

Weiwen Zhang and Liang Shi

Microbiology Department, Pacific Northwest National Laboratory, 902 Battelle Blvd, PO Box 999, Mail Stop: P7-50, Richland, WA 99352, USA

Correspondence
Weiwen Zhang
weiwen.zhang@pnl.gov

Liang Shi
liang.shi@pnl.gov

Received 9 July 2004
Revised 29 August 2004
Accepted 3 September 2004

Originally identified from eukaryotes, the Mg^{2+}- or Mn^{2+}-dependent protein phosphatases (PPMs) are a diverse group of enzymes whose members include eukaryotic PP2C and some prokaryotic serine/threonine phosphatases. In a previous study, unexpectedly large numbers of PPMs were identified in two *Streptomyces* genomes. In this work, a phylogenetic analysis was performed with all the PPMs available from a wide variety of microbial sources to determine the evolutionary origin of the *Streptomyces* PPM proteins. Consistent with earlier hypotheses, the results suggested that the microbial PPMs were relatively recent additions from eukaryotic sources. Results also indicated that the *Streptomyces* PPMs were divided into two major subfamilies at an early stage of their emergence in *Streptomyces* genomes. The first subfamily, which contains only six *Streptomyces* PPMs, possesses a catalytic domain whose sequence and architecture are similar to that of eukaryotic PPMs; the second subfamily contains 89 *Streptomyces* PPMs that lack the 5a and 5b catalytic domain motifs, similar to the PPMs SpoIIE and RsbU of *Bacillus subtilis*. Significant gene duplication was observed for the PPMs in the second subfamily. In addition, more than half (54 %) of the *Streptomyces* PPMs from the second subfamily were found to have at least one additional sensory domain, most commonly the PAS or the GAF domain. Phylogenetic analysis showed that these domains tended to be clustered according to the putative physiological functions rather than taxonomic relationship, implying that they might have arisen as a result of domain recruitment in a late evolutionary stage. This study provides an insight into how *Streptomyces* spp. may have expanded their PPM-based signal transduction networks to enable them to respond to a greater range of environmental changes.

INTRODUCTION

The Mg^{2+}- or Mn^{2+}-dependent protein phosphatase (PPM) family is the one of two major serine/threonine phosphatase superfamilies present in both eukaryotes and prokaryotes (Cohen, 1994). Although their sizes vary, all PPMs share a common catalytic domain consisting of nine to eleven signature sequence motifs containing eight conserved amino acid residues (Fig. 1). These conserved amino acids include four aspartates that are directly involved in coordination of the metal ions in the active site. However, some bacterial PPMs, such as SpoIIE and RsbU/X/P of *Bacillus subtilis*, and Icfg of *Synechocystis* PCC 6803, lack the 5a and 5b motifs (Fig. 1) (Bork *et al.*, 1996; Shi *et al.*, 1998). Despite the absence of these conserved motifs, the recombinant SpoIIE, RsbU/X/P and Icfg expressed in *Escherichia coli* are all functional protein serine phosphatases with a high degree of substrate specificity, suggesting that motifs 5a and 5b are not critical for either catalysis or substrate specificity (Shi, 2004). PPMs appear to be ubiquitous in eukaryotes, while the PPM genes are completely absent in some prokaryotic genomes. A recent survey showed that most archaeal genomes lacked PPM encoding genes, and only 51 % of the bacterial genomes surveyed contained ORFs encoding potential PPMs (Kennelly, 2002).

Abbreviations: PP, protein phosphatase; PPM, Mg^{2+}- or Mn^{2+}-dependent protein phosphatase.

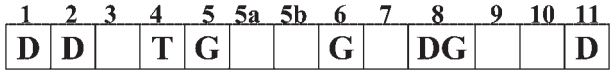

Fig. 1. Schematic representation of signature sequence of the PPM-family protein phosphatases. Single-letter amino acid code is used. The individual motifs of each family are indicated by the numbers; the sizes of these motifs are not drawn to scale.

Phylogenetic analysis of these bacterial PPM sequences suggested that they might have originated in eukaryotes, and were then acquired by bacteria and archaea via horizontal gene transfer events (Ponting *et al.*, 1999; Kennelly, 2002).

Most of the antibiotics currently used in medicine are produced by the family of soil-inhabiting bacteria *Streptomycetaceae*. Cultures of *Streptomyces* spp. synthesize these compounds as secondary metabolites, either following the rapid growth phase of liquid-grown cultures, or coupled to sporulation in plate-grown cultures. Numerous studies have shown that physiological differentiation and secondary metabolism are regulated by many environmental signals, such as nutrient availability, temperature or water stress (Chater, 1993; Hopwood, 1999). To regulate their complex life cycles, and to allow them to adapt to diverse environmental conditions, sensitive signal transduction systems must have evolved in *Streptomyces* species. Recent studies have shown that protein *O*-phosphorylation is a widespread phenomenon in streptomycetes (Umeyama *et al.*, 2002; Horinouchi, 2003; Nádvorník *et al.*, 1999). The involvement of eukaryotic-type protein phosphatases (PPs) in *Streptomyces* metabolism has been demonstrated by the observation that overexpression of the *Streptomyces coelicolor ptpA* protein phosphatase gene in *Streptomyces lividans* increases the production of actinorhodin and undecylprodigiosin (Li & Strohl, 1996; Umeyama *et al.*, 1996). Furthermore, the involvement of eukaryotic-type PPs in *Streptomyces* differentiation was demonstrated when the disruption of the *sppA* PP gene of *S. coelicolor* A3(2) impaired the vegetative growth and formation of hyphae and spores (Umeyama *et al.*, 2000).

