Multinuclear non-heme iron dependent oxidative enzymes: Landscape of their substrates, partner proteins and biosynthetic gene clusters

R. Antoine,[a#] L. Leprevost,[a] S. Jünger,[b] S. Zirah,[b] G. Lippens,[c]
Y. Li,[b] S. Dubiley,[c] F. Jacob-Dubuisson[a#]

[a] Univ. Lille, U1018-UMR9017-CIIL-Center for Infection and Immunity of Lille, Lille, FRANCE

[b] Unit Molecules of Communication and Adaptation of Microorganisms (MCAM), UMR 7245 CNRS, MNHN, Alliance Sorbonne University, Paris, FRANCE

[c] Toulouse Biotechnology Institute, CNRS/INRAE/INSA/UPS, 31077 Toulouse, FRANCE

Running head: Substrates and gene clusters of MNIO enzymes

[#] for correspondence: rudy.antoine@pasteur-lille.fr; francoise.jacob@ibl.cnrs.fr

Keywords: ribosomally synthesized post translationally modified peptides (RiPPs); multinuclear non heme iron-dependent oxidative (MNIO) enzymes; biosynthetic gene clusters; genome mining; sequence similarity network (SSN) analysis

Abstract. Proteins of the multinuclear non-heme iron-dependent oxidative (MNIO) enzyme superfamily catalyze various modification reactions on the precursors of ribosomally synthesized, post-translationally modified peptides (RiPPs). We recently identified two large families of MNIO-modified RiPPs called bufferins, which enhance bacterial growth under copper stress by chelating the excess metal ions. Here, we explored the diversity of potential MNIO substrates by performing extensive *in silico* studies. Analyses of MNIO-coding biosynthetic gene clusters (BGCs) identified various groups of putative precursors most of which are characterized by specific Cys-containing motifs, throughout the eubacterial phylogenetic tree. The precursors of most MNIO-modified RiPPs harbor N-terminal Sec-dependent signal peptides, a rare feature among bacterial RiPPs. Some precursors are very long relative to those of typical RiPPs, indicating that MNIO enzymes could modify both peptide and protein substrates. We also identified a distinct family of integral membrane proteins with large predicted extra-cytoplasmic domains mostly found in Actinomycetota, frequently but not systematically associated with MNIOs. Most MNIO BGCs harbor genes coding for DUF2063 domain-containing proteins or structurally related proteins, serving as partners of the enzymes for precursor modification. We uncovered a correlation between the presence or the absence of Sec signal peptides in the precursors and the types of partner proteins of the MNIO enzymes. This study depicts the global landscape of potential MNIO-dependent natural products by unveiling groups of peptides and proteins genetically associated with MNIOs. It reveals a treasure trove of potential new RiPP precursors which likely represent a widespread bacterial strategy to deal with copper stress, and most likely other stresses, in natural environments.

Impact statement. The multinuclear non-heme iron-dependent oxidative (MNIO) enzymes belong to an emerging superfamily of modification enzymes that catalyze various oxidation reactions on ribosomally synthesized post-translationally modified peptides (RiPP) precursors.

The largest families of MNIO-modified RiPPs, called bufferins, are involved in copper homeostasis. In this work we performed extensive *in silico* analyses to explore the diversity of RiPP precursors genetically associated with MNIO enzymes and identified major families. Some precursors are much larger than typical RiPP precursors, indicating that MNIO enzymes may also modify proteins. We defined subtypes of MNIO enzymes' partner proteins dedicated to specific families of precursors. Our analyses of the biosynthetic gene clusters unveiled functions beyond copper homeostasis, likely for the response to other metal stresses. The global landscape of MNIO-modified RiPP precursors will be a basis for investigations into new RiPP families and will undoubtedly lead to the discovery of new modifications and new functions.

Introduction

Peptide-based natural products can notably be synthesized by ribosomes and then enzymatically modified like ribosomally synthesized post-translationally modified peptides (RiPPs) or fully synthesized by enzymatic pathways (1, 2). RiPPs are highly prevalent natural products in bacteria, with several dozen families described over the last decades (2-5). Some eukaryotic organisms, particularly plants, cone snails and fungi, also produce RiPPs (6).

Bacterial RiPPs are encoded in biosynthetic gene clusters (BGCs) that comprise at least one precursor gene and one or more gene(s) coding for modification enzymes. The RiPP precursors harbor a core region that undergoes enzymatic installation of post-translational modifications (PTMs), generally flanked at its N-terminus by a leader peptide (7), or more rarely at its C terminus by a follower peptide (8). In many instances the leader peptide is recognized by a system-specific helix-winged-helix motif-containing 'RiPP recognition element' (RRE) that can be a discrete domain or the C- or N-terminal domain of a larger protein (9-11) and that presents the substrate to the modification enzyme(s) (9, 12). Several modes of interaction between the RiPP precursor and the RRE domain as well as RRE domain-independent modes

78 of precursor recognition have been reported (9, 11, 13-17). RiPP BGCs also frequently
79 comprise genes coding for maturation proteases and export systems (18, 19), as well as for
80 cognate immunity proteins in the case of RiPP bacteriocins (20).

81 RiPPs modified by members of the multinuclear non-heme iron oxidative (MNIO) enzymes
82 (Pfam short name MbnB_TglH_ChrH; Pfam entry PF05114) superfamily represent a vast,
83 emerging RiPP class, as described in a recent review (21) (Table 1). MNIO enzymes catalyze
84 various 2-or 4-electron oxidation reactions targeting mostly Cys residues and frequently work
85 with partner proteins, named DUF2063 (Domain of Unknown Function) (Pfam entry PF09836),
86 Anthrone_oxy (Pfam entry PF08592) or methbact_MbnC (NCBIFAM entry TIGR04160) (22-
87 24). The first MNIO-modified RiPPs identified, methanobactins, are chalkophores that
88 scavenge copper in the milieu for the assembly of cuproenzymes called methane
89 monooxygenases in Methanobacteria (24-26). Methanobactins were identified in non-
90 methanogenic bacteria as well (27). In complex with its partner protein MbnC, the MNIO
91 enzyme MbnB oxidizes specific internal Cys residues of the MbnA methanobactin precursor to
92 form copper-chelating oxazolone thioamide groups (26). Other MNIO enzymes modify C-
93 terminal Cys residues first added to precursor scaffold peptides, thus generating thia-amino acid
94 derivatives called 'pearlins' (28, 29). Yet another type of MNIO-modified RiPP, chryseobasin
95 harbors an imidazolidinedione heterocycle, a macrocycle, thioaminals and a thiomethyl group
96 (23). Non-Cys modifying MNIO enzymes were described to exert their catalytic activity on C-
97 terminal Asp or Asn residues, generating C-terminally amidated peptides or aminopyruvic acid,
98 respectively (27, 30).

99 We have described two large families of RiPPs modified by MNIO enzymes involved in
100 copper homeostasis and called them bufferins (22). Bufferins belong to the Buf1 (named
101 DUF2282; Pfam entry PF10048) or Buf2 (which we previously called Buf_6/12Cys) families.
102 Copper is a necessary transition metal that is toxic in excess (31), and bufferins use Cys-derived

103 thiooxazole heterocycles to sequester this metal in the periplasm (22). The recently reported
104 oxazolin represents another family of MNIO-modified RiPPs that bind copper, but it appears
105 to do so using oxazolone thioamide groups, similar to methanobactins (32).

106 Unlike the precursors of methanobactins, pearlins and aminopyruvatives, those of bufferins
107 and oxazolins feature Sec-dependent N-terminal signal peptides rather than typical leader
108 peptides, which is highly unusual for bacterial RiPP precursors (22). Although the presence of
109 Sec signal peptides was suggested in a few other RiPP precursors, the use of the general Sec
110 export machinery in the natural host has only been shown experimentally for bufferins (22, 33,
111 34). Thus, this appears to be an original feature among RiPP precursors. The partners of the
112 bufferin-modifying MNIO enzymes are proteins composed of a DUF2063 domain fused to an
113 RRE-type domain, another specific feature of this RiPP class (22).

114 Thousands of MNIOs have been identified *in silico* (22, 23, 27, 35, 36). In this work we
115 explored the global diversity of the MNIO BGCs by genome mining approaches and identified
116 new potential MNIO substrates including *bona fide* proteins, and new partner proteins.

117

118 **Results**

119 *Search for MNIO-associated RiPP precursors*. In order to identify the various types of putative
120 MNIO substrates, we collected the members of the MNIO family (approx. 14,000 proteins at <
121 90% sequence identity) using the Pfam MbnB_TglH_ChrH signature. These enzymes are
122 distributed across a number of bacterial taxonomic groups (Fig. 1A). Sequence similarity
123 network (SSN) analyses showed that Buf1- and Buf2-associated MNIO enzymes represent a
124 sizeable fraction of the largest sequence cluster of the representative node network (red and
125 yellow dots in Fig. 1B). MNIOs genetically associated with oxazolins (hereafter called
126 Buf_EGKCG/oxazolins, see below) are also part of this large cluster (pale blue dots in Fig. 1B).
127 Some of the small MNIO sequence clusters contain enzymes associated with methanobactins,

128 pearlins, methanobactin-like peptides from *Vibrio*, chryseobasins and aminopyruvatides (23,
129 26-28, 30). As evidenced from this analysis, many other types of potential MNIO substrates
130 remain to be described. It is likely that the small groups of MNIO enzymes mediate the
131 installation of a diversity of post-translational modifications.

132 We utilized a home-made RODEO-type approach (37, 38) to collect the BGCs of all these
133 MNIO enzymes, and we analyzed their compositions (Table S1). Known domains frequently
134 encoded in the vicinity of MNIO-coding genes notably include DUF2282, DUF2063, DoxX
135 (Pfam entry PF07681), Sigma-70 region 4 (Pfam entry PF08281), Sigma-70 region 2 (Pfam
136 entry PF04542), and NrsF (Pfam entry PF06532) , all found in bufferin BGCs (22). Because
137 no signatures were available yet for Buf2 or for Buf_EGKCG/oxazolins, we generated new
138 hmm profiles for these two groups (Suppl. Files S1 and S2).

139 Next, we performed two-step analyses to characterize the various types of putative RiPP
140 precursors (Fig. 2A). We first identified potential precursors in all BGCs collected, as described
141 below, sorted them through SSN analyses and where possible we built hmm profiles for new
142 protein families. Then, we used the available or the newly generated profiles to collect all
143 members of each family from the NCBI nr Protein database, irrespective of their genetic
144 environment, retrieved the corresponding BGCs and analyzed their composition.

145 From our initial set of BGCs, we first retrieved putative precursors shorter than 200 residues,
146 harboring Cys residues and a predicted Sec signal-peptide and coded by genes that precede the
147 MNIO genes. These criteria were motivated by the fact that the overwhelming majority of
148 MNIO substrates appear to modify Cys residues (21) (Fig. 1B), and by preliminary analyses
149 showing the strict conservation of the gene order *bufABC* (encoding the bufferin precursor, the
150 MNIO enzyme and the partner protein, respectively) in *bufferin* BGCs (22). Classification of
151 the >11,000 identified proteins by SSN analyses revealed large sequence clusters of Buf1, Buf2
152 and Buf_EGKCG/oxazolin homologs (Fig. 1C). In the latter cluster a sizeable proportion of

153 proteins correspond to the EF-hand_5 name (Pfam entry PF13202) (see below). Most other
154 proteins, which belong to small sequence clusters, lacked known domain signatures.
155 To enlarge our range of putative MNIO substrates, we also retrieved from our initial BGC
156 collection all proteins with Cys residues but devoid of predicted Sec signal-peptides and without
157 size restriction, whose genes precede MNIO genes. This yielded >1000 proteins, 40% of which
158 correspond to the NCBIfam entry TIGR04222. An SSN analysis with all these proteins revealed
159 a very scattered node network, with many sequence clusters harboring TIGR04222 proteins,
160 indicating the considerable sequence diversity of the latter (purple dots in Fig. 1D).
161
162 *General features of bufferins and their BGCs.* Next, we characterized the various groups of
163 proteins in more detail, starting with *bona fide* bufferins. We collected thousands Buf1 and
164 Buf2 bufferins (at < 90% redundancy) from the NCBI nr Protein database, in various taxonomic
165 groups (Fig. 3A,B). Of note, Buf2 BGCs were reported in uncultured Antarctic soil bacteria
166 (39), suggesting their large environmental distribution.
167 Genes for sigma factors and anti-sigma NsrF proteins were found in Buf1 but not in Buf2
168 BGCs (Fig. 3D,E). In both subfamilies, sizeable proportions of precursors are coded by genes
169 not genetically associated with MNIO genes. We also identified BGCs with tandem precursor
170 genes, suggesting that some MNIO enzymes might have more than one substrate. Additional
171 conserved genes were found in extremely low proportions, confirming that the bufferin
172 precursor-, MNIO-, DUF2063- and DoxX-coding genes constitute the cores of the bufferin
173 BGCs.
174 Buf1 precursors range from 30 to 835 residues and harbor the four conserved Cys residues
175 present in the founding member (22) (Fig. 3G; Suppl. Fig. S1). It is likely however that the
176 smallest of these proteins are truncated. Buf2 precursors are between 51 and 367 residues long,
177 with six conserved Cys residues in 90% of cases and twelve Cys in the remainders. An

