

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221823210>

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

ARTICLE *in* BIOCHEMISTRY · FEBRUARY 2012

Impact Factor: 3.02 · DOI: 10.1021/bi201763d · Source: PubMed

CITATIONS

17

READS

52

6 AUTHORS, INCLUDING:



Kimmen Sjölander

University of California, Berkeley

58 PUBLICATIONS 4,649 CITATIONS

[SEE PROFILE](#)



Judith P Klinman

University of California, Berkeley

279 PUBLICATIONS 13,138 CITATIONS

[SEE PROFILE](#)



NIH Public Access

Author Manuscript

Biochemistry. Author manuscript; available in PMC 2013 March 20.

Published in final edited form as:
Biochemistry. 2012 March 20; 51(11): 2265–2275. doi:10.1021/bi201763d.

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, University of California, Berkeley, CA, 94720, USA.

^{||}Department of Molecular and Cell Biology, University of California, Berkeley, CA, 94720, USA.

[§]Department of Bioengineering, University of California, Berkeley, CA, 94720, USA.

[¶]Department of Plant and Microbial Biology, University of California, Berkeley, CA, 94720, USA.

[△]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, USA.

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 concluded to be 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 possible new targets for antibiotic design and application.

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

*To whom correspondence should be addressed: Tel: 510-643-3668; Fax: 510-643-6232; klinman@berkeley.edu, kimmen@berkeley.edu.

¶Present address: Rockefeller University (H. Hang Lab), 1230 York Ave., New York, NY 10065

□These authors contributed equally to this work.

SUPPORTING INFORMATION AVAILABLE

Additionally included in supplemental information but not referenced in the text of this paper are the multiple sequence alignments selected from each core biosynthetic protein and its closest functionally distinct homologs (Figures S6, S7, and S8). This material is available free of charge via the internet at <http://pubs.acs.org>.

CONFLICT OF INTEREST DISCLOSURE

The authors declare no competing financial interest.

Keywords

pyrroloquinoline; quinone; pathogenicity; phylogenomic analysis; metallo-beta-lactamase; radical SAM domain; cofactorless oxidase

INTRODUCTION

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). 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 (Scheme 1) (8); however, the ability of or conditions required for PqqE to functionalize PqqA have not yet been demonstrated. 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 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, non-specific proteases can assume its role during cofactor biogenesis (5). One of the most enigmatic gene products is PqqB with high 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 suggests 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 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 homologs (18). These studies serve as a guide to predict putative roles for the open reading frames within the *pqq* operon and to probe the contribution of conserved amino acid side chains within the gene products with demonstrated function.

METHODS

Dataset

Five bacteria with demonstrated PQQ biosynthetic capacity were selected for this study: *Klebsiella pneumoniae*, *Methylobacterium extorquens AM1*, *Gluconobacter oxydans* 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 starting points 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 homologs, identify corresponding PFAM domains (24), and to differentiate homologs 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) homolog identification, (ii) multiple sequence alignment, (iii) phylogenetic tree construction and analysis, (iv) protein structure prediction, (v) functional site identification (see Scheme 2).

Homologs 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 greater than 70% gap characters. Maximum likelihood trees were estimated from the masked MSAs using RAxML (27).

A solved 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 solved three-dimensional structure 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 homolog with a solved structure (MoaA from *Staphylococcus aureus*, PDB ID 1TV8) was used to build a comparative model. The Phyre and Phyre2 servers (31) were used to construct comparative models for PqqB and PqqE; Modeller software (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 homologs 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 solved 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 important functionally.

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), PqqC (Figure S2), PqqD (Figure S3) and PqqE (Figure S4). 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 homologs were plotted on structures or models for these proteins.

Additional Details on Bioinformatics Methods

HMM clustering of homologs using FlowerPower global homology clustering: all parameters default except the number of subfamily HMM scoring iterations (set to 10). MAFFT MSA construction parameters: 5 maximum iterations and default parameters. RAxML parameters: JTT+ Γ model and 20 discrete γ -rate categories. The statistical support of branches was estimated using 100 bootstrap replicates. Due to 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 the whole genomes for these cases (34).

RESULTS AND DISCUSSION

Species with PQQ Biosynthetic Capability

Based on 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 that include PqqC–E, of which 125 also contain PqqB (Table S2), strongly implicating PqqB as essential to PQQ biosynthesis. Due to the short length of PqqA (less 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 the whole genomes for those species for which PqqB–E proteins had been detected. This identified an additional 37 *pqqA* genes which had been missed by the genome annotation pipelines for those species (Table S3). To summarize, of the 126 species for which our analyses support PQQ biosynthetic capability, a total of 106 contain an identifiable PqqA (Table S2); of these 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 homologs outside prokaryotes, the pathway is clearly specific to prokaryotes. The ACS Paragon Plus Environment 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). 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 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 *Chlamydiae*, were also found to contain the homologs 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 due to the availability of contigs including all *pqq* genes. Of these 96, all of which contain *pqqA, B, C, D, and 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 conserved order of the core genes with no large insertions.

Only in 17 species is *pqqF* clustered with *pqqB–E*, either in the order of *pqqB-C-D-E-F* or in the order of *pqqF-A-B-C-D-E*. In all other cases (27 species), *pqqF* is located remotely from the *pqqA–E* cluster, separated by genes unrelated to the pathway. The distance of *pqqF* from 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 OJ consecutively acting enzymes in proximity to 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 *M. chloromethanicum*. In addition, we have found a potential fusion of *pqqD* and *pqqE* in *Methylosinus trichosporium*. The sequence is annotated as PqqE (UniProt Accession: 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 to infer the evolution of the PQQ biosynthetic pathway (Figure 3). As suggested by the gene locus study, PqqB, PqqC, PqqD, and PqqE 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 for both gene identification and for reliable identification of short peptides (see above) PqqA was identified in most species (Figure 3).

Structure-Sequence Analyses of Core *pqq* Genes

PqqB, PqqC, PqqD, and PqqE 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). Evolutionary and structural analyses were performed to identify residues that are likely to be important functionally.

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 IDs 1XTO and 3JXP) has since confirmed this. Analysis of the phylogenetic tree of PqqB homologs shows that the closest functionally distinct homolog of PqqB is PhnP (Figure S1). PhnP is a phosphodiesterase that shares 26% sequence identity with 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*, in red in Figure 4A) are ranked 6, 1, and 3, respectively by the Discern algorithm (Table S4). The active site metals bound by PhnP are both manganese (43); however the metal ligands retained in PqqB represent a 2-His/1-carboxylate facial triad configuration, characteristic of the non-heme 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 $\text{Fe}^{\text{IV}}=\text{O}$ (44). This chemistry is preceded amongst members of the metallo- β -lactamase superfamily (45). Additionally, in the three-dimensional structure of PqqB (PDB ID 3JXP from *P. putida*), a zinc ion is bound several angstroms from this putative active site near the solvent interface by the motif CXXCX₂C (residues 19–24 in PqqB from *K. pneumoniae* in 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 it is proposed that the ion plays 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 (GXXXGGGPQWN residues 7–18 in PqqB from *K. pneumoniae*) located between the proposed structural zinc ion and the putative active site of PqqB and exposed to solvent, is conserved in 80% of PqqB sequences from the 126 proposed PQQ producers, hi 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 (46, 47), 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, 48). 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) involving transfer of 8 electrons and protons to molecular oxygen to form hydrogen peroxide/water. The closest functionally distinct homolog of PqqC identified in the phylogenetic tree is TenA (Figure S5). 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 (49). 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). 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 (in blue in Figure 5), with the remaining sixteen 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, in 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 in 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 fourth by Discern (50); His154, Tyr175 and Arg179 are proposed to form a core oxygen-binding pocket essential for oxygen activation (51) and are ranked tenth, seventh and twenty-sixth, respectively, by Discern (labeled with stars 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 protein from *K. pneumoniae*, and a homology model for *K. pneumoniae* (using PDB ID 3G2B from *Xanthomonas campestris* as a template with 29% sequence identity). In the case of PqqD, the global homology clustering criterion used to gather homologs resulted in no functionally divergent homologs being included in the tree (Figure S3). Thus, in this case, residues conferring functional *specificity* relative to functionally divergent homologs were unable to 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) suggest 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 homolog of PqqE is NirJ (Figure S4), sharing 26% sequence identity (using NirJ from *Sulfurovum Sp.*, UniProt accession A6X6Z2). NirJ catalyzes a key step in the biosynthesis of heme dI. PqqE and NirJ both have been experimentally verified to be members of the radical SAM superfamily (52, 53), catalytically cleaving the universal co-substrate 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 motif CX₃CX₂C; 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). The GGE motif (residues 66–68 in PqqE from *K.*

pneumoniae, ranked 10, 7, and 11 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) (54, 55).