In our previous study (Shi & Zhang, 2004), we identified 55 eukaryotic-type PPs in each of the *S. coelicolor* and *Streptomyces avermitilis* genomes. Considering that most of the previously surveyed prokaryotic genomes contained fewer than ten eukaryotic-type PPs (Shi *et al.*, 1998), the number of PPs found in *Streptomyces* was unexpectedly large. Analysis of the catalytic domains of the *Streptomyces* PPs showed that 49 of the PPs in *S. coelicolor* and 48 of the PPs in *S. avermitilis* belong to the PPM family. Only 29 of these PPMs have orthologues in both species (Shi & Zhang, 2004). As these streptomycetes have the largest number of PPMs ever identified from any single prokaryotic organism, the existence of PPMs in *Streptomyces* genomes raises questions of how they originated. Comparison of the *S. coelicolor* and *S. avermitilis* genomes revealed that the chromosomes appear to have expanded by internal duplication of DNA, and gene acquisition by horizontal transfer (Bentley *et al.*, 2002). This expansion has presumably allowed these organisms to adapt to a wider range of environmental conditions, and exploit a greater variety of nutrient sources. The preferential incorporation (and subsequent maintenance) of occasionally beneficial sequences outside the conserved ancestral core region of the chromosome has created arms comprising mostly 'non-essential' functions (Bentley *et al.*, 2002; Ikeda *et al.*, 2003). Consistent with this hypothesis, we have found that the majority of *Streptomyces* PPM genes that are unique to each genome are located outside of the core conserved region (Shi & Zhang, 2004). This suggests that recent gene acquisition or duplication events might be responsible, at least to a certain degree, for the presence of such a large number of PPMs in *Streptomyces* genomes. In this study, we have performed a detailed phylogenetic and domain structural analysis of the *Streptomyces* PPMs to show that these genes were divided into two major subfamilies after their establishment in the *Streptomyces* genome. Our analysis indicates that one PPM subfamily (designated subfamily I) has retained the original eukaryotic PPM domain architecture through the course of evolution. In contrast, the other PPM subfamily (designated subfamily II) has been subjected to significant evolutionary modification, including gene duplication and additional sensory domain recruitment. These observations suggest a mechanism by which *Streptomyces* species may have acquired a diverse population of PPMs to provide the robust signal transduction systems required for survival in the extremely diverse soil environments in which they are found.

METHODS

Search for prokaryotic PPMs. Most of the prokaryotic PPMs used in this study were obtained from previous studies by ourselves and others (Bork *et al.*, 1996; Shi *et al.*, 1998; Zhang *et al.*, 1998; Wang *et al.*, 2002; Shi & Zhang, 2004), but putative PPMs were also identified from additional microbial genomes (http://www.tigr.org) that had not been thoroughly searched prior to this study. To confirm that any tentatively identified ORFs indeed possessed the catalytic domain of PPMs, their DNA-derived amino acid sequences were used in new searches for conserved domains from the conserved domain databases SMART, Pfam, COG and KOG, with the reverse-position-specific BLAST algorithm provided by the National Center for Biotechnology Information (NCBI) using an *E* value $<0{\cdot}01$ (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer *et al.*, 2003). Applying the criteria described previously (Shi *et al.*, 1998), visual inspection was used to eliminate those domains lacking the minimum complement of conserved sequence features considered necessary to make a functional PPM enzyme.

Sequence alignment and phylogenetic reconstruction. The protein sequences of a total of 148 PPMs (146 from prokaryotes and 2 from eukaryotes) were retrieved from the NCBI or the TIGR database (http://www.ncbi.nlm.nih.gov or http://www.tigr.org). The domain structures were identified and analysed using the molecular architecture research tools provided by SMART (http://smart.embl-heidelberg.de) with *E* value $<0{\cdot}01$ (Letunic *et al.*, 2002). The sequence alignments were performed using default parameters of the CLUSTALW program available from the LaserGene software package (DNAStar), and PAUP* 4.0 beta version (Blumenberg, 1988). Confidence levels were determined by analysing 100 bootstrap replicates.

RESULTS

Identification of PPM proteins

Genes encoding serine/threonine-type protein phosphatase (PP) have been found in all three domains of life, *Archaea*, *Bacteria* and *Eukarya* (Kennelly, 2001). From 137 complete

microbial genomes (16 archaea and 121 bacteria, as of 22 February 2004) listed in the TIGR CMR (Comprehensive Microbial Resource; http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl), a total of 146 protein sequences of putative PPMs were identified, which are distributed in 1 archaeal and 27 bacterial genomes (Bork *et al.*, 1996). Among the genomes searched, the two *Streptomyces* species we studied previously contained the largest numbers of PPMs (49 PPMs for *S. coelicolor*, and 48 PPMs for *S. avermitilis*), followed by *Synechocystis* sp. PCC 6803, *Nostoc* sp. PCC 7120 (formerly *Anabaena* sp. PCC 7120) and *Bacillus subtilis*, which were found to possess five to eight PPMs. In addition, *Thermosynechococcus elongatus* and *Vibrio vulnificus* contained two or three PPMs. Of 16 archaeal genomes searched, only one PPM-encoding gene was detected (from *Thermoplasma volcanium* TVN0703), supporting the earlier observations that most archaeal genomes lack potential PPM genes (Shi *et al.*, 1998; Kennelly, 2002).

