BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

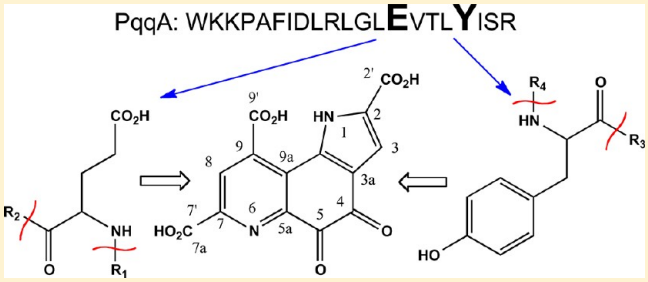
Distribution and Properties of the Genes Encoding the Biosynthesis of the Bacterial Cofactor, Pyrroloquinoline Quinone

Yao-Qing Shen,[⊥] Florence Bonnot,[†] Erin M. Imsand,[†] Jordan M. RoseFigura,[†,@] Kimmen Sjölander,*[,⊥,§,∥] and Judith P. Klinman*[,⊥,†,‡]

[†]Department of Chemistry, [‡]Department of Molecular and Cell Biology, [§]Department of Bioengineering, [∥]Department of Plant and Microbial Biology, and [⊥]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, California 94720, United States

Supporting Information

ABSTRACT: Pyrroloquinoline quinone (PQQ) is a small, redox active molecule that serves as a cofactor for several bacterial dehydrogenases, introducing pathways for carbon utilization that confer a growth advantage. Early studies had implicated a ribosomally translated peptide as the substrate for PQQ production. This study presents a sequence- and structure-based analysis of the components of the *pqq* operon. We find the necessary components for PQQ production are present in 126 prokaryotes, most of which are Gram-negative and a number of which are pathogens. A total of five gene products, PqqA, PqqB, PqqC, PqqD, and PqqE, are identified as being obligatory for PQQ production. Three of the gene products in the *pqq* operon, PqqB, PqqC, and PqqE, are members of large protein superfamilies. By combining evolutionary conservation patterns with information from three-dimensional structures, we are able to differentiate the gene products involved in PQQ biosynthesis from those with divergent functions. The observed persistence of a conserved gene order within analyzed operons strongly suggests a role for protein−protein interactions in the course of cofactor biosynthesis. These studies propose previously unidentified roles for several of the gene products, as well as identifying possible new targets for antibiotic design and application.



Pyrroloquinoline quinone (PQQ) is a low-molecular weight, redox active cofactor utilized by a number of prokaryotic dehydrogenases.[1,2] Although the cofactor is not required for bacterial survival, the presence of this molecule has been shown to enhance the rate of cell growth.[3] Some prokaryotic organisms are capable of synthesizing the redox active molecule, while other species rely on the environment for their supply. The biosynthesis of PQQ is accomplished by the gene products of a specific *pqq* operon. In *Klebsiella pneumoniae*, an organism with the experimentally demonstrated ability to produce PQQ, the *pqq* operon comprises six genes, *pqqA*−*F*[4] (the expressed genes from *K. pneumoniae* and four other demonstrated PQQ producers are summarized in Figure 1 and Table S1 of the Supporting Information). Genetic knockout studies show four of the six gene products (PqqA, PqqC, PqqD, and PqqE) are absolutely required for this pathway, while the role of PqqB is ambiguous.[5]

The successful in vitro characterization of two gene products, PqqC and PqqE, has demonstrated, first, that PqqC is a cofactorless, oxygen-activating enzyme catalyzing the final step in PQQ biosynthesis[6] and, second, that PqqE is a functional radical SAM enzyme capable of catalytic reductive cleavage of SAM to methionine and 5′-deoxyadenosine.[7] The putative substrate for PqqE is PqqA, a 22-amino acid peptide containing a conserved glutamate and tyrosine that provide the complement of carbon and nitrogen atoms required for PQQ synthesis

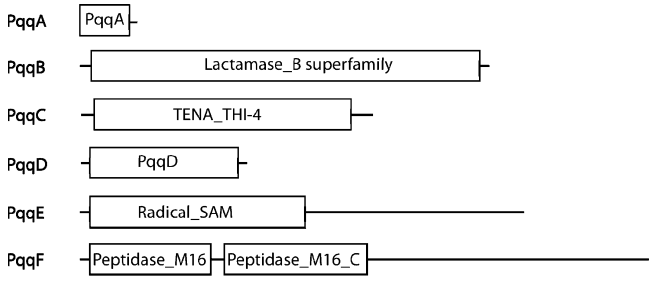


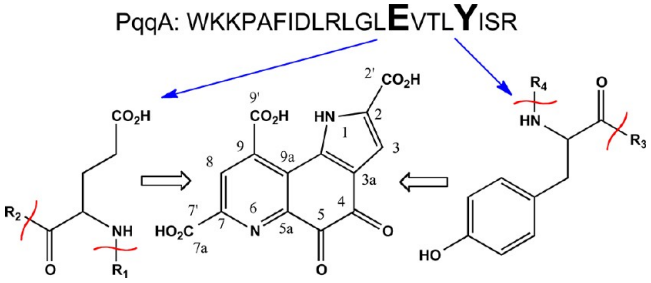
Figure 1. Domains in each *K. pneumoniae* Pqq protein. All domain information is from the Pfam database, except for that for the Lactamase_B superfamily domain in PqqB, which comes from the Conserved Domain Database of the National Center for Biotechnology Information.

(Scheme 1);[8] however, the ability of or conditions required for PqqE to functionalize PqqA have not yet been identified. The roles for PqqB, PqqD, and PqqF in PQQ production are even less clear. PqqD has recently been shown to interact physically with PqqE,[9] though the catalytic relevance of this interaction

Received: November 29, 2011
Revised: January 23, 2012
Published: February 10, 2012

dx.doi.org/10.1021/bi201763d | *Biochemistry* 2012, 51, 2265−2275

Scheme 1. Proposed Cross-Linking of Tyrosine and Glutamate To Form PQQ



has yet to be determined. Genetic knockout studies of PqqF, a protein with homology to zinc-dependent proteases, suggest it is not essential for PQQ production, with the implication that other cell-associated, nonspecific proteases can assume its role during cofactor biogenesis.[5] One of the most enigmatic gene products is PqqB with a high degree of sequence similarity to the family of metallo-β-lactamases.

It was determined in 1988 that PQQ is derived from a ribosomally translated peptide.[10] Since that time, several other biologically active molecules produced from amino acid precursors have been discovered.[11−13] Analysis of these biosynthetic pathways reveals several common gene products, including radical SAM enzymes (e.g., PqqE), metallo-β-lactamases (e.g., PqqB), small, cofactorless proteins (e.g., PqqD), and cofactorless oxygenases (e.g., PqqC), in addition to the expected peptidases (e.g., PqqF).[12−15] The common protein families and underlying structure of biosynthetic pathways suggest that elucidating the evolution of PQQ biosynthesis may be useful for the discovery of other natural products and the characterization of the pathways required for their biosynthesis. In particular, bioinformatic analysis of the growing number of genomic sequences for prokaryotic organisms has revealed orphan pathways with unknown products of potential therapeutic application.[11,16] Indeed, gene products from the *pqq* operon are already being used as guides for identifying such pathways.[17] Determining the evolution of each gene in the PQQ biosynthetic pathway may also contribute to understanding the ubiquitous use of certain protein families in the modification of peptides to form biologically active natural products.

This paper presents a bioinformatics analysis of the genes involved in PQQ biosynthesis to identify the essential, biosynthetic *pqq* genes and the species that contain the full complement of these genes (and, thus, are inferred to synthesize PQQ). Structural phylogenomic analyses were used to identify the sequence motifs and structural features that distinguish each biosynthetic protein from functionally divergent homologues.[18] These studies serve as a guide for predicting putative roles for the open reading frames within the *pqq* operon and probing the contribution of conserved amino acid side chains within the gene products with demonstrated function.

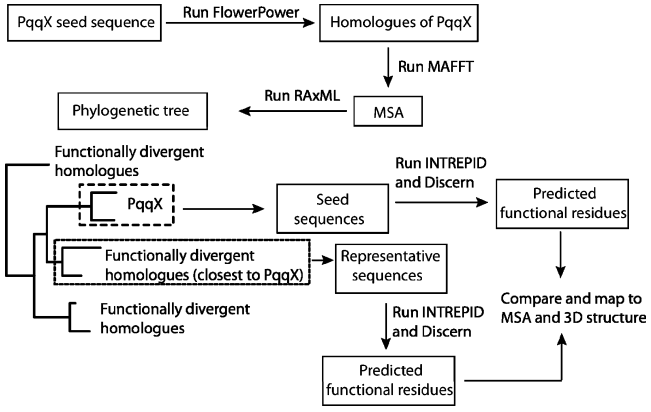
METHODS

Data Set. Five bacteria with demonstrated PQQ biosynthetic capacity were selected for this study: *K. pneumoniae*, *Methylobacterium extorquens* AM1, *Gluconobacter oxidans* 621H, *Rahnella aquatilis*, and *Streptomyces rochei*.[19−23] Each of the genes in the *pqq* operons in these five bacteria (shown in Table 1) was used as a starting point for bioinformatics analyses. *pqqA*−*E* are found in all five bacteria, while the sixth *pqq* gene, *pqqF*, is found in only two of the five species (*K. pneumoniae* and *R. aquatilis*). *M. extorquens* has a fused *pqqC/D* gene and also an extra gene, *pqqG* (not included in this study).

