Genomics, evolution, and crystal structure of a new family of bacterial spore kinases

Eric D. Scheeff,1 Herbert L. Axelrod,2,3 Mitchell D. Miller,2,3 Hsiu-Ju Chiu,2,3 Ashley M. Deacon,2,3 Ian A. Wilson,2,4 and Gerard Manning1*

1 Razavi Newman Center for Bioinformatics, Salk Institute for Biological Studies, La Jolla, California
2 Joint Center for Structural Genomics (JCSG)
3 Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California
4 Department of Molecular Biology, The Scripps Research Institute, La Jolla, California

INTRODUCTION

Many Gram-positive bacteria form endospores in response to stress. Spores are highly resistant to destructive agents such as heat, chemicals, and radiation, and can persist in harsh environments for many years.1 Although highly stable, spores can rapidly germinate when conditions become hospitable.2 The process of sporulation serves as a model for regulation in bacteria,3 and is important to the pathogenesis of species such as Bacillus anthracis and Clostridium botulinum.4

Bacterial spores have a layered structure which includes a protective protein coat. The coat must exclude harmful agents, while also allowing nutrients to enter to trigger germination.2 It is a complex structure, containing at least 70 different proteins in Bacillus subtilis.1 The roles of many coat proteins are poorly understood, but some are enzymes with known roles,5,6 or have sequence similarity to enzymes.1

YtaA (CotI) of B. subtilis is a member of a family of proteins specific to the phylum Firmicutes, which are implicated in spore formation and often form part of the spore coat.7–9 Apart from a brief mention of some members in InterPro,10 it has not been reported that these proteins form a single family, or that they are kinase homologues. We have therefore named this family the bacterial spore kinases (BSKs), and have carried out a comprehensive genomic and evolutionary analysis of the family, coupled to a combined analysis of sequence conservation and the crystal structure of YtaA.

The BSKs constitute a new family within the CAK kinases.11 CAKs adopt a protein kinase-like (PKL) fold, with distinctive CAK-specific structural elements.12 They usually phosphorylate small molecules, and are named for the choline and aminoglycoside kinase members, which were the first structures to be described.13,14 Although all CAKs share a similar fold, they span a wide sequence and phylogenetic space.11 Beyond the value to our understanding of sporulation, the structure of YtaA provides new insights into the evolution of the PKL superfamily. While similar to other CAKs, YtaA also

ABSTRACT

Bacterial spore formation is a complex process of fundamental relevance to biology and human disease. The spore coat structure is complex and poorly understood, and the roles of many of the protein components remain unclear. We describe a new family of spore coat proteins, the bacterial spore kinases (BSKs), and the first crystal structure of a BSK, YtaA (CotI) from Bacillus subtilis. BSKs are widely distributed in spore-forming Bacillus and Clostridium species, and have a dynamic evolutionary history. Sequence and structure analyses indicate that the BSKs are CAKs, a prevalent group of small molecule kinases in bacteria that is distantly related to the eukaryotic protein kinases. YtaA has substantial structural similarity to CAKs, but also displays distinctive features that broaden our understanding of the CAK group. Evolutionary constraint analysis of the protein surfaces indicates that members of the BSK family have distinct clade-conserved patterns in the substrate binding region, and probably bind and phosphorylate distinct targets. Several classes of BSKs have apparently independently lost catalytic activity to become pseudokinases, indicating that the family also has a major noncatalytic function.

Key words: protein kinase-like; PKL; CAK; endospore; pseudokinase; YtaA; CotS; YutH; YsxE; BSK.

Proteins 2010; 78:1470–1482.
© 2009 Wiley-Liss, Inc.

Additional Supporting Information may be found in the online version of this article.
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.
*Correspondence to: Gerard Manning, Razavi Newman Center for Bioinformatics, Salk Institute for Biological Studies, 10010 N. Torrey Pines Rd., La Jolla, CA 92037. E-mail: manning@salk.edu
Received 30 September 2009; Revised 4 November 2009; Accepted 16 November 2009
Published online 24 November 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22663
displays distinctive changes, exemplifying the array of innovations that have taken place in the PKL fold over long evolutionary timescales.

MATERIALS AND METHODS

**Crystallization**

YtaA was crystallized using the nanodroplet vapor diffusion method with standard JCSG crystallization protocols. Screening for diffraction was carried out using the Stanford Automated Mounting system at the Stanford Synchrotron Radiation Laboratory (SSRL, Menlo Park, CA). The crystallization reagent that produced the YtaA crystal used for structure determination consisted of 2.23 M ammonium sulfate, 0.1 M citric acid pH 5.5. Ethylene glycol was added as a cryoprotectant to a final concentration of 15% (v/v). The YtaA crystal was indexed in hexagonal space group P6_422 (Table I). Data collection, structure solution, and refinement

Multiple-wavelength anomalous diffraction (MAD) data were collected at the SSRL on beamline BL11-1 at wavelengths corresponding to the high-energy remote, \( \lambda_1 \), inflection \( \lambda_2 \), and peak \( \lambda_3 \) of a selenium MAD experiment. The datasets were collected at 100 K with a MarMosaic 325-mm CCD detector using Blu-Ice. The MAD data were integrated and reduced using MOSFLM and then scaled with the program SCALA. The selenium substructure solution and phasing were performed with SHELXD and SOLVE, and automatic model building was performed with iterative RESOLVE. Model completion and refinement were performed with Coot and REFMAC 5. Data and refinement statistics for YtaA are summarized in Table I.

