

Accepted Manuscript

Expression and purification of an adenylation domain from a eukaryotic nonribosomal peptide synthetase: Using structural genomics tools for a challenging target

T. Verne Lee, J. Shaun Lott, Richard D. Johnson, Vickery L. Arcus

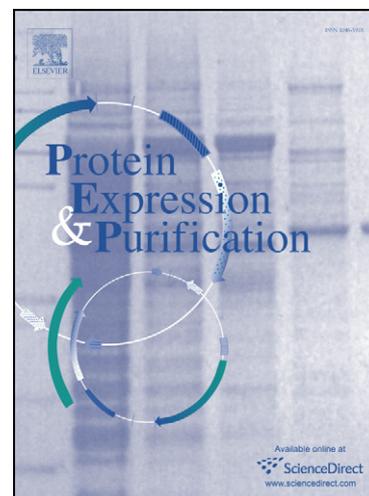
PII: S1046-5928(10)00236-6
DOI: [10.1016/j.pep.2010.08.004](https://doi.org/10.1016/j.pep.2010.08.004)
Reference: YPREP 3782

To appear in: *Protein Expression and Purification*

Received Date: 15 April 2010
Revised Date: 4 August 2010

Please cite this article as: T.V. Lee, J.S. Lott, R.D. Johnson, V.L. Arcus, Expression and purification of an adenylation domain from a eukaryotic nonribosomal peptide synthetase: Using structural genomics tools for a challenging target, *Protein Expression and Purification* (2010), doi: [10.1016/j.pep.2010.08.004](https://doi.org/10.1016/j.pep.2010.08.004)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Expression and purification of an adenylation domain from a eukaryotic nonribosomal peptide synthetase: Using structural genomics tools for a challenging target

T. Verne Lee^a, J. Shaun Lott^a, Richard D. Johnson^b, and Vickery L. Arcus^{a,c*}

^aAgResearch Structural Biology Laboratory, School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand.

^bAgResearch Ltd., Grasslands Research Centre, Tennent Drive, Private Bag 11008, Palmerston North 4442, New Zealand.

^cDepartment of Biological Sciences, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand.

Email addresses: t.lee@auckland.ac.nz (T.V. Lee), s.lott@auckland.ac.nz (J.S. Lott), richard.johnson@agresearch.co.nz (R.D. Johnson), varcus@waikato.ac.nz (V.L. Arcus)

*Corresponding author:

Vickery L. Arcus

Department of Biological Sciences

Faculty of Science and Engineering

University of Waikato

Private Bag 3105

Hamilton 3240

New Zealand

Ph: 64 7 838 4679

Fax: 64 7 838 4976

E-mail: varcus@waikato.ac.nz

Abstract

Nonribosomal peptide synthetases (NRPSs) are large multimodular and multidomain enzymes that are involved in synthesising an array of molecules that are important in human and animal health. NRPSs are found in both bacteria and fungi but most of the research to date has focused on the bacterial enzymes. This is largely due to the technical challenges in producing active fungal NRPSs, which stem from their large size and multidomain nature. In order to target fungal NRPS domains for biochemical and structural characterisation, we tackled this challenge by using the cloning and expression tools of structural genomics to screen the many variables that can influence the expression and purification of proteins. Using these tools we have screened 32 constructs containing 16 different fungal NRPS domains or domain combinations for expression and solubility. Two of these yielded soluble protein with one, the third adenylation domain of the SidN NRPS (SidNA3) from the grass endophyte *Neotyphodium lolii*, being tractable for purification using Ni-affinity resin. The initial purified protein exhibited poor solution behaviour but optimisation of the expression construct and the buffer conditions used for purification, resulted in stable recombinant protein suitable for biochemical characterisation, crystallisation and structure determination.

Keywords: protein screening, protein expression, nonribosomal peptide synthetases, fungi, natural products, siderophore

Introduction

Nonribosomal peptide synthetases (NRPSs)¹ make up an important class of biosynthetic enzymes that are involved in the production of a large array of short peptides, including many molecules that are used in the context of human and animal health [1]. To date, NRPSs have only been confirmed to be present in bacteria and fungi [2], although putative gene products with NRPS-like signatures have been identified in *Oryza sativa* and *Caenorhabditis elegans* [3]. NRPSs synthesise peptides by a multiple carrier thiotemplate mechanism [4, 5]. They are large, multimodular enzymes with each module catalysing the incorporation of one subunit into the peptide being synthesised. NRPS modules are, in turn, made up of independently-folding functional domains that catalyse the individual reactions of peptide synthesis. Three domains make up the minimum elongation module, the adenylation (A), peptidyl carrier protein (PCP) and condensation (C) domains. In addition to these core domains, there are a number of auxiliary domains, such as *N*-methylation (Met) and reduction (R) domains, that sometimes make up part of modules and chemically modify the peptide as it is being synthesised.

