Organic & Biomolecular Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2023, **21**, 4893

Direct pathway cloning and expression of the radiosumin biosynthetic gene cluster†

Xiaodan Ouyang, [a] Paul M. D'Agostino, [b] Matti Wahlsten, [a] Endrews Delbaje, [c] Jouni Jokela, [a] Perttu Permi, [d,e] Greta Gaiani, [a] Antti Poso, [f] Pija Bartos, [f] Tobias A. M. Gulder, [b] Hannu Koistinen [g] and David P. Fewer *[a]

Radiosumins are a structurally diverse family of low molecular weight natural products that are produced by cyanobacteria and exhibit potent serine protease inhibition. Members of this family are dipeptides characterized by the presence of two similar non-proteinogenic amino acids. Here we used a comparative bioinformatic analysis to identify radiosumin biosynthetic gene clusters from the genomes of 13 filamentous cyanobacteria. We used direct pathway cloning to capture and express the entire 16.8 kb radiosumin biosynthetic gene cluster from *Dolichospermum planctonicum* UHCC 0167 in *Escherichia coli*. Bioinformatic analysis demonstrates that radiosumins represent a new group of chorismate-derived non-aromatic secondary metabolites. High-resolution liquid chromatography-mass spectrometry, nuclear magnetic resonance spectroscopy and chemical degradation analysis revealed that cyanobacteria produce a cocktail of novel radiosumins. We report the chemical structure of radiosumin D, an *N*-methyl dipeptide, containing a special Aayp (2-amino-3-(4-amino-2-cyclohexen-1-ylidene) propionic acid) with *R* configuration that differs from radiosumin A–C, an *N*-Me derivative of Aayp (Amyp) and two acetyl groups. Radiosumin C inhibits all three human trypsin isoforms at micromolar concentrations with preference for trypsin-1 and -3 (IC_{50} values from 1.7 μM to >7.2 μM). These results provide a biosynthetic logic to explore the genetic and chemical diversity of the radiosumin family and suggest that these natural products may be a source of drug leads for selective human serine proteases inhibitors.

Received 10th March 2023,
Accepted 5th May 2023

DOI: 10.1039/d3ob00385j

rsc.li/obc

Introduction

Natural products are a source and an inspiration for the modern pharmaceutical industry.[1–4] Protease inhibitors are an important class of pharmaceuticals and many of them are natural products or mimic natural products.[5] The first protease inhibitor approved by the U.S. Food and Drug Administration (FDA), an AIDS drug saquinavir,[5] and a proteasome inhibitor bortezomib used for treatment of multiple myeloma are natural product mimics.[5–7] Cyanobacteria produce numerous chemically diverse natural products that display potent protease inhibition activity.[8–12] The most common cyanobacteria protease inhibitors are the linear and cyclic peptides that belong to the aeruginosin,[13] anabaenopeptin[14] and cyanopeptolin[15] families of natural products.[16] Many of these compounds are active in the low nanomolar range.[9,10] These chemically diverse protease inhibitors are synthesized through biosynthetic pathways that encode non-ribosomal peptide synthetase (NRPS) proteins.[17,18] These large modular enzymes select and condense a range of proteinogenic and non-proteinogenic amino acids into small peptides.[18] Non-proteinogenic amino acids, generated in diverse specialized metabolic pathways, are an important source of precursors for these peptides.[19] Natural product biosynthetic pathways also encode an assortment of tailoring enzymes that promote structural diversification, including halogenation, acylation, and glycosylation.[18,20–24]

Radiosumins are a chemically diverse family of unique dipeptides that contain two unusual non-proteinogenic amino acid residues that are produced by distantly related genera of cyanobacteria.[25–28] Radiosumin A (**1**), composed of Aayp

[a]*Department of Microbiology, University of Helsinki, Viikinkaari 9, FI-00014 Helsinki, Finland. E-mail: david.fewer@helsinki.fi*
[b]*Chair of Technical Biochemistry, Technical University of Dresden, Bergstraße 66, 01069 Dresden, Germany*
[c]*Center for Nuclear Energy in Agriculture, University of São Paulo, Avenida Centenário 303, Piracicaba 13400-970, São Paulo, Brazil*
[d]*Department of Chemistry, University of Jyväskylä, Finland*
[e]*Department of Biological and Environmental Science, Nanoscience Center, University of Jyväskylä, Finland*
[f]*Kuopio Campus, School of Pharmacy, University of Eastern Finland, Finland*
[g]*Department of Clinical Chemistry and Haematology, Faculty of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki FIN-00014, Finland*
† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3ob00385j

(2-amino-3-(4-amino-2-cyclohexen-1-ylidene) propionic acid), Aacp (2-amino-3-(4-amino-2-cyclohexylidene) propionic acid) and two acetyl groups, was discovered from the cyanobacterium *Plectonema radiosum* NIES-515 (Fig. 1).[25] The absolute configuration of radiosumin A (**1**) was unambiguously determined by total synthesis.[26] Radiosumin B (**2**), an *N*-methyl dipeptide, that contains Aayp and an *N*-Me derivative of Aayp (Amyp) in addition to two acetate groups, was produced from a bloom of the cyanobacterium *Microcystis aeruginosa* (Fig. 1).[28] Radiosumin C (**3**) consisting of two units of Aayp and two acetyl groups, was isolated from the cyanobacterium *Anabaena cylindrica* NIES-19 (Fig. 1).[27] Radiosumin A (**1**) and C (**3**) are porcine trypsin inhibitors with submicromolar IC_{50} values,[25,27] while no bioactivity has been reported for radiosumin B (**2**).[28] Radiosumins are predicted to be synthesized from two acetylated amino acids derived from *p*-aminophenylalanine.[29] Considerations based on the chemical structure of radiosumin suggested that it would be formed in nature by the coupling of two amino acid units with the same absolute configuration.[27] However, the exact biosynthetic origins of radiosumins remain unclear.

The reduced cost and increasing ease of next-generation sequencing has resulted in a rapid increase in the coverage of the cyanobacterium phylum.[30–33] Analysis of this public data using specialized bioinformatic tools now facilitate the rapid identification of natural product biosynthetic pathways.[34–36] However, linking natural product biosynthetic pathways to their predicted biosynthetic products requires experimental validation, such as biochemical assays or knockout mutagenesis.[23,37–40] Although cyanobacteria are slow-growing photoautotrophs that are frequently intractable to genetic manipulation,[41–43] a heterologous approach offers new opportunities for functional biosynthetic gene cluster validation. Direct pathway cloning was recently developed to streamline the capture and heterologous expression of biosynthetic gene clusters allowing the rapid refactoring and expression of small to medium-sized natural product biosynthetic pathways into any vector system of choice.[2,44–46]

Here, we identified 13 complete putative radiosumin biosynthetic gene clusters from microbial genomes and experimentally confirmed our bioinformatic predictions through direct pathway cloning and heterologous expression of the entire radiosumin pathway in *Escherichia coli* BAP1. We report the discovery of a novel radiosumin chemical variant based on high-resolution liquid chromatography mass spectrometry (LC-HRMS), nuclear magnetic resonance (NMR) spectroscopy and chemical analysis. Bioactivity studies revealed that radiosumin C (**3**) inhibits human trypsin isoforms at micromolar concentrations. Overall, these results provide insights into long-standing questions on the biosynthesis of radiosumins and greatly expand our knowledge regarding natural product biosynthesis in cyanobacteria.

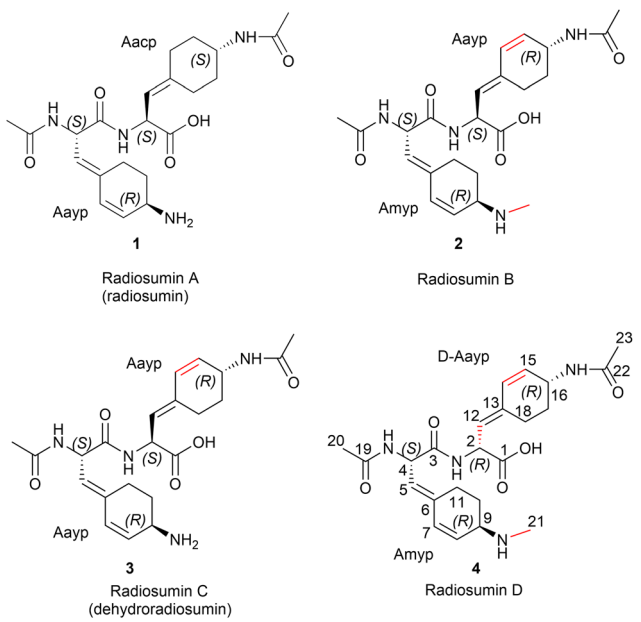

Fig. 1 Chemical structures of the four members of the radiosumin family of natural products. We propose renaming these chemical variants for convenience and present the old names of chemical variants in brackets. Differences between radiosumin A (**1**) and variants are marked by red colour.

Results and discussion

Identification of putative radiosumin biosynthetic gene clusters and screening of radiosumin producers

Radiosumin C (3) was originally reported from the freshwater cyanobacterium *A. cylindrica* NIES-19 in 1998.[27] We detected radiosumin C (**3**) (*m*/*z* 431.23, $[M + H]^+$) in *A. cylindrica* PCC 7122 by ultra-performance liquid chromatograph with quadrupole time-of-flight (UPLC-QTOF) spectrometry analysis, a pure strain co-identical to *A. cylindrica* NIES-19 (Fig. S1†).[32] The complete 7.1 Mb genome of *A. cylindrica* PCC 7122 is organized in a single chromosome and six plasmids.[32] We identified 15 secondary metabolite biosynthetic gene clusters encoded in the genome of *A. cylindrica* PCC 7122 using antiSMASH analysis.[34] We identified a putative 18.7 kb radiosumin biosynthetic gene cluster encoding nine proteins (Table 1), including two NRPS proteins with identical predicted adenylation domain Stachelhaus substrate specificity codes of DAEMSGGVLK (Table S1†).[47] Furthermore, this biosynthetic gene cluster encoded a predicted acetyltransferase and an amino transferase that we considered as likely to be involved in radiosumin biosynthesis based on its chemical structure (Table 1†). We annotated proteins encoded in the putative biosynthetic gene cluster as RadA-H (Table 1).

To explore the genetic diversity of the putative radiosumin biosynthetic pathways at the phylum level in cyanobacteria, we identified 11 further homologous putative radiosumin biosynthetic gene clusters from public databases from strains of *Dolichospermum* and *Aphanizomenon* using BLASTp searches using amino acid sequences of RadA-I enzymes from *A. cylindrica* PCC 7122 as a query (Fig. S2†). We obtained a draft genome of the freshwater cyanobacterium *P. radiosum* NIES-515, a radiosumin A (**1**) producer strain.[25] We also identi-

Table 1 Functional prediction of the proteins encoded in the radiosumin biosynthetic gene cluster from Anabaena cylindrica PCC 7122

Protein	Proposed function	aa	Homolog accession no.	Organism	Identity[d] (%)
RadA	4-Amino-4-deoxychorismate synthase	721	AAC44866[a]/F2RB79[b]	*Streptomyces pristinaespiralis*/*Streptomyces venezuelae* ATCC 10712	45.3/48.1
RadB	4-Amino-4-deoxyprephenate decarboxylase	198	A0A073CEA3[b]	*Planktothrix agardhii* NIVA-CYA 126/8	31
RadC	Isomerase	213	AerE[c]	*Microcystis aeruginosa strain* PCC7806	37.3
RadD	4-Amino-4-deoxychorismate mutase	96	AAC44868[a]/F2RB77[b]	*Streptomyces pristinaespiralis*/*Streptomyces venezuelae* ATCC 10712	40/40.9
RadE	Aminotransferase | *N*-acyltransferase	471	WP_052150012[a]	*Aphanizomenon flos-aquae*	85.8
RadG	Deacetylase	332	MBK5219837[a]	*Thermoleophilia bacterium*	44.0
RadI	A MATE efflux family protein	483	WP_196527063[a]	*Nostoc commune*	83.1
RadH1	Amino acid adenylation domain-containing protein	1100	WP_263746366	*Plectonema radiosum*	81.2
RadH2	Amino acid adenylation domain-containing protein	2041	WP_263746364	*Plectonema radiosum*	83.1