Bioinformatics Analyses. Bioinformatics analyses were performed to characterize the taxonomic distribution of homologues, to identify corresponding PFAM domains,[24] and to differentiate homologues sharing the same function from those that have divergent function. Core genes were identified from these combined analyses and used to identify species with probable PQQ biosynthetic capability.

For each of the genes in the PQQ operon for these five species, we performed the following analyses: (i) homologue identification, (ii) multiple sequence alignment, (iii) phylogenetic tree construction and analysis, (iv) protein structure prediction, and (v) functional site identification (see Scheme 2).

Scheme 2. Bioinformatics Analysis Used in This Study[a]



[a]Abbreviations: Pqq*X*, a representative of Pqq proteins; MSA, multiple sequence alignment.

Table 1. Reference Sequences of PqqA−F from Five Known PQQ-Forming Organisms[a]

	K. pneumoniae (X58778)	*M. extorquens* (NC_012808)	*G. oxidans* (CP000009)	*S. rochei* (AB088224)	*R. aquatilis* (FJ868974)
PqqA	P27503 (PQQA_KLEPN)	Q49148 (PQQA_METEA)	Q9L3B4 (PQQA_GLUOX)	Q83X96 (Q83X96_STRRO)	C3VIT4 (C3VIT4_RAHAQ)
PqqB	B5XX61 (PQQB_KLEP3)	Q49149 (PQQB_METEA)	Q9L3B3 (PQQB_GLUOX)	Q83XA0 (Q83XA0_STRRO)	C3VIT5 (C3VIT5_RAHAQ)
PqqC	B5XX60 (PQQC_KLEP3)	Q49150 (PQQCD_METEA)	Q9L3B2 (PQQC_GLUOX)	Q83X97 (Q83X97_STRRO)	C3VIT6 (C3VIT6_RAHAQ)
PqqD	B5XX59 (PQQD_KLEP3)	Q49150 (PQQCD_METEA)	Q9L3B1 (PQQD_GLUOX)	Q83X98 (Q83X98_STRRO)	O33505 (PQQD_RAHAQ)
PqqE	B5XX58 (PQQE_KLEP3)	P71517 (PQQE_METEA)	Q9L3B0 (PQQE_GLUOX)	P59749 (PQQE_STRRO)	O33506 (PQQE_RAHAQ)
PqqF	P27508 (PQQF_KLEPN)	C5AQL6 (C5AQL6_METEA)	absent	absent	C3VIT9 (C3VIT9_RAHAQ)

[a]The UniProt accession number for each protein is listed. In parentheses are UniProt identifiers.

dx.doi.org/10.1021/bi201763d | *Biochemistry* 2012, 51, 2265−2275

Homologues were retrieved from the UniProt protein database (release 2010_11) using the FlowerPower phylogenomic clustering software to select proteins sharing the same domain architecture.[25] Multiple sequence alignments (MSAs) were constructed using MAFFT,[26] followed by masking to remove columns with >70% gap characters. Maximum likelihood trees were estimated from the masked MSAs using RAxML.[27]

A determined three-dimensional structure was available for PqqC from *K. pneumoniae*;[6] protein structures were predicted for PqqB and PqqD from *K. pneumoniae* using a comparative modeling approach[28,29] with determined three-dimensional structures from other organisms as templates (PqqB from *Pseudomonas putida* and PqqD from *Xanthomonas campestris*[30]). In the case of PqqE, where no structures are yet available, the closest homologue with a determined structure {MoaA from *Staphylococcus aureus* [Protein Data Bank (PDB) entry 1TV8]} was used to build a comparative model. The Phyre and Phyre2 servers[31] were used to construct comparative models for PqqB and PqqE; Modeler[29] was used to predict the structure of PqqD.

Residues that were likely to be functionally important were identified via a combination of evolutionary conservation and structural information using the INTREPID and Discern algorithms.[32,33] INTREPID uses conservation signals over divergently related homologues organized by a phylogenetic tree to predict functional sites, while Discern uses INTREPID scores as well as information from protein three-dimensional structures. The multiple sequence alignments and phylogenetic trees were used as input to INTREPID, and the INTREPID scores and comparative models or determined structures were used as input to Discern. The Discern results are reported for PqqB–D. As the C-terminal portion of the PqqE model is not reliable, INTREPID results were reported for PqqE. These analyses identified a set of amino acids for each protein that are likely to be functionally important.

Most of the genes in the *pqq* operon are members of gene superfamilies, including paralogs that may have diverged functionally. To identify protein residues that are diagnostic of participation in PQQ biosynthesis, we performed identical analyses on representative sequences with divergent function from the most closely related sister clade in the reconstructed phylogenies for PqqB (Figure S1 of the Supporting Information), PqqC (Figure S2 of the Supporting Information), PqqD (Figure S3 of the Supporting Information), and PqqE (Figure S4 of the Supporting Information). Residues highly ranked by either INTREPID or Discern for the biosynthetic proteins were mapped onto structures and comparative models for *K. pneumoniae*, and the corresponding residues in the functionally divergent homologues were plotted on structures or models for these proteins.

Additional Details about Bioinformatics Methods. HMM global homology clustering of homologues used FlowerPower default parameters, the number of subfamily HMM scoring iterations (set to 10). MAFFT MSA construction parameters were as follows: maximum of five iterations and default parameters. RAxML parameters were as follows: JTT+Γ model and 20 discrete γ-rate categories. The statistical support of branches was estimated using 100 bootstrap replicates. Because of the short length of the *pqqA* gene, standard genome annotation pipelines frequently missed *pqqA* (i.e., these were false negatives in gene prediction pipelines). We supplemented our standard homology detection pipeline for PqqA using translated BLAST against whole genomes for these cases.[34]

RESULTS AND DISCUSSION

Species with PQQ Biosynthetic Capability. On the basis of genetic knockout studies, PqqC, PqqD, and PqqE were initially considered to be the most promising core set of proteins required for inferring PQQ biosynthetic capability.[20] The number of species containing different subsets of the core proteins is detailed in Figure 2. This analysis returned 126 species

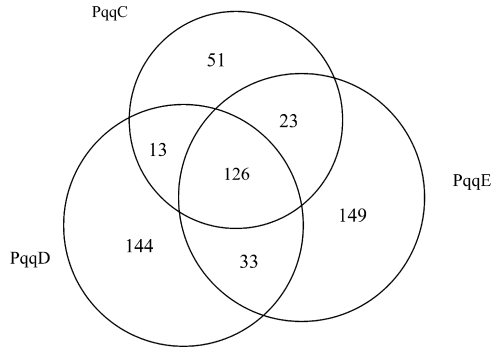


Figure 2. Venn diagram of core PQQ biosynthesis proteins (PqqC–E). Each circle represents the number of species containing the corresponding gene. PqqC homologues were found in 213 species. PqqD homologues were found in 316 species. PqqE homologues were found in 331 species. All three genes were found in 126 species.

that include PqqC–E, of which 125 also contain PqqB (Table S2 of the Supporting Information), strongly implicating PqqB as being essential for PQQ biosynthesis. Because of the short length of PqqA (fewer than 30 amino acids), standard genome annotation pipelines failed to detect an ORF for *pqqA* in many cases. The standard HMM-based pipeline (in which we searched for proteins deposited in the UniProt database for other biosynthetic proteins) was supplemented with translated BLAST against whole genomes for these cases. Specifically, translated BLAST was used with each of the PqqA peptides from our five seed organisms against whole genomes in which PqqB–E proteins had been detected. This identified an additional 37 *pqqA* genes that had been missed by the genome annotation pipelines for those species (Table S3 of the Supporting Information). To summarize, of the 126 species for which our analyses support PQQ biosynthetic capability, a total of 107 contain an identifiable PqqA (Table S2 of the Supporting Information); of these 126 species, 95 have whole genomes, of which 98% (93) contain PqqA (based on either the HMM methods or translated BLAST as described in Methods). This result strongly supports the hypothesis that PQQ biosynthetic capability requires PqqA–E and that PqqA is the substrate for the biosynthetic pathway.