Despite the fact that many of the important products of NRPSs are from fungi, most NRPS research to date has focused on enzymes of bacterial origin due to the technical difficulty of producing active fungal NRPS enzymes [6]. The most significant technical challenge in producing fungal NRPS proteins is the very large size of these enzymes. For example the peptaibol synthetase from *Trichoderma virens* is made up of 20,925 amino acid residues [7]. To mitigate this problem, research efforts have often focused on individual domains or combinations of small groups of domains. In bacteria, this is assisted by the fact that separate genes often encode individual domains. In fungi however, NRPS enzymes are usually encoded by a single large gene, which is translated into a multimodular, multidomain protein. This often necessitates the expression of individual domains (from the larger multidomain protein), which is a process that is fraught with difficulties [8, 9]. Hence, it is not surprising that there are few examples of the successful production of active fungal NRPS enzymes using heterologous expression systems [10] and the subsequent biochemical characterisation of these enzymes.

One potential strategy to tackle this technical challenge is to make use of the high-throughput cloning and expression screening systems [11, 12] that have been developed for structural genomics programs over the past decade. In the context of structural genomics, the high-throughput tools are used to screen large numbers of genes for the production of soluble proteins that can then be screened for crystal growth and structure determination. However, this approach can also be used to rapidly screen large numbers of constructs from targeted genes to find conditions under which domains can be stably and solubly expressed and purified. This can also facilitate the expression and purification of challenging proteins or protein classes by screening many other variables that can influence the expression and purification of proteins such as expression temperature and cell-lysis conditions [12].

A large number of uncharacterised NRPSs have recently been identified in the *Epichloë/Neotyphodium* complex (phylum Ascomycota, family Clavicipitaceae) of endophytic fungi [13]. The mutualistic relationship between the endophytes and the agronomic grasses that they colonise plays a vital role in pastoral agriculture by improving the grass plants' tolerance to biotic and abiotic stresses [14, 15]. The production of large amounts of soluble protein of domains from endophyte NRPSs would greatly facilitate their characterisation. In particular, the structural analysis of endophyte adenylation (A) domains will be of significant value in improving the substrate specificity prediction methods that currently work well for bacterial A domains but usually fail for A domains of fungal origin [6]. By using a cloning and expression screening system originally developed for structural genomics [16], we have successfully developed an expression construct and purification protocol that allows the production of pure, stable recombinant protein for an endophyte A domain. This, in turn has provided the basis for the first structural elucidation of a eukaryotic NRPS A domain [17].

Materials and Methods

Definition of domain borders of NRPS constructs

Domain border definition was accomplished by the generation of multiple sequence alignments using ClustalW via the ClustalX interface [18]. Known secondary structure elements (if available) were used to set gap penalties for ClustalW alignments. The locations of the domains were approximately estimated and the sequences for each domain were aligned with those from a number of previously published NRPS sequences from fungi and bacteria. The domain boundaries were defined from these alignments based on previously determined structures of bacterial A [19], C [20], and PCP [21] domains. Although no structures for NRPS Met and R domains were available, the boundaries of these domains could be defined based on the adjacent domains (the A domain into which the Met domain is inserted and the PCP domain preceding the R domain).

Gateway cloning

Cloning was carried out using the Gateway cloning system (Invitrogen) as described previously [16]. The NRPS fragments from four NRPS genes were amplified via PCR from *N. loli* Lp19 genomic DNA. A two-step nested PCR reaction was used to introduce the sequence encoding the recombinant Tobacco Etch Virus (rTEV) protease cleavage site into the PCR product which would allow the N-terminal affinity tags to be cleaved from the expressed protein. The PCR primers used are shown in Table S1 and Table S2. A BP reaction using the BP clonase enzyme (Invitrogen) was performed to recombine the amplified fragments into the pDONR221 plasmid. The correct clones were identified by restriction enzyme analysis and verified by sequencing. The fragments were then recombined into the pDEST15 and pDEST17 expression vectors via a LR reaction using the LR clonase enzyme (Invitrogen). As the cloned fragments contain rare codons, the resulting expression constructs were transformed into *E. coli* Rosetta (DE3) for expression. This expression strain harbours a plasmid that encodes the rare tRNA genes for Arg, Ile, Gly, Leu, Pro, Met, Thr and Tyr tRNAs.

