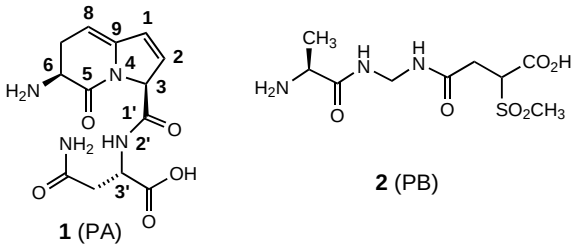
Antibiotic Discovery

The Biosynthetic Gene Cluster of Pantocin A Provides Insights into Biosynthesis and a Tool for Screening**

*Mi Jin, Sandra A. I. Wright, Steven V. Beer, and Jon Clardy**



The preceding communication described the isolation and structural characterization of pantocin A (PA, **1**), a member of the diverse family of antibiotics produced by *Pantoea agglomerans*, and its ability to block histidine biosynthesis by inhibiting L-histidinol phosphate aminotransferase. The unusual structure of pantocin A, especially its labile bicyclic core, suggested a previously uncharacterized biosynthetic pathway. Pantocin A (**1**) was discovered through the heterologous expression of a cosmid library, derived from the genomic DNA of *P. agglomerans* strain Eh318, in *Escherichia coli* XL2Blue.[1] This DNA-based approach to the discovery of biologically active small molecules, which facilitated the characterization of the Eh318 antibiotics pantocin A (**1**) and B (**2**),[2] also expedites the identification of the DNA involved in their production. Subcloning of the original cosmid clone pCPP702 resulted in a 3.5 kb DNA fragment that enables *E. coli* XL2Blue-pUC449 to produce **1**. Herein, we report a preliminary analysis of this 3.5-kb sequence that identifies the peptide precursor of **1**, two proteins responsible for its construction, and the resistance (self-immunity) mechanism. This analysis also provides the information needed to identify other strains producing antibiotics related to **1**.

Sequencing of the 3.5 kb insert revealed three major open reading frames (ORFs), which were named *paaA*, *paaB*, and *paaC* (Figure 1a).[3] Transposon mutagenesis indicated that *paaA* and *paaB* are required for the production of **1** and that disruption of *paaC* leads to reduced production of **1** and poor growth (Figure 1a).

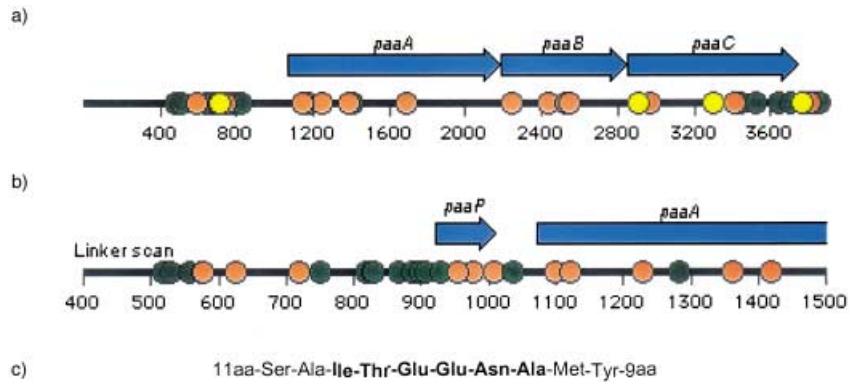


Figure 1. a) Transposon mutagenesis of pUC449. Circles indicate the position of transposon insertion and are color coded for anti-Ea273 activity: inactive: red, active: green, reduced activity: yellow. b) Linker-scan of pUC449 (region 449–1500). Circles indicate the position of a 15-bp linker insertion. c) Pantocin A precursor peptide. Bold symbols represent residues involved in the biosynthesis of **1** and molecules similar to **1**.

