DOI: 10.1002/anie.201202043

Biosynthetic Engineering

Introduction of a Non-Natural Amino Acid into a Nonribosomal Peptide Antibiotic by Modification of Adenylation Domain Specificity**

*Jenny Thirlway, Richard Lewis, Laura Nunns, Majid Al Nakeeb, Matthew Styles, Anna-Winona Struck, Colin P. Smith, and Jason Micklefield**

Nonribosomal peptides are a group of structurally diverse secondary metabolites, which include a number of important therapeutic agents and agrochemicals.[1,2] The modular architecture of the nonribosomal peptide synthetases (NRPS), responsible for the assembly of these complex peptides, provides an attractive platform from which to engineer the biosynthesis of "non-natural" nonribosomal peptides.[1] For example, NRPS module and domain exchanges have been used to produce cyclic lipopeptide variants of the clinically important antibiotic daptomycin.[3] Active-site modification of adenylation domains, which are responsible for activating substrate amino acids, has also been explored as an alternative approach for engineering NRPS assembly lines.[4–6] Using the archetypal structure of an A-domain, PheA from gramidicin S synthetase,[7] and multiple sequence alignments of known A-domains, it is possible to identify the key active-site residues that are responsible for recognition of substrate amino acids.[8] Whilst these structural insights have been very effective in guiding mutagenesis approaches to alter the substrate specificity of isolated A-domains in vitro,[9] there have been relatively few examples of how this approach can be used to engineer new nonribosomal peptide products in vivo.[4–6] Moreover, only conservative changes involving the exchange of proteinogenic amino acids of similar size and polarity, within peptide structures, have been effected to date using A-domain modifications in vivo. For instance, a directed evolution approach involving saturation mutagenesis of specific A-domain residues resulted in the production of andrimid analogues where the natural Val residue had been replaced by similar hydrophobic amino acids Leu, Ile, Ala, and Phe.[6]

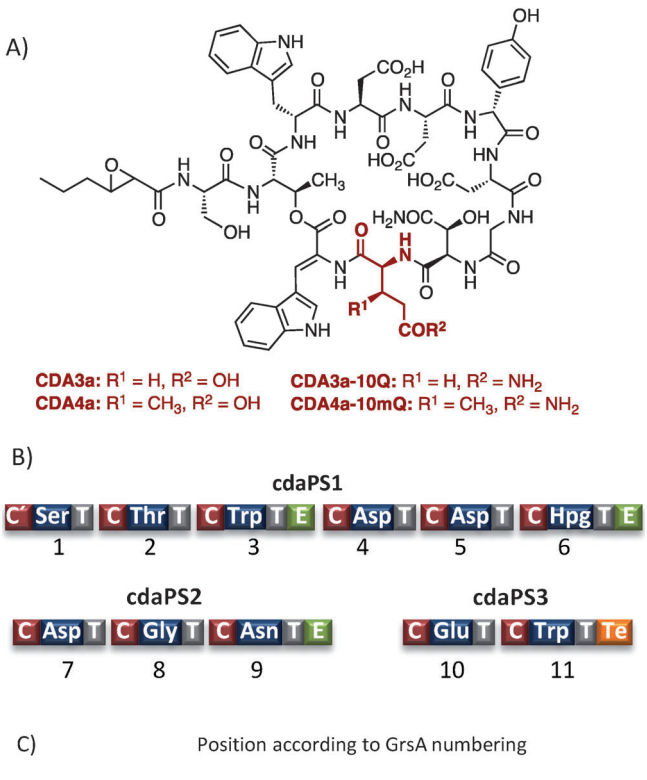
Previously we have been investigating the biosynthesis and biosynthetic engineering of calcium-dependent antibiotics (CDA),[4,10,11] members of the acidic lipopeptide family which also includes daptomycin.[12] Here we show how it is possible to change the specificity of an A-domain within the CDA NRPS assembly line to incorporate a synthetic non-natural amino acid into the decapeptide lactone core of CDA (Figure 1). Our strategy focused on altering the specificity of the module 10 A-domain of CdaPS3 so that it preferentially incorporates (2*S*,3*R*)-3-methyl glutamine (mGln) and Gln over the natural substrates (2*S*,3*R*)-3-methyl glutamic acid (mGlu) and Glu. Multiple sequence alignments of the predicted active site of the module 10 A-domain and related Glu- and Gln-activating A-domains (Figure 1 C) indicate that Glu-activating A-domains tend to possess a basic residue, Lys or His, at either position 239 or 278, which presumably stabilizes the side-chain carboxy group of the substrate through electrostatic interactions. On the other hand, the glutamine-activating A-domains tend to differ by the presence of a Gln rather than a basic residue, at the same relative positions 239 or 278. Furthermore, in vitro studies have shown that by changing Lys239 to a Gln residue within the glutamate-activating A-domain of the surfactin synthetase, SrfA, the specificity of this isolated A-domain is then changed from Glu to Gln.[5] In light of this, mutants of the CDA NRPS module 10 A-domain containing two single amino acid changes, Lys278Gln and Gln236Glu, as well as a double mutant comprising both mutations were constructed. It was envisaged that the complementary Gln236Glu mutation might further aid the recognition of glutamine or mGln substrates, with electrostatic interactions disfavoring interaction of glutamate or mGlu substrates.

