Studies on the biosynthesis and heterologous expression of complex secondary metabolites from streptomycetes

Dissertation
zur Erlangung des Grades des
Doktors der Naturwissenschaften (Dr.rer.nat.)
der Naturwissenschaftlich-Technischen Fakultät III
Chemie, Pharmazie, Bio- und Werkstoffwissenschaften
der Universität des Saarlandes

von

Tina Maria Binz
Saarbrücken
Februar 2009
Publications


Conference contributions

Binz, T.M., Wenzel, S.C., Mahmud, T., Müller, R., October 2005. Comparison of the myxochromide A and S biosynthetic gene cluster from *Myxococcus xanthus* and *Stigmatella aurantiaca*; International workshop: Biology of Bacteria Producing Natural Products, Dresden, Germany (poster presentation)


Binz, T.M., Wenzel, S.C., Bechthold, A., Müller, R., October 2006. Heterologous expression of the Phenalinolactone biosynthetic gene cluster in different *Streptomyces* sp. via Red/ET recombineering; International workshop: Biology of Bacteria Producing Natural Products, Tübingen, Germany (oral presentation)

Binz, T.M., Wenzel, S.C., Bechthold, A., Müller, R., October 2007. Heterologous expression of the Phenalinolactone biosynthetic gene cluster in different *Streptomyces* sp. via Red/ET recombineering; European Bioperpectives, Cologne, Germany (poster presentation)

Acknowledgements

First of all, I would like to express my sincere gratitude to my doctoral advisor Prof. Dr Rolf Müller, for his guidance, advice, and encouragement during the course of my Ph.D. thesis. His constant confidence and persistence in face of research uncertainties motivated me to continue my work and inspired me much.

At the same time I would like to thank my co-advisor Dr. Silke Wenzel for the very pleasant and successful team work, her support, advice, and motivation the entire time.

I would also like to thank Prof. Dr. Manfred Schmitt for gladly consenting to review my thesis.

I kindly thank my cooperation partners Margherita Sosio and Stefano Donadio from KtedoGen and Hans-Jörg Schnell and Andreas Bechthold from Freiburg University for their assistance and the successful cooperation.

I greatly thank Dr. Kira Weissman not only for the excellent proof-reading of the thesis but also for her skillful remarks and helpful discussions which inspired me much.

I especially wish to thank Daniel Krug for his unremitting help and assistance with the analytics and for his extremely valuable answers to all my questions.

Of course (!) I also thank all members from our research institute Pharmazeutische Biotechnologie and especially all members of my office. I am grateful for all their encouragements, futile discussions, and jokes; I really enjoyed the friendly work atmosphere!

Last but not least my heartfelt thanks to my family and my friends for always believing in me and their joyful and constant support.
Table of contents

TABLE OF CONTENTS 1
FIGURES VI
TABLES IX
ABBREVIATIONS XI

ZUSAMMENFASSUNG 1
ABSTRACT 2

1 INTRODUCTION 3
  1.1 Natural products as a robust source of new drugs and drug leads 3
  1.2 Streptomyces as important natural product producers 5
  1.3 Heterologous expression of natural product pathways 8
  1.4 The biosynthetic logic of nonribosomal peptide synthetases 13
  1.5 Outline of the dissertation 18

2 MATERIAL AND METHODS 20
  2.1 Chemicals 20
  2.2 Enzymes, kits and markers 21
  2.3 Buffers and stock solutions 22
    2.3.1 Solutions and buffers for molecular biology applications 22
    2.3.2 Solutions and buffers for Southern Blot and colony hybridization 23
    2.3.3 Solutions and buffers for enzymatic applications 24
  2.4 Media 26
    2.4.1 Media for E. coli cultivation 26
    2.4.2 Media for cultivation of pseudomonads 26
    2.4.3 Media for cultivation of streptomycetes 27
  2.5 Antibiotics 28
  2.6 Instruments and materials 29
  2.7 Bacterial strains, oligonucleotides and plasmids 31
    2.7.1 Bacterial strains 31
    2.7.2 Oligonucleotides 32
      2.7.2.1 Primers used in the phenalinolactone project 32
      2.7.2.2 Primers used in the GE81112 project 33
    2.7.3 Plasmids 34
      2.7.3.1 General plasmids 34
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.7.3.2</td>
<td>Plasmids used in the phenalinolactone project</td>
<td>35</td>
</tr>
<tr>
<td>2.7.3.3</td>
<td>Plasmids used in the GE81112 project</td>
<td>36</td>
</tr>
<tr>
<td>2.8</td>
<td>STREPTOMYCES</td>
<td>37</td>
</tr>
<tr>
<td>2.8.1</td>
<td>Mutants of S. lividans TK24</td>
<td>37</td>
</tr>
<tr>
<td>2.8.2</td>
<td>Mutants of S. coelicolor A3(2)</td>
<td>37</td>
</tr>
<tr>
<td>2.8.3</td>
<td>Mutants of S. coelicolor M512</td>
<td>38</td>
</tr>
<tr>
<td>2.8.4</td>
<td>Mutants of S.14386</td>
<td>38</td>
</tr>
<tr>
<td>2.9</td>
<td>CULTIVATION AND CONSERVATION OF STRAINS</td>
<td>38</td>
</tr>
<tr>
<td>2.9.1</td>
<td>E. coli/Pseudomonas putida</td>
<td>38</td>
</tr>
<tr>
<td>2.9.2</td>
<td>Streptomycetes</td>
<td>38</td>
</tr>
<tr>
<td>2.9.2.1</td>
<td>Spores</td>
<td>38</td>
</tr>
<tr>
<td>2.9.2.2</td>
<td>Mycelia</td>
<td>39</td>
</tr>
<tr>
<td>2.10</td>
<td>ISOLATION OF PROKARYOTIC DNA</td>
<td>39</td>
</tr>
<tr>
<td>2.10.1</td>
<td>Isolation of genomic DNA</td>
<td>39</td>
</tr>
<tr>
<td>2.10.2</td>
<td>Isolation of plasmid and cosmid DNA by alkaline lysis</td>
<td>39</td>
</tr>
<tr>
<td>2.10.3</td>
<td>Isolation of plasmid DNA with the NucleoSpin® Plasmid Kit</td>
<td>40</td>
</tr>
<tr>
<td>2.11</td>
<td>SEPARATION AND PURIFICATION OF DNA</td>
<td>40</td>
</tr>
<tr>
<td>2.11.1</td>
<td>Extraction of DNA from agarose gels</td>
<td>40</td>
</tr>
<tr>
<td>2.11.2</td>
<td>DNA precipitation</td>
<td>40</td>
</tr>
<tr>
<td>2.11.3</td>
<td>Phenol/chloroform extraction</td>
<td>40</td>
</tr>
<tr>
<td>2.12</td>
<td>POLYMERASE CHAIN REACTION</td>
<td>41</td>
</tr>
<tr>
<td>2.12.1</td>
<td>Composition of standard PCR setup</td>
<td>41</td>
</tr>
<tr>
<td>2.12.2</td>
<td>Colony PCR</td>
<td>42</td>
</tr>
<tr>
<td>2.13</td>
<td>ENZYMATIC MANIPULATION OF DNA</td>
<td>42</td>
</tr>
<tr>
<td>2.13.1</td>
<td>Restriction endonucleases</td>
<td>42</td>
</tr>
<tr>
<td>2.13.2</td>
<td>Ligation</td>
<td>42</td>
</tr>
<tr>
<td>2.13.3</td>
<td>Dephosphorylation</td>
<td>42</td>
</tr>
<tr>
<td>2.13.4</td>
<td>Cloning with TOPO-TA-Cloning kit</td>
<td>42</td>
</tr>
<tr>
<td>2.14</td>
<td>RED/ET RECOMBINATION</td>
<td>43</td>
</tr>
<tr>
<td>2.15</td>
<td>TRANSFORMATION OF BACTERIA</td>
<td>43</td>
</tr>
<tr>
<td>2.15.1</td>
<td>Electroporation of E. coli</td>
<td>43</td>
</tr>
<tr>
<td>2.15.2</td>
<td>Conjugation of streptomycetes with E. coli</td>
<td>44</td>
</tr>
<tr>
<td>2.15.3</td>
<td>Protoplast transformation</td>
<td>44</td>
</tr>
<tr>
<td>2.16</td>
<td>CONSTRUCTION OF A COSMID LIBRARY FROM STREPTOMYCES 14386</td>
<td>45</td>
</tr>
<tr>
<td>2.16.1</td>
<td>Preparation of the ligation reaction</td>
<td>45</td>
</tr>
<tr>
<td>2.16.2</td>
<td>Ligation and phage infection</td>
<td>45</td>
</tr>
<tr>
<td>2.16.3</td>
<td>Generation of the cosmid library</td>
<td>46</td>
</tr>
<tr>
<td>2.16.4</td>
<td>Robotically produced high-density colony arrays</td>
<td>46</td>
</tr>
<tr>
<td>2.17</td>
<td>DNA HYBRIDIZATION</td>
<td>47</td>
</tr>
<tr>
<td>2.17.1</td>
<td>Southern Blot</td>
<td>47</td>
</tr>
<tr>
<td>2.17.2</td>
<td>Vacuum Blot</td>
<td>47</td>
</tr>
</tbody>
</table>
2.17.3 Generation of DIG-labeled probes 
2.17.4 Hybridization and detection with the DIG labeling system 
2.17.5 Stripping of membranes 
2.17.6 Colony Hybridization 
2.17.7 Analysis of DNA- and protein sequences 

2.18 PROTEIN AND ENZYMATIC ANALYSIS 
2.18.1 Cultivation and cell disruption 
2.18.2 Protein purification 
2.18.2.1 GST-tagged proteins 
2.18.2.2 His-tagged proteins 
2.18.3 Bradford assay 
2.18.4 Protein concentration by centrifugation 
2.18.5 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) 
2.18.5.1 Preparing of SDS gels and electrophoresis 
2.18.5.2 Detection of protein bands 
2.18.6 Protein identification by MALDI-MS 
2.18.7 Cyclodeaminase assay 
2.18.8 ATP-PPi exchange assay 

2.19 ANALYSIS OF SECONDARY METABOLITES IN NATURAL PRODUCER STRAINS AND RECOMBINANT 
STREPTOMYCES STRAINS 
2.19.1 Cultivation in production medium 
2.19.1.1 Phenalinolactones 
2.19.1.2 GE81112 
2.19.2 Feeding experiments 
2.19.3 Extraction 
2.19.4 Purification of GE compounds 
2.19.5 Analysis of the secondary metabolites 
2.19.5.1 HPLC-UV/Vis, HPLC-MS, HPLC-MS/MS, LTQ-Orbitrap-MS 
2.19.5.2 NMR 

3 RESULTS 

3.1 HETEROLOGOUS EXPRESSION AND GENETIC ENGINEERING OF THE PHENALINOLACTONE BIOSYNTHETIC 
GENE CLUSTER USING RED/ET RECOMBINEERING 
3.1.1 Goals 
3.1.2 Reconstitution of the complete pla biosynthetic gene cluster 
3.1.2.1 Cloning strategy 
3.1.2.2 Cloning step 1: introduction of the Scal-oriT-tet-trpE box at the 5'-end of the pla biosynthetic gene 
cluster 
3.1.2.3 Cloning step 2: introduction of the zeocin resistance gene at the 3' end 
3.1.2.4 Cloning step 3: stitching of the pla biosynthetic gene cluster 
3.1.3 Transformation of the biosynthetic gene cluster into heterologous host strains 
3.1.3.1 Triparental conjugation of the expression construct CPhl8 into pseudomonads
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.3.2</td>
<td>Conjugation of the expression construct CPhl8 into streptomycetes</td>
</tr>
<tr>
<td>3.1.3.3</td>
<td>Verification of the mutants</td>
</tr>
<tr>
<td>3.1.4</td>
<td>Heterologous expression of the phenalinolactone biosynthetic gene cluster</td>
</tr>
<tr>
<td>3.1.4.1</td>
<td>Cultivation of pseudomonads</td>
</tr>
<tr>
<td>3.1.4.2</td>
<td>Cultivation of streptomycetes</td>
</tr>
<tr>
<td>3.1.4.3</td>
<td>HPLC-MS analysis</td>
</tr>
<tr>
<td>3.1.4.3.1</td>
<td>Production profile of the natural producer <em>Streptomyces</em> TÜ6071</td>
</tr>
<tr>
<td>3.1.4.3.2</td>
<td>Production profile of the <em>Pseudomonas</em> mutants</td>
</tr>
<tr>
<td>3.1.4.3.3</td>
<td>Production profile of the <em>Streptomyces</em> mutants</td>
</tr>
<tr>
<td>3.1.4.4</td>
<td>Optimization of heterologous PL production in <em>Pseudomonas</em></td>
</tr>
<tr>
<td>3.1.4.4.1</td>
<td>Promoter exchange</td>
</tr>
<tr>
<td>3.1.4.5</td>
<td>Optimization of heterologous PL production in <em>Streptomyces</em></td>
</tr>
<tr>
<td>3.1.5</td>
<td>Genetic modification of the <em>pla</em> biosynthetic gene cluster</td>
</tr>
<tr>
<td>3.1.5.1</td>
<td>Deletion of <em>plaA6</em></td>
</tr>
<tr>
<td>3.1.5.2</td>
<td>Deletion of <em>plaP2</em> and <em>plaP5</em></td>
</tr>
<tr>
<td>3.1.5.3</td>
<td>Verification of the new PL derivatives by high-resolution Orbitrap MS analysis</td>
</tr>
<tr>
<td>3.2</td>
<td>IDENTIFICATION OF THE GE81112 BIOSYNTHETIC GENE CLUSTER</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Goals</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Production profile of <em>Streptomyces</em> 14386</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Generation of a cosmid DNA library</td>
</tr>
<tr>
<td>3.2.3.1</td>
<td>Retrosynthetic analysis of GE81112 to design suitable probes</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Design of the probes</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Identification of an unknown NRPS biosynthetic gene cluster</td>
</tr>
<tr>
<td>3.2.5.1</td>
<td>Identification of two cosmids containing NRPS genes</td>
</tr>
<tr>
<td>3.2.5.2</td>
<td>Annotation of the unknown biosynthetic gene cluster</td>
</tr>
<tr>
<td>3.2.5.3</td>
<td>Analysis of the NRPS domains</td>
</tr>
<tr>
<td>3.2.5.3.1</td>
<td>A domains</td>
</tr>
<tr>
<td>3.2.5.3.2</td>
<td>PCP, C and E domains</td>
</tr>
<tr>
<td>3.2.6</td>
<td>Analysis of the cyclodeaminase from cosmid FD10</td>
</tr>
<tr>
<td>3.2.6.1</td>
<td>Heterologous expression of cosmids FD10 and A16</td>
</tr>
<tr>
<td>3.2.7</td>
<td>Isolation and attempted structure elucidation of three chlorinated substances</td>
</tr>
<tr>
<td>3.2.8</td>
<td>NMR structure elucidation</td>
</tr>
<tr>
<td>3.2.9</td>
<td>Identification of the GE81112 biosynthetic gene cluster</td>
</tr>
<tr>
<td>3.2.9.1</td>
<td>Identification of a cosmid containing GE81112 biosynthetic genes</td>
</tr>
<tr>
<td>3.2.9.2</td>
<td>Identification of overlapping cosmids</td>
</tr>
<tr>
<td>3.2.9.3</td>
<td>Annotation</td>
</tr>
<tr>
<td>3.2.9.4</td>
<td>Analysis of NRPS domains</td>
</tr>
<tr>
<td>3.2.9.4.1</td>
<td>Analysis of the A domains</td>
</tr>
<tr>
<td>3.2.9.4.2</td>
<td>Analysis of the C and PCP domains</td>
</tr>
<tr>
<td>3.2.9.4.3</td>
<td>Analysis of the TE domains</td>
</tr>
<tr>
<td>3.2.9.4.4</td>
<td>Docking domain analysis</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.2.10</td>
<td>Insertional mutagenesis of the GE81112 biosynthetic gene cluster</td>
</tr>
<tr>
<td>3.2.11</td>
<td>Feeding experiments</td>
</tr>
<tr>
<td>3.2.12</td>
<td>Biochemical characterization of the GE81112 biosynthetic gene cluster</td>
</tr>
<tr>
<td>3.2.12.1</td>
<td>Expression and purification of the lysine cyclodeaminase GeD</td>
</tr>
<tr>
<td>3.2.12.2</td>
<td>Cyclodeaminase assay</td>
</tr>
<tr>
<td>3.2.12.3</td>
<td>Expression of A domains</td>
</tr>
<tr>
<td>3.2.12.4</td>
<td>ATP-PP(_i) exchange assay</td>
</tr>
<tr>
<td>4</td>
<td>DISCUSSION</td>
</tr>
<tr>
<td>4.1</td>
<td>General scope of this work</td>
</tr>
<tr>
<td>4.2</td>
<td>Phenalinolactone - Heterologous expression of a complex secondary metabolite</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Heterologous expression of natural product pathways</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Heterologous expression of the PL pathway in pseudomonads</td>
</tr>
<tr>
<td>4.2.3</td>
<td>Heterologous expression of the PL pathway in streptomycetes</td>
</tr>
<tr>
<td>4.2.4</td>
<td>Genetic modification</td>
</tr>
<tr>
<td>4.2.5</td>
<td>Future aspects</td>
</tr>
<tr>
<td>4.3</td>
<td>GE81112 family compounds - Novel tetrapeptide inhibitors of bacterial protein synthesis</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Attempts to identify the GE81112 biosynthetic gene cluster</td>
</tr>
<tr>
<td>4.3.1.1</td>
<td>Generation of a cosmid library and hybridization strategy</td>
</tr>
<tr>
<td>4.3.1.2</td>
<td>Identification of an unknown NRPS biosynthetic gene cluster</td>
</tr>
<tr>
<td>4.3.1.3</td>
<td>Heterologous expression</td>
</tr>
<tr>
<td>4.3.2</td>
<td>New strategy-new probe</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Analysis of the GE81112 biosynthetic gene cluster</td>
</tr>
<tr>
<td>4.3.3.1</td>
<td>Biosynthetic proposal for the GE81112 compound family</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Summary and Future aspects</td>
</tr>
<tr>
<td>5</td>
<td>APPENDIX</td>
</tr>
<tr>
<td>6</td>
<td>REFERENCES</td>
</tr>
</tbody>
</table>
Figures