The structure of **1**, especially its asparagine side chain and the functionality and absolute configurations at C3 and C6, suggests a peptide origin. Feeding experiments with ^{14}C-labeled Glu and Gln were carried out with *E. coli* XL2Blue-pUC449 in minimal media to determine whether the N4–C9 fragment could arise from a cyclized glutamate or glutamine residue. Feeding with either ^{14}C-labeled Glu or Gln led to the production of ^{14}C-labeled **1** while control feeding with ^{14}C-labeled Leu did not. Uniformly ^{13}C-labeled Glu was fed to determine the labeling pattern of **1**, and ^{13}C NMR analysis revealed that all the carbon atoms in the bicyclic core (C1–C3 and C5–C9) were labeled to a similar extent, while significantly less incorporation was seen in the asparagine carbon atoms. Since Glu can serve as a general carbon source through its conversion into 2-oxoglutarate in the tricarboxylic acid cycle, the observed incorporation pattern likely reflects the partial conversion of Glu into Asn. Thus, **1** originates from two Glu/Gln residues and one Asp/Asn residue.

Two lines of evidence indicate that modification of a ribosomally assembled peptide[4,5] with the general structure -Glu/Gln-Glu/Gln-Asp/Asn- gives rise to **1**. First, the 3.5 kb

[*] Prof. J. Clardy,[++] M. Jin
Department of Chemistry and Chemical Biology
Cornell University
Ithaca, NY 14853-1301 (USA)
E-mail: jon_clardy@hms.harvard.edu
Prof. S. A. I. Wright,[+] Prof. S. V. Beer
Department of Plant Pathology
Cornell University
Ithaca, NY 14853 (USA)
[+] Current address:
Plant Pathology and Biocontrol Unit
SLU, 750 07 Uppsala (Sweden)
[++] Current address:
Department of Biological Chemistry and Molecular Pharmacology
Harvard Medical School
Boston, MA (USA)
Fax: (+1) 617-432-3702

[**] This work was supported by the National Institutes of Health (National Cancer Institute, R01-CA24487).

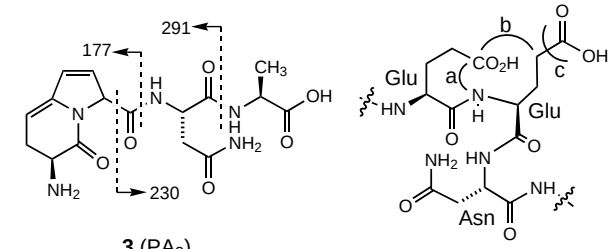
Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

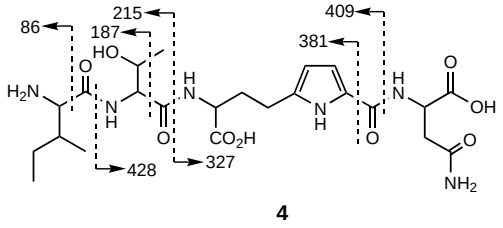
  DOI: 10.1002/ange.200351054

insert is too small to encompass the well-known nonribosomal peptide synthesis pathway with its multiple modules.[6] Second, cloning *paaA*, *paaB*, and *paaC* into a pET11a vector and expressing it under the control of a T7 promoter in *E. coli* strain BL21(DE3) did not produce a single clone with PA activity, which indicates that *paaABC* by itself is not sufficient for the biosynthesis of **1**.

We looked for the gene encoding the prepeptide upstream of *paaA* using a transposon-linker scan experiment with a 15-bp oligonucleotide insertion to saturate the 0.6 kb region upstream of the predicted start codon of *paaA* (Figure 1b). This approach revealed an additional small ORF, *paaP*, which is required for the production of **1**. The first indication for a *paaP* requirement came from a frame shift mutant of *paaP* that lost activity. In addition, four other linker insertions upstream of *paaP* abolished the activity of **1**, and these insertions are all in the vicinity of a predicted promoter region[7] (Figure 1b). Evidence for the gain of function was provided by cloning *paaPABC* into expression vector pET11b to transform *E. coli* strain BL21(DE3) gold and observing activity of **1** on Ea273-GA agar assay plates. *PaaP* encodes a 30-residue peptide with the sequence -Glu-Glu-Asn- near its midpoint (Figure 1c).