178 AlphaFold2 model of one of the latter indicates that it consists of duplicated 6-Cys-containing
179 domains (Fig. 3H; Suppl. Fig. S1). Approximately 16% Buf2 bufferins are predicted to harbor
180 lipoprotein signal peptides and might therefore be membrane-anchored after Sec-dependent
181 export. In both families we identified a few instances of large bufferin domain-containing
182 fusion proteins, including Buf1|Cytochrom_C|FGE-sulfatase (Pfam entries of the latter two
183 domains PF00034 and PF03781), MerC|Buf1 (Pfam entry of the former domain PF03203) and
184 MerC|Buf2 chimeras (Fig. 3J,K). The formylglycine-generating enzymes (FGE) are
185 cuproenzymes (40), consistent with copper-related functions. MerC-containing fusions are
186 found in BGCs related to mercury detoxification (41), indicating that some bufferins may
187 participate in stress responses to other metals.
188
189 *Other signal peptide-containing precursors.* Proteins of the third largest cluster of the putative
190 RiPP precursors node network (Fig. 1C) contain oxazolins (32). The newly generated hmm
191 profile for these proteins was called 'Buf_EGKCG/oxazolin' (Suppl. File S2), as they are
192 characterized by repeated Glu-Gly-Lys-Cys-Gly motifs. The majority of the approximately
193 2,000 proteins collected with this new signature are mostly found in *Pseudomonadota* (Fig.
194 3C). Analyses of their BGCs revealed that most are composed of the core RiPP-, MNIO-,
195 DUF2063 and DoxX-coding genes in various combinations (Fig. 3F). Small numbers of BGCs
196 harbor additional genes that code for proteins with the names oxidored_molybd (Pfam entry
197 PF00174) and Ferric reductase (Panther Database entry PTHR11972) domains, suggesting
198 metal-related or redox functions.
199 The Buf_EGKCG/oxazolin precursors are 31 to 506 residues long and are of two distinct
200 types. Proteins of the first type contain variable numbers of 17-residue repeats with two
201 EGKCG motifs each. An AlphaFold2-generated structural model of a mid-size protein showed
202 a β-helix fold (Fig. 3I; Suppl. Fig. S1), with the Cys residues putative targets of the MNIO

203 enzyme predicted to be oriented towards the inside of the helix, reminiscent of copper storage
204 proteins (42). The longest protein of the family, from *Pseudoalteromonas piscicida,* harbors 58
205 Cys (strain JCM_20779, protein acc. ATD05916.1). Proteins of the second type harbor two
206 Ca^{+2}-binding EF-hand_5 domains fused to the C terminus of a short Buf_EGKCG domain
207 devoid of 17-residue repeats and preceded by a predicted transmembrane helix (Fig. 3L).
208 AlphaFold2 predicted that the N-terminal Buf_EGKCG domains of the second type of proteins
209 are unstructured.

210 We also built an hmm profile called 'CxxxxC' for the fourth largest cluster of precursor
211 proteins (black dots in Fig. 1C), which notably contain signal peptides and conserved CxxxxC
212 motifs (Suppl. File S3). The newly built signature collected 76 proteins at < 90% sequence
213 identity, all from *Pseudomonadati*. In addition to the core RiPP-, MNIO, DUF2063- and DoxX-
214 coding genes, 54% of these BGCs contain genes for proteins named Lactamase_B_2 (Pfam
215 entry PF12706), which are metalloproteins.

216 Last, we searched for precursors in the BGCs of the approximately 4000 MNIO
217 enzymes of the SSN mega-cluster in Fig. 1B that have no Buf1, Buf2, Buf_EGKCG/oxazolin
218 or CxxxxC precursors encoded next to them. We identified hundreds of small annotated ORFs
219 immediately preceding the MNIO genes and coding for peptides containing between 1 and 18
220 Cys with predicted signal peptides (Suppl. Table S2). These ORFs probably represent new
221 precursors and might correspond to several small sequence clusters of Fig. 1C. For the
222 remaining BGCs, we performed RODEO analyses to detect other unannotated ORFs, which we
223 subjected to SSN analyses. However, these analyses did not group them into conserved
224 families. Which of these might be new precursors is currently difficult to establish.

225

226 *TIGR04222 proteins.* Using the TIGR04222 signature, > 3500 proteins were collected at < 90%
227 sequence identity from the nr NCBI database, most of them from Actinomycetota (Fig. 4A).

228 Analyses of the genetic environments of the TIGR04222-coding genes revealed a sizeable
229 proportion of BGCs devoid of MNIO genes (Fig. 4B). Furthermore, in more than 10% cases
230 TIGR04222 genes follow MNIO genes, in contrast with bufferin BGCs. No DUF2063-coding
231 genes were found in the TIGR04222 BGCs, some of which harbor other conserved genes of
232 unknown function.

233 The sizes of TIGR04222 proteins range from 70 to 643 residues (Fig. 4C). Alphafold2
234 models of selected proteins showed that they are composed of several transmembrane segments
235 and large periplasmic domain(s), with unstructured, C-terminal Ser- and Gly-rich regions
236 interspersed with Cys residues (Fig. 4D,E; Suppl. Fig. S1). Interestingly, these low-complexity
237 regions are predicted to be exported to the periplasm, like the bufferins. Together with their
238 frequent genetic association with MNIO enzymes, these features are consistent with
239 TIGR04222 proteins also being MNIO substrates. Nevertheless, more than a quarter of them
240 are devoid of C-proximal Cys residues, and many are not genetically associated with MNIO
241 enzymes. Thus, only some TIGR04222 proteins might represent new MNIO-modified
242 precursors.

243

244 *Other types of putative precursors.* The SSN analysis of the MNIO enzymes has revealed that
245 bufferins and related proteins are the majority substrates. Nevertheless, more diversity can be
246 expected from the small, uncharacterized sequence clusters. To expand our range of new
247 potential MNIO-modified precursors, we focused both on minor groups of MNIO enzymes,
248 corresponding to the small sequence clusters labeled 1 to 21 in Fig. 1B, and on the non-
249 characterized mid-size sequence clusters of putative precursors labelled I to IV in Fig. 1D.

250 Analyses of the BGCs that correspond to subclusters numbered 1 to 3 of the MNIO node
251 network identified new families of precursors with Cys-containing motifs (Fig. 5A). Those

genetically associated with the MNIO enzymes of the subcluster #3 harbor predicted Sec signal peptides, unlike those of clusters #1 and #2. Their BGCs are similar to those of bufferins.

We next analyzed the BGCs of MNIO enzymes found in clusters 4 to 21 (Fig. 5B). These BGCs are extremely diverse, even within each sequence cluster. Other than the already described precursors of chryseobasins, methanobactins, aminopyruvatides and pearlins, (23, 24, 26-28, 30), we identified potential signal peptide-less RiPP precursors in the BGCs of two other small clusters. MNIO enzymes of cluster #7 are genetically associated with putative precursors harboring a CxxC motif (Fig. 5B; Suppl. Fig. S1). Most BGCs corresponding to MNIO enzymes of cluster #9 also code for potential RiPP precursors devoid of Cys residues. Proteins with putative functions in RiPP biosynthesis were encoded in many of these BGCs (Pfam names listed in red in Fig. 5B). Thus, based on these analyses, many minor types of MNIO-modified RiPP precursors remain to be identified, but this is not straightforward as already acknowledged (43). In the absence of small ORFs it is hazardous to assign a precursor status to unknown ORFs with no distinctive features.

We also analyzed four sequence clusters of putative precursors, labeled I to IV in Fig. 1D. With 7 to 21 Cys residues, proteins of cluster #I plausibly represent a new group of precursors (Table 2). Similarly, those of cluster #IV harbor a CxxCC motif and might also be a new group of RiPP precursors. In contrast, no precursor genes were identified in the BGCs corresponding to clusters #II and #III. The genes that precede the MNIO enzyme genes in these latter two clusters code for peptidyl-tRNA hydrolases (Pfam entry PF01195), and for putative metalloproteins with a HExxH motif, respectively. Both types of proteins are thus more likely to be enzymes than precursors (28, 29, 44). In the BGCs encoding cluster #III proteins, we found genes coding for MNIO|Nitroreductase fusion proteins (Pfam entry of the latter: PF00881), suggesting the possibility of new chemical modifications.

277 *Identification of MNIO partner proteins.* Partners of MNIO enzymes include proteins that
278 possess DUF2063, Anthrone_oxyd or methbact_MbnC Pfam signatures. The former proteins
279 are by far the most frequent partner proteins of MNIO enzymes (approx. 60%), and we propose
280 to name them PME-type 1, for Partner protein of MNIO Enzyme, type 1. To identify other types
281 of partner proteins, we analyzed the proteins coded by genes that follow the MNIO genes, based
282 on observations that this is the most frequent gene order (Fig. 2B). Among the proteins thus
283 collected, 72% possess the DUF2063 signature, others have Anthrone_oxyd, methbact_MbnC,
284 DoxX or MNIO signatures, and 1700 others have no Pfam signature. SSN analyses of these
285 latter sequences revealed that they could be sorted into several sequence clusters (Fig. 6A).
286 To detect low-level similarity between these uncharacterized proteins and known domains,
287 we performed HHpred analyses to identify structural homologs. For many of them the *Neisseria*
288 *gonorrheae* protein NGO1945 (pdb# 3dee) and TglI (pdb#8hi7), the partner of the MNIO
289 enzyme TglH (28) were among the best hits. NGO1945 contains both a DUF2063 domain and
290 an RRE-type domain and is therefore a typical MNIO partner protein, although it was
291 previously predicted to be a transcription factor (45) (Fig. 6A, inset). Alphafold2 models of a
292 few of these uncharacterized proteins revealed that they harbor N-terminal α helix-rich domains
293 followed in some cases by RRE-fold type domains or truncated versions thereof (Fig. 6A,
294 models a to d). Therefore, many BGCs without *bona fide* DUF2063 genes encode DUF2063-
295 like proteins, missed by the DUF2063 signature because of a lack of sequence similarity. We
296 propose to call them PME-type 2 proteins. Note that PME-type 2 proteins appear to encompass
297 a variety of subtypes, and therefore future studies will be necessary to refine their classification.
298 The SSN analysis presented in Fig. 6A also indicated that yet additional types of partner
299 proteins likely exist. We analyzed the largest two sequence clusters of proteins not included in
300 PME-type 2 proteins (sequence clusters in gray in Fig. 6A). One of them contains small, all α-
301 helical proteins with no structural homology with other proteins (Fig. 6A, model e). They might

302 represent a minor type of MNIO partners, as most of the corresponding BGCs also encode
303 putative precursors, notably with C-terminal CxxC motifs. The proteins of the other large non-
304 PME type 2 cluster are coded by genes inserted between the MNIO and the *bona fide* PME-
305 type 1 genes in otherwise classical *buf1* operons. The function of these proteins in these RiPP
306 systems is unknown, but they are probably enzymes that introduce additional modifications.

307 We reanalyzed the products of unknown genes in the TIGR04222 BGCs. HHpred analyses
308 predicted that at least 10% of TIGR04222 BGCs code for PME-type 2 proteins. Similarly,
309 BGCs encoding MNIO enzymes from several small sequence clusters also encode PME-type 2
310 proteins, including those of pearlins and aminopyruvatides (circled in red in Fig. 5B).

311 We next determined the length distributions of the major two types of putative MNIO
312 partner proteins. PME-type 1 proteins have a narrow size distribution corresponding to full-
313 length DUF2063-RRE chimeras (Fig. 6B). In contrast, PME-type 2 proteins are very
314 heterogeneous in length. The group of proteins with similar lengths as the PME-type 1, around
315 260 residues, most likely harbor two full-length domains. However, a large proportion of PME-
316 type 2 proteins are too short to comprise both domains and probably consist of a DUF2063-
317 related domain alone or followed by a short putative RRE-type truncate or variant.