**Validation and deposition**

Analysis of the stereochemical quality of the model was accomplished using AutoDepsInputTool, MolProbity, SFCHECK, and WHATIF. Atomic coordinates and experimental structure factors of YtaA have been deposited in the PDB under the code 2Q83.

**Structure analysis**

Coordinates for structures other than YtaA were collected from the PDB as follows: homoserine kinase 2 (HSK2, PDB ID: 2PPQ), choline kinase (ChoK, PDB ID: 1CKP), aminoglycoside phosphotransferase (APH, PDB ID: 1LB T), and protein kinase A (PKA, PDB ID: 1CDK). MolProbity was used to add optimized hydrogen atoms; all suggested Asn/Gln/His flips were also accepted. Structural alignments were made with DaliLite.
**Sequence analysis**

Sequence homologs of YtaA were gathered using BLAST and HMM searches of the NCBI peptide non-redundant database\(^3\)\(^1\) and the IMG microbial genome database.\(^2\) Chromosomal clustering and operon structure was verified using IMG. Sequences from the BSK family were aligned with MUSCLE\(^3\)\(^3\) followed by manual curation, using the YtaA structure to determine appropriate gap locations. A nonredundant alignment was made by removing sequences from strain variants. A representative alignment of HSK2 sequences was made from sequences in Kannan et al.,\(^1\)\(^1\) filtered to retain sequences with <80% identity with cd-hit,\(^3\)\(^4\) followed by alignment with MUSCLE. The BSK and HSK2 alignments were merged by profile–profile alignment in ClustalX,\(^3\)\(^5\) and then manually edited to maximize agreement with the DaliLite alignment of YtaA and HSK2 structures (Supporting Information Figure S1). Sequence motif logos were made with the WebLogo server.\(^3\)\(^6\) Evolutionary constraints were mapped to the YtaA structure using the ConSurf server,\(^3\)\(^7\) using appropriate sections of the alignment as input. Conservation scores were calculated with the default Bayesian method, and positions that scored within the top 3 conservation bins in ConSurf were reported as conserved.

**Phylogenetic analysis**

The BSK/HSK2 family alignment was edited to remove sparsely populated (uninformative) columns and partial sequences (Supporting Information Figure S2). The alignment was evaluated with PHYML\(^3\)\(^8\) using the following settings: LG substitution model, four substitution categories, estimated gamma shape parameter, optimization of topology/branch length/substitution rate parameters, (the slower but more accurate) SPR tree topology search method, and 100 bootstraps (Supporting Information Figure S3). Taxonomy is from Bergey’s classification, based primarily on 16s rRNA.\(^3\)\(^9\)

**Raw data availability**

All supporting information is available (in its original file formats) at http://kinase.com/microbial/bsk.

**RESULTS AND DISCUSSION**

**BSK: A new family of spore-associated kinases**

We gathered over 220 homologous sequences from public databases that form a distinct new family (BSK) within the CAK kinases. Homologs were from the phylum *Firmicutes*, mostly within spore-forming species in the orders *Bacillales* and *Clostridiales* and largely absent from nonsporulating species (Supporting Information Tables S1 and S2). Multiple BSKs exist in many species, with four predominant within the *Bacillales*, while in *Clostridiales* six distinct BSKs are found in *Clostridium* and one in *Lachnospiraceae* (Fig. 1).

Four BSKs are found in *B. subtilis* and many other *Bacillaceae*: YutH and YsxE are present in almost all spore-forming species, whereas YtaA and CotS are more restricted. All four are experimentally implicated in sporulation. CotS and ytaA share a common promoter, controlled by the spore-specific factors σ\(^K\) and GerE\(^7\)\(^-\)\(^9\)\(^,\)\(^4\)\(^0\)\(^,\)\(^4\)\(^1\) and both are packaged into the spore coat in a CotE-dependent manner.\(^4\)\(^2\) CotS is not detectable in vegetative cells,\(^8\) indicating that its role is spore-specific. YutH and YsxE are also packaged into the coat,\(^4\)\(^2\)\(^,\)\(^4\)\(^3\) and ysxE shares an operon with the spore coat protein SpoVID\(^4\)\(^4\), both are regulated by the spore-specific σ\(^E\).\(^9\) CotS mutants produce morphologically normal spores,\(^7\)\(^,\)\(^8\) Mutants lacking yutH or ysxE also produce spores that are morphologically normal, but more sensitive to lysozyme, hypochlorite, and predation,\(^4\)\(^5\) indicating that BSKs are evolutionarily important for spore survival in natural environments.