In contrast to PqqB and PqqC, little experimental evidence is available for close homologs of PqqE. Low sequence identity amongst homologs 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 (53). The N-terminal region of superfamily members constitutes a highly conserved radical SAM core where *S*-adenosylmethionine is catalytically cleaved, and this region spans approximately 200 amino acids (residues 16 – 174 in PqqE). 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 overall sequence identity observed for the radical SAM superfamily (53, 55).

Homology between the C-terminal regions of PqqE (residues 206–380 from *K. pneumoniae*, UniProt accession B5XX58) and NirJ is supported by a 35% sequence identity (BLAST E-value of 5×10^{-6}) when comparing PqqE to a NirJ from *Methanosaerina acetivorans* (residues 249 – 346 from UniProt accession Q8TT64) (34); the annotation of NirJ from *M. acetivorans* was verified by locating the ORF for this gene product within the operon encoding heme dl biosynthesis in this organism. This homology between the C-termini of PqqE and NirJ allows inferences about the function of key amino acid residues in this region of the protein. 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 CX₂CX₂₇C 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 in order to identify the relative location of the two 4Fe-4S clusters. The evolutionarily closest three-dimensional structure to PqqE is of MoaA from *Staphylococcus aureus* with a BLAST E-value 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) (shown in grey) and the variable C-terminal region (shown in purple). A comparative model is also provided for NirJ in Figure 7B. Based on 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 CX₂CX₂₇C 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 (55).

CONCLUSIONS

This work presents use a novel bioinformatics protocol toward identifying species encoding a natural products biosynthesis pathway. The most conserved proteins previously shown as essential to PQQ production -- PqqC, D, and E -- were used to identify 126 bacterial species with PQQ biosynthetic capability. Genomes encoding these three proteins were further examined to determine the presence/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 non-essential to this pathway, allowing substitution by other proteases. The results further reveal a highly conserved gene order of *pqqA–E* within the operon, which along with recent evidence for a PqqD/E interaction, strongly suggests a role for macromolecular complex formation in function. A detailed sequence/structure analysis was employed to distinguish homologs that share a common multi-domain 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 non-heme iron monooxygenases. The presented findings may aid in the tailoring of antibiotics specifically toward the inhibition of PQQ biosynthesis, while furthering our understanding of related enzymes present in other pharmaceutically relevant biosynthetic pathways.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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

ABBREVIATIONS

PQQ	pyrroloquinoline quinone
MSA	Multiple Sequence Alignment
HMM	Hidden Markov Model
SAM	S-Adenosyl Methionine
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. Pyrroloquinoline quinone (PQQ) and quinoprotein enzymes. *Antioxid. Redox. Sign.* 2001; 3:757–774.
2. Goodwin PM, Anthony C. The biochemistry, physiology and genetics of PQQ and PQQ-containing enzymes. *Adv. Microb. Physiol.* 1998; 40:1–80. [PubMed: 9889976]
3. Sode K, Ito K, Witarto AB, Watanabe K, Yoshida H, Postma P. Increased production of recombinant pyrroloquinoline quinone (PQQ) glucose dehydrogenase by metabolically engineered *Escherichia coli* strain capable of PQQ biosynthesis. *J. Biotech.* 1996; 49:239–243.
4. Meulenbergh JJ, Sellink E, Riegman NH, Postma PW. Nucleotide sequence and structure of the *Klebsiella pneumoniae* *pqq* operon. *Mol. Gen. Genet.* 1992; 232:284–294. [PubMed: 1313537]

5. Velterop JS, Sellink E, Meulenberg JJ, David S, Bulder I, Postma PW. Synthesis of pyrroloquinoline quinone in vivo and in vitro and detection of an intermediate in the biosynthetic pathway. *J. Bacteriol.* 1995; 177:5088–5098. [PubMed: 7665488]
6. Magnusson OT, Toyama H, Saeki M, Rojas A, Reed JC, Liddington RC, Klinman JP, Schwarzenbacher R. Quinone biogenesis: Structure and mechanism of PqqC, the final catalyst in the production of pyrroloquinoline quinone. *Proc. Natl. Acad. Sci. U.S.A.* 2004; 101:7913–7918. [PubMed: 15148379]
7. Wecksler SR, Stoll S, Tran H, Magnusson OT, Wu SP, King D, Britt RD, Klinman JP. Pyrroloquinoline quinone biogenesis: demonstration that PqqE from Klebsiella pneumoniae is a radical S-adenosyl-L-methionine enzyme. *Biochemistry.* 2009; 48:10151–10161. [PubMed: 19746930]
8. Houck DR, Hanners JL, Unkefer CJ, van Kleef MA, Duine JA. PQQ: biosynthetic studies in *Methylobacterium AM1* and *Hyphomicrobium X* using specific ¹³C labeling and NMR. *Antonie van Leeuwenhoek.* 1989; 56:93–101. [PubMed: 2549867]
9. Wecksler SR, Stoll S, Iavarone AT, Imsand EM, Tran H, Britt RD, Klinman JP. Interaction of PqqE and PqqD in the pyrroloquinoline quinone (PQQ) biosynthetic pathway links PqqD to the radical SAM superfamily. *Chem. Commun.* 2010; 46:7031–7033.
10. Mazodier P, Biville F, Turlin E, Gasser F. Localization of a pyrroloquinoline quinone biosynthesis gene near the methanol dehydrogenase structural gene in *Methylobacterium organophilum* DSM 760. *J. Gen. Microbiol.* 1988; 134:2513–2524. [PubMed: 2855527]
11. McClellen AL, Cooper LE, Quan C, Thomas PM, Kelleher NL, van der Donk WA. Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proc. Natl. Acad. Sci. U.S.A.* 2006; 103:17243–17248. [PubMed: 17085596]
12. Milne JC, Eliot AC, Kelleher NL, Walsh CT. ATP/GTP hydrolysis is required for oxazole and thiazole biosynthesis in the peptide antibiotic microcin B17. *Biochemistry.* 1998; 37:13250–13261. [PubMed: 9748332]
13. Velasquez JE, van der Donk WA. Genome mining for ribosomally synthesized natural products. *Curr. Opin. Chem. Biol.* 2011; 15:11–21. [PubMed: 21095156]
14. Makris TM, Chakrabarti M, Muenck E, Lipscomb JD. A family of diiron monooxygenases catalyzing amino acid beta-hydroxylation in antibiotic biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 2010; 107:15391–15396. [PubMed: 20713732]
15. Fetzner S, Steiner RA. Cofactor-independent oxidases and oxygenases. *Appl. Microbiol. Biot.* 2010; 86:791–804.
16. Sudek S, Haygood MG, Youssef DT, Schmidt EW. Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium *Trichodesmium erythraeum*, predicted from the genome sequence. *Appl. Environ. Microb.* 2006; 72:4382–4387.
17. Haft DH. Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. *BMC Genomics.* 2011; 12
18. Sjolander K. Getting started in structural phylogenomics. *PLoS Comput. Biol.* 2010; 6:e1000621.
19. Arakawa K, Sugino F, Kodama K, Ishii T, Kinashi H. Cyclization mechanism for the synthesis of macrocyclic antibiotic lankacidin in *Streptomyces rochei*. *Chem. Biol.* 2005; 12:249–256. [PubMed: 15734652]
20. Velterop JS, Sellink E, Meulenberg JJ, David S, Bulder I, Postma PW. Synthesis of pyrroloquinoline quinone in vivo and in vitro and detection of an intermediate in the biosynthetic pathway. *J. Bacteriol.* 1995; 177:5088–5098. [PubMed: 7665488]
21. Toyama H, Chistoserdova L, Lidstrom ME. Sequence analysis of pqq genes required for biosynthesis of pyrroloquinoline quinone in *Methylobacterium extorquens* AM1 and the purification of a biosynthetic intermediate. *Microbiology-UK.* 1997; 143:595–602.
22. Hoelscher T, Goerisch H. Knockout and overexpression of pyrroloquinoline quinone biosynthetic genes in Gluconobacter oxydans 621H. *J. Bacteriol.* 2006; 188:7668–7676. [PubMed: 16936032]
23. Guo YB, Li J, Li L, Chen F, Wu W, Wang J, Wang H. 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. Microb.* 2009; 75:6792–6803.

24. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res. 2004; 32:D138–D141. [PubMed: 14681378]
25. Krishnamurthy N, Brown D, Sjolander K. FlowerPower: clustering proteins into domain architecture classes for phylogenomic inference of protein function. BMC Evol. Biol. 2007; 7(Suppl 1):S12. [PubMed: 17288570]
26. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002; 30:3059–3066. [PubMed: 12136088]
27. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22:2688–2690. [PubMed: 16928733]
28. Bennett-Lovsey RM, Herbert AD, Sternberg MJ, Kelley LA. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. Proteins. 2008; 70:611–625. [PubMed: 17876813]
29. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using Modeller. Curr. Prot. Bioinformatics. 2006; 5(Unit 56)
30. Tsai TY, Yang CY, Shih HL, Wang AH, Chou SH. *Xanthomonas campestris* PqqD in the pyrroloquinoline quinone biosynthesis operon adopts a novel saddle-like fold that possibly serves as a PQQ carrier. Proteins. 2009; 76:1042–1048. [PubMed: 19475705]
31. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 2009; 4:363–371. [PubMed: 19247286]
32. Sankararaman S, Sjolander K. INTREPID-Information-theoretic TREe traversal for Protein functional site identification. Bioinformatics. 2008; 24:2445–2452. [PubMed: 18776193]
33. Sankararaman S, Sha F, Kirsch JF, Jordan MI, Sjolander K. Active site prediction using evolutionary and structural information. Bioinformatics. 2010; 26:617–624. PMCID: PMC2828116. [PubMed: 20080507]
34. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller Q, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acid Res. 1997; 25:3389–3402. [PubMed: 9254694]
35. Duine JA. Quinoproteins: enzymes containing the quinonoid cofactor pyrroloquinoline quinone, topaquinone, or tryptophan-tryptophan quinone. Eur. J. Biochem./FEBS. 1991; 200:271–284.
36. Virdi NS, Price D, Ellison J, Yadalam S, Nigam S. Use of GDH-PQQ glucose meter systems in patients receiving maltose-containing therapies. Diabetologia. 2010; 53
37. Mushegian AR, Koonin EV. Gene order is not conserved in bacterial evolution. Trends Genet. 1996; 12:289–290. [PubMed: 8783936]
38. Omelchendo MV, Makarova KS, Wolf YI, Rogozin IB, Koonin EV. Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. Genome. Biol. 2003; 4:R55.
39. Dendekar T, Snel B, Huynen M, Bork P. Conservation of gene order: a fingerprint of proteins that physically interact. Trends Biochem. Sci. 1998; 23:432–328.
40. Fondi M, Emiliani G, Fani R. Origin and evolution of operons and metabolic pathways. Res. Microbiol. 2009; 160:502–512. [PubMed: 19465116]
41. Puehringer S, Metlitzky M, Schwarzenbacher R. The pyrroloquinoline quinone biosynthesis pathway revisited: as structural approach. BMC Biochemistry. 2008; 9:8. [PubMed: 18371220]
42. Podzelinska K, He S-M, Wathioer M, Yakunin A, Proudfoot M, Hove-Jenson B, Zechel DL, Jia Z. Structure of PhnP, a phosphodiesterase of the carbon-phosphorous lyase pathway for phosphonate degradation. J. Biol. Chem. 2009; 284:17216–17226. [PubMed: 19366688]
43. Vetting MW, Wackett LP, Que L Jr, Lipscomb JD, Ohlendorf DH. Crystallographic comparison of manganese- and iron-dependent homoprotocatechuate 2,3-dioxygenases. J. Bacteriology. 2004; 186:1945–1958.
44. Bruijnincx PCA, van Koten G, Gebbink RJMK. Mononuclear non-heme iron enzymes with the 2-His-l-carboxylate facial triad: recent developments in enzymology and modeling studies. Chem. Soc. Rev. 2008; 37:2716–2744. [PubMed: 19020684]

- NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript
45. Koehntop KD, Emerson JP, Que L Jr. The 2-His-1-carboxylate facial triad: a versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes. *J. Biol. Inorg. Chem.* 2005; 10:87–93. [PubMed: 15739104]
 46. Bellamacina CR. Protein motifs. 9. The nicotinamide dinucleotide binding motif: A comparison of nucleotide binding proteins. *FASEB J.* 1996; 10:1257–1269. [PubMed: 8836039]
 47. Dym O, Eisenber D. Sequence-structure analysis of FAD-containing proteins. *Protein Sce.* 2001; 10:1712–1728.
 48. Magnusson OT, Toyama H, Saeki M, Schwarzenbacher R, Klinman JP. The structure of a biosynthetic intermediate of pyrroloquinoline quinone (PQQ) and elucidation of the final step of PQQ biosynthesis. *J. Am. Chem. Soc.* 2004; 126:5342–5343. [PubMed: 15113189]
 49. Toms AV, Haas AL, Park JH, Begley TP, Ealick SE. Structural characterization of the regulatory proteins TenA and TenI from *Bacillus subtilis* and identification of TenA as a thiaminase II. *Biochemistry.* 2005; 44:2319–2329. [PubMed: 15709744]
 50. Magnusson OT, RoseFigura JM, Toyama H, Schwarzenbacher R, Klinman JP. Pyrroloquinoline quinone biogenesis: Characterization of PqqC and its H84N and H84A active site variants. *Biochemistry.* 2007; 46:7174–7186. [PubMed: 17523676]
 51. RoseFigura JM, Puehringer S, Scharzenbacher R, Toyama H, Klinman JP. Characterization of a protein-generated O(2) binding pocket in PqqC, a cofactorless oxidase catalyzing the final step in PQQ production. *Biochemistry.* 2011; 50:1556–1566. [PubMed: 21155540]
 52. Wecksler SR, Stoll S, Tran H, Magnusson OT, Wu S-P, King D, Britt RD, Klinman JP. Pyrroloquinoline quinone biogenesis: Demonstration that PqqE from *Klebsiella pneumoniae* is a radical S-adenosyl-L-methionine enzyme. *Biochemistry.* 2009; 48:10151–10161. [PubMed: 19746930]
 53. Brindley AA, Zajicek R, Warren MJ, Ferguson SJ, Rigsby SEJ. NirJ, a radical SAM family member of the d(1) heme biogenesis cluster. *FEBS Lett.* 2010; 584:2461–2466. [PubMed: 20420837]
 54. Vey JL, Drennan CL. Structural insights into radical generation by the radical SAM superfamily. *Chem. Rev.* 2011; 111:2487–2506. [PubMed: 21370834]
 55. Nicolet Y, Drennan CL. AdoMet radical proteins - from structure to evolution - alignment of divergent protein sequences reveals strong secondary structure element conservation. *Nucleic Acids Res.* 2004; 13:4015–4025. [PubMed: 15289575]