[a] GenBank. [b] UniProtKB/Swiss-Prot. [c] From reference article.[68] [d] Amino acid sequence identity between homolog and the proposed *rad* biosynthetic gene.

fied the putative radiosumin biosynthetic gene cluster from this strain (Fig. S2†), In total, we identified 13 putative radiosumin biosynthetic gene clusters (size range 16–19 kb) that encode 8–10 proteins (Fig. S2†). The radiosumin biosynthetic gene clusters exhibited a high degree of synteny and displayed a similar genetic organization (Fig. S2†). *A. cylindrica* PCC 7122 and *P. radiosum* NIES-515 both lack RadF, a methyltransferase, whilst *Dolichospermum flos-aquae* UHCC 0037 lacks RadI, a multidrug and toxic compound extrusion (MATE) efflux family protein (Fig. S2†). Each radiosumin biosynthetic gene cluster encodes two NRPS modules (Fig. S2†). However, these two modules are encoded as separate proteins in *Dolichospermum* sp. UHCC 0315A, *A. cylindrica* PCC 7122 and *P. radiosum* NIES-515 but are single bimodular proteins in the remaining cyanobacteria (Fig. S2†). Most radiosumin candidate producers were found in closely related genera of cyanobacteria and distributed through a phylogenomic tree based on 120 bacterial single-copy conserved marker proteins (Fig. S3†).[48]

We detected radiosumin A (**1**) (*m*/*z* 433.24, $[M + H]^+$) from *P. radiosum* NIES-515 and radiosumin C (**3**) (*m*/*z* 431.23, $[M + H]^+$) from *A. cylindrica* PCC 7122 (Fig. S1†). Given that there are four candidate radiosumin producer strains in the cyanobacterial collection at the University of Helsinki (UHCC), we decided to test these four candidate radiosumin producer strains for the presence of radiosumins. They were cultivated in corresponding medium at room temperature with continuous photon irradiation of 8 μmol m^{-2} s^{-1} for 30–40 days.[9,10,49,50] The methanol extracts from dried cell biomass were analysed with UPLC-QTOF. We identified a molecule (**4**) with a *m*/*z* 445.24 $[M + H]^+$ and an elemental formula of $C_{23}H_{33}N_4O_5^+$ that could correspond to radiosumin B (**2**) from *Aphanizomenon* sp. UHCC 0183, *D. planctonicum* UHCC 0167 and *Dolichospermum flos-aquae* UHCC 0037 (Fig. S4 and S5†). The MS^2 product ion profile of compound **4** is consistent with the annotated product ion structures of the radiosumin B (**2**) (Fig. S5†). Thus, we propose compound **4** may be a radiosumin B-like compound based on MS^2 data. We were unable to detect the expected radiosumin-like molecules from *Dolichospermum* sp. UHCC 0315A (data not shown).

Cyanobacteria blooms can consist of mixed assemblages even in cases where one member dominates the community.[51,52] Radiosumin B (**2**) was first reported from a bloom that was dominated by *Microcystis aeruginosa* but contained *Aphanizomenon* sp. as a minor constituent.[28] We did not detect the putative radiosumin biosynthetic gene cluster in any of the 175 public genomes of *Microcystis* (data not shown). The closely related diazotrophic genera, *Anabaena*, *Dolichospermum* and *Aphanizomenon* (ADA), frequently form toxic blooms in lakes and brackish waters where they can pose serious health risks to humans and animals.[52–55] We identified putative radiosumin biosynthetic gene clusters from the genomes of these bloom-forming cyanobacteria (Fig. S2†). Therefore, it is possible that the producer of radiosumin B (**2**) is a member of the ADA species complex that consists of difficult to distinguish members of the bloom-forming and potentially toxic cyanobacterial genera *Anabaena*, *Dolichospermum* and *Aphanizomenon*.[52–55] We also detected compound **4** in two bloom samples from Rajasaari (Helsinki, Finland) and Laaksolahti (Espoo, Finland) (Fig. S6†). The ADA species complex genera *Dolichospermum* and *Aphanizomenon* are known to dominate a large portion of the annual cyanobacterial blooms in the coastal regions of the Baltic Sea each summer.[56,57] Radiosumins were also reported from cyanobacterial blooms in man-made shallow fishponds in the Czech Republic.[58] Therefore, radiosumins may be a common component of blooms formed by members of this ADA species complex where they could act as protease inhibitors to affect grazers in freshwater ecosystems,[59] including some of the most important grazers of phytoplankton.[59,60]

Experimental validation of the radiosumin biosynthetic gene cluster in *Escherichia coli*

The biosynthetic mechanisms driving the chemical diversity of the most common classes of protease inhibitors in cyanobac-

teria are now known.[61] Direct pathway cloning was used to experimentally verify the involvement of the putative radiosumin biosynthetic gene cluster in radiosumin biosynthesis (Table S2 and S3†). We selected the radiosumin biosynthetic gene cluster from *D. planctonicum* UHCC 0167 for heterologous expression because we obtained high quality genomic DNA from this strain (data not shown). The 16.8 kb *radHABCDEFG* gene cluster from *D. planctonicum* UHCC 0167 was inserted into the pET28b-ptetO-*GFP*v2 expression vector using sequence- and ligation independent cloning (SLIC),[46] resulting in the pET28b-ptetO-*radHABCDEFG-GFP* expression construct (Fig. 2A). The biosynthetic gene cluster was placed under control of the $ptet_O$ tetracycline inducible promoter. Heterologous expression of several cyanobacterial biosynthetic gene clusters has been successful using this promoter.[2,44–46,62] *E. coli* BAP1 was used as the expression host and is a derivative of *E. coli* BL21(DE3) with genomic integration of the phosphopantetheinyl transferase *sfp*, thus promoting posttranslational activation of NRPS carrier proteins.[63] Successful translation was observed by fluorescence microscopy using the GFP reporter gene that is integrated into the applied vector backbone and positioned downstream of the radiosumin pathway (Fig. 2B).[64]

After cultivation in M9 mineral medium for three days, the culture supernatants were collected and freeze-dried. Organic solvent extraction followed by UPLC-QTOF analysis revealed a prominent peak of a compound with a retention time of 4.70 min and *m*/*z* 445.24 $[M + H]^+$, from the culture supernatant of the strain with pET28b-ptetO-*radHABCDEFG-GFP* (Fig. 2C). The peak was absent in the sample from the negative control strain with empty pET28b-ptetO-*GFP*v2 vector. A peak of compound **4** with identical retention time, *m*/*z* and product ion spectrum was observed from the pellet extract of *D. planctonicum* UHCC 0167, here used as a positive control (Fig. 2C and D). Therefore, compound **4** was produced in *E. coli* successfully. Although the production level of the compound **4** was much lower in *E. coli* BAP1 than in *D. planctonicum* UHCC 0167 when it was normalized with dry cell weights, *E. coli* grows much faster than cyanobacteria and offers a viable alternative for compound **4** production. The pET28b-ptetO-*radHABCDEFG-GFP* was then linearized and RadI was inserted using SLIC to construct another 24.7 kb expression vector pET28b-ptetO-*radHABCDEFGI-GFP*. The expression of RadI, the MATE efflux family protein that might involve in the transport of **4** across cellular membranes in cyanobacteria,[65] was found to increase the production of compound **4** (Fig. S7†). The low availability of chorismate, a precursor for various kinds of aromatic compounds produced in *E. coli* (such as L-phenylalanine, L-tyrosine, L-tryptophan, folate, ubiquinone, and siderophores),[66,67] likely limits the metabolic flux towards **4**. Thus, competition between chorismate-utilizing enzymes from *E. coli* may explain the lower radiosumin production level in the heterologous host.

Isolation and structure elucidation of a novel radiosumin variant

We purified 1.3 mg of compound **4** (*m*/*z* 445.24, $[M + H]^+$), which was also detected from *D. planctonicum* UHCC 0167 and *D. flos-aquae* UHCC 0037 (Fig. S4 and S5†), directly from 3 g of dried biomass from *Aphanizomenon* sp. UHCC 0183. NMR chemical shifts and atom connections from COSY, HSQC and HMBC NMR spectra matched with the radiosumin B (**2**) structure (Table 2 and Fig. S8–S12†). However, amino acid Aayp carbon chemical shifts at C-1, -12, -13, -15, as well as CH-2 and NH-2 proton chemical shifts and C-3 chemical shift of the Amyp building block differed significantly from radiosumin B (**2**) values,[28] indicating structural differences in amino acid Aayp (Table 2). Amino acid analysis by the Marfey method from oxidized and acid hydrolysed **4** yielded both L- and D-Ser (Fig. S13†), indicating that Aayp in this molecule is in the *R* configuration, which differs from the prototypical *S* stereochemistry of all known radiosumins A–C (**1**–**3**) (Fig. 1).[25,27,28] We therefore named **4** as radiosumin D. As previously reported, the chemical structure of radiosumin A (**1**) has been described by Matsuda and co-workers using mass spectrometry, 2D NMR and chemical degradation methods.[25] The absolute chemical structure and stereochemistry of radiosumin A (**1**) was precisely determined by total synthesis (Fig. 1).[26] Subsequently, radiosumin C (**3**), which has identical stereochemistry to radiosumin A (**1**), was analysed with the same methods as used for radiosumin A (**1**).[27] The slightly larger radiosumin B differs by having one extra methyl group when compared with radiosumin C (**3**) (Fig. 1).[28] In conclusion, based on mass spectrometric and amino acid analyses and NMR chemical shift comparisons, radiosumin D (**4**) is a novel radiosumin, a stereoisomer of radiosumin B (**2**) with an identical molecular formula but different Aayp stereochemistry when compared with radiosumins A (**1**), B (**1**) and C (**3**) (Fig. 1).

Proposed radiosumin biosynthesis

We propose the following biosynthetic scheme for radiosumin formation in *D. planctonicum* UHCC 0167 based on bioinformatic analysis and the chemical structure of radiosumin D (**4**) (Table 1 and Fig. 3). The 4-amino-4-deoxychorismate synthase homolog RadA catalyzes substitution of the secondary alcohol in chorismic acid (**5**) to the corresponding amine under retention of configuration, with glutamine as the nitrogen source, to form *p*-aminochorismic acid (4-amino-4-deoxychorismic acid) (**6**) (Fig. 3).[39,69] This is a well-characterized process as *p*-aminochorismic acid (**6**) is a key intermediate in the biosynthesis of *p*-aminobenzoic acid (PABA) and L-(*p*-aminophenyl) alanine.[70–73] Next, the 4-amino-4-deoxychorismate mutase RadD catalyzes the conversion of *p*-aminochorismic acid (**6**) to *p*-aminoprephenic acid (4-amino-4-deoxyprephenic acid) (**7**) (Fig. 3). This reaction sequence has exact precedence in both the pristinamycin IA and chloramphenicol biosynthetic pathways.[39,69,74] The 4-amino-4-deoxyprephenate decarboxylase RadB then catalyzes the decarboxylation of *p*-aminoprephenic acid (**7**) to produce intermediate **8** (Fig. 3). While such decarboxylation chemistry in chorismate-derived metabolites usually leads to aromatization (*e.g.*, in pristinamycin IA and chloramphenicol biosynthesis in secondary metabolism;[39,69] and in the bio-

