

Greg Mann,<sup>a</sup> Jesko Koehnke,<sup>a</sup>  
Andrew F. Bent,<sup>a</sup> Rachael  
Graham,<sup>a</sup> Wael Housen,<sup>b,c,d</sup>  
Marcel Jaspars,<sup>b</sup> Uli Schwarz-  
Linek<sup>a</sup> and James H. Naismith<sup>a\*</sup>

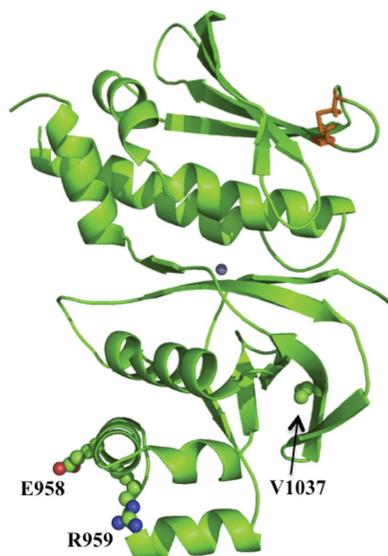
<sup>a</sup>Biomedical Sciences Research Complex,  
University of St Andrews, North Haugh,  
St Andrews, Fife KY16 8ST, Scotland, <sup>b</sup>Marine  
Biodiscovery Centre, Department of Chemistry,  
University of Aberdeen, Meston Walk,  
Aberdeen AB24 3UE, Scotland, <sup>c</sup>Institute of  
Medical Sciences, University of Aberdeen,  
Aberdeen AB25 2ZD, Scotland, and  
<sup>d</sup>Pharmacognosy Department, Faculty of  
Pharmacy, Mansoura University,  
Mansoura 35116, Egypt

Correspondence e-mail: jhn@st-andrews.ac.uk

Received 5 September 2014

Accepted 3 November 2014

**PDB reference:** PatG-DUF, 4uvq



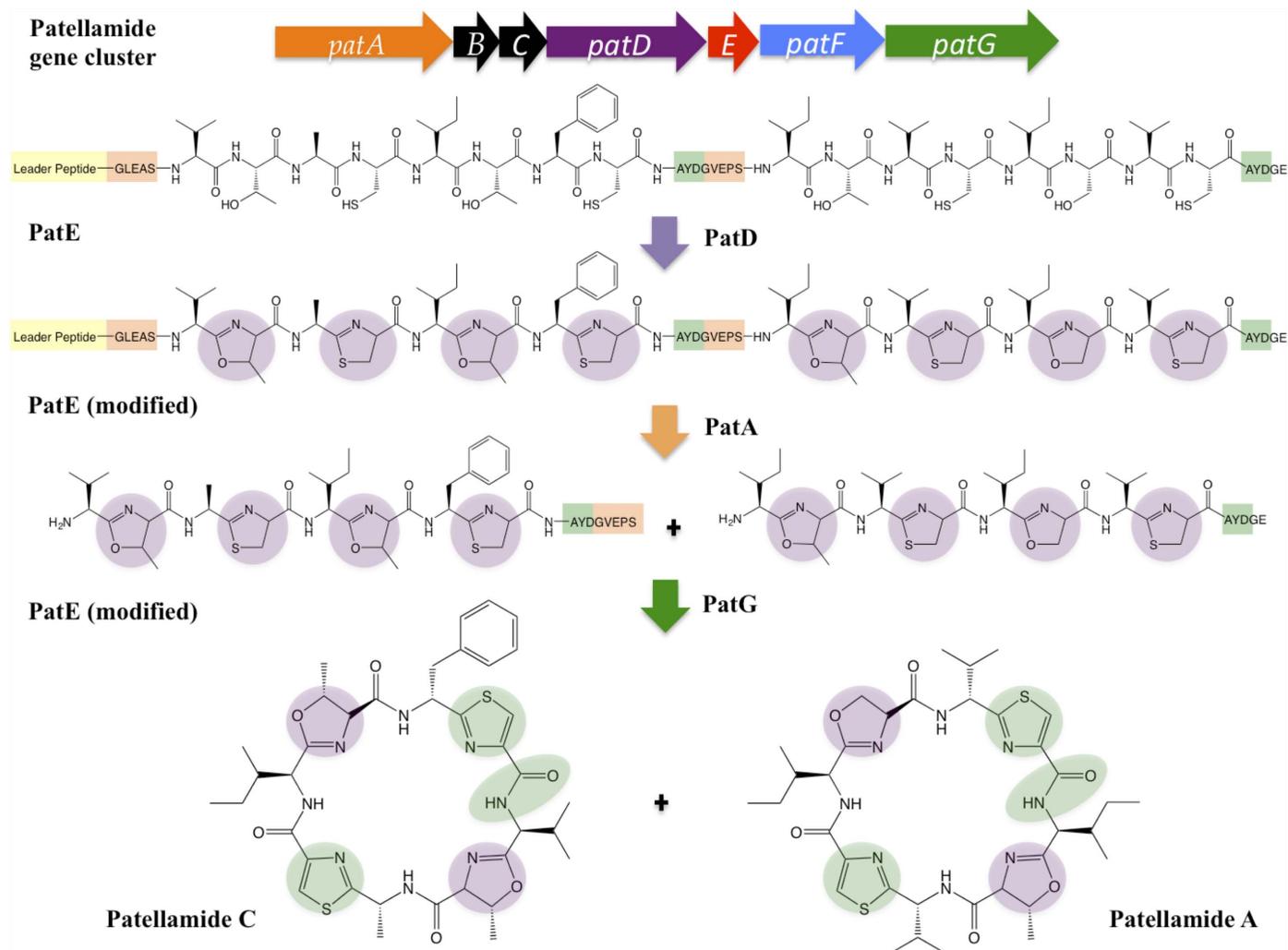
## The structure of the cyanobactin domain of unknown function from PatG in the patellamide gene cluster

Patellamides are members of the cyanobactin family of ribosomally synthesized and post-translationally modified cyclic peptide natural products, many of which, including some patellamides, are biologically active. A detailed mechanistic understanding of the biosynthetic pathway would enable the construction of a biotechnological ‘toolkit’ to make novel analogues of patellamides that are not found in nature. All but two of the protein domains involved in patellamide biosynthesis have been characterized. The two domains of unknown function (DUFs) are homologous to each other and are found at the C-termini of the multi-domain proteins PatA and PatG. The domain sequence is found in all cyanobactin-biosynthetic pathways characterized to date, implying a functional role in cyanobactin biosynthesis. Here, the crystal structure of the PatG DUF domain is reported and its binding interactions with plausible substrates are investigated.

