Genome mining- and synthetic biology-enabled production of hypermodified peptides

Agneya Bhushan[1], **Peter J. Egli**[2], **Eike E. Peters**[1], **Michael F. Freeman**[3], **Jörn Piel**[1,*]

[1]Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zurich, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland [2]Kantonsschule Obwalden, Rütistrasse 5, 6060 Sarnen, Switzerland [3]Department of Biochemistry, Molecular Biology, and Biophysics and BioTechnology Institute, University of Minnesota–Twin Cities, St Paul, Minnesota 55108, USA

Summary

The polytheonamides are among the most complex and biosynthetically distinctive natural products known to date. These potent peptide cytotoxins are derived from a ribosomal precursor processed by 49 mostly non-canonical posttranslational modifications. Since the producer is a "microbial dark matter" bacterium only distantly related to any cultivated organism, >70-step chemical syntheses have been developed to access these unique compounds. Here we mined prokaryotic diversity to establish a synthetic platform based on the new host *Microvirgula aerodenitrificans* that produces hypermodified peptides within two days. Using this system, we generated the aeronamides, new polytheonamide-type compounds with near-picomolar cytotoxicity. Aeronamides, as well as the polygeonamides produced from deep-rock biosphere DNA, contain the highest numbers of D-amino acids in known biomolecules. With increasing bacterial genomes being sequenced, similar host mining strategies might become feasible to access further elusive natural products from uncultivated life.

Introduction

Uncultivated bacteria comprise a major portion of biological diversity, encompassing numerous deep-branching taxa with poorly known functional properties. Such "microbial dark matter" has been proposed as an enormous untapped resource of bioactive natural products for pharmaceutical applications[1], based on a wealth of detected biosynthetic gene clusters (BGCs) in metagenomic datasets[2,3], new natural product families[4,5], and hidden, talented producers with a rich chemistry[6]. Showcase compounds that embody all three

aspects are the polytheonamides (Fig. 1a, b), marine sponge-derived peptide cytotoxins that are chemically distinct from any other known natural product type[7]. Produced by a member of the sponge microbiome, the chemically rich symbiont 'Candidatus Entotheonella factor'[4,6], the 49-residue polytheonamides are β-helical pore-forming peptides that feature numerous non-proteinogenic amino acids, including 18 diverse D-amino acids that alternate with L-configured residues. Counterintuitively considering this extraordinary complex structure, polytheonamides are of ribosomal biosynthetic origin and belong to a large new family of ribosomally synthesized and posttranslationally modified peptides (RiPPs)[8] termed proteusins, of which polytheonamides are currently the only characterized members[4].

Functional studies revealed a complex but astonishingly streamlined pathway (Fig. 1b)[4,9]. Acting on PoyA, a precursor protein containing standard L-amino acids, only 7 enzymes introduce 49 posttranslational modifications in a highly promiscuous and precisely controlled fashion. PoyA comprises an N-terminal leader region and a C-terminal core that is the target of the modifications. A single radical *S*-adenosylmethionine (rSAM) enzyme, PoyD, generates all 18 D-amino acids via iterative epimerizations. Further enzymes install 8 *N*-methylations of Asn side chains (PoyE), 4 hydroxylations (PoyI), 1 dehydration at Thr (PoyF), and 17 methylations at diverse non-activated carbon atoms (PoyB and PoyC), including 4 methylations to construct a *t*-butyl unit (PoyC). Ultimately, proteolytic cleavage by PoyH releases the core and triggers hydrolysis of an N-terminal enamine function at the *t*-butylated Thr to the α-keto moiety of polytheonamides[9,10].

Marine sponges are a treasure trove of bioactive natural products[11], but further pharmacological development is impeded by limited supply and synthetically challenging chemical structures. Impressive total syntheses of polytheonamides achieved by the Inoue group required over 70 steps for the optimized route[12,13]. Sustainable production was proposed based on the suspected or known roles of 'Entotheonella' and other symbiotic bacteria as actual sources of many sponge compounds[14,15]. However, to date these have not been realized, since the known producers in sponges remain uncultured[15], are only distantly related to established bacterial gene expression hosts[6], and often use unconventional, poorly studied enzymes for natural product biosynthesis. Considering their structural complexity, potent bioactivity, and noncanonical biosynthesis, polytheonamides represent an informative model to learn how bacterial production systems for sponge bioactives can be established. The producer 'E. factor' belongs to the candidate phylum 'Tectomicrobia'[6], a to-date uncultured "microbial dark matter" taxon. As an alternative method to cultivation, we encountered multiple challenges when attempting to reconstitute the complete symbiont pathway in heterologous bacterial hosts. For example, although the epimerase PoyD acts irreversibly at each amino acid center, it processed only the C-terminal half of the core in *E. coli*[9]. The C-methyltransferases PoyB and PoyC remained completely inactive in *E. coli*. Both are cobalamin-dependent rSAM methyltransferases, a highly challenging protein family in the context of biotechnological applications[9]. Functional *poyB* and *poyC* expressions were ultimately possible in the non-standard host *Rhizobium leguminosarum*[9], which unlike *E. coli* contains a complete cobalamin biosynthetic pathway. Here, C methylations occurred at most core positions, but with low efficiency, to yield complex mixtures of products carrying only 1-4 of the 17 methyl groups. Thus, only early polytheonamide intermediates were accessible with these hosts.

Here we show that a combination of genome mining enabled by functional and molecular dynamics insights and bacterial host development provides a highly efficient production system for polytheonamide-type compounds. The platform can be used to produce various hypermodified peptides and generates terminal products within 2 days. Among these, the aeronamides with near-picomolar cytotoxicity and the polygeonamides, compounds accessed from deep-rock biosphere DNA, feature the highest number of D-amino acids currently known for biologically produced molecules. With massively increasing sequencing efforts that reveal a large BGC diversity outside the established natural product sources[16], similar production strategies can be expected provide access to diverse elusive chemistries from microbial dark matter.

Results and Discussion

Genome mining reveals candidates for polytheonamide-type BGCs in unusual, taxonomically diverse bacteria

To search for potential alternative expression hosts that contain candidate polytheonamide-type clusters, we considered the requirements for β-helix formation as the characteristic, bioactivity-conferring feature. An alternating DL-amino acid pattern, as introduced by the epimerase PoyD, is known to promote β-helix formation[17], but since the epimerization patterns mainly depend on the precursor sequence and currently elude prediction[18,19], the mere presence of a *poyD* homolog in a proteusin BGC is not indicative of a polytheonamide-type compound. However, previous NMR[7] and molecular dynamics studies[20] suggested that a stable β-helix only results from side-chain Asn N-methylation in repeated NX_5N motifs. This modification by the N-methyltransferase PoyE at every sixth position[4,9], corresponding to one helix turn, promotes formation of a stabilizing long-range hydrogen bond clamp between the methylamide moieties. Polytheonamides contain a second, interlocked NX_5N motif that might further increase stability[7] (Fig. 1c). Based on these more characteristic features, we analyzed (meta)genomic datasets for clustered NX_5N-proteusin precursor, epimerase, and methyltransferase genes, which revealed candidate polytheonamide-type BGCs in three distantly related bacteria (Fig. 1b, Supplementary Fig. 1, Supplementary Table 1): a deep-rock subsurface metagenomic bin assigned to the uncultivated *Rhodospirillaceae* (Alphaproteobacteria) bacterium BRH-c57 (*geo* cluster)[21], a single-cell genome of the marine uncultivated Verrucomicrobia member SCGC AAA164-I21 (*vep* cluster, lacking a methyltransferase gene)[22], and the cultivable Betaproteobacterium *M. aerodenitrificans* DSM 15089 isolated from activated sludge (*aer* cluster)[23]. The single-cell *vep* locus is positioned at the edge of a sequenced contig and likely incomplete. Every core except for GeoA3 featured NX_5N repeats (Fig. 1c), with the *geo* cluster encoding two such precursors among a total of three. All clusters also encode 1-2 cobalamin-dependent rSAM methyltransferases homologs (PoyB and PoyC). The *aer* and *geo* clusters also contain candidates gene for a PoyF-like dehydratase as well as a protease, although the protease families differ from that of the polytheonamide cluster protease PoyH[10]. The extensive similarities to the *poy* system suggested that polytheonamide-type natural products are more widespread in diverse bacteria and that the culturable *M. aerodenitrificans* might be an attractive production host for these rare compounds.

