SUPPORTING INFORMATION

A two-component protease in *Methylorubrum extorquens* with high activity toward the peptide precursor of the redox cofactor pyrroloquinoline quinone

Ana M. Martins[1], John Latham[2], Paulo Martel[3], Ian Barr[1,4#], Anthony Iavarone[1,4], Judith P. Klinman[1,4,5*]

From the [1]California Institute for Quantitative Biosciences, University of California Berkeley, Berkeley CA 94720 USA; [2]Department of Chemistry and Biochemistry, University of Denver, Denver CO 80208 USA; [3]Center for Molecular and Structural Biomedicine, University of Algarve, 8005-139 Faro, Portugal; [4]Department of Chemistry, University of California Berkeley, Berkeley CA 94720 USA; [5]Department of Molecular and Cell Biology, University of California Berkeley, Berkeley CA 94720 USA.

Table S1. Summary of the bacterial genera and species studied in Shen *et al.* (1).

Class/phylum/order	Genus	Species
α-proteobacteria	30	41
β-proteobacteria	18	30
γ-proteobacteria	23	45
ε-proteobacteria	1	1
Actinobacteria	5	6
Verrucomicrobia	2	2
Acidobacteriia	1	1
Aquificales	1	1

Table S2. Bacterial species analyzed for Pqq-encoding genes by Shen and co-workers (1). The Excel Table shows the bacterial species, grouped by phylum and class (in the case of the phylum Proteobacteria), the PQQ production of each strain (according to (2)), the gene arrangement from the analysis of Shen *et al.* (1), the gene arrangement found in this analysis the PqqF and PqqG definitions, GenBank accession numbers, number of amino acid residues, zinc-binding consensus sequences (for PqqF) and whether the genes encoding PqqF and PqqG are contiguous and the location concerning the *pqq* operon.