### 1. Introduction

Prokaryotic secondary metabolism creates an array of bioactive molecules, some of which can be exploited for pharmaceutical development. Cyanobactins, secondary metabolites of cyanobacteria, are a superfamily of ribosomally synthesized and post-translationally modified peptides (RiPPs) that are highly variable with a wide range of bioactivities (Sivonen *et al.*, 2010). Patellamides, eight-residue cyclic peptides that contain oxazoline, thiazoline and thiazole heterocycles and D-amino acids (Schmidt *et al.*, 2005), belong to this superfamily and have attracted considerable interest owing to the ability of patellamide D to reverse the effects of multidrug resistance in human leukaemia cells (Williams & Jacobs, 1993; Housen & Jaspars, 2010). These natural products are too structurally complex for cost-effective *de novo* chemical synthesis, and this prevents the creation of novel and diverse analogues with desired properties. Manipulation of the natural product’s biosynthetic pathway can offer an alternative, more cost-effective, route to such analogues, but does require a comprehensive understanding of the individual enzymes involved. Moreover, the use of the isolated enzymes can allow the introduction of non-natural functionalities.

Patellamides originate from a longer precursor peptide, which is modified by a series of post-translational modifying (PTM) enzymes to yield the final cyclic product. Their biosynthesis can be attributed to a single operon (*patA–patG*) encoding the precursor peptide (PatE) and the enzymes required for processing (PatA, PatD and PatG; Schmidt *et al.*, 2005; Fig. 1). The precursor peptide includes a core peptide (or multiple core peptides) that becomes the natural product, as well as conserved recognition sequences and an N-terminal leader peptide, which are discarded during maturation but are required to direct post-translational tailoring. It is assumed, but has not been proven, that all of the chemical tailoring steps are required for biological activity. All but one of the chemical transformations for the tailoring steps have been assigned to domains



**Figure 1** Cartoon schematic summarizing patellamide biosynthesis. All of the required biosynthetic machinery is clustered into a single operon. The precursor peptide (PatE) is processed by PatD, generating thiazoline and oxazoline heterocycles, cleaved at the N-terminus by PatA and cleaved at the C-terminus, macrocyclized and oxidized by PatG to form highly modified patellamides.

within the processing enzymes (Agarwal *et al.*, 2012; Schmidt *et al.*, 2005). The first step is the heterocyclization of PatE core peptide cysteines, serines and threonines by PatD. This is followed by removal of the N-terminal recognition elements and leader peptide by the protease domain of PatA. Finally, PatG catalyzes the macrocyclization of the modified core peptide, simultaneously removing the C-terminal recognition sequence. Additionally, PatG is responsible for the oxidation of thiazolines to thiazoles (Fig. 1). Structural and mechanistic insights have been reported for the protease domain of PatA (Houssen *et al.*, 2012; Agarwal *et al.*, 2012), the macrocyclization domain of PatG (Koehnke *et al.*, 2012; Agarwal *et al.*, 2012), the heterocyclase TruD (Koehnke *et al.*, 2013; a close homologue of PatD from the trunkamide pathway) and the apparently inactive prenyl transferase PatF (Bent *et al.*, 2013). Only epimerization (which is likely to precede thiazoline oxidation) is unassigned to a protein domain, but it has been suggested that this could be chemically spontaneous (Milne *et al.*, 2006). Whether the epimerization and oxidation occur in the linear or the macrocyclic form is unknown. Two gene products, PatB and PatC, have been reported as non-essential to synthesis, and their role is unknown (Schmidt *et al.*, 2005).

PatA and PatG both contain C-terminal domains of unknown function (DUFs) and these domains are homologous (56% identical

to each other (Fig. 2a). The gene encoding PatG-DUF (amino-acid residues 914–1191) was amplified from gDNA isolated from *Prochloron* sp. that were isolated from Pacific reef samples. The exact strain could not be determined and the amino-acid sequence of PatG-DUF (PatG-DUF<sub>sp.</sub>) differs at three points (E958D, R959Y and V1037M) from the deposited amino-acid sequence of PatG-DUF isolated from *P. didemni* (PatG-DUF<sub>di.</sub>; accession No. AAY21556.1; Fig. 2b). It is unclear as to whether these differences arose as a result of natural variance amongst species or owing to a cloning artefact.

The DUF domain is conserved in all patellamide-like cyanobactin biosynthetic pathways characterized to date (Supplementary Fig. S1<sup>1</sup>), suggesting it may play a functional role. To gain some insight into the DUF domain, we have determined the crystal structure of PatG-DUF<sub>sp.</sub> to a resolution of 1.72 Å, and using isothermal titration calorimetry (ITC) and nuclear magnetic resonance spectroscopy (NMR) we have probed its binding to some plausible substrates: the precursor peptide and intermediates formed during biosynthesis (Fig. 1).

<sup>1</sup> Supporting information has been deposited in the IUCr electronic archive (Reference: SX5117).

2. Methods

2.1. Expression of native and SeMet PatG-DUF

The C-terminal domain of PatG from *Prochloron* sp. (PatG-DUF<sub>sp.</sub>; amino-acid residues 914–1191) was cloned from full-length *patG* into the pHISTEV plasmid with an N-terminal His<sub>6</sub> tag and a *Tobacco etch virus* (TEV) protease site (Liu & Naismith, 2009). The protein was expressed in *Escherichia coli* BL21 (DE3) cells grown in Luria–Bertani (LB) medium supplemented with 50 µg ml<sup>-1</sup> kanamycin. Cultures were grown at 37°C and 200 rev min<sup>-1</sup> until the OD<sub>600</sub> reached 0.6; the cells were then induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG; final concentration 1 mM) and further grown at 18°C and 200 rev min<sup>-1</sup> overnight.