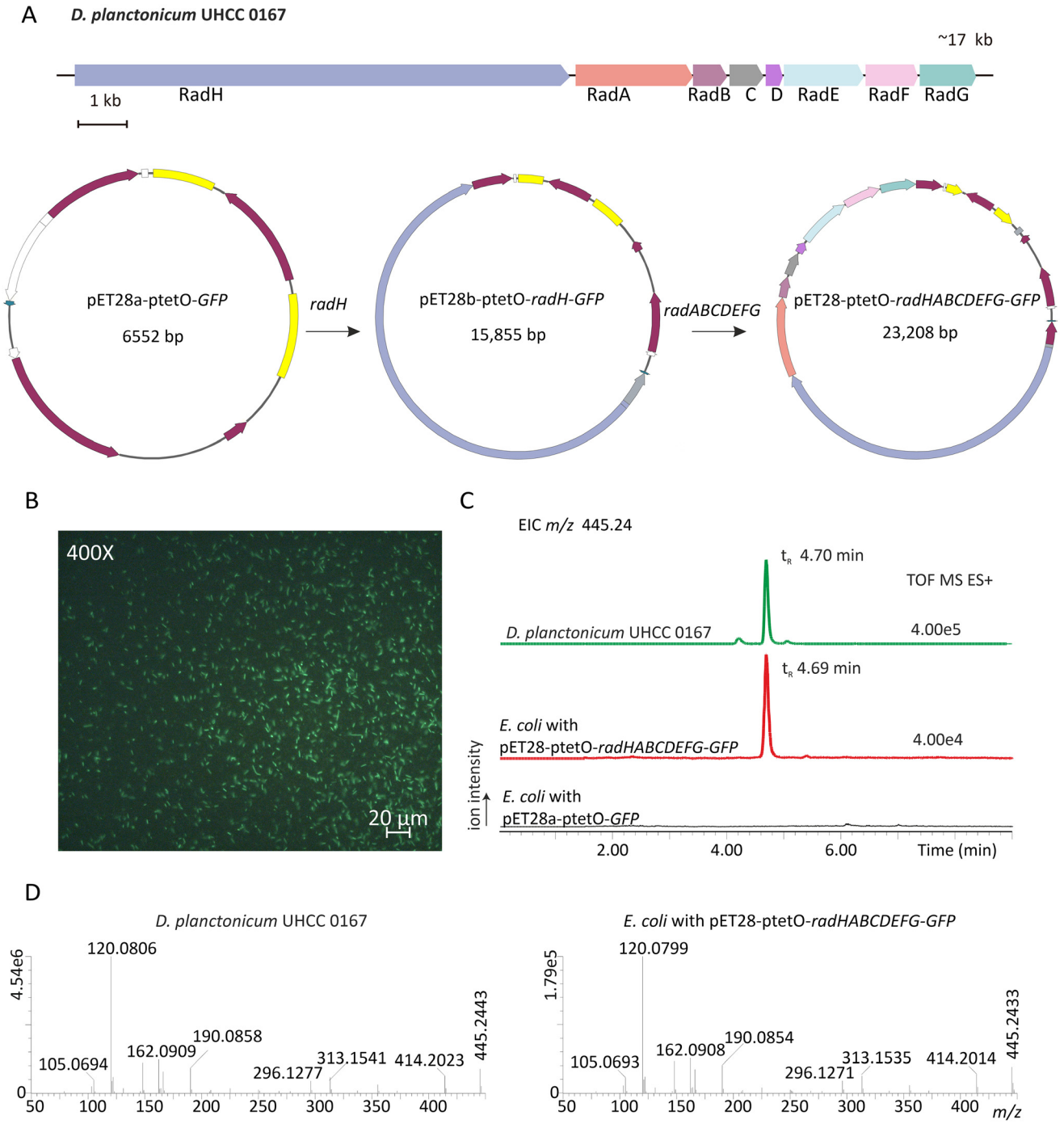

Fig. 2 Direct pathway cloning of the 16.8 kb radiosumin biosynthetic gene cluster from *D. planctonicum* UHCC 0167 and heterologous expression in *E. coli* BAP1 (A) The *radHABCDEFG* biosynthetic gene cluster was cloned into the pET28b-ptetO-*GFP*v2 vector under the control of $ptet_O$ promoter and with a *GFP* transcriptional reporter gene by SLIC. (B) Fluorescence microscopic analysis of *E. coli* BAP1 harboring pET28b-ptetO-*radHABCDEFG-GFP* after three days induction with 0.5 µg mL^{-1} tetracycline. *E. coli* cells were visualized using a fluorescence microscope (×400 magnification, excitation at 475–490 nm and emission at 505–535 nm). (C) Extracted ion (*m/z* 445.24 $[M + H]^+$) chromatograms of methanol extracts of *D. planctonicum* UHCC 0167 and *E. coli* with pET28b-ptetO-*radHABCDEFG-GFP* or pET28b-ptetO-*GFP* empty vector. Mass spectra of the 4.70 min peaks showed protonated compound **4** with −0.2 and −1.4 ppm error. (D) UPLC-QTOF secondary mass spectrometry analysis. MS^2 spectra (parent ion *m/z* 445.24) from *m/z* 50 to 450 of the 4.70 min peak showed diagnostic product ions for compound **4** (Fig. S5†).

genesis of phenylalanine and tyrosine from phenylpyruvate and 4-hydroxyphenylpyruvate in primary metabolism[75,76]), non-aromatizing prephenate decarboxylases do also have precedence in natural product biosynthesis. These include BacA, AerD, SalX and Plu3043, in the biosynthesis of bacilysin, aeruginoside 126A and B, salinosporamide A, and

Table 2 NMR data of radiosumin D in CD_3OD and comparison of shift values with data from Coleman and Wright[28]

Pos	δ_C	δ_H	Multi, J (Hz)	HMBC	ROESY	$\Delta\delta_C$	$\Delta\delta_H$
1	176.7					2.9	
2	54.5	4.97	d 9.2	1, 3, 12, 13	18, 18′, 12	2.4	−0.18
NH-2		7.72		—			−0.66
3	170.4					−1.4	
4	52.7	5.22	d 9.4	3, 5, 6, 19	11, 11′, 5	0.0	−0.01
NH-4		8.42		—			0.01
5	127.2	5.50	d 9.6	3, 4, 7, 11	4, 7	0.0	0.01
6	137.7					−0.1	
7	136.4	6.41	dd 10.0, 1.4	5, 6, 9, 11	5, 8	−0.2	0.02
8	124.1	5.84	dd 10.0, 2.1	6, 9, 10	7	−0.1	0.02
9	55.7	3.88	m	7, 8, 10	8, 10′, 21	−0.2	0.01
NH-9		—		—			
10	25.7	1.73	m	9, 11	10′	−0.1	0.02
10′		2.22	m	6, 8, 9, 11	10		
11	23.0	2.51	m	—	11′	0.0	0.04
11′		2.84	m	6, 7, 9, 10	4, 11		
12	126.9	5.28	d 9.4	1, 2, 14, 18	14	3.9	−0.08
13	137.8					−2.2	
14	133.4	6.11	dd 10.0, 1.8	12, 13, 16, 18	12. 15	0.8	−0.03
15	129.9	5.62	dd 9.8, 3.5	13, 17	14, 16	−2.0	−0.07
16	46.3	4.44	m	14, 15, 17	15, 17, 17′, 18	−0.3	0.00
NH-16		8.14		—			−0.09
17	29.3	1.61	m 4.2	13, 15, 16, 18	17′, 18′	−0.1	0.05
17′		1.90	m 4.2	13, 15, 16, 18	17, 18		−0.03
18	23.4	2.45	m	12, 13, 17	17′, 18′	−0.5	0.11
18′		2.73	m	12, 13, 14, 16, 17	18		0.07
19	172.6					−0.1	
20	22.3	2.00	s	4, 19		0.0	0.03
21	30.4	2.72	s	9	8, 9	−0.3	0.00
22	172.4					−0.2	
23	22.4	1.94	s	22		0.0	0.01

Δδ represents δ (this study)- δ (C&W).

dihydrostilbene.[38,76,77] RadB thus likely represents a new member of the family of non-aromatizing *p*-aminoprephenate decarboxylases. RadE has two domains that putatively function as an aminotransferase and *N*-acyltransferase and thus is likely to catalyze the reaction sequence from intermediate **8** into **9** and **10** (Fig. 3). The isomerase RadC is a homolog of AerE from *Microcystis aeruginosa* strain PCC 7806. The latter was shown to accelerate 1,3-allylic isomerization of the non-aromatic decarboxylation product of prephenate, dihydro-4-hydroxyphenylpyruvate (H_2HPP).[68] The product of isomerized H_2HPP shares structural similarity to intermediate **11** (Fig. 3). Thus, RadC is proposed to be responsible for the conversion of intermediate **10** to **11** (Fig. 3). This step is followed by action of the deacetylase RadG, which catalyzes cleavage of the *N*-acetyl group resulting in intermediate **12** (Fig. 3). The latter is methylated by the methyltransferase RadF to produce intermediate **13** (Fig. 3).

The RadH protein encodes a starter condensation domain (Fig. S2,† and Fig. 3). Starter domains are found on initiation modules in NRPS proteins where they are responsible for acylation of the first amino acid incorporated in natural product biosynthetic pathways.[78] For example, starter condensation domains install β-hydroxyl fatty acid in the biosynthesis of the lipopeptides surfactin,[79] lichenysin,[80] fengycin[81] and arthrofactin[82] by *N*-acetylation. We propose that RadH starter condensation domains are responsible for the *N*-acetylation of radiosumins (Fig. 3). It is quite common that non-ribosomal peptides show variation at the N-terminus, especially regarding acylation, which can protect the N-terminus from degradation, to modulate polarity, and to confer specific properties such as membrane insertion.[18] The RadH protein encodes two modules with almost identical adenylation domain substrate specificity codes, DAEMAGGVLK and DAETSGGVLK (Fig. 3 and Table S1†). RadH likely activates similar compounds, non-proteinogenic amino acids **11** and **13**, to produce various radiosumin analogues as a last step of synthesis (Fig. 3). For example, the low efficiency of the epimerase domain in RadH could lead to analogues with different stereochemistry. However, the exact order and timing of biosynthetic reactions needed to produce the radiosumins will require more detailed experimental validation. The truncated RadH protein from *Dolichospermum* sp. UHCC 0315A, which is split into two proteins, lacks the AMP-binding C-terminal domain, phosphopantetheine attachment site and TubC N-terminal docking domain (Fig. S2B†). In NRPS adenylation domains, the catalytic residue is positioned on a loop in the C-terminal domain, the phosphopantetheine attachment site is necessary for the transporter units of NRPSs, and multiple NRPS subunits interact with each other in a specific linear order mediated by specific docking domains.[83,84] Therefore, the radiosumin bio-