Six distinct BSKs (bacterial spore kinase *Clostridiales*, BSK1–BSK6) are found in members of the spore-forming genus *Clostridium* within the family *Clostridiales* I (the genus *Clostridium* is paraphyletic, with some members in other families within *Clostridiales*\(^3\)\(^9\)). Expression profiling in *Clostridium acetobutylicum* shows that BSK4 is selectively expressed during sporulation.\(^4\)\(^6\) Reanalysis of these data suggests that BSKC3 and BSKC5 may also be induced during sporulation, with BSKC3 having the stronger pattern. Accordingly, we find plausible conserved σ\(^K\) binding sites in the BSKC4 and BSKC3 promoters, a weakly conserved site in the BSKC5 promoter, and no site in the single operon containing BSKC1, 2, and 6 (data not shown). This suggests that the expanded family in *Clostridium* may have diverged into spore-associated and nonspore-associated functions.

Most sporulating species within *Clostridiales* and *Bacillales* have BSKs and vice versa, but there are exceptions. A single gene, BSKC7, is present in the *Lachnospiraceae*, in both spore formers and nonspore formers (Supporting Information Table S1). Conversely, *Clostridium difficile* (family *Peptostreptococcaceae*\(^3\)\(^9\)) forms spores but has no BSKs.

Sequence similarity between BSKs is low (highest pairwise identity is ~30%), making phylogenetic reconstruction difficult, with low-bootstrap values at many basal branches (Supporting Information Figure S3). However, when coupled to known species relationships, our results suggest that the most parsimonious evolutionary scenario requires independent expansions in *Bacillales* and *Clostridiales* (see Fig. 1).

In addition to the highly represented BSKs, several divergent members are found in some species (Supporting
Information Table S1). Most notable are four homologs seen within *Heliobacterium modesticaldum* (bacterial spore kinase *Heliobacterium*, BSKH1-4), an unusual phototrophic member of a distinct family in *Clostridiales*. Similarity is weak between the BSKHs and the BSKCs, suggesting that the BSKHs represent an independent
expansion in *H. modesticaldum* that may be related to the shared sporulating phenotype (Supporting Information Figure S3).

Several sporulating species in *Clostridiales* may form a bridge between this order and *Bacillales*. *Clostridium thermocellum* contains YtaA, CotS, and BSKC4. *Symbiobacterium thermophilum* and *Desulfotomaculum acetoxidans* have no BSKCs, but have YtaA (*D. acetoxidans* also has CotS) (Supporting Information Table S1). Although horizontal transfer cannot be ruled out, *C. thermocellum* could represent an ancestral state, from which expansion in *ytaA* could produce the *Bacillales* genes, and expansion in *BSKC4* could produce *Clostridiales* genes.

**Several BSKs are predicted to be catalytically inactive**

The sequence motifs required for enzymatic activity in PKL kinases have been extensively explored and mapped to the structure of PKA, the prototype of PKL kinases. Although the CAK family displays considerable plasticity in these motifs relative to other PKL families, a few key residues have remained nearly invariant, most notably D239YtaA (D166PKA), which coordinates the target substrate hydroxyl group in substrate-bound structures, and is believed to be required for catalytic activity.

Five BSKs have lost D239YtaA, along with other motifs generally required for enzymatic function, and we predict that they are pseudokinases. Assuming that their common ancestor was active, our evolutionary model indicates that BSKs lost catalytic activity independently in *Bacillales* and *Clostridiales*, and possibly in *Lachnospiraceae* (see Fig. 1). This pattern suggests a common nonenzymatic function for BSKs, coupled to a sometimes dispensable role as a kinase.

These five pseudokinase BSKs have a variety of inactivating mutations, in addition to the loss of D239YtaA (see Fig. 1). Three eukaryotic protein kinase (*ePK*) pseudokinase structures have recently been published, all show a selective loss of catalytic residues, coupled with retention of residues required for folding, resulting in structures that are highly similar to their catalytically active relatives. A similar pattern is seen in the BSK pseudokinases, though the sequence changes that occur on inactivation are even more extreme. For example, despite the poor overall sequence similarity between BSKs, H148YtaA (H158PKA) and D269YtaA (D220PKA), which form hydrogen bonds to stabilize the fold of the C-terminal lobe, are conserved in almost all BSK pseudokinases, indicating that this interaction is critical for a family-wide function. However, H237YtaA (H164PKA) which forms hydrogen bond interactions that directly stabilize the geometry of the active site, is only partially conserved in YutH and lost completely in BSKC6. In fact, BSKC6 has lost the entire N-terminal lobe of the kinase domain (see discussion of structure later), similar to some viral PKL kinases, and the KIND domain in metazoans. This change appears to be relatively recent: There are BSK6-like proteins (BSKC6L) with intact N-terminal lobes in *Caldicellulosiruptor saccharolyticus* and *Anaerocellum thermophilum* (Supporting Information Tables S1 and S3). BSKC6L appears to be the ancestral form of BSK6, though it still lacks residues required for enzymatic activity.

The remaining members of the BSK family display substantial selective conservation of known CAK catalytic motifs (though some aspects of the YtaA active site are unusual, see discussion later). This pattern strongly suggests that these proteins will be active kinase enzymes (see Fig. 1). However, no BSK has been experimentally assayed for catalytic activity, and we therefore define these members only as putatively active BSKs.