Expression testing

Expression testing was performed in 96-well format as described previously [16]. The expression cultures were grown in 1 mL of autoinduction medium [22] containing ampicillin (100 µg/mL) and chloramphenicol (25 µg/mL) in each well of a 96-well deep-well plate. Six replicates each of sixteen expression constructs can conveniently be processed at once. The cultures were incubated with shaking at 800 rpm on a Thermomixer Comfort (Eppendorf) that allows expression to be performed at various temperatures. The cultures were grown at 37 °C for 3 h 30 min and the temperature was then set to the desired expression temperature. The culture was grown at the desired expression temperature for 13 h to 16 h for 37 °C expression, 15 h to 18 h for 28 °C expression, and 24 h to 26 h for 18 °C and 10 °C expression. The cells were harvested by centrifugation (1800 × *g* for 30 min) and resuspended in various lysis buffers. Lysozyme (1 mg/mL) and DNase I (10 µg/mL) were added, the cell suspension was incubated on ice for 30 min and MgCl₂ (1 mM) was added. The cell suspensions were sonicated on ice using an eight-element probe for 4-5 15 s bursts at 80 W using a Vibra-Cell ultrasonic processor (Sonics). The lysed cell suspensions were transferred to a 96-well PCR plate and the plate centrifuged at 3,600 × *g* for 1 h at 4 °C to separate the soluble and insoluble fractions for analysis by SDS-PAGE.

Small-scale affinity tag binding tests

Cultures were grown in autoinduction medium [22] containing ampicillin (100 µg/mL) and chloramphenicol (25 µg/mL). Expression temperatures of 10 °C and 18 °C were used for the Nrps5Av1 and SidNA3 proteins, respectively. The lysis buffers contained 10 mM imidazole and after centrifugation to separate the soluble and insoluble fractions, 30 µL of Profinity IMAC Ni-charged resin (Bio-Rad) was added to the soluble fraction. The mixture was incubated on ice for 1 h with periodic mixing. The sample was centrifuged at 16,000 × *g* for 1 min at 4 °C. The supernatant was discarded and the resin washed by resuspending the resin in 600 µL of lysis buffer, centrifuging the sample at 16,000 × *g* for 1 min at 4 °C, and discarding the supernatant. This washing cycle was repeated two more times and the samples analysed by SDS-PAGE.

Large-scale expression and purification of the SidNA3 domain

The expression culture was grown in 500 mL of autoinduction medium [22] containing ampicillin (100 µg/mL) and chloramphenicol (25 µg/mL) at 37 °C for 3 h 30 min with shaking at 200 rpm. The culture was cooled to 18 °C and grown for 24 h to 26 h at this expression temperature. The cells were harvested by centrifugation (at 3000 × g for 30 min at 4 °C) and resuspended in 6 mL of lysis buffer (25 mM sodium phosphate pH 6.5 or pH 8.0, 500 mM NaCl, 5 mM β-mercaptoethanol). Lysozyme (1 mg/mL), DNase I (10 µg/mL), and a Complete Mini EDTA-free (Roche) tablet were added. The cell suspension was incubated on ice for 30 min and MgCl₂ (1 mM) was added. The cells were lysed by high-pressure cell disruption at 18.5 kPa using a One Shot cell disruptor (ConstantSystems Ltd.). The lysed cell suspension was transferred to a SS-34 tube and centrifuged at 20,000 × g for 30 min at 4 °C.

The first step of the purification of the SidNA3 protein was via immobilised metal affinity chromatography (IMAC). The supernatant was loaded onto a 5 mL HiTrap Chelating HP column (GE Healthcare) loaded with Ni²⁺ ions and pre-equilibrated with lysis buffer. The column was washed with 15 ml to 30 mL lysis buffer containing 10 mM imidazole. The bound proteins were eluted from the column by running a gradient from 10 mM to 500 mM imidazole over 50 mL. The fractions containing the SidNA3 protein were pooled and dialysed against lysis buffer at 4 °C overnight to remove the imidazole. The rTEV protease (recombinant Tobacco Etch Virus NIa protease modified to contain a poly-Histidine tag) was used to cleave the poly-Histidine tag from the SidNA3 protein. The protein sample was supplemented with 0.1 mM EDTA and the rTEV protease added in an approximate rTEV to SidNA3 ratio of 1:20 (w/w). The mixture was incubated overnight at room temperature.

Subtractive IMAC was performed to separate the rTEV, cleaved poly-Histidine tags, and undigested SidNA3 protein from the digested SidNA3 protein. The protein solution was loaded onto a 5 mL HiTrap Chelating HP column (GE Healthcare) loaded with Ni²⁺ ions and pre-equilibrated with lysis buffer. The column was washed with 30 mL lysis buffer

containing 10 mM imidazole. The flow-through and wash (containing the SidNA3 protein) were pooled and concentrated using a 20 mL Vivaspin concentrator (Sartorius AG) with a molecular weight cut off of 30 kDa to a volume of ~0.5 mL.