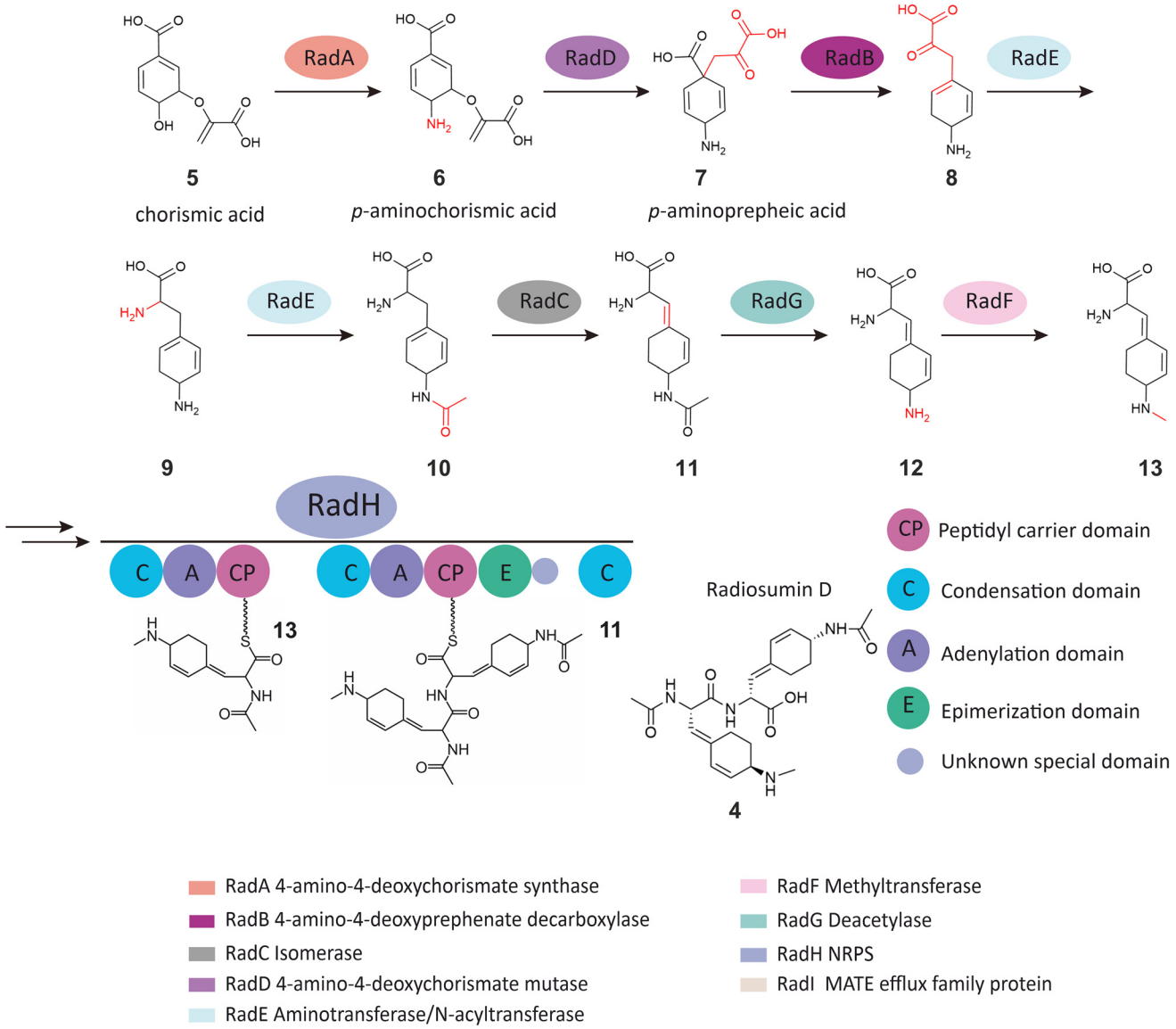

Fig. 3 Putative pathway for the synthesis of radiosumin D in *Dolichospermum planctonicum* UHCC 0167.

synthetic gene cluster from *Dolichospermum* sp. UHCC 0315A is likely to be a naturally inactive mutant.

Bioactivity of radiosumins

Proteases are ubiquitous enzymes expressed in cells and regulate a wide range of different cellular functions.[85–87] Trypsin, one of the best-characterized serine proteases, is not only involved in physiological processes, but is also important in pathological processes, such as atherosclerosis, inflammation, and cancer.[88,89] Previous studies have shown that trypsin promotes proliferation, invasion, and metastatic dissemination of cancer cells.[90–93] In humans, there are three genes (*PRSS1*, *PRSS2* and *PRSS3*) that encode the trypsin-precursors trypsinogen-1, -2, and -3, respectively.[88,90,94] Trypsin-3 has been reported as a potential target for cancer therapy, to repress cancer progression and metastatic dissemination in prostate,[87,95] breast,[96] pancreatic,[97] and lung cancers.[92] Highly similar structures of all three human trypsin-isoenzymes and minor differences between the substrate specificities make it challenging but possible to find selective inhibitors for trypsin-3.[9,10,98] Metabolites derived from cyanobacteria have often been tested with bovine and porcine trypsins, whose structures are different from human trypsins.[99] Thus, such studies are not very helpful for the development of selective inhibitors of human trypsin-isoproteases.[99] Among cyanobacterial peptides, only three aeruginosin family trypsin inhibitors, *i.e.*, suomilide and two varlaxins, have been tested with human trypsins.[9,10] While radiosumin A (**1**) and radiosumin C (**3**) were reported to be inhibitors of porcine pancreatic trypsin[100] with IC_{50} values of 0.32 μM and 0.23 μM

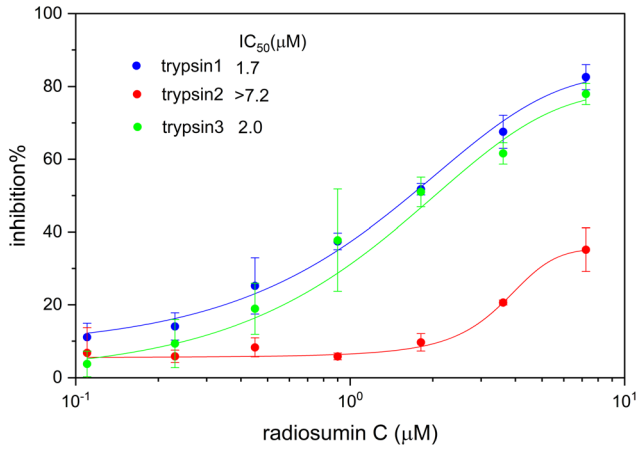

Fig. 4 Inhibition of human trypsin-isoenzymes by radiosumin C (**3**).

respectively,[25,27] those have not been studied using human trypsins. To explore whether radiosumins inhibit human trypsins, radiosumin C (**3**) and radiosumin D (**4**) were purified from *A. cylindrica* PCC 7122 and *Aphanizomenon* sp. UHCC 0183, respectively. They were tested for inhibition of proteolytic activity of recombinant human trypsin-1, -2 and -3 using a chromogenic trypsin substrate.[9,10,101] Radiosumin C (**3**) was found to inhibit trypsin-1 and -3 with IC_{50} values of 1.7 μM and 2.0 μM, respectively, while trypsin-2 was less potently inhibited (IC_{50} value >7.2 μM) (Fig. 4). The radiosumin D (**4**) preparation also showed trypsin-inhibitory activity, with similar preference for trypsin-1 and -3 inhibition (Fig. S14A†), however, this may be partly or entirely due to impurities in the preparation. The purified radiosumin D (**4**) sample from *D. planctonicum* UHCC 0183 included 11% impurities (Fig. S14B†). Based on MS^2 ion fragmentation data analysis, we assume that the larger impurity at *m*/*z* 445.24 represents radiosumin B (**2**) (Fig. S14B†). The less than 100% efficiency of the epimerization domain of the NRPS yields a mixture of radiosumin B (**2**) and D (**4**).[102] While radiosumin B (**2**) activity against trypsins has not been studied previously,[28] it is possible that it contributes to the detected trypsin-inhibitory activity of the radiosumin D (**4**) preparation. The different stereochemistry of radiosumin D (**4**) from radiosumin B (**2**) and C (**3**) may be less favourable to trypsin inhibition. Thus, regarding radiosumin D (**4**), we can only conclude that it is not as potent inhibitor of human trypsins as radiosumin C (**3**). The preferential inhibition of trypsin-1 and -3 by radiosumin C (**3**) is interesting as, other cyanobacterial compounds we have tested (suomilide and varlaxins)[9,10] inhibited trypsin-2 and -3 more potently than trypsin-1. This could result from the differences in active sites and specificity of those trypsin-isoenzymes.[98,101] Three small molecules with the bis-benzamidine substructure were reported to inhibit human trypsins based on a structure-based docking screen.[103] However, these molecules either inhibited all three human trypsins equally or preferentially inhibited trypsin-1 and -2.[9,103] Considering the development of trypsin-3 inhibitory drugs, radiosumins, together with other cyanobacterial compounds, may aid identification of inhibitors that are highly selective for different trypsin-isoenzymes and, thus, can be used as lead molecules for the drug development.

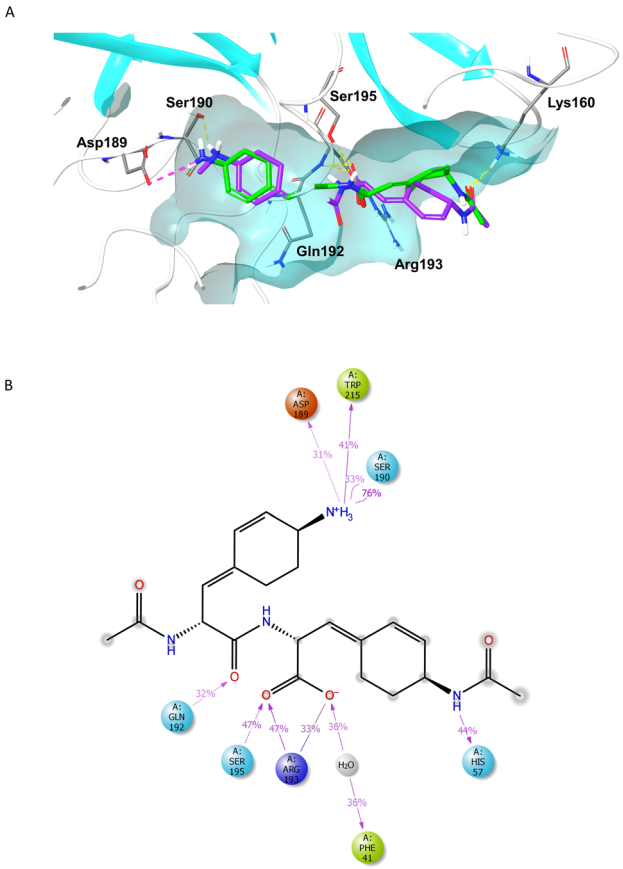

Fig. 5 Computational studies of trypsin-3 interactions with radiosumins. (A) The binding modes of radiosumin C (**3**) and D (**4**) on trypsin-3 structure. Radiosumin C (**3**) shown in green and radiosumin D (**4**) in purple. The stereochemistry of radiosumin D (**4**) forces the cyclohexene–acetamide moiety away from the protein, towards the solvent. This leads to a loss of hydrophobic interactions which partially causes the lower inhibitor activity. The inhibitory activity is also lowered by the methyl substituent at the Asp189 and Ser190 binding amino group, which lowers the electrostatic interactions between these residues and radiosumin D (**4**). (B) The most common interactions of trypsin-3 with radiosumin C (**3**) in the molecular dynamic simulations. The percentages correspond to the frequency of the interaction in the simulations. In case there are two different kinds of interactions (*e.g.*, ionic and hydrogen bonding with Ser190), two interaction arrows are displayed. The dotted arrows depict interaction to the sidechain and the solid arrows depict interaction to the main chain. The atoms with darker background are exposed to the solvent.

Modelling of the interaction between radiosumins and human trypsins

Molecular docking was used to check the binding poses of the inhibitors and the stability of the binding was studied with molecular dynamics (MD) simulations of inhibitor-trypsin complexes (Fig. 5). All radiosumins were successfully docked to trypsin-2 and trypsin-3 models. The structure of trypsin-1 is slightly more closed, which creates challenges in the calcu-

lations, particularly in the docking of radiosumin D (**4**) (data not shown). If the docking to trypsin-1 failed, the docking pose for the simulations was taken from the trypsin-3 results. Due to differences in stereochemistry, radiosumin C (**3**) showed higher docking scores with all trypsins and displayed better complementarity with the binding site than radiosumin (Fig. 5A). The stereochemistry of radiosumin D (**4**) forces the cyclohexene–acetamide moiety away from the protein, towards the solvent (Fig. 5). This lowers the hydrophobic interaction surface with the protein which partially explains the lower inhibitory activity of radiosumin D (**4**). The high root-mean-square deviations (RMSD) of the MD simulations suggest that the binding mode of the radiosumins is not very stable, but instead generally fluctuates (Table S4†). Of note, suomilide showed highly stable binding in MD simulations, which was also confirmed experimentally.[9] The high RMSD of radiosumins with trypsins is not caused by protein conformational changes, as the protein backbone does not move considerably during the simulations (RMSD < 2 Å) (Table S4†). Generally, the RMSD of the ligand stays below 10 Å from the starting pose, which corresponds to the other parts of the ligand rotating, while the aminocyclohexene moiety remains in its spot interacting with Asp189, Trp215 and Ser190 (trypsin-3 numbering) almost all the time in all simulations (Fig. 5, and Table S5†). The Asp189-amino interaction is the only ionic interaction in the complexes and this also partially explain the slightly higher inhibitory activity of radiosumin C (**3**) compared to radiosumin D (**4**) (Table S5†). The amino group is methylated in radiosumin D (**4**), which reduces the interaction possibilities (Fig. 5A). Whereas radiosumin C (**3**) has ionic interactions or hydrogen bonding with both Asp189 and Ser190 throughout the simulation (Fig. 5B), these interactions are less common in the radiosumin D (**4**) simulations (Table S5†). Asp189 and Ser190 have previously been shown to be important for the interaction between trypsin-3 and the trypsin inhibitor suomilide.[9] Asp189 located at the base of the specificity pocket of trypsins, is also particularly vital to the interaction between trypsin-3 and the small trypsin inhibitor diminazene or peptide inhibitors with 10 amino acids.[101,103] These results also highlight the importance of pH for the inhibitory activity of radiosumins. At lower pH, the amino group is more likely ionized and forms a more stable complex, and thus the inhibitory activity should be better. We used pH 7.6 in biological activity assays; the lower pH condition could be tested in future studies. These modelling analysis results suggest that the higher activity of radiosumin C (**3**) is due to its unsubstituted amino group and better fitting stereochemistry. However, modelling was not able to explain the different potencies of radiosumins for inhibition of different human trypsin-isoenzymes.[9] It is rather common that the relatively small differences in inhibitory activity (within one order of magnitude) cannot be explained by computational simulations. Previous reports and data taken together show the importance of the selective human trypsin inhibitors, especially for inhibitors targeting trypsin-3 related to different cancers.[87,92,95–97] Evette S. Radisky's group has made efforts to identify modified natural serine protease inhibitors with improved affinity to trypsin-3 based on structure-based and directed evolution protein engineering.[103–106]