Extensive polytheonamide-type epimerization of the *M. aerodenitrificans* precursor AerA

Since epimerases can install diverse epimerization patterns[18, 19], we first focused on the homolog AerD from *M. aerodenitrificans* to investigate whether the cluster belongs to polytheonamide-type or unrelated compounds. An alignment of all identified precursors with PoyA suggested for the AerD substrate AerA a predicted 100 aa N-terminal leader and a 46 aa C-terminal core region (Supplementary Fig. 2). The residue numbering used in the following assigns residue one to the first of the core peptide, and negative numbers count backwards towards the N terminus[8]. Instead of the QAAGG cleavage site before Thr1 in PoyA[10], AerA as well as the three *geo* precursors contain an AVAPQ site, possibly due to different protease types encoded in the clusters (*aerH*, *geoH*, *geoP*), while the *vep* precursor has an AVAGG site. Heterologous expression of *aerA* in *E. coli* as N-terminally His$_6$-tagged protein (designated Nhis-AerA) gave soluble product when coexpressed with *aerD* and was partially soluble when expressed alone (Supplementary Figs. 3 and 4), which compared very favorably to *poyA* expressions that consistently failed in the absence of *poyD*. For *aerD* coexpressions, proteinase K digest and mass spectrometric (MS) analysis of the affinity-purified precursor revealed a unique ion matching the C-terminal 46 aa region of AerA (AerA 1-46, *m/z* = 1038.3335 Da, calc. 1038.3273 Da, [M+4H]$^{4+}$) (Fig. 2a,b), an assignment supported by MS/MS (MS^2) data (Supplementary Fig. 5). To test whether AerD protects this region from proteolysis by introducing epimerizations, we used a slightly modified method of the orthogonal D$_2$O induction based system (ODIS), previously developed and validated in our laboratory[19]. In this method, the epimerase is induced in a deuterated water background following precursor expression. Upon epimerization, each epimerized residue incorporates a deuterium ion, introducing a mass shift of 1 Da that can be localized via high-performance liquid chromatography-MS^2 (HPLC-MS^2). The proteinase K-protected core region detected in these samples exhibited a mass shift of +21 Da only in *aerD* coexpressions, pointing to 21 epimerizations that even eclipse the 18 epimerizations in polytheonamides (Fig. 2b). By MS^2 analysis, fragmentation data were obtained that permitted localization of 18 of the 21 epimerized residues, with two more epimerizations localized by MS^2 analysis of the GluC-digested fragment (Supplementary Fig. 5). The remaining epimerization was localized by MS^2 analysis to the core region 44-46 and would be on Val45 in case the alternating pattern is continued. Thus, AerD introduces remarkably efficient polytheonamide-like epimerization over almost the entire length of the peptide, including all five Asn residues, and only interrupted by the single achiral Gly11 (Fig. 2c).

Reporter assays suggest production conditions in *M. aerodenitrificans*

We next investigated under what conditions the *aer* cluster is expressed in *M. aerodenitrificans*. To permit use of a reporter assay for this genetically poorly studied genus, we established a conjugation-based method for plasmid introduction. Subsequently, the taurine-inducible plasmid pLMB509, first developed for heterologous expression in rhizobia[24], was tested by introducing the putative promoter region of the *aer* cluster (located upstream of *aerC*, Supplementary Table 2) in front of the glucuronidase reporter gene *gusA*. Of the conditions tested, cultivation of the *M. aerodenitrificans* reporter strain at 30 °C in terrific broth (TB) medium resulted in considerably higher GusA activity over other conditions already after one day (Supplementary Fig. 6). To further test for the presence of

active *aer* enzymes, we incubated epimerized Nhis-AerA, purified from *E. coli aerA+aerD* expressions, with cell-free *M. aerodenitrificans* lysates prepared at induction conditions. Gratifyingly, this experiment resulted in mass shifts corresponding to five methylations (Fig. 3a), further confirming that the cluster is expressed in the native host. Four of the five methylations were localized to Asn25, Asn31, Asn37 and Asn41 (Supplementary Fig. 7). The fifth methylation could only be localized to a C-terminal four-residue fragment and was proposed to occur on Asn43, since rSAM methyltransferase activity was neither apparent nor likely in these experiments performed under aerobic conditions (Fig. 3b, Supplementary Fig. 7).

Characterization of aeronamide A

The cell-free conversion data supported the existence of polytheonamide-type products that might be present in *M. aerodenitrificans* under induction conditions. However, extensive analyses of cultures using diverse *in situ*- and extraction-based MS- and HPLC methods as well as cytotoxicity assays failed to identify a candidate, perhaps due to absorbance or aggregation of the putatively highly lipophilic compounds. We therefore tested a "tagged-bait" strategy next to generate fully modified, but uncleaved precursors in *Microvirgula*. The precursor bait for the *aer* enzymes was an AerA variant carrying a His$_6$ tag for reisolation and an AVAGG instead of the AVAPQ site at the leader-core interface to prevent core release by the endogenous protease. After conjugation of the encoding gene *nhis-aerA(GG)* under control of the *aer* promoter into *M. aerodenitrificans* and cultivation at aeronamide production conditions, proteins were affinity-purified from cultures periodically collected over a period of three days and treated with the protease GluC to simplify MS-based peptide analysis. The LC-MS^2 data (Fig. 4a, Supplementary Fig. 8) revealed a major product at *m/z* 1165.68 ([M+4H]$^{4+}$), corresponding to 11 methylations, and a minor peak at *m/z* 1164.68 ([M+4H]$^{4+}$), corresponding to 12 methylations and 1 dehydration. MS^2 data pinpointed single methylations to 4 Asn residues, with the fifth methylation on Asn43 again proposed, similar to the polytheonamides, along with the remaining 7 C-methylations on 5 Val (positions 4, 10, 18, 34), Ile2, and on Leu6. Similar to the *poy* system, in which Thr1 methylation by PoyC is observed only after dehydration,[9] the methylated AerA Ile2 was present only when Thr1 was dehydrated, suggesting this to be one of the final reactions prior to core release. Strikingly, expression of Nhis-AerA with the native PQ cleavage site in *M. aerodenitrificans* for only a single day was sufficient to produce a major product at 1192.44 (m/z, [M+4H]$^{4+}$) that carried the attached leader, all 12 methylations and 1 dehydration (Fig. 4b). No intermediate species lacking the dehydration was observed, perhaps because the dehydratase AerF prefers the native PQ to the engineered GG motif adjacent to Thr1. Another minor product (12% relative abundance) corresponding to an extra methylation was also present, which was localized to Ile2 (Supplementary Fig. 9), in effect adding two methyl groups to the residue. Furthermore, treatment of purified Nhis-AerA with proteinase K left the core intact, suggesting full epimerization in *M. aerodenitrificans* (Supplementary Fig. 10). The product had an observed mass shift of +1 Da (expected mass = 4299.4673 Da, observed mass = 4300.4836 Da) as compared to the hypothetical Thr1 enamine core, consistent with an α-keto function due to spontaneous hydrolysis as for polytheonamides. We furthermore detected the native, untagged AerA precursor after purification of Nhis-AerA(GG) from *M. aerodenitrificans* (Supplementary Fig. 11), as well as in subsequent