This analysis shows that while some individual PQQ biosynthetic proteins have distant homologues outside prokaryotes, the pathway is clearly specific to prokaryotes. The vast majority (approximately 88%) of the species predicted to be PQQ-forming are proteobacteria, with the α-, β-, and γ-classes of this phylum well represented (Figure 3 and Table S2 of the Supporting Information). Altogether, only six Gram-positive bacteria were found to be PQQ-forming, showing that PQQ production is more prevalent in Gram-negative than Gram-positive organisms. These results are consistent with the fact that either the enzymes or catalytic domains of enzymes requiring

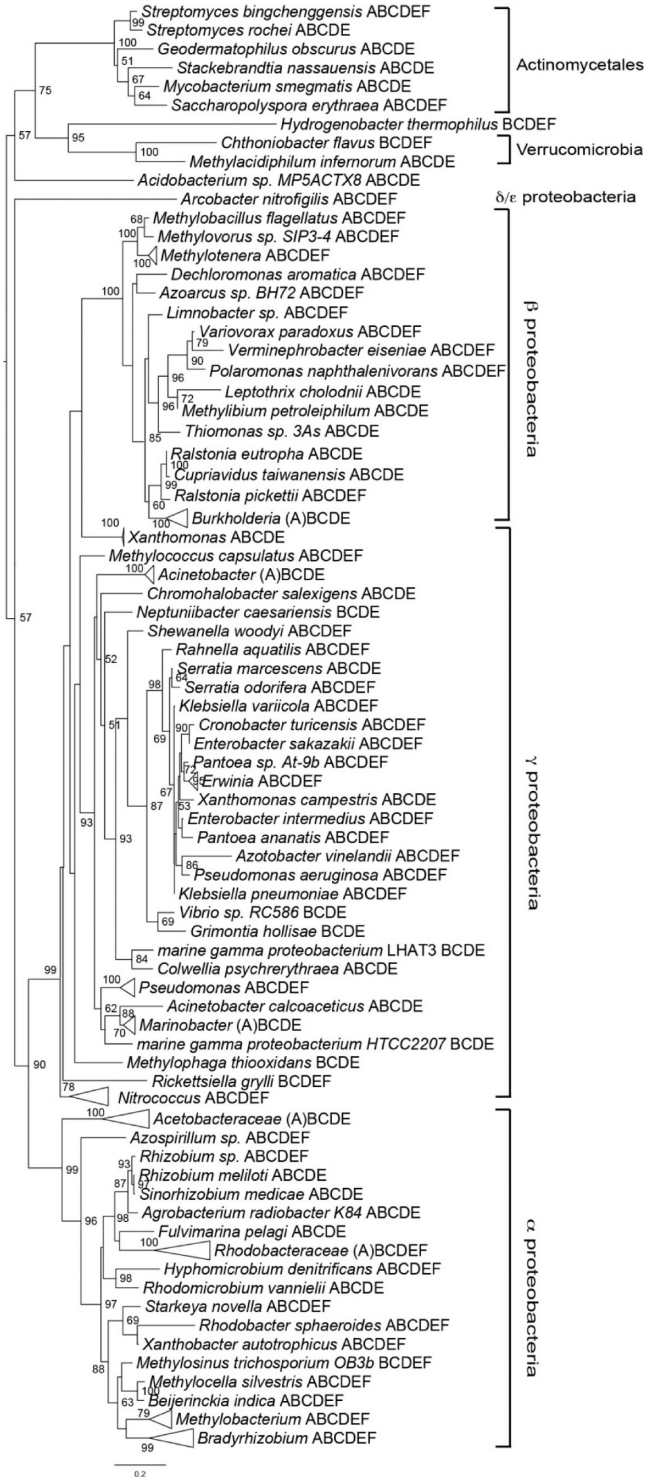


Figure 3. Distribution of PQQ-synthesis proteins mapped to a phylogenetic tree based on 16s rRNA. The letters following species names indicate Pqq proteins. A–F represent PqqA–PqqF, respectively. Tree nodes were collapsed if all the species have the same set of Pqq proteins (PqqA was labeled with A if it was found in at least one species in the collapsed nodes).

PQQ as a cofactor have, to date, been described in the context of a localization in the periplasm of Gram-negative bacteria.[1,35] Two species of *Verrumicrobia*, a recently discovered phylum that is a sister to *Chylamidae*, were also found to contain the homologues of PqqB–E.

While a role for bacterially derived PQQ in mammalian homeostasis has been proposed,[36] the exclusive mapping of the PQQ-generating enzymes to bacteria, together with the growth advantage conferred to selected prokaryotes by PQQ-dependent metabolism, makes the PQQ biosynthetic pathway an enticing target for the design of inhibitors for use in antibiotic cocktails. A number of the PQQ-forming species identified in this study can operate as opportunistic pathogens (Table 2), suggesting that inhibitors of the pathway may be especially well suited to a cocktail of antibiotics administered to immune-suppressed patients.

Organization and Evolution of the PQQ Operon. Of the 126 species identified here as synthesizing PQQ, 96 have gene order information, either based on whole genomes or because of the availability of contigs including all *pqq* genes. Of these 96, all of which contain *pqqA–E*, 93 have no large insertions within the operon. Among these 93, 91 have the conserved gene order *pqqA-B-C-D-E*. In total, 96% of whole genomes having *pqq* operons have a conserved order of the core genes with no large insertions.

In only 17 species is *pqqF* clustered with *pqqB–E*, either in the order *pqqB-C-D-E-F* or in the order *pqqF-A-B-C-D-E*. In all other cases (27 species), *pqqF* is remote from the *pqqA–E* cluster, separated by genes unrelated to the pathway. The distance from *pqqF* to other *pqq* genes suggests that *pqqF* is out of the evolutionary and regulatory constraints exerted on other *pqq* genes. This finding also agrees with the hypothesis that the function of PqqF may be replaced by other peptidases.

The specific ordering of *pqqA–E* in the operon suggests that these gene products may form a catalytically relevant complex. Conservation of gene order within an operon is not characteristic of prokaryotic organisms,[37,38] and when gene order is conserved in an operon, the corresponding gene products have been shown to physically interact.[39,40] Protein–protein interactions are common in metabolic pathways, and these interactions can serve to increase the efficiency of the overall process by positioning consecutively acting enzymes in the proximity of one another and by protecting sensitive intermediates from degradation via direct channeling between these enzymes. In the PQQ biosynthetic pathway, protecting PqqA during the biosynthetic process would shield the peptide from proteolytic degradation before completion of the necessary modifications. The absence or remoteness of *pqqF* from the operon suggests that this gene product is not involved in protein–protein interactions with other members of the pathway.

The proposed interaction of *pqq* gene products is further supported by the fusion of *pqqC* and *pqqD* in four *Methylobacterium* species: *M. radiotolerans*, *M. populi*, *M. extorquens*, and *Methylobacterium chloromethanicum*. In addition, we have found a potential fusion of *pqqD* and *pqqE* in *Methylosinus trichosporium*. The sequence is annotated as PqqE (UniProt entry D5QP91), but it clearly possesses a PqqD domain.

The phylogenetic profile of each *pqq* gene was mapped to the 16s rRNA tree of PQQ-forming species identified in this study as inferring the evolution of the PQQ biosynthetic pathway (Figure 3). As suggested by the gene locus study, PqqB–E appear to have evolved together in the pathway. PqqF appears to have been lost several times (e.g., in *Acetobacteraceae*, *Burkholderia*, and *Acinetobacter*), and presumably, its function has been replaced by other peptidases in these species. Despite the limitations of bioinformatics methods in gene identification and in the reliable identification of short peptides (see above), PqqA was identified in most species (Figure 3).

Table 2. Potential Pathogens or Symbionts Containing PqqB−E, Ordered by Species Name[a]