SEC was used to further purify proteins that had been purified by IMAC and also to remove large soluble aggregates of the target protein. The concentrated protein solution was loaded onto a HiLoad 16/60 Superdex 200 pg column (GE Healthcare) column pre-equilibrated with running buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 5 mM β -mercaptoethanol or 25 mM MES pH 6.5, 500 mM NaCl, 5 mM β -mercaptoethanol) and eluted with one column volume (120 mL) of running buffer at a flow rate of 1 mL/min.

Dynamic light scattering measurements and screening of buffer conditions

Dynamic light scattering (DLS) was used to analyse the state of aggregation and polydispersity of protein samples [23]. Protein samples were centrifuged at $16,000 \times g$ for 10 min at 4 °C, 12 μ L to 15 μ L was placed into a clean, dry quartz cuvette and DLS analysis performed on a DynaPro MTSC Dynamic Light Scattering Instrument (Protein Solutions Inc.).

For screening of buffer conditions via DLS, protein solutions from the fractions collected from the monomer peak of the SEC run were immediately diluted into a series of buffers with a range of pH values (pH 5.5, pH 6.5, pH 7.5 and pH 9.0) and salt concentrations (150 mM NaCl and 500 mM NaCl). The diluted protein solutions were analysed via DLS. The protein solutions were concentrated to ~1.5 mg/mL using 600 μ L Vivaspin concentrators (Sartorius AG) with molecular weight cut offs of 30 kDa and DLS analysis was performed on them again, and a third time after storage at 4 °C for 2 days.

Results and Discussion

We screened a range of constructs of fungal NRPS domains. The boundaries of the domains were estimated based on multiple sequence alignments and comparison to previously published structures of the three core NRPS domains from bacterial enzymes [19-21]. Gateway cloning was used to screen sixteen different domains and domain combinations from four endophyte NRPSs. The four NRPSs were PerA, a recently discovered NRPS which synthesises the insect-feeding deterrent peramine [24]; LpsA, a three-module NRPS which is involved in the biosynthesis of the toxic ergopeptine alkaloid, ergovaline [25, 26]; Nrps5, a novel single-module NRPS with an unknown product [13]; and SidN, a three-module NRPS which is involved in the synthesis of an extracellular siderophore that is essential for the maintenance of the mutualistic character of the grass-endophyte relationship (Linda Johnson, unpublished results). At the time that the domains were cloned, only a fragment of the SidN NRPS gene, encompassing the second and third modules of this NRPS, had been sequenced. Hence, only the SidN domains that were encoded by this DNA fragment were included. The domains and amino acids included in each fragment are detailed in Table 1. These constructs were cloned from *N. lolii* Lp19 genomic DNA using the Gateway cloning system and the pDEST17 and pDEST15 vectors which also incorporate cleavable poly-Histidine and GST tags respectively. The Gateway cloning system results in additional amino acid residues between the fusion tag and target protein due to the required recombination sequences that flank the target sequence. Although, these extra residues have the potential to affect the solubility of the expressed proteins [16], we find that the convenience of this cloning system outweighs this possible downside. To mitigate against these extra residues causing ongoing problems during biochemical characterisation, our approach uses nested PCR to introduce an rTEV protease cleavage site between the N-terminus of the protein and the Gateway recombination site. This results in just three extra amino acids at the N-terminus of the protein following cleavage with rTEV protease.

Expression testing using a range of lysis buffers and at a range of temperatures showed that protein was expressed from all of the constructs (Table 2). Following lysis, soluble protein was produced for two of the constructs (Table 2; Fig. 1A, B), SidNA3 (the third

adenylation domain from the three-module SidN NRPS) and Nrps5Av1 (the adenylation domain from the one-module Nrps5 NRPS and an extra 232 residues at the N-terminus with no homology to any characterised NRPS domain). Low expression temperatures, with a maximum of 18 °C for SidNA3 and 10 °C for Nrps5Av1, were required in order for the proteins to be expressed in soluble form. A general increase in the production of soluble protein by expression at low temperatures has been widely observed [11], and although the reasons for this increase remain unclear, it is likely to be due to the kinetics of protein folding in *E. coli* and the upregulation of chaperones at these temperatures.