expressions of the non-native cores described below, suggesting a leader-mediated association of the precursor proteins. Importantly, the presence of the native modified peptide in the non-native core expressions also suggests complete conversion for the product observed during Nhis-AerA expressions. The combined data assign functions to all posttranslational enzymes in the *aer* pathway. The remarkably clean and rapid conversion of Nhis-AerA to a single major product suggests that the characterized core contains all modifications prior to release from the leader. Compared with the polytheonamides, but in agreement with the absence of a close *poyC* homolog in the *aer* cluster, aeronamide A lacks the *tert*-butyl moiety on Thr1, which has been implicated as a cytotoxicity determinant of polytheonamides[25].

Generating the terminal compound

The clean conversion of the aeronamide precursor to one main product with modifications accounting for all *aer* genes suggested that the leader-attached product is the final intermediate in the pathway. We therefore aimed to generate leaderless aeronamide A and determine its bioactivity. While proteinase K leaves the core intact, the observed amounts were unsatisfactory. After disappointing trials with other commercial peptidolytic enzymes, we decided to heterologously produce the cluster-encoded protease AerH in *E. coli*, which was predicted to be a 37 kDa trypsin-like serine protease. We expressed and purified Nhis-AerH, with the protease eluting in two fragments of ~24 and ~13 kDa (Supplementary Fig. 12), a known feature of self-cleaving proteases[26]. The purified Nhis-AerH components were then incubated with Nhis-AerA from *M. aerodenitrificans* for 18 hours in the presence of $CaCl_2$ and analyzed via LC-MS. Conveniently, the analysis showed the presence of one major product (Supplementary Fig. 13) corresponding to aeronamide A. Of note is that the hydrophobic aeronamides precipitated out after cleavage from the leader and needed to be dissolved in *n*-propanol for further analysis. For isolation, Nhis-AerA from a 5.5 L *M. aerodenitirificans* culture and treated with Nhis-AerH was purified on a C-18 solid phase exchange column followed by further separation by HPLC to yield 600 μg of pure aeronamide A (Fig. 5a).

The bioactivity of aeronamide A was tested in *E. coli, Bacillus subtilis, Saccharomyces cerevisiae* and HeLa assays. Similar to polytheonamides, aeronamide A showed potent cytotoxic activity against HeLa cells with an IC_{50} value of 1.48 nM (polytheonamide B: 0.58 nM for these cells[27]), but not towards the bacteria and fungi. To test whether the cytotoxicity is based on a similar pore-forming mechanism as for polytheonamides, we performed a H^+/Na^+ ion exchange activity assay on artificial liposomes. Gratifyingly, a similar capability as exhibited by polytheonamides for transporting H^+ and Na^+ ions was induced by aeronamide A (Fig. 5b). The low nanomolar activity was unexpected given that aeronamide A lacks the *tert*-butyl moiety on Thr1 (Fig. 5c), implicated as being a key factor in polytheonamide cytotoxicity. While studies on synthetic polytheonamide variants had tested various lipophilic moieties at this position[25,28], an aeronamide-type α-ketobutyryl terminus was not evaluated.

Core switching yields various hypermodified peptides

M. aerodenitrificans cleanly generates hypermodified products after only 1-2 days in contrast to the previously tested *E. coli* and rhizobial expressions. To further explore the synthetic scope of this remarkable host, core peptides from the deep-rock metagenome (GeoA1-A3), the Verrucomicrobial single-cell genome (VepA), and of PoyA were swapped with the native core sequence of Nhis-AerA and Nhis-AerA(GG). Following a two-day expression in *M. aerodenitrificans*, samples were collected, purified, and analyzed via LC-MS following GluC treatment. Information about epimerizations was gained from independent *E. coli* ODIS experiments, which could not be conducted in *Microvirgula*. For as yet unknown reasons, neither modified nor unmodified precursors were observed for the GeoA3 and VepA variants. In contrast, extensively modified peptides with various dehydration, methylation, and epimerization patterns were obtained for the other six constructs, which are summarized for the three cores in Fig. 6 and detailed in Supplementary Figs. 14-17. For the polytheonamide core, the GG as well as the PQ variant yielded products with up to 10 methylations. LC-MS^2 data for the major ions localized up to 6 methylations to Ile and Val residues, but no methylations were detected on six Asn residues with observable MS^2 fragments (Supplementary Fig. 14). ODIS experiments in *E. coli* with the PQ variant revealed epimerized product carrying 6 D-residues, but surprisingly the pattern was shifted compared to that of polytheonamides and left all Asn residues in the L-configuration (Supplementary Fig. 14). This result and the fact that all previously observed N-methylations occurred on D-Asn suggests that the D-configuration is an important prerequisite for methylation by PoyE and AerE.

Cores of the deep-rock precursors GeoA1 and GeoA2 yielded hypermodified "polygeonamides" (Supplementary Fig. 15), of which the variants carrying the natural PQ instead of the GG motif were dehydrated and also showed more extensive, up to 17-fold methylations. For the most abundant MS ion corresponding to 14 methylations it was possible to localize the dehydration to Thr1 and methyl groups to at least 5 Asn and diverse Val, Ile, and Thr residues. Due to kinetic isotope effects, ODIS epimerization experiments in D$_2$O usually generates a mixture of partially epimerized products that were challenging to deconvolute for the GeoA peptides, since multiple species eluted in LC-MS runs at the same time (Supplementary Fig. 16). However, in all cases a clear alternating pattern was observed (Supplementary Fig. 17) after treatment of purified precursor protein with GluC or proteinase K. For both polygeonamide cores, the pattern was similar as for AerA, with 16 (of 24 possible, based on theoretical alternation) epimerized residues in the GeoA1 core and 23 (of 23) from GeoA2.