species	pathogenicity	set of *Pqq* genes
Acinetobacter baumannii ATCC 19606	opportunistic human pathogen	ABCDE
Acinetobacter hemolyticus ATCC 19194	rare human pathogen	BCDE
Azoarcus sp. (strain BH72)	plant symbiont	ABCDEF
Bradyrhizobium japonicum	plant symbiont	ABCDEF
Bradyrhizobium sp. (strain BTAi1/ATCC BAA-1182)	plant symbiont	ABCDEF
Burkholderia cenocepacia (strain HI2424)	human cystic fibrosis pathogen	ABCDE
Burkholderia cepacia (strain J2315/LMG 16656)	animal and plant pathogen	ABCDE
Burkholderia glumae (strain BGR1)	plant pathogen	ABCDE
Burkholderia multivorans (strain ATCC 17616/249)	animal pathogen in mammals	ABCDE
Burkholderia phymatum (strain DSM 17167/STM815)	plant symbiont	ABCDE
Colwellia psychrerythraea (strain 34H/ATCC BAA-681)	plant pathogen	ABCDE
Cronobacter turicensis (strain DSM 18703/LMG 23827/z3032)	neonatal pathogen	ABCDEF
Cupriavidus taiwanensis (strain R1/LMG 19424)	plant pathogen	ABCDE
Dinoroseobacter shibae (strain DFL 12)	animal symbiont	ABCDEF
Enterobacter intermedius	opportunistic pathogen	ABCDEF
Enterobacter sakazakii (strain ATCC BAA-894)	neonatal pathogen	ABCDEF
Erwinia amylovora (strain ATCC 49946/CCPPB 0273/Ea273/27-3)	plant pathogen	ABCDEF
Erwinia pyrifoliae	plant pathogen	ABCDEF
Erwinia tasmaniensis (strain DSM 17950/Et1/99)	plant commensal	ABCDEF
Gluconacetobacter diazotrophicus (strain ATCC 49037/DSM 5601/PAl5)	plant symbiont	ABCDE
Granulibacter bethesdensis (strain ATCC BAA-1260/CGDNIH1)	human pathogen	ABCDE
Grimontia hollisae CIP 101886	human pathogen	BCDE
K. pneumoniae	opportunistic pathogen	ABCDEF
Klebsiella sp. 1_1_55	opportunistic pathogen	BCDEF
Marinobacter algicola DG893	plant symbiont	ABCDE
Methylobacterium nodulans (strain ORS2060/LMG 21967)	plant saprophyte and symbiont	ABCDEF
Methylobacterium populi (strain ATCC BAA-705/NCIMB 13946/BJ001)	plant endophyte	ABCDEF
Methylobacterium radiotolerans (strain ATCC 27329/DSM 1819/JCM 2831)	plant symbiont	ABCDEF
Mycobacterium smegmatis [strain ATCC 700084/mc(2)155]	commensal in mammals	ABCDE
Pantoea ananatis	plant pathogen	ABCDEF
Pseudomonas aeruginosa	opportunistic human pathogen	ABCDEF
Pseudomonas entomophila (strain L48)	insect pathogen	ABCDEF
Pseudomonas fluorescens	potential pathogen to birds	ABCDEF
Pseudomonas mendocina (strain ymp)	rare human pathogen	ABCDEF
Pseudomonas savastanoi pv savastanoi NCPPB 3335	tumor-inducing pathogen	ABCDEF
Pseudomonas syringae pv phaseolicola (strain 1448A/Race 6)	plant symbiont	ABCDEF
R. aquatilis	opportunistic pathogen	ABCDEF
Ralstonia pickettii (strain 12D)	opportunistic pathogen	ABCDEF
Rhizobium meliloti	plant symbiont	ABCDE
Rhizobium sp. (strain NGR234)	plant symbiont	ABCDEF
Rickettsiella grylli	arthropod pathogen	BCDEF
Serratia marcescens	opportunistic human pathogen	ABCDE
Serratia odorifera 4Rx13	oppotunistic pathogen	ABCDEF
Verminephrobacter eiseniae (strain EF01-2)	animal endosymbiont	ABCDEF
Xanthomonas axonopodis pv citri (citrus canker)	plant pathogen	ABCDE
Xanthomonas campestris pv Campestris	plant pathogen	ABCDE
Xanthomonas oryzae pv Oryzae	rice bacterial blight pathogen	ABCDE

[a]See Table S2 of the Supporting Information for a complete list of predicted PQQ synthesis species.

Structure−Sequence Analyses of Core *pqq* Genes. PqqB−E are each found in superfamilies whose functions may have diverged from their common ancestor by gene duplication events, allowing members of the family to acquire novel functions or to partition the ancestral function (Figure 1 and Table S1 of the Supporting Information). Evolutionary and structural analyses were performed to identify residues that are likely to be functionally important.

Structure−Sequence Analysis of PqqB. Initially, sequence analysis of PqqB suggested that the enzyme was a member of the metallo-β-lactamase superfamily.[41] The determined three-dimensional structure of PqqB (PDB entries 1XTO and 3JXP) has since confirmed this. Analysis of the phylogenetic tree of PqqB homologues shows that the closest functionally distinct homologue of PqqB is PhnP (Figure S1 of the Supporting Information). PhnP is a phosphodiesterase with a sequence 26% identical with that of PqqB.[42] The comparison of PqqB to PhnP (Figure 4) shows the overall structural similarity between PhnP and PqqB; however, the ligands for one of the two metals observed to be bound in PhnP are not present in PqqB. The predicted metal ligand residues that are retained in PqqB (Asp92, His93, and His269 in PqqB from *K. pneumoniae*, colored red in Figure 4A) are ranked 6, 1, and 3, respectively, by the Discern algorithm (Table S4 of the Supporting Information). The active site metals bound by PhnP are both manganese;[42] however, the metal ligands retained in PqqB represent a 2-His/1-carboxylate facial triad configuration, characteristic of the nonheme ferrous binding family. While future studies with recombinant PqqB are necessary to demonstrate metal binding and function, this arrangement of ligands suggests that PqqB likely catalyzes oxidative chemistry via a conserved chemical mechanism that involves the generation of an active site Fe^{IV}═O.[43−45] This chemistry has precedent among members of the metallo-β-lactamase superfamily.[46] Additionally, in the three-dimensional structure of PqqB (PDB entry 3JXP from *P. putida*), a zinc ion is bound several angstroms from this putative active site near the solvent interface by the $CXCX_2C$ motif (residues 19−24 in PqqB from *K. pneumoniae*, colored blue in Figure 4). This metal is conserved in both PqqB and PhnP. Studies of PhnP have shown that this zinc is not necessary for catalysis, and the ion is proposed to play a role in maintaining the structure of the enzyme.[42] The comparable location of this zinc ion in PqqB suggests that it may also play a structural role. Lastly, it can be noticed that a glycine-rich motif (GXXXGGGXPQWN, residues 7−18 in PqqB from *K. pneumonia*), located between the proposed structural zinc ion and the putative active site of PqqB

dx.doi.org/10.1021/bi201763d | *Biochemistry* 2012, 51, 2265−2275

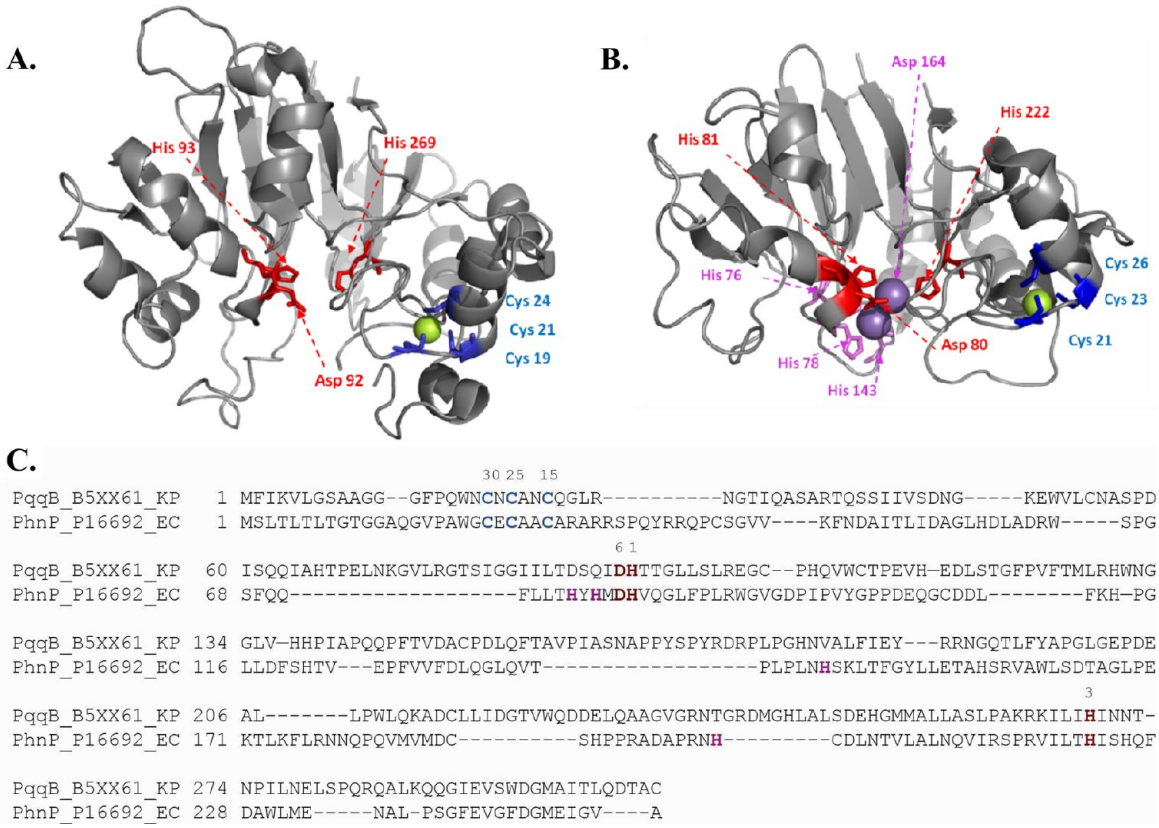


Figure 4. Sructure–sequence analysis of PqqB. The spheres of Mn ions in the active site are colored purple, and the sphere of the zinc ion is colored green. Residues involved in putative structural Zn ion binding are colored blue; Residues involved in Mn ions binding to the PhnP active site are colored red if conserved in PqqB and purple if not. (A) Homology model of the structure of PqqB from *K. pneumoniae*, KP (UniProt entry B5XX61) modeled onto the determined structure of PqqB from *P. putida* (66% identical). (B) Structure of the PhnP sequence from *E. coli*, EC (UniProt entry P16692, PDB entry 3G1P). (C) Pairwise alignment of PqqB and PhnP. The numbers above the columns indicate the ranking predicted by Discern for the corresponding residue of PqqB (Table S4 of the Supporting Information).

and exposed to solvent is conserved in 80% of PqqB sequences from the 126 proposed PQQ producers. In PhnP, this motif is less conserved, indicating that it might impart a functional specificity to PqqB. This motif has features (G repeats) common to nucleotide-binding pockets,[47,48] suggesting a possible role for a nucleotide cofactor as an electron donor in a PqqB-catalyzed activation of O_2. This bioinformatics analysis indicates that PqqB may be the missing hydroxylase, performing a role in the oxidation of the tyrosine of PqqA prior to its cross-linking with glutamate (Scheme 1).