Both the poly-Histidine tagged and GST tagged SidNA3 proteins were soluble but only the poly-Histidine tagged version of the Nrps5Av1 was soluble. To check whether the soluble poly-Histidine tagged SidNA3 and Nrps5Av1 proteins were tractable for purification, small-scale affinity tag binding tests were conducted on cultures expressed at 18 °C and 10 °C, respectively. The SidNA3 protein bound to the affinity resin but the Nrps5Av1 protein did not. Repeating the small-scale affinity tag binding tests using a set of different lysis buffers did not result in binding of the Nrps5Av1 protein to the resin. One possible reason why the soluble Nrps5Av1 protein does not bind to Ni-affinity resin is that the N-terminal poly-Histidine tag is buried in, or otherwise obscured by, the structure of the protein and is therefore inaccessible for binding. The additional 232 residues at the N-terminus of the Nrps5Av1, which presumably form a domain of unknown function and structure, may be occluding the short poly-Histidine tag or causing aggregation. Although further work on the Nrps5Av1 protein (such as cloning it into a C-terminal poly-Histidine tag expression vector) may allow it to be purified, we chose to focus on the SidNA3 protein, as it appeared to be more tractable.

Large-scale expression of the poly-Histidine tagged SidNA3 protein was carried out at 18 °C and the protein was purified via immobilised metal affinity chromatography (IMAC). Following dialysis to remove imidazole, the purified SidNA3 protein was digested with the rTEV protease to remove the poly-Histidine tag. Subtractive IMAC was carried out to remove the cleaved poly-Histidine tag, undigested SidNA3 protein and rTEV, although removal of the undigested SidNA3 protein was incomplete. To improve

the purity of the final protein sample, size exclusion chromatography (SEC) using a 16/60 S200 column was performed using 25 mM HEPES pH 7.5 buffer with 150 mM NaCl and 5 mM β -mercaptoethanol. There were two peaks on the 280 nm absorbance trace from SEC (Fig. 2A). The first peak was in the void volume of the column and consisted of protein aggregates (>600 kDa) containing the digested and undigested SidNA3 protein, while the second peak had a retention volume of ~70 mL and consisted of the monomeric digested SidNA3 protein (62.6 kDa; Fig. 2B). When the fractions under the monomeric peak were pooled and SEC repeated, the same two peaks with a similar peak size ratio were observed, showing that the SidNA3 protein was continuing to aggregate. Dynamic light scattering (DLS) analysis of fractions from under the monomeric peak confirmed that the SidNA3 protein was aggregating in solution, with molecular weight estimates of >1000 kDa and multimodal size distributions. In addition, the protein also precipitated slowly on storage at 4 °C.

Aggregation and precipitation are detrimental for biochemical characterisation and crystallisation so efforts were made to improve the solution behaviour of the protein. One potential cause for the aggregation was incorrect identification of the domain boundaries. The presence or absence of even a few residues can have a profound impact on protein folding and stability when individual domains from a multidomain protein are expressed separately [8, 9]. In particular, if the constructs are too large, unstructured stretches at N- or C-terminus may lead to aggregation of the protein. Hence, two shorter constructs of SidNA3 were cloned using the Gateway system. The first, SidNA3v2 (residue range: 2270-2826), had 14 residues removed from the N-terminus compared to the original fragment (residue range: 2256-2826), while the second, SidNA3v3 (residue range: 2270-2811), had a further 15 residues removed from the C-terminus. The use of protein disorder prediction servers such as DISOPRED2 [27] can be helpful in predicting whether regions of protein are likely to be disordered. In the case of SidNA3, the 14 N-terminal residues that were removed from the expression construct were not predicted to be disordered while the 15 residues removed from the C-terminus were predicted to be disordered. Expression testing showed that soluble protein was produced from both of the shorter constructs (data not shown).

Large-scale expression and purification of the poly-Histidine tagged SidNA3v2 and SidNA3v3 proteins was carried out as for the original SidNA3 protein. In both cases the SEC results showed a large reduction in the proportion of the protein that was aggregated when compared to the original SidNA3 construct (Fig. 3A, B). DLS measurements of fractions from under the monomeric peak showed that only a small amount of aggregation was occurring with typical average molecular weight estimates of about 100 kDa (the expected molecular weights are 61.1 kDa and 59.4 kDa) and polydispersity indices (Cp/Rh) of about 28%. However, when the SidNA3v2 and SidNA3v3 protein solutions were concentrated, rapid aggregation was observed via DLS (molecular weight estimates of >1000 kDa). DLS was then used to screen a set of buffers with a range of pH values and NaCl concentrations. It was found that a pH of 6.5 and a NaCl concentration of 500 mM greatly reduced the aggregation of the protein when concentrated and large scale purification by SEC using 25 mM MES pH 6.5 buffer containing 500 mM NaCl and 5 mM β -mercaptoethanol confirmed this (Fig. 4A, B). Protein from fractions under the peaks containing monomeric protein produced excellent DLS results with typical average molecular weight estimates of about 65 kDa and polydispersity indices (Cp/Rh) of between 7% and 15% (indicating a largely monodisperse protein solution). Concentration of the protein up to as much as 15 mg/mL did not result in any detectable aggregation, although over a period of several hours slow aggregation was evident by DLS analysis. The SidNA3v3 protein did, however, show a tendency to precipitate when concentrated. Thus, SidNA3v2 protein was selected for further experiments. SidNA3v2 has been shown to be active and the structure of the domain has been solved by X-ray crystallography; the biochemical characterisation and structure solution is described in detail elsewhere [17].