Chemical evidence coupled with site-directed mutagenesis also implicated *paaP* in the biosynthesis of **1**. Large-scale fermentations of clones producing **1** uncovered a minor metabolite, pantocin A_2 (**3**, PA_2), which is produced along with **1**. PA_2 (**3**) has the same UV spectrum as **1**, the same level





of antibiotic activity against *E. amylovora* test strain Ea273, and a different HPLC retention time. Combined ESI-MS/MS and NMR analyses revealed that **3** is in fact **1** modified by the addition of a C-terminal alanine group. An alanine in this position is encoded by *paaP* (Figure 1c).

In an attempt to produce modified PA-type molecules by altering *paaP*, site-directed mutagenesis of Ala to Val was used to produce clone *E. coli*-pUC449AtV. This clone has reduced antibiotic activity in its culture supernatant. Characterization of all the compounds produced by this mutant clone proved difficult because of low production levels and lability problems, but a stable inactive compound (**4**) was isolated and characterized by NMR spectroscopy and ESI-MS/MS. Compound **4** is the hydrolysis product of the core of **1** bearing additional Ile and Thr residues at the N-terminus, and these N-terminal residues are also predicted by *paaP*. These combined experiments implicate six consecutive residues in the center of the *paaP* gene product with production of **1**. A predicted stem-loop structure follows *paaP*, and the production of the *paaP* gene product might not be tightly coupled to the production of the *paaA–C* gene products.

PaaC is the resistance gene. Transposon insertions in *paaC* do not completely abolish production of **1**,[3] but clones with such insertions grow poorly. Cloning *paaC* into a PA-susceptible *E. coli* strain confers resistance to **1** (Figure 2). *PaaC* is predicted to be a transmembrane protein with 46% sequence similarity to a known *E. coli* efflux pump[9] (accession no. BAB35563) for intermediates on the cysteine biosynthetic pathway. Thus, resistance is based on the transport of **1** out of the producing cell.

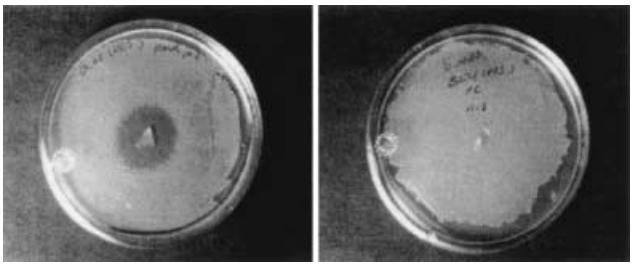


Figure 2. Left: *E. coli* BL21(DE3)-pMR102.[8] Right: *E. coli* BL21(DE3)-pMR102*PaaC* showing resistance to **1**.

The roles of the *paaA* and *paaB* gene products are not well-defined, but they must carry out the series of transformations schematically indicated in **5**: a) formation of the C9–N4 bond, b) formation of the C9–C1 bond, and c) oxidative cleavage of the carboxylate bond at C1. The predicted *paaA* gene product, PaaA, has significant sequence similarity to the bacterial ThiF/MoeB family of proteins (37 to 43% identity). ThiF[10] and MoeB[11,12] both adenylate carboxyl groups to activate them for an addition–elimination sequence, and PaaA has the conserved ATP-binding site found in ThiF and MoeB. Thus PaaA is the likely candidate to perform the cyclization shown in **5** (a). BLAST searches on PaaB have shown only low sequence similarities to proteins of unknown function. A Pfam search identified a domain match ($E = 0.025$) in PaaB to the 2OG-Fe^{II} oxygenase superfamily (pfam03171). Since the formation of **1** from the precursor peptide requires an oxidation, this similarity to Pfam might be significant.