Structure–Sequence Analysis of PqqC. The activity of PqqC from *K. pneumoniae* has been determined.[6,49] PqqC catalyzes the last step of PQQ biosynthesis, oxidation of 3a-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic acid (AHQQ), which involves the transfer of eight electrons and protons to molecular oxygen to form hydrogen peroxide/water. The closest functionally distinct homologue of PqqC identified in the phylogenetic tree is TenA (Figure S2 of the Supporting Information). TenA is a thiaminase II that catalyzes the hydrolysis of 4-amino-5-aminomethyl-2-methylpyrimidine as part of a pathway for salvaging base-degraded thiamin.[50] Although both proteins contain buried active sites, the reaction catalyzed by PqqC is very different from that catalyzed by TenA. The comparison of the structures of PqqC and TenA (Figure 5) shows their overall structural similarity, but the active site of PqqC is distinct from that of TenA. Discern analysis of PqqC revealed a cluster of residues in and around the active site (Table S5 of the Supporting Information). In the top 20 ranked residues, four are outside the active site and could play a role in opening the active site to the substrate (colored blue in Figure 5), with the 16 remaining residues located in the active site. Only one of these residues is conserved in TenA at the corresponding position (according to the structural alignment constructed by VAST, colored red in Figure 5). The residues of potential catalytic importance in PqqC (Tyr23, His24, Arg50, Gln54, Arg80, His84, Tyr128, Glu147, His154, Arg157, Tyr175, Arg179, and Lys214, colored green in Figure 5) that are not conserved in TenA likely contribute to the specificity of PqqC activity. This suggests that each enzyme in the family has evolved a distinct function highly specific to the pathway in which it is active, as evidenced by the differing active site side chains of PqqC and TenA. Residues previously shown to be important in PqqC by biochemical experiments were identified as the most likely important functional residues in our study. His84 is proposed to be a proton donor and is ranked 4th by Discern;[51] His154, Tyr175, and Arg179 are proposed to form a core oxygen-binding pocket essential for oxygen activation[52] and are ranked 10th, 7th, and 26th, respectively, by Discern (labeled with asterisks in Figure 5).

Structure–Sequence Analysis of PqqD. PqqD is a small protein (90 amino acids on average) with no detectable cofactor. In this study, the Discern algorithm was used with the PqqD sequence from *K. pneumoniae* and a constructed homology model for *K. pneumoniae* (using PDB entry 3G2B from

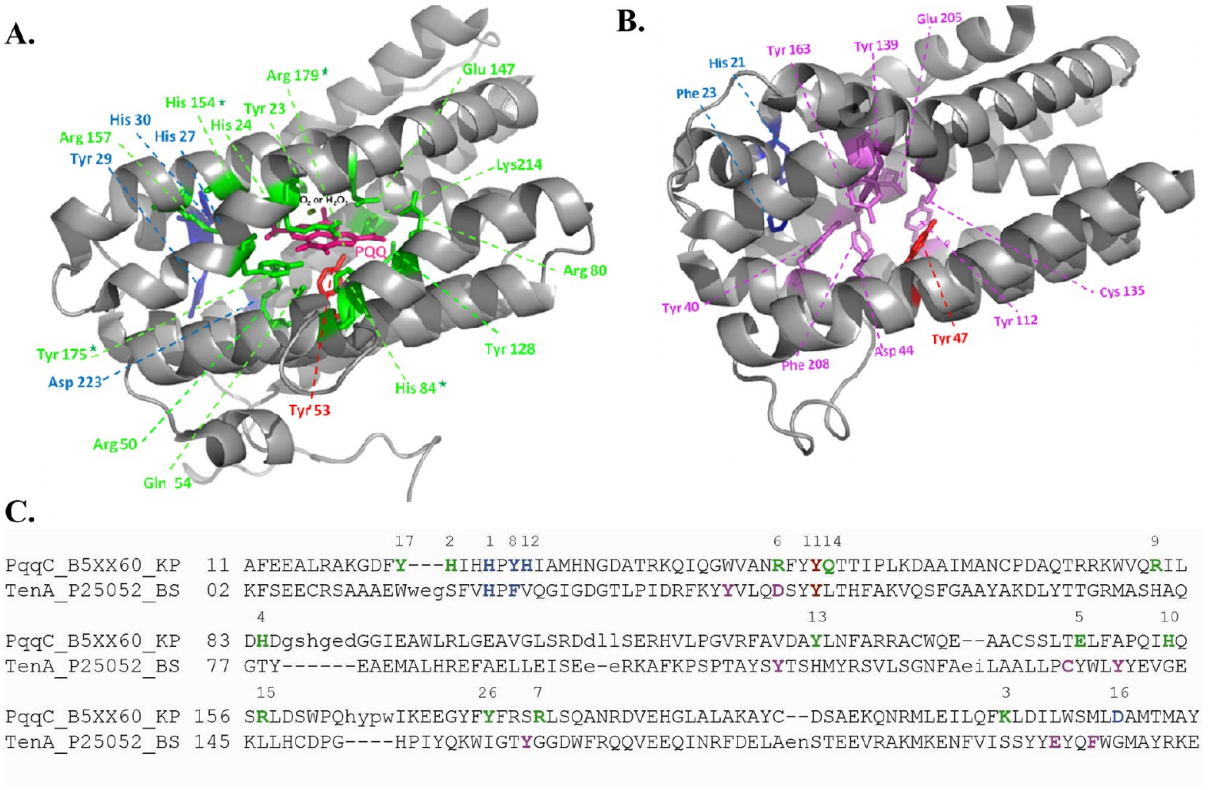


Figure 5. Structure–sequence analysis of PqqC. (A) Structure of PqqC (B5XX60 from *K. pneumoniae*, KP, PDB entry 1OTW). The top-ranked residues predicted by Discern are labeled. Residues outside the active site are labeled in blue; the residue in the active site is labeled in red if conserved in TenA and in green if not. The residues characteristic of biochemical experiments are marked with asterisks. (B) Structure of TenA from *B. subtilis*, BS (P25052, PDB entry 1yaf). Predicted functional residues are colored pink. (C) Pairwise structural alignment of PqqC and TenA (created with VAST[53]). The numbers above the columns indicate the ranking of catalytic residues in PqqC predicted by Discern (Table S5 of the Supporting Information).

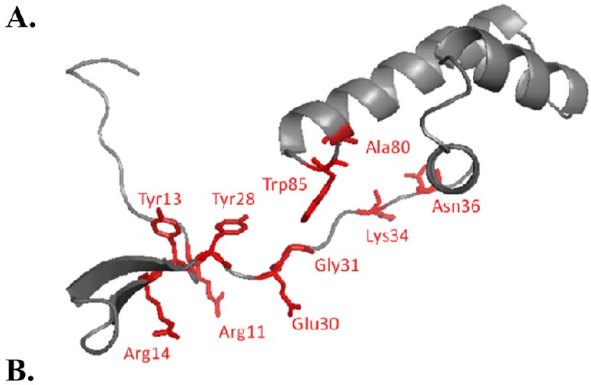
X. campestris as a template with a sequence that is 29% identical). In the case of PqqD, the global homology clustering criterion used to gather homologues resulted in no functionally divergent homologues being included in the tree (Figure S3 of the Supporting Information). Thus, in this case, residues conferring functional specificity relative to functionally divergent homologues could not be identified. However, Discern identifies a set of positions that are clearly involved in function. The finding that conserved residues are largely constrained to an exposed, unstructured region of PqqD (Figure 6 and Table S6 of the Supporting Information) suggests that the previously observed interaction of this protein with PqqE may necessitate significant restructuring of PqqD.[9]

Structure–Sequence Analysis of PqqE. The closest functionally divergent homologue of PqqE is NirJ (Figure S4 of the Supporting Information), sharing 26% sequence identity (using NirJ from *Sulfurovum* sp., UniProt entry A6X6Z2). NirJ catalyzes a key step in the biosynthesis of heme d1. PqqE and NirJ both have been experimentally verified to be members of the radical SAM superfamily,[54,55] catalytically cleaving the universal cosubstrate of this superfamily, *S*-adenosylmethionine, to form a highly oxidizing radical used for hydrogen abstraction on the remaining substrate.[7] Members of the radical SAM superfamily coordinate a [4Fe-4S] cluster using the CX_3CX_2C motif; the three cysteine ligands for this SAM-binding cluster were ranked within the top three evolutionarily conserved residues by INTREPID (Cys22, Cys26, and Cys29 in PqqE from *K. pneumoniae*) (Table S7 of the Supporting Information). The GGE motif (residues 66–68 in PqqE from *K. pneumoniae*, ranked 10th, 7th, and 11th, respectively, by INTREPID) that is characteristic of radical SAM enzymes and serves to stabilize SAM in the active site is found in both enzymes as well (Table S7 of the Supporting Information).[56,57]