Interestingly, while CotS is a putatively active BSK, it may also be a pseudokinase in some species. Although it conserves D239YtaA, it frequently loses the DxD motif (Dxe across all CAKs, first position is D257YtaA) and N244YtaA (see Fig. 1). These motifs are not universally required for PKL kinase activity, but they are almost completely conserved in putatively active BSKs. CotS also sometimes loses H237YtaA. Remarkably, three species outside the *Bacillaceae* (including *C. thermocellum*) have a CotS that retains all active site residues (Supporting Information Tables S1 and S3). Thus, as with BSKC6, we can directly observe an apparent ongoing process of loss of functionality within CotS through the examination of current genome sequences.

**YtaA and cotS form a conserved chromosomal cluster with genes involved in nucleotide sugar metabolism**

YtaA and cotS form a conserved chromosomal cluster with a pair of related glycosyl transferases (cotSA and **Table II**

| Gene symbol | COG | Name                                      | Reaction                                      |
|-------------|-----|-------------------------------------------|-----------------------------------------------|
| ytaA        | COG1210 | UDP-glucose pyrophosphorylase | Glucose-1-phosphate + UTP → UDP-glucose       |
| ytcA        | COG1004 | UDP-glucose 6-dehydrogenase | UDP-glucose + NAD → UDP-glucuronate + NADH    |
| ytcB        | COG0451 | UDP-glucose epimerase | Interconverts UDP-glucose and UDP-galactose   |
| cotSA, ytcC | COG0438 | Spore-associated glycosyl transferase | Transfers NDP-sugars to protein or small molecule acceptors |
| cotS, ytaA  | COG2344 | BSK                                      | Kinase (CotS may be inactive)                  |
ytCC, 47% sequence ID) and a set of enzymes involved in nucleotide sugar metabolism (ytC-B and ytdA; Table II, Fig. 2). While the order and operon membership of these genes varies between species, their conserved tight clustering and predicted coregulation suggests functional interactions. Indeed, both CotSA and YtcC are found in the spore and CotSA requires CotS to be assembled into the spore, suggesting that they might be binding partners.

This linkage is further supported by coordinated gene loss in several species. No genome has a ytaA/cotS without this class of glycosyl transference, or a glycosyl transference without ytaA or cotS (Supporting Information Table S1). Our phylogenetic model suggests that there have been multiple coordinated losses of these genes, sometimes linked to losses of the ytc and ytd genes. Four of the 17 sequenced strains of B. thuringiensis have a chromosomal cluster containing both BSKs and both glycosyl transferences, ytcB and ytdA. The other 13 strains lack all six genes. Similarly, of six Geobacillus species, one (WCH70) lacks both BSKs and both glycosyl transferences, whereas another (Y412MC10) has lost one of each and both have also lost some of the ytc/ytd genes. Two of eight B. cereus species have both BSKs and both glycosyl transferences, and the rest lack both.

A BSK-glycosyl transferase link is also seen in most Clostridium species, where BSKCs 2, 6, and 1 are clustered in a single operon and BSKC4 is nearby (see Fig. 2). This operon is flanked by two glycosyl transferences which are distantly related to cotSA and ytcC. By contrast, the conserved linkage of BSKC7 in the mostly non-sporulating Lachnospiraceae is to unrelated genes.

The crystal structure of YtaA indicates that BSKs are CAK kinases with unusual features

The crystal structure of YtaA, at 2.5 Å resolution (Table I), has an overall similarity to previously determined CAK structures. YtaA contains the PKL biloled fold, with a smaller, mostly β-stranded N-terminal lobe, and a larger, mostly α-helical C-terminal lobe (see Fig. 3). The electron density indicated that a molecule containing an adenosine moiety was bound in the interlobe cleft, where the adenosine moieties of ATP/ADP are found in other PKL structures. This molecule was present due to copurification with YtaA; it was not provided in the crystallization conditions (see “Materials and Methods” section). There was no electron density beyond the ribose of the adenosine moiety, and the ribose was partially disordered (Supporting Information Figure S4). Thus, it was unknown if the expected phosphate groups were genuinely absent or simply disordered. We therefore modeled the ligand as adenosine. In addition, the phos-
phosphate-binding loop between Q52 and A57 had poor electron density, and could not be reliably modeled. This loop normally interacts with the phosphates of the ATP and is assumed to be disordered. The N-terminus of the protein, G1-E21, was also not resolved in the electron density maps.

Despite its overall similarity, YtaA is not closely related to any previously characterized CAK structures. The closest similarity is to homoserine kinase 2 (HSK2), an enzyme involved in threonine biosynthesis (Table III). Here, we compare the structure of YtaA to those of choline kinase (ChoK) and aminoglycoside phosphotransferase (APH), prototypical CAKs with well-characterized structures that also have a bound adenosine-derived cofactor, enabling comparison of the ATP binding sites. Of these two structures, YtaA is moderately more similar to ChoK (Table III). We also compare YtaA to a representative ePK, PKA. A substantial structural core, which encompasses the essential residues for ATP binding and phosphotransfer, is shared between ePKs and CAKs (see Fig. 4). We name shared structural elements by ePK conventions, and use an “i” suffix to label elements unique to CAKs.