The instructions for conferring the ability to both synthesize and resist the remarkable antibiotic **1** in an *E. coli* host are encoded on a 3.5 kb stretch of DNA. The 30 residue peptide precursor of **1** is the product of gene *paaP*; this peptide is modified by PaaA, PaaB, and possibly by unidentified host proteases, which convert the precursor peptide into **1**. The producing organism avoids the toxic effects of **1** by pumping it into the extracellular environment with PaaC where it presumably confers a selective advantage to the producing organism.[13]

This preliminary understanding of the biosynthesis of **1** also provided a tool for establishing the occurrence of the same or related pathways in other *P. agglomerans* strains. Since the *paaB* gene has the least similarity with other reported genes, it was used to generate a digoxigenin-labeled probe for southern blotting of 88 antibiotic-producing *P. agglomerans* strains.[13] A total of 61 strains hybridized with *paaB*, and these strains were the ones most effective in suppressing *E. amylovora* in tests on immature pears and in apple orchards. Like Eh318, most of the strains hybridizing with *paaB* have histidine-suppressible antibiotic activity, but a few produced antibiotics that were not histidine-suppressible.[13] These strains are likely to produce small molecule antibiotics that differ from **1**. Studies on the structure and mechanism of PaaA and PaaB are underway, as are attempts to characterize new antibiotics related to **1**.

Experimental Section

Transposon mutagenesis: Transposon mutagenesis was performed on pUC449 plasmid DNA using the GPS-1 genome priming system following the manufacturer's protocol (New England Biolabs, Inc.). Either pGPS1.1 or pGPS5 were used as the transprimer donor plasmid, and *E. coli* XL2-blue or DH5α cells were used for transformation. Mutants were selected on Luria–Bertani plates supplemented with 30 μg mL^{-1} kanamycin. Anti-Ea273 activity was tested using a soft agar overlay assay.[3] The positions and orientations of transposon insertions were identified by sequencing. The gene sequence of **1** has GenBank accession no. U81376.

Linker scanning experiment: The GPS-LS linker scanning system was used according to the manufacturer's protocol (New England BioLabs, Inc.). pGPS5 transposon mutants were pooled for plasmid extraction. Plasmid DNA was digested with restriction enzymes XbaI and XcmI. A 2.7-kb DNA fragment was purified on gel from the digestion mixture and ligated into a 5.1-kb XbaI-XcmI fragment from plasmid pUC449. The ligated plasmid DNA was amplified once by transformation into *E. coli* and plasmid extraction on pooled colonies. The derived plasmid DNA was digested with PmeI, religated, and used to transform *E. coli* XL1-blue cells. The resulting mutant clones were assayed for anti-Ea273 activity and sequenced to identify positions of linker insertion.

Cloning of *E. coli*-pUC449AtV: Plasmid pUC449 was amplified using the following primer pair to generate an Ala to Val point mutation: 5′-CTGCTATCACTGAAGAAAATG**T**TATGTATACCAAAGGTCAGG-3′ and 5′-CCTGACCTTTGGTATACATA**A**CATTTTCTTCAGTGATAGCAG-3′.The PCR product was digested with the restriction enzyme DpnI, and the digested product was used to transform *E. coli* strain XL1-Blue to give *E. coli*-pUC449AtV. The mutant plasmid was sequenced to confirm the site of mutation.