**Figure 1.1:** Prominent natural products used in clinical applications. 3
**Figure 1.2:** New substances derived from natural products. 4
**Figure 1.3:** Streptomycetes colonies. 5
**Figure 1.4:** Natural compounds from actinomycetes and streptomycetes which serve as important medicines. 6
**Figure 1.5:** Mechanism of Red/ET recombineering. 10
**Figure 1.6:** Red/ET recombineering. 11
**Figure 1.7:** General strategy for the heterologous expression of large biosynthetic pathways. 12
**Figure 1.8:** Reactions catalyzed by the essential domains in NRPS multimodular assembly lines. 14
**Figure 1.9:** The multiparticle thiotemplate mechanism illustrated with the example of tyrocidine A synthesis. 15
**Figure 1.10:** The enterobactin assembly line as example for iterative NRPSs. 16
**Figure 1.11:** The non-linear yersiniabactin NRPS. 17
**Figure 2.1:** Spotting array of the high-density colony filters. 46
**Figure 3.1:** Description of the cloning strategy. 58
**Figure 3.2:** Red/ET Cloning step 1. 59
**Figure 3.3:** Red/ET Cloning step 2. 60
**Figure 3.4:** Red/ET Cloning step 3. 61
**Figure 3.5:** Restriction analysis of the Red/ET recombination constructs. 62
**Figure 3.6:** Verification of *P. putida* KT2240::CPhl8 mutants by colony PCR. 63
**Figure 3.7:** Verification of *S. lividans* TK24::CPhl8 mutants by PCR. 64
**Figure 3.8:** Chemical structures of phenalinolactones A, B, C and D. 65
**Figure 3.9:** HPLC-MS analysis of the natural producer *Streptomyces* Tü6071. 66
**Figure 3.10:** ESI-MS/MS fragmentation patterns of PL A and PL D. 67
**Figure 3.11:** HPLC/MS analysis of streptomycetes mutants carrying the PL pathway. 68
**Figure 3.12:** Introduction of the inducible Pm-promoter in front of *plaM2* by Red/ET. 69
**Figure 3.13:** Introduction of two *ermE* promoters between *plaO2* and *plaM2* by Red/ET recombineering. 72
**Figure 3.14:** Deletion of the glycosyl transferase encoding gene *plaA6*. 74
**Figure 3.15:** HPLC/MS analysis of *S. coelicolor::CPhl18*. 76
**Figure 3.16:** Verification of the unglycosylated derivative PL E by LTQ-Orbitrap MS analysis. 77
**Figure 3.17:** Analysis of the extracts of strain *S. 14386* by high-resolution MS on an LTQ Orbitrap mass spectrometer. 80
**Figure 3.18:** HPLC/MS analysis of an extract of strain *S. 14386*. 81
**Figure 3.19:** Analysis by high resolution MS of the substances with molecular masses 631, 645, 659. 82
**Figure 3.20:** General overview of the strategy for generation of a cosmid library. 84
**Figure 3.21:** Digestion of genomic DNA from *S. 14386* with *Sau3A*. 85
**Figure 3.22:** Analysis of 19 randomly chosen cosmid clones. 85
Figure 3.23: Structures of GE81112 metabolites.

Figure 3.24: Example of conversion of L-lysine to L-pipeolic acid by the action of a lysine cyclodeaminase as it is known from rapamycin.

Figure 3.25: High density colonies filters.

Figure 3.26: Sequencing results from cosmids FD10 and AI6.

Figure 3.27: Consensus core motifs of A domains from cosmid FD10.

Figure 3.28: Alignment of A domains from cosmid FD10 with the A domain of GrsA.

Figure 3.29: PCP consensus motif.

Figure 3.30: Alignment of the cyclodeaminase from cosmid FD10 with LipE, RapL and TubZ.

Figure 3.31: Analysis of S. lividans::FD10 extract by high-resolution MS.

Figure 3.32: Purification procedure for substances 631, 645 and 659.

Figure 3.33: HPLC-MS analysis of heptane and MeOH fractions.

Figure 3.34: HPLC/MS analysis of different purified fractions after preparative HPLC.

Figure 3.35: HPLC-MS analysis of different fractions after preparative HPLC.

Figure 3.36: Postulated structural elements of compound 631 identified by NMR spectroscopy.

Figure 3.37: Southern Blot analysis of genomic DNA from S. 14386.

Figure 3.38: Screening of S. 14386 cosmid library with the specific cyclodeaminase probe.

Figure 3.39: Screening of the cosmid library to identify cosmids which overlapped the T7 end of cosmid BI11.

Figure 3.40: Southern Blot analysis of cosmid BI11 and the overlapping cosmid.

Figure 3.41: GE81112 biosynthetic gene cluster on cosmids BI11 and BA23.

Figure 3.42: The GE81112 biosynthetic gene cluster.

Figure 3.43: GE81112 biosynthetic gene cluster.

Figure 3.44: Predicted interaction of NRPS enzymes from the GE81112 biosynthetic gene cluster.

Figure 3.45: Proposed NRPS interaction and order of subunits in the GE81112 biosynthesis.

Figure 3.46: High resolution MS analysis of extracts of S. 14386 wild type and S. 14386::pKC1132 _BI11_PipA mutant.

Figure 3.47: Construct pGEX_cyclo for the expression of geD as a GST fusion protein in E. coli.

Figure 3.48: Expression of the construct pGEX_cyclo in E. coli BL21.

Figure 3.49: Expression of the cyclodeaminase GeD.

Figure 3.50: Purified protein GeD.

Figure 3.51: Cyclodeaminase assays.

Figure 3.52: Expression of A domains in E. coli Rosetta BL21 (DE3) pLysS/RARE at 16 °C.

Figure 3.53: SDS-PAGE of adenylation domains overexpressed in E. coli.

Figure 3.54: Relative substrate specificities of the adenylation domains GeEA1, GeGA2.

Figure 4.1: Proposed biosynthetic pathway for phenalinolactone formation.

Figure 4.2: Proposed biosynthesis of possible PL precursors. IPP: isopentenyl pyrophosphate, GGDP: geranylgeranyl diphosphate.

Figure 4.3: PL derivatives expected after deletion of different biosynthetic genes, in heterologous streptomyces hosts.
| Figure 4.4: | Module and domain arrangement from NRPS biosynthetic gene cluster on cosmid AI6 and FD10. | 142 |
| Figure 4.5: | Nonlinear arrangement of the GE81112 biosynthetic gene cluster. | 147 |
| Figure 4.6: | GE81112 tetrapeptide structure and corresponding amino acid sequence Pip-Orn/Gln-His-His. | 148 |
| Figure 4.7: | A) Nonlinear genetic arrangement of the GE81112 NRPS and B) proposed interaction of NRPS proteins to yield a linear assembly line. | 149 |
| Figure 4.8: | Proposed biosynthetic mechanisms. | 150 |
| Figure 4.9: | Proposed biosynthetic mechanisms for generation of the second building block. | 151 |
| Figure 4.10: | Proposed mechanism for the PCP-bound cyclization of arginine by a putative arginine cyclase | 152 |
| Figure 4.11: | A linear model for the GE81112 NRPS templated assembly from four amino acids. | 154 |
| Figure 4.12: | Proposed mechanism for the biosynthesis and incorporation of L-hydroxyvaline in the zorbamycin biosynthetic pathway. | 155 |
| Figure 4.13: | Sequence alignment of acyltransferase/thioesterases found in NRPS clusters. | 156 |
| Figure 4.14: | Proposed GE81112 biosynthetic pathway of factor B1. | 157 |
## Tables

| Table 2.1: | Chemicals used and their sources | 20 |
| Table 2.2: | Enzymes, kits and markers and their manufacturers | 21 |
| Table 2.3: | Solutions and buffers for molecular biology | 22 |
| Table 2.4: | Solution and buffers for Southern Blot and colony hybridization | 23 |
| Table 2.5: | Solution and buffers for enzymatic applications | 24 |
| Table 2.6: | Media for \textit{E. coli} cultivation | 26 |
| Table 2.7: | Media for cultivation of pseudomonads | 26 |
| Table 2.8: | Media for cultivation of streptomycetes | 27 |
| Table 2.9: | Antibiotics used in this work | 28 |
| Table 2.10: | Bacterial strains used in this work | 31 |
| Table 2.11: | Primers used in the phenalinolactone project, DST= desalted | 32 |
| Table 2.12: | Primers used in the GE81112 project | 33 |
| Table 2.13: | General plasmids used in this work | 34 |
| Table 2.14: | Plasmids used in the phenalinolactone project | 35 |
| Table 2.15: | Plasmids used in the GE81112 project | 36 |
| Table 2.16: | Mutants of \textit{S. lividans} TK24 generated in this work | 37 |
| Table 2.17: | Mutants of \textit{S. coelicolor} A3(2) generated in this work | 37 |
| Table 2.18: | Mutants of \textit{S. coelicolor} M512 generated in this work | 38 |
| Table 2.19: | Mutants of \textit{S. 14386} generated in this work | 38 |
| Table 2.20: | PCR protocols | 41 |
| Table 2.21: | Imidazol gradient applied for Äkta system. | 51 |
| Table 3.1: | Phenalinolactone production by \textit{Streptomyces} Tü6071, \textit{S. lividans} TK24, \textit{S. coelicolor} A3(2) and \textit{S. coelicolor} M512 during cultivation in different media at 28 °C | 71 |
| Table 3.2: | Phenalinolactone production by \textit{Streptomyces} Tü6071, \textit{S. lividans} TK24, \textit{S. coelicolor} A3(2) and \textit{S. coelicolor} M512 during cultivation in NL111 medium at 28 °C and 37 °C | 71 |
| Table 3.3: | \textit{Streptomyces} mutants obtained after conjugation with the modified constructs | 75 |
| Table 3.4: | Principal fragments of the standard substance PL D and the unglycosylated derivative PL | 78 |
| Table 3.5: | GE81112 compounds produced under different cultivation conditions | 79 |
| Table 3.6: | End sequencing results for cosmids FD10 and AI6 with T4 and T7 primers | 89 |
| Table 3.7: | List of NRPS genes from cosmids FD10 and AI6 | 90 |
| Table 3.8: | Putative A domain ‘code residues’ within A domains from both clusters | 94 |
| Table 3.9: | Putative A domain ‘code residues’ within A domains from cosmids | 106 |
| Table 3.10: | Substrate specificity pocket of pipecolic acid incorporating A domains | 106 |
| Table 3.11: | End sequencing results of cosmid BI11 with T7 and T4 primers | 107 |
| Table 3.12: | BLAST results for genes on the cosmids BI11 and BA23 | 109 |
| Table 3.13: | Putative A domain ‘code residues’ within A domains in the GE81112 cluster | 114 |
| Table 3.14: | Transformation efficiency of \textit{S. 14386} by biparental conjugation | 119 |
| Table 5.1: | Sequence alignment of putative modules 1-5 from the GE81112 biosynthetic pathway with two modules of the NRPS pathway from cosmid FD10 | 159 |
| Table 5.2: | Sequence alignment of proteins GeJ and GeM from the GE81112 biosynthetic pathway | 161 |
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A domain</td>
<td>Adenylation domain</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>AMP</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>Apra</td>
<td>Apramycin</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosome</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumine</td>
</tr>
<tr>
<td>BPC</td>
<td>Base peak chromatogram</td>
</tr>
<tr>
<td>C</td>
<td>Carbon</td>
</tr>
<tr>
<td>C domain</td>
<td>Condensation domain</td>
</tr>
<tr>
<td>Cm</td>
<td>Chloramphenicol</td>
</tr>
<tr>
<td>CoA</td>
<td>CoenzymeA</td>
</tr>
<tr>
<td>Cos</td>
<td>Cosmid</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DIG</td>
<td>Digoxigenin</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Desoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Desoxyribonucleosid triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>1,4-Dithiothreitol</td>
</tr>
<tr>
<td>E domain</td>
<td>Epimerization domain</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediamine tetraacetic acid</td>
</tr>
<tr>
<td>EIC</td>
<td>Extracted ion chromatogram</td>
</tr>
<tr>
<td>h</td>
<td>Hour</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>Kan</td>
<td>Kanamycin</td>
</tr>
<tr>
<td>kb</td>
<td>Kilo base pairs</td>
</tr>
<tr>
<td>kV</td>
<td>Kilovolt</td>
</tr>
<tr>
<td>l</td>
<td>Liter</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>Mbp</td>
<td>Mega base pairs</td>
</tr>
<tr>
<td>Me</td>
<td>Methyl</td>
</tr>
<tr>
<td>Mg</td>
<td>Magnesium</td>
</tr>
<tr>
<td>mg</td>
<td>Milligram</td>
</tr>
<tr>
<td>min</td>
<td>Minutes</td>
</tr>
<tr>
<td>ml</td>
<td>Milliliter</td>
</tr>
<tr>
<td>mM</td>
<td>Millimolar</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>Tandem mass spectrometry</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass-to-charge-ratio</td>
</tr>
<tr>
<td>N</td>
<td>Nitrogen</td>
</tr>
<tr>
<td>NRPS</td>
<td>Nonribosomal peptide synthetase</td>
</tr>
<tr>
<td>O</td>
<td>Oxygen</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>Orf</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>oriT</td>
<td>Origin of transfer</td>
</tr>
<tr>
<td>P</td>
<td>Phosphor</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamid gelelectrophoresis</td>
</tr>
<tr>
<td>PCP</td>
<td>Peptidyl carrier protein</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PKS</td>
<td>Polyketide synthase</td>
</tr>
<tr>
<td>Ppant</td>
<td>Phosphopantethein</td>
</tr>
<tr>
<td>PP_i</td>
<td>Pyrophosphate</td>
</tr>
<tr>
<td>RBS</td>
<td>Ribosome binding site</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>Rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>Rt</td>
<td>Retention time</td>
</tr>
<tr>
<td>RT</td>
<td>Room temperature</td>
</tr>
<tr>
<td>S</td>
<td>Sulfur</td>
</tr>
<tr>
<td>SAP</td>
<td>Shrimps alkaline phosphatase</td>
</tr>
<tr>
<td>Sec</td>
<td>Seconds</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>sp.</td>
<td>Species</td>
</tr>
<tr>
<td>TE</td>
<td>Thioesterase</td>
</tr>
<tr>
<td>Tet</td>
<td>Tetracycline</td>
</tr>
<tr>
<td>U</td>
<td>Units</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>Volt</td>
</tr>
<tr>
<td>v/v</td>
<td>Volume to volume</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
<tr>
<td>Zeo</td>
<td>Zeocin</td>
</tr>
<tr>
<td>°C</td>
<td>degree Celcius</td>
</tr>
<tr>
<td>μF</td>
<td>micro Farad</td>
</tr>
<tr>
<td>μ</td>
<td>micro</td>
</tr>
<tr>
<td>Ω</td>
<td>Ohm</td>
</tr>
</tbody>
</table>
Zusammenfassung