Phylogenetic analysis of the catalytic domains of *Streptomyces* PPMs

To infer the evolutionary histories of *Streptomyces* PPMs, a phylogenetic analysis was performed using sequences of PPM catalytic domains rather than complete coding sequences of whole proteins, in order to increase the confidence level of sequence alignment. The protein sequences of the 49 identified microbial PPMs, 2 eukaryotic PPMs, and 97 PPMs identified from the *S. coelicolor* and *S. avermitilis* genomes, were subjected to domain identification using the molecular architecture research tools provided by SMART. The PPM catalytic domain sequences, which were about 220 aa with all the conserved residues (Fig. 1), were then extracted and used for phylogenetic tree construction. A phylogenetic tree was generated from the aligned catalytic domain sequences, and the confidence of the tree topology was evaluated by bootstrap analysis (Fig. 2).

The conventional view of the universal evolutionary tree is that eukaryotes are rooted in the bacteria (Woese *et al.*, 1990). However, the phylogenies we derived for PPMs clearly deviated from this canonical view, since all of the prokaryotic PPMs were rooted in eukaryotic PPMs. This observation is consistent with an earlier assertion that PPMs may have originated in eukaryotes after their separation from the last common ancestor with bacteria, and later radiated into bacteria through multiple horizontal gene-transfer events (Ponting *et al.*, 1999; Kennelly, 2002). This postulated gene acquisition from eukaryotes would have to have occurred at an early stage of bacterial evolution, since a similar PPM subfamily I catalytic domain architecture was observed to be widespread in diverse bacterial species (Fig. 1).

Another interesting observation that can be made from our phylogenetic analysis is that there was strong bootstrap support for the division of *Streptomyces* PPMs into two major subfamilies. Subfamily I contained only 6 *Streptomyces* PPMs (three pairs of homologues from genomes of *S. coelicolor* and *S. avermitilis*), and 31 other microbial PPMs from 24 different microbial species. The catalytic domains of the PPM subfamily I all shared the same domain architecture as eukaryotic PPMs, in that they contained 11 conserved motifs, with 8 absolutely conserved residues. The diversification of the PPM subfamily I functional domain between species might be due to the subsequent bacterial speciation and functional specialization. No obvious sign of gene duplication was found for the PPMs in this subfamily. Subfamily II, however, contained 91 *Streptomyces* PPMs and 18 PPMs from other microbial sources, mainly from cyanobacteria and *Bacillus*. PPMs from this subfamily showed a relatively high degree of similarity, indicating that some might have arisen by gene duplication events occurring after the original acquisition of the progenitor PPM gene. Comparison of the architecture of catalytic domains showed that they all lacked the 5a and 5b motifs, and showed much less sequence similarity to eukaryotic PPMs. These results suggested that the diversification of architecture in the PPM subfamily II may be due more to functional specialization than to bacterial speciation.

Phylogenetic analysis of additional sensory domains of *Streptomyces* PPMs

Our previous study reported that 63 % of PPMs from *S. coelicolor* and 75 % of the PPMs from *S. avermitilis* possess at least one extra sensory domain, in addition to their PP catalytic domain (Shi & Zhang, 2004). Interestingly, all of the PPMs with additional sensory domains were from subfamily II (Fig. 2), while none of the PPMs from subfamily I contained any additional sensory domain. Two types of sensory domains were most frequently found to be added in these *Streptomyces* subfamily II PPMs. The first was a PAS domain that binds the chromophore 4-hydroxycinnamyl, and is involved in sensing energy-related environmental factors such as oxygen, redox potential or light; the second is the GAF domain involved in binding cyclic nucleotides (Shi & Zhang, 2004; Taylor & Zhulin, 1999; Taylor *et al.*, 1999; Aravind & Ponting, 1997). Five of the *Streptomyces* PPMs with an additional sensory domain were found to contain the PAS domain only, 14 contained the GAF domain only, and 29 contained both PAS and GAF domains (Fig. 2). Outside of the *Streptomyces* PPMs, a PAS domain was also found in the *B. subtilis* RsbP PPM gene, and GAF domains were found in PPMs from *Synechocystis* sp., *Thermosynechococcus elongatus* and *Synechococcus* sp. WH8102. In addition to the PAS and GAF domains, all four of the PPMs identified in *Nostoc* sp. PCC 7120 contained REC domains. The REC domain encodes a *cheY*-homologous receiver domain involved in receiving signals from the sensor partner in bacterial two-component signal transduction systems. The observation that these REC domains were primarily present in subfamily II PPMs in *Nostoc* suggests that there may be species preferences in