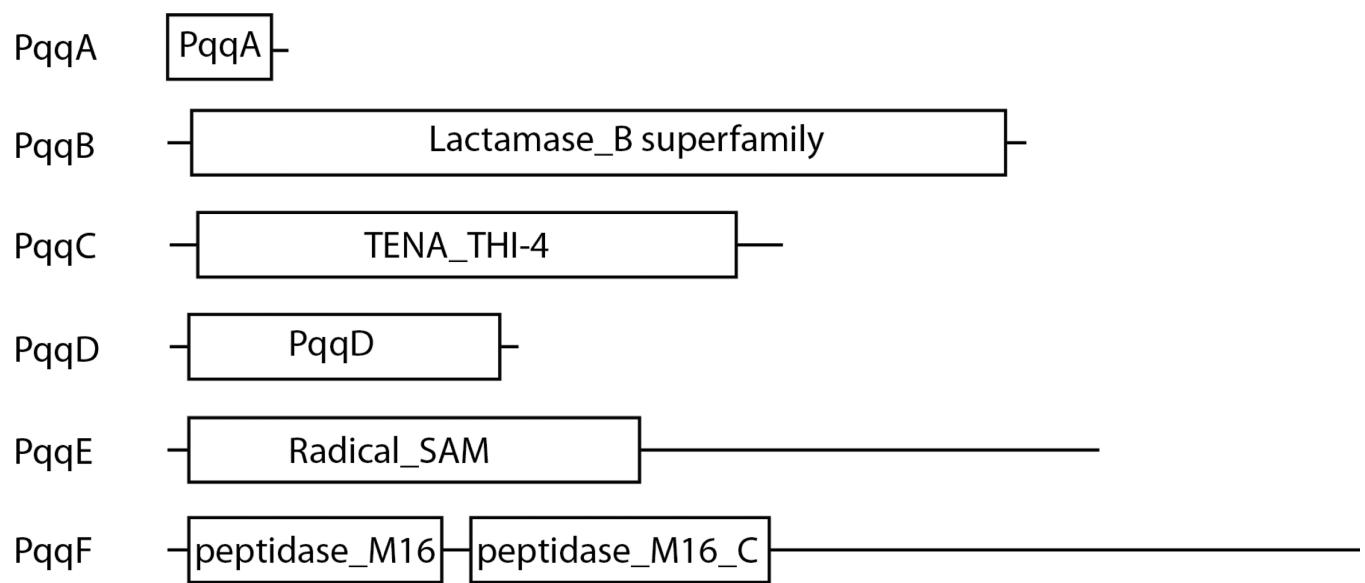
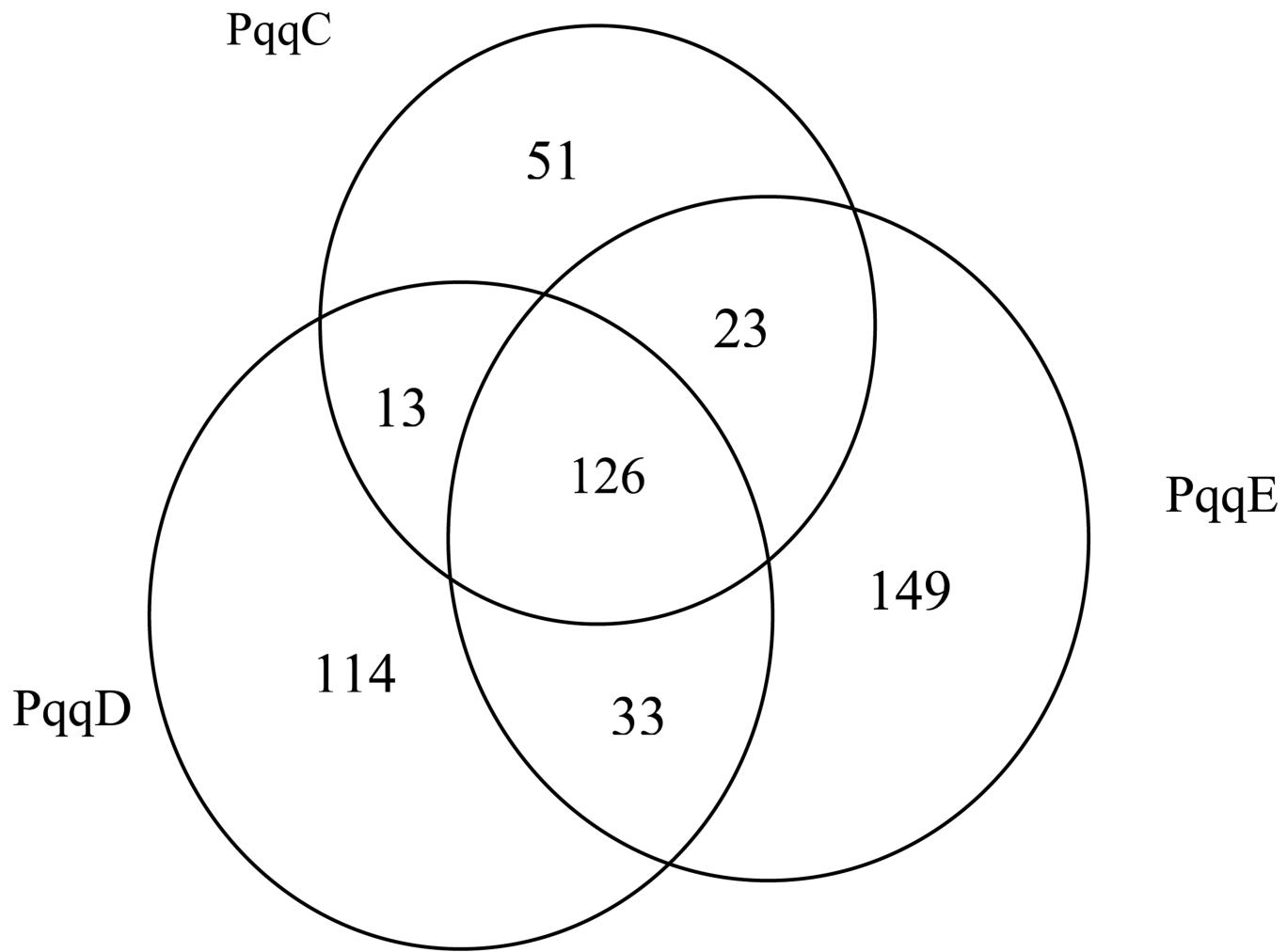
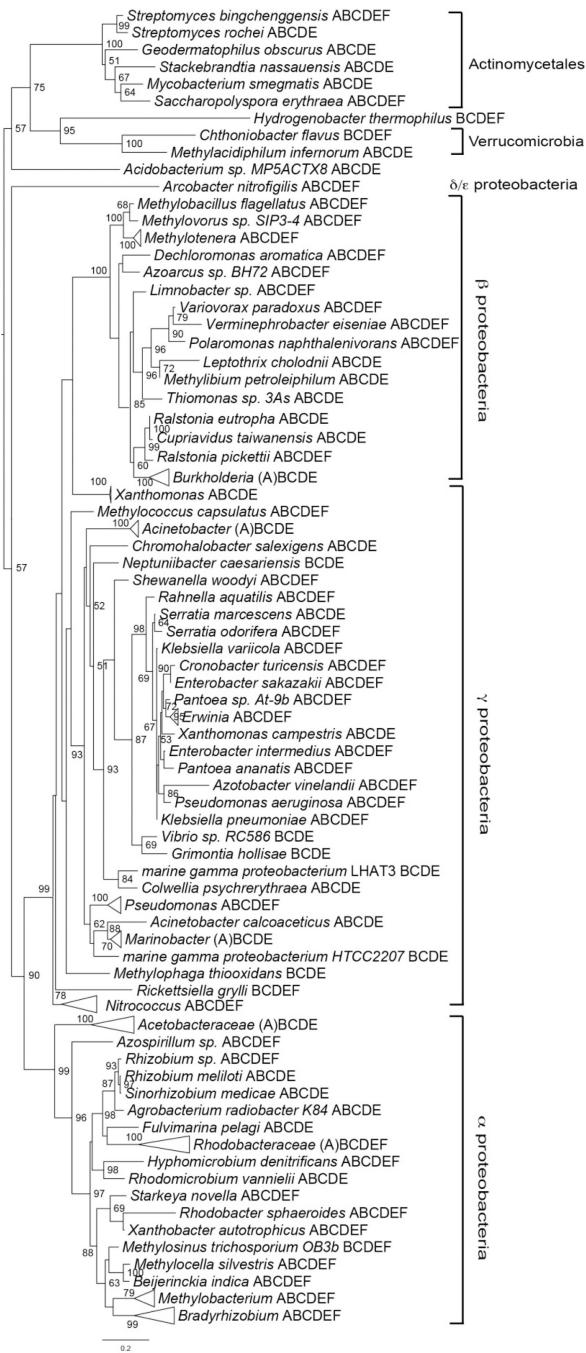