In contrast to PqqB and PqqC, little experimental evidence is available for close homologues of PqqE. A low level of sequence identity among homologues is, indeed, a common feature of the radical SAM superfamily, though nearly all of its members adopt a half to full TIM barrel fold. The structure of radical SAM enzymes can be described as two distinct regions, each catalyzing a specific half-reaction.[56] The N-terminal region of superfamily members constitutes a highly conserved radical SAM core in which *S*-adenosylmethionine is catalytically cleaved, and this region spans approximately 200 amino acids (residues 16–174 in PqqE), Figure S5. The second region (located at the C-terminus) is the location of the specific half-reaction catalyzed by each enzyme of the superfamily and, consequently, is highly variable in both length and sequence identity. This variability gives rise to the low level of overall sequence identity observed for the radical SAM superfamily.[56,57]

Homology between the C-terminal regions of PqqE (residues 206–380 from *K. pneumoniae*, UniProt entry B5XX58) and NirJ is supported by a 35% sequence identity (BLAST *E* value of 5×10^{-6}) upon comparison of PqqE to a NirJ from *Methanosarcina acetivorans* (residues 249–346 from UniProt entry Q8TT64);[34] the annotation of NirJ from *M. acetivorans* was verified by locating the ORF for this gene product within the operon encoding heme d1 biosynthesis in this organism. This homology between the C-termini of PqqE and NirJ allows inferences about the

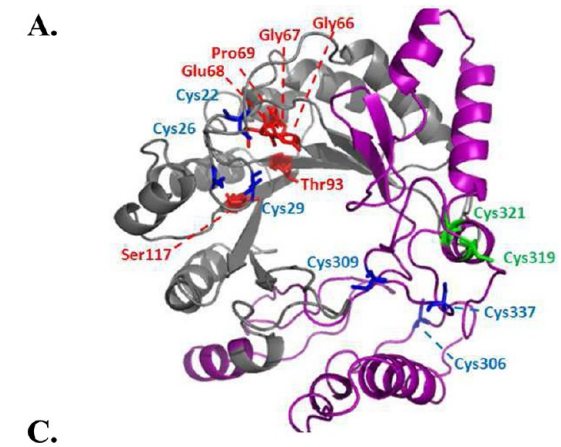
dx.doi.org/10.1021/bi201763d | *Biochemistry* 2012, 51, 2265–2275

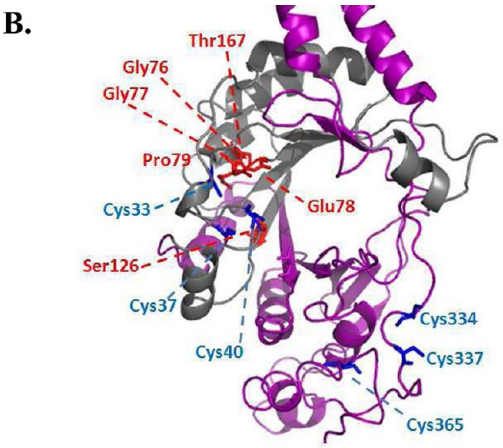


B.

```
                                    4 65                 8 17  3 10
PqqD_B5XX59_KP     1 MQKTS-------IVAFRRGYRLQWEAAQESHVILYPEGMAKLNETAAAILELVDGRRDVAAIIAMLNERFPEAGGV-DDD
PqqCD_Q49150_ME    1 L---------------PRGVRLRFDEVRNKHVLLAPERTFDLDDNAVAVLKLVDGRNTVSQIAQILGQTYDADPAIIEAD
PqqD_Q9L3B1_GO     1 MTEAPHVVAEGTVLSFARGHRLQHDRVRDVWIVQAPEKAFVVEGAAPHILRLLDGKRSVGEIIQQLAIEFSAPREVIAKD
PqqD_Q83X98_SR     1 MTGLP----EPTVPRLRPGVRLTRDPARGELALL-PERVVVLNDTAAAVLAHCDGTTSLAGIVERLAEEYE---GVSAED
PqqD_O33505_RA     1 MITIT----EHYTPMFRRGYRMQFEKTQDCHVILYPEGMAKLNDSATFILQLVDGERTIANIIDELNERFPEAGGV-NDD
                                 2    9
PqqD_B5XX59_KP    73 VVEFLQIACQQKWITCREPE 92
PqqCD_Q49150_ME   66 ILPMLAGLAQKRVLE----R 81
PqqD_Q9L3B1_GO    81 VLALLSELTEKNVLH----T 96
PqqD_Q83X98_SR    73 VRELLLRLAQRRVVDLH--G 90
PqqD_O33505_RA    76 VKDFFAQAHAQKWITFREPA 95
```

Figure 6. Structure–sequence analysis of PqqD. (A) Homology model of PqqD from *K. pneumoniae* (UniProt entry B5XX59) modeled onto the three-dimensional structure of PqqD from *X. campestris* (PDB entry 3G2B). Residues ranked within the top 10 by Discern are colored red (Table S6 of the Supporting Information). (B) MSA of the five PqqD seed sequences constructed using MAFFT and displayed using Belvu. Residues ranked within the top 10 by Discern are colored red. Abbreviations: KP, *K. pneumoniae*; ME, *M. extorquens* AM1; GO, *G. oxidans* 621H; RA, *R. aquatilis*; SR, *S. rochei* (Table S6 of the Supporting Information).





C.

```
                                                                2   1   3                                  8  7
PqqE_B5XX58_KP     1 MSQNKPAVNP-----P------LWLLAELTYRCPLQCPYCSNPLDFARQEKELTTEQWIEVFRQARAMGSVQLGFSGGEP
NirJ_A6Q6Z2_SS     1 MFRLSNLLKSVTAGKPARVLDGSIAIWNFTNRCNLSCLHCYS-KADLDAVDTLTTENIMETLPKLKANGVKFLIFSGGEP
                                          13                    5
PqqE_B5XX58_KP    70 LTRKDLPELIRAARDLGFYTNLITSGIGLTESKLDAFSEAGLDHIQISFQASDEVLNAALAGNKKAFQQKLAMARAVKAR
NirJ_A6Q6Z2_SS    80 LTRKDLFDIAARCKELGIVTYLSTNGLYVKKSNAEKILDT-FDYIGISIDGSPEV-HDAFRGLKGSFVESMKAVDLLNSF

PqqE_B5XX58_KP   150 D-YPMVLNFVLHRHNIDQLDKIIELCIELEADDVELATCQFYGWAFLNREGLLPTREQIARAEQVVADYRQKMAASGNLT
NirJ_A6Q6Z2_SS   158 GKTKVGIRFTITKDTYDDLQFIFELAEEHNIPKVYISHLVYSGRGLENLEMDLSKEQRITAVNYILDKAFEYHESG---R

PqqE_B5XX58_KP   229 NLLFVTPDY-----------------YEERPKGCMGGW-----GSIFLSVTPEGTALPCHSARQLPVAFPSVLEQSLESI
NirJ_A6Q6Z2_SS   235 DIEIVTGNMEMDAILFYDRFEKKYPEYAEEMKRRLIEWGGNSAGRKLLNIDSEGFVKPDPFFP---VKIGNILTQDFSDI
                                                  21  20          25 26                24
PqqE_B5XX58_KP   287 WYD--SFGFNRYRGYDWMPEP------CRSCDEKEKDFGGCRCQAFMLTGSADNADPVCSKSPHHHKILEARREAACSDI
NirJ_A6Q6Z2_SS   312 WTNEPTELLQKLR-----VHPRELGGKCVECQYLNICNGGSRSRAYAIYGDMWAEDPSCYLTEEQIKGN-----------

PqqE_B5XX58_KP   359 KVSQLQFRNRTRSQLIYKTREL 380
NirJ_A6Q6Z2_SS   376 ----------------------N 376
```

Figure 7. Structure–sequence analysis of PqqE. (A) Homology models of PqqE from *K. pneumoniae*, KP (UniProt entry B5XX58). (B) Homology models of NirJ from *Sulfurovum* sp, SS (UniProt entry A6X6Z2) based on the three-dimensional structure of MoaA (PDB entry 1TV8). (C) MSA of PqqE (UniProt entry B5XX58) and NirJ (UniProt entry A6X6Z2). The INTREPID ranking of important functional residues is indicated by numbers above the alignment (Table S7 of the Supporting Information). Colors in panels A–C correspond to the following groups: blue for iron–sulfur cluster ligands, red for residues located within the vicinity of the SAM-binding [4Fe-4S] cluster, green for residues highly conserved in PqqE that are not present in NirJ, pink for regions in which the homology is less reliable and structural information is more speculative.

dx.doi.org/10.1021/bi201763d | *Biochemistry* 2012, 51, 2265–2275

function of key amino acid residues in this region of the protein to be made. Both PqqE and NirJ have been experimentally determined to bind an additional [4Fe-4S] cluster, likely located outside of the N-terminal radical SAM core. Three C-terminal cysteine residues were ranked in the top 25 by INTREPID (Cys306, Cys309, and Cys337 in PqqE from *K. pneumoniae*) for both PqqE and NirJ, and analysis of the MSA of PqqE and NirJ proteins indicates that these residues are the only C-terminal cysteines conserved for both proteins (Figure 7C). These residues form a $CX_2CX_{27}C$ motif; this analysis suggests that this motif serves to coordinate the non-SAM binding [4Fe-4S] cluster.