318

319 *Specificity of MNIO enzymes for RiPP precursors and partner proteins.* The discovery of other
320 types of partner proteins prompted us to investigate specific associations between the subgroups
321 of MNIO enzymes, the families of RiPP precursors, and the types of partner proteins. To
322 facilitate visualization of these relationships, we colored the nodes of the large sequence cluster
323 of MNIO enzymes of Fig. 1B according to the types of precursors and partner proteins encoded
324 in the same BGCs. This confirmed that specific areas of this large cluster of MNIO enzymes
325 correspond to specific precursor families (Fig. 7A). In addition, coloring of the nodes of MNIO
326 enzymes as a function of their partner proteins showed that those genetically associated with

327 Sec signal peptide-containing precursors (i.e., the Buf1, Buf2, Buf_EGKCG/oxazolins and
328 CxxxxC families) have authentic PME-type 1 proteins (Fig. 7B). In contrast, those associated
329 with signal peptide-less TIGR04222 proteins appear rather to have PME-type 2 partners.
330 This correlation was confirmed for three small MNIO sequence clusters numbered 1 to
331 3. Thus, the precursors associated with the MNIO enzymes of clusters #1 and #2 are devoid of
332 predicted Sec signal peptides, and the partners of their MNIO enzymes are PME-type 2 proteins
333 (Fig. 7A,B). In contrast, the precursors associated with MNIO enzymes of cluster #3 harbor
334 predicted Sec signal peptides, and their MNIO enzymes have *bona fide* PME-type 1 proteins.
335 Thus, there is a correlation between the presence of a signal peptide in the precursor and the
336 type of partner protein.
337
338 *Co-occurrence analyses.* Finally, we mined > 40,000 fully assembled bacterial genomes in the
339 NCBI database to determine the maximal numbers of occurrences of bufferin or bufferin-like
340 BGCs in a single genome. The most extreme case is that of *Legionella pneumophila subsp.*
341 *fraseri*, with five Buf1 (DUF2282), two Buf2 (BUF_6/12Cys)-, one BUF_EGKCG-, seven
342 MNIO-, six DUF2063- and three DoxX-coding genes (Suppl. Fig. S2). The presence of bufferin
343 precursor genes not associated with MNIO genes suggests that these precursors might be
344 modified by MNIO enzymes encoded at a different locus or that they are not modified.
345 However, we do not know whether all these BGCs are functional.
346 Our genomic analysis further revealed that very few Buf1, Buf2 or Buf_EGKCG/oxazolin
347 precursor genes are present without MNIO and PME-type 1 genes in the same genome (Table
348 3). In contrast, a third of genomes harboring TIGR04222 genes are devoid of MNIO genes,
349 revealing a looser association between these two families and further supporting the idea that
350 TIGR04222 proteins form several subfamilies only some of which are likely to be modified by
351 MNIO enzymes.

352

353 **Discussion**

354 We explored the substrate diversity of MNIO enzymes and described several groups of
355 potential precursors in various eubacterial taxonomic groups. We revealed that the BGCs of the
356 largest groups of RiPP precursors are composed of a small number of core genes. This limited
357 gene set and the overwhelming conservation of Cys-containing motifs indicate that these
358 residues are the most likely targets of MNIO modifications, and that secondary PTMs are rare
359 for the major precursor families. In contrast, the BGCs corresponding to the small clusters of
360 MNIO enzymes are larger and more diverse, with many comprising genes for putative tailoring
361 enzymes. They are likely to code for new MNIO-catalyzed chemistry, including on non-Cys
362 residues, as well as for various secondary modifications. In fact, it is in some of those small
363 MNIO enzyme groups that precursors were found to be modified on Asp or Asn residues or to
364 carry additional PTMs catalyzed by other enzymes (23, 30). In many cases however, peptide
365 precursors are difficult to identify, suggesting that some of these MNIO enzymes may modify
366 *bona fide* proteins, which is likely the case for subsets of TIGR04222 proteins.

367 We have revealed that the Buf1 bufferin of *C. vibrioides* is involved in copper homeostasis
368 and harbors Cys-derived thiooxazole groups mediating the chelation of Cu(I) and Cu(II),
369 thereby enhancing bacterial growth in conditions of copper excess (22). Buf1 likely promotes
370 survival of environmental bacteria subjected to copper stress in soils or water polluted by
371 metals, notably from anthropic sources (46), or when they are engulfed by amoeba (47).
372 Similarly, the use of copper as an antimicrobial agent in innate immunity (48) might account
373 for the presence of bufferins in bacterial pathogens of mammals. Although we have not
374 identified the chemical modifications carried by Buf2 bufferins, preliminary characterization
375 of that of *C. vibrioides* has shown that it has a similar function as Buf1, its PTMs result in the
376 same mass loss per Cys residue, and its UV/vis spectrum shows a similar 305-nm absorbance

377 maximum (22). Furthermore, the structural models of Buf and Buf2 bufferins are related, and
378 their associated MNIO enzymes belong to the same large sequence cluster, suggesting that they
379 are iso-functional. These elements argue in favor of similar PTMs and functions for both
380 bufferin families.

381 The PTMs and the function of a member of the MNIO-modified Buf_EGKCG/oxazolin
382 family were recently reported (32). Similar to Buf1, it binds Cu(I) but appears to do so using
383 Cys-derived oxazolone thioamide groups, similar to methanobactins. These proteins constitute
384 the third largest sequence cluster in the SSN node network of RiPP precursors. Interestingly,
385 the MNIO enzymes associated with these precursors are closely related to those associated with
386 bufferin precursors, and their partners are PME-type 1 proteins, like for bufferins. Whether the
387 MNIO enzymes that belong to the large sequence cluster are all iso-functional remains to be
388 seen. Nevertheless, the discrepancy between the PTMs of bufferins and Buf_EGKCG/oxazolins
389 is puzzling and warrants further investigations.

390 In contrast, TIGR04222 membrane proteins are unrelated putative precursors. Those that
391 harbor C-proximal Cys-rich segments predicted to be exported to the periplasm may be
392 modified by MNIO enzymes. However, one cannot speculate on the nature of their PTMs, as
393 the MNIO enzymes and partner proteins genetically associated with TIGR04222 proteins are
394 somewhat distant from those of bufferins. Similarly, the functions of these proteins in the
395 producing organisms remain to be investigated. Some TIGR04222 proteins are not even
396 associated with MNIOs and/or have no Cys residues. Given their diversity, TIGR04222
397 proteins are likely to have a variety of functions and to form several distinct families, only some
398 of which may be post-translationally modified by MNIO enzymes.

399 Methanobactins are involved in copper acquisition and bufferins and
400 Buf_EGKCG/oxazolins in the protection against copper. However, the diversity of the
401 precursors and of their BGCs indicates that many putative RiPPs identified in this study likely

402 play roles beyond adaptation to excess copper. This was already suggested by the denomination
403 '*gig*' (gold-induced genes) operons, which are actually bufferin BGCs, although their role in
404 the protection against gold remains to be shown (49, 50). The identification of fusion proteins
405 between bufferins and metal-binding domains such as MerC further supports a role in the
406 homeostasis of various other metals. As for Buf_EGKC|EF-hand fusion proteins, in addition to
407 a role in copper homeostasis they might contribute to Ca^{2+} homeostasis or signaling, as
408 exemplified by EfhP in *Pseudomonas aeruginosa* (51, 52). Furthermore, the presence of *doxX*
409 genes in many bufferin BGCs suggests protection from other stresses commonly encountered
410 by bacteria. A DoxX homolog in a different genetic context was shown to protect bacteria
411 against oxidative or sulfur stress (53). As for most other types of MNIO-modified RiPPs, their
412 functions remain unexplored thus far.

413 The presence of signal peptides in a large proportion of MNIO substrates is an original
414 feature among bacterial RiPPs. As the signal peptide directs the precursor to the Sec machinery,
415 there must be a mechanism to avoid premature export before the installation of the PTMs. How
416 recognition between the precursor and the MNIO-partner protein complex operates has been
417 described for methanobactin (24) and pearlin precursors (54), neither of which have signal
418 peptides. In this work, we found out that MNIO enzymes whose substrates are precursors with
419 signal peptides have PME-type 1 proteins, unlike those associated with signal peptide-less
420 precursors. It is tempting to speculate that interactions between the precursor and the PME-type
421 1 protein might mask the signal peptide until after installation of the PTMs, thereby preventing
422 an early interaction with the Sec machinery. We have preliminary experimental results in our
423 Buf1 model system supporting the idea that the signal peptide is indeed part of the recognition
424 sequence by the BufB-BufC complex. In contrast, many BGCs harbouring TIGR04222 and
425 MNIO genes appear not to harbor partner protein genes. Whether and how the corresponding
426 precursors are modified remains to be established. Future studies will also reveal which MNIO

427 enzymes can work without partner proteins. Of note, the second MNIO enzyme of
428 methanobactin-like BGCs in *Vibrio*, MovX, appears to do so (27). Nevertheless, MovX
429 comprises a winged helix-turn-helix domain in addition to the catalytic domain, which might
430 fulfill a partner function.

431 Bufferin or Buf_EGKCG/oxazolin precursor genes are rarely present in genomes without
432 MNIO and partner protein genes. Presumably, in these rare occurrences the peptides have
433 activities that do not need PTMs, or they are no longer expressed or functional. We also found
434 genomes with several bufferin-like precursor genes but a single MNIO and PME-type 1 gene
435 pair. It is possible that the lifestyles of these bacteria have fostered RiPP expansion, and that a
436 single MNIO-PME type 1 complex can modify several related precursors. Alternatively, only
437 some of the precursors are modified, and the unmodified ones serve a different purpose. Further
438 studies are needed to address these issues.

439 Our study suggests that MNIO enzymes modify not only peptides but also proteins that
440 respond to the same signatures as smaller precursors but contain repeated motif or additional
441 domains. In particular, many Cys residues are found in very large precursors of the
442 Buf_EGKCG/oxazolin family. If these residues are all modified before export, there must be a
443 mechanism to thread the precursor through the MNIO active site for its sequential processing,
444 similar to the iterative mechanisms reported for several other families of RiPP precursors (17,
445 55, 56). The identification of BGCs encoding long precursors followed by two MNIO-coding
446 genes indicates that more than one enzyme might be needed to extensively modify such long
447 proteins.

448

449 **Materials and Methods**

450 *Database searches.* The non-redundant NCBI protein database was searched with various
451 hidden Markov models (Pfam https://www.ebi.ac.uk/interpro/ (57), TIGRFAMs

452 https://www.jcvi.org/research/tigrfams (58) and home-built HMM models), using the
453 hmmsearch program of the HMMER3 suite (59). Each set of results were clustered by CD-HIT
454 (60) with default parameters to remove redundant sequences sharing at least 90% sequence
455 identity, to avoid overrepresentation of well-studied bacterial genera. The size, the domain
456 composition, and the taxonomy were determined for each protein.
457
458 *Sequence similarity network analysis (SSN).* Sequence similarity networks were generated
459 using the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST)
460 (https://efi.igb.illinois.edu/efi-est/) (61). We followed the guidelines of the EFI-EST site and
461 chose rather stringent parameters (alignment score corresponding to 40% pairwise identity),
462 to better separate the groups of interest. The selected sequence alignment score threshold
463 (AST) is given in each figure. Networks were visualized by Cytoscape (62) by using the
464 organic layout. The ID numbers of all entries found in the SSN node networks are provided in
465 Suppl. Table S3.
466
467 *Local database construction and characterization of genetic environments.* A local DNA
468 database was constituted by downloading all available bacterial genomic sequences for both
469 Eubacteria and Archeae from the NCBI RefSeq, GenBank and WGS databases. A Python script
470 was written to extract the genetic environments of all MNIO-coding genes by retrieving the
471 sequences of the five genes on either side of the gene of interest in a Rodeo-type approach. The
472 resulting table was used to collect all genes coding for proteins of interest. To define the RiPP
473 BGCs, we imposed the following criteria. All genes encoding proteins known to be relevant to
474 MNIO BCGs (with Pfam short names DUF2282, MbnB_TglH_ChrH (MNIO), DUF2063,
475 DoxX, Sigma70_r4_2 (sigma factor), Sigma70_r2 (sigma factor), NrsF, or NCBIfam name
476 methanobac_OB3b (= MbnA), or corresponding to the new signatures Buf_EGKCG/oxazolin,

477 Buf_6/12Cys, or CxxxxC) in the vicinity of the MNIO genes were considered part of the BGC,
478 irrespective of the intergenic distances. For unrelated genes, we set up a stringent intergenic
479 distance criterion (maximum 10 bp from any relevant gene) for inclusion to ensure that these
480 new genes are in operons with at least one known gene. Although somewhat arbitrary, this
481 choice proved to provide the most information while avoiding too much background noise. For
482 the smaller sequence clusters of MNIO enzymes, RODEO analyses were performed, and all
483 ORFs identified were considered as part of the BCGs irrespective of the intergenic distances.
484 The numbers of occurrences of each type of operonic structures were computed. Prediction of
485 signal-peptides and their cleavage sites was performed using SignalP V6 (63).
486
487 *Generation of hmm profiles.* We generated hmm profiles for the Buf_6/12Cys,
488 Buf_EGKCG/oxazolin and CxxxxC families according to the procedure described in (64).
489 Briefly, a few dozen proteins (33, 36 and 41 for Buf2, Buf_EGKCG and CxxxxC proteins,
490 respectively) were picked from each cluster of interest of the Representative Node Network
491 70% displayed with Cytoscape. We chose nodes well distributed across each cluster. Only
492 nodes with a 'number of IDs in Rep Node' equal to 1 were kept. Outliers with respect to length
493 were also removed. For each family we first determined the most frequent protein size in the
494 cluster, and we selected sequences from this major subset to generate the seed alignment. The
495 sequences were aligned with MAFFT (L-INS-i), an iterative refinement method incorporating
496 local pairwise alignment information. The non-conserved regions (mostly the N- and C-termini)
497 were edited manually in Jalview, and a hmm profile was generated with this seed alignment
498 using hmmbuild (59). We then performed new SSN analyses with all proteins collected using
499 a given hmm profile to ensure that only *bona fide* members were retained. Outliers found in
500 small, isolated SSN clusters were discarded. The hmm profile files are provided in
501 Supplementary materials (Files S1 to S3).