Abstract

The heterologous expression of natural product biosynthetic pathways is of increasing interest in biotechnology and drug discovery. This approach enables the production of complex metabolites in more amenable host organisms and provides the basis for the generation of novel analogs through genetic engineering. In this thesis, we describe a straightforward strategy for the heterologous expression of the highly complex phenalinolactone biosynthetic pathway, which was recently cloned from *Streptomyces* sp. Tü6071. Using Red/ET recombineering, the phenalinolactone pathway was reconstituted from two cosmids and heterologously expressed in several *Streptomyces* strains. The established expression system now provides a convenient platform for functional investigations of the biosynthetic genes and the generation of novel analogs, by genetic engineering of the pathway in *Escherichia coli*.

The second part of this thesis describes work on a distinct class of secondary metabolites, the GE81112 tetrapeptide family. We developed a strategy for the cloning and identification of the GE81112 biosynthetic gene cluster, in order to investigate the biosynthetic pathway in detail. Generation of a cosmid library enabled us to identify the corresponding biosynthetic gene cluster on two overlapping cosmids. In parallel, we established methods to manipulate the strain genetically, allowing us to verify the identity of the GE81112 gene cluster by gene inactivation experiments. In addition, we characterized several proteins from the pathway using enzymatic assays *in vitro*. Taken together, these data have enabled us to propose a preliminary model for GE81112 biosynthesis. The results also open the door to developing new derivatives of these promising compounds by genetic engineering.
1 Introduction

1.1 Natural products as a robust source of new drugs and drug leads

Microorganisms produce a vast number of secondary metabolites with useful biological activities for both human and veterinary medicine, including compounds which function as antibiotics, immunosuppressive and antitumor agents, as well as herbicides, insecticides and antiparasitic agents. Figure 1.1 exemplifies a further characteristic feature of these so-called ‘natural products’: their significant structural diversity. The recruitment of many different building blocks, as well as the activity of diverse modifying enzymes, leads to the construction of highly complex chemical structures.

Figure 1.1: Prominent natural products used in clinical applications.
Natural product research, as we know it today, began with the serendipitous discovery of penicillin in 1929 by Alexander Fleming – an event which inaugurated a new era in the treatment of bacterial diseases [1]. Since then, many natural products and synthetically modified natural product derivatives have been successfully developed for clinical use to treat human diseases, across a broad range of therapeutic areas [2]. Even today, natural products continue to play a dominant role in the discovery of leads for the development of new drugs. In fact, approximately 50% of the pharmaceuticals currently in clinical use are of natural product origin.

A special focus in the field of natural product research is the development of new antibiotics [3]. Motivating these efforts is the fact that infectious diseases remain the second-leading cause of death worldwide and the third-leading cause of mortality in economically-advanced countries [4]. In order to deal with growing bacterial resistance to existing drugs, there is a pressing need for the discovery and development of new antibiotics [5]. Despite the fact that large pharmaceutical companies have significantly de-emphasized natural product discovery over the past several years, many new and interesting molecules with unique scaffolds and/or novel modes of action have been found during this time by research efforts in academia and small pharma companies [6]. Smaller biotechnology companies, in particular, have increased their focus on natural product research, mainly in the infectious diseases but also in anticancer areas (Figure 1.2) [6]. These findings emphasize that natural product research continues to provide significant value in the discovery of novel chemical structures and bioactive leads for clinical development.

![Platensimycin](Antibacterial.png)  ![Sorbicillactone](Anticancer.png)  ![Cyanosporaside A](Anticancer.png)

*Figure 1.2: New substances derived from natural products.*

Among the natural producers of secondary metabolites, beside plants, microorganisms are a particularly promising source of novel bioactive entities. In nature, the diversity of
microorganisms is enormous and only a very minor portion of bacterial and fungal species have been cultured to date and examined for secondary metabolite production [7]. The majority of bioactive compounds were isolated from, non-filamentous bacteria, fungi and actinomycetes [7;8].

With respect of secondary metabolism, actinomycetes are the most-intensively studied group; among them, the genus *Streptomyces* has been proven to be the most proficient producer of secondary metabolites, at least under standard laboratory conditions [9]. A wide array of bioactive molecules has been isolated from streptomycetes, leading to their historical importance in natural product discovery [3].

1.2 *Streptomycetes* as important natural product producers

Streptomycetes are soil-dwelling bacteria and ubiquitous in nature. Their ability to colonise the soil is greatly facilitated by growth as a vegetative hyphal mass which can differentiate into spores that assist in spread and persistence (Figure 1.3). The spores are a semi-dormant stage in the life-cycle, and can survive in soil for long periods (several years).

![Streptomycetes colonies.](image)

**Figure 1.3:** Streptomycetes colonies.  
(a) Typical colony morphologies of *Streptomyces* sp. isolated from the soil. (b) A panoramic view of the colony morphology of *Streptomyces coelicolor*. Both peripheral and aerial mycelia develop from the central mass of the colony. Metabolites, including the blue antibiotic actinorhodin, are excreted into the medium and into aqueous droplets on the hydrophobic surface of the colony. (Figures reproduced from reference [9]).

Streptomycetes produce many extracellular enzymes, such as cellulases, xylanases, amylases, maltases, and others. These catalysts make them important players in soil biodegradation processes, as they are able to decompose complex mixtures of polymers in dead plants, animal and fungal material [10]. Further characteristics of these strains are their high G+C content (72% on average), linear chromosomes and large genome sizes. The sequenced genome of *Streptomyces coelicolor* comprises 8.6 Mbp [11], whereas the *S. avermitilis*
The genome contains 9 Mbp [12]. However, their most notable characteristic is their capability to produce biologically active secondary metabolites, a feature which has made them the focus of natural product research.

Streptomycetes were first recognized as natural product producers some 60 years ago, with the discovery of streptomycin [13]. Since then, the majority of known antibiotics have been isolated from streptomycetes. Of the 12000 or so antibiotics known by 1995, 55% were derived from streptomycetes and 11% from other actinomycetes [14]. Among these antibiotics are drugs which have been in clinical use as antibiotics for many years, including the macrolide antibiotic erythromycin A synthesized by the actinomycete *Saccharopolyspora erythraea* [15] and the aromatic polyketide tetracycline [16] (Figure 1.4).

![Figure 1.4: Natural compounds from actinomycetes and streptomycetes which serve as important medicines.](image)

Another prominent example of a therapeutically relevant streptomycese natural product is the cyclic lipopeptide antibiotic daptomycin (Cubicin®) from *S. roseosporus*. Daptomycin was approved in the United States in 2003 and in Germany in 2006 for the treatment of skin infections caused by Gram-positive pathogens. In addition to antibiotics, streptomycetes have
also yielded important anticancer compounds, such as the cytostatic doxorubicin [17], which is widely used in chemotherapy (Figure 1.4).

As the vast majority of therapeutically important secondary metabolites are produced from *Streptomyces* sp., making them a proven source for therapeutically-useful compounds, they should be the continued focus of research to identify new bioactive substances. In the last years intense efforts have also been made to discover new groups of microbial producers with the assumption that an increase to discover new organisms correlates with the potential to discover chemical diversity (new structures). These investigations dealt with marine bacteria, pathogenic bacteria (pseudomonads) or organisms that are difficult to cultivate (cyanobacteria) [18-20]. Among the bacteria discovered in these screenings especially the group of myxobacteria have attracted attention as proficient producers of novel bioactive compounds [21].

In the majority of the cases, these secondary metabolites are the end products of complex, multistep biosynthetic processes. Understanding the mechanisms involved in the biosynthesis of the compounds is an essential prerequisite for optimizing production and yield. Furthermore, directed manipulation of the genes governing secondary metabolism offers a promising alternative to total or semi-synthesis, to generate altered natural products. The success of this approach depends on the cloning and subsequent genetic and biochemical characterization of the corresponding biosynthetic pathways.

Fortunately, in the majority of cases, the compounds are produced by sets of genes which are clustered in the bacterial chromosome; this feature presumably allows the microorganisms to co-regulate the expression of the many genes. The first proof that the genes for antibiotic biosynthesis are clustered was provided by Malpartida and Hopwood in 1984 [22]. The researchers isolated and cloned a continuous DNA fragment from *S. coelicolor* which apparently carried the complete genetic information required for the biosynthesis of the antibiotic actinorhodin. Indeed, the cloned DNA fragment could complement several actinorhodin non-producing mutants. Furthermore, introduction of the DNA fragment into the heterologous host *S. pavulus* directed actinorhodin biosynthesis – providing proof that the cluster of genes located on this DNA fragment was responsible for the production of the antibiotic actinorhodin. The advantage for natural product researchers is that if a single gene within the gene cluster is located, the others can mostly be identified fairly straightforwardly by chromosomal walking.

Many molecular tools have been developed for targeted genetic manipulation in streptomyces over the last several decades [10] and hundreds of biosynthetic gene clusters...
have been identified to date. Together, these developments have enabled detailed analysis of the molecular principles that govern multiple pathways [23]. Recently, this field has been revolutionized by advances in sequencing technologies, which now allow the rapid sequencing of entire bacterial genomes in a matter of days. The first *Streptomyces* strain to be sequenced was *S. coelicolor* [11]. Before the genome of *S. coelicolor* was sequenced, it was known that the strain produces three antibiotics, actinorhodin, prodiginin and the nonribosomal peptide calcium-dependent antibiotic (CDA) [10;11]. However, the genome sequence revealed a further 23 gene clusters (comprising approximately 5% of the total genome), which are likely to be dedicated to secondary metabolism.

Thirty gene clusters were identified in the second sequenced strain, *S. avermitilis*, which could be correlated to secondary metabolism, corresponding to 6.5% of the genome. Although antibiotic production by the *Streptomyces* has been studied intensively over the last century, mathematical models predict that only 3% of antibacterial compounds synthesized by these bacteria have been discovered to date [24]. These facts indicate that the genomic capacity of these microorganisms to synthesise natural products is much higher than originally anticipated, a hypothesis which is supported by the available genome sequencing data. In order to identify, isolate and evaluate these as yet uncharacterized biosynthetic gene clusters and exploit them for analogue production, techniques must be developed to allow the straightforward genetic manipulation of the strains. Genetic manipulation in natural producer strains is straightforward in some model organisms (for example *S. coelicolor*) but for many other *Streptomyces* strains it remains more challenging, although there has been some progress [10]. Many of these strains are slow-growing or not yet culturable, which further hinders the genetic manipulation of the biosynthetic pathways.

One very useful approach is to express entire gene clusters in genetically more tractable production hosts, which facilitates the genetic manipulation of the secondary metabolite pathways. This so-called ‘heterologous expression’ technique represents a promising tool for both obtaining deeper insights into the biosynthetic pathways, and for generating novel derivatives of natural compounds for evaluation as drug leads.

### 1.3 Heterologous expression of natural product pathways

There has been considerable recent interest in both the natural product research and drug discovery communities, in developing heterologous expression techniques for complete secondary metabolite pathways. Furthermore, heterologous expression can also be used to
screen genomic DNA obtained from microbial niches (metagenomes) for the production of novel secondary metabolites [25;26].

Efforts towards achieving heterologous expression have met with considerable success during the last decade [27-29]. However, in most cases, only small biosynthetic gene clusters were expressed in hosts which were closely-related to the producing strains [30]. To date, the heterologous expression of secondary metabolite pathways from *Streptomyces* has mainly been achieved in related actinomycete species, and the relatively small biosynthetic gene clusters (less than 30 kbp) were often located on a single cosmid or bacterial artificial chromosome (BAC) within a genomic library [31-33]. Expression of larger and more complex pathways remains more demanding, but can be facilitated by constructing several plasmids that harbor subsets of the biosynthetic genes [34].

As this approach requires time consuming classical cloning steps, however, a new method for the expression and engineering of large biosynthetic pathways was developed in our working group by Wenzel *et al.* [35]. In this case, Red/ET (λ-mediated) recombineering technology [36-38] was used to reconstitute the hybrid nonribosomal peptide synthetase/polyketide synthase (NRPS/PKS) myxochromide S pathway from the myxobacterium *Stigmatella aurantiaca* on one expression construct and to integrate genetic elements for expression in *Pseudomonas putida*. Heterologous expression was successful, and the construct yielded higher amounts of myxochromide in *P. putida* than in the natural producer *Stigmatella aurantiaca* [35;39;40].