A comparative (homology) structural model for PqqE was constructed to identify the relative location of the two [4Fe-4S] clusters. The three-dimensional structure most evolutionarily similar to PqqE is of MoaA from *S. aureus* with a BLAST *E* value of 1×10^{-6} and 23% identity and 41% positives (identical and similar amino acids) for PqqE residues 1−206. The PHYRE2 server produced a comparative model for the entire three-dimensional structure of PqqE, extending beyond the conserved N-terminus into the highly variable C-terminus. Figure 7A presents the PqqE structural model, differentiating between the region anticipated to be accurately modeled (the first 206 amino acids, colored gray) and the variable C-terminal region (colored purple). A comparative model is also provided for NirJ in Figure 7B. On the basis of these structural models, the second [4Fe-4S] cluster is located opposite the N-terminal SAM-binding [4Fe-4S] cluster, across the putative hydrophobic active site channel. Also highlighted by INTREPID and the MSA of PqqE proteins were Cys319 and Cys321 that are not present in NirJ. These residues are located proximal to the $CX_2CX_{27}C$ motif at the C-terminus of PqqE and so might be involved in a function specific to the role of PqqE in the biosynthetic pathway. Possible roles for these two residues will depend on whether these residues face the active site or solvent interface. In the former case, they may act to stabilize the peptide substrate in the active site; in the latter case, they may play a role in the protein−protein interactions demonstrated with PqqD.[9]

CONCLUSION

This work presents the use of a novel bioinformatics protocol for identifying species encoding a natural product biosynthetic pathway. The most conserved proteins previously shown as being essential to PQQ production (PqqC−E) were used to identify 126 bacterial species with PQQ biosynthetic capability. Genomes encoding these three proteins were further examined to determine the presence or absence of additional *pqq* genes. These analyses revealed the conservation of *pqqA* and *pqqB*, substantiating a previously unclear essential role for PqqB. The ORF for *pqqF* appears to be frequently lost and is assumed to be nonessential to this pathway, allowing substitution with other proteases. These results further reveal a highly conserved *pqqA−E* gene order within the operon, which, along with recent evidence of a PqqD−PqqE interaction strongly suggests a role for macromolecular complex formation in function. A detailed sequence−structure analysis was employed to distinguish homologues that share a common multidomain architecture from those that have only partial homology and to identify sequence motifs that are diagnostic of each of the separate biosynthetic proteins. From such a phylogenetic−structure−sequence analysis, it is seen that PqqB contains a 2-His, 1-Asp active site configuration characteristic of nonheme iron oxygenases. These findings may aid in the tailoring of antibiotics specifically for the inhibition of PQQ biosynthesis, while improving our understanding of related enzymes present in other pharmaceutically relevant biosynthetic pathways.

ASSOCIATED CONTENT

Supporting Information

Supplemental data mentioned in the text, together with multiple sequence alignments selected from each core biosynthetic protein and its closest functionally distinct homologues (Figures S6−S8). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*Telephone: (510) 643-3668. Fax: (510) 643-6232. E-mail: klinman@berkeley.edu or kimmen@berkeley.edu.

Present Address

[@]Rockefeller University (H. Hang lab), 1230 York Ave., New York, NY 10065.

Author Contributions

F.B., E.M.I., and J.M.R. contributed equally to this work.

Funding

This work was supported by funding from the National Institutes of Health (Grant GM039296 to J.P.K.), the National Science Foundation (Grant 0732065 to K.S.), and the U.S. Department of Energy (Grant DE-SC0004916 to K.S.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Ruchira Datta, Glen Javis, and Shailen Tuli for technical advice on different aspects of this work.

ABBREVIATIONS

PQQ, pyrroloquinoline quinone; MSA, multiple sequence alignment; HMM, hidden Markov model; SAM, *S*-adenosylmethionine; ORF, open reading frame; AHQQ, 3a-(2-amino-2-carboxyethyl)-4,5-dioxo-4,5,6,7,8,9-hexahydroquinoline-7,9-dicarboxylic acid.

REFERENCES

(1) Anthony, C. (2001) Pyrroloquinoline quinone (PQQ) and quinoprotein enzymes. *Antioxid. Redox Signaling 3*, 757−774.

(2) Goodwin, P. M., and Anthony, C. (1998) The biochemistry, physiology and genetics of PQQ and PQQ-containing enzymes. *Adv. Microb. Physiol. 40*, 1−80.

(3) Sode, K., Ito, K., Witarto, A. B., Watanabe, K., Yoshida, H., and Postma, P. (1996) Increased production of recombinant pyrroloquinoline quinone (PQQ) glucose dehydrogenase by metabolically engineered *Escherichia coli* strain capable of PQQ biosynthesis. *J. Biotechnol. 49*, 239−243.

(4) Meulenberg, J. J., Sellink, E., Riegman, N. H., and Postma, P. W. (1992) Nucleotide sequence and structure of the *Klebsiella pneumoniae pqq* operon. *Mol. Gen. Genet. 232*, 284−294.

(5) Velterop, J. S., Sellink, E., Meulenberg, J. J., David, S., Bulder, I., and Postma, P. W. (1995) Synthesis of pyrroloquinoline quinone in vivo and in vitro and detection of an intermediate in the biosynthetic pathway. *J. Bacteriol. 177*, 5088−5098.

(6) Magnusson, O. T., Toyama, H., Saeki, M., Rojas, A., Reed, J. C., Liddington, R. C., Klinman, J. P., and Schwarzenbacher, R. (2004) Quinone biogenesis: Structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone. *Proc. Natl. Acad. Sci. U.S.A. 101*, 7913−7918.