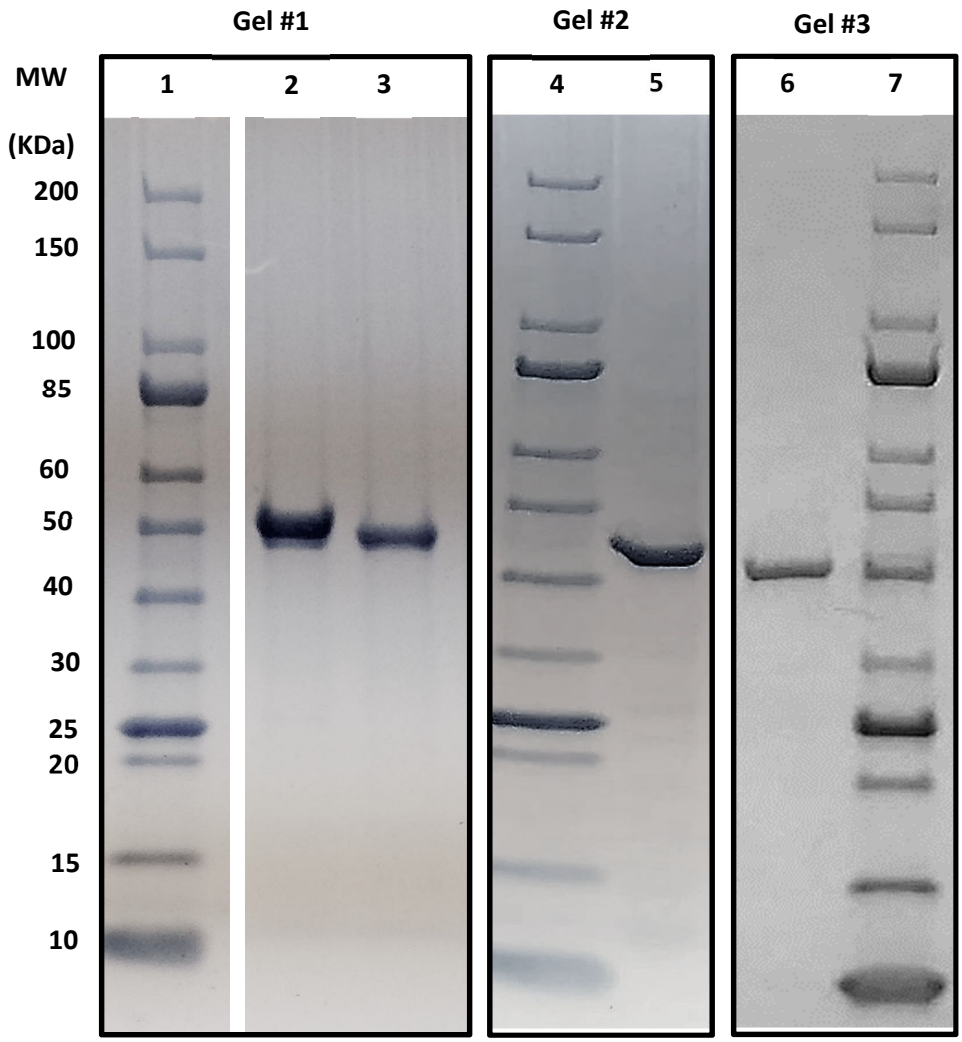


Figure S1. SDS-PAGE gels of purified, his-tagged and untagged PqqF and PqqG. The lanes are as follows: (1), (4), (7) protein ladder (10 to 200 KDa), (2) His_6PqqF, (3) PqqF, (5) His_6PqqG and (6) PqqG. The molecular weights predicted from the protein sequences are: His_6PqqF 53.14 kDa, PqqF 51.26 kDa, His_6PqqG 47.90 kDa, PqqG 46.01 kDa. Gels #1, #2 and #3 were obtained in different experiments. Gel #1 shows the ladder lane separated from the protein sample lanes.

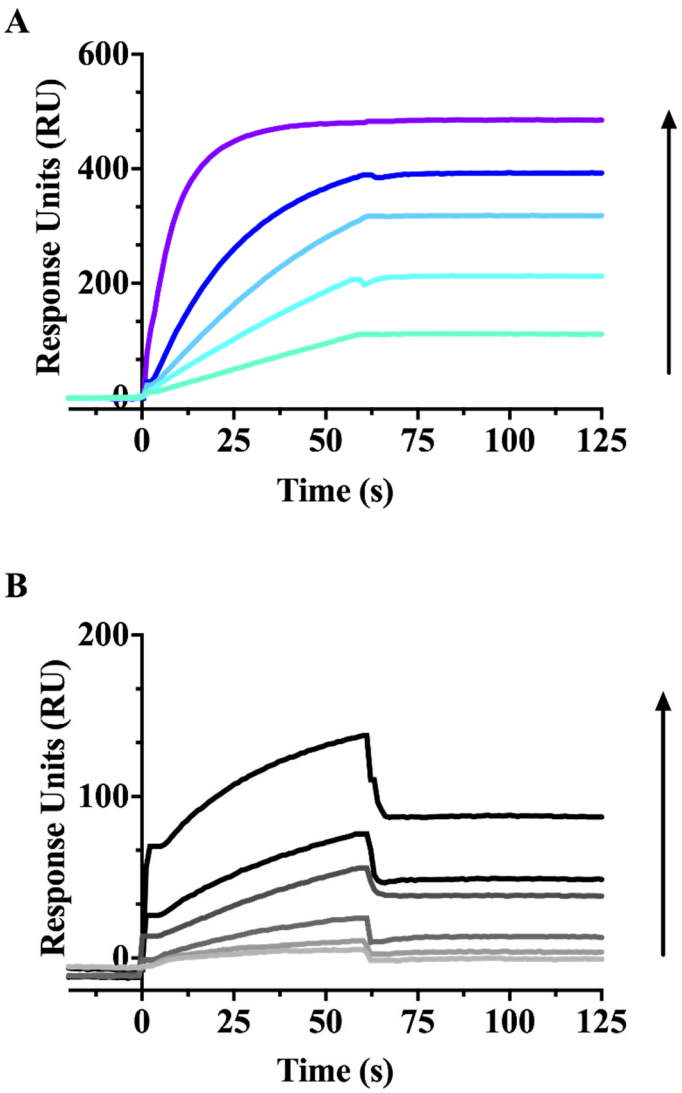


FIGURE S2. Binding of PqqF and PqqG observed by SPR. Representative datasets of each experiment are shown. (*A*) His_6PqqG was used as a ligand and PqqF was used as analyte, (*B*) His_6PqqF was used as a ligand and PqqG as used as analyte. In both cases the SPR response increased as the analyte concentration increased, from 45 nM to 9 μM. However, both PqqF and PqqG also showed non-specific adsorption to the SPR chip, leading us to perform K_d measurements by ITC. (see Fig. 4 in main text).