3: Compound **3** was isolated from *E. coli*-pUC449 minimal medium culture, by using the same isolation procedure as that for **1**, in a yield of around 0.2 mg L^{-1}. HRFAB-MS: *m*/*z*: 418.1130 [*M*+K]$^+$, calcd *m*/*z*: 418.1129 for $C_{16}H_{21}N_5O_6K$. ^{1}H NMR (D_2O, 500 MHz, 25 °C): δ = 6.44 (1H, dd, *J* = 6.4, 2.0 Hz, H1), 6.22 (1H, brd, *J* = 6.4 Hz, H2), 5.34 (1H, m, H-8), 5.33 (1H, brs, H3), 4.71 (1H, dd, *J* = 8.5, 5.0 Hz, H3′), 4.25 (1H, dd, *J* = 14.5, 8.0 Hz, H6), 4.08 (1H, q, *J* = 7.3 Hz, H8′), 2.90–2.76 (2H, m, H7), 2.84 (1H, dd, *J* = 15.5, 5.0 Hz, H4′a), 2.72 (1H, dd, *J* = 15.5, 8.5 Hz, H4′b), 1.29 ppm (3H, d, *J* = 7.3 Hz, H9′); ^{13}C NMR (D_2O + 5% CD_3OD, 100 MHz, 25 °C): δ = 180.6 (C10′ COOH), 175.6 (C6′ CO), 171.8 (C5′, $CONH_2$), 170.6 (C5), 165.8 (C1′), 143.7 (C9), 131.0 (C2), 128.4 (C1), 97.6 (C8), 68.0 (C3), 52.2 (C8′), 51.8 (C3′), 50.2 (C6), 37.3 (C4′), 26.3 (C7), 18.6 ppm (C9′); IR (KBr): $\tilde{\nu}$ = 3700–2400, 1676, 1588, 1404 cm^{-1}; UV (H_2O): λ_{max} (ε) = 201 (9547), 273 (4039); $[\alpha]_D^{25} = -202°$, (*c* = 0.5, H_2O); ESI-MS/MS: *m*/*z*: 380 [*M*+H]$^+$, 291 [*M*−Ala]$^+$, 263, 230, 184, 177 [*M*−Ala−Asn]$^+$, 151. To determine the absolute configuration of the C-terminal alanine, 1 mg of **3** was hydrolyzed in 6 M HCl at 90 °C for 20 h, then dried. The resulting compounds were resolved on a CHIRALPLATE (Macherey–Nagel) using the solvent mixture acetone:H_2O:CH_3OH (10:2:2) and compared with authentic samples of L- and D-Ala treated in the same manner. Ala derived from **3** migrated the same as L-Ala (R_f = 0.57) and was readily distinguished from control D-Ala (R_f = 0.52).

4: ^{1}H NMR (D_2O, 500 MHz, 25 °C): δ = 6.78 (1H, d, *J* = 3.5 Hz, H3), 6.03 (1H, d, *J* = 3.5 Hz, H4), 4.64 (1H, dd, *J* = 8.5, 5.0 Hz, H3′), 4.35 (1H, d, *J* = 7.0 Hz, Thr αH), 4.19 (1H, dd, *J* = 8.0, 4.0 Hz, H3″), 4.10 (1H, m, Thr βH), 3.94(1H, d, *J* = 6.0 Hz, Ile αH), 2.83 (1H, dd, *J* = 15.0, 5.0 Hz, H4′a), 2.70 (1H, dd, *J* = 15.0, 8.5 Hz, H4′b), 2.67(2H, t, *J* = 7.5 Hz, H1″), 2.11–1.90 (2H, m, H2″), 1.42 (1H, m, Ile βH), 1.23 (3H, d, *J* = 6.0 Hz, Thr γH), 1.20–1.14 (2H, m, Ile γCH_2), 0.94 (3H, d, *J* = 7.0 Hz, Ile γCH_3), 0.84 ppm (3H, t, *J* = 7.0 Hz, Ile δCH_3); ^{13}C NMR (D_2O, externally referenced with 5% CD_3OD in D_2O, 100 MHz, 25 °C): δ = 179.1 (C4″, COOH), 178.6 (Asn COOH), 177.1 (Asn $CONH_2$), 171.5 and 170.6 (Ile and Thr CO) 163.4 (C1′), 138.5 (C5), 124.7 (C2), 113.4 (C3), 108.6 (C4), 68.1 (Thr βCH), 60.7 (Thr αCH), 58.8 (Ile αCH), 55.7 (C3″), 53.0 (Asn αCH), 39.2 (Asn βCH_2), 37.6 (Ile βCH), 32.4 (C2″), 25.1 (Ile γCH_2), 24.3 (C1″), 20.0 (Thr γCH_3), 15.2 (Ile γCH_3), 11.6 ppm (Ile δCH_3); ESI-MS/MS: *m*/*z*: 541 [*M*+H]$^+$, 524, 428 [*M*−N-terminal Ile]$^+$, 409 [*M*−C-terminal Asn]$^+$, 391, 381, 327, 296, 278, 252, 234, 215 [*M*−N-terminal Ile-Thr]$^+$, 209, 195, 187, 169, 86. Compound **4** was isolated from *E. coli*-pUC449AtV minimal media culture in a yield of 0.1 mg L^{-1}.