(7) Wecksler, S. R., Stoll, S., Tran, H., Magnusson, O. T., Wu, S. P., King, D., Britt, R. D., and Klinman, J. P. (2009) Pyrroloquinoline quinone biogenesis: Demonstration that PqqE from *Klebsiella pneumoniae* is a radical S-adenosyl-L-methionine enzyme. *Biochemistry 48*, 10151−10161.
(8) Houck, D. R., Hanners, J. L., Unkefer, C. J., van Kleef, M. A., and Duine, J. A. (1989) PQQ: Biosynthetic studies in *Methylobacterium* AM1 and *Hyphomicrobium* X using specific ^{13}C labeling and NMR. *Antonie van Leeuwenhoek 56*, 93−101.
(9) Wecksler, S. R., Stoll, S., Iavarone, A. T., Imsand, E. M., Tran, H., Britt, R. D., and Klinman, J. P. (2010) Interaction of PqqE and PqqD in the pyrroloquinoline quinone (PQQ) biosynthetic pathway links PqqD to the radical SAM superfamily. *Chem. Commun. 46*, 7031−7033.
(10) Mazodier, P., Biville, F., Turlin, E., and Gasser, F. (1988) Localization of a pyrroloquinoline quinone biosynthesis gene near the methanol dehydrogenase structural gene in *Methylobacterium organophilum* DSM 760. *J. Gen. Microbiol. 134*, 2513−2524.
(11) McClerren, A. L., Cooper, L. E., Quan, C., Thomas, P. M., Kelleher, N. L., and van der Donk, W. A. (2006) Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proc. Natl. Acad. Sci. U.S.A. 103*, 17243−17248.
(12) Milne, J. C., Eliot, A. C., Kelleher, N. L., and Walsh, C. T. (1998) ATP/GTP hydrolysis is required for oxazole and thiazole biosynthesis in the peptide antibiotic microcin B17. *Biochemistry 37*, 13250−13261.
(13) Velasquez, J. E., and van der Donk, W. A. (2011) Genome mining for ribosomally synthesized natural products. *Curr. Opin. Chem. Biol. 15*, 11−21.
(14) Makris, T. M., Chakrabarti, M., Muenck, E., and Lipscomb, J. D. (2010) A family of diiron monooxygenases catalyzing amino acid β-hydroxylation in antibiotic biosynthesis. *Proc. Natl. Acad. Sci. U.S.A. 107*, 15391−15396.
(15) Fetzner, S., and Steiner, R. A. (2010) Cofactor-independent oxidases and oxygenases. *Appl. Microbiol. Biotechnol. 86*, 791−804.
(16) Sudek, S., Haygood, M. G., Youssef, D. T., and Schmidt, E. W. (2006) Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium *Trichodesmium erythraeum*, predicted from the genome sequence. *Appl. Environ. Microbiol. 72*, 4382−4387.
(17) Haft, D. H. (2011) Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. *BMC Genomics 12*, 21.
(18) Sjolander, K. (2010) Getting started in structural phylogenomics. *PLoS Comput. Biol. 6*, e1000621.
(19) Arakawa, K., Sugino, F., Kodama, K., Ishii, T., and Kinashi, H. (2005) Cyclization mechanism for the synthesis of macrocyclic antibiotic lankacidin in *Streptomyces rochei*. *Chem. Biol. 12*, 249−256.
(20) Velterop, J. S., Sellink, E., Meulenberg, J. J., David, S., Bulder, I., and Postma, P. W. (1995) Synthesis of pyrroloquinoline quinone in vivo and in vitro and detection of an intermediate in the biosynthetic pathway. *J. Bacteriol. 177*, 5088−5098.
(21) Toyama, H., Chistoserdova, L., and Lidstrom, M. E. (1997) Sequence analysis of pqq genes required for biosynthesis of pyrroloquinoline quinone in *Methylobacterium extorquens* AM1 and the purification of a biosynthetic intermediate. *Microbiology (Reading, U.K.) 143*, 595−602.
(22) Hoelscher, T., and Goerisch, H. (2006) Knockout and overexpression of pyrroloquinoline quinone biosynthetic genes in *Gluconobacter oxydans* 621H. *J. Bacteriol. 188*, 7668−7676.
(23) Guo, Y. B., Li, J., Li, L., Chen, F., Wu, W., Wang, J., and Wang, H. (2009) Mutations that disrupt either the pqq or the gdh gene of *Rahnella aquatilis* abolish the production of an antibacterial substance and result in reduced biological control of grapevine crown gall. *Appl. Environ. Microbiol. 75*, 6792−6803.
(24) Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) The Pfam protein families database. *Nucleic Acids Res. 32*, D138−D141.
(25) Krishnamurthy, N., Brown, D., and Sjolander, K. (2007) FlowerPower: Clustering proteins into domain architecture classes for phylogenomic inference of protein function. *BMC Evol. Biol. 7* (Suppl. 1), S12.
(26) Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res. 30*, 3059−3066.
(27) Stamatakis, A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics 22*, 2688−2690.
(28) Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J., and Kelley, L. A. (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins 70*, 611−625.
(29) Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006) Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics*, Vol. 5, Unit 56, Wiley, New York.
(30) Tsai, T. Y., Yang, C. Y., Shih, H. L., Wang, A. H., and Chou, S. H. (2009) *Xanthomonas campestris* PqqD in the pyrroloquinoline quinone biosynthesis operon adopts a novel saddle-like fold that possibly serves as a PQQ carrier. *Proteins 76*, 1042−1048.
(31) Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: A case study using the Phyre server. *Nat. Protoc. 4*, 363−371.
(32) Sankararaman, S., and Sjolander, K. (2008) INTREPID: Information-theoretic TREe traversal for Protein functional site identification. *Bioinformatics 24*, 2445−2452.
(33) Sankararaman, S., Sha, F., Kirsch, J. F., Jordan, M. I., and Sjolander, K. (2010) Active site prediction using evolutionary and structural information. *Bioinformatics 26*, 617−624.
(34) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, Q., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389−3402.
(35) Duine, J. A. (1991) Quinoproteins: Enzymes containing the quinonoid cofactor pyrroloquinoline quinone, topaquinone, or tryptophan-tryptophan quinone. *Eur. J. Biochem. 200*, 271−284.
(36) Virdi, N. S., Price, D., Ellison, J., Yadalam, S., and Nigam, S. (2010) Use of GDH-PQQ glucose meter systems in patients receiving maltose-containing therapies. *Diabetologia*, Vol. 53, Springer, Berlin.
(37) Mushegian, A. R., and Koonin, E. V. (1996) Gene order is not conserved in bacterial evolution. *Trends Genet. 12*, 289−290.
(38) Omelchendo, M. V., Makarova, K. S., Wolf, Y. I., Rogozin, I. B., and Koonin, E. V. (2003) Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. *Fenome Biol. 4*, R55.
(39) Dendekar, T., Snel, B., Huynen, M., and Bork, P. (1998) Conservation of gene order: A fingerprint of proteins that physically interact. *Trends Biochem. Sci. 23*, 432−328.
(40) Fondi, M., Emiliani, G., and Fani, R. (2009) Origin and evolution of operons and metabolic pathways. *Res. Microbiol. 160*, 502−512.
(41) Puehringer, S., Metlitzky, M., and Schwarzenbacher, R. (2008) The pyrroloquinoline quinone biosynthesis pathway revisited: A structural approach. *BMC Biochem. 9*, 8.
(42) Podzelinska, K., He, S.-M., Wathioer, M., Yakunin, A., Proudfoot, M., Hove-Jenson, B., Zechel, D. L., and Jia, Z. (2009) Structure of PhnP, a phosphodiesterase of the carbon-phosphorous lyase pathway for phosphonate degradation. *J. Biol. Chem. 284*, 17216−17226.
(43) Vetting, M. W., Wackett, L. P., Que, L. Jr., Lipscomb, J. D., and Ohlendorf, D. H. (2004) Crystaollographic comparison of manganese- and iron-dependent homoprotocatechuate 2,3-dioxygnases. *J. Bacteriol. 186*, 1945−1958.
(44) Bruijnincx, P. C. A., van Koten, G., and Gebbink, R. J. M. K. (2008) Mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad: recent developments in enzymology and modeling studies. *Chem. Soc. Rev. 37*, 2716−2744.
(45) Koehntop, K. D., Emerson, J. P., and Que, L. Jr. (2005) The 2-His-1-carboxylate facial triad: A versatile platform for dioxyygen activation by mononuclear non-heme iron(II) enzymes. *J. Biol. Inorg. Chem. 10*, 87−93.

(46) Makris, T. M., Chakrabarti, M., Münck, E., and Lipscomb, J. D. (2010) A family of diiron monooxygenases catalyzing amino acid beta-hydroxylation in antibiotic biosynthesis. *Proc. Natl. Acad. Sci. U.S.A. 107*, 15391−15396.
(47) Bellamacina, C. R. (1996) Protein motifs. 9. The nicotinamide dinucleotide binding motif: A comparison of nucleotide binding proteins. *FASEB J. 10*, 1257−1269.
(48) Dym, O., and Eisenber, D. (2001) Sequence-structure analysis of FAD-containing proteins. *Protein Sci. 10*, 1712−1728.
(49) Magnusson, O. T., Toyama, H., Saeki, M., Schwarzenbacher, R., and Klinman, J. P. (2004) The structure of a biosynthetic intermediate of pyrroloquinoline quinone (PQQ) and elucidation of the final step of PQQ biosynthesis. *J. Am. Chem. Soc. 126*, 5342−5343.
(50) Toms, A. V., Haas, A. L., Park, J. H., Begley, T. P., and Ealick, S. E. (2005) Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II. *Biochemistry 44*, 2319−2329.
(51) Magnusson, O. T., RoseFigura, J. M., Toyama, H., Schwarzenbacher, R., and Klinman, J. P. (2007) Pyrroloquinoline quinone biogenesis: Characterization of PqqC and its H84N and H84A active site variants. *Biochemistry 46*, 7174−7186.
(52) RoseFigura, J. M., Puehringer, S., Scharzenbacher, R., Toyama, H., and Klinman, J. P. (2011) Characterization of a protein-generated O_2 binding pocket in PqqC, a cofactorless oxidase catalyzing the final step in PQQ production. *Biochemistry 50*, 1556−1566.
(53) Gibrat, J. F., Madej, T., and Bryant, S. H. (1996) Surprising similarities in structure comparison. *Struct. Biol. 6*, 377−385.
(54) Wecksler, S. R., Stoll, S., Tran, H., Magnusson, O. T., Wu, S.-P., King, D., Britt, R. D., and Klinman, J. P. (2009) Pyrroloquinoline quinone biogenesis: Demonstration that PqqE from *Klebsiella pneumoniae* is a radical S-adenosyl-L-methionine enzyme. *Biochemistry 48*, 10151−10161.
(55) Brindley, A. A., Zajicek, R., Warren, M. J., Ferguson, S. J., and Rigsby, S. E. J. (2010) NirJ, a radical SAM family member of the d_1 heme biogenesis cluster. *FEBS Lett. 584*, 2461−2466.
(56) Vey, J. L., and Drennan, C. L. (2011) Structural insights into radical generation by the radical SAM superfamily. *Chem. Rev. 111*, 2487−2506.
(57) Nicolet, Y., and Drennan, C. L. (2004) AdoMet radical proteins—from structure to evolution—alignment of divergent protein sequences reveals strong secondary structure element conservation. *Nucleic Acids Res. 13*, 4015−4025.