Conclusion

Uncultivated bacteria are recognized as a vast resource of new natural products.[1–3] Some of the chemistry is visible in host-microbiome associations such as sponges, tunicates, and other macro-organisms with high drug discovery potential[29]. However, despite the wealth of sequenced environmental BGCs with known and unknown functions, accessing the encoded chemistry through bacterial synthesis is a challenge. For a number of cases, heterologous gene expression[1] and cultivation of previously uncultivated microbes[30–32] have been successful, but these methods become particularly demanding for phylogenetically distant

producers and non-canonical pathways. To our knowledge, biological synthesis of natural products from invertebrate symbionts has to date only been achieved for patellamide- and divamide-type RiPPs from tunicate-associated cyanobacteria [33–35]. An alternative production method is based on the rationale that BGCs are subject to extensive horizontal gene transfer and that culturable sources therefore likely exist even for idiosyncratic compounds from deep-branching, uncultured divisions. Previous work on the serendipitous discovery of culturable bacterial producers of invertebrate polyketides[36–39] support this concept. As shown here, a targeted, BGC-based search can not only reveal such producers, but also yield novel synthetic biology chassis for previously demanding biosynthetic steps. Even during the preparation of this manuscript, we identified another polytheonamide-like cluster in the cultivable myxobacterium *Cystobacter fuscus* DSM 52655 along with the *aer* cluster in two other *Microvirgula* strains (*Microvirgula* sp. AG722 and *M. aerodenitrificans* strain BE2.4). With several 100,000 genomes currently being sequenced in big-data initiatives on uncultured and cultivated bacteria [40–43], mining the cultured biodiversity will likely be applicable to a much broader range of environmental BGCs to access novel chemistry.

Our study shows that even for "unique" compounds like polytheonamides, closely related pathways are present in taxonomically and ecologically remarkably diverse organisms that include animal symbionts, marine plankton, activated sludge microbes, and even rock-dwelling organisms, none of them belonging to the classically studied natural product sources. Polygeonamide B1 with a D-residue content of almost 50% and 23 epimerized moieties is to the best of our knowledge the biomolecule with the highest number of D-amino acids reported to date. The actual compound generated by the uncultivated producer remains unknown. It is likely related but might contain more C-methyl groups than the *Microvirgula* products, since the *geo* BGC is architecturally, virtually identical to the *aer* system except for one additional rSAM methyltransferase gene. The endolithic producer of polygeonamides, assigned as a member of the Rhodospirillaceae by metagenomic binning, was consistently detected in all examined borehole samples and thus proposed as an indigenous member of the deep-rock ecosystem[21]. Deep subsurface habitats have recently been recognized to contain some of the most diverse biomes on Earth[44], estimated to comprise 90% of the planet's bacterial biomass[45]. The data on polygeonamides provide a rare glimpse into chemical functions of these elusive life forms, suggesting that endolithic communities engage in bioactive natural product synthesis and chemical warfare. It is intriguing to speculate on the ecological function of polygeonamides. Based on their high similarity to polytheonamides and aeronamides, they likely target eukaryotes, which are indeed present in endolithic biomes[46]. The Rhodospirillaceae producer, which according to meta-omic data is an autotroph growing on H_2 and CO_2[21], might benefit from chemical defenses conferred by polygeonamides in the extremely nutrient-limited rock habitat.

The *Microvirgula* platform transformed the *aer* and *geo* precursors to hypermodified polytheonamide-like peptides with remarkable efficiency. Surprisingly, however, the *poy* precursor showed a shifted epimerization pattern as compared to natural polytheonamides. This effect was, with fewer epimerizations, previously observed in *E. coli* for epimerase coexpressions with a truncated PoyA variant[9]. In both cases, the number of core residues was shifted from odd to even or vice versa as compared to the correctly epimerized cores. It

will be interesting to test whether the spacing of core residues relative to the epimerase-binding leader[47] is responsible for this effect. In addition, further development of *M. aerodenitrificans* as a heterologous host to include new promoter and plasmid systems could enable whole pathway expression for not only the *poy* system but also other challenging natural product gene clusters. One of the most attractive features of the *Microvirgula* system is the ease to introduce C-methylations. Cobalamin-dependent rSAM enzymes that methylate unactivated carbon centers are biochemically intriguing, but technically demanding enzymes with barely utilized synthetic potential[48,49]. With more than 7000 members predicted from sequenced genes, they are involved in the biosynthesis of vitamins and many bioactive compounds including carbapenems, gentamicin, fosfomycin, novobiocins, moenomycins, thiostrepton, and bottromycins[50–53]. Most of the few characterized enzymes catalyze 1-2 C-methylations, and CysS was shown to install up to 3 C-methyl groups into cystobactamid[54]. Comparatively, AerC in *Microvirgula* methylates numerous residues in various peptide cores at high efficiency, suggesting that the host might also be suited for other C-methylation systems. Beyond C-methylations, the dozens of noncanonical modifications introduced into peptides in a single day of cultivation suggests that *M. aerodenitrificans* exhibits considerable synthetic potential as a plug and play platform for peptides that cover distinct structural space.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

## References

1. Charlop-Powers Z, Milshteyn A, Brady SF. Metagenomic small molecule discovery methods. Curr Opin Microbiol. 2014; 19:70–75. [PubMed: 25000402]

2. Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature. 2018; 558:440–444. [PubMed: 29899444]

3. Charlop-Powers Z, Owen JG, Reddy BV, Ternei MA, Brady SF. Chemical-biogeographic survey of secondary metabolism in soil. Proc Natl Acad Sci U S A. 2014; 111:3757–3762. [PubMed: 24550451]

4. Freeman MF, et al. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. Science. 2012; 338:387–390. [PubMed: 22983711]

5. Iqbal HA, Low-Beinart L, Obiajulu JU, Brady SF. Natural product discovery through improved functional metagenomics in *Streptomyces*. J Am Chem Soc. 2016; 138:9341–9344. [PubMed: 27447056]

6. Wilson MC, et al. An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature. 2014; 506:58–62. [PubMed: 24476823]

7. Hamada T, et al. Solution structure of polytheonamide B, a highly cytotoxic nonribosomal polypeptide from marine sponge. J Am Chem Soc. 2010; 132:12941–12945. [PubMed: 20795624]

8. Arnison PG, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep. 2013; 30:108–160. [PubMed: 23165928]

9. Freeman MF, Helf MJ, Bhushan A, Morinaka BI, Piel J. Seven enzymes create extraordinary molecular complexity in an uncultivated bacterium. Nat Chem. 2017; 9:387–395. [PubMed: 28338684]

10. Helf MJ, Freeman MF, Piel J. Investigations into PoyH, a promiscuous protease from polytheonamide biosynthesis. J Ind Microbiol Biotechnol. 2019; doi: 10.1007/s10295-018-02129-3

11. Carroll AR, Copp BR, Davis RA, Keyzers RA, Prinsep MR. Marine natural products. Nat Prod Rep. 2019; 36:122–173. [PubMed: 30663727]

12. Inoue M, et al. Total synthesis of the large non-ribosomal peptide polytheonamide B. Nat Chem. 2010; 2:280–285. [PubMed: 21124508]

13. Hayata A, Itoh H, Inoue M. Solid-phase total synthesis and dual mechanism of action of the channel-forming 48-mer peptide polytheonamide B. J Am Chem Soc. 2018; 140:10602–10611. [PubMed: 30040396]

14. Bewley CA, Faulkner DJ. Lithistid sponges: Star performers or hosts to the stars. Angew Chem Int Ed. 1998; 37:2162–2178.

15. Piel J. Metabolites from symbiotic bacteria. Nat Prod Rep. 2009; 26:338–362. [PubMed: 19240945]

16. Cimermancic P, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell. 2014; 158:412–421. [PubMed: 25036635]

17. Navarro E, Tejero R, Fenude E, Celda B. Solution NMR structure of a d,l-alternating oligonorleucine as a model of beta-helix. Biopolymers. 2001; 59:110–119. [PubMed: 11373724]

18. Morinaka BI, et al. Radical S-adenosyl methionine epimerases: regioselective introduction of diverse d-amino acid patterns into peptide natural products. Angew Chem Int Ed. 2014; 53:8503–8507.