L-Selenomethionine-labelled (SeMet) PatG-DUF<sub>sp.</sub> was expressed from *E. coli* BL21 (DE3) cells grown in minimal medium supplemented with glucose-free nutrient mix, 50 µg ml<sup>-1</sup> kanamycin and 5% glycerol. This medium was inoculated with an overnight culture grown in LB, which was washed three times in minimal medium. After 15 min growth at 37°C, 60 mg l<sup>-1</sup> L-selenomethionine was added. The cultures were returned to 37°C and 200 rev min<sup>-1</sup> and grown until the OD<sub>600</sub> reached 0.6, when 100 mg l<sup>-1</sup> lysine, phenylalanine, threonine and 50 mg l<sup>-1</sup> isoleucine and valine were added. After incubation for a further 20 min, expression was induced with IPTG (final concentration 1 mM) and the cells were grown at 18°C for 24 h. For both native and SeMet variants, the cells were harvested by centrifugation (4000g, 4°C, 15 min).

2.2. Purification of native and SeMet PatG-DUF

Cell pellets were resuspended in lysis buffer [20 mM bis-tris pH 6.8, 150 mM NaCl, 20 mM imidazole, 3 mM β-mercaptoethanol (BME), 0.1% (v/v) Triton X-100] supplemented with 0.4 mg DNase per gram of wet cell pellets and cOmplete EDTA-free protease-inhibitor tablets (Roche; one tablet per 50 ml resuspension). The cells were lysed *via* passage through a cell disruptor at 207 MPa (Constant Systems) and the cell debris was removed by centrifugation (40 000g, 4°C, 20 min). The supernatant was loaded onto a pre-equilibrated Ni-NTA column at 4°C and the column was washed with 80 column volumes of lysis buffer. The protein was eluted from the column in elution buffer [20 mM bis-tris pH 6.8, 150 mM NaCl, 250 mM imidazole, 3 mM BME, 0.1% (v/v) Triton X-100] and fractions containing protein were pooled and passed over a desalting column (Desalt 16/10, GE Healthcare) to exchange them into ‘low imidazole’ buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 mM imidazole, 3 mM BME, 10% glycerol). TEV protease was added at a mass:mass ratio of 1 mg of TEV to 10 mg of protein and the sample was incubated for 2 h at 20°C to remove the N-terminal His<sub>6</sub> tag. The digested sample was passed over a second Ni-NTA column and the flow-through was collected and loaded onto a Superdex 75 gel-filtration column (GE Healthcare) equilibrated in buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP, 10% glycerol). The purity of the protein was determined by SDS-PAGE and its integrity was determined by mass spectrometry. Analysis of the mass spectrum of SeMet

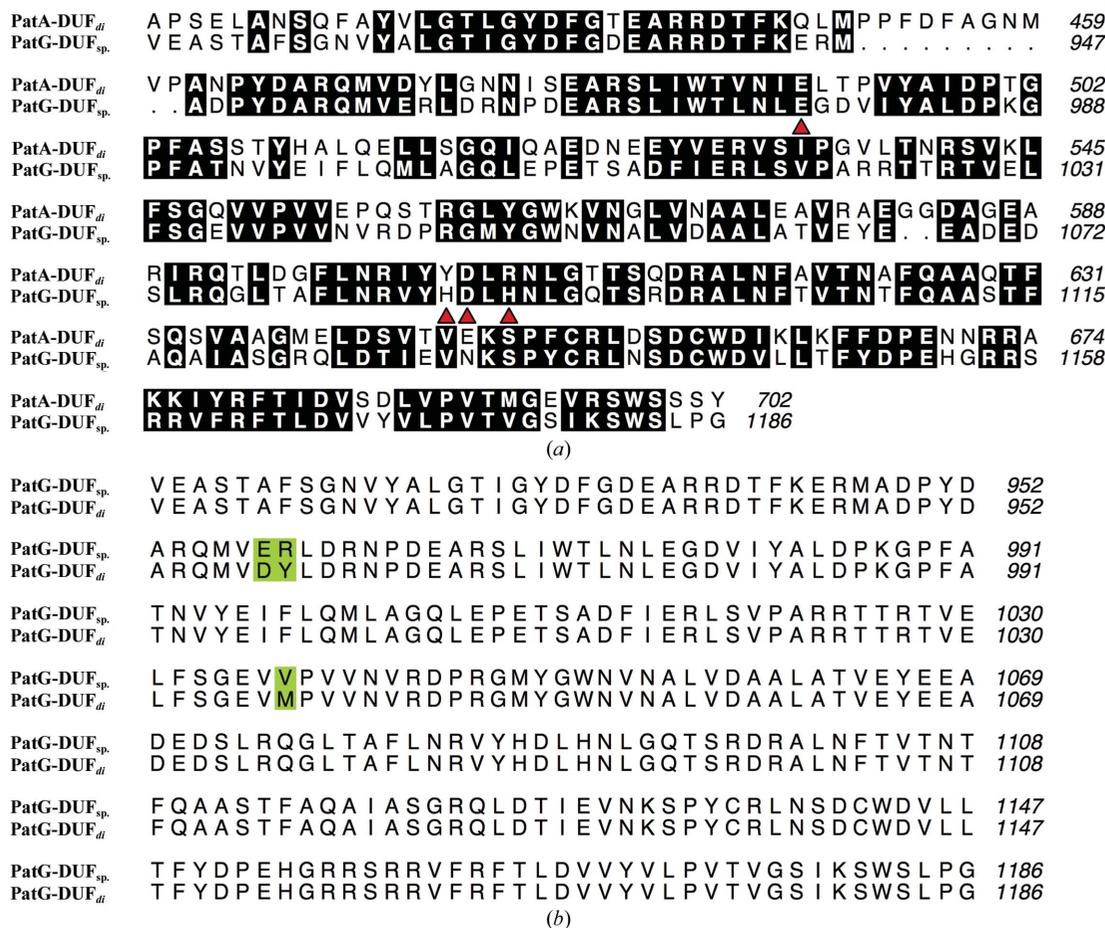


Figure 2 (a) Sequence alignment of PatA-DUF and PatG-DUF<sub>sp.</sub>. The conserved C-terminal DUFs of PatA and PatG share 56% identity. Residues found to coordinate Zn<sup>2+</sup> in PatG-DUF<sub>sp.</sub> are identified by red triangles. (b) Sequence alignment of PatG-DUF cloned from DNA isolated from *Prochloron* sp. (PatG-DUF<sub>sp.</sub>) and PatG-DUF isolated from *P. didemni* (PatG-DUF<sub>di</sub>). Sequence differences are highlighted in green.