Experimental

Strains, media and genetic materials

Bacterial strains and plasmids constructed in this study are listed (Table S2†). *E. coli* DH5α was used for gene cloning and plasmid maintenance. *E. coli* BAP1 was used for gene cluster expression. *D. planctonicum* UHCC 0167, *D. flos-aquae* UHCC 0037, *Aphanizomenon* sp. UHCC 0183 and *Dolichospermum* sp. UHCC 0315A were obtained from the University of Helsinki culture collection (UHCC, Finland), *P. radiosum* NIES-515 was purchased from the Microbial Culture Collection at the National Institute for Environmental Studies (NIES Collection, Tsukuba, Japan) and *A. cylindrica* PCC 7122 was purchased from the Pasteur Culture Collection of Cyanobacteria (PCC, France). All recombinant *E. coli* strains were cultured in Luria–Bertani (LB) medium or M9 mineral medium with 50 μg mL^{-1} kanamycin. *D. planctonicum* UHCC 0167 was grown in Z8 medium at room temperature with continuous photon irradiation of 8 μmol m^{-2} s^{-1}. *P. radiosum* NIES-515 was grown in 50 mL Csi medium.[107] *Aphanizomenon* sp. UHCC 0183 and *A. cylindrica* PCC 7122 were grown in 50 mL Z8 medium without nitrogen, under the same conditions as *D. planctonicum* UHCC 0167. In the summer of 2022, naturally occurring cyanobacterial blooms were sampled from two locations in Finland: Rajasaari (Helsinki, 60°10′59.1″N 24°54′35.7″E) and Laaksolahti (Espoo, 60°14′47.4″N 24°45′00.0″E). The samples were designated as GG20 and GG68, respectively. *In situ*, a volume of 1 L of water from the water surface was collected for each sample using a 3 L bucket. In the laboratory, the samples were transferred to 1 L graduated cylinders and allowed to decant overnight at 4 °C. The biomass that floated to the top was subsequently carefully transferred into plastic boxes (Orthex™; dimensions: 12 × 12 × 4 cm) and stored at −20 °C until freeze-drying.

Identification of radiosumin biosynthetic gene clusters

The complete genome data of *A. cylindrica* PCC 7122 was downloaded from NCBI (accession: CP003659) and was analysed with antiSMASH 6.0[34] to predict the presence of biosynthetic gene clusters. Once the candidate radiosumin biosynthetic gene cluster was identified in the genome, the amino acid sequences of the predicted ORFs close to the NRPS were used as queries in BLASTp searches against the NCBI protein database and NCBI CD Database (CDD). Based on the functional predictions of these genes and chemical structure analysis of radiosumin, the radiosumin biosynthetic gene cluster of *A. cylindrica* PCC 7122 was manually annotated using Artemis. Artemis was used to inspect and annotate the identified biosynthetic gene clusters from other strains. The substrate specificity of the NRPS adenylating domains was predicted based on the results provided by antiSMASH 6.0.

Clinker 0.0.23 was used to compare and visualize the similarity of identified biosynthetic gene clusters.[108] BiG-SCAPE 1.1.14 was used to annotate the NRPS domains against the pfam database.[35]

Genome sequencing and assembly

D. planctonicum UHCC 0167 was grown as described above for four weeks. *Plectonema radiosum* NIES-515 was grown in Csi medium under the same condition. Genomic DNA from *D. planctonicum* UHCC 0167 was extracted by the phenol–chloroform method as previously described.[12] Genomic DNA from *P. radiosum* NIES-515 was extracted using DNeasy PowerSoil kit (QIAGEN, Netherlands) according to the manufacturer's instructions. A gap identified in the *radH* gene from *D. planctonicum* UHCC 0167 was closed by Sanger sequencing. Whole-genome sequencing of *P. radiosum* NIES-515 was performed using Illumina MiSeq. The quality of the Illumina reads was checked by running FastQC v0.10.1 (**https://www.bioinformatics.babraham.ac.uk/projects/fastqc**). Cutadapt 1.18[109] was run to remove adapter contamination. Further trimming was performed with Prinseq v0.20.4.[110] *De novo* genome assembly was performed using Spades v3.15.4.[111] with default parameters. Non-cyanobacterial contigs were removed from the assembly using Kaiju 1.8.2[112] in combination with in-house scripts, with the level set to phylum and the taxon set to Cyanobacteria. The technical quality of the resulting assemblies was evaluated with Quast v5.0.2.[113] Genome completeness and contamination were calculated using CheckM 1.1.3.[114] The coverage of the assembled genome was calculated by mapping the reads with Bowtie2 v2.4.5.[115] Sorting and indexing were performed with SAMtools 1.15.1.[116] Bedtools v2.27.1[117] was used to calculate the sequence coverage.

Phylogenomic tree

The amino acid sequences of 120 bacterial single-copy conserved marker proteins from the strains with the radiosumin biosynthetic gene cluster and other closely related cyanobacterial genomes were selected. These sequences were aligned with GTDB-Tk 0.3.2,[118] and used to infer the maximum-likelihood phylogenomic tree with RAxML 8.0.0[119] applying 1000 bootstraps using the PROTGAMMAIGTR model, and assigned as the best by ProtTest 3.4.2.[120]

Liquid chromatography–mass spectrometry

Cyanobacterial biomass was collected from 40 mL of culture by centrifugation at 7000*g* for 7 min. For *E. coli*, both supernatants and pellets were collected by centrifugation (Centrifuge 5804R, Eppendorf, Germany) at 7000*g* for 5 min. The biomass and supernatant were frozen at −80 °C for 30 min and freeze-dried overnight (Christ Beta 2-8 LSCplus, LyoCube 4-8, Germany). Approximately 200 μL of 0.5 mm glass beads (Scientific Industries Inc., USA) and 1 mL methanol (HiperSolv for high-performance liquid chromatography, BDH Laboratory supplies) with 0.1% trifluoroacetic acid (TFA) were added to the dried pellets in 2 mL screw tube. Cells were broken using a FastPrep cell disrupter (Bio 101, Thermo Electron Corporation, Qbiogene, Inc., USA) at a speed of 6.5 ms^{-1}, 30 s for twice. The dried supernatants were dissolved in 3 mL methanol with 0.1% TFA. The suspension from cell extracts or growth media extracts was centrifuged at 12 000 rpm for 1 min. Each sample consisted of 50 μL methanolic supernatant mixed with 150 μL acetonitrile and was subsequently filtered into a vial using a 0.2 μm filter (13 mm syringe filter, PTFE, VWR International, USA) and stored at 4 °C.

Sample analysis was performed with ultra-performance liquid chromatograph with quadrupole time-of-flight (UPLC-QTOF) mass spectrometry (Acquity I-Class UPLC-SynaptG2-Si, Waters Corp., Milford, MA). An ACQUITY UPLC BEH Amide column (1.7 μm, 2.1 mm × 100 mm, 130 Å, Waters Corp., Milford, MA) was used for separation. During analysis, a total of 2 μL of sample was injected for analysis. The mobile phase was 0.2% ammonium formate (A1) or 0.2% ammonium acetate (A2) and acetonitrile (B). The flow rate was 0.3 mL min^{-1}. The method was as following: (1) from 10% A1 to 40% A1 in 9 min, then maintained 40% A1 and 60% B for 1 min, and back to 10% A1 in 0.1 min. The total run time was 16 min; (2) from 15% A2 to 35% A2 in 7 min, then maintained 35% A2 and 65% B for 2 min, and back to 15% A2 in 0.1 min. The total run time was 15 min. The major instrument parameters for MS and MS/MS analysis were capillary voltage 2.5 kV, source temperature 120 °C, sampling cone 20.0 V, source offset 60.0, desolvation temperature 600 °C, desolvation gas flow of 1000 L h^{-1}, nebuliser gas flow pressure 6.0 bar, ion energy 1.0, and trap collision energy 4.0. Samples were detected at positive polarity Resolution Mode mass range m/z 50–2000. QTOF was calibrated using sodium formate and Ultramark® 1621. Leucine Enkephalin was used at 10 s intervals as a lock mass reference compound. In MS/MS mode Trap Collision Energy Ramp proceeded from 10.0 eV to 30.0 eV.

Primer design, PCR and cloning

All expression vectors in this study were generated by using the direct pathways cloning mediated (DiPaC) sequence- and ligation-independent cloning (SLIC) method as described previously.[2,46] Primers for DiPaC were designed using the NEBuilder assembly tool. Homology arms were composed of at least 25 nt with a calculated melting temperature of >50 °C. DNA assemblies were simulated using Snapgene v2.3.2. The bioinformatic tools OligoAnalyzer and OligoEvaluator were used to check primer quality,[46] avoiding severe secondary hairpin structures and primer dimer formation. The putative radiosumin biosynthetic gene cluster was checked with ARNold tool to predict Rho-independent transcription terminators within intergenic regions.[46,121] All oligonucleotide primers used in this study are listed (Table S3†) and were synthesized by Eurofins Genomics (Germany).

Linearized vectors and insert gene fragments were obtained by PCR. A standard 50 μL PCR reaction batch for amplicon cycling reactions consisted of: 1 × Q5 reaction buffer, 100–200 μM deoxynucleotide triphosphates, 500 nM of forward and reverse primer, DNA template and 0.01–0.02 U μL^{-1} Q5 High-Fidelity DNA polymerase (New England Biolabs,

United Kingdon). Cycling was performed in a Thermal Cycler (Bio-rad, USA) as follows: (1) initial denaturation, 98 °C for 1 min; (2) denaturation, 98 °C for 10 s; (3) annealing, for 30 s; (4) extension, 72 °C for tExt; step (2) to (4) were repeated in total for 30 cycles. (5) Final extension, 72 °C for 2 min (6) 4 °C on hold.

The presence of PCR products was checked with 1% agarose gel electrophoresis. Both linearized vector and gene fragments were purified by PureLink™ PCR Purification Kit (Invitrogen, USA). To allow inducible expression, the entire radiosumin gene cluster was placed behind a $ptet_O$ promoter and lac operator. The *radH* gene was inserted into pET28b-ptetO-*GFP*v2 to create pET28b-ptetO-*radH-GFP* by using SLIC method, which was conducted by 1× buffer 2.1 (NEB) and 0.5 μL of T4 polymerase (NEB) in a 10 μL total reaction volume. The mixture was incubated for 2.5 min at room temperature, followed by a 10 min incubation on ice and was transformed by heat shock into competent *E. coli* DH5α. Based on pET28b-ptetO-*radH-GFP* construct, *radABCDEFG* was inserted into the vector and generated pET28b-ptetO-*radHABCDEFG-GFP*. Then *radI* was introduced into pET28b-ptetO-*radHABCDEFG-GFP*, resulting in the pET28b-ptetO-*radHABCDEFGI-GFP* expression plasmid. Positive clones for constructs were initially screened by colony PCR followed by restriction digest analysis and Sanger sequencing.