19. Morinaka BI, Verest M, Freeman MF, Gugger M, Piel J. An orthogonal D_2O-based induction system that provides insights into d-amino acid pattern formation by radical S-adenosylmethionine peptide epimerases. Angew Chem Int Ed. 2017; 56:762–766.

20. Renevey A, Riniker S. The importance of N-methylations for the stability of the $\beta^{6.3}$-helical conformation of polytheonamide B. Eur Biophys J. 2017; 46:363–374. [PubMed: 27744521]

21. Bagnoud A, et al. Reconstructing a hydrogen-driven microbial metabolic network in Opalinus Clay rock. Nat Commun. 2016; 7

22. Labonte JM, et al. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. ISME J. 2015; 9:2386–2399. [PubMed: 25848873]

23. Patureau D, et al. *Microvirgula aerodenitrificans* gen. nov., sp. nov., a new gram-negative bacterium exhibiting co-respiration of oxygen and nitrogen oxides up to oxygen-saturated conditions. Int J Syst Bacteriol. 1998; 48(Pt 3):775–782. [PubMed: 9734031]

24. Tett AJ, Rudder SJ, Bourdes A, Karunakaran R, Poole PS. Regulatable vectors for environmental gene expression in Alphaproteobacteria. Appl Environ Microbiol. 2012; 78:7137–7140. [PubMed: 22820336]

25. Shinohara N, Itoh H, Matsuoka S, Inoue M. Selective modification of the N-terminal structure of polytheonamide B significantly changes its cytotoxicity and activity as an ion channel. ChemMedChem. 2012; 7:1770–1773. [PubMed: 22489077]

26. Blair WS, Semler BL. Self-cleaving proteases. Curr Opin Cell Biol. 1991; 3:1039–1045. [PubMed: 1814362]

27. Iwamoto M, Shimizu H, Muramatsu I, Oiki S. A cytotoxic peptide from a marine sponge exhibits ion channel activity through vectorial-insertion into the membrane. FEBS Lett. 2010; 584:3995–3999. [PubMed: 20699099]

28. Itoh H, Matsuoka S, Kreir M, Inoue M. Design, synthesis and functional analysis of dansylated polytheonamide mimic: an artificial peptide ion channel. J Am Chem Soc. 2012; 134:14011–14018. [PubMed: 22861006]

29. Morita M, Schmidt EW. Parallel lives of symbionts and hosts: chemical mutualism in marine animals. Nat Prod Rep. 2018; 35:357–378. [PubMed: 29441375]

30. Partida-Martinez LP, Hertweck C. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. Nature. 2005; 437:884–888. [PubMed: 16208371]

31. Kampa A, et al. Metagenomic natural product discovery in lichen provides evidence for specialized biosynthetic pathways in diverse symbioses. Proc Natl Acad Sci U S A. 2013; 110:E3129–E3127. [PubMed: 23898213]

32. Ling LL, et al. A new antibiotic kills pathogens without detectable resistance. Nature. 2015; 517:455–459. [PubMed: 25561178]

33. Schmidt EW, et al. Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. Proc Natl Acad Sci U S A. 2005; 102:7315–7320. [PubMed: 15883371]

34. Long PF, Dunlap WC, Battershill CN, Jaspars M. Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. Chembiochem. 2005; 6:1760–1765. [PubMed: 15988766]

35. Smith TE, et al. Accessing chemical diversity from the uncultivated symbionts of small marine animals. Nat Chem Biol. 2018; 14:179–185. [PubMed: 29291350]

36. Schleissner C, et al. Bacterial production of a pederin analogue by a free-living marine alphaproteobacterium. J Nat Prod. 2017; 80:2170–2173. [PubMed: 28696720]

37. Kust A, et al. Discovery of a pederin family compound in a nonsymbiotic bloom-forming cyanobacterium. ACS Chem Biol. 2018; 13:1123–1129. [PubMed: 29570981]

38. Hoffmann T, Müller S, Nadmid S, Garcia R, Müller R. Microsclerodermins from terrestrial myxobacteria: an intriguing biosynthesis likely connected to a sponge symbiont. J Am Chem Soc. 2013; 135:16904–16911. [PubMed: 24124771]

39. Tao Y, et al. Samholides, swinholide-related metabolites from a marine cyanobacterium cf. *Phormidium* sp. J Org Chem. 2018; 83:3034–3046. [PubMed: 29457979]

40. Weimer BC. 100K pathogen genome project. Gen Announc. 2017; 5:e00594–e00517.

41. Kyrpides NC, et al. Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. PLoS Biol. 2014; 12

42. Gilbert JA, Jansson JK, Knight R. Earth microbiome project and global systems biology. mSystems. 2018; 3:e00217. [PubMed: 29657969]

43. Sunagawa S, Karsenti E, Bowler C, Bork P. Computational eco-systems biology in Tara Oceans: translating data into knowledge. Mol Syst Biol. 2015; 11

44. Magnabosco C, et al. The biomass and biodiversity of the continental subsurface. Nat Geosci. 2018; 11:707–717.

45. Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. Proc Natl Acad Sci U S A. 2018; 115:6506–6511. [PubMed: 29784790]

46. Borgonie G, et al. Eukaryotic opportunists dominate the deep-subsurface biosphere in South Africa. Nat Commun. 2015; 6

47. Fuchs SW, et al. A lanthipeptide-like N-terminal leader region guides peptide epimerization by radical SAM epimerases: Implications for RiPP evolution. Angew Chem Int Ed. 2016; 55:12330–12333.

48. Wang SC. Cobalamin-dependent radical S-adenosyl-L-methionine enzymes in natural product biosynthesis. Nat Prod Rep. 2018; 35:707–720. [PubMed: 30079906]

49. Bauerle MR, Schwalm EL, Booker SJ. Mechanistic diversity of radical S-adenosylmethionine (SAM)-dependent methylation. J Biol Chem. 2015; 290:3995–4002. [PubMed: 25477520]

50. Werner WJ, et al. *In vitro* phosphinate methylation by PhpK from *Kitasatospora phosalacinea*. Biochemistry. 2011; 50:8986–8988. [PubMed: 21950770]

51. Marous DR, et al. Consecutive radical S-adenosylmethionine methylations form the ethyl side chain in thienamycin biosynthesis. Proc Natl Acad Sci U S A. 2015; 112:10354–10358. [PubMed: 26240322]

52. Lanz ND, et al. Enhanced solubilization of cass B radical S-adenosylmethionine methylases by improved cobalamin uptake in *Escherichia coli*. Biochemistry. 2018; 57:1475–1490. [PubMed: 29298049]

53. McLaughlin MI, van der Donk WA. Stereospecific radical-mediated B_{12}-dependent methyl transfer by the fosfomycin biosynthesis enzyme Fom3. Biochemistry. 2018; 57:4967–4971. [PubMed: 29969250]