PatG-DUF<sub>sp</sub>, confirmed the successful incorporation of the four expected SeMet residues. Native and SeMet PatG-DUF<sub>sp</sub> were concentrated to 4.75 mg ml<sup>-1</sup> for crystallography.

2.3. Mutagenesis

Mutagenesis of PatG-DUF<sub>sp</sub> was performed using established protocols (Liu & Naismith, 2008).

2.4. CD spectroscopy

Near-UV CD spectra were recorded on a Jasco J-810 spectropolarimeter with samples in 10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP, 10% glycerol.

2.5. Crystallography

Crystal screens were set up with the sitting-drop vapour-diffusion method using a Gryphon robot (Art Robbins). The protein was screened against sparse-matrix screens composed of a range of known crystallization conditions at 20 and 4°C (Jancarik & Kim, 1991). Diffraction-quality crystals were grown in a condition consisting of 0.04 M potassium phosphate, 16%(w/v) PEG 8000, 20%(v/v) glycerol. Crystals of SeMet PatG-DUF<sub>sp</sub> grew using the same condition as the native protein. A single SeMet PatG-DUF<sub>sp</sub> crystal was cryoprotected in mother liquor and flash-cooled in liquid nitrogen. A single-wavelength anomalous dispersion (SAD) data set was collected at the Se K absorption edge at 100 K on beamline I02 at Diamond Light Source. The structure was solved using *AutoSol* and the chains were built into electron density using *AutoBuild* from the *PHENIX* crystallography suite (Adams *et al.*, 2010). Examination of the identified anomalous scattering atoms revealed the Se atoms of the four methionine residues as expected, but also an additional atom at the interface between one monomer and its symmetry mate. The atom was coordinated by two histidine residues, an aspartic acid and a

glutamic acid. Based on the anomalous scattering and coordination, we assign this atom as a Zn<sup>2+</sup> ion. A map using the phases calculated from the anomalous scattering atoms and a map after density modification of these phases are shown in Supplementary Fig. S2. The structure was manually rebuilt in *Coot* (Emsley & Cowtan, 2004) and refined using *REFMAC5* (Murshudov *et al.*, 2011). TLS restraints were generated for refinement using the *TLSMD* server (Painter & Merritt, 2006) and the structure was validated using *MolProbity* (Chen *et al.*, 2010). The Zn<sup>2+</sup> ion refined with a temperature factor similar to those of the ligating atoms and no additional *F<sub>o</sub> - F<sub>c</sub>* density was observed at the site, supporting our identification of the atom as a Zn<sup>2+</sup> ion. Full structural statistics are given in Table 1.

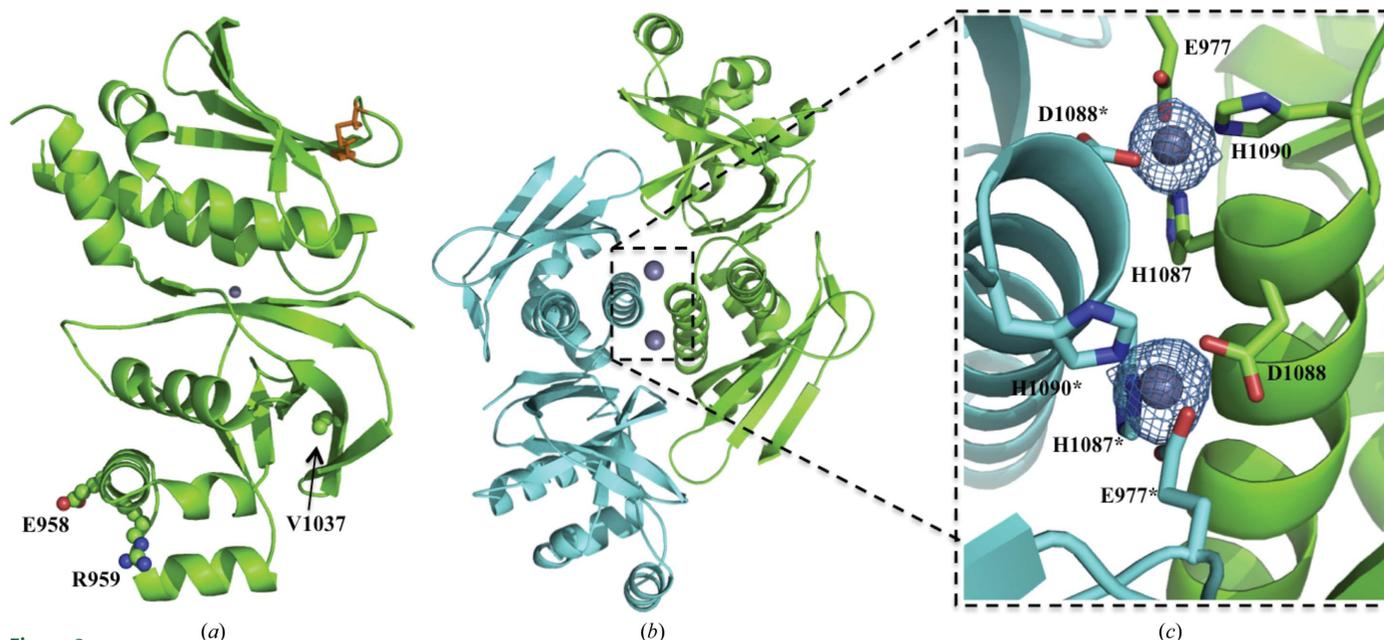
All sequence alignments were created using *Clustal Omega* (Sievers *et al.*, 2014) and presented using *ALINE* (Bond & Schüttelekopf, 2009). All structural images were generated using *PyMOL* (DeLano, 2002).