Figure 1.

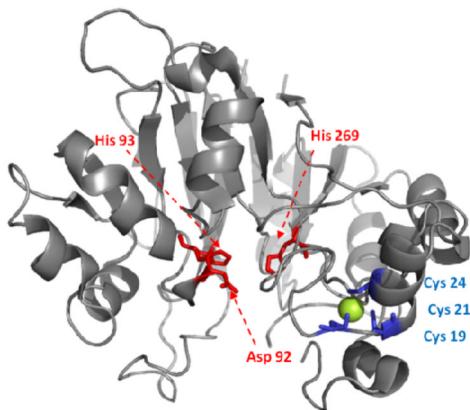
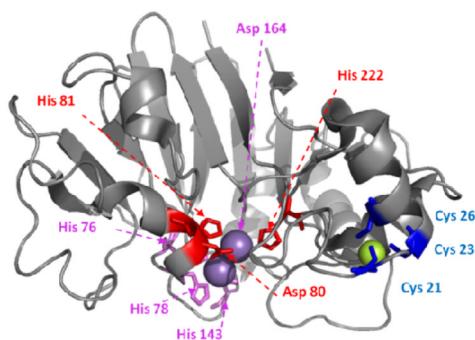
Domains in each *K. pneumoniae* Pqq proteins. All domain information is from the Pfam database, except for the Lactamase_B superfamily domain in PqqB, which comes from the Conserved Domain Database of NCBI.

**Figure 2.**

Venn diagram of core PQQ biosynthesis genes (PqqC, PqqD and PqqE). Each circle represents the number of species containing the corresponding gene. *pqqC* homologs were found in 213 species; *pqqD* homologs were found in 316 species; *pqqE* homologs were found in 331 species. All three genes were found in 126 species.

**Figure 3.**

Distribution of PQQ synthesis proteins mapped to a phylogenetic tree based on 16s ribosomal RNA. The letters following species names indicate Pqq proteins. A–F: PqqA–PqqF. Tree nodes were collapsed if all the species have the same set of Pqq proteins (PqqA was not considered in this set as it is easily missed in gene finding procedures, and labeled with (A) if it was found in at least one species in the collapsed nodes).

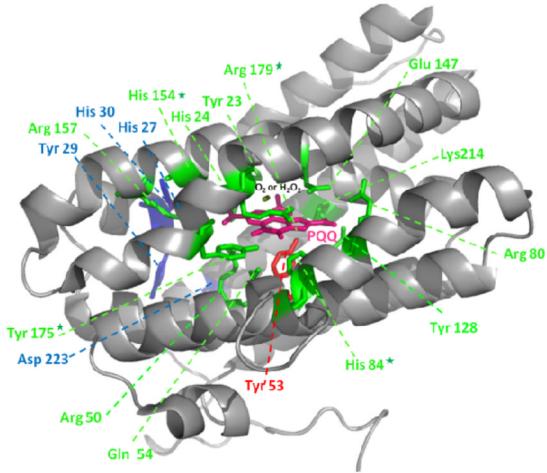
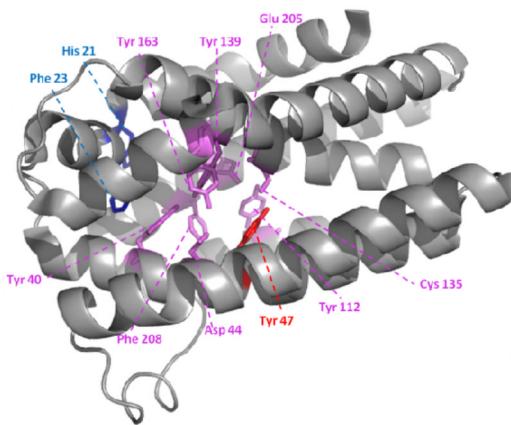
A.**B.****C.**

<pre> PqqB_B5XX61_KP 1 MFIKVLGSAAGG--GFPQWNCNCANCQGLR-----NGTIQASARTQSSIIVSDNG----KEWVLCNASPD PhnP_P16692_EC 1 MSLTTLTGTTGAQGVPAWGCECAACARARRSPQYRRQPCSGVV---KFNDAITLIDAGLHDLADRW----SPG </pre>	6 1
<pre> PqqB_B5XX61_KP 60 ISQQIAHTPELNKGVLRGTSIGGIILTDSIDHITGLLSLREGC--PHQVWCTPEVH--EDLSTGFVFMLRHNG PhnP_P16692_EC 68 SFQQ-----FLLTHYHMDHVQGLFPLRWGVGDPIPVYGPPDEQGCDDL-----FKH--PG </pre>	3
<pre> PqqB_B5XX61_KP 134 GLV--HHPIAPQQPFTVDACPDLQFTAVPIASNAPPYSPYRDRPLPGHNVALFIEY---RRNGQTLYAPGLGEPE PhnP_P16692_EC 116 LLDFSHSTV---EPFVVFDLQGLQVT-----PLPLNHSKLTFGYLLETAHSRVAWLSTAGLPE </pre>	3
<pre> PqqB_B5XX61_KP 206 AL-----LPWLQKADCLLIDGTVWQDDELQAAGVGRTGRDMGHLALSDEHGMMALLASLPAKRKILIHINNT- PhnP_P16692_EC 171 KTLKFLRNNQPQVMMDC-----SHPPRADAPRNH-----CDLNTVLALNQVIRSPRVILTHISHQF </pre>	
<pre> PqqB_B5XX61_KP 274 NPILNELSPQRQALKQQGIEVSWDGMAITLQDTAC PhnP_P16692_EC 228 DAWLME----NAL-PSGFEVGFDGMEIGV---A </pre>	

Figure 4.

Comparison of the PqqB three-dimensional structure model and PhnP structure. The spheres of Mn ions in the active site are in purple and the sphere of zinc ion is in green. Residues involved in putative structural Zn ion binding are in blue, Mn binding of the PhnP active site are in red if conserved in PqqB and in purple if not. **A:** model of the structure of PqqB from *K. pneumoniae* (UniProt accession B5XX61) using the solved structure of PqqB from *P. putida* (66% identity). **B:** structure of PhnP sequence from *E. coli* (UniProt accession PI6692, PDB ID 3G1P). **C:** The pairwise alignment of PqqB and PhnP. The numbers above

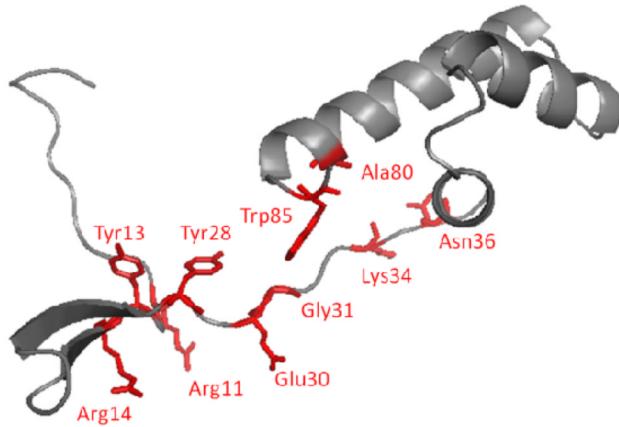
columns indicate the ranking predicted by Discern for the corresponding residue of PqqB (Table S4).