Southern blot experiments: A nylon membrane (Pall Biodyne B) from Nalge Nunc International and a DIG-high prime DNA labeling and detection starter kit I from Roche Applied Bioscience were used. All test strains were inoculated on a nylon membrane placed on top of an LB agar plate. Eh318 and Ea273 were included as positive and negative controls, respectively. Colonies were grown to at least 2 mm in diameter, then the cells were lysed, and the DNA denatured using 0.4 M NaOH solution. After neutralization with tris(hydroxymethyl)-aminomethane-HCl (tris-HCl) buffer (0.5 M, pH 7.0) and washing (2 × saline sodium citrate solution), the DNA was cross-linked with UV light to the membrane. The *paaB* gene (ca. 800 bp) was amplified by a polymerase chain reaction using primer pair 5′-ATGAGCATTGATAGTATTACCTTTAACGACAAAG-3′ and 5′-CATGAATCTAGAGAACCCGCCATTC-3′, and purified on gel. The amplified *paaB* fragment (1 μg, 16 μL) was denatured and used to generate the DIG-labeled probes at 37 °C for 16 h. Southern blotting was performed according to the manufacturer's protocol. Hybridization was performed using 2.6 μL of denatured probes in 8.5 mL of solution at 45 °C for 16 h. Stringency washes (2x15 m) were performed at 65 °C using 0.5 × saline sodium citrate.

Received: January 29, 2003 [Z51054]

Keywords: amino acids · antibiotics · isotopic labeling

[1] S. A. I. Wright, C. H. Zumoff, L. Schneider, S. V. Beer, *Appl. Environ. Microbiol.* **2001**, *67*, 284.
[2] S. F. Brady, S. A. Wright, J. C. Lee, A. E. Sutton, C. H. Zumoff, R. S. Wodzinski, S. V. Beer, J. Clardy, *J. Am. Chem. Soc.* **1999**, *121*, 11912.
[3] S. A. Wright, Ph.D. thesis, Cornell University (Ithaca), **1997**.
[4] R. Kolter, F. Moreno, *Annu. Rev. Microbiol.* **1992**, *46*, 141.
[5] J. E. Gonzalez-Pastor, J. L. San Millan, F. Moreno, *Nature* **1994**, *369*, 281.
[6] M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem. Rev.* **1997**, *97*, 2651.
[7] BDGP neural network promoter prediction (**1998** version). M. G. Reese, N. L. Harris, F. H. Eeckman, *Biocomputing:*

Proceedings of the 1996 Pacific Symposium (Ed.: L. Hunter, T. E. Klein), World Scientific Publishing Co, Singapore, **1996**.
[8] M. Munson, P. F. Predki, L. Regan, *Gene* **1994**, *144*, 59.
[9] T. Dassler, T. Maier, C. Winterhalter, A. Bock, *Mol. Microbiol.* **2000**, *36*, 1101.
[10] S. V. Taylor, N. L. Kelleher, C. Kinsland, H. J. Chiu, C. A. Costello, A. D. Backstrom, F. W. McLafferty, T. P. Begley, *J. Biol. Chem.* **1998**, *273*, 16555.
[11] M. W. Lake, M. M. Weubbens, K. V. Rajagopalan, H. Schindelin, *Nature* **2001**, *414*, 325.
[12] S. Leimkuhler, M. M. Wuebbens, K. V. Rajagopalan, *J. Biol. Chem.* **2001**, *276*, 34695.
[13] M. Jin, Ph.D. thesis, Cornell University (Ithaca), **2003**.