SAV7478
SCO0767
SAV568
SAV7324
SCO7354
SAV924
NT02SC4120
SAV437
SAV1391
SAV7172
SAV4151
SCO0676
SAV786
SCO7158
SAV586
SCO0464
SAV587
SAV3223
SCO5040
SAV1554
SAV1902
SCO6484
SAV3185
SCO5104
SCO7063
SAV1054
SAV7330
SCO7009
SAV767
SCO7220
SAV4108
SCO4120
SAV1142
NT02SC1139
SAV7288
SAV4394
SCO3796
SAV7504
SCO0751
SCO0451
SAV350
SCO7541
SAV7498
SCO0675
SAV1086
SCO7326
SAV1235
SAV1098
SCO7320
SAV4229
SCO3971
SCO7448
SCO7759
SCO3679
NT02SC4665
SAV3928
SCO0542
SAV698
SAV808
SAV4009
SCO4201
SCO0872
SAV5429
SCO2624
SAV5983
Nostoc sp PCC7120 All1703
Bacillus subtilis RsbU
Synechocystis sp PCC6803 Icfg
Synechocystis sp PCC6803 Slr1983
Synechocystis sp PCC6803 Slr2031
Thermosynechococcus elongates Tlr2236
Synechococcus sp WH8102 Synew0012
Synechocystis sp PCC6803 Sll1365
Synechocystis sp PCC6803 Slr0114
SCO7748
SAV2513
SCO5747
SAV5565
SCO2560
SAV922
SCO3691
Nostoc sp PCC7120 All3759
Nostoc sp PCC7120 Alr1086
Nostoc sp PCC7120 Alr2280
Thermosynechococcus elongatus Tlr1153
Bacillus subtilis RsbP
SAV1334
SCO6390
SCO5933
SCO1327
SCO4107
SCO4447
SAV6744
SCO0717
SAV4415
SCO3785
SCP1.41c
SCP1.313
SAV321
SAV956
SAV1093
SCO7747
SCO7380
SAV1097
SCO7321
Vibrio vulnificus NY01VVA0038
Bacillus subtilis SpoIIE
Thermoanaerobacter tengcongensis TTE2396
Bacillus subtilis RsbX
Sub-family II
Bacillus anthracis Ba4001
Myxococcus xanthus Pph1
Vibrio vulnificus VV20439
Bacillus subtilis PrpC
Streptococcus pyogenes SpyM3-1370
Streptococcus agalactiae SA-Stp1
Lactococcus lactis NT01LL2265
Clamydophila caviae CCA00399
Chlamydia muridarum TC0530
Clamydia pneumoniae CP0358
Synechocystis sp PCC6803 Sll1033
Thermosynechococcus elongatus Tlr1644
Nostoc sp PCC7120 Alr0547
Clostridium acetobutylicum CAC0035
Mycoplasma pneumoniae MG108
Nostoc sp PCC7120 All1731
Synechocystis sp PCC6803 Sll0602
Nostoc sp PCC7120 All2470
Thermosynechococcus elongates Tll0923
SAV5374
SCO2667
Synechocystis sp PCC6803 PphA
Nostoc sp PCC7120 Alr4516
Thermosynechococcus elongates Tlr2243
Bradyrhizobium japonicum NTBJ4564
Mesorhizobium loti NT02ML1926
Pseudomonas aeruginosa PA-Stp1
Mycobacterium bovis Mb0018c
Mycobacterium tuberculosis PstP
SAV4341
SCO3845
Corynebacterium efficiens NT01CE0042
Corynebacterium glutamicum Cgl0045
SAV5727
SCO7538
Thermoplasma volcanium TVN0703
Archaeal PPM
Aquifer aeolicus Aq-575
Human P2CA
Rat P2CA
Eukaryotic PPMs
Sub-family I
0.1 change

recruiting additional sensory domains, possibly due to different environmental selective pressures.

For *Streptomyces* spp. living in particularly complex and variable soil environments, having multiple sensory domains added to the PPMs may provide a selective advantage by allowing these PPMs to sense multiple environmental signals and/or interact with multiple regulatory proteins simultaneously. The observation that the additional sensory domains in *Streptomyces* PPMs are found exclusively in subfamily II raises the possibility that the additional sensory domains were recruited into the subfamily II PPM molecules at a late evolutionary stage. To test this hypothesis, we performed a detailed phylogenetic analysis of PAS and GAF domains to see whether they cluster with genes of unrelated lineage. Forty-two PAS domain sequences (mean of 65–100 aa) and 45 sequences of GAF domain (mean of 150–200 aa) were identified from *Streptomyces* PPMs using the SMART program, and were used to construct phylogenetic trees, along with a few dozen PAS or GAF domains from other bacterial sources. The bootstrap analysis was performed to check the reliability of the cluster assignments (Fig. 3). Analysis of PAS domains showed seven recognizable clusters, each with strong bootstrap support. Although individual exceptions were present, significant clustering based on their putative physiological function, rather than taxonomic relationship, was observed within each PAS cluster (Fig. 3A). This suggests that PAS domains with different functional specialities were either recruited into *Streptomyces* PPMs directly from other species, or that they were already present in *Streptomyces* and were later recruited into PPM molecules.

In the phylogenetic tree constructed using GAF domains, most of the known GAF domains from other microbial sources clustered in one clade, in which they also tended to be grouped based on their putative functions rather than taxonomic relationship; for example, GAF domains from CyaB of *Nostoc* sp. and *Anabaena* sp. were clustered together, while distanced from CyaC of same species in the phylogenetic tree (cluster B in Fig. 3B) (Aravind & Ponting, 1997). The majority of the GAF domains from *Streptomyces* PPMs were grouped together in another clade (cluster A in Fig. 3B), in which they shared the same root, suggesting that they might originate from the same ancestor, and then be subjected to lateral functional specialization. It is also notable that the GAF domains from non-PPM *Streptomyces* proteins were clustered differently from those from *Streptomyces* PPMs (Fig. 3B), suggesting they had different courses of origination and evolution.