Accordingly, a 1.8 kb DNA fragment spanning the A-domain of module 10 was generated and the desired mutations incorporated by site-directed mutagenesis. The three mutant DNA fragments were each transferred to pMAH,[13] which operates in *Streptomyces coelicolor* as a suicide vector, to generate three plasmids pKQ, pQE, and pKQQE. The wild-type strain *S. coelicolor* MT1110 was chosen as a host as it tends to produce non-phosphorylated CDAs which reduces the complexity of product analysis. In addition, the previously described mutant strain MT1110 Δ*glmT*[11] was used as a second host strain, because it lacks the *glmT* gene (*SCO3215*) that encodes a methyltransferase required for mGlu biosynthesis. MT1110 Δ*glmT* is unable to produce the

[*] Dr. J. Thirlway,[+] L. Nunns, Dr. M. Al Nakeeb, M. Styles, Dr. A.-W. Struck, Prof. J. Micklefield
School of Chemistry & Manchester Interdisciplinary Biocentre
The University of Manchester
131 Princess Street, Manchester M1 7DN (UK)
E-mail: j.micklefield@manchester.ac.uk
Dr. R. Lewis,[+] Prof. C. P. Smith
FHMS, University of Surrey
Guildford, Surrey, GU27XH (UK)
[+] These authors contributed equally to this work.

[**] This work was supported by the BBSRC (BB/C503662, BB/C507210, and PhD studentships to L.N. and M.S.). We also thank Dr. Barrie Wilkinson (Biotica) for his valuable advice and Prof. Gareth Morris's group (University of Manchester) for assistance with NMR experiments.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201202043.



C) Position according to GrsA numbering

Protein	236	239	278	299	301	322	330	331
CdaPS3 (Glu)[a]	Gln	Gly	Lys	Thr	Gly	Val	Gly	His
SrfA (Glu)	Ala	Lys	Asp	Leu	Gly	Val	Val	Asp
FenA (Glu)	Ala	Trp	His	Phe	Gly	Ser	Val	Glu
LicA (Gln)	Ala	Gln	Asp	Leu	Gly	Val	Val	Asp
TycC (Gln)	Ala	Trp	Gln	Phe	Gly	Leu	Ile	Asp

Figure 1. A) Structures of CDA lipopeptides. B) Organization of the CDA NRPS: C, condensation domain (red); adenylation domain with cognate substrate inset (blue); T, thiolation domain; E, epimerization domain (green); Te, thioesterase domain (orange). C) Alignment of the module 10 A-domain of CdaPS3 with selected Glu and Gln activating A-domains: gramicidin S synthetase A (GrsA); surfactin synthetase A (SrfA); fengycin synthetase A (FenA); lichenysin synthetase A (LicA); tyrocidine synthetase C (TycC). [a] Activates mGlu as well as Glu.