Heterologous expression in *E. coli*

Heterologous expression assays for the strains containing pET28b-$ptet_O$-*GFP*v2 series plasmids were conducted according to a protocol described previously[2,46,62] with minor modifications. Generally, *E. coli* BAP1 was transformed with pET28b-ptetO-*GFP*v2, pET28b-ptetO-*radHABCDEFG-GFP* or pET28b-ptetO-*radHABCDEFGI-GFP* separately using the chemical method. Clones were selected from LB agar plates containing 50 μg mL^{-1} kanamycin. Three transformants and a negative control were transferred to liquid LB medium supplemented with 50 μg mL^{-1} kanamycin and incubated at 37 °C with shaking (200 rpm). Overnight precultures were inoculated 1 : 50 into 20 ml M9 mineral medium. Expression cultures were incubated at 37 °C with shaking (200 rpm) until an OD_{600} of 0.4–0.5 was reached. Cultures were then cooled on ice for at least 10 min, induced with tetracycline to a final concentration of 0.5 μg mL^{-1} and incubated in the dark for 3 days at 20 °C with shaking at 200 rpm.

Fluorescence microscopy

To preliminarily estimate the proteins expression level in *E. coli* by fluorescence microscopy, approximately 20 μL cell culture grown for three days were pipette on glass slides (76 × 26 mm, MENZEL-GLASER, VWR, USA) with microscope cover glasses (22 × 22 mm, VWR, USA). Images were taken on a ZEISS Axioskop 2 Routine microscope that was equipped with an AxioCam 305 color camera and an EC Plan-NEOFLUA 40× Ph2 objective. Digital images were acquired and analysed with the ZEN 3.1blue edition software, the fluorescence imaging of GFP upon excitation at 475–490 nm and emission detection between 505 and 535 nm.

Purification of radiosumin

Aphanizomenon sp. UHCC 0183 strain was grown in 30 L Z8 culture medium without nitrogen divided into 6 batches, each batch was grown for 4 weeks. All cell biomasses were collected by centrifuging at 7000*g* for 7 min and frozen at −80 °C for 1 h then put into the freeze drier for 2 days, yielding 9.5 g of dried biomass. The purification process was conducted as follows: 60 mL methanol was used to extract 1 g of dry cells with a Heidolph Silentcrusher M at 20 000 rpm for 30 s. The supernatant was collected after centrifugation at 7000*g* for 5 min. The extraction process was repeated twice. A total of 180 mL of the crude extract was dried with a rotary evaporator (Büchi Rotavapor R-200, USA) at 30 °C by adding 5 g chomatorex (Fuji-Davison Chemical Ltd, Aichi, Japan) chromatography silica ODS powder. The silica gel bound sample was fractionated in two batches using a solid-phase Strata SI-1 silica cartridge 5 g per 20 mL (Phenomenex, USA). The sample was loaded on a preconditioned cartridge and sequentially eluted with *n*-heptane, acetonitrile, acetone, ethyl acetate, and methanol. Fractions were dried with a nitrogen gas flow and re-dissolved in 1 mL of methanol and analysed by LC-MS as described previously. The fractions that contained target radiosumin were combined in the same tube and dried under nitrogen gas stream and then re-dissolved in 1 mL of methanol. The target radiosumin was further purified with an XBridge BEH Amide OBD Prep Column (130 Å, 5 μm, 19 mm × 150 mm) and a Waters Auto Purification System. The positive-ion mode was used. The injection volume was 100 μL with a flow rate of 20 mL min^{-1}. The mobile phase was consisted of 0.2% formic acid (solvent A) and acetonitrile (solvent B). A gradient from 15% to 35% A over 6 minutes was used. The combined organic solvents fractions were dried under nitrogen gas stream again. The solid material was dissolved in methanol to be analysed with the UPLC-QTOF system.

NMR spectroscopy

NMR spectra for radiosumin D were collected using the Bruker Advance III HD 800 MHz NMR spectrometer, equipped with a helium-cooled TCI ^{1}H, ^{13}C, ^{15}N triple resonance cryoprobe and a *z*-gradient coil. Data were collected at 25 °C using MeOD as a solvent. In addition to ^{1}H experiment with weak pre-saturation field applied to the residual water signal frequency during the T_1 recovery delay, a panoply of two-dimensional (2D) homo- and heteronuclear NMR experiments were collected to obtain a full assignment of radiosumin D ^{1}H, ^{13}C and ^{15}N resonances (Table S6†). 2D total correlation spectroscopy (TOCSY) experiment with mixing time (τ_m) of 90 ms together with a double-quantum filtered correlation spectroscopy (DQF-COSY) experiment were collected to identify spin-systems. Through space correlations were measured using 2D ROESY experiment with τ_m 200 ms. Accidental overlap of ^{1}H resonances was lifted using 2D ^{13}C HSQC (heteronuclear single quantum coherence) and 2D ^{13}C HSQC-TOCSY with τ_m = 90 ms. Long-range ^{1}H–^{13}C

and ^{1}H–^{15}N connectivities were established using 2D ^{13}C HMBC (heteronuclear multiple bond correlation) and 2D ^{15}N HMBC experiments, respectively. The transfer delay for long-range correlations based on $^{n}J_{C,H}$ and $^{n}J_{N,H}$ couplings was set to 62.5 ms in both ^{13}C HMBC and ^{15}N HMBC experiments. All experiments had a T_1 recovery time of 1 second.

Amino acid analysis

The absolute configuration of radiosumin D (**4**) was assigned by comparing the retentions of Marfey's derivatives of oxidized and hydrolysed radiosumin D (**4**) to commercial amino acids standards. Oxidation of isolated radiosumin D (**4**) was performed as described previously.[25] Hydrolysis and derivatization were performed with the Marfey method as described previously.[122] Reference amino acids L-Ser and D-Ser (Sigma, Switzerland) and radiosumin D (**4**) hydrolysate were derivatized with 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (FDAA).

Samples were analysed with UPLC-QTOF equipped with a Kinetex C8 column (2.1 × 100 mm, 1.7 μm, 100 Å, Phenomenex, Torrance, CA). Two μL samples were injected and then eluted at 40 °C with 0.1% formic acid in water (solvent A) and acetonitrile/isopropanol (1 : 1, +0.1% formic acid, solvent B) at a flow rate of 0.3 mL min^{-1}. A gradient from 5% B went to 25% B in 10 min, then to 100% B in 0.1 min and maintained 3.9 min, then back to 5% of B in 0.5 min and after 5.5 min ready for new injection. QTOF was operated in positive electrospray ionization resolution mode.

Biological activity testing

The inhibition of human trypsins was conducted as described previously for suomilide and varlaxins.[9,10] Recombinant trypsinogens-1, -2 and -3 were produced in *E. coli* and activated to trypsin-1, -2, and -3 as described previously.[101] Enzyme inhibitory activities of radiosumin C (**3**) and D (**4**) were determined in 96-well plates by adding 10 μL of radiosumin C (**3**) or D (**4**) diluted in ultra-purified H_2O, 15 μL of 0.1% bovine serum albumin in 50 mM Tris-buffered saline (BSA/TBS), pH 7.6, and 25 μL of each trypsin (final concentration 1.8 nM). Ultra-purified H_2O (10 μL) was used as a control. As radiosumins were dissolved in methanol, the reactions contained up to 2% methanol, which did not have any effect on trypsin activity (data not shown). The radiosumins were pre-incubated with trypsins at room temperature for 15 min before the addition of 50 μL of 0.2 mM chromogenic substrate S-2222 (ChromogeniX) in ultra-purified H_2O. After the addition of substrates, the tested radiosumin C (**3**) and D (**4**) concentrations, ranged from 0.11 to 7.23 μM and 0.95 to 60 μM, respectively. The change of absorbance at 405 nm was followed for 15 min (VICTOR X4 Multiple plate reader, PerkinElmer, Finland). The experiment was performed three times, each with two replicates (n = 3). The absorbance variation was measured during the substrate reaction phase, where the control reaction showed a linear increase in absorbance over time. The absolute IC_{50} values were estimated using Origin 2021b or Quest Graph™ (**https://www.aatbio.com/tools/ic50-calculator**).

Molecular modelling

All modelling was conducted with Maestro Small-Molecule Drug Discovery Suite 2021-2 (Schrödinger Release 2021-2: Maestro, Schrödinger, LLC, 2021) with an OPLS4 force field. For the docking and MD simulations, we used the PDB structures 2RA3 (trypsin-1) and 1H4 W (trypsin-3), which were prepared using Protein Preparation Wizard. The amino acid numbering used here was based on the numbering in these PDB structures. The crystal structure mutations were manually changed to match the Uniprot sequence (trypsin1: His99 → Arg99 and Ala177 → Ser177, trypsin3: Val148 → Ala128 and Ala188 → Thr188). Trypsin-2 was constructed as a homology model with 1H4W as a template using the Prime homology modelling tool. The sequence identity was 89% and the bound Ca^{2+} ion was included in the homology model. To keep the structures comparable, this bound Ca^{2+} ion was also inserted into the trypsin-1 model. The initial coordinates for the ligand–protein complexes were obtained by Glide XP docking using default settings. The ligands were first preprocessed using LigPrep with default settings, retaining the stereochemistry and producing different protonation states at pH 7 using Epik. The ligands were docked into the protein models using the Glide eXtra precision (XP) mode with default settings. The grid box centre was defined by residues corresponding to His224 and Gln192 (trypsin-1 numbering). The most visually plausible and high-scoring docking poses were selected as the starting structures for the simulations. MD simulations were conducted with Desmond using default parameters Maestro Small-Molecule Drug Discovery Suite 2021-3 on the CSC (IT Center for Science, Finland) supercomputer Puhti. The systems were solvated in an orthorhombic box (edges 10 Å from the protein), neutralized with counter ions, and 0.15 M NaCl salt was added. The water was described with the SPC water model. The default relaxation protocol of Desmond was used before the 1000 ns production simulations, which were conducted in NPT ensemble (300 K, thermostat: Nosé–Hoover chain; 1.01325 bar, barostat: Martyna–Tobias–Klein). The default time step of 2 fs and cutoff radius of 9.0 Å for coulombic interactions were used. The total simulation time was 5 μs (5 × 1000 ns) for all systems. For analysis, the simulations were concatenated using trj_merge.py script.

Conclusions

We identified the biosynthetic gene cluster encoding radiosumins from 13 cyanobacteria based on bioinformatic analysis. The function of the entire radiosumin biosynthetic gene cluster was confirmed by direct pathway cloning and heterologous expression. Characterization of the radiosumin biosynthesis pathway based on bioinformatic analysis revealed a new non-aromatizing route from chorismic acid (**5**) to nonproteinogenic amino acids. High-resolution LC-MS and NMR data revealed that radiosumin D (**4**) is a new *N*-methyl dipeptide, composed of an *N*-Me derivative of Aayp, two acetate groups and an Aayp with *R* configuration that differs from the *S*

stereochemistry of Aayp substructure in radiosumins A–C (**1**–**3**). Radiosumin C (**3**) inhibits all three human trypsin-isoenzymes but exhibits somewhat different relative potency in inhibition of different trypsin-isoenzymes, as compared to other cyanobacterial trypsin-inhibitors analysed thus far. The pioneering studies of the structure and biosynthesis of radiosumins here not only shed light on how cyanobacteria produce these diverse dipeptides in nature, but will also promote development of these compounds as pharmaceutical drugs targeting trypsin-isoenzymes. Our results now provide a biosynthetic logic to explore the genetic and chemical diversity of the radiosumin family.