Further evidence for lateral recruitment of additional sensory domains was provided by the phylogenetic trees of catalytic domains and PAS or GAF domains. An obvious incongruity is observed when comparing these trees, in that two PPMs with homologous catalytic domains do not necessarily possess the same additional sensory domains. For example, the SAV4009 PPM (from *S. avermitilis*) contains both PAS and GAF domains, while its homologous PPM SCO4201 in *S. coelicolor* has no additional sensory domain, suggesting that the additional domain recruitment occurred after the *Streptomyces* speciation. These results suggest that both PAS and GAF domain recruitment in *Streptomyces* PPMs may not have originated from a simple scheme of vertical transmission occurring concurrently with speciation, but may be the result of domain recruitment in later evolutionary stages. This domain recruitment may have occurred as a result of adaptive function specialization occurring after duplication of the catalytic domain of another PPM gene.

DISCUSSION

Streptomycetes are among the most numerous and ubiquitous soil bacteria. Unusually for bacteria, streptomycetes exhibit complex multicellular development, and they have the ability to synthesize numerous natural compounds with complex chemical structures (Bentley *et al.*, 2002). It is thus expected that *Streptomyces* spp. must have very delicate and robust signal transduction mechanisms (Chater, 1993). However, even considering the complexity of their life style, it is still a surprise to find that two *Streptomyces* genomes contain five to ten times more PPM-encoding genes than any of the other sequenced bacterial genomes. This is particularly striking when contrasted with the number of PPs from three other identified PP superfamilies (the phosphoprotein phosphatase family, the conventional protein tyrosine phosphatase family, and the low-molecular-mass protein tyrosine phosphatase family), which occur at frequencies similar to those found in other bacterial genomes. In this study, a detailed phylogenetic analysis of *Streptomyces* PPMs was undertaken to explore their origin and evolutionary history.

PPMs are functionally diverse enzymes, and they have been

Fig. 2. Phylogenetic analysis of the catalytic domains of *Streptomyces* PPMs. The PPMs of *S. coelicolor* A3(2) are those with locus accession numbers beginning with 'SCO' (annotated by Sanger Sequencing Center), 'SCP' (located in plasmid) and 'NT' (annotated by TIGR). The PPMs of *S. avermitilis* are those locus accession numbers beginning with 'SAV'. The *Streptomyces* PPMs from subfamily I are marked in bold. The major nodes supported by a bootstrap value of $\geqslant$50 % are indicated by black spots. The *Streptomyces* PPMs with secondary functional domain are marked in colour, with red indicating the proteins with the PAS domain, green indicating those with the GAF domain, and blue indicating those with both PAS and GAF domains. The PPM proteins with the REC domain are marked in purple.