YtaA retains two CAK-specific structural elements in the C-terminal lobe (see Fig. 4). First, α1i-α2i is a large helical insertion [blue in Fig. 4(B–D)] after αE, which spatially replaces the ePK-specific activation segment [blue in Fig. 4(A)]. YtaA adds a new helix to the end of this insertion [α2bi, red in Fig. 4(D)]. Second, distinctive helices at the C-terminus join with the insertion to form a putative substrate binding site [α3i–α5i, yellow in Fig. 4(B–D)]. In YtaA, the number and position of these helices is more similar to that seen in ChoK than APH. The observed structural similarity in the C-terminal lobe of YtaA and the other CAKs suggests that YtaA uses this region to bind small molecule substrates, as was previously shown for ChoK and APH. While all three CAKs retain these distinctive elements, the structural similarity is lowest in these areas, with different number and placement of elements. It is likely that these differences at least partially reflect changes in the substrate specificity of the three enzymes.

YtaA binds adenosine in a distinctive manner

The YtaA structure reveals a distinctive ATP binding pocket that is broadly similar to other CAKs, but has key elements that help to define BSks as a distinct family. In some aspects, the YtaA pocket is more like that of ChoK, but in others it is more like the APH pocket.

In CAKs, the adenine ring of ATP usually interacts with another aromatic ring from the N-terminal lobe, but the specific interactions and the orientation of the ATP are different in each structure (see Fig. 5). This pattern is in contrast to ePKs, which have a stereotypical ATP binding mechanism: The primary hydrophobic interaction from the N-terminal lobe to the adenine ring of ATP is almost always via V57PKA in β2 and A70PKA in β3. In ChoK and YtaA, the interacting ring comes

Table III
Superposition of YtaA with Representative Structures from a Search with the Dali Server

| Structure                | Classification | PDB ID: chain | Dali Z-score | RMSD (Å) | Aligned positions | %ID |
|-------------------------|----------------|---------------|--------------|----------|-------------------|-----|
| Homoserine kinase 2     | CAK            | 2PPQ:A        | 23.5         | 3.2      | 287               | 16  |
| Choline/ethanolamine kinase | CAK          | 3DXQ:A        | 18.9         | 3.9      | 265               | 13  |
| YihE                    | CAK            | 1ZYL:A        | 18.4         | 4.1      | 273               | 14  |
| Choline kinase α-2 (ChoK) | CAK           | 1CKP:A        | 15.6         | 3.9      | 259               | 14  |
| Methylthioribose kinase (MTRK) | CAK         | 2PUN:A        | 15.1         | 4.0      | 260               | 12  |
| Aminoglycoside phosphotransferase (APH) | CAK       | 1L8T:A        | 13.1         | 5.1      | 219               | 14  |
| RI101                   | RIO           | 1ZTH:C        | 9.9          | 3.6      | 168               | 15  |
| Protein kinase A (PKA)  | ePK           | 1CDK:A        | 7.7          | 3.4      | 162               | 14  |

Structures with IDs in italics were used in analysis and comparisons with YtaA.
from the interlobe linker (W123\textsuperscript{YtaA} and F208\textsuperscript{ChoK}). Although this side chain emanates from the same backbone location, in YtaA aromatic $\pi$-$\pi$ stacking is observed, while in ChoK the rings interact in a perpendicular manner. As a result, in ChoK the face of the adenine ring also packs against L144\textsuperscript{ChoK} in $\beta$3 (see Fig. 5). In APH, the interacting ring instead comes from Y42\textsuperscript{APH} in $\beta$3 (corresponding to L144\textsuperscript{ChoK} and A70\textsuperscript{PKA}), and stacks atop the adenine ring in a similar fashion to W123 in YtaA (see Fig. 5).\textsuperscript{66} In all three structures (as in PKA), the adenine ring also forms H-bonds to the protein backbone in the linker region. The changes in the interaction patterns produce substantial changes in the positioning and orientation of the adenosine moiety (see Fig. 5).

Residues forming the ATP binding pocket from the C-terminal lobe also vary between CAKs. ChoK forms a primary interaction from L313\textsuperscript{ChoK} in $\beta$7 and the ATP ribose also hydrogen bonds with the protein backbone in the linker region. YtaA instead forms the hydrophobic interaction with L256\textsuperscript{YtaA} from $\beta$8, and the ribose moiety forms no hydrogen bonds with the protein. APH uses both hydrophobic sites (F197\textsuperscript{APH} from $\beta$7 and L207\textsuperscript{APH} from $\beta$8). While the ribose of ATP still forms a hydrogen bond, it is to the backbone upstream from $\beta$7 (see Fig. 5).

Consideration of these three CAKs demonstrates that the ATP binding pockets are quite variable within the CAK family. It is possible, particularly in the case of YtaA, that the ATP molecule could shift position and ori-
The YtaA active site indicates it is likely to be a functional enzyme

PKL kinases share a tightly integrated active site where the ATP phosphates are coordinated and positioned for optimal phosphotransfer. In ePKs, the K72\(^{PKA}\)-E91\(^{PKA}\) ion pair links \(\beta3\) and \(\alpha C\), while also providing a positive charge (K72) to interact with the negatively charged ATP phosphates. APH retains a similar K44\(^{APH}\)-E60\(^{APH}\) ion pair, which fulfills a similar role in the APH structure.
ChoK replaces the Lys residue with R146ChoK, a common substitution in CAKs. In contrast to other CAKs, YtaA retains the Lys (K72YtaA), but completely lacks the Glu partner, replacing it with S84YtaA (see Fig. 1), which does not form an analogous interaction. The missing Glu may be functionally replaced by D259YtaA in the Dxe motif (DLD in YtaA, Fig. 6). The side-chain carboxyl groups of D259YtaA and E60APH occupy the same spatial location (see Fig. 6), indicating that this sort of compensation is plausible. Further, D259 is often conserved as an acidic residue in CAKs and is highly conserved in most BSKs (see Fig. 1).