Author contributions

XO, DF and PMD designed the study. XO, MW, PP, and PB performed the experiments. GG collected and provided the bloom samples. XO, DF, ED, JJ, PP, HK and PB analysed and interpreted the data. XO, DF, PMD and HK wrote the manuscript, which was revised and approved by all authors.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

We thank Lyudmila Saari and Annikki Löfhjelm for their valuable help in handling the cyanobacterial cultures and for excellent technical assistance in activity assays. We thank Leena A. Räsänen and Vy A. Huynh (University of Helsinki) for critically reading the manuscript and providing comments to improve the manuscript. We thank the IT Centre for Science (CSC) of Finland for offering a platform for modelling work. We would also like to thank the China Scholarship Council for providing a grant (Grant 201906150148) and the University of Helsinki's Doctoral Programme in Microbiology and Biotechnology for providing travel funding to XO. This work was also supported by a funding from the NordForsk NCoE program NordAqua (project no. 82845) (DPF), the Nessling Foundation funding (Grant no. 202200182) to DPF, a grant from the Sigrid Jusélius Foundation (HK), a funding from Magnus Ehrnrooth Foundation (HK), the Competitive Funding to Strengthen University Research Profiles, 5th call, funding to University of Eastern Finland, funded by the Academy of Finland (grant number 325022) (PB), a PRINT Scholarship from the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (CAPES) (88887.572010/2020-00) (ED), a funding from the Technical University of Dresden Research Pool and the Hans Fischer Society (PMD), and a funding from Jane and Aatos Erkko foundation and the Academy of Finland (grant 323435) (PP).

References

1 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2016, **79**, 629–661.
2 C. Greunke, E. R. Duell, P. M. D'Agostino, A. Glockle, K. Lamm and T. A. M. Gulder, *Metab. Eng.*, 2018, **47**, 334–345.
3 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2020, **83**, 770–803.
4 J. A. van Santen, E. F. Poynton, D. Iskakova, E. McMann, T. A. Alsup, T. N. Clark, C. H. Fergusson, D. P. Fewer, A. H. Hughes, C. A. McCadden, J. Parra, S. Soldatou, J. D. Rudolf, E. M. Janssen, K. R. Duncan and R. G. Linington, *Nucleic Acids Res.*, 2022, **50**, D1317–D1323.
5 F. H. Al-Awadhi and H. Luesch, *Nat. Prod. Rep.*, 2020, **37**, 827–860.
6 C. J. l. Porte, *Expert Opin. Drug Metab. Toxicol.*, 2009, **5**, 1313–1322.
7 S. Ito, *Cancers*, 2020, **12**, 1–19.
8 J. Demay, C. Bernard, A. Reinhardt and B. Marie, *Mar. Drugs*, 2019, **17**, 1–49.
9 M. N. Ahmed, M. Wahlsten, J. Jokela, M. Nees, U.-H. Stenman, D. O. Alvarenga, T. Strandin, K. Sivonen, A. Poso, P. Permi, M. Metsä-Ketelä, H. Koistinen and D. P. Fewer, *ACS Chem. Biol.*, 2021, **16**, 2537–2546.
10 L. M. P. Heinila, J. Jokela, M. N. Ahmed, M. Wahlsten, S. Kumar, P. Hrouzek, P. Permi, H. Koistinen, D. P. Fewer and K. Sivonen, *Org. Biomol. Chem.*, 2022, **20**, 2681–2692.
11 D. A. Gallegos, J. Sauri, R. D. Cohen, X. Wan, P. Videau, A. O. Vallota-Eastman, L. A. Shaala, D. T. A. Youssef, R. T. Williamson, G. E. Martin, B. Philmus, A. E. Sikora, J. E. Ishmael and K. L. McPhail, *J. Nat. Prod.*, 2018, **81**, 1417–1425.
12 J. Jokela, L. M. P. Heinila, T. K. Shishido, M. Wahlsten, D. P. Fewer, M. F. Fiore, H. Wang, E. Haapaniemi, P. Permi and K. Sivonen, *Front. Microbiol.*, 2017, **8**, 1963.
13 K. Ersmark, J. R. Del Valle and S. Hanessian, *Angew. Chem., Int. Ed.*, 2008, **47**, 1202–1223.
14 P. R. Monteiro, S. C. do Amaral, A. S. Siqueira, L. P. Xavier and A. V. Santos, *Toxins*, 2021, **13**, 1–37.
15 H. Mazur-Marzec, A. Fidor, M. Ceglowska, E. Wieczerzak, M. Kropidlowska, M. Goua, J. Macaskill and C. Edwards, *Mar. Drugs*, 2018, **16**, 1–19.
16 G. E. Chlipala, S. Mo and J. Orjala, *Curr. Drug Targets*, 2011, **12**, 1654–1673.
17 S. Agrawal, D. Acharya, A. Adholeya, C. J. Barrow and S. K. Deshmukh, *Front. Pharmacol.*, 2017, **8**, 828.
18 R. D. Sussmuth and A. Mainz, *Angew. Chem., Int. Ed. Engl.*, 2017, **56**, 3770–3821.
19 J. B. Hedges and K. S. Ryan, *Chem. Rev.*, 2020, **120**, 3161–3209.
20 S. Cadel-Six, C. Dauga, A. M. Castets, R. Rippka, C. Bouchier, N. Tandeau de Marsac and M. Welker, *Mol. Biol. Evol.*, 2008, **25**, 2031–2041.
21 C. J. B. Micha Fridman, T. Lupoli, D. Kahne, C. T. Walsh and S. Garneau-Tsodikova, *Biochemistry*, 2007, **46**, 8462–8471.

22 K. Ishida, G. Christiansen, W. Y. Yoshida, R. Kurmayer, M. Welker, N. Valls, J. Bonjoch, C. Hertweck, T. Borner, T. Hemscheidt and E. Dittmann, *Chem. Biol.*, 2007, **14**, 565–576.
23 C. Clemente, N. Johnson, X. Ouyang, R. V. Popin, S. Dall'Angelo, M. Wahlsten, J. Jokela, A. Colombano, B. Nardone, D. P. Fewer and W. E. Houssen, *Chem. Commun.*, 2022, **58**, 12054–12057.
24 D. P. Fewer, J. Jokela, E. Paukku, J. Osterholm, M. Wahlsten, P. Permi, O. Aitio, L. Rouhiainen, G. V. Gomez-Saez and K. Sivonen, *PLoS One*, 2013, **8**, e73618.
25 H. Matsuda, T. Okino, M. Murakami and K. Yamaguchi, *J. Org. Chem.*, 1996, **61**, 8648–8650.
26 H. Noguchi, T. Aoyama and T. Shioiri, *Tetrahedron Lett.*, 1997, **38**, 2883–2886.
27 S. Kodani, K. Ishida and M. Murakami, *J. Nat. Prod.*, 1998, **61**, 854–856.
28 J. E. Coleman and J. L. C. Wright, *J. Nat. Prod.*, 2001, **64**, 668–670.
29 M. Welker and H. von Dohren, *FEMS Microbiol. Rev.*, 2006, **30**, 530–563.
30 A. Gkazi, *Technology Networks*, 2021.
31 S. Goodwin, J. D. McPherson and W. R. McCombie, *Nat. Rev. Genet.*, 2016, **17**, 333–351.
32 P. M. Shih, D. Wu, A. Latifi, S. D. Axen, D. P. Fewer, E. Talla, A. Calteau, F. Cai, N. Tandeau de Marsac, R. Rippka, M. Herdman, K. Sivonen, T. Coursin, T. Laurent, L. Goodwin, M. Nolan, K. W. Davenport, C. S. Han, E. M. Rubin, J. A. Eisen, T. Woyke, M. Gugger and C. A. Kerfeld, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 1053–1058.
33 A. Calteau, D. P. Fewer, A. Latifi, T. Coursin, T. Laurent, J. Jokela, C. A. Kerfeld, K. Sivonen, J. Piel and M. Gugger, *BMC Genomics*, 2014, **15**, 1471–2164.
34 K. Blin, S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. van Wezel, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2021, **49**, W29–W35.
35 J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. Mullowney, S. A. Kautsar, J. H. Tryon, E. I. Parkinson, E. L. C. De Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker, A. Roeters, W. Lokhorst, A. Fernandez-Guerra, L. T. D. Cappelini, A. W. Goering, R. J. Thomson, W. W. Metcalf, N. L. Kelleher, F. Barona-Gomez and M. H. Medema, *Nat. Chem. Biol.*, 2020, **16**, 60–68.
36 C. L. M. Gilchrist, T. J. Booth, B. van Wersch, L. van Grieken, M. H. Medema and Y.-H. Chooi, *Bioinform. Adv.*, 2021, **1**, 1–10.
37 M. Li, L. Chen, Z. Deng and C. Zhao, *FEBS Open Bio.*, 2016, **6**, 603–609.
38 S. A. Mahlstedt and C. T. Walsh, *Biochemistry*, 2010, **49**, 912–923.
39 V. Blanc, P. Gil, N. BamasJacques, S. Lorenzon, M. Zagorec, J. Schleuniger, D. Bisch, F. Blanche, L. Debussche, J. Crouzet and D. Thibaut, *Mol. Microbiol.*, 1997, **23**, 191–202.
40 A. Mattila, R. M. Andsten, M. Jumppanen, M. Assante, J. Jokela, M. Wahlsten, K. M. Mikula, C. Sigindere, D. H. Kwak, M. Gugger, H. Koskela, K. Sivonen, X. Liu, J. Yli-Kauhaluoma, H. Iwai and D. P. Fewer, *ACS Chem. Biol.*, 2019, **14**, 2683–2690.
41 P. Lindblad, D. Fuente, F. Borbe, B. Cicchi, J. A. Conejero, N. Couto, H. Čelešnik, M. M. Diano, M. Dolinar, S. Esposito, C. Evans, E. A. Ferreira, J. Keller, N. Khanna, G. Kind, A. Landels, L. Lemus, J. Noirel, S. Ocklenburg, P. Oliveira, C. C. Pacheco, J. L. Parker, J. Pereira, T. K. Pham, F. Pinto, S. Rexroth, M. Rögner, H.-J. Schmitz, A. M. S. Benavides, M. Siurana, P. Tamagnini, E. Touloupakis, G. Torzillo, J. F. Urchueguía, A. Wegelius, K. Wiegand, P. C. Wright, M. Wutschel and R. Wünschiers, *Algal Res.*, 2019, **41**, 1–15.
42 R. Laurenceau, C. Bliem, M. S. Osburne, J. W. Becker, S. J. Biller, A. Cubillos-Ruiz and S. W. Chisholm, *Access Microbiol.*, 2020, **2**, 1–13.
43 D. Liu and H. B. Pakrasi, *Microb. Cell Fact.*, 2018, **17**, 48.
44 N. Eusébio, R. Castelo-Branco, D. Sousa, M. Preto, P. D'Agostino, T. A. M. Gulder and P. N. Leão, *ACS Synth. Biol.*, 2022, **11**, 3493–3503.
45 P. M. D'Agostino, C. J. Seel, X. Ji, T. Gulder and T. A. M. Gulder, *Nat. Chem. Biol.*, 2022, **18**, 652–658.
46 P. M. D'Agostino and T. A. M. Gulder, *ACS Synth. Biol.*, 2018, **7**, 1702–1708.
47 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 1999, **6**, 493–505.
48 D. H. Parks, M. Chuvochina, D. W. Waite, C. Rinke, A. Skarshewski, P. A. Chaumeil and P. Hugenholtz, *Nat. Biotechnol.*, 2018, **36**, 996–1004.
49 L. M. P. Heinilä, D. P. Fewer, J. K. Jokela, M. Wahlsten, X. Ouyang, P. Permi, A. Jortikka and K. Sivonen, *Org. Biomol. Chem.*, 2021, **19**, 5577–5588.
50 L. Dalponte, A. Parajuli, E. Younger, A. Mattila, J. Jokela, M. Wahlsten, N. Leikoski, K. Sivonen, S. A. Jarmusch, W. E. Houssen and D. P. Fewer, *Biochemistry*, 2018, **57**, 6860–6867.
51 J. Lehtimäki, C. Lyra, S. Suomalainen, M. Salkinoja-Salonen, P. Sundman, L. Rouhiainen and A. K. S. Lars Paulin, *Int. J. Syst. Evol. Microbiol.*, 2000, **50**, 1043–1053.
52 E. Dittmann, D. P. Fewer and B. A. Neilan, *FEMS Microbiol. Rev.*, 2013, **37**, 23–43.
53 C. S. Sheik, K. E. Natwora, E. E. Alexson, J. D. Callaghan, A. Sailer, K. M. Schreiner, B. A. Steinman, M. S. Finkenbinder, C. T. Filstrup and A. J. Bramburger, *J. Great Lakes Res.*, 2022, **48**, 1191–1205.
54 C. B. Driscoll, K. A. Meyer, S. Sulcius, N. M. Brown, G. J. Dick, H. Cao, G. Gasiunas, A. Timinskas, Y. Yin, Z. C. Landry, T. G. Otten, T. W. Davis, S. B. Watson and T. W. Dreher, *Harmful Algae*, 2018, **77**, 93–107.
55 J. Österholm, R. V. Popin, D. P. Fewer and K. Sivonen, *Toxins.*, 2020, **12**, 1–16.
56 K. Sivonen, S. I. Niemelä, R. M. Niemi, L. Lepistö, T. H. Luoma and L. A. Räsänen, *Hydrobiologia*, 1990, **190**, 267–275.
57 J. E. Teikari, R. V. Popin, S. Hou, M. Wahlsten, W. R. Hess and K. Sivonen, *Sci. Rep.*, 2019, **9**, 4888.