502

503 *Characterization of unknown proteins.* To gain insight into proteins of interest devoid of Pfam
504 or TIGRfam domain signatures, we picked several proteins in each cluster of the representative
505 node network and first looked for Interpro signatures. If no Interpro signature was identified,
506 we searched for related proteins using HHpred (65) and selected the best hits. If necessary, we
507 generated AlphaFold2 models of selected members and used DALI (66) to find related
508 structures in the RCSB Protein Data Bank (67).

509

510 *Analysis of assembled bacterial genomes.* The proteomes of the fully assembled bacterial
511 genomes were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/assembly). The
512 bacterial proteomes were searched with the hmmscan program of the HMMER3 suite (59) using
513 the hmm profile signatures DUF2282, DUF692, DUF2063, DoxX, EF-hand_5, Buf_6/12_Cys,
514 Buf_EGKCG, and TIGR04222 using home-made scripts. The domains of all the corresponding
515 proteins were determined, and the occurrences of each type of domain of interest were
516 computed (Table S4).

517

518 **Data summary**

519 A local DNA database was constituted by downloading all available bacterial genomic
520 sequences for both Eubacteria and Archeae from the NCBI RefSeq, GenBank and WGS
521 databases. Protein sequences were collected from the non-redundant NCBI protein database.
522 We also mined the NCBI assembled genomes database. Accession numbers of all the protein
523 sequences used in this work and all genome identification numbers are provided in
524 Supplementary materials in the form of Excel tables.

525

526 **Funding statement**

This project was funded by the ANR grant CuRiPP (ANR-22-CE44-0001-02) to FJD. L. Leprevost and S. Jünger were supported by PhD fellowships of Lille University and of the French Ministry of Education (ED 227 MNHN-SU), respectively. S. Dubiley was supported as a Visiting Scientist by a fellowship of the Collège de France.

Conflict of interest statement

The authors declare that they have no conflict of interest.

References

1. Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. Chem Rev. 2006;106(8):3468-96.
2. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep. 2013;30(1):108-60.
3. Kloosterman AM, Cimermancic P, Elsayed SS, Du C, Hadjithomas M, Donia MS, et al. Integration of machine learning and pan-genomics expands the biosynthetic landscape of RiPP natural products. PLoS Biol. 2021;18:e3001026.
4. Li Y, Rebuffat S. The manifold roles of microbial ribosomal peptide-based natural products in physiology and ecology. J Biol Chem. 2020;295(1):34-54.
5. Montalban-Lopez M, Scott TA, Ramesh S, Rahman IR, van Heel AJ, Viel JH, et al. New developments in RiPP discovery, enzymology and engineering. Nat Prod Rep. 2021; 38(1):130-239.
6. Luo S, Dong SH. Recent Advances in the Discovery and Biosynthetic Study of Eukaryotic RiPP Natural Products. Molecules. 2019;24(8).
7. Oman TJ, van der Donk W. Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat Chem Biol. 2010;6(1):9-18.
8. Schwalen CJ, Hudson GA, Kosol S, Mahanta N, Challis GL, Mitchell DA. In Vitro Biosynthetic Studies of Bottromycin Expand the Enzymatic Capabilities of the YcaO Superfamily J Am Chem Soc. 2017;139(50):18154-7.
9. Burkhart BJ, Hudson GA, Dunbar KL, Mitchell DA. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. Nat Chem Biol. 2015;11(8):564-70.
10. Ortega MA, Hao Y, Zhang Q, Walker MC, van der Donk WA, Nair SK. Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. Nature. 2015; 517(7535):509-12.
11. Sumida T, Dubiley S, Wilcox B, Severinov K, Tagami S. Structural Basis of Leader Peptide Recognition in Lasso Peptide Biosynthesis Pathway. ACS Chem Biol. 2019; 14(7):1619-27.
12. Chekan JR, Ongpipattanakul C, Nair SK. Steric complementarity directs sequence promiscuous leader binding in RiPP biosynthesis. Proc Natl Acad Sci U S A. 2019; 116(48):24049-55.

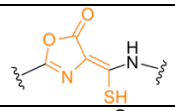
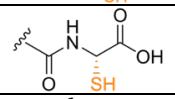
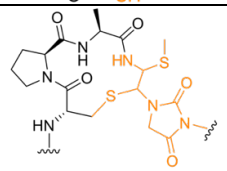
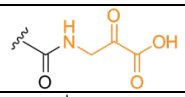
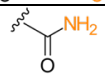
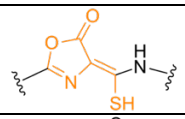
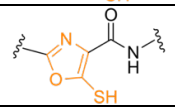
13. Dong SH, Kulikovsky A, Zukher I, Estrada P, Dubiley S, Severinov K, et al. Biosynthesis of the RiPP trojan horse nucleotide antibiotic microcin C is directed by the N-formyl of the peptide precursor. Chem Sci. 2019;10(8):2391-5.
14. Hernandez Garcia A, Nair S. Structure and Function of a Class III Metal-Independent Lanthipeptide Synthetase. ACS Central Sci. 2023;9(10):1944-56.
15. Li K, Condurso HL, Li G, Ding Y, Bruner SD. Structural basis for precursor protein-directed ribosomal peptide macrocyclization. Nat Chem Biol. 2016;12(11):973-9.
16. Chigumba DN, Mydy LS, de Waal F, Li W, Shafiq K, Wotring JW, et al. Discovery and biosynthesis of cyclic plant peptides via autocatalytic cyclases. Nat Chem Biol. 2022;18:18-28.
17. Miller FS, Crone KK, Jensen MR, Shaw S, Harcombe WR, Elias MH, et al. Conformational rearrangements enable iterative backbone N-methylation in RiPP biosynthesis. Nat Commun. 2021;12:5355.
18. Severi E, Thomas GH. Antibiotic export: transporters involved in the final step of natural product production. Microbiology (Reading). 2019;165(8):805-18.
19. Kenney GE, Rosenzweig AC. Methanobactins: Maintaining copper homeostasis in methanotrophs and beyond. J Biol Chem. 2018;293(13):4606-15.
20. Duquesne S, Destoumieux-Garzon D, Peduzzi J, Rebuffat S. Microcins, gene-encoded antibacterial peptides from enterobacteria. Nat Prod Rep. 2007;24(4):708-34.
21. Chen JY, van der Donk WA. Multinuclear non-heme iron dependent oxidative enzymes (MNIOs) involved in unusual peptide modifications. Curr Opin Chem Biol. 2024;80:102467.
22. Leprevost L, Jünger S, Lippens G, Guillaume C, Sicoli G, Oliveira L, et al. A widespread family of ribosomal peptides metallophores involved in bacterial adaptation to metal stress. submitted. 2024.
23. Ayikpoe RS, Zhu L, Chen JY, Ting CP, van der Donk WA. Macrocyclization and Backbone Rearrangement During RiPP Biosynthesis by a SAM-Dependent Domain-of-Unknown-Function 692. ACS Cent Sci. 2023;9(5):1008-18.
24. Dou C, Long Z, Li S, Zhou D, Jin Y, Zhang L, et al. Crystal structure and catalytic mechanism of the MbnBC holoenzyme required for methanobactin biosynthesis. Cell Res. 2022;32(3):302-14.
25. Kenney GE, Rosenzweig AC. Genome mining for methanobactins. BMC Biol. 2013;11:17.
26. Kenney GE, Dassama LMK, Pandelia ME, Gizzi AS, Martinie RJ, Gao P, et al. The biosynthesis of methanobactin. Science. 2018;359(6382):1411-6.
27. Chioti VT, Clark KA, Ganley JG, Han EJ, Seyedsayamdost MR. N-Calpha Bond Cleavage Catalyzed by a Multinuclear Iron Oxygenase from a Divergent Methanobactin-like RiPP Gene Cluster. J Am Chem Soc. 2024;146(11):73132-7323.
28. Ting CP, Funk MA, Halaby SL, Zhang Z, Gonen T, van der Donk WA. Use of a scaffold peptide in the biosynthesis of amino acid-derived natural products. Science. 2019;365(6450):280-4.
29. Yu Y, van der Donk WA. Biosynthesis of a 3-thia-alpha amino acids on a carrier peptide. Proc Natl Acad Sci USA. 2022(119):e22052851119.
30. Nguyen DT, Zhu L, Gray DL, Woods TJ, Padhi C, Flatt KM, et al. Biosynthesis of macrocyclic peptides with C-terminal beta-amino-alpha-keto acid groups by three different metalloenzymes. Acs Central Science. 2024;10:2022-32.
31. Andrei A, Ozturk Y, Khalfaoui-Hassani B, Rauch J, Marckmann D, Trasnea PI, et al. Cu Homeostasis in Bacteria: The Ins and Outs. Membranes (Basel). 2020;10(9).
32. Manley OM, Shriver TJ, Xu T, Melendrez IA, Palacios P, Robson SA, et al. A multi-iron enzyme installs copper-binding oxazolone/thioamide pairs on a nontypeable

Haemophilus influenzae virulence factor. Proc Natl Acad Sci U S A. 2024; 121(28):e2408092121.
33. Herranz C, Driessen AJM. Sec-Mediated Secretion of Bacteriocin Enterocin P by *Lactococcus lactis*. Appl Env Microbiol. 2005;71(4):1958-63.
34. Widdick DA, Dodd HM, Barraille P, White J, Stein TH, Chater KF, et al. Cloning and engineering of the cinnamycin biosynthetic gene cluster from *Streptomyces cinnamoneus cinnamoneus* DSM 40005. Proc Natl Acad Sci U S A. 2003;100(7):4316-21.
35. Clark KA, Seyedsayamdost MR. Bioinformatic Atlas of Radical SAM Enzyme-Modified RiPP Natural Products Reveals an Isoleucine-Tryptophan Crosslink. J Am Chem Soc. 2022;144(39):17876-88.
36. Zallot R, Oberg N, Gerlt JA. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. Biochemistry. 2019;58(41):4169-82.
37. Tietz JI, Schwalen CJ, Patel PS, Maxson T, Blair PM, Tai HC, et al. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. Nat Chem Biol. 2017; 13(5):470-8.
38. Rivera-Millot A, Slupek S, Chatagnon J, Roy G, Saliou JM, Billon G, et al. Streamlined copper defenses make *Bordetella pertussis* reliant on custom-made operon. Commun Biol. 2021;4(1):46.
39. Waschulin V, Borsetto C, James R, Newsham KK, Donadio S, Corre C, et al. Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. ISME J. 2022;16(1):101-11.
40. Knop M, Dang TQ, Jeschke G, Seebeck FP. Copper is a Cofactor of the Formylglycine-Generating Enzyme. ChemBioChem. 2017;18:161-5.
41. Sone Y, Nakamura R, Pan-Hou H, Itoh T, Kiyono M. Role of MerC, MerE, MerF, MerT, and/or MerP in resistance to mercurials and the transport of mercurials in *Escherichia coli*. Biol Pharm Bull. 2013;36(11):1835-41.
42. Straw ML, Chaplin AK, Hough MA, Paps J, Bavro VN, Wilson MT, et al. A cytosolic copper storage protein provides a second level of copper tolerance in *Streptomyces lividans*. Metallomics. 2018;10(1):180-93.
43. Yu Y, van der Donk WA. Genome mining for natural products made by multinuclear iron-dependent oxidation enzymes (MNIOs). Meth Enzymol. 2025; in press.
44. Morishita Y, Ma S, De La Mora E, Li H, Chen H, Ji X, et al. Fused radical SAM and αKG-HExxH domain proteins contain a distinct structural fold and catalyse cyclophane formation and β-hydroxylation. Nat Chem. 2024;16:1882-93.
45. Das D, Grishin NV, Kumar A, Carlton D, Bakolitsa C, Miller MD, et al. The structure of the first representative of Pfam family PF09836 reveals a two-domain organization and suggests involvement in transcriptional regulation. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2010;66(Pt 10):1174-81.
46. Staehlin BM, Gibbons JG, Rokas A, O'Halloran TV, Slot JC. Evolution of a Heavy Metal Homeostasis/Resistance Island Reflects Increasing Copper Stress in Enterobacteria. Genome Biol Evol. 2016;8(3):811-26.
47. Hao X, Luthje F, Ronn R, German NA, Li X, Huang F, et al. A role for copper in protozoan grazing - two billion years selecting for bacterial copper resistance. Mol Microbiol. 2016;102(4):628-41.
48. Sheldon JR, Skaar EP. Metals as phagocyte antimicrobial effectors. Curr Opin Immunol. 2019;60:1-9.
49. Hirth N, Gerlach MS, Wiesemann N, Herzberg M, Grosse C, Nies DH. Full Copper Resistance in *Cupriavidus metallidurans* Requires the Interplay of Many Resistance Systems. Appl Environ Microbiol. 2023;89(6):e0056723.