The remaining active site residues, which are required for metal binding and catalysis, are conserved in YtaA (as in APH and ChoK), indicating that this protein is very likely to be catalytically active. The three key residues D239YtaA (D166PKA), N244YtaA (N171PKA), and D257YtaA (D184PKA), are in approximately standard conformations for a PKL kinase (see Fig. 6).

Putatively active BSKs have a distinctive and highly conserved linker motif

Sequence comparison between putatively active and inactive BSKs reveals a highly conserved, mostly hydrophobic structural linker motif including F83YtaA, Y90YtaA, S151YtaA, and Y154YtaA, which is strongly associated with likely enzymatic activity (Fig. 1, hydrophobic linker section), though the motif does not directly interact with the active site (see Fig. 7). These conserved residues form a network of hydrophobic and H-bond interactions that link together αE, αC, β6, and the loop linking αE and α1i, which forms a convoluted structure along the “back” of the enzyme, opposite the active site (see Fig. 7). This motif effectively connects key portions of the two lobes. It also links to the hydrogen bond network that stabilizes the catalytic region of many PKL kinases (mentioned earlier) through an H-bond to H148YtaA. Given the strong correlation with conservation of catalytic motifs, and its linkage to known highly conserved residues, we propose that this motif is likely to stabilize the protein for proper enzymatic function.

The motif is also present, though not fully conserved, in HSK2 (see Fig. 1), but not other CAKs, further demonstrating the relatively close relationship between these two families. Interestingly, it is also present in CotS, suggesting that this enzyme may indeed be active, despite unusual sequence changes in some species.

Conservation patterns in the putative BSK substrate binding site indicate a variety of distinct substrates

Previous structures of CAKs bound to substrate have defined a substrate binding region incorporating residues from α1i–α2i, the catalytic loop, αF, and α4i–α5i. Evolutionary constraint analysis of the entire BSK family with ConSurf reveals a conserved surface region in this

### Table IV

| YtaA residue | Secondary structure element |
|--------------|----------------------------|
| R176         | α1i                        |
| Q179         | α1i                        |
| Q238*        | Catalytic loop             |
| R268*        | αF                         |
| R271*        | αF                         |
| K272*        | αF                         |
| M273         | αF                         |
| I275         | αF                         |
| P276         | αF                         |
| Y318         | α4i                        |
| E319         | α4i                        |
| Y221         | α4i                        |
| D322         | α4i                        |
| R325         | α4i                        |

*Highly conserved specifically in YtaA.
same location, defining the putative substrate binding site for BSKs (Table IV). This site forms a bowl-shaped pocket with a complex surface (see Fig. 8). Remarkably, the electron density maps for the YtaA structure show additional density for an unknown ligand in this region. However, it is unlikely that this ligand could be the biological substrate of YtaA: it is distant from the key active site residue D239YtaA (D166PKA), which directly coordinates the substrate hydroxyl group.49

Although the substrate binding region is generally conserved throughout the BSKs, each group of orthologs in the family displays distinct subsets of highly conserved residues within the pocket (Fig. 1, putative substrate binding section). YtaA has four highly conserved residues, of which three are basic (Table IV), forming a patch directly adjacent to D239 (blue in Fig. 8). CotS has weaker conservation in this patch, but adds an additional region of higher conservation. Remarkably, YutH retains a highly conserved site, despite its catalytic inactivity, indicating that it is very likely to still bind a pseudosubstrate. By contrast, the YsxE binding site is very poorly conserved, and probably does not bind a ligand. This general pattern is repeated in Clostridiales, with BSKC6 having a particularly poorly conserved site, further emphasizing its rapid evolutionary degradation. This pattern strongly suggests that each protein in the family has distinct substrate binding properties. The equivalent residues in HSK2 also differ substantially, (see Fig. 1), indicating that BSKs are unlikely to be homoserine kinases.

**CONCLUSIONS**

The BSKs are a new family of bacterial kinases with distinctive structural features and an unusual subcellular location, with most members packaged into the bacterial spore coat. Although the precise functions of BSKs are unknown, our integrated genomic, phylogenetic, and structural approach has highlighted several attributes of the family.

The dynamic evolutionary pattern seen in BSKs suggests that they provide multiple specific functional enhancements to different species, rather than acting as core structural elements of the coat. This notion is supported by the absence of BSKs from some sporulating species, such as *C. difficile*, and the mild phenotypes of BSK mutants.7,8,45 Further diversity comes from frequent apparent loss of catalytic activity, suggesting that BSKs have a common, and perhaps predominant, nonenzymatic function. The diverse sequence conservation patterns in the putative substrate binding site also suggest distinct functions, and indicate that some BSK pseudokinases could function through binding pseudosubstrates. The tight association between some BSKs and glycosyl transferases and predicted nucleotide sugar metabolizing enzymes suggests that they may bind or phosphorylate one of these reactants, correlating with the aminoglycoside substrates of the related APHs. Given the relevance of spore formation to bacterial pathogenesis, and the demonstrated drugability of PKL kinases,67,68 we believe further experimental characterization of this family is warranted.