mGlu precursor, thereby circumventing any competition with the preferred natural substrate for the module 10 A-domain. The host strains were then transformed with the plasmids pKQ, pQE and pKQQE and using the standard double crossover methodology several independent mutant strains were generated for each of the desired mutations. Mutant strains derived from at least two independent transformations were obtained in all cases and these strains were grown in liquid culture and then analyzed by LC-MS for CDA production. This indicates that the Gln236Glu mutation has little effect on CDA production compared with the parental strains (Supporting Information, Figure S1). On the other hand the Lys278Gln mutation did have a significant effect. In the case of MT1110-KQ a new major product with a retention time of 6.7 min was evident, with a minor amount of CDA3a and CDA4a (retention time 7.0 min) that are both produced as the major products in the parental strain, MT1110 (Figure S2). The new product exhibited protonated, sodiated and potassiated singly charged ions which are consistent with a lipopeptide variant CDA3a-10Q (Figure 1 A) possessing a glutamine residue at position 10. The new product was subsequently purified by HPLC and subjected to high-resolution mass spectrometry, confirming the proposed molecular formula (Table 1).

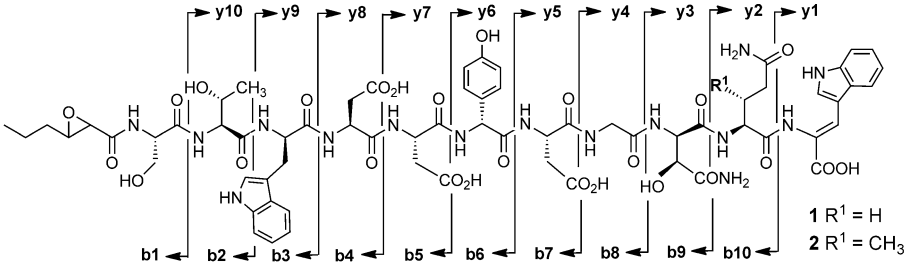
Table 1: HRMS of the new cylic lipopeptides.

CDAs	Ion	Formula	Observed[a]	Calculated
CDA3a-10Q	$[M+H]^+$	$C_{66}H_{78}N_{15}O_{25}^+$	1480.5309	1480.5288
	$[M+2H]^{2+}$	$C_{66}H_{79}N_{15}O_{25}^{2+}$	740.7686	740.7681
CDA4a-10mQ	$[M+H]^+$	$C_{67}H_{80}N_{15}O_{25}^+$	1494.5513	1494.5444
	$[M+2H]^{2+}$	$C_{67}H_{81}N_{15}O_{25}^{2+}$	747.7768	747.7759

[a] Observed masses are all within 5 ppm of calculated mass values.

During isolation of the new product CDA3a-10Q, a second product was observed which possesses a slightly shorter retention time and exhibits *m/z* 749.77 and 1498.53, which corresponds to the $[M+2H]^{2+}$ and $[M+H]^+$ ions of a linear form of CDA3a-10Q (**1**, Scheme 1) which is proposed to arise from hydrolysis of the lactone ring or failure of the thioesterase domain to effect complete cyclization. Similar byproducts have been identified during production of CDAs[11] and other related lipopeptides.[12c] The proposed linear variant **1**, was subjected to tandem MS yielding a product ion spectrum (Figure S4). From the products ions, complete sequence data was obtained (Table 2) confirming the presence of Gln at position 2 in place of Glu. The MT1110 Δ*glmT*-KQ mutant also produces CDA3a-10Q as the major product, with CDA3a as a minor component (Figure S3). Neither of the strains containing double point mutations, MT1110-KQQE or MT1110 Δ*glmT*-KQQE, produced any CDA lipopeptides as observed by LC-MS analysis. In light of this, the MT1110 Δ*glmT*-KQ strain was selected for further studies aimed at incorporating mGln into CDA.