50. Wiesemann N, Mohr J, Grosse C, Herzberg M, Hause G, Reith F, et al. Influence of copper resistance determinants on gold transformation by *Cupriavidus metallidurans* strain CH34. J Bacteriol. 2013;195(10):2298-308.
51. Kayastha BB, Kubo A, Burch-Konda J, Dohmen RL, McCoy JL, Rogers RR, et al. EF-hand protein, EfhP, specifically binds Ca(2+) and mediates Ca(2+) regulation of virulence in a human pathogen *Pseudomonas aeruginosa*. Sci Rep. 2022;12(1):8791.
52. Sarkisova SA, Lotlikar SR, Guragain M, Kubat R, Cloud J, Franklin MJ, et al. A *Pseudomonas aeruginosa* EF_Hand protein EfhP (PA4107), modulates stress responses and virulence at high calcium concentrations. PLoS One. 2014;9(6):e98985.
53. Nambi S, Long JE, Mishra BB, Baker R, Murphy KC, Olive AJ, et al. The Oxidative Stress Network of *Mycobacterium tuberculosis* Reveals Coordination between Radical Detoxification Systems. Cell Host Microbe. 2015;17(6):829-37.
54. Zheng Y, Xu X, Fu X, Zhou X, Dou C, Yu Y, et al. Structures of the holoenzyme TglHI required for 3-thiaglutamate biosynthesis. Structure. 2023;31(10):1220-32 e5.
55. Ghilarov D, Stevenson CEM, Travin DY, Piskunova J, Serebryakova M, Maxwell A, et al. Architecture of Microcin B17 Synthetase: An Octameric Protein Complex Converting a Ribosomally Synthesized Peptide into a DNA Gyrase Poison. Mol Cell. 2019;73(4):749-62 e5.
56. Habibi Y, Weerasinghe NW, Uggowitzer KA, Thibodeaux CJ. Partially Modified Peptide Intermediates in Lanthipeptide Biosynthesis Alter the Structure and Dynamics of a Lanthipeptide Synthetase. J Am Chem Soc. 2022;144(23):10230-40.
57. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412-D9.
58. Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. Nucleic Acids Res. 2003;31(1):371-3.
59. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011; 7(10):e1002195.
60. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28(23):3150-2.
61. Oberg N, Zallot R, Gerlt JA. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. J Mol Biol. 2023; 435(14):168018.
62. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.
63. Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023-5.
64. Shelton KE, Mitchell DA. Bioinformatic prediction and experimental validation of RiPP recognition elements. Methods Enzymol. 2023;679:191-233.
65. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005;33(Web Server issue):W244-8.
66. Holm L, Laiho A, Toronen P, Salgado M. DALI shines a light on remote homologs: One hundred discoveries. Protein Sci. 2023;32(1):e4519.
67. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000;28(1):235-42.

716 **Table 1. Known types of MNIO-modified RiPPs**

717

RiPP name	RiPP Precursor	MNIO & partner	Target residue	PTM	PTM chemical structure	Ref
Methanobactin	MbnA	MbnBC	Cys	Oxazolone thioamide		26
Pearlin	TglA	TglHI	C-terminal Cys	3 thia-Glu		28
Chryseobasin	ChrA	ChrHI*	Cys	Peptide macrocycle, Imidazoline-2,4 dione, thiomethyl		23
Aminopyruvatide	ApyA	ApyHI	C-terminal Asp	Aminopyruvic acid		30
Methanobactin-like	MovA	MovX #	C-terminal Asn	C-terminal amide		27
Oxazolin	HvfA	HvfBC	Cys	Oxazolone thioamide		32
Bufferin	BufA	BufBC	Cys	Thiooxazole		22

718
719 ** The activity of ChrH requires a SAM radical; ChrI was reported to be dispensable in a recent*
720 *review (ref. 21). # No partner protein required but MovX has an additional, winged helix-turn-helix*
721 *domain.*
722

723

724 **Table 2. Main features of the proteins found in clusters I to IV of Fig. 1D.**
725

Clusters	Number of seq.	Length (residues)	Number of Cys	Signatures	Major BGC organisations
I [1]	100	122-293	7 to 21	-	Prot Cluster I - MNIO - PME-type 2
II [2]	13	233-256	1	peptidyl-tRNA hydrolase	Prot Cluster II - MNIO - TIGR04222
III [3]	11	267-386	3 to 4	HEXXH motif	Prot Cluster III - MNIO/Nitroreductase
IV [1]	10	119-146	5 to 6	-	Prot Cluster IV - MNIO

726
727 *[1] Based on the presence of conserved Cys and the composition of their BGCs, the proteins found in*
728 *clusters I and IV likely represent new types of MNIO-modified precursors. [2] This is less plausible for the*
729 *proteins of cluster II given their predicted function. [3] Proteins of cluster III are more likely to be*
730 *processing or modification metalloenzymes. Putative precursors remain to be found in the BGCs*
731 *corresponding to clusters II and III proteins.*

732

733 **Table 3. Analysis of fully assembled bacterial genomes**

Type of precursor	Genomes with precursor genes	Max numbers of precursor genes per genome	Numbers of genomes with precursor genes and w/o MNIO genes	Number of genomes with precursor and MNIO genes and w/o PME-type 1 genes
DUF2282	4072	5	21	27
Buf_6-12Cys	864	4	0	1
Buf_EGKCG	2583	4	67	56
TIGR04222	1372	5	490	ND

734

735 *ND, not determined as TIGR04222 genes are not associated with PME-type 1 genes.*

736

737 # Figure legends

738

739 **Figure 1. The MNIO superfamily and associated families of precursors. A.** Taxonomic
740 distribution of the MNIO enzymes. The absolute numbers of proteins that belong to each
741 taxonomic group are provided. For the sake of clarity, only the groups that contain more than
742 ten members are shown. **B,C,D.** SSN analyses. **B.** Representative node network of MNIO
743 enzymes (alignment score threshold (AST) of 50). MNIO enzymes genetically associated with
744 Buf1 (DUF2282), Buf2 (BUF_6/12Cys), Buf_EGKCG/oxazolins (see below) and
745 Buf_EGKC|EF-hand_5 (see below) peptides are colored red, yellow, light blue and dark blue,
746 respectively. The small clusters containing MNIO enzymes that modify the precursors of
747 chryseobasins (#4), methanobactins (#11), pearlins (#13), aminopyruvatides (#17) and *Vibrio*
748 methanobactin-like RiPPs (#21) are indicated. The clusters numbered 1 to 21 are further
749 described later in the text and in Fig. 5. **C**. Representative node network of putative precursors
750 genetically associated with MNIO enzymes and harboring predicted signal-peptides (AST=10).
751 The largest clusters by decreasing sizes correspond to precursors of the Buf1 (DUF2282), Buf2
752 (Buf_6/12_Cys), Buf_EGKCG/oxazolin, CxxxxC (see below) and chryseobasin families. **D**.
753 Representative node network of putative precursors genetically associated with MNIO enzymes
754 and lacking predicted signal-peptides (AST=10). TIGR04222 proteins (purple nodes) and
755 methanobactins (green nodes) were notably found in this search. Methanobactins do not form
756 clusters in the node network, most likely because of their small size and low sequence identity.
757 The largest four sequence clusters of unknown proteins (labeled I to IV) are further described
758 in the text and in Table 2.

759

760 **Figure 2. Flow chart for the identification of putative MNIO substrates and partners. A**.
761 Search for MNIO substrates (putative RiPP precursors). After collecting all MNIO ORFs, the
762 ORFs that precede them were collected and analyzed for domain signatures. SSN analyses
763 revealed new putative families, for which hmm signatures were generated. Available and new
764 signatures were used to collect all family members in the NCBI nr protein database, irrespective
765 of their genetic environments. These proteins and their BCGs were analyzed. **B**. Search for
766 partner proteins of MNIO enzymes. The ORFs that follow MNIO ORFs were analyzed for
767 domain signatures, and SSN analyses were performed with the unknown proteins. HHpred
768 analyses were carried out for all proteins from the various sequence clusters of the SSN
769 network, and AlphaFold2 models were generated for a few of them. Only proteins less than
770 90% identical in sequence were used in all analyses.

771

772 **Figure 3. Features of precursors with signal peptides. A-C**. Taxonomic distributions of the
773 Buf1 (DUF2282) bufferins (A), Buf2 (Buf_6/12Cys) bufferins (B), and Buf-
774 EGKCG/oxazolins (C). **D-F.** Major types of genetic organizations with their relative
775 frequencies in the NCBI nr protein database. Some of the stand-alone precursor genes
776 genuinely form single-gene transcriptional units, whereas others are flanked by genes that are
777 not conserved to a significant extent and are thus not shown. **G-I.** Structural models of a Buf1
778 bufferin (WP_011519119 of *Cupriavidus metallidurans*; G), two Buf2 bufferins
779 (WP_010947964 of *Legionella pneumophila* subsp. *fraseri* and WP_077285100 of
780 *Cognaticolwellia aestuarii*; H) and a mid-size Buf_EGKCG protein (WP_070991527 of
781 *Pseudoalteromonas byunsanensis*; I). The Cys residues are shown in stick representation. In
782 Buf1, the central Cys residues are the targets of PTMs, whereas the first and last ones form a
783 disulfide bond. In the Buf_EGKCG/oxazolin protein, the Cys side chains that might be the
784 targets of PTMs are predicted to point toward the inside of the helix. The Buf_EGKC proteins
785 that form the N-terminal domains of the Buf_EGKC|EF-hand_5 chimeras are predicted to be
786 unstructured and are thus not represented. **J-L**. Numbers of occurrences and schematic
787 representation of the fusion proteins harboring Buf1 (J), Buf2 domains (K) and Buf_EGKCG
788 domains (L). Examples of these proteins include WP_160059411.1
789 (Buf1/Cytochrom_C/FGE-sulfatase); WP_017136223.1 (Buf1/Cytochrom_C); OAD22792.1
790 (Buf1/Buf1); WP_200808486.1 (MerC/Buf1); OOO01496.1 (MerC/Buf2) and
791 WP_187711723.1 (Buf_EGKCG/Ef-hand_5/ Ef-hand_5). The MerC fusions are predicted to