A.**B.****C.**

	17	2	1	812	6	1114	9
PqqC_B5XX60_KP	11	AFEEALRAKGDFY---	HIHHPYHIAMHNGDATRKQIQGWVANRFFYQTTIPLKDAAIMANC	PDAQTRRKWVQRIL			
TenA_P25052_BS	02	KFSEECRSAAAEWwegSFVHPFVQGIGDGLPIDRFKYVVLQDSYYLTHFAKVQSFGAAYAKDLYTTGRMASHAQ					
	4				13		5 10
PqqC_B5XX60_KP	83	DHDgshgedGGIEAWLRLGEAVGLSRDd11SERHVLPGVRFAVDAYLN	FARRACWQE--AACSSLTELFAPQIHQ				
TenA_P25052_BS	77	GTY-----EAEMALHREFEAELLEISEe-eRKAFKPSPTAYS	YTSHMYRSVLSGNFAeiLAALLPCYWLYEVGE			3 16	
	15						
PqqC_B5XX60_KP	156	SRLDSPWQhypwIKEGYFYFRSRLSQANRDVEHGLALAKAYC--DSAEKQNRMLEILQFKLDILWSMLDAMT	MAY				
TenA_P25052_BS	145	KLLHCDPG---HPIYQKWIGTYGGDWFRQVVEQINRFDELAn	STEEVRAKMKENFVISSYYEYQFWGMAYRE				

Figure 5.

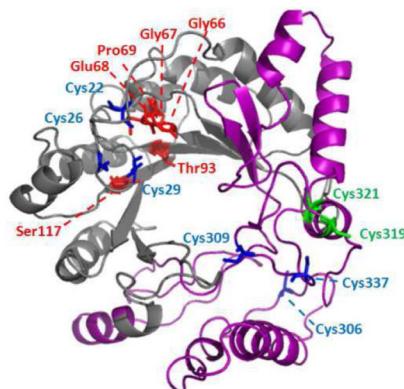
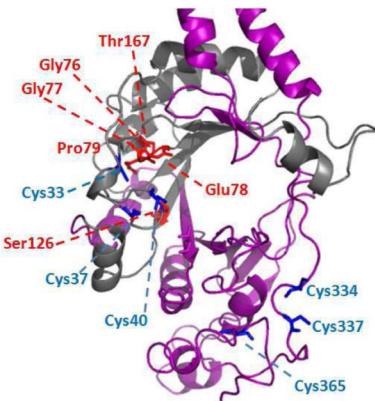
Comparison of the structures of PqqC and TenA. **A:** Structure of PqqC (B5XX60 from *K. pneumoniae*, PDB ID 1OTW). The top ranked residues predicted by Discern are labeled. Residues outside the active site are labeled in blue, the residue in the active site are in red if conserved in TenA and in green if not. The residues characteristic by biochemical experiments are marked with a star. **B:** Structure of TenA from *B. subtilis* (P25052, PDB ID lyaf). Predicted functional residues are in purple. **C:** The pairwise structural alignment of PqqC and TenA (made by VAST). The numbers above columns indicate the ranking of catalytic residues in PqqB predicted by Discern (Table S5).

A.**B.**

	4	6	5	8	17	3	10
PqqD_B5XX59_KP	MQKTS	-----	IVAFRRGYRLQWEAAQESHVILYPEGMAKLNETAAAILELVDGRRDVAIIAMLNERFPEAGGV-DDD				
PqqCD_Q49150_ME	1	L	-----PRGVRLRFDEVRNKHVLLAPERTFDLDDNAVAVLKLVDRNTVSQIAQILGQTYDADPAIIEAD				
PqqD_Q9L3B1_GO	1	MTEAPHVVAEGTVLSFARGHRLQHDRVRDVWIVQAPEKAFVVEGAAPHTLRLLDGKRSVGEIIQQLAIEFSAPREVIAKD					
PqqD_Q83X98_SR	1	MTGLP-----EPTVPLRPVGVRLTRDPARGEALL-PERVVVLNDTAAAVLAHCAGTTSLAGIVERLAEYE---GVSAED					
PqqD_O33505_RA	1	MITIT-----EHYTPMFRRGYRQMFEKTQDCHVILYPEGMAKLNDSATFILQLVDGERTIANIIDELNERFPEAGGV-NDD					
	2	9					
PqqD_B5XX59_KP	73	VVEFLQIACQQKWITCREPE	92				
PqqCD_Q49150_ME	66	ILPMLAGLAQKRVLE	-----R	81			
PqqD_Q9L3B1_GO	81	VLALLSELTEKNVLIH	-----T	96			
PqqD_Q83X98_SR	73	VRELLLRLAQRRVVDLH	--G	90			
PqqD_O33505_RA	76	VKDFFAQAHAKWITFREPA	95				

Figure 6.

Structure-sequence analysis of PqqD. **A:** Homology model of PqqD from *K. pneumoniae* (UniProt accession number B5XX59) modeled onto the three-dimensional structure of PqqD from *X. campestris* (PDB ID 3G2B). Residues ranked within the top ten by Discern are colored in red (Table S6). **B:** MSA of the five PqqD seed sequences constructed using MAFFT and displayed using Belvu. Residues ranked within the top ten by DESCERN are colored in red. KP, *K. pneumoniae*; ME, *M. extorquens AM1*; GO, *G.r oxydans 621H*; RA, *R. aquatilis*; SR, *S. rocheri* (Table S6)