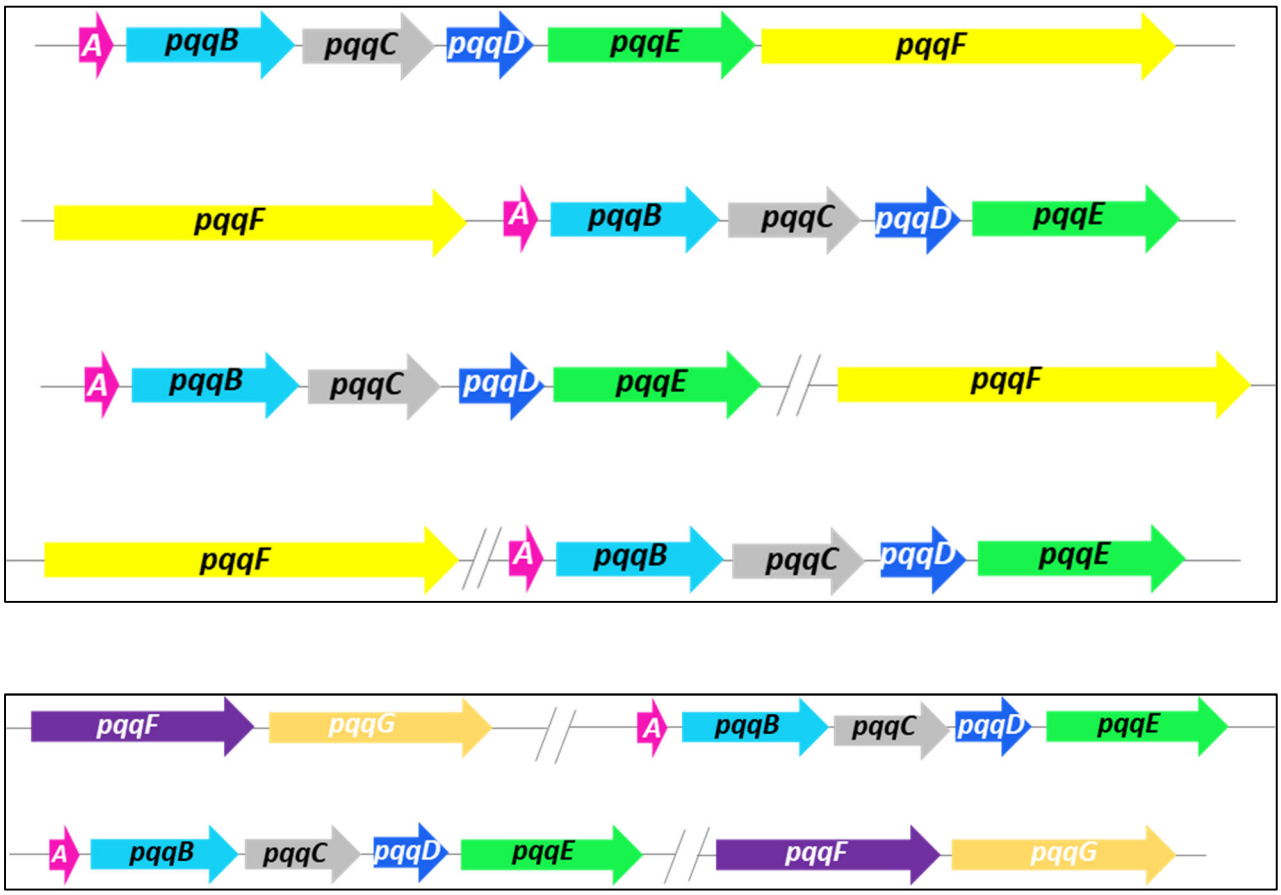


FIGURE S3. The most usual PQQ gene arrangements in the genome of *(top panel)* γ-proteobacteria and (*bottom panel)* α-proteobacteria.