54. Wang Y, Schnell B, Baumann S, Müller R, Begley TP. Biosynthesis of branched alkoxy groups: Iterative methyl group alkylation by a cobalamin-dependent radical SAM enzyme. J Am Chem Soc. 2017; 139:1742–1745. [PubMed: 28040895]
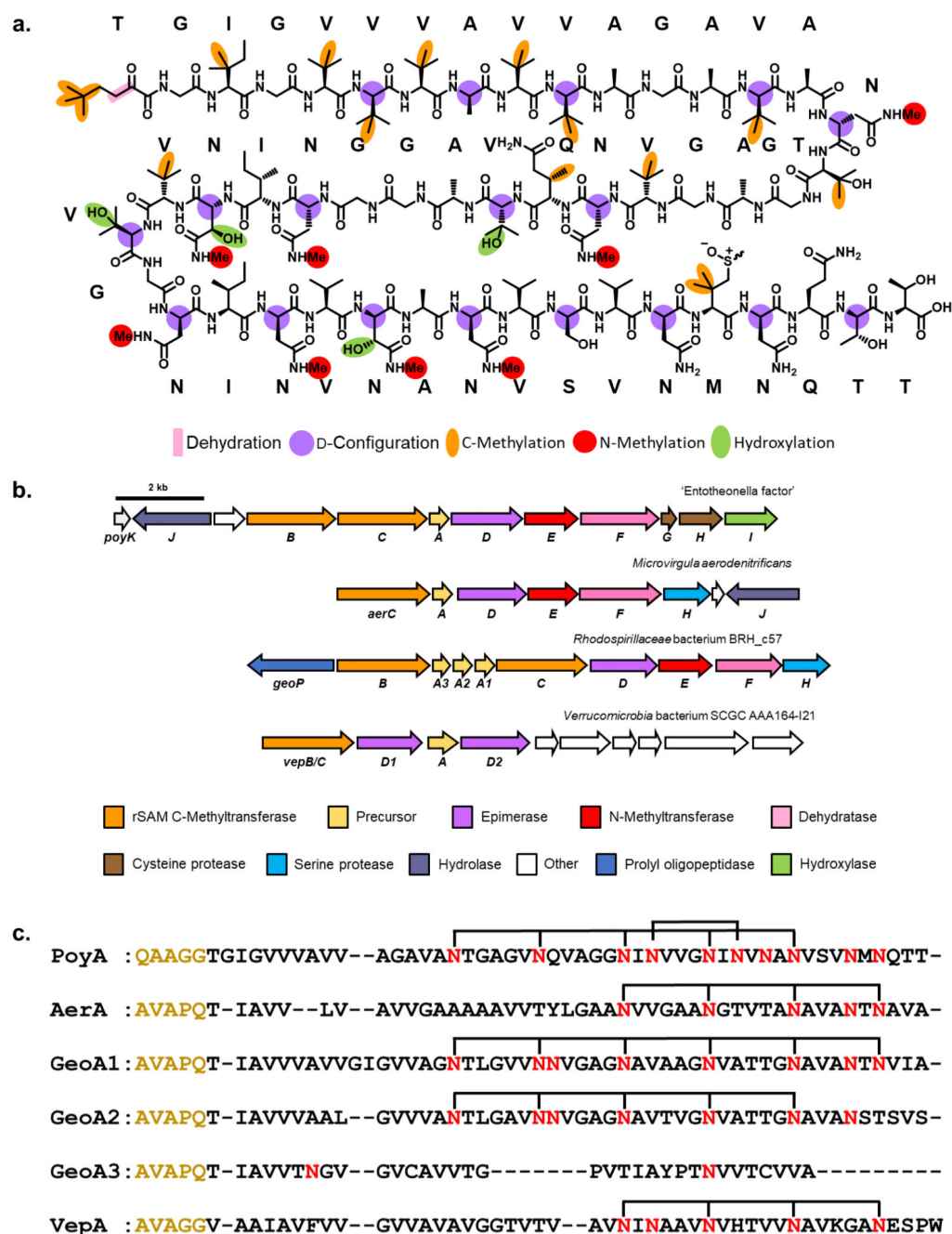
Figure 1. Structure and BGC of polytheonamides.
(a) Structures of polytheonamide A and B are shown, differing in the configuration of the methionine sulfoxide. Posttranslational modifications on the canonical amino acids are shown in the legend below. (b) Organization of the polytheonamide (*poy*) BGC and the clusters identified in this study (*aer, geo*, and *vep*). Gene colors indicate homologous encoded proteins with functions proposed in the legend. White arrows refer to genes unrelated to natural product biosynthesis. (c) Alignment of the core sequences encoded in all
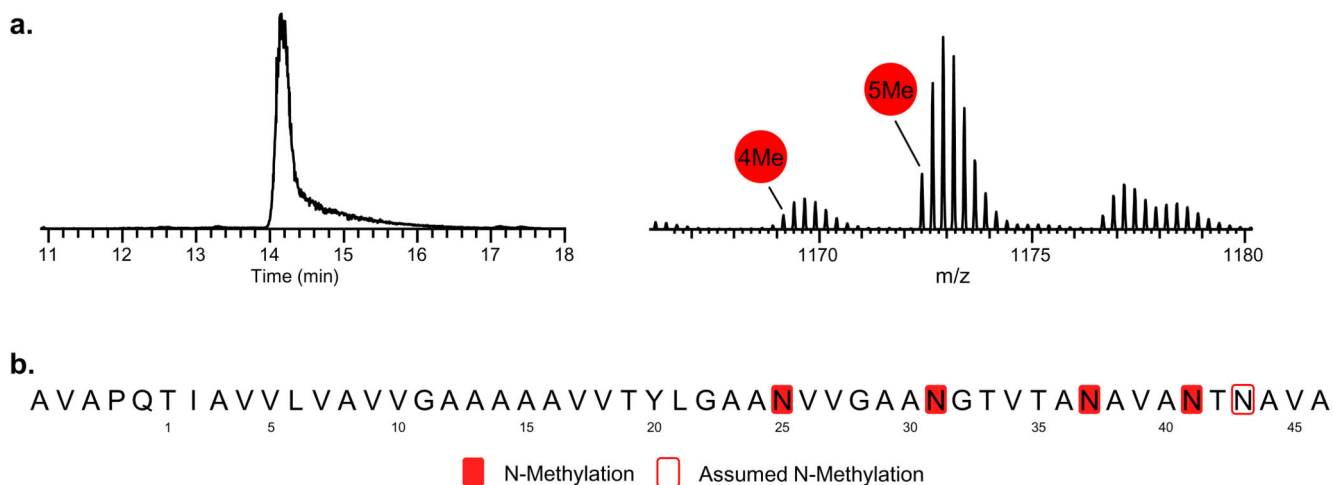
clusters with additional upstream residues in gold belonging to the leader C-terminus. Asn residues predicted to be N-methylated are in red, with predicted helix clamps shown.
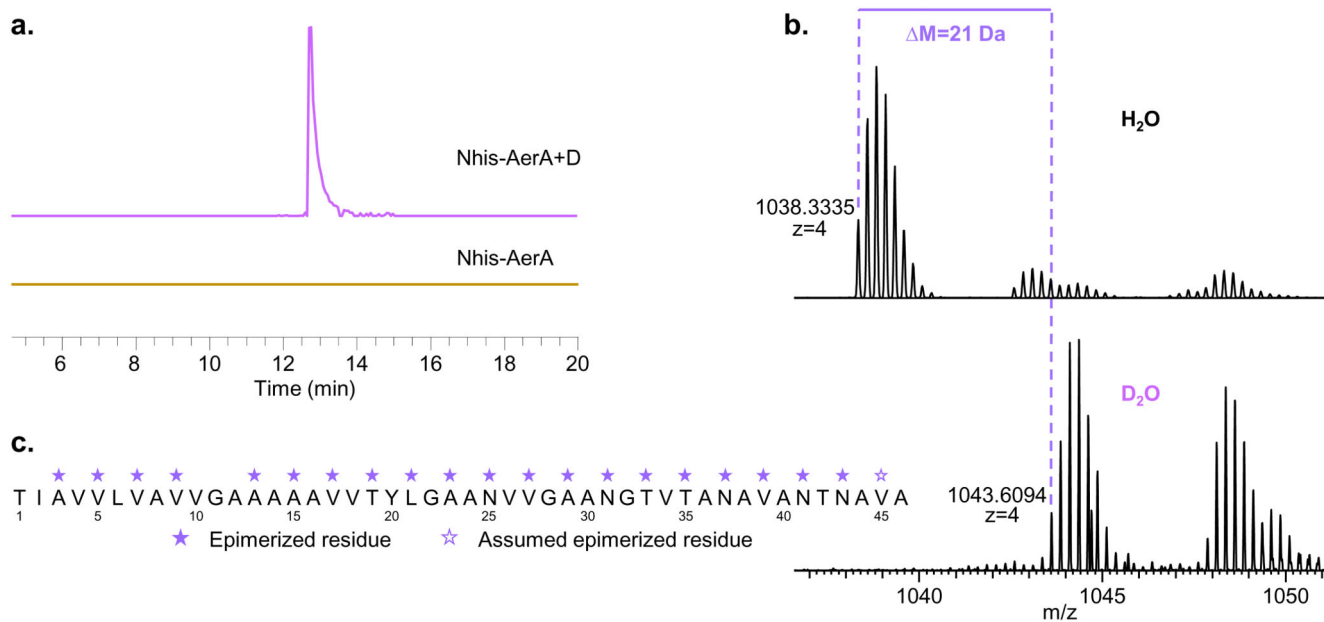
Figure 2. Epimerization of AerA in *E. coli*.
(a) Extracted ion chromatogram (EIC) looking for the protected AerA core (m/z = 1038.33 $[M+4H]^{4+}$ residues 1-46) after proteinase K digestion of Nhis-AerA and Nhis-AerAD purified from *E. coli*. **(b)** Ion of the protected core fragment (top) after expression in H_2O-based medium (top) compared with the corresponding ion after ODIS expression in D_2O (below). A mass shift of 21 Da was observed that was localized to the residues indicated by asterisks in **(c)**. Assumed modifications refer to modifications that could not be localized by MS^2 analysis to a particular residue, but for which ion fragments of smaller peptides contained the modification. Representative experiments were repeated independently two times with similar results, while ODIS was performed one time.