Conclusions

In summary, using a cloning and expression screening system originally developed for a structural genomics effort [16], we have successfully identified an expression construct and a purification protocol to produce large amounts of a soluble, stable recombinant fungal NRPS A domain - SidNA3 from the grass endophyte *N. lolii*. This is one of the

few examples of success in expressing and purifying active fungal NRPS enzymes [10]. We screened 16 NRPS domains or domain combinations for expression and solubility. These were cloned into two protein expression vectors (using Gateway-based approaches) giving a total of 32 constructs. These 32 constructs were screened at different expression temperatures and under different cell-lysis conditions yielding two different soluble domains. Of these two, one was tractable for purification using Ni-affinity resin, named SidNA3, the third adenylation domain from SidN. However, the initial SidNA3 domain, when purified, exhibited poor solution behaviour, which was unsuitable for structure determination or biochemical characterisation. Optimisation of the SidNA3 expression construct, together with optimisation of the buffer conditions used for purification, resulted in SidNA3 protein that was stable in solution and suitable for both crystallisation and biochemical analysis. This success demonstrates the utility of high-throughput tools, originally developed for structural genomics, in tackling expression and solubility problems for challenging proteins. These tools can be used for both the initial screening of potential targets and optimising problematic expression and purification of particular proteins. Greater use of these tools has the potential to accelerate research efforts in cases where the production of the desired protein is a limiting factor.

Acknowledgements

We thank Linda Johnson for providing up-to-date sequence information for the endophyte NRPS genes and Christine Voisey and Damien Fleetwood for providing the *Neotyphodium lolii* Lp19 genomic DNA. We are grateful to Rachael Goldstone and Nicole Moreland for assistance with the high-throughput cloning and expression system. Verne Lee was supported by an Agricultural and Marketing Research and Development Trust (AGMARDT) doctoral scholarship and a University of Auckland doctoral scholarship. This work was supported by the NZ Foundation for Research Science and Technology contract CX10X0203.

References

- [1] D. Schwarzer, R. Finking, M.A. Marahiel, Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* 20 (2003) 275-287.
- [2] F. Tiburzi, P. Visca, F. Imperi, Do nonribosomal peptide synthetases occur in higher eukaryotes? *IUBMB Life* 59 (2007) 730-733.
- [3] L. Di Vincenzo, I. Grgurina, S. Pascarella, *In silico* analysis of the adenylation domains of the freestanding enzymes belonging to the eucaryotic nonribosomal peptide synthetase-like family. *FEBS J.* 272 (2005) 929-941.
- [4] R. Finking, M.A. Marahiel, Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* 58 (2004) 453-488.
- [5] S.A. Sieber, M.A. Marahiel, Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics. *Chem. Rev.* 105 (2005) 715-738.
- [6] T. Schwecke, K. Goettling, P. Durek, I. Duenas, N.F. Kaeufer, S. Zock-Emmenthal, E. Staub, T. Neuhof, R. Dieckmann, H. von Doehren, Nonribosomal peptide synthesis in *Schizosaccharomyces pombe* and the architectures of ferrichrome-type siderophore synthetases in fungi. *Chembiochem* 7 (2006) 612-622.
- [7] A. Wiest, D. Grzegorski, B.W. Xu, C. Goulard, S. Rebuffat, D.J. Ebbola, B. Bodo, C. Kenerley, Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase. *J. Biol. Chem.* 277 (2002) 20862-20868.
- [8] A.E. Meekhof, S.J. Hamill, V.L. Arcus, J. Clarke, S.M.V. Freund, The dependence of chemical exchange on boundary selection in a fibronectin type III domain from human tenascin. *J. Mol. Biol.* 282 (1998) 181-194.
- [9] A.S. Politou, M. Gautel, C. Joseph, A. Pastore, Immunoglobulin-type domains of titin are stabilized by amino-terminal extension. *FEBS Lett.* 352 (1994) 27-31.
- [10] T. Hornbogen, S.-P. Riechers, B. Prinz, J. Schultchen, C. Lang, S. Schmidt, C. Mügge, S. Turkanovic, R.D. Süssmuth, E. Tauberger, R. Zocher, Functional characterization of the recombinant *N*-methyltransferase domain from the multienzyme enniatin synthetase. *Chembiochem* 8 (2007) 1048-1054.
- [11] N.S. Berrow, K. Bussow, B. Coutard, J. Diprose, M. Ekberg, G.E. Folkers, N. Levy, V. Lieu, R.J. Owens, Y. Peleg, C. Pinaglia, S. Quevillon-Cheruel, L. Salim, C. Scheich, R. Vincentelli, D. Busso, Recombinant protein expression and