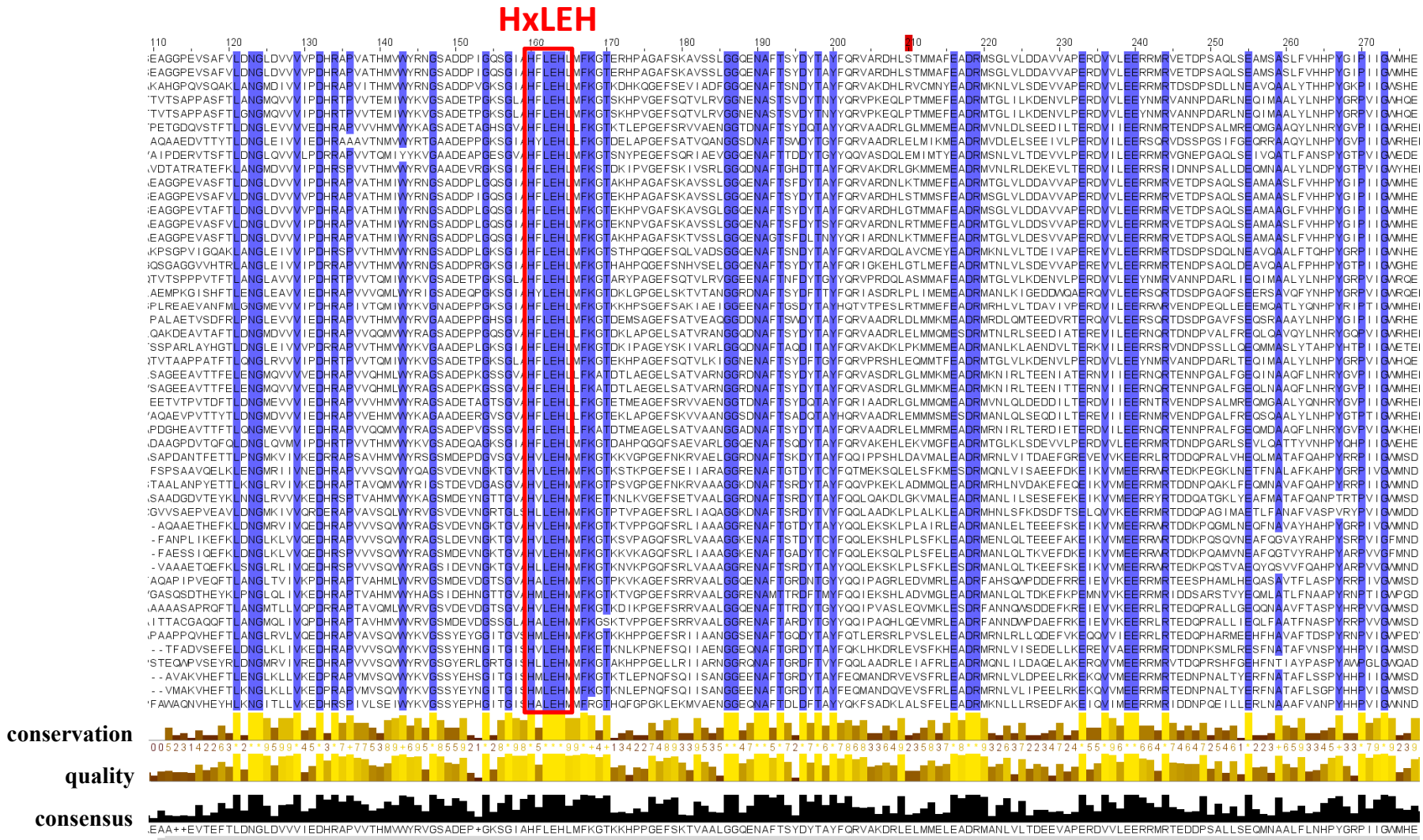


Figure S4. Segment of the sequence alignment of the N-terminal of PqqF (M16B peptidases) from α-proteobacteria. Highlighted in red is the zinc-binding consensus HxLEH. Sequences were aligned using Clustal Omega and the figure was generated and analyzed in Jalview (3).