Initially Boc-protected (2*S*,3*R*)-3-methyl glutamine was synthesised (Scheme S1). However, removal of the Boc group



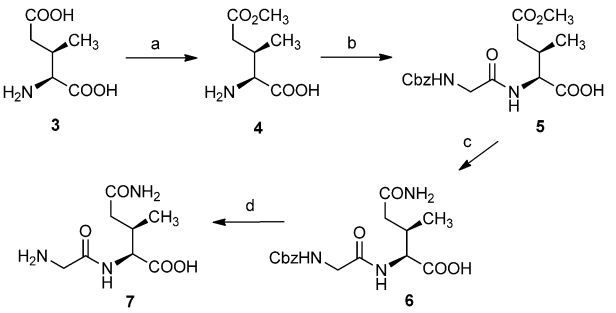
Scheme 1. Linear variants of CDA3a-10Q **1** and CDA4a-10mQ **2**.

Table 2: MS-MS product ion series observed for CDA3a-10Q linear peptide **1**.

y Ion[a]	Observed	Calcd	b Ion	Observed	Calcd
y10	1281.433	1281.444	b10	–	1296.466
y9	1180.415	1180.397	b9	1168.421	1168.407
y8	994.313	994.317	b8	1038.359	1038.369
y7	879.289	879.290	b7	981.361	981.348
y6	764.264	764.263	b6	866.308	866.321
y5	615.217	615.216	b5	717.268	717.273
y4	500.191	500.189	b4	602.243	602.246
y3	443.171	443.167	b3	487.223	487.219
y2[b]	313.130	313.130	b2	–	301.140
y1[b]	185.071	185.071	b1	–	200.092

[a] The masses of the y ion series are consistent with C-terminal acylium ions, which are typically generated upon fragmentation of peptides possessing N-terminal acyl groups. [b] Observed masses for y2 and y1 ions are consistent with a loss of Gln at this position.

proved problematic and any (2*S*,3*R*)-3-methyl glutamine (mGln) that may have been produced was subsequently transformed into 3-methyl glutamic acid during work-up and purification. Similar observations have been made previously during the synthesis of other glutamine analogues.[14] We suggest that any mGln initially produced on Boc-deprotection undergoes acid-catalyzed cyclization to generate a pyroglutamate (lactam) intermediate which is subsequently hydrolyzed to give 3-methyl glutamic acid. In light of this we chose to synthesize the Gly-mGln dipeptide **7** (Scheme 2), which we



Scheme 2. Synthesis of Gly-mGln: a) $SOCl_2$, MeOH; 30 min at −20 °C to 5 °C under N_2 (91%); b) Cbz-Gly-OSu, $NaHCO_3$ (10%) aqueous solution, 1,4-dioxane; 30 min at RT (78%); c) 28% aqueous NH_4OH; 18 h at RT (65%); d) H_2/10% Pd/C in MeOH.

anticipated would be more stable to side-chain amide hydrolysis, during fermentation, but would be susceptible to proteolysis inside the cell releasing mGln for incorporation by the CDA NRPS assembly line. The synthesis of the dipeptide **7** was achieved starting with mGlu **3**,[11] which was selectively esterified to give the monomethyl ester **4**. Coupling with Cbz-protected Gly gave dipeptide **5**. Ammoniolysis of the ester group of **5** provided Cbz-Gly-mGln **6** which was deprotected by hydrogenolysis. The Gly-mGln dipeptide **7**, which proved to be stable to hydrolysis, was then fed to MT1110 Δ*glmT*-KQ grown in liquid culture. LC-MS analysis of the culture supernatant revealed production of CDA3a-10Q as the major product, with an additional new product that exhibited doubly charged and singly charged molecular ions in MS that are consistent with a CDA derivative possessing a mGln residue CDA4a-10mQ (Figure S5). This product was isolated by HPLC and subjected to HRMS, which revealed doubly charged and singly charged molecular ions that were consistent with the proposed molecular formula of CDA4a-10mQ (Table 1). The linear forms of CDA3a-10Q **1** and CDA4a-10mQ **2** were also evident in the culture supernatant (Figure S6). Tandem MS analysis of CDA4a-10mQ (Figure 1) resulted in a product ion spectrum (Figure S7 and Table 3)