792 harbour five transmembrane segments, with the C-terminal bufferin domains of the chimeras
793 on the periplasmic side of the membrane.
794
795 **Figure 4. Features of TIGR04222 proteins. A.** Taxonomic distribution. **B.** Major types of
796 genetic organizations with their relative frequencies in the NCBI nr protein database. For this
797 analysis only the TIGR04222 proteins harboring Cys residues were considered. See below (text
798 and Fig. 6) for the PME-type 2 proteins. **C.** Size distribution of the TIGR04222 proteins
799 showing the wide range of lengths. **D.** AlphaFold2 models of selected TIGR04222 proteins
800 (WP_179806563 of *Micromonospora purpureochromogenes*, 300 residues, and ADB18956 of
801 *Pirellula staley*i DSM 6068, 516 residues). The transmembrane segments are in orange, and the
802 Cys residues are shown as yellow balls. E. Sequences of the C-terminal Cys-rich domains of
803 the latter two proteins, which are predicted to be exposed to the periplasm.
804
805 **Figure 5. Characterization of small sequence clusters of MNIO enzymes to identify new**
806 **precursors. A, B.** Analyses of the MINO sequence clusters denoted #1, #2 and #3 (**A**) and
807 zoom on the sequence clusters #4 to #21 (**B**) of Fig. 1B. The taxonomic groups of these MNIO
808 enzymes and the salient features of their BGCs are shown, including the proteins frequently
809 found encoded in these BGCs (Pfam short names in red, panel **B**). In **A**, the AlphaFold2 models
810 of representative RiPPs are those of WP_073215760 *Massilia sp*. CF038 (cluster #1) and
811 WP_127967852.1 *Hahella sp*. KA22 (cluster #2), with the Cys residues shown as yellow
812 spheres. The precursors genetically associated with MNIO enzymes of sequence cluster #3 are
813 predicted to be unstructured, and therefore no model is shown. **B.** The sequence clusters #4,
814 #11, #13, #17 and #21 contain MNIO enzymes that modify the precursors of chryseobasins,
815 methanobactins, pearlins, aminopyruvatides and *Vibrio* methanobactin-like RiPPs,
816 respectively. In addition, new putative precursors were identified for MNIO enzymes of clusters
817 #7 and #9. AlphaFold models of representative proteins are shown (WP_078758616 of
818 *Novilysobacter spongiicola* and WP_053243439 of *Clostridium sp.,* respectively). Note that no
819 Cys residues are present in the latter. Among these clusters of MNIO enzymes, only the BGCs
820 of the cluster #3 (panel **A**) encode *bona fide* DUF2063 domain-containing partner proteins
821 (called PME-type 1; see below). In contrast, the BGCs of the ten clusters circled in red (panels
822 **A** and **B**), including those of TglH and ApyH for the biosynthesis of pearlins and
823 aminopyruvatides, respectively, encode PME-type 2 proteins (see below).
824

825 **Figure 6. Characterization of new MNIO partner proteins. A**. Representative node network
826 (AST= 5) and structural models of the unknown proteins encoded by genes that follow the
827 MNIO-coding genes. The nodes of proteins found to be structurally related to the PME-type 1
828 *NGO1945* protein of *Neisseria* (pdb #3dee) signature but that do not possess the Pfam
829 DUF2063 signature are colored in red. They are hereafter called PME-type 2 proteins. The inset
830 shows the X-ray structure of the *NGO1945* protein, with its DUF2063 and RRE domains
831 colored orange and green, respectively. The Alphafold models a to d of representative PME-
832 type 2 proteins found in the major sequence cluster are shown with the same color code: a,
833 WP_033442786 of *Saccharothrix* sp. NRRL B-16314 (UNIPROT A0A7W9M5Z5, 115
834 residues); b, WP_091237309 of *Micromonospora matsumotoense* (UNIPROT A0A246RJP9,
835 158 residues); c, WP_151034108 of *Cellvibrio* sp. KY-GH-1 (UNIPROT A0A5J6PC27, 215
836 residues); d, WP_006975235 of *Plesiocystis pacifica* (UNIPROT A6GEU5, 276 residues). The
837 blue model shown in e (WP_035891099 of *Legionella norrlandica*) represents a potential new
838 type of partner proteins that belong neither to PME-type 1 nor to PME-type 2 proteins. **B.** Size
839 distribution of the PME-type 1and PME-type 2 proteins. The majority length of 250-260
840 residues of PME-type 1 proteins in the left-hand graph corresponds to a fusion between a
841 DUF2063 domain and a complete RRE-fold type domain. In the right-hand graph, the shorter
842 proteins that are heterogeneous in length most likely do not contain full-length RRE-type
843 domains.
844
845 **Figure 7. Subgroups of MNIO enzymes are associated with specific precursors and**
846 **partner proteins.** Members of the large sequence cluster of the node network of MNIO
847 enzymes in Fig. 1B are colored according to their genetically linked precursors and partner
848 proteins. In **A**, the colors of the dots represent the MNIO enzymes according to their putative
849 precursor substrates. The pale green dots include both the Buf_EGKCG/oxazolin proteins and
850 the Buf_EGKCG|EF-hand_5 chimeras. The precursors with signal peptides are circled with a
851 stippled blue line. In **B**, the orange and cyan dots represent MNIO enzymes associated with
852 PME-type 1 and with PME-type 2 partner proteins, respectively.
853
854
855
856

Figure 1

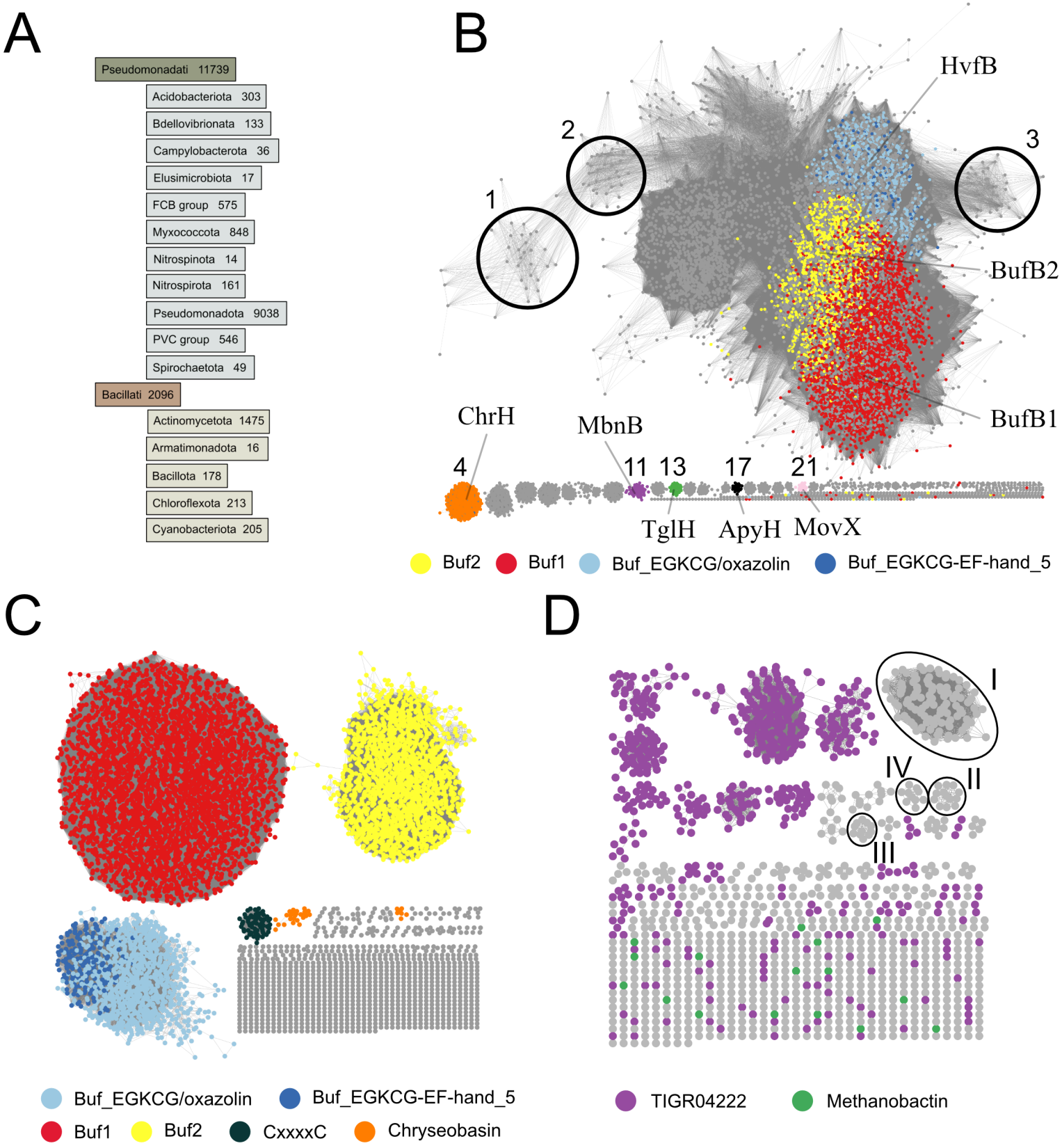

A
Pseudomonadati 11739
Acidobacteriota 303
Bdellovibrionata 133
Campylobacterota 36
Elusimicrobiota 17
FCB group 575
Myxococcota 848
Nitrospinota 14
Nitrospirota 161
Pseudomonadota 9038
PVC group 546
Spirochaetota 49
Bacillati 2096
Actinomycetota 1475
Armatimonadota 16
Bacillota 178
Chloroflexota 213
Cyanobacteriota 205
B
HvfB
1
2
3
BufB2
BufB1
ChrH
MbnB
4
11
13
17
21
TglH
ApyH
MovX
Buf2
Buf1
Buf_EGKCG/oxazolin
Buf_EGKCG-EF-hand_5
C
Buf_EGKCG/oxazolin
Buf_EGKCG-EF-hand_5
Buf1
Buf2
CxxxxC
Chryseobasin
D
I
IV
II
III
TIGR04222
Methanobactin

Figure 2

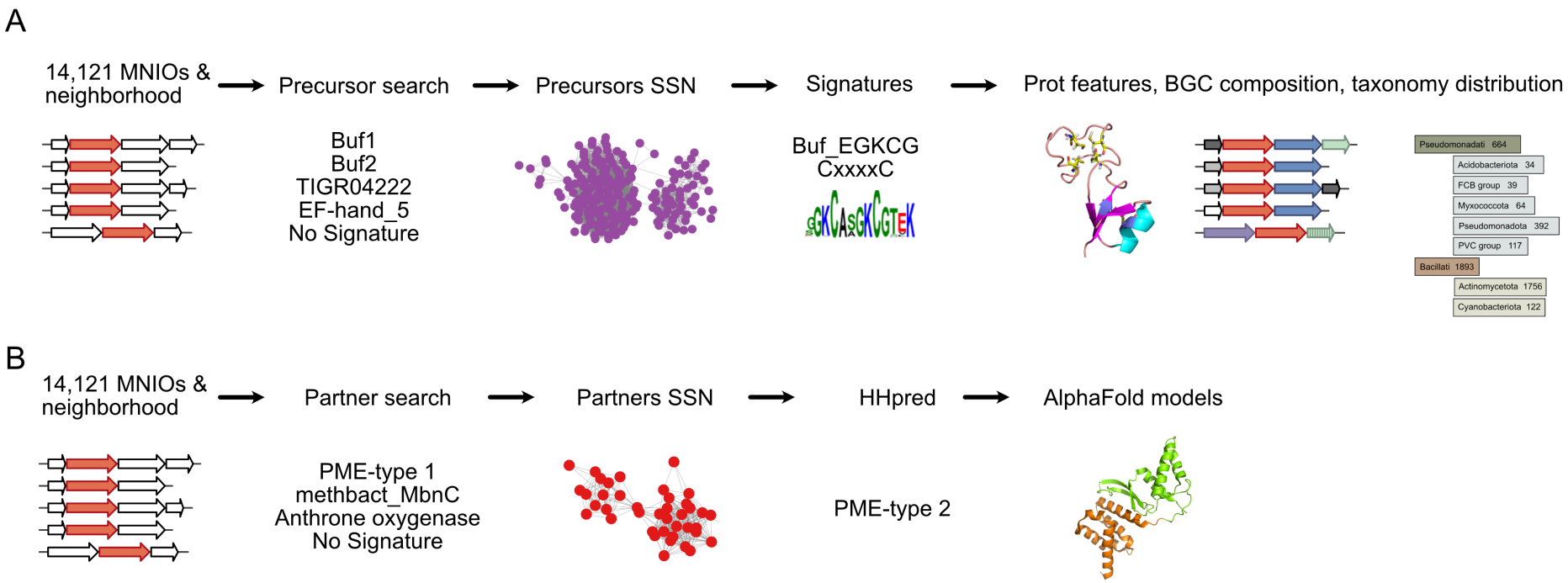

A
14,121 MNIOs & neighborhood
Precursor search
Precursors SSN
Signatures
Prot features, BGC composition, taxonomy distribution
Buf1
Buf2
TIGR04222
EF-hand_5
No Signature
Buf_EGKCG
CxxxxC
Pseudomonadati 664
Acidobacteriota 34
FCB group 39
Myxococcota 64
Pseudomonadota 392
PVC group 117
Bacillati 1893
Actinomycetota 1756
Cyanobacteriota 122
B
14,121 MNIOs & neighborhood
Partner search
Partners SSN
HHpred
AlphaFold models
PME-type 1
methbact_MbnC
Anthrone oxygenase
No Signature
PME-type 2