a.

b.

A V A P Q T I A V V L V A V V G A A A A A V V T Y L G A A N V V G A A N G T V T A N A V A N T N A V A

N-Methylation   Assumed N-Methylation

Figure 3. Identifying conditions for expression of the *aer* BGC.
(a) EIC following cell-free assays performed at induction conditions. The product peak in the chromatogram (left) at the top belongs to epimerized Nhis-AerA that was treated with M. aerodenitrificans cell-free extract and GluC (corresponding to 5 methylations). The corresponding mass spectrum is shown on the right. **(b)** Position of methylations located to core Asn residues based on MS^2 data (Supplementary Fig. 7). Methylation on Asn43 was not localized but proposed based on observed y-ion fragment masses. Representative experiments were repeated independently two times with similar results.
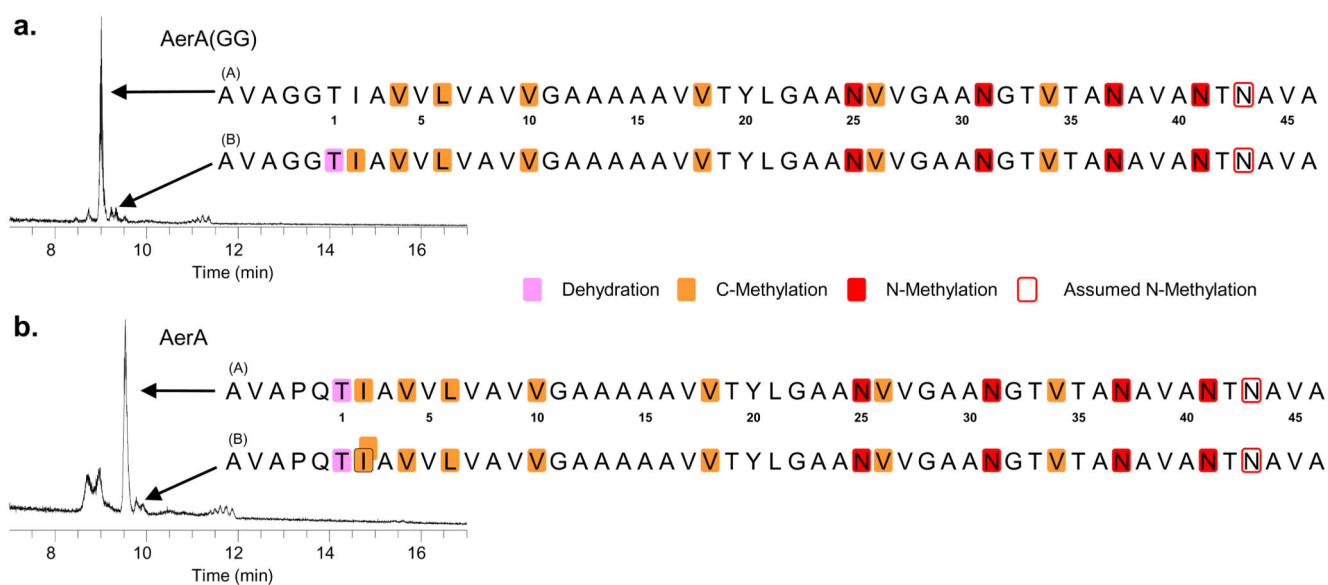
Figure 4. Hypermodified aeronamide peptides from expressions in *M. aerodenitrificans*.
(a) Total ion chromatogram (TIC) of GluC-treated Nhis-AerA**(GG)** after a two-day expression. A, major product; B, minor product (Supplementary Fig. 8). **(b)** TIC of GluC-treated Nhis-AerA after one day of expression. A, major product; B, minor product (Supplementary Fig. 9). Modifications localized to residues as described in the legend. Representative experiments were repeated independently two times with similar results.
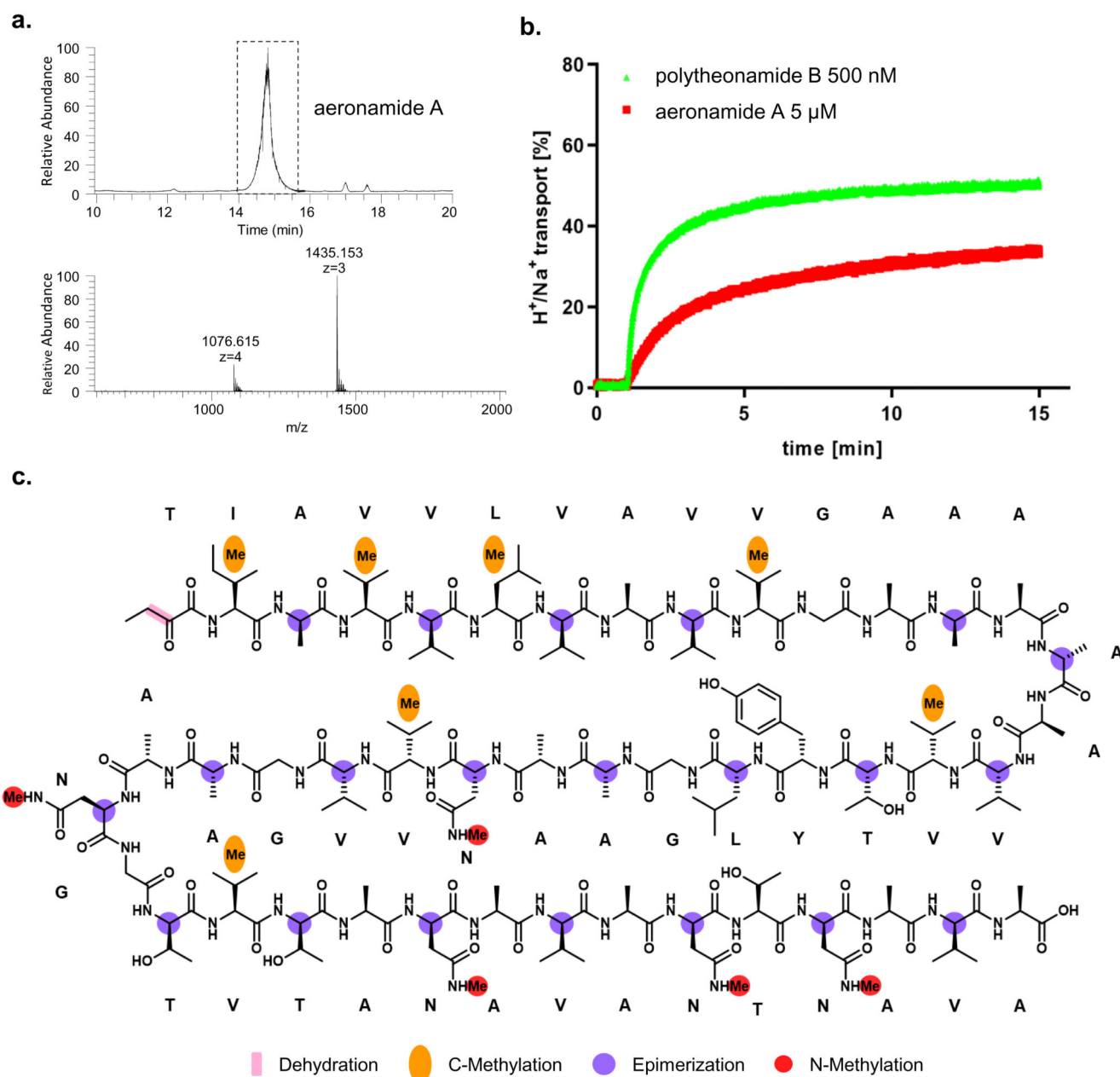
Figure 5. Generating aeronamide A and characterizing its ion-transport activity.