Table 3: MS-MS product ion series observed for CDA4a-10mQ.[a]

y Ion	Observed	Calcd	b Ion	Observed	Calcd
y10	1295.250	1295.460	b10	–	1310.481
y9	–	1194.412	b9	1168.422	1168.407
y8	1008.233	1008.333	b8	1038.495	1038.369
y7	893.238	893.306	b7	981.211	981.348
y6	778.254	778.279	b6	866.319	866.321
y5	629.214	629.231	b5	717.255	717.273
y4	514.193	514.204	b4	602.247	602.246
y3	–	457.183	b3	487.212	487.219
y2[b]	327.146	327.145	b2	301.099	301.140
y1[b]	185.069	185.071	b1	200.084	200.092

[a] The product ions series is derived from the doubly charged cyclic parent ion $[M+2H]^{2+}$. The y ion series masses are consistent with C-terminal acylium ions. [b] Observed masses for y2 and y1 ions are consistent with the presence of mGln at position 10.

that clearly indicates a peptide sequence with a mGln residue at position 10, consistent with the proposed structure of CDA4a-10mQ. The major product CDA3a-10Q was also purified and subjected to additional characterization by ^{1}H NMR spectroscopy. TOCSY and NOESY experiments, and comparison of NMR data with the parent lipopeptide CDA3a, provide further evidence to support the structural assignment obtained using tandem MS (Tables S2 and S3 and Figures S10–S17).

In summary, we have shown that introduction of a single Lys278Gln mutation into the module 10 A-domain of the CDA NRPS results in a novel glutamine-containing CDA variant. Moreover, feeding the dipeptide Gly-mGln to a Lys278Gln mutant lacking the natural mGlu precursor (MT1110 Δ*glmT*-KQ) also leads to the incorporation of synthetic mGln into CDA. The fact that the mGln-containing CDA4a-10mQ was produced at lower levels than the natural CDA4a from the parental strain may be due to the facile hydrolysis of the amide side chain of the mGln precursor released upon proteolysis. Nevertheless, (2*S*,3*R*)-configured mGln has not been identified in nature to date and consequently the work presented here provides the first example of how active-site modification of an adenylation domain within an NRPS can be used to introduce a synthetic non-natural amino acid into a nonribosomal peptide natural product. The combination of the mutasynthesis strategy described here and directed evolution methods for selection of A-domains with altered specificity[6] may prove to be a powerful approach to expand the structural diversity of nonribosomal peptides which are too complex for effective

synthetic modification. Ultimately such engineering approaches might be useful for fine tuning the physicochemical properties and biological activities of complex nonribosomal peptides, which are urgently required for pharmaceutical, agrochemical, and other applications.

Received: March 14, 2012
Revised: May 18, 2012
Published online: June 18, 2012

Keywords: adenylation domain · antibiotics · biosynthetic engineering · natural products · nonribosomal peptides

[1] a) S. A. Sieber, M. A. Marahiel, *Chem. Rev.* **2005**, *105*, 715 – 738; b) C. T. Walsh, *Science* **2004**, *303*, 1805 – 1810; c) B. Wilkinson, J. Micklefield, *Nat. Chem. Biol.* **2007**, *3*, 379 – 386; d) M.-C. Wu, B. Law, B. Wilkinson, J. Micklefield, *Curr. Opin. Biotechnol.* **2012**, DOI: 10.1016/j.copbio.2012.03.008.

[2] a) J. H. Lee, B. S. Evans, G. Li, N. L. Kelleher, W. A. van der Donk, *Biochemistry* **2009**, *48*, 5054 – 5056; b) D. Schwartz, N. Grammel, E. Heinzelmann, U. Keller, W. Wohlleben, *Antimicrob. Agents Chemother.* **2005**, *49*, 4598 – 4607; c) M. Ongena, P. Jacques, *Trends Microbiol.* **2008**, *16*, 115 – 125.