Red/ET is a novel technique for DNA manipulation, which is precise and independent of the presence of restriction sites and the size of the DNA molecule. These features make it ideal for the manipulation of large DNA fragments. DNA molecules targeted by Red/ET recombineering are precisely altered by homologous recombination in strains of *E. coli* which express phage-derived protein pairs, either RecE/RecT from the Rac prophage, or Redα/Redβ from λ phage. These protein pairs are functionally and operationally equivalent. RecE and Redα are 5’→3’ exonucleases, while RecT and Redβ are DNA annealing proteins (Figure 1.5).
A functional interaction between RecE and RecT, or between Redα and Redβ, is also required in order to catalyze the homologous recombination reaction. Recombination is mediated by homology regions, which are stretches of identical DNA sequences found on both molecules to be recombined (Figure 1.6). As the sequence of the homology regions can be generated in a chemically synthesized oligonucleotide, any position on a target molecule can be specifically altered. Red/ET recombination thus allows, at least in principle, every type of possible DNA modification (insertions, deletions, substitutions, fusions, point mutations, direct cloning and subcloning) regardless of the size of the target molecule (e.g. cosmids or BACs) [41].
Perlova, et al. [42] also used Red/ET to stitch together a myxobacterial biosynthetic pathway. Here, Red/ET recombineering was employed to modify two cosmids containing the entire myxothiazol pathway. In a final ligation step, the whole pathway was reconstituted on one expression construct, which was subsequently expressed in *Myxococcus xanthus*. Production of myxothiazol was comparable to that of the natural producer *Stigmatella aurantiaca* [42]. The same pathway was also expressed in *Pseudomonas putida* [43]. These and subsequent examples [44;45] demonstrate how this promising technique can now be used for the reconstitution of large and complex biosynthetic pathways in *E. coli* and their subsequent heterologous expression in related or non-related heterologous hosts. Figure 1.7 gives a general overview of the strategy for cloning and heterologous expression of large natural product assembly lines [35;39]
After screening of a BAC or cosmid library for constructs containing parts of the investigated natural product biosynthetic gene cluster, overlapping fragments can be stitched together in *E. coli* using Red/ET recombineering. The resulting construct harboring the complete natural product assembly line can be further modified in *E. coli* by recombineering, e.g., by inserting genes needed for transformation into the heterologous host and/or insertion of promoter region(s). The final construct can be transferred into a suitable heterologous host for expression of the biosynthetic pathway. Natural product formation can ultimately be analyzed by conventional techniques, e.g., HPLC-MS and/or TLC [35].

The successful heterologous expression of several classes of natural products, including the polyketides (PKs) and non-ribosomal polypeptides (NRPs) and their hybrids, presented a special challenge, because the biosynthesis of these complex structures typically requires large gene sets [46;47]. The core structures in each case are assembled by gigantic multienzymes called polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), respectively. In both systems, the proteins have to be posttranslationally modified in order to be active [48]. These multifunctional enzymes can reach remarkable sizes, for example mycolactone A and B from the the bacterium *Mycobacterium ulcerans* are synthezised by PKS multienzymes housing eight and seven extension modules containing 16990 amino acids and 14130 amino acids, respectively [49].

Once the proteins are expressed as functional proteins within the cell, they require a specific pool of substrates, including but not limited to short-chain fatty acids, proteinogenic and non-proteinogenic amino acids, mevalonate, shikimate pathway intermediates, and sugars to enable natural product formation. Therefore, the required substrates must be available in the heterologous hosts in adequate amounts, at the correct time point.
1.4 The biosynthetic logic of nonribosomal peptide synthetases

NRPS multienzymes are composed of repetitive, multicatalytic units called modules which carry out several reactions in a coordinated manner. Typically, each module in the assembly line performs one cycle of chain extension (incorporation of one amino acid into the growing peptide chain). One cycle of chain extension comprises three essential enzymatic reactions: 1) recognition of an amino acid as substrate and its activation as its aminoacyl-adenylate; 2) covalent binding of the amino acid as a thioester to the multienzyme; and 3) condensation with the amino acyl or peptidyl group tethered to the neighbouring module [50]. This process is schematized in Figure 1.8. Overall, peptide synthesis proceeds in an N- to C-terminal direction.

Each of the steps in chain extension is carried out by a specific domain within the module. A typical module consists minimally of an adenylation (A) domain, a peptidyl carrier protein (PCP) domain (also referred to as a thiolation (T) domain) and a condensation (C) domain [47]. The A domain is responsible for recognizing and activating the starter/extender unit, which is then covalently linked to the PCP. The attachment point is the terminal sulfhydryl of a phosphopantetheine arm attached to a highly conserved serine residue [51]. This ‘swing arm’ is added post-translationally to the PCP by a phosphopantetheinyl transferase (PPTase). The Ppant arm (20 Å in length) is characterized by its flexibility, which presumably facilitates the transport of the intermediates to the reaction centers [51]. The condensation domain then catalyses peptide bond formation between the aminoacyl-S-PCP and the incoming aminoacyl or peptidyl-S-PCP (see Figure 1.8). The release of the intermediates in a linear or cyclic form is normally performed by a thiosterase domain, which is located at the end of the last module of the assembly line [52;53].
A variety of optional domains increase the structural diversity of the products, including methyl transferase (MT), epimerase (E), oxidation (Ox) and heterocyclization (HC) activities [54]. These tailoring enzymes are embedded in the NRPS assembly lines and modify the elongating chains while they are still covalently tethered to the proteins as peptidyl-S-enzyme intermediates. Other tailoring enzymes act in trans, recognizing either or both the peptidyl chains or the PCP domains. A third group of tailoring enzymes act after the assembly-line stage of the biosynthesis, and further transform the released peptides (e.g., by introducing cross-linking or addition of glycosyl groups) [55].
In general, nonribosomal peptide synthesis is a linear process in which each module is responsible for one particular chain elongation step and the specific order of the modules defines the sequence of the incorporated amino acids. However, recent characterization of many NRPS systems has revealed several examples where the arrangement of modules and domains within the enzyme is not co-linear with the sequence of transformations required to generate the observed product. Therefore, Mootz et al. proposed a classification of NRPS systems into three groups, according to their biosynthetic logic [56]: linear NRPSs, iterative NRPSs and nonlinear NRPSs. Representative examples for each type of biosynthesis are shown in Figures 1.9–1.11.

Figure 1.9: The multiple carrier thiotemplate mechanism illustrated with the example of tyrocidine A synthesis.

Three peptide synthetases encoded by the genes tycA, tycB and tycC act together to assemble the cyclic decapeptide. Diagram reproduced from [57].

Figure 1.9 illustrates a linear NRPS, which is involved in the biosynthesis of the decapeptide tyrocidine [58]. In a linear NRPS, the core domains are arranged in the order C-A-PCP. The initiation module, which lacks the C domain, is responsible for the incorporation of the first amino acid. The terminal module normally contains a TE domain for release of the full-length peptide chain. Other examples of linear NRPSs are those responsible for biosynthesis of surfactin [59] and pristinamycin [60].
In iterative NRPSs, modules or domains are used more than once for the assembly of a particular product. This strategy is employed to construct peptide chains that consist of repeated smaller units. The key step is then the oligomerization and cyclization of the monomers by the thioesterase domain. For example, in enterobactin biosynthesis (Figure 1.10), the complete NRPS assembly line acts iteratively to generate a common structural unit (trimer of dihydroxybenzoylserine units) [61], which is then oligomerized to the final product. Bacillibactin of *B. subtilis*, a cyclotrimerized dihydroxybenzoyl-glycyl-threonine-peptide, is structurally very similar [62]. An iterative NRPS is also involved in the assembly of gramicidin [63].
Over the last several years, it has become clear that a number of biosynthetic gene clusters deviate in their domain organization from the standard C-A-PCP architecture. In fact, these nonlinear NRPS clusters comprise a major proportion of the NRPS systems found in nature [64]. The main characteristics of these systems are the unusual arrangement of the core domains C, A and PCP. Such a non-linear NRPS is depicted in Figure 1.11. The NRPS responsible for the siderophore yersiniabactin is a striking example of nonlinear NRPS assembly, where one A domain is used to load three different PCPs [65]. Due to the unusual domain organization of nonlinear NRPS biosynthetic gene clusters, it is very difficult to predict the product structures with any confidence. Therefore, detailed biochemical studies are needed to understand the function and the interplay between the enzymes, which makes research with nonlinear NRPS very challenging.
1.5 Outline of the dissertation

Studies on the biosynthesis and the heterologous expression of complex secondary metabolites from streptomycetes

The present thesis deals with experiments to heterologously express the phenalinolactone biosynthetic gene cluster from *Streptomyces* Tü6071, and to investigate the biosynthesis of the GE81112s from *Streptomyces* 14386. Figure 1.12 shows the GE81112 family and the phenalinoactones, of which the main metabolites are PL A and PL D.

![GE81112 and Phenalinolactone](image)

**Figure 1.12:** Structures of the GE81112 metabolites and the phenalinolactones.

The phenalinolactones are structurally intriguing tricyclic terpene glycosides which are produced by *Streptomyces* sp. Tü6071. The compounds exhibit promising antibacterial activity [66]. In addition to a highly oxidized γ-butyrolactone ring, which is most likely derived from pyruvate, the oxidatively functionalized terpenoid backbone is decorated with a 5-methylpyrrole-2-carboxylic acid as well as the rare deoxyhexose, 4-0-methyl-L-amicetose. The recent cloning and sequence analysis of the *pla* biosynthetic gene cluster from *Streptomyces* Tü6071 [66] revealed a highly complex genetic architecture that consists of 35 *orf s*, which are organized into 11 putative operons.

Our goal was to heterologously express the cluster using Red/ET recombineering, as the technique had already been shown to be suitable for the reconstitution and expression of natural product pathways. However, at the outset of this work, the approach had not been applied to non-PKS/NRPS gene clusters from streptomycetes. Initially the PL cluster was
located on two different cosmids. Therefore, the first aim of the work was to collect the biosynthetic genes together onto one expression construct, and then to introduce genetic elements for expression in different heterologous hosts. The established expression system was then expected to provide a convenient platform for functional investigations of the biosynthetic genes and the generation of novel analogues, by genetic engineering of the pathway in *E. coli*.

The second part of this thesis describes work on a second class of secondary metabolites, the GE81112 tetrapeptide family. The GE81112 compounds were isolated from a *Streptomyces* sp., and to date, three variant GE factors have been identified (GE81112 factors A, B1 and B). The metabolites were found in the course of microbial product screening experiments aimed at discovering novel antibiotics acting on bacterial protein synthesis. Extensive NMR and MS studies revealed that the GEs incorporate 4 uncommon amino acids: the nonproteinogenic amino acids hydroxypipecolic acid and hydroxypentanoic acid, as well as an (amino)histidine and a hydroxychlorohistidine [67] (Figure 1.12). Although the core structure of the GEs could be predicted straightforwardly to arise from a nonribosomal peptide synthetase (NRPS) multienzyme, the origin of several amino acids and functional groups, including the aminohistidine as well as the chlorination and carbamoylation, was not obvious from considerations of classical assembly-line biosynthesis. As the biosynthetic gene cluster had not been identified, the first objective was to clone the GE81112 biosynthetic gene cluster in order to allow studies of the underlying biosynthesis. Once the cluster was identified, it would then be annotated, as the first step towards deciphering the biosynthetic pathway. In parallel, we aimed to develop methods to genetically manipulate the strain, as they were not available. A final goal was to investigate GE biosynthesis by heterologous expression.
2 Material and Methods

2.1 Chemicals

All chemicals used in this work were obtained from the manufacturers, at analytical grade.

<table>
<thead>
<tr>
<th>Chemical product</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetone</td>
<td>Sigma Aldrich</td>
</tr>
<tr>
<td>Acetonitrile</td>
<td></td>
</tr>
<tr>
<td>Chloroform</td>
<td></td>
</tr>
<tr>
<td>Ethanol</td>
<td></td>
</tr>
<tr>
<td>Ethyl acetate</td>
<td></td>
</tr>
<tr>
<td>Hexane</td>
<td></td>
</tr>
<tr>
<td>Methanol</td>
<td></td>
</tr>
<tr>
<td>Magnesium sulfate</td>
<td></td>
</tr>
<tr>
<td>2-Propanol</td>
<td></td>
</tr>
<tr>
<td>Succinic acid</td>
<td></td>
</tr>
<tr>
<td>Bacto agar</td>
<td>Becton Dickinson and Co., USA</td>
</tr>
<tr>
<td>Casitone</td>
<td></td>
</tr>
<tr>
<td>Casaminoacids</td>
<td></td>
</tr>
<tr>
<td>Malt extract</td>
<td></td>
</tr>
<tr>
<td>Tryptone</td>
<td></td>
</tr>
<tr>
<td>Peptone</td>
<td></td>
</tr>
<tr>
<td>Yeast extract</td>
<td></td>
</tr>
<tr>
<td>Ammonium acetate</td>
<td>Fluka</td>
</tr>
<tr>
<td>Sodium-EDTA</td>
<td></td>
</tr>
<tr>
<td>Potassium hydroxide</td>
<td></td>
</tr>
<tr>
<td>Acetic acid</td>
<td>Merck</td>
</tr>
<tr>
<td>Boric acid</td>
<td></td>
</tr>
<tr>
<td>Bromphenolblue</td>
<td></td>
</tr>
<tr>
<td>Calcium chloride-di-hydrate</td>
<td></td>
</tr>
<tr>
<td>Glucose monohydrate</td>
<td></td>
</tr>
<tr>
<td>Glycerin</td>
<td></td>
</tr>
<tr>
<td>Magnesium chloride</td>
<td></td>
</tr>
<tr>
<td>Magnesium sulfate hepta-hydrate</td>
<td></td>
</tr>
<tr>
<td>Potassium acetate</td>
<td></td>
</tr>
<tr>
<td>Potassium hydrogen phosphate</td>
<td></td>
</tr>
<tr>
<td>Sodium chloride</td>
<td></td>
</tr>
<tr>
<td>Sodium citrate-di-hydrate</td>
<td></td>
</tr>
<tr>
<td>Sodium hydroxide</td>
<td></td>
</tr>
<tr>
<td>BSA (Bovine Serum Albumin)</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>Ammonium persulfate</td>
<td></td>
</tr>
<tr>
<td>5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside</td>
<td></td>
</tr>
<tr>
<td>Coomassie Brilliantblue 250R</td>
<td></td>
</tr>
<tr>
<td>Dithiothreitol</td>
<td></td>
</tr>
<tr>
<td>Ethidium bromide solution (1%)</td>
<td></td>
</tr>
<tr>
<td>Formamide</td>
<td></td>
</tr>
<tr>
<td>Maleic acid</td>
<td></td>
</tr>
<tr>
<td>Sodium dodecyl sulfate (SDS)</td>
<td></td>
</tr>
<tr>
<td>Rotiphorese®/Gel 30</td>
<td></td>
</tr>
<tr>
<td>TEMED</td>
<td></td>
</tr>
<tr>
<td>Triton X-100</td>
<td>ROTH</td>
</tr>
</tbody>
</table>
2.2 Enzymes, Kits and Markers

All enzymes used in this work are listed in Table 2.2 and were used according to the manufacturers’ protocols.

