Qeios

Peer Review

# Review of: "Multinuclear Non-Heme Iron-Dependent Oxidative Enzymes: Landscape of Their Substrates, Partner Proteins"

**Daniel H. Haft**[1]

1. National Institutes of Health, Bethesda, United States

This paper provides a thorough analysis of multinuclear nonheme iron-dependent oxidase (MNIO) enzymes found in known or predicted ribosomally synthesized and post-translationally modified peptide (RiPP)-type biosynthetic gene clusters (BGC). Most of the larger clusters of MNIO enzymes are now reasonably well understood because of successful experimental work, including work on several forms of bufferin-type RiPPs by the authors of this paper, but most of the smaller clusters still await their first description in the literature.

The comprehensive review in the introductory materials of RiPP systems dependent on MNIO enzymes is clear, highly informative, and easy to read. Throughout the paper, the authors cite actual accession numbers both for existing protein family-defining Hidden Markov Models (HMMs) and for individual proteins. This courtesy makes it much easier for readers to go from the text of the paper to target regions in bacterial genomes. It is deeply appreciated.

The authors' use of a workflow of MNIO enzyme identification, subclustering done by building sequence similarity networks (SSN), examination of conserved gene neighborhoods surrounding MNIO enzymes based on prediction of protein roles based on Pfam domain content shared with characterized BGC, gene order, protein length, and cysteine-rich regions is a sound method. Experience in biochemical characterization, protein family curation, or both can provide critical insights for aiding in the interpretation of bioinformatics. This study was crisp and interesting to read and will add significantly to our understanding of MNIO enzyme biochemistry.



doi.org/10.32388/GR3WXT

Major comments:

For a minority of the 21 MNIO clusters most closely examined, the apparent BGC was one not yet characterized, and furthermore, one of the conserved genes in the cluster was proposed to encode a new family of RiPP precursor peptides. These are, arguably, the highest interest novel findings in the paper. At least the three or four most promising of these should be shown in a multiple sequence alignment. My personal preference would be an alignment colored by residue type, so that Cys residues could have their own distinctive color. The MNIO clusters said to have candidate RiPP precursors adjacent included those designated 1, 2, 3, 7, and 9, as well as clusters designated I and IV. The alignment I was able to build around WP_073215760 from Cluster 1 in Figure 5 was surprising in several ways, including the central location of the GGCNCG motif. Building an alignment for cluster 2 around WP_127967852.1 also brings surprises (more than half of the members have a Sec-independent "twin-Arginine" type of signal peptide). The Cys-containing motif, GCxxCGxxGxD, looks plausible for heme-binding, but also for post-translational modification. While the logic of the paper was sound and easy enough to follow, and the figures were both clear and expressive, the impact of the new findings is somewhat diminished by the lack of multiple sequence alignments for the suggested new RiPP families.

The new information provided about family TIGR04222 is certainly intriguing. The frequent proximity to MNIO enzymes was previously not described, and it strongly supports the notion that the Cys-rich extension regions most members carry undergo MNIO-catalyzed post-translational modification in that C-terminal region. However, non-fragment members of TIGR04222 are much longer than the leader peptide regions of typical RiPPs. That region, with multiple hydrophobic stretches, may serve some maturation function such as RRE, or transport across the plasma membrane, and may be able to act on RiPP precursors that are not fused. The suggestion that TIGR04222 lacking the RiPP-like extension region may function as a target, not a modification accessory protein, may be incorrect. Perhaps the discussion could be adjusted to further explore the meaning of the shared region of TIGR04222 family proteins.

Minor comments:

When citing TIGRFAMs models, please recognize that the database has become part of NCBIFAMs. The proper citation now is PMID:37962425, and a URL is

https://www.ncbi.nlm.nih.gov/refseq/annotation_prok/tigrfams/

Each model has a web page, e.g.

https://www.ncbi.nlm.nih.gov/genome/annotation_prok/evidence/TIGR04160/

https://www.ncbi.nlm.nih.gov/genome/annotation_prok/evidence/TIGR04222/

It would be good to add URLs for Pfam models as well,

MbnB_TglH_ChrH is https://www.ebi.ac.uk/interpro/entry/pfam/PF05114/ etc. (the former DUF692).

The alignment should be shown for the putative RiPP family of Cluster 3, even if AlphaFold balks at predicting a structure with confidence. That should be okay for a RiPP family because chaperones bind the precursor, the leader peptide typically is lost, and the mature product may have cross-links that fold prediction software should not be expected to reproduce. At a minimum, a protein accession should be provided.

The AlphaFold prediction for the putative RiPP from Cluster 9 probably should not be shown, as the separation of the long helices from each other suggests the fold prediction overall did very little packing of secondary structure and has very low confidence.

I found several HMMs in supplementary materials, as text in a PDF document. However, the preparation of those materials introduced newline characters that corrupted the HMM and made them unusable. Please find a different format for ma

king the HMMs available. Providing curated and trimmed multiple sequence alignments, from which HMMs can be made in one step, would be preferable to the HMMs alone.

It is virtually certain that protein families will be built around the proposed new RiPP precursor proteins that this work has identified and reported. Even if the RiPP hypothesis does not prove true, these families are likely to have some role in the pathway that involves the neighboring MNIO. The protein families that are built, and will serve as annotation rules in bacterial genome pipelines, and that will cite this paper, will need a standard name to use for member proteins. It would be good to have the authors suggest names for these families, even if they are as contrived as, say, "adjamnio-1 family protein", (ADJAcent to MNIO cluster 1). Names for the new families would help improve genome annotation for these BGCs more rapidly than could be done in their absence.

## Declarations

**Potential competing interests:** No potential competing interests to declare.