58 A. Kust, K. Rehakova, J. Vrba, V. Maicher, J. Mares, P. Hrouzek, M. C. Chiriac, Z. Benedova, B. Tesarova and K. Saurav, *Toxins*, 2020, **12**, 1–21.
59 A. Schwarzenberger, M. Ilic and E. Von Elert, *Harmful Algae*, 2021, **106**, 102062.
60 A. Schwarzenberger, C. J. Kuster and E. Von Elert, *Mol. Ecol.*, 2012, **21**, 4898–4911.
61 E. Dittmann, M. Gugger, K. Sivonen and D. P. Fewer, *Trends Microbiol.*, 2015, **23**, 642–652.
62 T. Liu, R. Mazmouz, S. E. Ongley, R. Chau, R. Pickford, J. N. Woodhouse and B. A. Neilan, *ACS Chem. Biol.*, 2017, **12**, 2021–2029.
63 B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science*, 2001, **291**, 1790–1792.
64 E. R. Duell, P. M. D'Agostino, N. Shapiro, T. Woyke, T. M. Fuchs and T. A. M. Gulder, *Microb. Cell Fact.*, 2019, **18**, 1–11.
65 T. Kusakizako, H. Miyauchi, R. Ishitani and O. Nureki, *Biochim. Biophys. Acta, Biomembr.*, 2020, **1862**, 1–14.
66 O. Kerbarh, E. M. M. Bulloch, R. J. Payne, T. Sahr, F. Rébeillé and C. Abell, *Biochem. Soc. Trans.*, 2005, **33**, 763–766.
67 S. Noda, T. Shirai, S. Oyama and A. Kondo, *Metab. Eng.*, 2016, **33**, 119–129.
68 X. Qiu, W. Zhu, W. Wang, H. Jin, P. Zhu, R. Zhuang and X. Yan, *J. Struct. Biol.*, 2019, **205**, 44–52.
69 L. T. Fernandez-Martinez, C. Borsetto, J. P. Gomez-Escribano, M. J. Bibb, M. M. Al-Bassam, G. Chandra and M. J. Bibb, *Antimicrob. Agents Chemother.*, 2014, **58**, 7441–7450.
70 B. G. Chia-Yu, P. Teng, S. Z. Doktor, B. P. Nichols, R. K. Bhatnagar and L. C. Vining, *J. Am. Chem. Soc.*, 1985, **107**, 5008–5009.
71 B. P. Nichols, A. M. Seibold and S. Z. Doktor, *J. Biol. Chem.*, 1989, **264**, 8597–8601.
72 J. F. Parsons, P. Y. Jensen, A. S. Pachikara, A. J. Howard, E. Eisenstein and J. E. Ladner, *Biochemisrty*, 2002, **41**, 2198–2208.
73 Z. H. a. M. D. Toney, *Biochemistry*, 2006, **45**, 5019–5028.
74 C. T. Walsh, R. V. O'Brien and C. Khosla, *Angew. Chem., Int. Ed.*, 2013, **52**, 7098–7124.
75 A. R. Knaggs, *Nat. Prod. Rep.*, 1999, **16**, 525–560.
76 S. Mahlstedt, E. N. Fielding, B. S. Moore and C. T. Walsh, *Biochemistry*, 2010, **49**, 9021–9023.
77 J. M. Crawford, S. A. Mahlstedt, S. J. Malcolmson, J. Clardy and C. T. Walsh, *Chem. Biol.*, 2011, **18**, 1102–1112.
78 C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, *BMC Evol. Biol.*, 2007, **7**, 78.
79 K. Arima, A. Kakinuma and G. Tamura, *Biochem. Biophys. Res. Commun.*, 1968, **31**, 488–494.
80 S. Horowitz and W. M. Griffin, *J. Ind. Microbiol. Biotechnol.*, 1991, **7**, 45–52.
81 V. Tosato, A. M. Albertini, M. Zotti, S. Sonda and C. V. Bruschi, *Microbiology*, 1997, **143**, 3443–3450.
82 M. Morikawa, H. Daido, T. Takao, S. Murata, Y. Shimonishi and T. Imanaka, *J. Bacteriol.*, 1993, **175**, 6459–6466.
83 R. Iacovelli, R. A. L. Bovenberg and A. J. M. Driessen, *J. Ind. Microbiol. Biotechnol.*, 2021, **48**, 1–30.
84 C. Hacker, X. Cai, C. Kegler, L. Zhao, A. K. Weickhmann, J. P. Wurm, H. B. Bode and J. Wohnert, *Nat. Commun.*, 2018, **9**, 4366.
85 C. Lopez-Otin and J. S. Bond, *J. Biol. Chem.*, 2008, **283**, 30433–30437.
86 J. S. Bond, *J. Biol. Chem.*, 2019, **294**, 1643–1651.
87 H. Koistinen, R. M. Kovanen, M. D. Hollenberg, A. Dufour, E. S. Radisky, U. H. Stenman, J. Batra, J. Clements, J. D. Hooper, E. Diamandis, O. Schilling, A. Rannikko and T. Mirtti, *IUBMB Life*, 2023, 1–21.
88 S. Rakashanda, F. Rana, S. Rafiq, A. Masood and S. Amin, *Biotechnol. Mol. Biol. Rev.*, 2012, **7**, 90–101.
89 S. P. Gupta and S. D. Gupta, *Cancer-Leading Proteases*, 2020, ch. 1, pp. 1–11.
90 A. Paju and U. H. Stenman, *Crit. Rev. Clin. Lab. Sci.*, 2006, **43**, 103–142.
91 P. Nyberg, M. Ylipalosaari, T. Sorsa and T. Salo, *Exp. Cell Res.*, 2006, **312**, 1219–1228.
92 E. S. Radisky, *Asian J. Androl.*, 2013, **15**, 439–440.
93 K. Soreide, E. A. Janssen, H. Korner and J. P. Baak, *J. Pathol.*, 2006, **209**, 147–156.
94 U. Wiegand, S. Corbach, A. Minn, J. Kang and B. Miiller-Hill, *Gene*, 1993, **136**, 167–175.
95 A. Hockla, E. Miller, M. A. Salameh, J. A. Copland, D. C. Radisky and E. S. Radisky, *Mol. Cancer Res.*, 2012, **10**, 1555–1566.
96 A. Hockla, D. C. Radisky and E. S. Radisky, *Breast Cancer Res. Treat.*, 2010, **124**, 27–38.
97 G. Jiang, F. Cao, G. Ren, D. Gao, V. Bhakta, Y. Zhang, H. Cao, Z. Dong, W. Zang, S. Zhang, H. H. Wong, C. Hiley, T. Crnogorac-Jurcevic, N. R. Lemoine and Y. Wang, *Gut*, 2010, **59**, 1535–1544.
98 O. Schilling, M. L. Biniossek, B. Mayer, B. Elsasser, H. Brandstetter, P. Goettig, U. H. Stenman and H. Koistinen, *Biol. Chem.*, 2018, **399**, 997–1007.
99 K. W. Jared, C. Roach, Lu Gan and L. Hood, *J. Mol. Evol.*, 1997, **45**, 640–652.
100 S. J. K. Richard, J. P. Cannellt, A. M. Owsianka and J. M. Walker, *Planta Med.*, 1988, **54**, 10–14.
101 P. Wu, J. Weisell, M. Pakkala, M. Perakyla, L. Zhu, R. Koistinen, E. Koivunen, U. H. Stenman, A. Narvanen and H. Koistinen, *Biol. Chem.*, 2010, **391**, 283–293.
102 U. Linne, S. Doekel and M. A. Marahiel, *Biochemistry*, 2001, **40**, 15824–15834.
103 O. Kayode, Z. Huang, A. S. Soares, T. R. Caulfield, Z. Dong, A. M. Bode and E. S. Radisky, *PLoS One*, 2017, **12**, 1–15.
104 I. Cohen, O. Kayode, A. Hockla, B. Sankaran, D. C. Radisky, E. S. Radisky and N. Papo, *Biochem. J.*, 2016, **473**, 1329–1341.
105 M. A. Salameh, A. S. Soares, A. Hockla, D. C. Radisky and E. S. Radisky, *Biochem. J.*, 2011, **440**, 95–105.

106 M. A. Salameh, A. S. Soares, D. Navaneetham, D. Sinha, P. N. Walsh and E. S. Radisky, *J. Biol. Chem.*, 2010, **285**, 36884–36896.
107 S. Ishizawa, Y. Yamamoto, T. Denboh and H. Ueda, *Fish Sci.*, 2010, **76**, 669–676.
108 C. L. M. Gilchrist and Y. H. Chooi, *Bioinformatics*, 2021, **37**, 2473–2475.
109 M. Martin, *EMBnet*, 2011, **17**, 10–12.
110 R. Schmieder and R. Edwards, *Bioinformatics*, 2011, **27**, 863–864.
111 A. Prjibelski, D. Antipov, D. Meleshko, A. Lapidus and A. Korobeynikov, *Curr. Protoc. Bioinf.*, 2020, **70**, 1–29.
112 P. Menzel, K. L. Ng and A. Krogh, *Nat. Commun.*, 2016, **7**, 1–9.
113 A. Gurevich, V. Saveliev, N. Vyahhi and G. Tesler, *Bioinformatics*, 2013, **29**, 1072–1075.
114 D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz and G. W. Tyson, *Genome Res.*, 2015, **25**, 1043–1055.
115 B. Langmead and S. L. Salzberg, *Nat. Methods*, 2012, **9**, 357–359.
116 P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies and H. Li, *Gigascience*, 2021, **10**, 1–4.
117 A. R. Quinlan and I. M. Hall, *Bioinformatics*, 2010, **26**, 841–842.
118 P. A. Chaumeil, A. J. Mussig, P. Hugenholtz and D. H. Parks, *Bioinformatics*, 2019, **36**, 1925–1927.
119 A. Stamatakis, *Bioinformatics*, 2014, **30**, 1312–1313.
120 D. Darriba, G. L. Taboada, R. Doallo and D. Posada, *Bioinformatics*, 2011, **27**, 1164–1165.
121 M. Naville, A. Ghuillot-Gaudeffroy, A. Marchais and D. Gautheret, *RNA Biol.*, 2011, **8**, 11–13.
122 J. Jokela, L. M. P. Heinilä, T. K. Shishido, M. Wahlsten, D. P. Fewer, M. F. Fiore, H. Wang, E. Haapaniemi, P. Permi and K. Sivonen, *Front. Microbiol.*, 2017, **8**, 1–14.