Table 2.2: Enzymes, kits and markers and their manufacturers

<table>
<thead>
<tr>
<th>Product</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page Ruler™ Prestained Protein Ladder</td>
<td>MBI Fermentas</td>
</tr>
<tr>
<td>Page Ruler™ Protein Ladder</td>
<td></td>
</tr>
<tr>
<td>1 kb DNA ladder</td>
<td></td>
</tr>
<tr>
<td>T4 DNA ligase</td>
<td></td>
</tr>
<tr>
<td>Shrimp alkaline phosphatase</td>
<td></td>
</tr>
<tr>
<td>Restriction endonucleases</td>
<td></td>
</tr>
<tr>
<td>Plasmid purification kit</td>
<td></td>
</tr>
<tr>
<td>Nucleotides</td>
<td></td>
</tr>
<tr>
<td>Cre recombinase</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>Phusion DNA Polymerase</td>
<td>Finzymes</td>
</tr>
<tr>
<td>Triple Master Polymerase</td>
<td>Eppendorf</td>
</tr>
<tr>
<td>HotStarTaq Ploymerase</td>
<td>Quiagen</td>
</tr>
<tr>
<td>Topo TA Cloning® Kit</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>Gel-Dry™ Drying Kit</td>
<td></td>
</tr>
<tr>
<td>Nucleospin® Extract</td>
<td>Macherey &amp; Nagel</td>
</tr>
<tr>
<td>Ribonuclease A</td>
<td>Roth</td>
</tr>
<tr>
<td>Proteinase K</td>
<td></td>
</tr>
<tr>
<td>Lysozyme</td>
<td></td>
</tr>
<tr>
<td>DIG labeled DNA Molecular Weight Marker III</td>
<td>Roche</td>
</tr>
<tr>
<td>DIG-Labeling and Detection Kit</td>
<td></td>
</tr>
<tr>
<td>DIG-Blocking solution</td>
<td></td>
</tr>
<tr>
<td>DIG EasyHyb</td>
<td></td>
</tr>
<tr>
<td>DIG PCR Labeling Mix</td>
<td></td>
</tr>
<tr>
<td>BugBuster® Protein Extraction Reagent</td>
<td>Novagen</td>
</tr>
<tr>
<td>Gigapack® III Gold Packaging Extract</td>
<td>Stratagene</td>
</tr>
</tbody>
</table>
### 2.3 Buffers and stock solutions

#### 2.3.1 Solutions and buffers for molecular biology applications

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Constituents</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Lysis buffer (P1)</td>
<td>NaOH</td>
<td>20 mM</td>
</tr>
<tr>
<td></td>
<td>10% SDS</td>
<td>10 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 100 ml</td>
</tr>
<tr>
<td>Cell suspension buffer (P2)</td>
<td>Glucose</td>
<td>5 mM</td>
</tr>
<tr>
<td></td>
<td>Tris</td>
<td>2.5 mM</td>
</tr>
<tr>
<td></td>
<td>EDTA</td>
<td>1 mM</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 100 ml</td>
</tr>
<tr>
<td>DNA Loading buffer (6×)</td>
<td>Glycerin (87%)</td>
<td>3 ml</td>
</tr>
<tr>
<td></td>
<td>Bromphenol blue</td>
<td>25 mg</td>
</tr>
<tr>
<td></td>
<td>Xylene cyanol</td>
<td>25 mg</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 10 ml</td>
</tr>
<tr>
<td>0.5 M EDTA</td>
<td>EDTA</td>
<td>186 g</td>
</tr>
<tr>
<td></td>
<td>NaOH</td>
<td>20 g</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 1L</td>
</tr>
<tr>
<td>Glycerol (50%)</td>
<td>Glycerol</td>
<td>115 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 1L</td>
</tr>
<tr>
<td>Glycerol (20%)</td>
<td>Glycerol</td>
<td>57.5 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 100 ml</td>
</tr>
<tr>
<td>0.1 M IPTG</td>
<td>IPTG</td>
<td>240 mg</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 10 ml</td>
</tr>
<tr>
<td>5 M Sodium chloride</td>
<td>NaCl</td>
<td>580 g</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 100 ml</td>
</tr>
<tr>
<td>Neutralisation buffer (P 3)</td>
<td>3 M Potassium acetate</td>
<td>60 ml</td>
</tr>
<tr>
<td></td>
<td>Acetic acid</td>
<td>11.5 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 100 ml</td>
</tr>
<tr>
<td>10% SDS solution</td>
<td>SDS</td>
<td>10 g</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to100 ml</td>
</tr>
<tr>
<td>SET buffer</td>
<td>Tris-HCl (pH 7.5)</td>
<td>20 mM</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>75 mM</td>
</tr>
<tr>
<td></td>
<td>EDTA (pH 8)</td>
<td>25 mM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>To 100 ml</td>
</tr>
</tbody>
</table>
### Buffer Constituents

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Constituents</th>
<th>Amounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBE buffer (1x)</td>
<td>TRIZMA base</td>
<td>242 g</td>
</tr>
<tr>
<td></td>
<td>Boric acid</td>
<td>57.1 ml</td>
</tr>
<tr>
<td></td>
<td>Na-EDTA</td>
<td>100 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 1L</td>
</tr>
<tr>
<td>SM-buffer</td>
<td>Tris-HCl (1 M, pH 7.4)</td>
<td>50 ml</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>5.4 g</td>
</tr>
<tr>
<td></td>
<td>MgSO₄.₇ H₂O</td>
<td>2 g</td>
</tr>
<tr>
<td></td>
<td>Gelatine-solution (2%)</td>
<td>5 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 1 L</td>
</tr>
<tr>
<td>1 M Tris-HCl</td>
<td>TRIZMA Base</td>
<td>60.5 g</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 500 ml</td>
</tr>
<tr>
<td>P (protoplast) buffer</td>
<td>Sucrose</td>
<td>103 g</td>
</tr>
<tr>
<td></td>
<td>K₂SO₄</td>
<td>0.25 g</td>
</tr>
<tr>
<td></td>
<td>MgCl₂.₆ H₂O</td>
<td>2.02 g</td>
</tr>
<tr>
<td></td>
<td>Trace element solution</td>
<td>2 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 800 ml</td>
</tr>
<tr>
<td></td>
<td>Add after autoclaving:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>KH₂PO₄ (0.5%)</td>
<td>1 ml</td>
</tr>
<tr>
<td></td>
<td>CaCl₂.2H₂O (3.68%)</td>
<td>10 ml</td>
</tr>
<tr>
<td></td>
<td>TES buffer (5.73%, pH 7.2)</td>
<td>10 ml</td>
</tr>
</tbody>
</table>

### Solutions and buffers for Southern Blot and colony hybridization

#### Table 2.4: Solution and buffers for Southern Blot and colony hybridization

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Constituents</th>
<th>Amounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depurination buffer</td>
<td>HCl (1 M)</td>
<td>250 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 1L</td>
</tr>
<tr>
<td>Denaturation buffer</td>
<td>NaOH</td>
<td>20 g</td>
</tr>
<tr>
<td></td>
<td>NaCl (5 M)</td>
<td>333 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 1 L</td>
</tr>
<tr>
<td>Neutralisation buffer</td>
<td>Tris HCl (1 M, pH 7.4)</td>
<td>500 ml</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>175.5 g</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 1 L</td>
</tr>
<tr>
<td>SSC (20×)</td>
<td>NaCl</td>
<td>155.3 g</td>
</tr>
<tr>
<td></td>
<td>Sodium citrate.2 H₂O</td>
<td>88.2 g</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 1 L</td>
</tr>
</tbody>
</table>
### Buffer Constituents

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Constituents</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybridization buffer</td>
<td>DIG Easy Hyb Granules, H₂O</td>
<td>to 64 ml</td>
</tr>
<tr>
<td>2× wash solution</td>
<td>SSC (20×), 10% SDS, H₂O</td>
<td>100 ml, 10 ml, to 1L</td>
</tr>
<tr>
<td>1× wash solution</td>
<td>SSC (20×)</td>
<td>50 ml, 10 ml, to 1L</td>
</tr>
<tr>
<td>Maleic acid buffer</td>
<td>Maleic acid, NaCl, H₂O</td>
<td>11.61 g, 8.8 g, to 1L</td>
</tr>
<tr>
<td>Blocking solution</td>
<td>DIG blocking reagent, Maleic acid buffer</td>
<td>50 g, to 500 ml</td>
</tr>
<tr>
<td>Detection buffer</td>
<td>Tris-HCl (1 M), NaCl, H₂O</td>
<td>100 ml, 5.8g, to 1L</td>
</tr>
<tr>
<td>Stripping buffer</td>
<td>NaOH, SDS (10%), H₂O</td>
<td>8 g, 10 ml, to 1L</td>
</tr>
</tbody>
</table>

### 2.3.3 Solutions and buffers for enzymatic applications

#### Table 2.5: Solution and buffers for enzymatic applications

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Constituents</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x buffer</td>
<td>Tris-HCl, NaCl, MgCl₂, H₂O</td>
<td>750 mM, 1 M, 100 mM, to 50 ml</td>
</tr>
<tr>
<td>Binding buffer</td>
<td>Tris-HCl (pH 7.8), NaCl, Glycerol, Imidazole, H₂O</td>
<td>20 mM, 200 mM, 100 ml, 60 mM, to 1L</td>
</tr>
<tr>
<td>Borate buffer</td>
<td>Boric acid, H₂O</td>
<td>200 mM, to 1L</td>
</tr>
<tr>
<td>Buffer</td>
<td>Constituents</td>
<td>Concentration</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Elution buffer</td>
<td>Tris-HCl (pH 7.8)</td>
<td>20 mM</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>200 mM</td>
</tr>
<tr>
<td></td>
<td>Glycerol</td>
<td>100 ml</td>
</tr>
<tr>
<td></td>
<td>Imidazole</td>
<td>500 mM</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 1 L</td>
</tr>
<tr>
<td>Stop-mix</td>
<td>Activated charcoal</td>
<td>6 g</td>
</tr>
<tr>
<td></td>
<td>Tetrasodium pyrophosphate</td>
<td>50 ml</td>
</tr>
<tr>
<td></td>
<td>Perchloric acid</td>
<td>19.2 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 500 ml</td>
</tr>
<tr>
<td>Stripping buffer</td>
<td>Sodium phosphate (pH 7.4)</td>
<td>20 mM</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>500 mM</td>
</tr>
<tr>
<td></td>
<td>EDTA</td>
<td>50 mM</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 1 L</td>
</tr>
<tr>
<td>Nickel solution</td>
<td>NiSO₄</td>
<td>100 mM</td>
</tr>
<tr>
<td>Coomassie Brilliant Blue</td>
<td>Brilliant Blue 250</td>
<td>2 g</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>450 ml</td>
</tr>
<tr>
<td></td>
<td>Methanol</td>
<td>450 ml</td>
</tr>
<tr>
<td></td>
<td>Acetic Acid</td>
<td>100 ml</td>
</tr>
<tr>
<td>4× Protein loading buffer</td>
<td>Tris (1 M, pH 6.8)</td>
<td>5 ml</td>
</tr>
<tr>
<td></td>
<td>87% Glycerol</td>
<td>4 ml</td>
</tr>
<tr>
<td></td>
<td>Bromophenol blue</td>
<td>10 mg</td>
</tr>
<tr>
<td></td>
<td>SDS</td>
<td>1 g</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 10 ml</td>
</tr>
<tr>
<td></td>
<td>DTT 500 mM</td>
<td>800 μl</td>
</tr>
<tr>
<td></td>
<td>The DTT is added directly before use.</td>
<td></td>
</tr>
<tr>
<td>PBS buffer</td>
<td>1 PBS tablet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 500 ml</td>
</tr>
<tr>
<td>Wash buffer</td>
<td>Tetrasodium pyrophosphate</td>
<td>50 ml</td>
</tr>
<tr>
<td></td>
<td>Perchloric acid</td>
<td>19.2 ml</td>
</tr>
<tr>
<td></td>
<td>H₂O</td>
<td>to 500 ml</td>
</tr>
</tbody>
</table>
2.4 Media

2.4.1 Media for *E. coli* cultivation

Table 2.6: Media for *E. coli* cultivation (for agar plates 16 g/L agar were added to the medium)

<table>
<thead>
<tr>
<th>Medium</th>
<th>Constituents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LB medium</td>
<td>Tryptone</td>
<td>10 g</td>
</tr>
<tr>
<td></td>
<td>Yeast extract</td>
<td>5 g</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>5 g</td>
</tr>
<tr>
<td></td>
<td>H$_2$O</td>
<td>to 1 L</td>
</tr>
<tr>
<td>2YT medium</td>
<td>Tryptone</td>
<td>16 g</td>
</tr>
<tr>
<td></td>
<td>Yeast extract</td>
<td>10 g</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>5 g</td>
</tr>
<tr>
<td></td>
<td>H$_2$O</td>
<td>to 1 L</td>
</tr>
<tr>
<td>Freezing solution</td>
<td>MgSO$_4$ x 7 H$_2$O</td>
<td>760 mg</td>
</tr>
<tr>
<td></td>
<td>Na-Citrat x 2 H$_2$O</td>
<td>4.5 g</td>
</tr>
<tr>
<td></td>
<td>NH$_4$SO$_4$</td>
<td>9 g</td>
</tr>
<tr>
<td></td>
<td>Glycerol</td>
<td>440 g</td>
</tr>
<tr>
<td></td>
<td>H$_2$O</td>
<td>to 800 ml</td>
</tr>
<tr>
<td></td>
<td>Add after autoclaving:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K$_2$HPO$_4$</td>
<td>47 g</td>
</tr>
<tr>
<td></td>
<td>KH$_2$PO$_4$</td>
<td>18 g</td>
</tr>
<tr>
<td></td>
<td>H$_2$O</td>
<td>to 200 ml</td>
</tr>
</tbody>
</table>

2.4.2 Media for cultivation of pseudomonads

Table 2.7: Media for cultivation of pseudomonads

<table>
<thead>
<tr>
<th>Medium</th>
<th>Constituents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>LB medium</td>
<td>Tryptone</td>
<td>10 g</td>
</tr>
<tr>
<td></td>
<td>Yeast extract</td>
<td>5 g</td>
</tr>
<tr>
<td></td>
<td>NaCl</td>
<td>5 g</td>
</tr>
<tr>
<td></td>
<td>H$_2$O</td>
<td>to 1 L</td>
</tr>
<tr>
<td>PMC medium</td>
<td>K$_2$HPO$_4$</td>
<td>6 g</td>
</tr>
<tr>
<td></td>
<td>KH$_2$PO$_4$</td>
<td>5 g</td>
</tr>
<tr>
<td></td>
<td>NH$_4$(SO$_4$)$_2$</td>
<td>1 g</td>
</tr>
<tr>
<td></td>
<td>Na-Citrate</td>
<td>8.8 g</td>
</tr>
<tr>
<td></td>
<td>MgSO$_4$</td>
<td>800 mM</td>
</tr>
<tr>
<td></td>
<td>H$_2$O</td>
<td>to 1 L</td>
</tr>
</tbody>
</table>
### Media for cultivation of streptomycetes