2.6. Expression and purification of <sup>15</sup>N PatE and PatD

For biochemical characterization an engineered PatE variant was used denoted PatE'. PatE' harbours a single core peptide with the sequence ITACITFC (corresponding to the natural product patellamide D) and a C-terminal His<sub>6</sub> tag. PatE' and <sup>15</sup>N-labelled PatE' were produced as described previously (Koehnke *et al.*, 2013). The enzyme PatD was purified in the same way as previously described for its homologue TruD (Koehnke *et al.*, 2013).

2.7. Heterocyclization of PatE with PatD

<sup>15</sup>N PatE' (100 μM) was incubated with PatD (5 μM) in PatE' gel-filtration buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 1 mM TCEP) and the product was purified using size-exclusion chromatography as described previously (Koehnke *et al.*, 2013). The reaction was monitored using MALDI-TOF MS as described previously (Koehnke *et al.*, 2013).



**Figure 3** (a) X-ray crystal structure of PatG-DUF<sub>sp</sub>, represented as a cartoon. The disulfide bond between Cys1136 and Cys1142 is shown as orange sticks and the Zn<sup>2+</sup> ion is shown as a grey sphere. PatG-DUF<sub>sp</sub> amino acids which differ from those in PatG-DUF<sub>at</sub> are highlighted in ball-and-stick representation for clarity. (b) X-ray crystal structure of the PatG-DUF<sub>sp</sub> dimer represented as a cartoon. (c) Enlargement of the Zn<sup>2+</sup>-coordination site. Glu977, His1087, Asp1088 and His1090 are shown as green sticks and Glu977\*, His1087\*, Asp1088\* and His1090\* are shown as cyan sticks. Difference electron density (*F<sub>o</sub> - F<sub>c</sub>*) contoured at 3σ with phases calculated from a model which was refined without Zn<sup>2+</sup> present is shown as a blue isomesh.

**Table 1**Data-collection and refinement statistics for SeMet PatG-DUF<sub>sp</sub>.

Anomalous data were collected using a single crystal on beamline I02 at the Diamond Light Source. Statistics are average values; values for the highest resolution shell are given in parentheses.	
Data collection	
Wavelength (Å)	0.9797
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>2</sub>
Unit-cell parameters (Å, °)	<i>a</i> = 64.1, <i>b</i> = 96.2, <i>c</i> = 40.2, $\alpha = \beta = \gamma = 90.0$
Resolution (Å)	53.17–1.72 (1.77–1.72)
$\langle I/\sigma(I) \rangle$	12.1 (3.4)
<i>R</i> <sub>merge</sub> (%)	14.5 (89.6)
Completeness (%)	99.9 (100)
Multiplicity	14.0 (11.3)
Anomalous completeness (%)	99.9 (99.9)
Anomalous multiplicity	7.3 (5.7)
Initial map correlation	0.4
Refinement	
<i>R</i> factor (%)	17.5 (23.9)
<i>R</i> <sub>free</sub> (%)	20.0 (29.5)
R.m.s.d., bond lengths (Å)	0.005
R.m.s.d., bond angles (°)	0.932
No. of non-H atoms	
Protein atoms	2109
Solvent atoms	269
Heterogen atoms	1
<i>B</i> factors (Å <sup>2</sup> )	
All	26.0
Protein	25.1
Ligand (Zn <sup>2+</sup> )	12.9
Water	32.8

## 2.8. ITC and data analysis

ITC experiments were performed using a VP-ITC instrument (MicroCal) in PatE' gel-filtration buffer at 20°C. Both PatE' and PatG-DUF<sub>di</sub> were dialyzed into freshly prepared PatE' gel-filtration buffer for 2 d (changing to fresh buffer after the first day) at 4°C to ensure that the buffers for the experiment were matched. A cell solution of 40 μM PatG-DUF<sub>di</sub> and a syringe solution of 600 μM PatE' were prepared (using dialysis buffer to dilute the samples) and degassed at 18°C for 15 min. The PatE' solution was titrated into the PatG-DUF<sub>di</sub> solution as follows: an initial injection of 2 μl followed by injections of 5 μl at a speed of 0.5 μl min<sup>-1</sup> with a delay of 4 min between injections. The syringe was stirred at 307 rev min<sup>-1</sup> for the duration of the experiment. The raw data were processed using the MicroCal *Origin* software. The baseline was adjusted and the integrations were performed manually.