(a) TIC (top) and mass spectra (bottom) of HPLC purified aeronamide A following *in vitro* cleavage of Nhis-AerA (purified from *M. aerodenitrificans*) with Nhis-AerH (purified from *E. coli*). (b) Results of H^+/Na^+ ion exchange activity assay on artificial liposomes for aeronamide A and polytheonamide B. (c) Structure of aeronamide A. Modified residues are labeled according to the legend below. The orange balloons above the residues point to a methylation localized to this amino acid, but the exact position of the modification in the residue is unknown. Representative experiments were repeated independently two times with similar results.
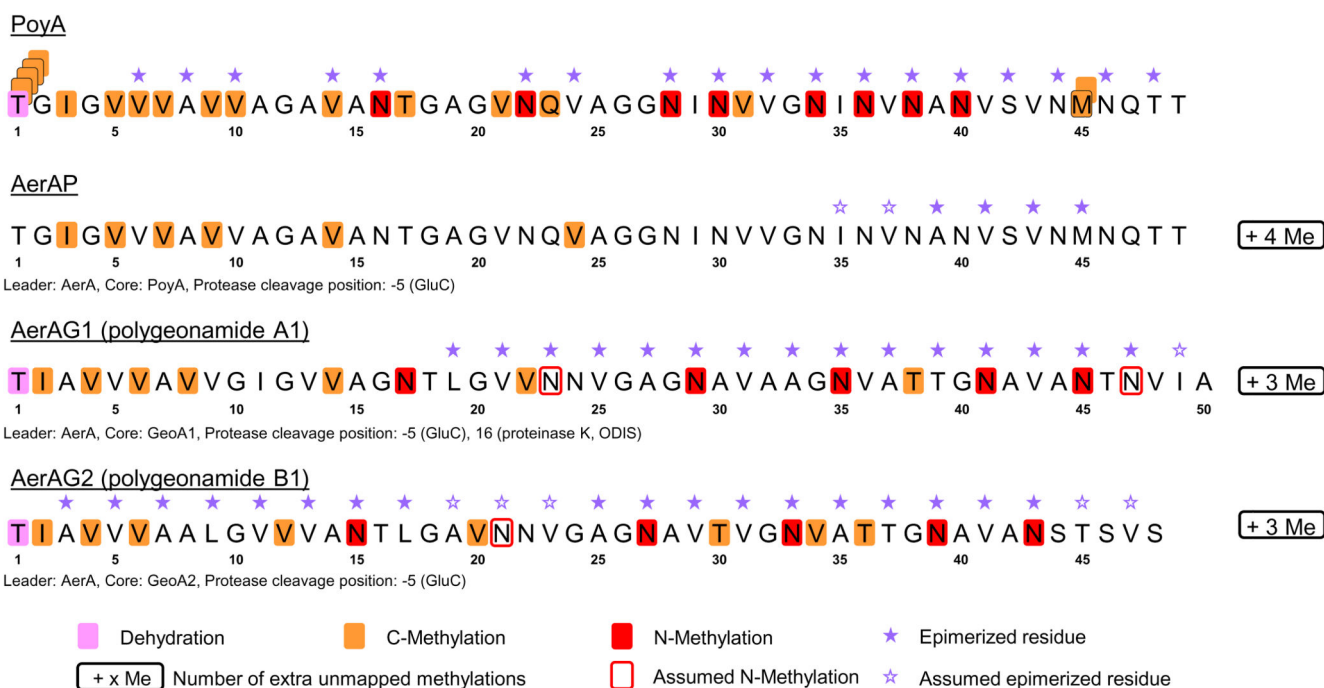
Figure 6. Peptides of non-native cores modified by the *M. aerodenitrificans* platform.
Epimerizations were localized via ODIS expressions in *E. coli*. Assumed modifications refer to modifications that could not be localized to a particular residue by MS^2 analysis, but for which fragmentation supported the existence of such a modification. For reference, all modifications that occur in polytheonamides (except hydroxylations) are shown on top (PoyA). AerAP refers to hybrid precursor carrying the aeronamide leader and the polytheonamide core, AerAG contains polygeonamide cores from the deep-rock biosphere. Additionally tested precursor variants with a GG instead of the PQ site show a similar modification pattern but lack the dehydration at Thr1. Leader, core and cleavage sites of proteases used to assign the modifications are indicated below each core sequence. Detailed MS^2 analyses localizing the various modifications are shown in Supplementary Figs. 14-17.