- solubility screening in *Escherichia coli*: a comparative study. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 62 (2006) 1218-1226.
- [12] C. Kambach, Pipelines, Robots, Crystals and biology: what use high throughput solving structures of challenging targets? *Curr. Protein Pept. Sci.* 8 (2007) 205-217.
- [13] R. Johnson, C. Voisey, L. Johnson, J. Pratt, D. Fleetwood, A. Khan, G. Bryan, Distribution of NRPS gene families within the *Neotyphodium/Epichloë* complex. *Fungal Genet. Biol.* 44 (2007) 1180-1190.
- [14] C. Gimenez, R. Cabrera, M. Reina, A. Gonzalez-Coloma, Fungal endophytes and their role in plant protection. *Curr. Org. Chem.* 11 (2007) 707-720.
- [15] R. Rodriguez, R. Redman, More than 400 million years of evolution and some plants still can't make it on their own: plant stress tolerance via fungal symbiosis. *J. Exp. Bot.* 59 (2008) 1109-1114.
- [16] N. Moreland, R. Ashton, H.M. Baker, I. Ivanovic, S. Patterson, V.L. Arcus, E.N. Baker, J.S. Lott, A flexible and economical medium-throughput strategy for protein production and crystallization. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 61 (2005) 1378-1385.
- [17] T.V. Lee, L.J. Johnson, R.D. Johnson, A. Koulman, G.A. Lane, J.S. Lott, V.L. Arcus, The structure of a eukaryotic non-ribosomal peptide synthetase adenylation domain which activates a large hydroxamate amino acid in siderophore biosynthesis. *J. Biol. Chem.* 285 (2010) 2415-2427.
- [18] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24 (1997) 4876-4882.
- [19] E. Conti, T. Stachelhaus, M.A. Marahiel, P. Brick, Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *Embo J.* 16 (1997) 4174-4183.
- [20] T.A. Keating, C.G. Marshall, C.T. Walsh, A.E. Keating, The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* 9 (2002) 522-526.

- [21] T. Weber, R. Baumgartner, C. Renner, M.A. Marahiel, T.A. Holak, Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure Fold. Des.* 8 (2000) 407-418.
- [22] F.W. Studier, Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* 41 (2005) 207-234.
- [23] W.W. Wilson, Light scattering as a diagnostic for protein crystal growth - A practical approach. *J. Struct. Biol.* 142 (2003) 56-65.
- [24] A. Tanaka, B.A. Tapper, A. Popay, E.J. Parker, B. Scott, A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory. *Mol. Microbiol.* 57 (2005) 1036-1050.
- [25] P. Damrongkool, A.B. Sedlock, C.A. Young, R.D. Johnson, K.E. Goetz, B. Scott, C.L. Schardl, D.G. Panaccione, Structural analysis of a peptide synthetase gene required for ergopeptine production in the endophytic fungus *Neotyphodium lolii*. *DNA Seq.* 16 (2005) 379-385.
- [26] D.G. Panaccione, R.D. Johnson, J.H. Wang, C.A. Young, P. Damrongkool, B. Scott, C.L. Schardl, Elimination of ergovaline from a grass-*Neotyphodium* endophyte symbiosis by genetic modification of the endophyte. *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 12820-12825.
- [27] J.J. Ward, J.S. Sodhi, L.J. McGuffin, B.F. Buxton, D.T. Jones, Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* 337 (2004) 635-645.

¹*Abbreviations used:* A domain: adenylation domain; C domain: condensation domain; DDM: n-dodecyl- β -D-maltoside; DLS: dynamic light scattering; GST: glutathione S-transferase; IMAC: immobilised metal affinity chromatography; Met or M domain: N-methylation domain; NRPS: nonribosomal peptide synthetase; PCP or P domain: peptidyl carrier domain; R domain: reduction domain; rTEV: recombinant Tobacco Etch Virus NIa; SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis; SEC: size exclusion chromatography.