## 2.9. NMR binding experiments

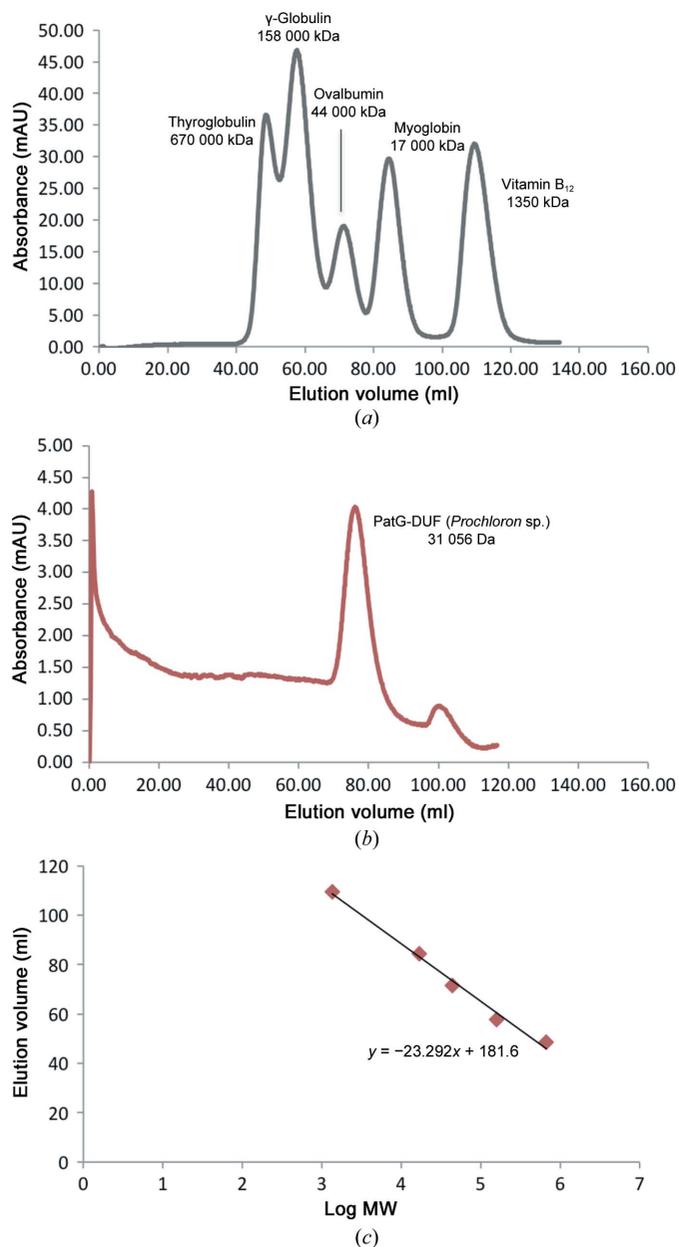
NMR experiments were performed at 10°C in a Bruker DRX500 spectrometer equipped with a 5 mm TXIz probe. The instrument was run using the *TopSpin* software (Bruker). The sample for binding experiments consisted of 100 μM <sup>15</sup>N PatE' in PatE' gel-filtration buffer supplemented with 5% D<sub>2</sub>O. and a <sup>1</sup>H-<sup>15</sup>N HSQC spectrum was recorded. Aliquots of 1.6 mM PatG-DUF<sub>di</sub> were added to final concentrations of 50, 100 and 200 μM and HSQCs were recorded after each addition. The spectra were overlaid using *TopSpin* to identify changes (if any) in the spectra subsequent to the addition of PatG-DUF<sub>di</sub>. The <sup>1</sup>H-<sup>15</sup>N HSQC spectra were acquired with Watergate water suppression (Piotto *et al.*, 1992) at 1024 × 128 points and a digital resolution of 9.8 and 19.0 Hz for the <sup>1</sup>H and <sup>15</sup>N dimensions, respectively. The experiment was repeated using the modified <sup>15</sup>N PatE' containing four heterocycles under the same conditions. This resulted in a poorer signal-to-noise ratio, and therefore the experiment was repeated using 250 μM modified <sup>15</sup>N PatE' and aliquots of 2 mM PatG-DUF<sub>di</sub> were added to final concentrations of 125, 250 and 500 μM.

## 3. Results

### 3.1. Crystal structure of PatG-DUF

The gene encoding PatG-DUF<sub>sp</sub> (amino-acid residues 914–1191) was amplified from gDNA isolated from *Prochloron* sp. Since we cannot be certain as to whether the variations between PatG-DUF<sub>sp</sub> and PatG-DUF<sub>di</sub> reflect natural variation amongst *Prochloron* species or have arisen from errors in the original amplification, we used site-directed mutagenesis to generate PatG-DUF with the sequence from *P. didemni*. Both PatG-DUF<sub>sp</sub> (Glu958, Arg959 and Val1037) and PatG-DUF<sub>di</sub> (Asp958, Tyr959 and Met1037) were purified as described in §2 and a near-UV CD spectrum revealed no significant difference (Supplementary Fig. S3). We concluded that either sequence was suitable for use in structural characterization. Crystallography was carried out on PatG-DUF<sub>sp</sub> as it had been purified first and gave crystals. We carried out biophysical studies with PatG-DUF<sub>di</sub> as we felt that these data would be more informative with a known database match.

The absence of PatG-DUF homologues in the PDB required the generation of a SeMet PatG-DUF<sub>sp</sub> derivative, and the crystal structure was solved using single-wavelength anomalous dispersion. The SeMet PatG-DUF<sub>sp</sub> crystal belonged to space group *P*2<sub>1</sub>2<sub>2</sub> and contained one monomer in the asymmetric unit. The structure has a fold comprising seven α-helices and 13 β-strands. Analysis with secondary-structure matching at the European Bioinformatics Institute (<http://www.ebi.ac.uk/msd-srv/ssm/>) and *TopSearch* (<https://topsearch.services.came.sbg.ac.at>) both suggest that the fold is novel, with the closest matches being a series of unrelated enterotoxins giving r.m.s.d. values of between 3 and 4 Å over ~80 residues (Krissinel & Henrick, 2004; Wiederstein *et al.*, 2014). The refined model contains residues 920–1064 and 1077–1184 (the residue numbers correspond to full-length PatG). The missing residues are at the N- and C-termini of the model and in a connecting loop between α-helices 5 and 6, and are presumed to be disordered. The model includes one disulfide bond between Cys1136 and Cys1142 and a single Zn<sup>2+</sup> ion (Fig. 3a). Glu958 and Arg959 are located on the opposite face from the Zn<sup>2+</sup> ion, whilst Val1037 is located on the edge of a hydrophobic pocket remote from the zinc (Fig. 3a). The changes in sequence would seem to be unlikely to alter the structure, which is consistent with the near-UV CD spectra. Since zinc was not deliberately added to any of the purification buffers or crystallization conditions, it is presumed the Zn<sup>2+</sup> ion was bound during protein expression. The Zn<sup>2+</sup> ion is coordinated in a tetrahedral arrangement by residues Glu977, His1087 and His1090. The fourth coordination site is occupied by residue Asp1088\* from a symmetry mate (PatG-DUF<sub>sp</sub>\*), creating a dimer with two Zn<sup>2+</sup>-binding sites (Figs. 3b and 3c). Analysis with *PISA* (<http://www.ebi.ac.uk/pdbe/pisa/>; Krissinel & Henrick, 2007) reveals that there are almost no contacts between the protein residues, and removing the Zn<sup>2+</sup> ions *in silico* reduces the complex formation significance score to 0.056, indicating that Zn<sup>2+</sup> is crucial in mediating dimer interactions. Since Zn<sup>2+</sup> was not added, its presence in the crystal could suggest that the dimer is present in solution. However, PatG-DUF (both PatG-DUF<sub>sp</sub> and PatG-DUF<sub>di</sub>) eluted from the gel-filtration column with a retention time consistent with a monomer (Fig. 4), preventing us from a firm assignment of dimer *versus* monomer in solution. Interestingly, the residues involved in Zn<sup>2+</sup> binding are not all conserved in PatA-DUF (Fig. 1a). While the H1087Y substitution might still coordinate Zn<sup>2+</sup>, the H1090R substitution would not. Additionally, these residues are not well conserved in the majority of the PatG-DUF homologues (with notable exceptions being TruG-DUF and ArtG-DUF) or in any of the PatA-DUF homologues (Supplementary Fig. S1). Consequently, we