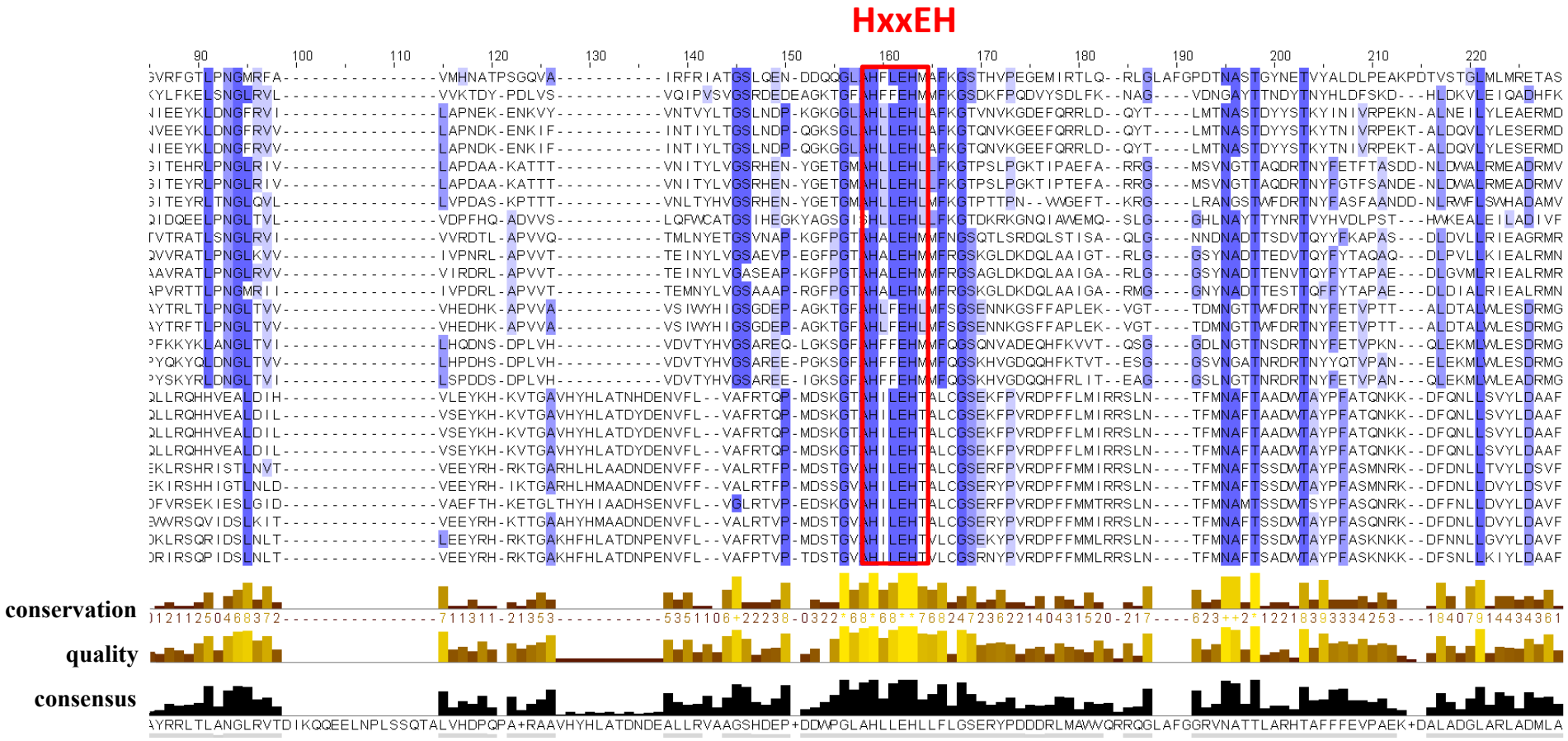


Figure S5A. Segment of the sequence alignment of the N-terminal of PqqF (M16A peptidases) from γ-proteobacteria. Highlighted in red is the zinc-binding consensus HxxEH. Sequences were aligned using Clustal Omega and the figure was generated and analyzed in Jalview (3).