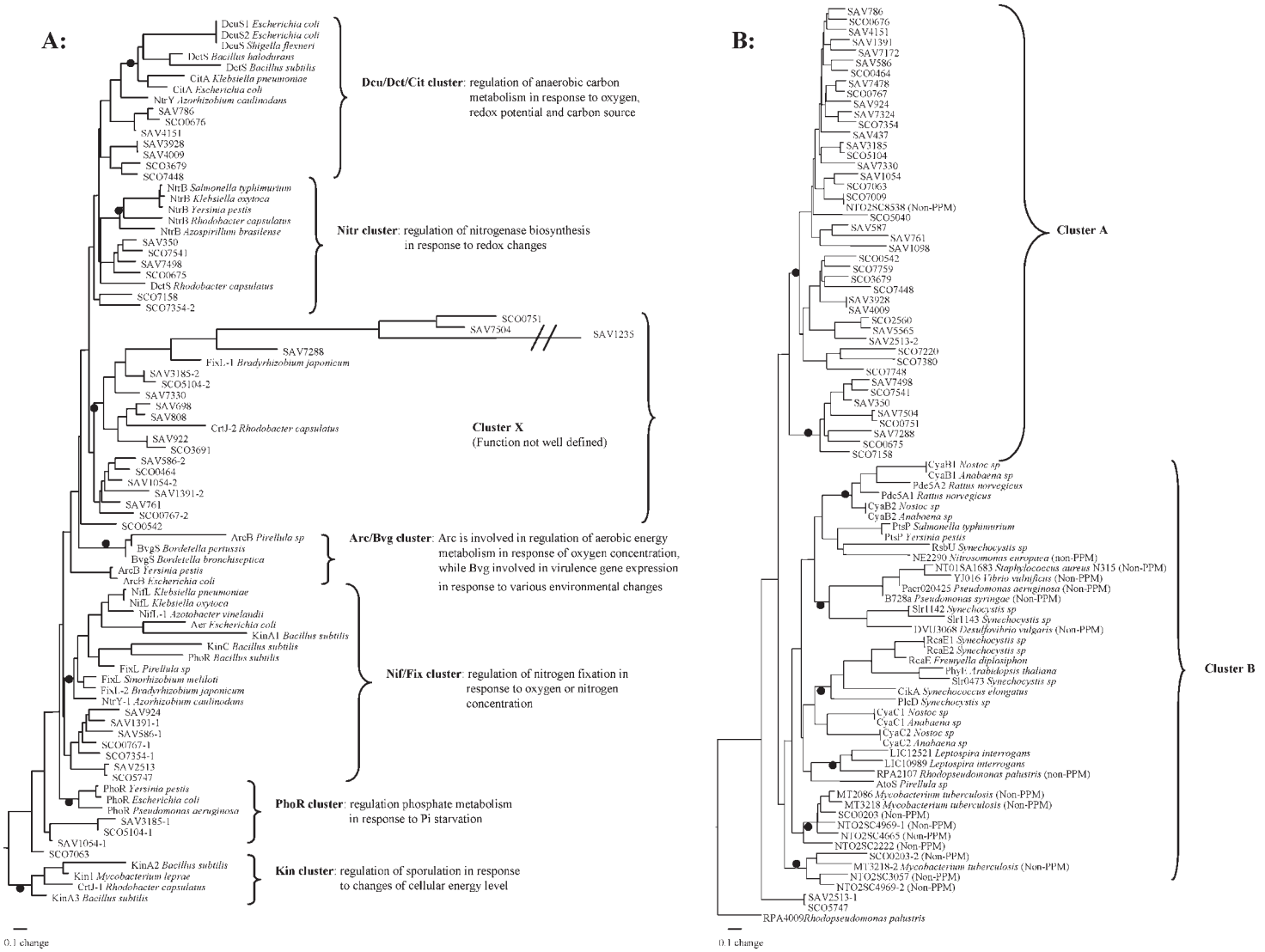

Fig. 3. Phylogenetic analysis of the additional functional domains from *Streptomyces* PPMs. (A) Phylogenetic tree of PAS domains. (B) Phylogenetic tree of GAF domains. The domains from *S. coelicolor* A3(2) PPMs are those with locus accession numbers beginning with 'SCO' (annotated by Sanger Sequencing Center), and 'NT' (annotated by TIGR). The domains from *S. avermitilis* PPMs are those locus accession numbers starting with 'SAV'. Several major branching points supported by a bootstrap value of $\geqslant$50 % are indicated by black spots. The domains from other microbial sources are indicated by gene names followed by species number. The functional definition for the clusters in the PAS phylogenetic tree was according to Taylor & Zhulin (1999). The lack of information on the physiological function of GAF domains hindered the function definition of clusters.

found to be involved in regulation of cellular responses to environmental stress in mammalian and yeast cells, and to the growth regulator abscisic acid in plant cells. The functions of the microbial PPMs that have been characterized include regulation of spore formation, stress response, cell density during stationary phase, carbon and nitrogen assimilation, vegetative growth, development of fruiting bodies and cell segregation (Beuf *et al.*, 1994; Duncan *et al.*, 1995; Gaidenko *et al.*, 2002; Irmler & Forchhammer, 2001; Rajagopal *et al.*, 2003; Shi *et al.*, 1999; Treuner-Lange *et al.*, 2001; Yang *et al.*, 1996). It is notable that most bacterial species containing relatively large numbers of PPM-encoding genes, such as *Anabaena* sp. PCC 7120, *Bacillus* spp. and *Streptomyces* spp., are able to undergo complex morphological differentiation, similar to that seen in multicellular eukaryotes. Although it still needs proof, preliminary analysis points to a possible link between the complexity of a micro-organism's life style, and the distribution of PPM-encoding genes in its genome. In the past several years, extensive knowledge has been accumulating on the cellular roles of PAS-domain-based signalling systems and, on the basis of this knowledge, it is often possible to propose a role for a particular PAS domain based on known functions in similar signalling systems (Taylor & Zhulin, 1999). Our observation that most of the PAS domains from *Streptomyces* PPMs are clustered with PAS domains of known function from other species might thus provide some guidance for further investigations of the physiological functions of *Streptomyces* PPMs.

Although the physiological function of most *Streptomyces* PPMs is still unknown, studies in *B. subtilis* suggest that sigma factors might be one of the predominant targets regulated by PPM-mediated signal transduction pathways (Mittenhuber, 2002). In support of this, a *B. subtilis* PPM containing a PAS domain was recently found to be necessary for conveying signals of energy stress to the sigma B transcription factor in this organism (Vijay *et al.*, 2000). A possible functional link between PPMs and sigma factors is also suggested by the correlation between the number of PPMs and sigma factors in each genome, with most PPM-rich micro-organisms also containing a large number of sigma factors. For example, *B. subtilis* contains 18 sigma factors, *Synechocystis* sp. PCC 6803 contains 7, and *Anabaena* sp. PCC 7120 contains 9 (Gruber & Gross, 2003). A recent survey has identified an astonishing 63 sigma-factor-encoding genes from the *S. coelicolor* genome, and phylogenetic analysis suggested that gene duplications might be responsible for the origin of some of these sigma factors. The gene duplication in group 4, which contains 49 out of the total of 63 sigma factors, was particularly evident. Among this group are the sigma factors σ^{BldN} and σ^{E} involved in regulation of aerial hyphae formation and the integrity of the cell envelope (Gruber & Gross, 2003). Although still requiring biochemical and genetic verification, the correlation between the numbers of PPMs and sigma factors, and their apparently similar evolutionary histories, strongly suggests that the PPMs in *Streptomyces*, especially those from the second subfamily, are most likely to be involved in regulating the activities of at least some sigma factors.