Figure 3

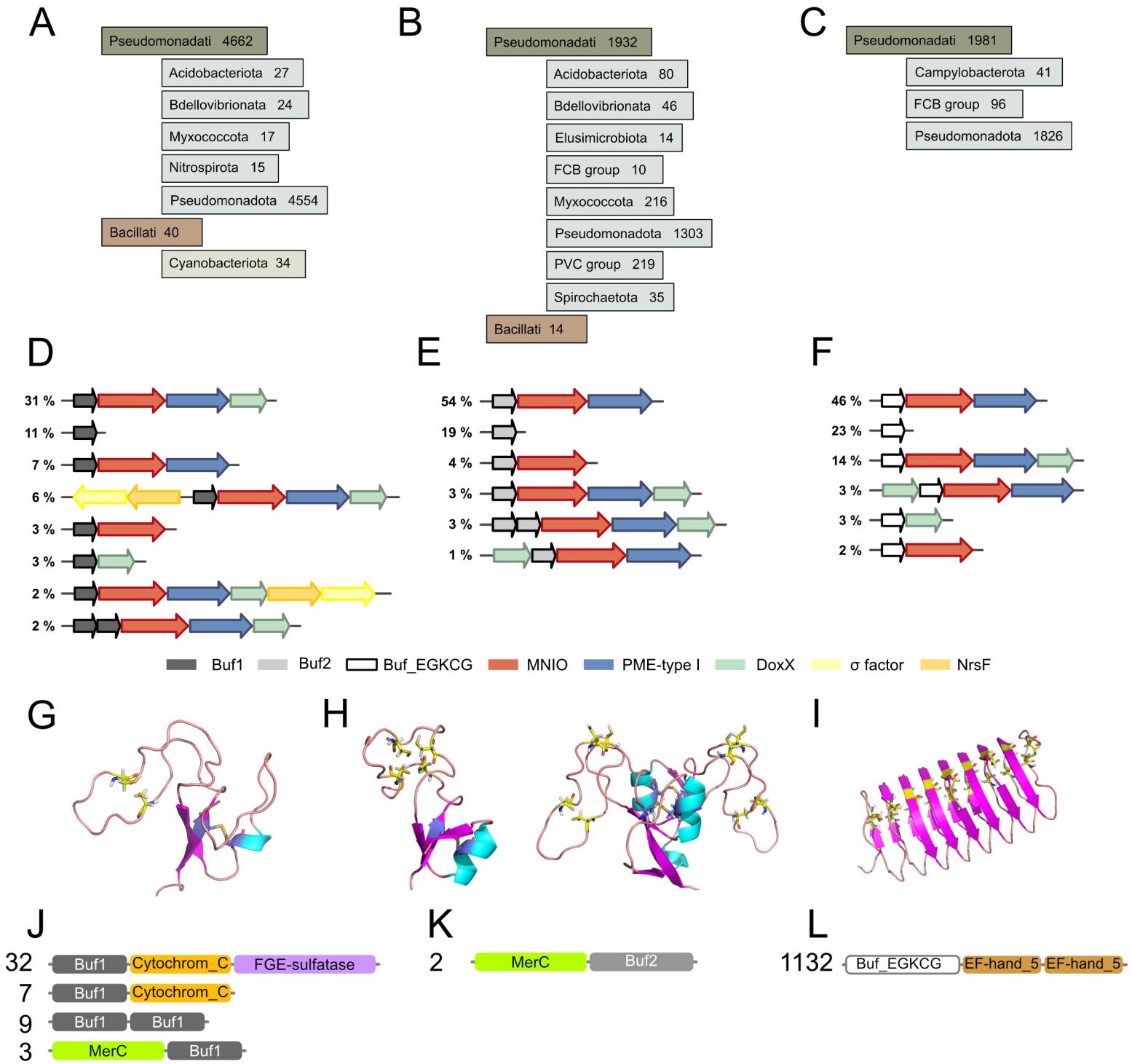
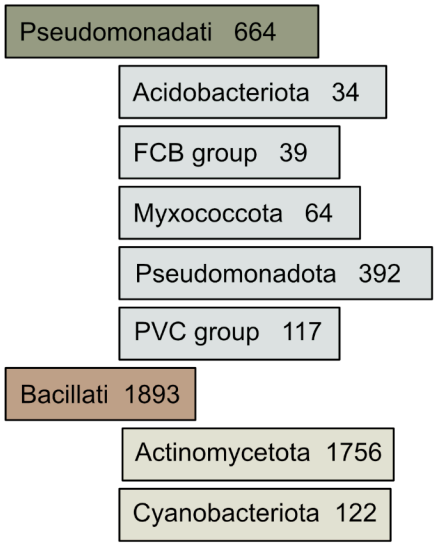
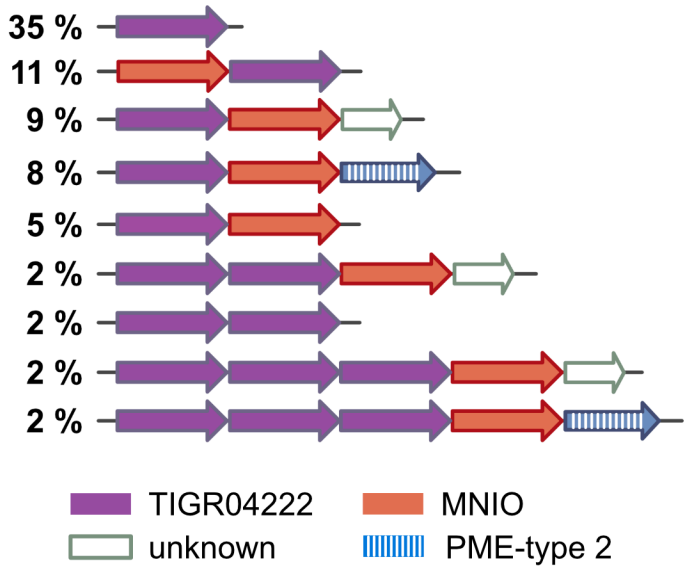

A
Pseudomonadati 4662
Acidobacteriota 27
Bdellovibrionata 24
Myxococcota 17
Nitrospirota 15
Pseudomonadota 4554
Bacillati 40
Cyanobacteriota 34
B
Pseudomonadati 1932
Acidobacteriota 80
Bdellovibrionata 46
Elusimicrobiota 14
FCB group 10
Myxococcota 216
Pseudomonadota 1303
PVC group 219
Spirochaetota 35
Bacillati 14
C
Pseudomonadati 1981
Campylobacterota 41
FCB group 96
Pseudomonadota 1826
D
31 %
11 %
7 %
6 %
3 %
3 %
2 %
2 %
E
54 %
19 %
4 %
3 %
3 %
1 %
F
46 %
23 %
14 %
3 %
3 %
2 %
Buf1
Buf2
Buf_EGKCG
MNIO
PME-type I
DoxX
σ factor
NrsF
G
H
I
J
32 Buf1 Cytochrom_C FGE-sulfatase
7 Buf1 Cytochrom_C
9 Buf1 Buf1
3 MerC Buf1
K
2 MerC Buf2
L
1132 Buf_EGKCG EF-hand_5 EF-hand_5