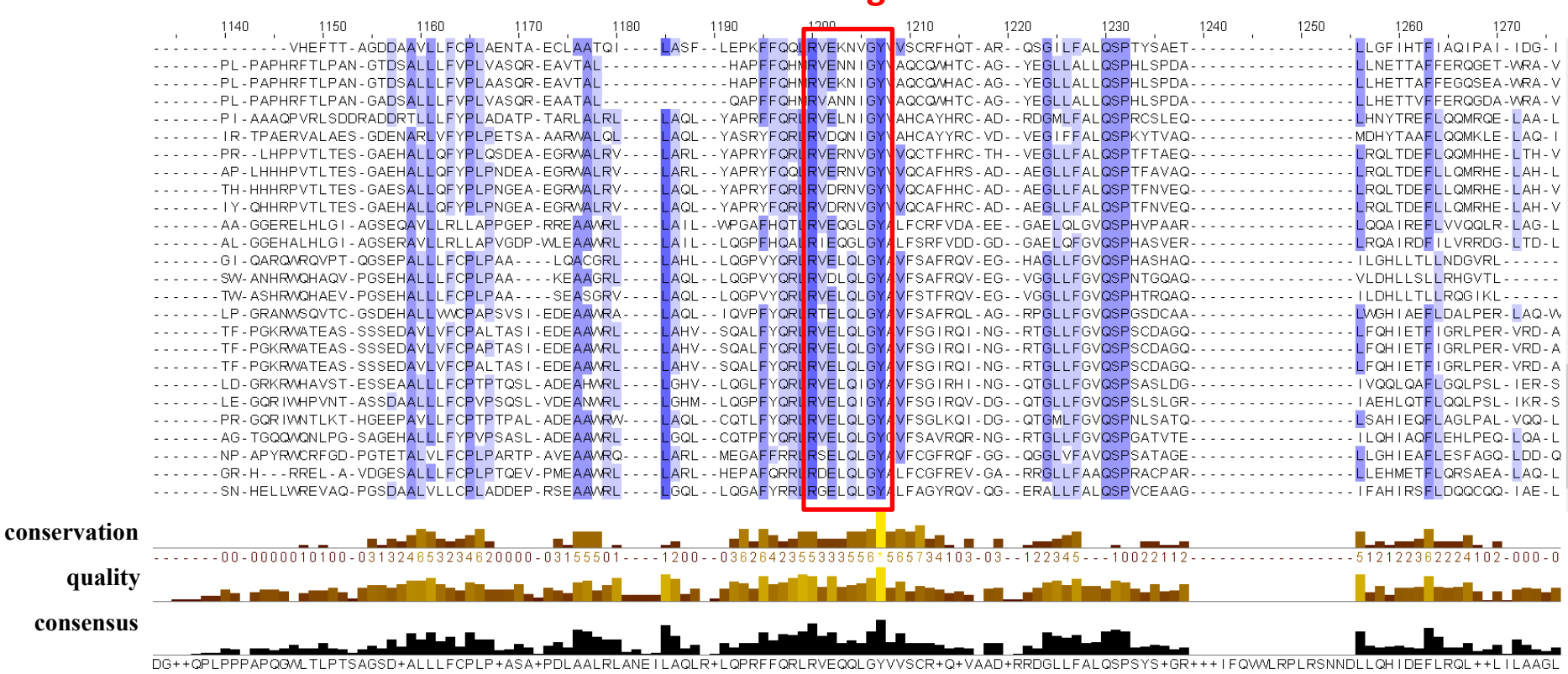


Figure S5B. Segment of the sequence alignment of the C-terminal of PqqF (M16A peptidases) from γ-proteobacteria. Highlighted in red is the RY region characteristic of the C-terminal of this type of M16 peptidases. Sequences were aligned using Clustal Omega and the figure was generated and analyzed in Jalview (3).