Like the PPMs previously characterized from other bacteria (Shi, 2004), *Streptomyces* PPMs were clearly divided into two major subfamilies. However, the unbalanced distribution of these *Streptomyces* PPMs into the two subfamilies is unusual. The first subfamily contains only six *Streptomyces* PPMs, each with a conserved catalytic domain architecture similar that of eukaryotic PPMs, and showing no sign of either gene duplication or additional domain recruitment. One possible explanation could be that these six PPMs are involved in the regulation of critical 'house-keeping' functions, and their regulatory functions were specialized at a very early stage of evolution. If this were the case, any gene duplication and mutation in this subfamily would be likely to have an adverse effect on the fitness of the cell. In contrast, the PPM genes in the second subfamily might be involved in the regulation of less essential metabolic activities, and would therefore be flexible enough to serve as a basis for gene duplication and domain recruitment to generate novel sensory and regulatory activities to deal with highly variable environments. Several recent studies comparing multiple complete genomes from phylogenetically distant species have concluded that the number of universally conserved gene families is very small, and that multiple events of horizontal gene transfer and domain recruitment within or across species constitute a major evolutionary path to the generation of novel genes (Copley *et al.*, 2003; Koonin & Galperin, 1997). This is especially applicable in the case of the non-essential genes that tend to be recruited to deal with the survival needs in new niches. The phenomenon of horizontal gene transfer between streptomycetes and other bacteria and eukaryotes has been suggested to be responsible for genomic changes ranging from acquisition of individual genes, to entire antibiotic gene clusters (Egan *et al.*, 2001; Coque *et al.*, 1991). Such gene transfers have been demonstrated to occur at high frequencies, even in the soil environment (Wellington *et al.*, 1992). It is especially noteworthy that the large-scale domain recruitment we observed in *Streptomyces* PPMs represents regulatory rather than structural genes, since it has been proposed that regulatory functions are not routinely acquired by means of heterogeneous acquisition (Jain *et al.*, 1999; Ma & Zeng, 2004). Our study provides an insight into how *Streptomyces* spp. may have expanded their PPM-based signal transduction networks to enable them to respond to a greater range of environmental changes.

ACKNOWLEDGEMENTS

We would like to thank Dr David E. Culley for his critical reading of this manuscript. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the US Department of Energy through contract DE-AC06-76RLO 1830.

REFERENCES

Aravind, L. & Ponting, C. P. (1997). The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem Sci* **22**, 458–459.

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A.-M. & 40 other authors (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147.

Beuf, L., Brown, N. P., Hegyi, H. & Schultz, J. (1994). A protein involved in co-ordinated regulation of inorganic carbon and glucose metabolism in the facultative photoautotrophic cyanobacterium *Synechocystis* PCC 6803. *Plant Mol Biol* **25**, 855–864.

Blumenberg, M. (1988). Concerted gene duplications in the two keratin gene families. *J Mol Evol* **27**, 203–211.

Bork, P., Brown, N. P., Hegyi, H. & Schultz, J. (1996). The protein phosphatase 2C (PP2C) superfamily: detection of bacterial homologues. *Protein Sci* **5**, 1421–1425.

Chater, K. F. (1993). Genetics of differentiation in *Streptomyces*. *Annu Rev Microbiol* **47**, 683–713.

Cohen, P. T. W. (1994). Nomenclature and chromosomal localization of human protein serine/threonine phosphatase genes. *Adv Protein Phosphatases* **8**, 371–376.

Copley, R. R., Goodstadt, L. & Ponting, C. (2003). Eukaryotic domain evolution inferred from genome comparisons. *Curr Opin Genet Dev* **13**, 623–628.

Coque, J. J., Martin, J. F., Calzada, J. G. & Liras, P. (1991). The cephamycin biosynthetic genes *pcb*AB, encoding a large multidomain peptide synthetase, and *pcb*C of *Nocardia lactamdurans* are clustered together in an organization different from the same genes in *Acremonium chrysogenum* and *Penicillium chrysogenum*. *Mol Microbiol* **5**, 1125–1133.

Duncan, L., Alper, S., Arigoni, F., Losick, R. & Stragier, P. (1995). Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division. *Science* **270**, 641–644.

Egan, S., Wiener, P., Kallifidas, D. & Wellington, E. M. (2001). Phylogeny of *Streptomyces* species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. *Antonie Van Leeuwenhoek* **79**, 127–133.

Gaidenko, T., Kim, T. J. & Price, C. W. (2002). The PrpC serine-threonine phosphatase and PrkC kinase have opposing physiological roles in stationary-phase *Bacillus subtilis* cells. *J Bacteriol* **184**, 6109–6114.

Gruber, T. M. & Gross, C. A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57**, 441–466.

Hopwood, D. A. (1999). Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*. *Microbiology* **145**, 2183–2202.

Horinouchi, S. (2003). AfsR as an integrator of signals that are sensed by multiple serine/threonine kinases in *Streptomyces coelicolor* A3(2). *J Ind Microbiol Biotechnol* **20**, 462–467.

Ikeda, H., Ishikawa, J., Hanamoto, K. & 7 other authors (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature Biotechnol* **21**, 526–531.

Irmler, A. & Forchhammer, K. (2001). A PP2C-type phosphatase dephosphorylates the P_{II} signaling protein in the cyanobacterium *Synechocystis* PCC 6803. *Proc Natl Acad Sci U S A* **98**, 12978–12983.

Jain, R., Rivera, M. C. & Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A* **96**, 3801–3806.

Kennelly, P. J. (2001). Protein phosphatase – a phylogenetic perspective. *Chem Rev* **101**, 2291–2312.