The structure of YtaA also illuminates the remarkable innovations that have occurred in the active site of CAK kinases. Although the ePK family is very diverse, the mode of ATP binding and the conservation of active site residues is almost identical across the entire family, with only a few narrow exceptions.69,70 In contrast, CAKs often display substantial changes to the residues in the ATP binding pocket and catalytic region.11 Our comparison of YtaA with the structures of other CAKs reveals the substantial degree of structural variability coincident with these sequence changes. Consideration of the broader PKL superfamily has revealed a wide range of structural changes in substrate binding regions,12 and the CAKs now demonstrate that such changes can even propagate into the active sites. The variability in CAK active sites may be due to the wide variety of molecules that they must phosphorylate: each CAK may be enzymatically

---

**Figure 8**

Predicted substrate binding region of the BSK family, mapped onto the YtaA surface. The orange and blue surface show the region generally conserved throughout the family. The blue region is specifically highly conserved in YtaA (Table IV). The key active site residue D239YtaA is shown with a red space-filling shell. The unknown ligand in the YtaA structure (green) and the adenosine are shown in ball-and-stick.
optimized for its specific substrate. 71 In contrast, while the ePKs display a broad range of peptide motif specificity, 72 the ultimate catalytic target of these enzymes is much more restricted: the hydroxyl groups of serine, threonine, and tyrosine residues. Thus, as with YtaA, new structures of CAs should continue to provide insights into the true catalytic plasticity of the PKL fold.

ACKNOWLEDGMENTS

This work was supported by NHGRI grant HG004164-01 (G.M.) and by National Institutes of Health, Protein Structure Initiative grant number U54 GM074898 (JCSG) from the National Institute of General Medical Sciences (www.nigms.nih.gov). The authors specifically acknowledge the invaluable work of other members of the JCSG who contributed to the high-throughput JCSG pipeline and to target selection and protein production. Portions of this research were carried out at the Stanford Synchrotron Radiation Lightsource (SSRL). The SSRL is a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health (National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences).

REFERENCES

1. Henriques AO, Moran CP, Jr, Structure, assembly, and function of the spore surface layers. Annu Rev Microbiol 2007;61:553–588.
2. Setlow P. Spore germination. Curr Opin Microbiol 2003;6:550–556.
3. Driks A. Overview: development in bacteria: spore formation in Bacillus subtilis. Cell Mol Life Sci 2002;59:389–391.
4. Driks A. Maximum shields: the assembly and function of the bacterial spore coat. Trends Microbiol 2002;10:251–254.
5. Bagyan I, Setlow P. Localization of the cortical lytic enzyme CowJ in spores of Bacillus subtilis. J Bacteriol 2002;184:1219–1224.
6. Hullo ME, Mosser I, Danchin A, Martin-Veraete I. CotA of Bacillus subtilis is a copper-dependent laccase. J Bacteriol 2001;183:5426–5430.
7. Abe A, Koide H, Kohno T, Watabe K. A Bacillus subtilis spore coat polypeptide gene, cotS. Microbiology 1995;141 (Pt 6):1433–1442.
8. Takamatsu H, Chikaiho Y, Kodauma T, Koide H, Kozuka S, Tochikubo K, Watabe K. A spore coat protein. Cot S, of Bacillus subtilis is synthesized under the regulation of sigma K and GreE during development and is located in the inner coat layer of spores. J Bacteriol 1998;180:2968–2974.
9. Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS, Losick R. The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis. PLoS Biol 2004;2:e328.
10. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Birns D, Bork P, Das U, Daugherty L, Duqueynne L, Finn RD, Gough J, Haft D, Hulo H, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen I, Mccaula C, McDowell J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimmia M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. InterPro: the integrative protein signature database. Nucleic Acids Res 2009;37:D211–D215.
11. Kannan N, Taylor SS, Zhai Y, Venter JC, Manning G. Structural and functional diversity of the microbial kinome. PLoS Biol 2007;5:e17.
12. Scheeff ED, Bourne PE. Structural evolution of the protein kinase-like superfamily. PLoS Comput Biol 2005;1:e49.
13. Hon WC, McKay GA, Thompson PR, Sweet RM, Yang DS, Wright GD, Berghuis AM. Structure of an enzyme required for aminoglycoside antibiotic resistance reveals homology to eukaryotic protein kinases. Cell 1997;88:887–895.
14. Peisach D, Gee P, Kent C, Xu Z. The crystal structure of choline kinase reveals a eukaryotic protein kinase fold. Structure (Camb) 2003;11:707–713.
15. Santarsiero BD, Yegian DT, Lee CC, Sparaggon G, Gu J, Scheibe D, Uber DC, Cornell EW, Nordmeyer RA, Kolbe WF, Jin J, Jones AL, Jaklevic JM, Schultz PG, Stevens RC. An approach to rapid protein crystallization using nanodroplets. J Appl Crystallogr 2002;35:278–281.
16. Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kirsch S, Sparaggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, McPhillips TM, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Eiblmer MA, Wooley J, Taylor SS, Hodgson KO, Wilson IA, Schultz PG, Stevens RC. Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline. Proc Natl Acad Sci USA 2002;99:11644–11649.
17. McPhillips TM, McPhillips SE, Chiu HJ, Cohen AE, Deacon AM, Ellis PJ, Garman E, Gonzalez A, Sauter NK, Phizackerley RP, Solits SM, Kuhn P. Blu-ice and the distributed control system: software for data acquisition and instrument control at macromolecular crystallography beamlines. J Synchrotron Radiat 2002;9:401–406.
18. CCP4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
19. Tickle IL, Laskowski RA, Moss DS. Error estimates of protein structure coordinates and deviations from standard geometry by full-matrix refinement of gamma2- and beta2-crystallin. Acta Crystallogr D Biol Crystallogr 1998;54:243–252.
20. Leslie A. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4+ESF-EACMB Newsletter Protein Crystallogr 1992;26.
21. Schneider TR, Sheldrick GM. Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 2002;58:1772–1779.
22. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 1999;55:849–861.
23. Terwilliger TC. Improving macromolecular atomic models at moderate resolution by automated iterative model building, statistical density modification and refinement. Acta Crystallogr D Biol Crystallogr 2003;59:1174–1180.
24. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60:2126–2132.
25. Winn MD, Murshudov GN, Papaz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol 2003;374:300–321.
26. Yang H, Guranovic V, Dutta S, Feng C, Berman HM, Westbrook JD. Automated and accurate deposition of structures solved by X-ray diffraction to the protein data bank. Acta Crystallogr D Biol Crystallogr 2004;60:1833–1839.
27. Davis IW, Leaver-Fay A, Chen VB, Block KN, Kapral GJ, Wang X, Murray AW, Arendall WB, III, Sneyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35:W375–W383.
28. Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:52–56.
29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
30. Holm L, Park J, Daliliw workbench for protein structure comparison. Bioinformatics 2000;16:566–567.
31. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.