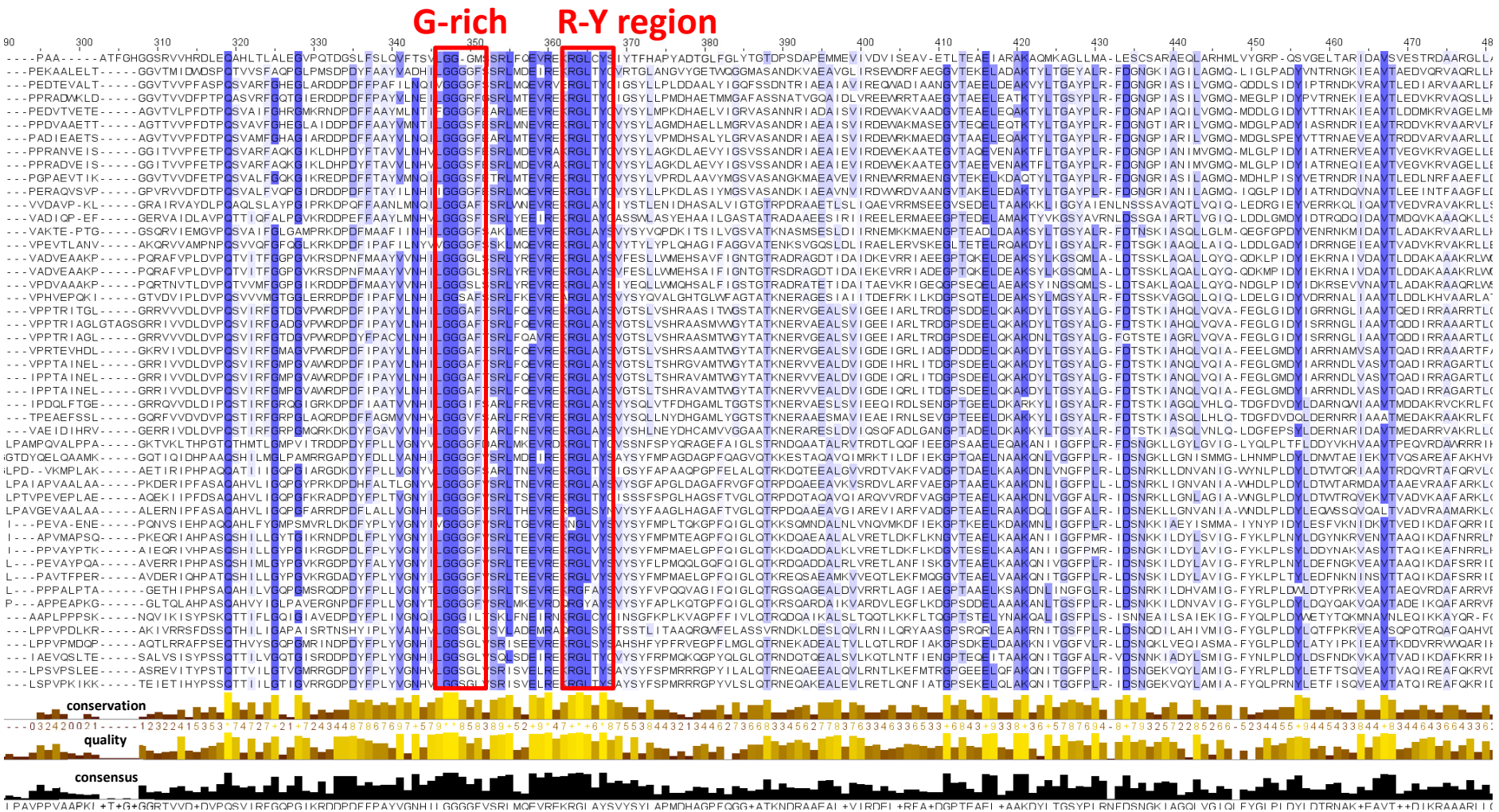
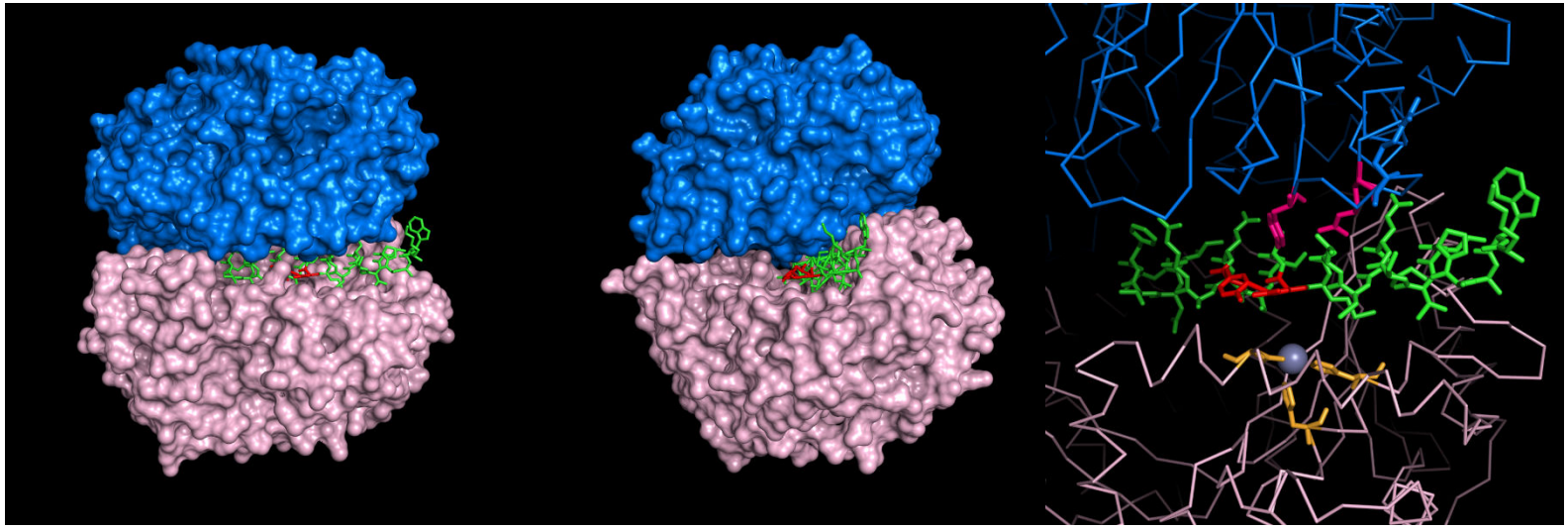