**Table 2.8: Media for cultivation of streptomycetes**

<table>
<thead>
<tr>
<th>Medium</th>
<th>Constituents</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptone soy broth medium</td>
<td>Tryptone soy broth, H₂O</td>
<td>30 g to 1L</td>
</tr>
<tr>
<td>(TSB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeast extract-malt extract medium</td>
<td>Yeast extract, Peptone, Malt extract, Glucose, Sucrose, H₂O</td>
<td>3 g, 5 g, 3 g, 10 g, 340 g to 1L</td>
</tr>
<tr>
<td>(YEME)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soya mannitol medium</td>
<td>Mannitol, Soya flour, H₂O</td>
<td>20 g, 20 g, 20 g to 1L</td>
</tr>
<tr>
<td>(SM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INA5 medium</td>
<td>Glycerol, Soya flour, CaCO₃, NaCl, H₂O</td>
<td>30 g, 15 g, 5 g, 2 g, 2 g to 1L</td>
</tr>
<tr>
<td>T6 medium</td>
<td>Glycerol, Soya flour, CaCO₃, H₂O</td>
<td>45 g, 25 g, 2 g to 1 L</td>
</tr>
<tr>
<td>R2YE medium</td>
<td>Sucrose, K₂SO₄, MgCl₂·6 H₂O, Glucose, Casaminoacids, H₂O</td>
<td>51.1 g, 0.125 g, 5.06 g, 5 g, 0.05 g to 400 ml</td>
</tr>
<tr>
<td></td>
<td>Add after autoclaving:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>KH₂PO₄ (0.5%)</td>
<td>5 ml</td>
</tr>
<tr>
<td></td>
<td>CaCl₂·2H₂O (3.68%)</td>
<td>40 ml</td>
</tr>
<tr>
<td></td>
<td>L-Proline (2 %)</td>
<td>7.5 ml</td>
</tr>
<tr>
<td></td>
<td>TES-Puffer (5.73 %, pH 7.2)</td>
<td>50 ml</td>
</tr>
<tr>
<td></td>
<td>Trace element solution [10]</td>
<td>1 ml</td>
</tr>
<tr>
<td></td>
<td>1 N NaOH</td>
<td>2.5 ml</td>
</tr>
<tr>
<td></td>
<td>Yeast extract (10%)</td>
<td>25 ml</td>
</tr>
</tbody>
</table>
### Material and Methods

#### Medium and Constituents

<table>
<thead>
<tr>
<th>Medium</th>
<th>Constituents</th>
<th>Amounts</th>
</tr>
</thead>
<tbody>
<tr>
<td>V6 medium</td>
<td>Glucose, Meat extract, Yeast extract, Peptone, Casein, NaCl, pH 7.5</td>
<td>20 g, 5 g, 5 g, 5 g, 3 g, 1.5 g to 1L</td>
</tr>
<tr>
<td>NBG</td>
<td>Peptone, Meat extract, Glucose, NaCl, H₂O</td>
<td>10 g, 3 g, 10 g, 5 g to 1L</td>
</tr>
<tr>
<td>NL111</td>
<td>Meat extract, CaCO₃, Malt extract, H₂O, pH 7.2</td>
<td>20 g, 10 g, 100 g to 1L</td>
</tr>
</tbody>
</table>

#### Antibiotics

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Stock solution</th>
<th>End concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>100 mg/ml</td>
<td>100 µg/ml</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>34 mg/ml</td>
<td>34 µg/ml</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>60 mg/ml</td>
<td>60 µg/ml</td>
</tr>
<tr>
<td>Tetracyline</td>
<td>12.5 mg/ml</td>
<td>12.5 µg/ml</td>
</tr>
<tr>
<td>Zeocin</td>
<td>100 mg/ml</td>
<td>25 µg/ml</td>
</tr>
<tr>
<td>Apramycin</td>
<td>60 mg/ml</td>
<td>60 µg/ml</td>
</tr>
<tr>
<td>Nalidixic acid</td>
<td>25 mg/ml</td>
<td>25 µg/ml</td>
</tr>
</tbody>
</table>
2.6 Instruments and materials

- **Agarose gel electrophoresis**
  - Electrophoresis chamber SUB-CELL® GT (Biorad)
  - Electrophoresis chamber MINI-SUB-CELL® GT (Biorad)

- **Centrifuges**
  - Cool centrifuge 5805R (Eppendorf)
  - Tabletop centrifuge 5415D (Eppendorf)
  - Centrifuge Avanti JE (Beckmann Coulter)
  - Biofuge Pico (Heraeus)

- **Electroporator**
  - Gene pulser XCell (Biorad)
  - Electroporation cuvette 0.1 cm (Biorad)

- **Electro blotter**
  - Trans-Blot SD Semi Dry Transfer Cell (Biorad)

- **French Press** (SLM Aminco)

- **HPLC**
  - DAD-coupled HPLC (Dionex)

- **HPLC-MS**
  - Agilent 110 series HPLC system (Agilent)
  - Bruker HCTplus (Bruker)
  - Bruker micrOTOF (Bruker)
  - LTQ-Orbitrap (Thermo Finnigan)

- **Image documentation**
  - N-1000 Darkroom (Peqlab)
  - Camera biovision (Peqlab)

- **Incubators**
  - Incubators (Binder)
  - Hybridization oven APT Line Series BFED (Binder)
  - Multitron Shakers (Infors)

- **Laminar flow**
  - HeraeusLaminAir® (Kendro)
  - Hera Safe (Kendro)
• NMR
  Bruker Advance 500 (Bruker)

• PCR
  Mastercycler Gradient (Eppendorf)
  PCR Sprint Thermocycler (Thermo Electron Corp)

• pH-measurements
  pH-Meter 766 Calimatic (Knick)

• Photometer
  Helios Epsilon Spectrophotometer (Thermo)

• Protein purification
  Äkta Prime system (GEHealthcare)

• TLC
  Alugram®SL G/UV<sub>254</sub> TLC Platten (Macherey & Nagel)

• Thermomixer
  Thermomixer compact (Eppendorf)
  Thermomixer comfort (Eppendorf)

• Sterilisation
  Systec VX-150 (Systec)

• Ultrasonic bath
  Sonarex (Bandelin)

• Water processing
  Milli-Q water purification system (Millipore)
  PURELAB ultra (ELGA)

• Others materials
  Membrane filter 0.22 µm (Millipore)
  Nylon membrane, positively charged (Roche)
  Microspin Columns (Amersham Pharmacia)
### 2.7 Bacterial strains, oligonucleotides and plasmids

#### 2.7.1 Bacterial strains

**Table 2.10: Bacterial strains used in this work**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype/characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E.coli</em> DH10B</td>
<td>F-mcrAΔ(mrr-hdsRMS-mcrBC) Ø80dlacZΔM15ΔlacX74 deoR recA1 endA1 araD 139Δ (ara,leu) 7696 galU galK rpsL nupG</td>
<td>[68]</td>
</tr>
<tr>
<td><em>E. coli</em> HS 996</td>
<td>F-mcrAΔ(mrr-hdsRMS-mcrBC) Ø80dlacZΔM15ΔlacX74 deoR recA1 endA1 araD 139Δ (ara,leu) 7696 galU galK rpsL nupG flnuA::IS</td>
<td>Invitrogen</td>
</tr>
<tr>
<td><em>E. coli</em> ET 12567/pUZ8002</td>
<td>dam13::Tn9 dcm-6 hsdM hsdR recF143 zjj-201::Tn10 galK2 galT22 ara14 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tss-78 mtlI glnV44, carrying the oriT mobilising plasmid pUZ8002</td>
<td>[10]</td>
</tr>
<tr>
<td><em>E. coli</em> GB2005</td>
<td>HS996 ΔrecT and redα</td>
<td>Genebridges</td>
</tr>
<tr>
<td><em>E. coli</em> GB2005 red</td>
<td>GB2005 insertion of the pBAD recE recT γ recA operon</td>
<td>Genebridges</td>
</tr>
<tr>
<td><em>E. coli</em> SCS110</td>
<td>rpsL (Strr) thr leu endA thi-1 lacY galK galT ara tonA tss dam dcm supE44 Δ(lac-proAB) [F’ traD36 proAB lacIqZM15]</td>
<td>Stratagene</td>
</tr>
<tr>
<td><em>E. coli</em> BL21(DE3)</td>
<td>F’ ompT gal dcm lon hsdS&lt;sub&gt;r&lt;/sub&gt; galK galT hsdS&lt;sub&gt;q&lt;/sub&gt; (r&lt;sub&gt;q&lt;/sub&gt; m&lt;sub&gt;q&lt;/sub&gt;) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])</td>
<td>[69]</td>
</tr>
<tr>
<td><em>E. coli</em> Rosetta BL21 (DE3)</td>
<td>F’ ompT hsd S&lt;sub&gt;r&lt;/sub&gt; (R&lt;sub&gt;r&lt;/sub&gt; m&lt;sub&gt;r&lt;/sub&gt;) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (Cm&lt;sup&gt;R&lt;/sup&gt;)</td>
<td>Novagen</td>
</tr>
<tr>
<td><em>E.coli</em> SURE</td>
<td>endA1 glnV44 thi-1 gyrA96 relA1 lac recB recI sbcC umuC::Tn5 uvrC e14 Δ(mcrCB-hsdSMR-mrr)171 F’[proAB’ lacI lacZM15 Tn10(tet&lt;sup&gt;R&lt;/sup&gt;)]</td>
<td>Stratagene</td>
</tr>
<tr>
<td><em>P. putida</em> K72440</td>
<td>wild type</td>
<td>[71]</td>
</tr>
<tr>
<td><em>S. lividans</em> TK24</td>
<td>Derived from <em>S. lividans</em> TK66</td>
<td>[72]</td>
</tr>
<tr>
<td><em>S. coelicolor</em> A3(2)</td>
<td>wild type</td>
<td>[73]</td>
</tr>
<tr>
<td><em>S. coelicolor</em> M512</td>
<td>ΔredD, ΔactI-ORF4 SCP1 SCP2&lt;sup&gt;+&lt;/sup&gt;</td>
<td>[74]</td>
</tr>
<tr>
<td><em>S. Tü6071</em></td>
<td>phenalinolactone producer</td>
<td>[66]</td>
</tr>
<tr>
<td><em>S. 14386</em></td>
<td>GE81112 producer</td>
<td>[75]</td>
</tr>
</tbody>
</table>
### 2.7.2 Oligonucleotides

#### 2.7.2.1 Primers used in the phenalinolactone project

Table 2.11: Primers used in the phenalinolactone project, DST= desalted

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5′→3′)</th>
<th>Purification</th>
</tr>
</thead>
<tbody>
<tr>
<td>PhlET6</td>
<td>CGAGCCGGGGATGCGCACAATTCGCCCCTTTCTTTCTTCAA GGCGCAATCTCGTGATCAGCAGTCGTT</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET7</td>
<td>GTAGAGGATCTCTAGATATAATTGGGAGTGATTTCCTTTGTT TAAAGGATCTCCTCGCTCTCGGCCCA</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET10</td>
<td>GTAGAGGATCTCTAGATATAATTGGGAGTGATTTCCTTTGTT TAAAGGATCAGTACTCCACTAGTGTGAATGTAATACGAC TCACCATAGGGGAGTCTCCTCGGAGAGCAG</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET11</td>
<td>GGGCGCTCTGGTTGTGGAGAGCGAGTTCGCTGACCAGTGCGC GACGACGTACTTTACTGCGGAAGTCTCGCT</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET18</td>
<td>CAGCAGTGGTACGAGGCGCTGTCGCTCCGCCGCAATGCG GTGACGCTGACGTCGCTGTAACCTGTA</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET19</td>
<td>CCGGGGATGGGATGCTGATACGCGCTTTGCTGCTCGCCGTCT CGGCATCGACGTGTTGCGTGATACCTCGA</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET20</td>
<td>CCTCGGATCTGCTCGCTGAGGCGCCTCCGACACGCGCTTGGAAAGCCTCTTAATTCTCGGAGGAGTC</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET21</td>
<td>CGAATGGGCGCCAGTTCGTTGGCTGTTCCGCTCGGTAGGCTG TCAGGCGTCGTCGTTGAC</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET22</td>
<td>GAGGCCAGCTGGGCGCCACGCAGTCGTGCTGAGGAGA AATGGTGCAGGGGAAAGTTCTCTTAATCTAGAAGATAGGAA CTTCTCGAGCGACGACGTGTTGAC</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET23</td>
<td>TCCGGCTGACAGCGCGCTCTGCCTCGGCTGCGCGGACGCCTCT CGCTCTCTCTGTAAGTCTATGAGGAAATAGGAC TTACAGGGCGTCCGTACCGCCG</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET28</td>
<td>CGGCCCAAGAAGTGAGGACTGCGCCCGCAGAGGAGG TGATTTAAATCCCTGTGTCGTCCCTG</td>
<td>HPLC</td>
</tr>
<tr>
<td>PhlET29</td>
<td>AGACGCCCAAGAAGTGACGTGGCTGCTGAAATTCTCCGCTGACCAGGCTACTTCCCTACCAACCGACGATTTGT GCCACAAAGCGACATTCCGCCCCGCTCCGACCTC</td>
<td>HPLC</td>
</tr>
<tr>
<td>Apra_for</td>
<td>GTGCAATACGAATGGGCGAAAA</td>
<td>DST</td>
</tr>
<tr>
<td>Apra_rev</td>
<td>TCGCCAAATCGACTGGCGAG</td>
<td>DST</td>
</tr>
<tr>
<td>ermE</td>
<td>GCCAGGCTTCTGGCCCATCT</td>
<td>DST</td>
</tr>
<tr>
<td>CPhl19_for</td>
<td>GCCAGGCTTCTGGCCCATCT</td>
<td>DST</td>
</tr>
<tr>
<td>ermE</td>
<td>ACGTGCGCGCTCCGCGCCGGA</td>
<td>DST</td>
</tr>
<tr>
<td>CPhl19_rev</td>
<td>ACGTGCGCGCTCCGCGCCGGA</td>
<td>DST</td>
</tr>
</tbody>
</table>
### 2.7.2.2 Primers used in the GE81112 project