Figure 4

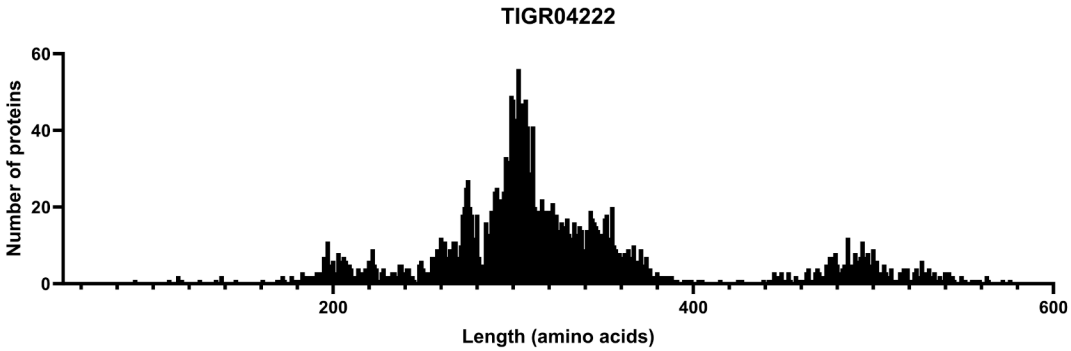
A



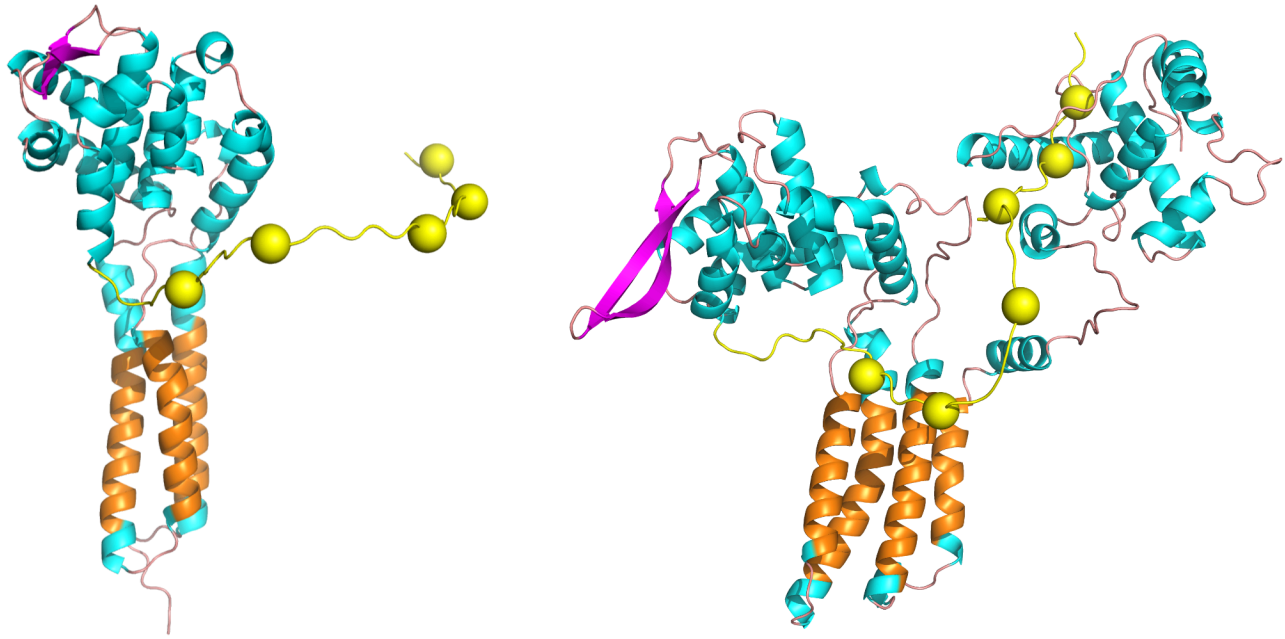
B



C



D



E

SSGGTGGGSCGGGSSCGGGGSSCGGGGGCGGGGGCGG

SSDVSSGCSSGCSGGSGCGGGGGCGGSGCGGCGGD

Figure 5

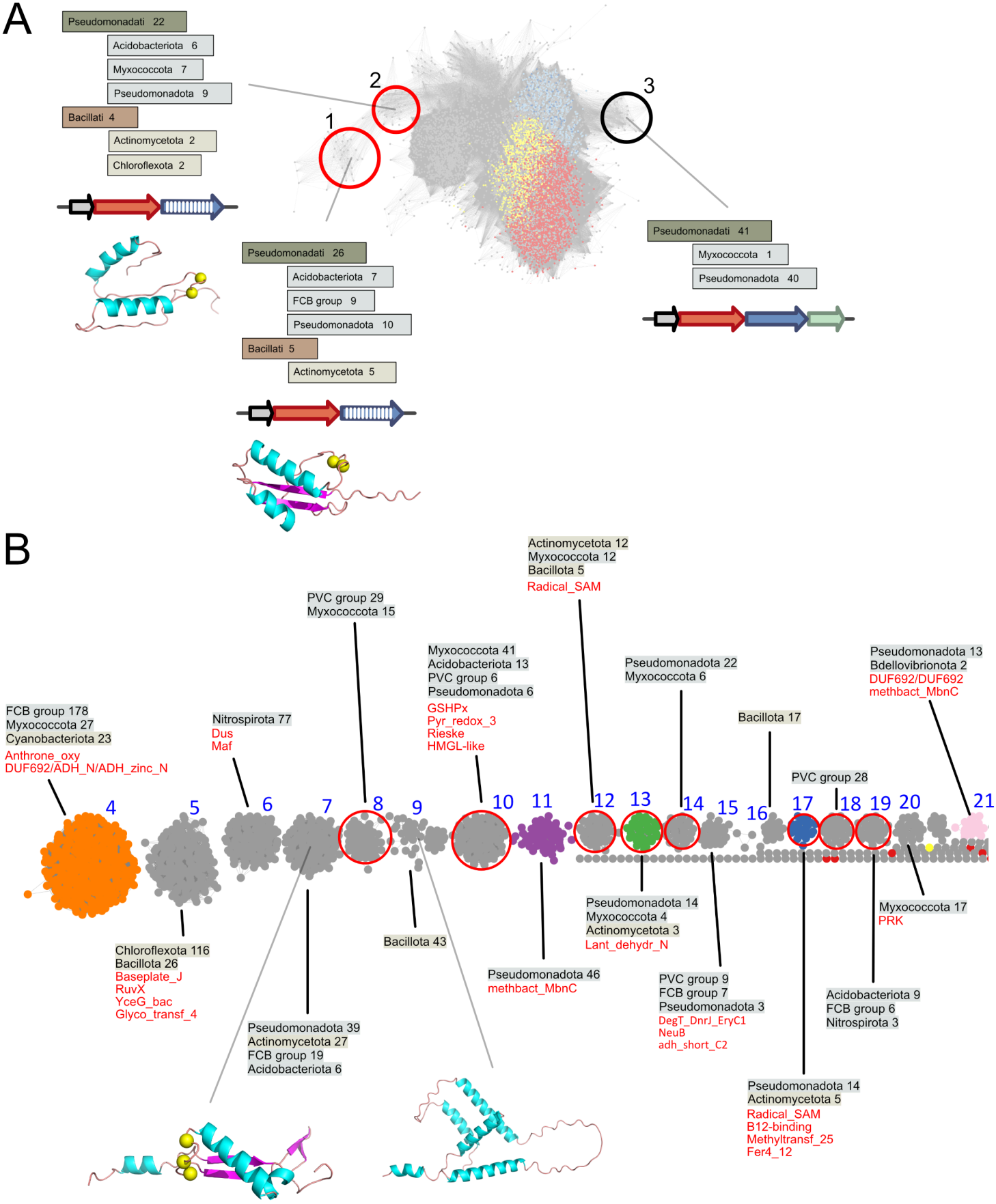

A
Pseudomonadati 22
Acidobacteriota 6
Myxococcota 7
Pseudomonadota 9
Bacillati 4
Actinomycetota 2
Chloroflexota 2
1
2
3
Pseudomonadati 26
Acidobacteriota 7
FCB group 9
Pseudomonadota 10
Bacillati 5
Actinomycetota 5
Pseudomonadati 41
Myxococcota 1
Pseudomonadota 40
B
FCB group 178
Myxococcota 27
Cyanobacteriota 23
Anthrone_oxy
DUF692/ADH_N/ADH_zinc_N
Nitrospirota 77
Dus
Maf
PVC group 29
Myxococcota 15
Myxococcota 41
Acidobacteriota 13
PVC group 6
Pseudomonadota 6
GSHPx
Pyr_redox_3
Rieske
HMGL-like
Actinomycetota 12
Myxococcota 12
Bacillota 5
Radical_SAM
Pseudomonadota 22
Myxococcota 6
Bacillota 17
PVC group 28
Pseudomonadota 13
Bdellovibrionota 2
DUF692/DUF692
methbact_MbnC
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Chloroflexota 116
Bacillota 26
Baseplate_J
RuvX
YceG_bac
Glyco_transf_4
Pseudomonadota 39
Actinomycetota 27
FCB group 19
Acidobacteriota 6
Bacillota 43
Pseudomonadota 46
methbact_MbnC
Pseudomonadota 14
Myxococcota 4
Actinomycetota 3
Lant_dehydr_N
PVC group 9
FCB group 7
Pseudomonadota 3
DegT_DnrJ_EryC1
NeuB
adh_short_C2
Pseudomonadota 14
Actinomycetota 5
Radical_SAM
B12-binding
Methyltransf_25
Fer4_12
Acidobacteriota 9
FCB group 6
Nitrospirota 3
Myxococcota 17
PRK

Figure 6

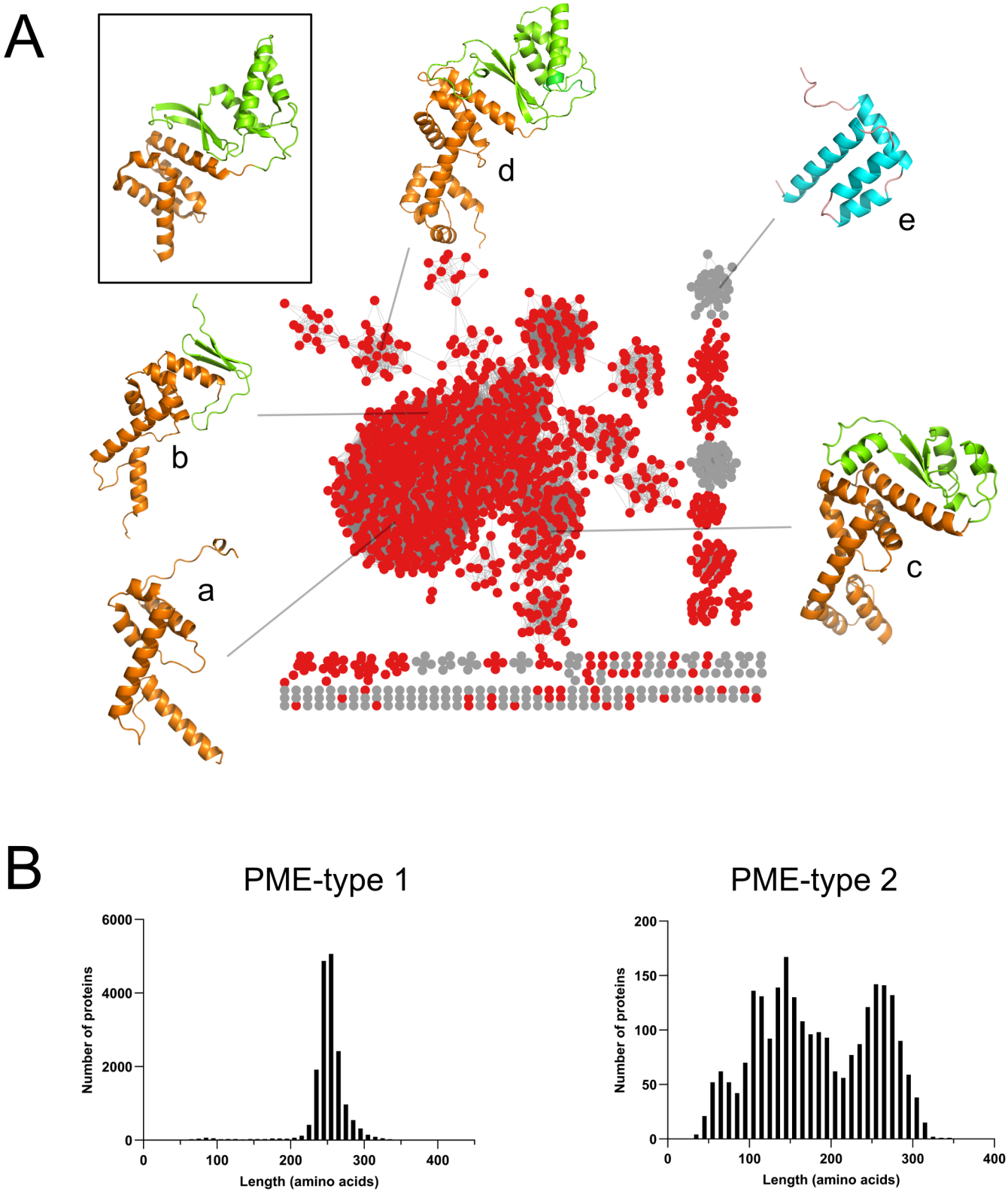
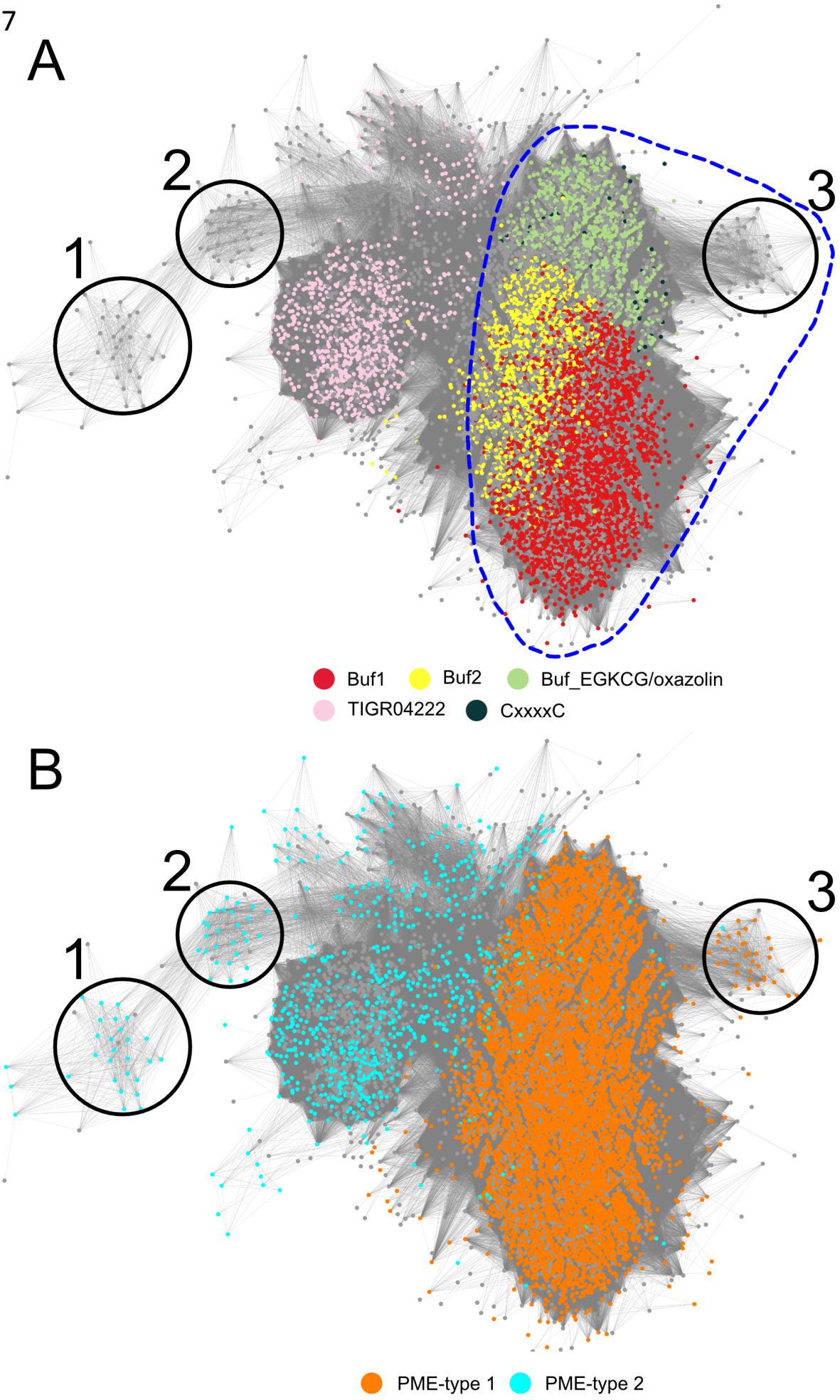

A
a
b
c
d
e
B
PME-type 1
Number of proteins
Length (amino acids)
0
2000
4000
6000
0
100
200
300
400
PME-type 2
Number of proteins
Length (amino acids)
0
50
100
150
200
0
100
200
300
400

Figure 7


A
1
2
3
Buf1
Buf2
Buf_EGKCG/oxazolin
TIGR04222
CxxxxC
B
1
2
3
PME-type 1
PME-type 2