Figure S6. Segment of the sequence alignment of PqqG (M16B peptidase, non-peptidase homolog chain) from α-proteobacteria. Highlighted in red is the RY region characteristic of the C-terminal of this type of M16 peptidases. Sequences were aligned using Clustal Omega and the figure was generated and analyzed in Jalview (3).



A **B** **C**

Figure S7. Idealized representation of the interaction between the crosslinked PpqA* peptide (green sticks) and the PqqF/PqqG heterodimer protease (blue and pink surfaces). The dimer structure was modeled on the open conformation of *Sphingomonas* sp. A1 dimer (PDB file 3amj, chains C+D) to show the peptide fitting in the active site. (*A*) Front view looking straight at the dimer interface; (*B*) Rotated view (45º degrees about the vertical axis); (*C*) Close-up of the peptide. The gap near the metal center (grey with coordinating residues in orange) and R/Y pair (magenta) suggest a possible ingress path for the peptide. Because the heterodimer is represented in the open clam shell conformation it is unlikely that the R/Y pair is interacting with PqqA* in this situation. Also note that the cross-link between residues Glu15 and Tyr19 of PqqA*, in red, may help in stabilizing a helical conformation for the peptide.

REFERENCES

1. Shen, Y. Q., Bonnot, F., Imsand, E. M., RoseFigura, J. M., Sjolander, K., and Klinman, J. P. (2012) Distribution and properties of the genes encoding the biosynthesis of the bacterial cofactor, pyrroloquinoline quinone. *Biochemistry* 51, 2265-2275
2. RoseFigura, J. M. (2010) *Investigation of the structure and mechanism of a PQQ biosynthetic pathway component, PqqC, and a bioinformatics analysis of potential PQQ producing organisms.* Ph.D., UC Berkeley
3. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189-1191