[3] K. T. Nguyen, D. Ritz, J.-Q. Gu, D. Alexander, M. Chu, V. Miao, P. Brian, R. H. Baltz, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17462 – 17467; S. Doekel, M.-F. Coëffet-Le Gal, J.-Q. Gu, M. Chu, R. H. Baltz, P. Brian, *Microbiology* **2008**, *154*, 2872 – 2880.

[4] G. C. Uguru, C. Milne, M. Borg, F. Flett, C. P. Smith, J. Micklefield, *J. Am. Chem. Soc.* **2004**, *126*, 5032 – 5033.

[5] K. Eppelmann, T. Stachelhaus, M. A. Marahiel, *Biochemistry* **2002**, *41*, 9718 – 9726.

[6] B. S. Evans, Y. Chen, W. W. Metcalf, H. Zhao, N. L. Kelleher, *Chem. Biol.* **2011**, *18*, 601 – 607.

[7] E. Conti, T. Stachelhaus, M. A. Marahiel. P. Brick, *EMBO J.* **1997**, *16*, 4174 – 4183.

[8] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493 – 505; G. L. Challis, J. Ravel, C. A. Townsend, *Chem. Biol.* **2000**, *7*, 211 – 224; C. Rausch, T. Weber, O. Kohlbacher, W. Wohlleben, D. H. Huson, *Nucleic Acids Res.* **2005**, *33*, 5799 – 5808.

[9] a) C.-Y. Chen, I. Georgiev, A. C. Anderson, B. R. Donald, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3764 – 3769; b) B. Villiers, F. Hollfelder, *Chem. Biol.* **2011**, *18*, 1290 – 1299.

[10] a) Z. Hojati, C. Milne, B. Harvey, L. Gordon, M. Borg, F. Flett, B. Wilkinson, C. P. Smith, J. Micklefield, *Chem. Biol.* **2002**, *9*, 1175 – 1187; b) A. Powell, M. Borg, B. Amir-Heidari, J. M. Neary, J. Thirlway, B. Wilkinson, C. P. Smith, J. Micklefield, *J. Am. Chem. Soc.* **2007**, *129*, 15182 – 15191; c) C. Mahlert, F. Kopp, J. Thirlway, J. Micklefield, M. A. Marahiel, *J. Am. Chem. Soc.* **2007**, *129*, 12011 – 12018; d) A. Powell, M. Al Nakeeb, B. Wilkinson, J. Micklefield, *Chem. Commun.* **2007**, 2683 – 2685; e) B. Amir-Heidari, J. Thirlway, J. Micklefield, *Org. Biomol. Chem.* **2008**, *6*, 975 – 978; f) B. Amir-Heidari, J. Thirlway, J. Micklefield, *Org. Lett.* **2007**, *9*, 1513 – 1516; g) J. Neary, A. Powell, L. Gordon, C. Milne, F. Flett, B. Wilkinson, C. P. Smith, J. Micklefield, *Microbiology* **2007**, *153*, 768 – 776; h) B. Amir-Heidari, J. Micklefield, *J. Org. Chem.* **2007**, *72*, 8950 – 8953.

[11] C. Milne, A. Powell, M. Al Nakeeb, C. P. Smith, J. Micklefield, *J. Am. Chem. Soc.* **2006**, *128*, 11250 – 11259.

[12] a) R. H. Baltz, *Curr. Opin. Chem. Biol.* **2009**, *13*, 144 – 151; b) L. Robbel, M. A. Marahiel, *J. Biol. Chem.* **2010**, *285*, 27501 – 27508; c) R. H. Baltz, V. Miao, S. K. Wrigley, *Nat. Prod. Rep.* **2005**, *22*, 717 – 741; d) J. Micklefield, *Chem. Biol.* **2004**, *11*, 887 – 888.

[13] G. Bucca, A. M. E. Brassington, G. Hotchkiss, V. Mersinias, C. P. Smith, *Mol. Microbiol.* **2003**, *50*, 153 – 166.

[14] a) R. Dave, B. Badet, P. Meffre, *Amino Acids* **2003**, *24*, 245 – 261; b) P. Meffre, R. H. Dave, J. Leroy, B. Badet, *Tetrahedron Lett.* **2002**, *43*, 6279.