**Table 2.12: Primers used in the GE81112 project**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5'→3')</th>
<th>Purification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyclo_14386_for</td>
<td>AGCGCCAGGAGAGACCCCTTG</td>
<td>DST</td>
</tr>
<tr>
<td>Cyclo_14386_rev</td>
<td>GCCTGCAACTGGGTGACGCG</td>
<td>DST</td>
</tr>
<tr>
<td>Cyclo Probe_for</td>
<td>CCCCGGGCCCGGAACGCATG</td>
<td>DST</td>
</tr>
<tr>
<td>Cyclo Probe_rev</td>
<td>GAACGCCATGACCATGCGG</td>
<td>DST</td>
</tr>
<tr>
<td>Tubz_up</td>
<td>GGAAGTGGCCGTCAGAAGGGATGTC</td>
<td>DST</td>
</tr>
<tr>
<td>TubZ_down</td>
<td>CTGCACGCGCTGATGGTGAGGTGTC</td>
<td>DST</td>
</tr>
<tr>
<td>RapL_up</td>
<td>GCGCGCTCGGCGATCGGTG</td>
<td>DST</td>
</tr>
<tr>
<td>RapL_down</td>
<td>CTTGGGCGAGCTCGTGCTTT</td>
<td>DST</td>
</tr>
<tr>
<td>CycloKO_for</td>
<td>ACTGCAAGATCTGCGAAGCGCCGCCGCCGCG</td>
<td>DST</td>
</tr>
<tr>
<td>CycloKO_rev</td>
<td>TACTGAGAATTTCGGTGCTACGCGCGGGCT</td>
<td>DST</td>
</tr>
<tr>
<td>NRPS-A1-up</td>
<td>CGGCTCAACCGACACCGCAARGGNGG</td>
<td>DST</td>
</tr>
<tr>
<td>NRPS-H1-dn</td>
<td>CGGCCGAGGAGCGGCGGACCTCC</td>
<td>DST</td>
</tr>
<tr>
<td>RevA3</td>
<td>CTCGCCGSCSACCCGCGCCCAGCGGGGCCAAGGG</td>
<td>DST</td>
</tr>
<tr>
<td>PSLGG 2000</td>
<td>GCCGCGTGGGCTACCGAGGCGAGGCCG</td>
<td>DST</td>
</tr>
<tr>
<td>BI11-T7end-2-for</td>
<td>TGGAGCGTCAGGCGCATCCC</td>
<td>DST</td>
</tr>
<tr>
<td>BI11-T7end-2-rev</td>
<td>GGACGGCTTTCACCGCGG</td>
<td>DST</td>
</tr>
<tr>
<td>pGEX_Cyclo_for_EcoRI</td>
<td>CTTTGGGAAATTCACAGGCGGCTTAACACGCGG</td>
<td>DST</td>
</tr>
<tr>
<td>pGEX_Cyclo_for_NotI</td>
<td>GCCCGTGCGCCGCACCGCGGTCCCGGCTTCC</td>
<td>DST</td>
</tr>
<tr>
<td>pET28_Cyclo_for_NdeI</td>
<td>GCCCGCATATGACAAAACCGCCACTGAGGCG</td>
<td>DST</td>
</tr>
<tr>
<td>pET28_Cyclo_for_BamHI</td>
<td>GCCCGTGAGATCTCAAGGCGCGCAGCTC</td>
<td>DST</td>
</tr>
<tr>
<td>BI11A1_pip_for</td>
<td>CGCGGGATACCCCTACCGCGAGGCGGTCCCGT</td>
<td>DST</td>
</tr>
<tr>
<td>BI11A1_pip_rev</td>
<td>GCCGCCATATGCTCGTCCCGGCAACGGGCGGCG</td>
<td>DST</td>
</tr>
<tr>
<td>BI11A2_for</td>
<td>GCCGCCATATGCTCGTCCCGGCAACGGGCGGCG</td>
<td>DST</td>
</tr>
<tr>
<td>BI11A2_rev</td>
<td>GGTACGGATCCGCCGCCGCCGCCGCCGCCGCC</td>
<td>DST</td>
</tr>
<tr>
<td>BI11A3_for</td>
<td>CGGACCATTAGCCCTCGGCCGCGCGCGAGGCG</td>
<td>DST</td>
</tr>
<tr>
<td>BI11A3_rev</td>
<td>ACGCGGATACCCACGATCCAGACCCGCCGCCAG</td>
<td>DST</td>
</tr>
<tr>
<td>FD10_KO_for</td>
<td>GCCAGTCCAGGACGCGACGCGAGGAGGACGCGGCG</td>
<td>HPLC</td>
</tr>
<tr>
<td>FD10_KO_rev</td>
<td>GCCGTATGTGATTTATACGTCGGGTGCACGCGG</td>
<td>HPLC</td>
</tr>
<tr>
<td>Sonde BI11 SEQ1</td>
<td>CATTGTTGAGACTGCTGCTCTTTT</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ2</td>
<td>CGAGTTCGCCACCTCCAAGGCGG</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ3</td>
<td>GCTGGACGCGGACCTGAGGCGG</td>
<td>DST</td>
</tr>
</tbody>
</table>
### Material and Methods

#### Primer Sequence (5’→3’)

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’→3’)</th>
<th>Purification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sonde BI11 SEQ4</td>
<td>GCGCCCTGATTCCTCCACGG</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ5</td>
<td>TGGGGCAGTGCGTGCAGCATG</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ6</td>
<td>GCCGTCGCACGGAGACCCTTG</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ7</td>
<td>CATACGCTCCCGGGTCGTGGT</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ7_rev</td>
<td>GTGATCGCAGGGCGGCTGA</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ6_rev</td>
<td>GCCGTCGCCCGGGCCCGCC</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ5_rev</td>
<td>AGTTCGTGCAGGGCGGCGCC</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ4_rev</td>
<td>TCACTGGCGGGCCACCGAGG</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ 3_rev</td>
<td>CTCACGACTCAGTATAGGCG</td>
<td>DST</td>
</tr>
<tr>
<td>Sonde BI11 SEQ 2_rev</td>
<td>GGCGCGGGCGTGCTCGCCGA</td>
<td>DST</td>
</tr>
<tr>
<td>T7</td>
<td>TAATACGACTCAGTATAGGCG</td>
<td>DST</td>
</tr>
<tr>
<td>T4</td>
<td>CACCGTGGGCGGCGGTGATG</td>
<td>DST</td>
</tr>
<tr>
<td>LacZ1</td>
<td>CTTGGGCTGAGGTGCAC</td>
<td>DST</td>
</tr>
<tr>
<td>LacZ2</td>
<td>GTGTGGATATGAGCGGG</td>
<td>DST</td>
</tr>
</tbody>
</table>

#### 2.7.3 Plasmids

##### 2.7.3.1 General plasmids

Table 2.13: General plasmids used in this work

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Relevant characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSC101-BAD-γβα-Apra</td>
<td>ET cloning vector</td>
<td>Genebridges</td>
</tr>
<tr>
<td>pSC101-BAD-γβα-Amp</td>
<td>ET cloning vector</td>
<td>Genebridges</td>
</tr>
<tr>
<td>pCR2.1-TOPO</td>
<td>Cloning vector, lacZa, t7, f1 ori, neoR, bla, pUC origin</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>pKC1132</td>
<td>Knockout vector for Streptomyces, oriT RK2, lacZa, aac(3)IV</td>
<td>[10]</td>
</tr>
<tr>
<td>pKC1139</td>
<td>Knockout vector for Streptomyces, oriT RK2 lacZa, aac(3)IV, ori psG5</td>
<td>[10]</td>
</tr>
<tr>
<td>pRK2013</td>
<td>Tn903, RK2</td>
<td>[76]</td>
</tr>
<tr>
<td>pGEX-6-P-1</td>
<td>Expression vector, for expression with N-terminal GST Tag</td>
<td>GE Healthcare</td>
</tr>
<tr>
<td>pET28b(+)</td>
<td>Expression vector, for expression with N-terminal His Tag</td>
<td>Novagen</td>
</tr>
<tr>
<td>pOJ436</td>
<td>Cosmid vector, oriT RK2 aac(3)IV, λcos attP ØC31</td>
<td>[10]</td>
</tr>
</tbody>
</table>
### 2.7.3.2 Plasmids used in the phenalinolactone project

Table 2.14: Plasmids used in the phenalinolactone project

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Vector</th>
<th>Relevant characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPhl7</td>
<td>pOJ436</td>
<td>ET cloning product, Cos 3-1O12 + zeocin resistance gene, apra&lt;sup&gt;R&lt;/sup&gt;, zeo&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>CPhl8</td>
<td>pOJ436</td>
<td>ET cloning product, complete PL biosynthetic gene cluster containing the zeocin resistance gene and the oriT-tet-trpE cassette, apra&lt;sup&gt;R&lt;/sup&gt;, tet&lt;sup&gt;R&lt;/sup&gt;, zeo&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>CPhl9</td>
<td>pOJ436</td>
<td>ET cloning product, Cos 10-4D08 + oriT-tet-trpE cassette, tet&lt;sup&gt;R&lt;/sup&gt;, apra&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>CPhl10</td>
<td>pOJ436</td>
<td>ET cloning product, CPhl8 + P&lt;sub&gt;mr&lt;/sub&gt;-promotor in front of the gene plaM2, tet&lt;sup&gt;R&lt;/sup&gt;, apra&lt;sup&gt;R&lt;/sup&gt;,cmR</td>
</tr>
<tr>
<td>CPhl13</td>
<td>pOJ436</td>
<td>ET cloning product, complete PL biosynthetic gene cluster containing loxP-kan-cassette, apraR, tetR, zero, kanR</td>
</tr>
<tr>
<td>CPhl14</td>
<td>pOJ436</td>
<td>ET cloning product, complete PL biosynthetic gene cluster, knockout of plaP5, apraR, tetR, zero, kanR</td>
</tr>
<tr>
<td>CPhl15</td>
<td>pOJ436</td>
<td>ET cloning product, complete Pf biosynthetic gene cluster containing loxP-kan-cassette, apraR, tetR, zero, kanR</td>
</tr>
<tr>
<td>CPhl16</td>
<td>pOJ436</td>
<td>ET cloning product, complete PL biosynthetic gene cluster, knockout of plaP2, apraR, tetR, zero, kanR</td>
</tr>
<tr>
<td>CPhl18</td>
<td>pOJ436</td>
<td>ET cloning product, complete PL biosynthetic gene cluster, knockout of plaA6, apraR, tetR, zero, kanR</td>
</tr>
<tr>
<td>CPhl19</td>
<td>pOJ436</td>
<td>ET cloning product, complete PL biosynthetic gene cluster containing bidirectional ermE promoter in front of the genes plaM2 and plaO2</td>
</tr>
<tr>
<td>Cos3-1O12</td>
<td>pOJ436</td>
<td>Cosmid containing part of the PL biosynthetic gene cluster, apra&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cos10-4D08</td>
<td>pOJ436</td>
<td>Cosmid containing part of the PL biosynthetic gene cluster, apra&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
### 2.7.3.3 Plasmids used in the GE81112 project

**Table 2.15: Plasmids used in the GE81112 project**

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Vector</th>
<th>Relevant characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cos FD10</td>
<td>pOJ436</td>
<td>Cosmid from <em>Streptomyces</em> 14386 cosmid library, pOJ436 backbone, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cos AI6</td>
<td>pOJ436</td>
<td>Cosmid from <em>Streptomyces</em> 14386 cosmid library, pOJ436 backbone, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cos BI11</td>
<td>pOJ436</td>
<td>Cosmid from <em>Streptomyces</em> 14386 cosmid library, pOJ436 backbone, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cos BA23</td>
<td>pOJ436</td>
<td>Cosmid from <em>Streptomyces</em> 14386 cosmid library, pOJ436 backbone, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cos EB22</td>
<td>pOJ436</td>
<td>Cosmid from <em>Streptomyces</em> 14386 cosmid library, pOJ436 backbone, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKC1132_fd10</td>
<td>pKC1132</td>
<td>pKC1132 derivative containing an internal 1.3 kb fragment of NRPS from FD10, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKC1132_AI6</td>
<td>pKC1132</td>
<td>pKC1132 derivative containing an internal 1.3 kb fragment of NRPS from AI6, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKC1132_BI11_PipA</td>
<td>pKC1132</td>
<td>pKC1139 derivative containing an internal 0.6 kb fragment of piperolic acid incorporating A-domain from BI11, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKC1139_fd10</td>
<td>pKC1139</td>
<td>pKC1132 derivative containing an internal 1.3 kb fragment of NRPS FD10, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKC1139_AI6</td>
<td>pKC1139</td>
<td>pKC1132 derivative containing an internal 1.3 kb fragment of NRPS AI6, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>pKC1139_BI11_PipA</td>
<td>pKC1139</td>
<td>pKC1132 derivative containing an internal 0.6 kb fragment of piperolic acid incorporating A-domain from BI11, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>BI11-PipA-Topo</td>
<td>pCR2.1-TOPO</td>
<td>Topo derivative containing an internal 0.6 kb fragment of piperolic acid incorporating A-domain from BI11, <em>apra</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cyclo-pGEX</td>
<td>pGEX6-P-1</td>
<td>pGEX-6P-1 derivative for expression of the cyclodeaminase gene with N-terminal GST Tag, <em>amp</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>BI11A1</td>
<td>pET28b(+)</td>
<td>pET28b derivative for expression of BI11 A1 with N-terminal His tag, <em>kan</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>BI11A2</td>
<td>pET28b(+)</td>
<td>pET28b derivative for expression of BI11 A2 with N-terminal His tag, <em>kan</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
<tr>
<td>BI11A3</td>
<td>pET28b(+)</td>
<td>pET28b derivative for expression of BI11 A3 with N-terminal His tag, <em>kan</em>&lt;sup&gt;R&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
Material and Methods

2.8 Streptomyces

2.8.1 Mutants of *S. lividans* TK24

Table 2.16: Mutants of *S. lividans* TK24 generated in this work

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Relevant characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. lividans</em> TK24::CPhl8</td>
<td>Integration of CPhl8 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. lividans</em> TK24::CPhl14</td>
<td>Integration of CPhl14 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. lividans</em> TK24::CPhl16</td>
<td>Integration of CPhl16 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. lividans</em> TK24::CPhl18</td>
<td>Integration of CPhl18 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. lividans</em> TK24::CPhl19</td>
<td>Integration of CPhl19 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. lividans</em> TK24::FD10</td>
<td>Integration of FD10 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. lividans</em> TK24::BI11</td>
<td>Integration of BI11 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. lividans</em> TK24::BA23</td>
<td>Integration of BA23 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

2.8.2 Mutants of *S. coelicolor* A3(2)

Table 2.17: Mutants of *S. coelicolor* A3(2) generated in this work

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Relevant characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. coelicolor</em> A3(2)::CPhl8</td>
<td>Integration of CPhl8 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. coelicolor</em> A3(2)::CPhl14</td>
<td>Integration of CPhl14 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. coelicolor</em> A3(2)::CPhl16</td>
<td>Integration of CPhl16 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. coelicolor</em> A3(2)::CPhl18</td>
<td>Integration of CPhl18 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><em>S. coelicolor</em> A3(2)::CPhl19</td>
<td>Integration of CPhl19 at the <em>attP</em> site for heterologous expression, <em>apra</em>&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
2.8.3 Mutants of *S. coelicolor* M512

Table 2.18: Mutants of *S. coelicolor* M512 generated in this work

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Relevant characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. coelicolor</em> M512::CPhl8</td>
<td>Integration of CPhl8 at the <em>attP</em> site for heterologous expression, <em>apra</em>^R^</td>
</tr>
<tr>
<td><em>S. coelicolor</em> M512::CPhl14</td>
<td>Integration of CPhl14 at the <em>attP</em> site for heterologous expression, <em>apra</em>^R^</td>
</tr>
<tr>
<td><em>S. coelicolor</em> M512::CPhl16</td>
<td>Integration of CPhl16 at the <em>attP</em> site for heterologous expression, <em>apra</em>^R^</td>
</tr>
<tr>
<td><em>S. coelicolor</em> M512::CPhl18</td>
<td>Integration of CPhl18 at the <em>attP</em> site for heterologous expression, <em>apra</em>^R^</td>
</tr>
<tr>
<td><em>S. coelicolor</em> M512::CPhl19</td>
<td>Integration of CPhl19 at the <em>attP</em> site for heterologous expression, <em>apra</em>^R^</td>
</tr>
</tbody>
</table>

2.8.4 Mutants of *S. 14386*

Table 2.19: Mutants of *S. 14386* generated in this work

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Relevant characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. 14386</em>::pKC1132_BI11_PipA</td>
<td>Insertion of pKC1132_BI11_PipA in <em>S. 14386</em> to create a knockout of the GE81112 cluster</td>
</tr>
</tbody>
</table>

2.9 Cultivation and conservation of strains

2.9.1 *E. coli/Pseudomonas putida*

3 ml of sterile LB medium were inoculated with a few cells from a single colony and cultivated at 30 °C (*Pseudomonas*) or 37 °C (*E. coli*) at 180 rpm overnight. For storage at −80 °C, 1 ml of the culture was mixed with 1 ml of 50% glycerol. To restart cultivation, a few cells from the frozen cultures were inoculated into fresh LB medium with an inoculating loop. Antibiotics were added to the growth medium as appropriate.