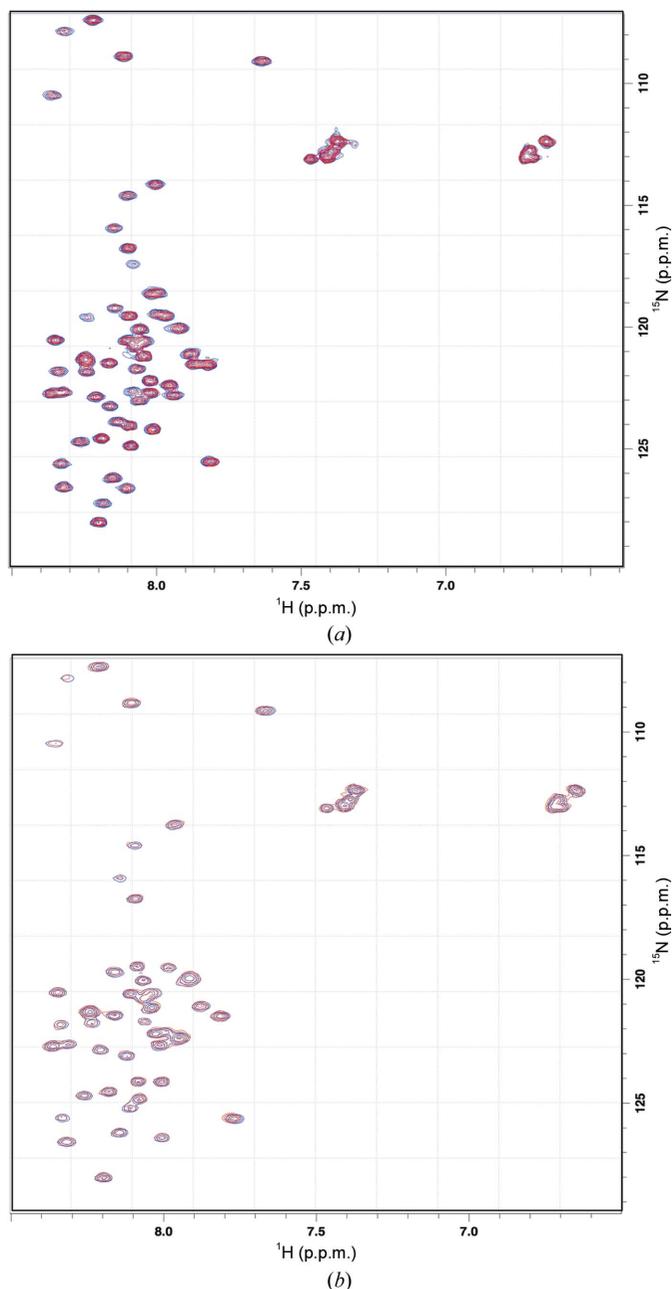
**Figure 4** (a) Superdex 75 gel-filtration chromatogram of gel-filtration standards (Bio-Rad). (b) Superdex 75 gel-filtration chromatogram of PatG-DUF. (c) Standard curve (elution volume versus log MW). Interpolating from the standard curve, we can determine that PatG-DUF elutes from the gel-filtration column as a monomer.

cannot make any claims regarding the biological relevance of a dimeric assembly for the DUF domain.

The structure of PatG-DUF<sub>sp</sub> is novel and therefore provides no immediate insight into its function within patellamide biosynthesis. In an attempt to determine whether the C-terminal DUFs are directly involved in the biosynthesis, binding experiments between PatG-DUF<sub>di</sub> and the precursor peptide PatE', and a modified precursor subsequent to heterocyclization by PatD, were performed using ITC and NMR.

### 3.2. Investigating potential binding partners for PatG-DUF<sub>di</sub>

Titration unmodified PatE' into a solution of PatG-DUF<sub>di</sub> reproducibly results in a binding curve (Supplementary Fig. S4). However,



**Figure 5** (a) Overlaid <sup>1</sup>H-<sup>15</sup>N HSQC NMR spectra of PatE' before (blue) and after (red) the addition of 2.0 equivalents of PatG-DUF<sub>di</sub>. (b) Overlaid <sup>1</sup>H-<sup>15</sup>N HSQC NMR spectra of heterocyclized PatE' before (blue) and after (red) the addition of 2.0 equivalents of PatG-DUF<sub>di</sub>.