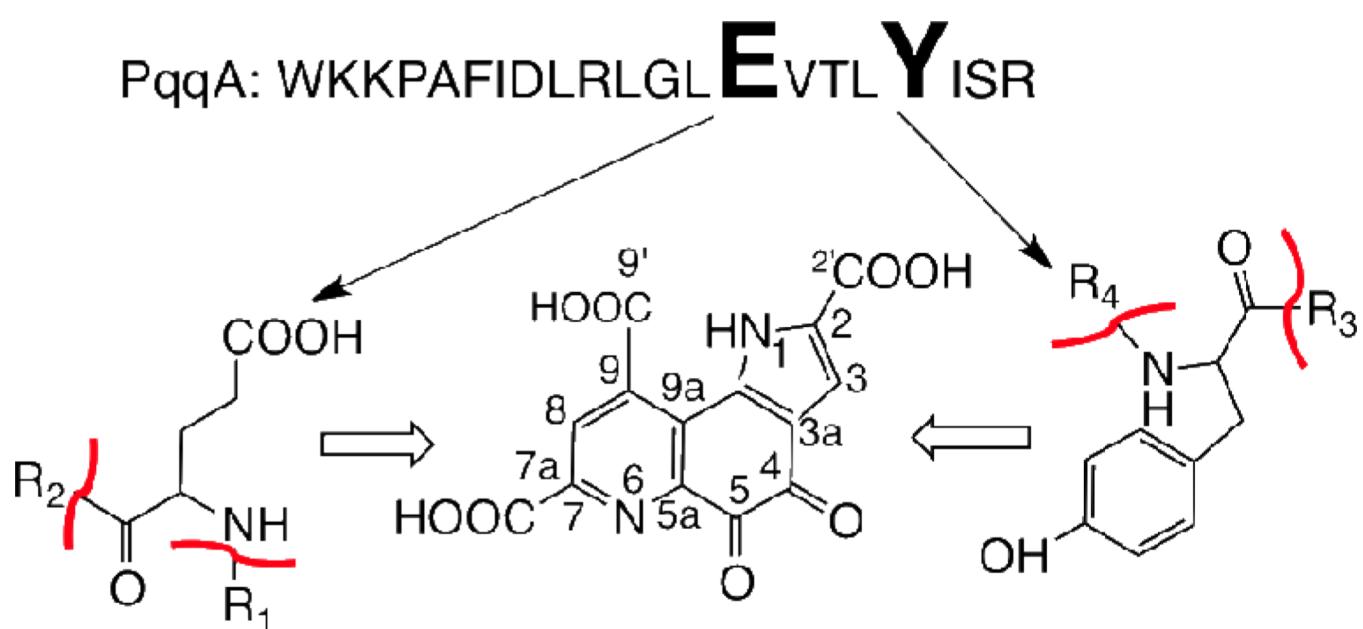
A.**B.****C.**

		2	1	3	8	7
PqqE_B5XX58_KP	1	MSQNKP AVNP-----P-----	LWLLAELTYR	CPLQCPYCSNPLDFARQEKE LTTEQWIEVFRQARAMGSVQLGFS	GGEP	
NirJ_A6Q6Z2_SS	1	MFRLSNL LSVTAGKPARVLDGS TAIWNFTNRCNLSC LHCYS -KADLD	A	DVDTLTTENIMETLP PLKANGVKFLIFS	GGEP	
PqqE_B5XX58_KP	70	LTRKDLP ELIRAARDLGFYT NLI	TSGIGL TESKLD AFSEAGLDH IQIS	SFQASD EVLNAA LAGNKKAFQQKLAMARAVK AR		
NirJ_A6Q6Z2_SS	80	LTRKD LFDIAARCKELGIVT YLSTNGLYV KKSNAEK KILD T-FDYIGIS	IDGSPEV-HDAFRGLKG FVESMKAVDLLNSF			
PqqE_B5XX58_KP	150	D-YPMVLNFVLHRHNIDQLDK IIELCIELEADDV EATCQFYGWAFLNREG	LLPTREQI ARAEQVVADYRQKMAASGNLT			
NirJ_A6Q6Z2_SS	158	G TKVGIRFTITKDTYDLQFIFELAEHHNIPK VISHLV YSGR LENLEM DLSKEQRITAV NY L DKAFEHESG--R				
PqqE_B5XX58_KP	229	NLLFVTPDY-----	YEERPK KGCMGGW-----GSIFL SVTP EGTALPCHSARQLPVAFPSV LEQSLESI			
NirJ_A6Q6Z2_SS	235	DIEIVTG NMEMDAILYDRFEKKY PEYAEEMKRR LIEWGGNSAGR KLLNIDSE GFVKPDFFP --VKIGNIL TQDFSDI				
PqqE_B5XX58_KP	287	WYD--SFGFNRYRGYDWMP EPE-----CRSCDEKEKDFGGCRQAFMLTG SADNADP VCSKSPHHHKILEAR REAACSDI				
NirJ_A6Q6Z2_SS	312	WTNEPE TLLQKLR-----VHPRELGGKC VECY QLNICNGGSR RAYAIY GDMWAEDPSC YLT EQIKGN-----				
PqqE_B5XX58_KP	359	KVSQQL FRNRTSQLIYKTREL	380			
NirJ_A6Q6Z2_SS	376	-----N	376			

Figure 7.

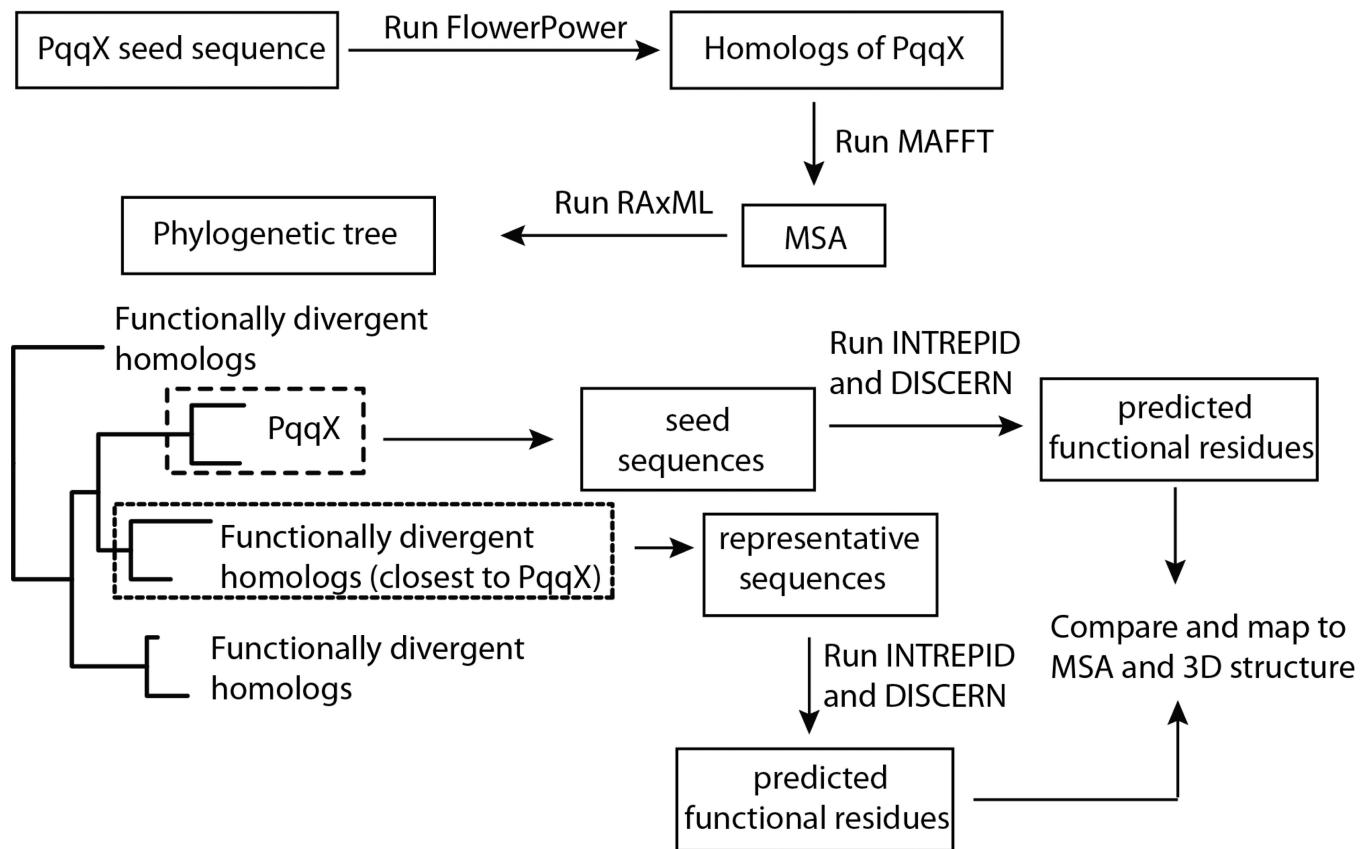
Structure-sequence analysis of PqqE. **A:** Homology models of PqqE from *K. pneumoniae* (UniProt accession number B5XX58). **B:** Homology models of NirJ from *Sulfurovum sp.* (UniProt accession A6X6Z2) constructed using the Phyre2 webserver based on the three-dimensional structure of MoaA (PDB ID 1TV8). **C:** MSA of PqqE (UniProt accession number B5XX58) and NirJ (UniProt accession number A6X6Z2) constructed using MAFFT and displayed using Belvu. The INTREPID ranking of important functional residues is indicated by numbers above the alignment (Table S7). Colors in panels **A**, **B**, and **C** correspond to the following groups: *blue*: iron sulfur cluster ligands, *red*: residues located

within the vicinity of the SAM-binding 4Fe-4S cluster, *green*: residues highly conserved in PqqE that are not present in NirJ, *purple*: regions where the homology is less reliable and so structural information is speculative. KP, *K. pneumoniae*; SS, *Sulfurovum sp.*



Scheme 1.