32. Markowitz VM, Szeto E, Palaniappan K, Grechkin Y, Chu K, Chen JM, Dubchak I, Anderson I, Lykidis A, Mavromatis K, Ivanova NN, Kyrpides NC. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. Nucleic Acids Res 2008;36:D528–D533.

33. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–1797.

34. Li W, Jaroszewsli L, Godzik A. Tolerating some redundancy significantly speeds up clustering of large protein databases. Bioinformatics 2002;18:77–82.

35. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics 2002;Chapter 2:Unit 2.3.

36. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res 2004;14:1188–1190.

37. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 2005;33:W299–W303.

38. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2005;54:597–609.

39. Ludwig W, Schleifer KH, Whitman WB. Revised road map to the domain Bacteria. Bergey’s Manual of Systematic Bacteriology: Volume 3:The Firmicutes. New York: Springer; 2009.

40. Lai EM, Phadke ND, Kachman MT, Giorno R, Vazquez S, Vazquez JA, Goodrum JR, Driks A. Proteomic analysis of the spore coats of Bacillus subtilis sporulation genes in Bacillus subtilis. Proc Natl Acad Sci USA 2003;100:74–79.

41. Raymond J, Chen M, Blankenship RE, Touchman JW. The genome sequence of the photosynthetic alga Tetrahymena thermophila. Biochim Biophys Acta 2005;174:201–206.

42. Clepet C, Borne F, Krishnapillai V, Baird C, Patte JC, Cami B. Isolation, organization and expression of the Pseudomonas aeruginosa threonine genes. Mol Microbiol 1992;6:3109–3119.

43. Holm L, Sander C. Mapping the protein universe. Science 1996;273:595–603.

44. Malito E, Sekulic N, Too WC, Konrad M, Lavie A. Elucidation of human choline kinase crystal structures in complex with the products ADP or phosphocholine. J Mol Biol 2006;364:136–151.

45. Bossemeyer D, Engh RA, Kinzel V, Ponstingl H, Huber R. Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 A structure of the complex with Mn2+ adenyl imidodiphosphate and inhibitor peptide PKI(5-24). Embo J 1993;12:849–859.

46. Eichenberger P, Radzio-Andzelm E. Three protein kinase structures define a common motif. Structure 1994;2:345–355.

47. Kannan N, Neuwald AF. Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? J Mol Biol 2005;351:956–972.

48. Nolen B, Taylor S, Ghosh G. Regulation of protein kinases: controlling activity through activation segment conformation. Mol Cell 2005;19:661–675.

49. Sattley WM, Madigan MT, Connolly EM, van Ooij C, Eichenberger P, Driks A. The sigmaE regulon and the identification of additional sporulation genes in Bacillus subtilis. Mol Microbiol 2006;59:487–502.

50. Madhusudan, Akamine P, Xuong NH, Taylor SS. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase. Nat Struct Biol 2002;9:273–277.

51. Madhusudan, Akamine P, Xuong NH, Taylor SS. Crystal structure of a transition state mimic of the catalytic subunit of cAMP-dependent protein kinase. Nat Struct Biol 2002;9:273–277.

52. Fong DH, Berghuis AM. Substrate promiscuity of an aminoglycoside antibiotic resistance enzyme via target mimicry. Embo J 2002;21:2323–2331.