Figure Legends

Fig. 1. Expression of soluble protein from the SidNA3 and Nrps5Av1 constructs. SDS-PAGE of the soluble fractions from expression tests of the poly-Histidine tagged (A) SidNA3 and (B) Nrps5Av1 proteins. The expression tests were conducted at 18 °C for the SidNA3 protein and 10 °C for the Nrps5Av1 protein. A negative control was run on each gel and for the SidNA3 protein a sample of a culture that was not induced was also run. The pH and the NaCl concentration in each of the lysis buffers are indicated. For the SidNA3 protein one sample was lysed in buffer containing 0.2% n-dodecyl- β -D-maltoside (DDM). The expressed protein is marked with arrows. The gels were stained with Coomassie Blue and key protein standard sizes (kDa) are indicated.

Fig. 2. Size exclusion chromatography of the SidNA3 protein using a 16/60 S200 column. (A) The 280 nm absorbance trace from size exclusion purification of the SidNA3 protein. The fractions collected are marked and numbered. (B) SDS-PAGE of the fractions from the SEC run. The combined fractions (Final) and sample that was loaded onto the column (Load) have also been run on the gel. The gels were stained with Coomassie Blue and key protein standard sizes (kDa) are indicated.

Fig. 3. Size exclusion chromatography of the SidNA3v2 and SidNA3v3 proteins using a 16/60 S200 column and pH 7.5 buffer containing 150 mM NaCl. The 280 nm absorbance traces from the SEC of the (A) SidNA3v2 and (B) SidNA3v3 proteins are shown.

Fig. 4. Size exclusion chromatography of the SidNA3v2 and SidNA3v3 proteins using a 16/60 S200 column and pH 6.5 buffer containing 500 mM NaCl. The 280 nm absorbance traces from the SEC of the (A) SidNA3v2 and (B) SidNA3v3 proteins are shown.

Tables

Table 1

Details of the NRPS fragments cloned using Gateway cloning.

Construct	Gene	NRPS product	Domains included	Residue range included
PerA1P	<i>perA</i>	Peramine	A domain 1, PCP domain 1	262-911
LpsA1	<i>lpsA</i>	Ergovaline	A domain 1	297-876
LpsA1P	<i>lpsA</i>	Ergovaline	A domain 1, PCP domain 1	297-958
LpsA2	<i>lpsA</i>	Ergovaline	A domain 2	1257-1846
LpsA2P	<i>lpsA</i>	Ergovaline	A domain 2, PCP domain 2	1257-1932
LpsA3	<i>lpsA</i>	Ergovaline	A domain 3	2368-2936
LpsA3P	<i>lpsA</i>	Ergovaline	A domain 3, PCP domain 3	2368-3021
Nrps5Av1	<i>nrps5</i>	Unknown	Unknown domain ^a , A domain	1-790
Nrps5Av2	<i>nrps5</i>	Unknown	A domain	233-790
SidNPC2	<i>sidN</i>	Siderophore	PCP domain 2, C domain 2	1710-2268
SidNC2	<i>sidN</i>	Siderophore	C domain 2	1814-2268
SidNA3	<i>sidN</i>	Siderophore	A domain 3	2256-2826
SidNA3P	<i>sidN</i>	Siderophore	A domain 3, PCP domain 3	2256-2912
SidNPC3	<i>sidN</i>	Siderophore	PCP domain 3, C domain 3	2822-3375
SidNC3	<i>sidN</i>	Siderophore	C domain 3	2947-3375
SidNC3P	<i>sidN</i>	Siderophore	C domain 3, PCP domain 4	2947-3473

^aThe Nrps5 NRPS has and additional 232 residues at the N-terminus that do not show homology any characterised NRPS domain.

Table 2

Results of testing the expression constructs made using Gateway cloning.

Construct	pDEST17 (poly-Histidine tag)		pDEST15 (GST tag)	
	Expressed	Soluble ^a	Expressed	Soluble ^a
PerA1P	Yes	No	Yes	No
LpsA1	Yes	No	Yes	No
LpsA1P	Yes	No	Yes	No
LpsA2	Yes	No	Yes	No
LpsA2P	Yes	No	Yes	No
LpsA3	Yes	No	Yes	No
LpsA3P	Yes	No	Yes	No
Nrps5Av1	Yes	Yes (10 °C)	Yes	No
Nrps5Av2	Yes	No	Yes	No
SidNPC2	Yes	No	Yes	No
SidNC2	Yes	No	Yes	No
SidNA3	Yes	Yes (18 °C)	Yes	Yes (18 °C)
SidNA3P	Yes	No	Yes	No
SidNPC3	Yes	No	Yes	No
SidNC3	Yes	No	Yes	No
SidNC3P	Yes	No	Yes	No

^aThe maximum expression temperature at which the protein is soluble is in brackets.