The proposed cross-linking of the tyrosine and the glutamate to form PQQ.

**Scheme 2.**

The bioinformatics analysis in this study. PqqX: a representative of Pqq proteins. MSA: multiple sequence alignment.

Reference sequences of PqqA–F from five known PQQ-forming organisms.*

Table 1

	<i>K. pneumoniae</i> (X58778)	<i>M. extorquens</i> (NC_012808)	<i>G. oxydans</i> (CP000009)	<i>S. rochei</i> (AB083224)	<i>R. aquatilis</i> (FJ868974)
PqqA	P27503 (PQQA_KLEPN)	Q49148 (PQQA_METEA)	Q9L3B4 (PQQA_GLUOX)	Q83X96 (Q83X96_STRRRO)	C3VIT4 (C3VIT4_RAHAQ)
PqqB	B5XXX61 (PQQB_KLEP3)	Q49149 (PQQB_METEA)	Q9L3B3 (PQQB_GLUOX)	Q83XA0 (Q83XA0_STRRRO)	C3VIT5 (C3VIT5_RAHAQ)
PqqC	B5XXX60 (PQQC_KLEP3)	Q49150 (PQQCD_METEA)	Q9L3B2 (PQQC_GLUOX)	Q83X97 (Q83X97_STRRRO)	C3VIT6 (C3VIT6_RAHAQ)
PqqD	B5XXX59 (PQQD_KLEP3)	Q49150 (PQQCD_METEA)	Q9L3B1 (PQQD_GLUOX)	Q83X98 (Q83X98_STRRRO)	O33505 (PQQD_RAHAQ)
PqqE	B5XXX58 (PQQE_KLEP3)	P71517 (PQQE_METEA)	Q9L3B0 (PQQE_GLUOX)	P59749 (PQQE_STRRRO)	O33506 (PQQE_RAHAQ)
PqqF	P27508 (PQQF_KLEPN)	C5AQL6 (C5AQL6_METEA)	absent	absent	C3VIT9 (C3VIT9_RAHAQ)

* UnitProt accession number for each protein is listed. In parentheses are UnitProt identifiers.

Table 2

Potential pathogens or symbionts containing PqqB, PqqC, PqqD and PqqE, ordered by species name (see Table S2 for a complete list of predicted PQQ-synthesis species).

Species	Pathogenicity	Set of <i>Pqq</i> genes
<i>Acinetobacter baumannii</i> ATCC 19606	Opportunistic human pathogen	ABCDE
<i>Acinetobacter haemolyticus</i> ATCC 19194	Rare human pathogen	BCDE
<i>Azoarcus</i> sp. (strain BH72)	Plant symbiont	ABCDEF
<i>Bradyrhizobium japonicum</i>	Plant symbiont	ABCDEF
<i>Bradyrhizobium</i> sp. (strain BTail / ATCC BAA-1182)	Plant symbiont	ABCDEF
<i>Burkholderia cenocepacia</i> (strain HI2424)	Human cystic fibrosis pathogen	ABCDE
<i>Burkholderia cepacia</i> (strain J2315 / LMG 16656)	Animal and plant pathogen	ABCDE
<i>Burkholderia glumae</i> (strain BGR1)	Plant pathogen	ABCDE
<i>Burkholderia multivorans</i> (strain ATCC 17616 / 249)	Animal pathogen in Mammalia	ABCDE
<i>Burkholderia phymatum</i> (strain DSM 17167 / STM815)	Plant symbiont	ABCDE
<i>Colwellia psychrerythraea</i> (strain 34H / ATCC BAA-681)	Plant pathogen	ABCDE
<i>Cronobacter turicensis</i> (strain DSM 18703/LMG 23827/z3032)	Neonatal pathogen	ABCDEF
<i>Cupriavidus taiwanensis</i> (strain R1 / LMG 19424)	Plant pathogen	ABCDE
<i>Dinoroseobacter shibae</i> (strain DFL 12)	Animal symbiont	ABCDEF
<i>Enterobacter intermedius</i>	Opportunistic pathogen	ABCDEF
<i>Enterobacter sakazakii</i> (strain ATCC BAA-894)	Neonatal pathogen	ABCDEF
<i>Erwinia amylovora</i> (strain ATCC 49946/CCPPB 0273/Ea273/ 27-3)	Plant pathogen	ABCDEF
<i>Erwinia pyrifoliae</i>	Plant pathogen	ABCDEF
<i>Erwinia tasmaniensis</i> (strain DSM 17950 / Et1/99)	Plant commensal	ABCDEF
<i>Gluconacetobacter diazotrophicus</i> (strain ATCC 49037/DSM 5601/PA15)	Plant symbiont	ABCDE
<i>Granulibacter bethesdensis</i> (strain ATCC BAA-1260/ CGDNIH1)	Human pathogen	ABCDE
<i>Grimontia holisae</i> CIP 101886	Human pathogen	BCDE
<i>Klebsiella pneumoniae</i>	Opportunistic pathogen	ABCDEF
<i>Klebsiella</i> sp. 1_1_55	Opportunistic pathogen	BCDEF
<i>Marinobacter algicola</i> DG893	Plant symbiont	ABCDE
<i>Methylobacterium nodulans</i> (strain ORS2060 / LMG 21967)	Plant saprophyte and symbiont	ABCDEF
<i>Methylobacterium populi</i> (strain ATCC BAA-705/NCIMB 13946/BJ001)	Plant endophyte	ABCDEF
<i>Methylobacterium radiotolerans</i> (strain ATCC 27329/DSM 1819/JCM 2831)	Plant symbiont	ABCDEF
<i>Mycobacterium smegmatis</i> (strain ATCC 700084/mc(2)155)	Commensal in Mammalia	ABCDE
<i>Pantoea ananatis</i>	Plant pathogen	ABCDEF
<i>Pseudomonas aeruginosa</i>	Opportunistic human pathogen	ABCDEF
<i>Pseudomonas entomophila</i> (strain L48)	Insect pathogen	ABCDEF
<i>Pseudomonas fluorescens</i>	Potential pathogen to birds	ABCDEF
<i>Pseudomonas mendocina</i> (strain ymp)	Rare human pathogen	ABCDEF
<i>Pseudomonas savastanoi</i> pv. <i>savastanoi</i> NCPPB 3335	tumor-inducing pathogen	ABCDEF
<i>Pseudomonas syringae</i> pv. <i>phaseolicola</i> (strain 1448A/Race 6)	Plant symbiont	ABCDEF

Species	Pathogenicity	Set of <i>Pqq</i> genes
<i>Rahnella aquatilis</i>	Opportunistic pathogen	ABCDEF
<i>Ralstonia pickettii</i> (strain 12D)	Opportunistic pathogen	ABCDEF
<i>Rhizobium meliloti</i>	Plant symbiont	ABCDE
<i>Rhizobium</i> sp. (strain NGR234)	Plant symbiont	ABCDEF
<i>Rickettsiella grylli</i>	Athropod pathogen	BCDEF
<i>Serratia marcescens</i>	Opportunistic human pathogen	ABCDE
<i>Serratia odorifera</i> 4Rx13	Opportunistic pathogen	ABCDEF
<i>Verminephrobacter eiseniae</i> (strain EF01-2)	Animal endosymbiont	ABCDEF
<i>Xanthomonas axonopodis</i> pv. <i>citri</i> (Citrus canker)	Plant pathogen	ABCDE
<i>Xanthomonas campestris</i> pv. <i>Campestris</i>	Plant pathogen	ABCDE
<i>Xanthomonas oryzae</i> pv. <i>Oryzae</i>	Rice bacterial blight <i>pathogen</i>	ABCDE