Kennelly, P. J. (2002). Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. *FEMS Microbiol Lett* **206**, 1–8.

Koonin, E. V. & Galperin, M. Y. (1997). Prokaryotic genomes: the emerging paradigm of genome-based microbiology. *Curr Opin Genet Dev* **7**, 757–763.

Letunic, I., Goodstadt, L., Dickens, N. J. & 7 other authors (2002). Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res* **30**, 242–244.

Li, Y. & Strohl, W. R. (1996). Cloning, purification, and properties of a phosphotyrosine protein phosphatase from *Streptomyces coelicolor* A3(2). *J Bacteriol* **178**, 136–142.

Ma, H. W. & Zeng, A. P. (2004). Phylogenetic comparison of metabolic capacities of organisms at genome level. *Mol Phylogenet Evol* **31**, 204–213.

Marchler-Bauer, A., Anderson, J. B., DeWeese-Scott, C. & 24 other authors (2003). CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* **31**, 383–387.

Mittenhuber, G. (2002). A phylogenomic study of the general stress response sigma factor sigmaB of *Bacillus subtilis* and its regulatory proteins. *J Mol Microbiol Biotechnol* **4**, 427–452.

Nádvorník, R., Vomastek, T., Janecek, J., Techniková, Z. & Branny, P. (1999). Pkg2, a novel transmembrane protein Ser/Thr kinase of *Streptomyces granaticolor*. *J Bacteriol* **181**, 15–23.

Ponting, C. P., Aravind, L., Schultz, J., Bork, P. & Koonin, E. V. (1999). Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J Mol Biol* **1289**, 729–745.

Rajagopal, L., Clancy, A. & Ruhens, C. E. (2003). A eukaryotic type serine/threonine kinase and phosphatase in *Streptococcus agalactiae* reversibly phosphorylate an inorganic pyrophosphatase and affect growth, cell segregation, and virulence. *J Biol Chem* **278**, 14429–14441.

Shi, L. (2004). Manganese-dependent protein *O*-phosphatases in prokaryotes and their biological functions. *Front Biosci* **9**, 1382–1397.

Shi, L. & Zhang, W. (2004). Comparative analysis of eukaryotic-type protein phosphatases in two streptomycete genomes. *Microbiology* **150**, 2247–2256.

Shi, L., Potts, M. & Kennelly, P. J. (1998). The serine threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms: a family portrait. *FEMS Microbiol Rev* **22**, 229–253.

Shi, L., Bischoff, K. M. & Kennelly, P. J. (1999). The *icfG* gene cluster of *Synechocystis* sp. strain PCC 6803 encodes an Rsb/Spo-like protein kinase, protein phosphatase, and two phosphoproteins. *J Bacteriol* **181**, 4761–4767.

Taylor, B. L. & Zhulin, I. B. (1999a). PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* **63**, 479–506.

Taylor, B. L., Zhulin, I. B. & Johnson, M. S. (1999b). Aerotaxis and other energy-sensing behavior in bacteria. *Annu Rev Microbiol* **53**, 103–128.

Treuner-Lange, A., Ward, M. J. & Zusman, D. R. (2001). Pph1 from *Myxococcus xanthus* is a protein phosphatase involved in vegetative growth and development. *Mol Microbiol* **40**, 126–140.

Umeyama, T., Tanabe, Y., Aigle, B. D. & Horinouchi, S. (1996). Expression of the *Streptomyces coelicolor* A3(2) *ptpA* gene encoding a phosphotyrosine protein phosphatase leads to overproduction of secondary metabolites in *S. lividans*. *FEMS Microbiol Lett* **144**, 177–184.

Umeyama, T., Naruka, A. & Horinouchi, S. (2000). Genetic and biochemical characterization of protein phosphatase with dual substrate specificity in *Streptomyces coelicolor* A3(2). *Gene* **258**, 55–62.

Umeyama, T., Lee, P. C. & Horinouchi, S. (2002). Protein serine/threonine kinases in signal transduction for secondary metabolism and morphogenesis in *Streptomyces*. *Appl Microbiol Biotechnol* **59**, 419–425.

Vijay, K., Brody, M. S., Fredlund, E. & Price, C. W. (2000). A PP2C phosphatase containing a PAS domain is required to convey signals of energy stress to the sigmaB transcription factor of *Bacillus subtilis*. *Mol Microbiol* **35**, 180–188.

Wang, L., Sun, Y.-P., Chen, W.-L., Li, J.-H. & Zhang, C. C. (2002). Genomic analysis of protein kinases, protein phosphatases and two-component regulatory systems of the cyanobacterium *Anabaena* sp. strain PCC 7120. *FEMS Microbiol Lett* **217**, 155–165.

Wellington, E. M., Cresswell, N. & Herron, P. R. (1992). Gene transfer between streptomycetes in soil. *Gene* **115**, 193–198.

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci U S A* **87**, 4576–4579.

Yang, X., Kang, C. M., Brody, M. S. & Price, C. W. (1996). Opposing pairs of serine protein kinases and phosphatases transmit signals of environmental stress to activate a bacterial transcription factor. *Genes Dev* **10**, 2265–2275.

Zhang, C. C., Gonzalez, L. & Phalip, C. (1998). Survey, analysis and genetic organization of genes encoding eukaryotic-like signaling proteins on a cyanobacterial genome. *Nucleic Acids Res* **26**, 3619–3625.