analysis of the data yields unlikely stoichiometries of around ten molecules of PatG-DUF<sub>di</sub> binding to one molecule of PatE'. Close inspection of the cell revealed heavy precipitation at the end of the experiment, suggesting that we were measuring protein aggregation. Using <sup>15</sup>N-labelled PatE', a <sup>1</sup>H-<sup>15</sup>N HSQC spectrum was recorded in the absence and presence of 0.5, 1.0 and 2.0 molar equivalents of PatG-DUF<sub>di</sub> (Supplementary Fig. S5a). The spectra of pure <sup>15</sup>N PatE' in the absence (blue) and presence (red) of 2.0 equivalents of PatG-DUF<sub>di</sub> have been overlaid (Fig. 5a), and clearly show an identical chemical shift pattern, showing that PatG-DUF<sub>di</sub> does not bind the unmodified precursor peptide. The NMR experiments were repeated using <sup>15</sup>N PatE' that had been processed with PatD, introducing four heterocycles (two thiazolines and two oxazolines) within the context

of the core peptide. As before, the addition of up to 2.0 equivalents of PatG-DUF<sub>di</sub> to the modified PatE' resulted in an identical NMR spectrum (Fig. 5b and Supplementary Fig. S5b), indicating that the heterocycle-containing peptide does not bind PatG-DUF<sub>di</sub> either. The known disorder of the leader peptide of PatE (Koehnke *et al.*, 2013) would suggest that blocking of PatG-DUF<sub>di</sub> binding by the leader peptide is unlikely. At present, we are unable to produce labelled cleaved precursor or macrocycle in sufficient quantities and purity for NMR experiments.

#### 4. Conclusions

The patellamide proteins PatA and PatG contain C-terminal domains of unknown function which are 56% identical to each other and are conserved in related cyanobactin biosynthetic pathways. X-ray crystallography revealed the structure of PatG-DUF<sub>sp</sub> to be a novel fold and, although a dimer is observed in the crystal, the residues involved in Zn<sup>2+</sup> coordination (which is required for dimerization) are not well conserved in PatA-DUF or in homologues from other pathways. Consequently, it is difficult to speculate as to the functional importance of the dimer. We have been unable to detect any binding to linear substrates (simple peptides or peptides with heterocycles), indicating that these are unlikely to be substrates or ligands of the DUFs. The pitfall in using one technique to assess binding was shown by the reproducible but functionally meaningless ITC binding curves that we observed. Further study is required with macrocycles before we can rule out any binding of the DUF domain to patellamide substrates. It remains possible that the dimerization of the domain indicates an interaction between the DUFs, potentially leading to homodimers of PatA/PatG or even heterodimers of PatA and PatG, which might be important for the activity of these enzymes *in vivo*. Further study is required to investigate such hypotheses. If the DUF domain has no enzymatic role, coupled with the observation that the protease domains of PatA and PatG are active in isolation, it is interesting as to why it is conserved throughout cyanobactin biosynthetic pathways. However, it also suggests that it could be excluded from *in vitro* enzyme-based synthesis systems. This simplifies the design and use of a biotechnological 'toolkit' for bioactive cyclic and linear peptides.

We thank the staff of beamline I02 at the Diamond Light Source for their support with data collection. Thanks go to the NMR and BSRC mass-spectrometry facilities at the University of St Andrews for crucial support. This work was funded by the BBSRC (BB/

K015508/1) and ERC (TNT-LEAP), and the University of St Andrews infrastructure is supported by a Wellcome Trust Capital Award (086036). WH is the recipient of the SULSA Leaders award.

#### References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Agarwal, V., Pierce, E., McIntosh, J., Schmidt, E. W. & Nair, S. K. (2012). *Chem. Biol.* **19**, 1411–1422.
- Bent, A. F., Koehnke, J., Houssen, W. E., Smith, M. C. M., Jaspars, M. & Naismith, J. H. (2013). *Acta Cryst.* **F69**, 618–623.
- Bond, C. S. & Schüttelkopf, A. W. (2009). *Acta Cryst.* **D65**, 510–512.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Houssen, W. E. & Jaspars, M. (2010). *Chembiochem*, **11**, 1803–1815.
- Houssen, W. E., Koehnke, J., Zollman, D., Vendome, J., Raab, A., Smith, M. C. M., Naismith, J. H. & Jaspars, M. (2012). *Chembiochem*, **13**, 2683–2689.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Koehnke, J., Bent, A., Houssen, W. E., Zollman, D., Morawitz, F., Shirran, S., Vendome, J., Nneoyigbe, A. F., Trembleau, L., Botting, C. H., Smith, M. C. M., Jaspars, M. & Naismith, J. H. (2012). *Nature Struct. Mol. Biol.* **19**, 767–772.
- Koehnke, J., Bent, A. F., Zollman, D., Smith, K., Houssen, W. E., Zhu, X., Mann, G., Lebl, T., Scharff, R., Shirran, S., Botting, C. H., Jaspars, M., Schwarz-Linek, U. & Naismith, J. H. (2013). *Angew. Chem. Int. Ed. Engl.* **52**, 13991–13996.
- Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* **D60**, 2256–2268.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Liu, H. & Naismith, J. H. (2008). *BMC Biotechnol.* **8**, 91.
- Liu, H. & Naismith, J. H. (2009). *Protein Expr. Purif.* **63**, 102–111.
- Milne, B. F., Long, P. F., Starcevic, A., Hranueli, D. & Jaspars, M. (2006). *Org. Biomol. Chem.* **4**, 631–638.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Painter, J. & Merritt, E. A. (2006). *Acta Cryst.* **D62**, 439–450.
- Piotto, M., Saudek, V. & Sklenář, V. (1992). *J. Biomol. NMR*, **2**, 661–665.
- Schmidt, E. W., Nelson, J. T., Rasko, D. A., Sudek, S., Eisen, J. A., Haygood, M. G. & Ravel, J. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 7315–7320.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2014). *Mol. Syst. Biol.* **7**, 539.
- Sivonen, K., Leikoski, N., Fewer, D. P. & Jokela, J. (2010). *Appl. Microbiol. Biotechnol.* **86**, 1213–1225.
- Wiederstein, M., Gruber, M., Frank, K., Melo, F. & Sippl, M. J. (2014). *Structure*, **22**, 1063–1070.
- Williams, A. B. & Jacobs, R. S. (1993). *Cancer Lett.* **71**, 97–102.