2.9.2 *Streptomyces*

2.9.2.1 Spores

1 ml of sterile 20% glycerol was added to a fresh well-sporulating plate. The surface of the culture was scraped vigorously with an inoculating loop to resuspend the spores. The crude suspension was then poured into a sterile falcon tube and diluted with 20% glycerol to obtain
a homogenous suspension. 1 ml of this spore suspension was then aliquoted into 1.5 ml eppendorf tubes and frozen at –20 °C.

2.9.2.2 Mycelia

A square from a well-sporulating plate was used for inoculation of 30 or 50 ml medium (typically TSB) as preculture. The primary growth cultures were inoculated 1:100 with the preculture, and incubated at 28 °C or 30 °C at 180 rpm. For storage at –80 °C, 1 ml of the culture was mixed with 1 ml of 50% glycerol. To restart cultivation, a few cells were inoculated on fresh agar plates.

2.10 Isolation of prokaryotic DNA

2.10.1 Isolation of genomic DNA

Genomic streptomycetes DNA was isolated using the salting out procedure [10]. Mycelia from a 30 ml culture were harvested at 3000 rpm for 10 min. The mycelia were resuspended in 5 ml SET buffer, and then 100 µl lysozyme solution (1 mg/ml lysozyme in P-buffer) was added. The mixture was incubated at 37 °C for 30–60 min. 140 µl proteinase K solution (20 mg/ml in H2O) was then added, and then the solution was mixed by inversion. 600 µl of 10% SDS was added and again the sample was mixed by inversion. The mixture was incubated at 55 °C for 2 h. 2 ml 5M NaCl were added, and again mixed by inversion. After cooling to 37 °C, 5 ml chloroform were added, and the sample was mixed by inversion for 30 min at 20 °C. The sample was then centrifuged for 15 min at 3000 rpm at room temperature. The supernatant was transferred to a fresh tube, 0.6 vol isopropanol were added, and then the sample was mixed by inversion. After ca. 3 min, the DNA could be spooled onto a sealed Pasteur pipette. The DNA was rinsed in ca. 2 ml 70% ethanol, air dried and dissolved in 200 µl H2O by heating at 55 °C.

2.10.2 Isolation of plasmid and cosmid DNA by alkaline lysis

An overnight culture of 1–3 ml of *E. coli* was harvested by centrifugation in 1.5 ml eppendorf tubes. The supernatant was removed and the pellet was resuspended in 250 µl buffer P1 + RNAse (100µg/ml). The cells were then lysed in 250 µl P2 (NaOH/SDS). The lysate was then neutralised by adding 250 µl potassium acetate. The samples were then centrifuged for 10 min at 13000 rpm. The plasmid DNA remained in the supernatant was precipitated with 600 µl
isopropanol. After 10 min of centrifugation at 13000 rpm, the DNA pellets were washed with 700 µl 70% EtOH, air dried, and then dissolved in 20 µl H₂O.

2.10.3 Isolation of plasmid DNA with the NucleoSpin® Plasmid Kit
Isolation of plasmid DNA with the NucleoSpin® Plasmid Kit was performed according to the manufacturer’s instructions.

2.11 Separation and purification of DNA
A 0.8% agarose gel containing ethidium bromide (EtBr, ~0.25 µg/ml) was prepared and the DNA samples, prepared in 6× DNA loading dye, were loaded into the slots of the gel. A size standard (GeneRuler™ 1 kb DNA ladder, Fermentas) was used to determine the length of DNA samples from PCR or specific fragments from restrictions. Electrophoresis was performed in 1× TBE buffer with a voltage of approximately 10 V per cm gel length until the dye front reached the edge of the gel. DNA was visualized under UV light (245 nm) using a gel documentation (peqlab).

2.11.1 Extraction of DNA from agarose gels
After separation by gel electrophoresis, desired DNA fragments were excised from the gel under UV light. Extraction from the gel was carried out with the NucleoSpin® Extract Kit, according to the manufacturer’s protocol.

2.11.2 DNA precipitation
DNA dissolved in H₂O was mixed with 1/10 vol 3M NaOAc solution. After addition of 1 volume isopropanol, the sample was incubated 20 min at –20 °C. After centrifugation for 10 min at 13000 rpm, the pellet was washed with 700 µl 70% EtOH, air dried and redissolved in the desired volume of H₂O.

2.11.3 Phenol/chloroform extraction
An equal volume of phenol/chloroform was added to the DNA sample, followed by vortexing. After centrifugation the upper aqueous phase was separated cleanly by decanting. Phenol was quantitatively removed by subsequent precipitation with isopropanol (see part 2.10.2).
2.12 Polymerase chain reaction

For polymerase chain reaction or PCR different polymerases were used which are listed in Table 2.20.

2.12.1 Composition of standard PCR setup

Different annealing temperatures were tested for the PCR reactions

<table>
<thead>
<tr>
<th>Table 2.20: PCR protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Buffers</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>MgCl₂ (25 mM)</td>
</tr>
<tr>
<td>dNTPs (1.2 mM/dNTP)</td>
</tr>
<tr>
<td>DMSO</td>
</tr>
<tr>
<td>Glycerol (50%)</td>
</tr>
<tr>
<td>Primers 1/2 (50 pmol/µl)</td>
</tr>
<tr>
<td>Polymerase</td>
</tr>
<tr>
<td>Template</td>
</tr>
<tr>
<td>Sterile water</td>
</tr>
<tr>
<td>Total volume</td>
</tr>
<tr>
<td>Primary denaturation</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Annealing</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Extension</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Final extension</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Cycles</td>
</tr>
</tbody>
</table>
2.12.2 Colony PCR

Cells were removed from an agar plate with an inoculating loop or 10 µl of cell suspension was taken from a liquid culture and transferred into an eppendorf tube containing 10 µl H₂O. The cell mixture was then boiled at 95 °C for 10 min in an eppendorf thermomixer. 1 µl of this mixture was used as template in the PCR reaction. Hot Star Taq Polymerase was used for all colony PCRs following the protocol in Table 2.20.

2.13 Enzymatic manipulation of DNA

2.13.1 Restriction endonucleases

Reactions were performed in a total volume of 20 µl containing 2 µl 10× buffer, 5 µl DNA, 0.5 µl restriction enzyme and 12.5 µl H₂O. The restriction mix was incubated for 2–4 hours or overnight at 37 °C (or at other incubation temperatures recommended by the manufacturer). All restriction enzymes and corresponding buffers were purchased from Fermentas GmbH.

2.13.2 Ligation

Ligations were carried out in a total reaction volume of 10 µl, containing 1 µl vector, 7 µl insert DNA, 1 µl 10× ligation buffer and 1 µl T4 DNA ligase. The reaction was incubated overnight at 16 °C, and then heat-inactivated for 20 min at 65 °C (Thermomixer comfort) following verification of the successful ligation by agarose gel electrophoresis.

2.13.3 Dephosphorylation

Dephosphorylation was carried out within the restriction reaction. 1 µl shrimp alkaline phosphatase (SAP) was added to the restriction digestion. Incubation was carried out at 37 °C for 1 h. Heat-inactivation of the enzyme was then performed for 15 min at 65 °C.

2.13.4 Cloning with TOPO-TA-Cloning kit

Direct cloning of PCR products was carried out using the TOPO TA Cloning® kit (Invitrogen). TOPO TA Cloning® provides a highly efficient, 5-minute, one-step cloning strategy ("TOPO® Cloning") for the direct insertion of Taq polymerase-amplified PCR products into a plasmid vector. Taq polymerase has a non-template-dependent terminal transferase activity that adds a single deoxyadenosine (A) to the 3´ ends of PCR products. The linearized vector supplied in this kit has single, overhanging 3´ deoxythymidine (T) residues.
This allows PCR inserts to ligate efficiently with the vector. If PCR reactions were carried out with other than Taq polymerase, addition of 3’ A-overhangs is needed. All reactions were performed according to the manufacturer’s protocol.

2.14 Red/ET recombination

For Red/ET recombination the target cosmid/plasmid was first electroporated into competent cells containing the Red/ET catalyzing plasmid GB2005/pSC101-BAD-γβαA-amp. Selection was carried out using ampicillin (50 µg/ml) and the appropriate antibiotic from the incoming target construct. Cells were grown at 30 °C overnight and used as preculture. 1.4 ml LB medium were inoculated with 30 µl of the cells and cells grown at 30 °C until an OD$_{600}$ = 0.2 was reached. Expression of Red/ET proteins was induced by adding 25 µl of 10% L-arabinose. Further incubation was carried out at 37 °C until an OD$_{600}$ = 0.4 was obtained. Cells were then harvested by centrifugation for 1 min at 13000 rpm and washed twice with 1 ml chilled H$_2$O. After the last washing step, the cells were resuspended in 30 µl H$_2$O and electroporated with purified PCR product. Electroporation conditions were as follows: 1350 V, 10 µF, 600 Ohms. After electroporation, 1 ml pre-warmed LB was added and the cells were cultivated for 1 h at 37 °C without antibiotic. The cells were then plated out on LB-agar containing the appropriate antibiotics. Recombination products were checked by restriction digest with several enzymes (selected on the basis of in silico digests of the expected products).

2.15 Transformation of bacteria

2.15.1 Electroporation of E. coli

To obtain electrocompetent cells, cells were grown at 37 °C to an OD$_{600}$ = 0.6–0.8 in 1.5 ml Eppendorf tubes. The cell suspension was then centrifuged and washed twice with chilled H$_2$O. Cells were then resuspended in 30 µl H$_2$O and used for electroporation with a 0.1 cm cuvette. Electroporation parameters were as follows: 1250 V, 10 µF, 600 Ohms, and the expected time constant was 5 ms. After electroporation, 1 ml of pre-warmed LB was added and the cells were cultivated for 1 h at 37 °C without antibiotic. The cells were then plated out on LB-agar containing the appropriate antibiotics.
2.15.2 Conjugation of streptomycetes with E. coli

Electrocompeent cells of E. coli ET12567/pUZ8002 grown in the presence of kanamycin (kan) (25 µg/ml) and chloramphenicol (cm) (25 µg/ml) to maintain selection for pUZ8002, were prepared as described previously (2.15.1). The competent cells were transformed with the oriT-containing vector, and selected for the incoming plasmid only. A colony was inoculated in 1 ml of LB-medium containing cm and kan and the antibiotic used to select for the oriT-containing plasmid, and grown overnight. The overnight culture was diluted 1:100 in fresh LB containing antibiotics, and grown at 37 °C to an OD600 of 0.4–0.6. The cells were washed twice with an equal volume of LB and resuspended in 0.1 volume of LB. During the washing of the E. coli cells, for each conjugation approximately 10^8 Streptomyces spores were added to 500 µl 2YT broth, heat shocked at 50 ºC for 10 min and then cooled. 500 µl E. coli cells were then added to 500 µl of the heat-shocked spores or mycelial fragments. The mixture was spun briefly and most of the supernatant was poured off. The pellet was then resuspended in the residual liquid. The cells were plated out on SM agar plates without antibiotic. After 16–20 h incubation at 30 ºC, cells were scraped from the plate and resuspended in 500 µl LB medium. The mixture was then plated in different dilutions on SM agar plates containing nalidixic acid (25 µg/ml) and apramycin (60 µg/ml). Incubation at 30 ºC was continued until the appearance of potential exconjugants.

2.15.3 Protoplast transformation

25 ml YEME medium were inoculated with 0.1 ml spore suspension and incubated for 36–40 h in a 30 ºC shaker at 180 rpm. Cultures of S. lividans and S. coelicolor were poured into a 50 ml falcon tube and centrifuged for 10 min at 1000×g. The supernatant was discarded and the pellet was resuspended in 15 ml of 10.3% sucrose. After 10 min of centrifugation, the supernatant was again discarded and the washing step was repeated. The mycelia were resuspended in 4 ml lysozyme solution (1 mg/ml in P-buffer) and incubated at 30 ºC for 15–60 min. 5 ml P-buffer were added and the protoplasts were sedimented gently by spinning at 1000×g for 7 min. The supernatant was discarded and the protoplasts were resuspended in 1 ml P-buffer. 300 µl PEG (sterile filtered) were mixed with 900 µl P-buffer. 400 µl of this solution were mixed with 100 µl of the freshly-prepared protoplasts and 10 µl DNA. Dilutions of this mixture were plated out on R2YE agar plates, and the plates incubated at 30 ºC. After 16–20 h, the plates were overlaid with 1 ml H2O containing 1 mg/ml apramycin. Resistant colonies typically appeared after 3 days.
2.16 Construction of a cosmid library from *Streptomyces 14386*

This method was used to clone large genomic DNA fragments (35–45 kb) into cosmid vectors [77] using phage transfection.

2.16.1 Preparation of the ligation reaction

Genomic DNA of *Streptomyces 14386* (see part 2.10.1) was partially digested with *Sau3A*. After separation by gel electrophoresis, fragments showing an average size between 35 and 45 kb were pooled and extracted with phenol/chloroform (2.11.3) and precipitated. The DNA was subsequently dephosphorylated (2.10.3) and extracted with phenol/chloroform again. The aqueous phase was later precipitated together with the vector pOJ436.

The *E. coli/Streptomyces* shuttle vector pOJ436 was digested with the restriction endonuclease *PvuII* and subsequently dephosphorylated with SAP. The DNA was extracted with phenol/chloroform and precipitated. The vector was subsequently digested with *BamHI*, extracted with phenol/chloroform again and precipitated together with the partially-digested genomic DNA.

2.16.2 Ligation and phage infection

The precipitated DNA pellet from 2.16.1 was resuspended in 15 µl H₂O and ligated at 16 °C (2.13.2). The DNA was precipitated again and resuspended in 20 µl H₂O. 4 µl from the ligation reaction were mixed with the GigapackIII Gold Packaging Extract (Stratagene), incubated for 1.5 h at room temperature and subsequently mixed with 500 µl SM-buffer and 50 µl chloroform. The mixture was centrifuged briefly and stored at 4 °C.

In parallel *E. coli* SURE was grown in LB medium supplemented with 10 mM Mg₂SO₄ and 0.2% maltose, until an OD₆₀₀ = 0.7 was reached. Various amounts of packaging extract were then mixed with *E. coli* SURE as follows:

- 2 µl packaging extract + 23 µl SM buffer + 25 µl *E. coli* SURE
- 5 µl packaging extract + 20 µl SM buffer + 25 µl *E. coli* SURE
- 10 µl packaging extract + 15 µl SM buffer + 25 µl *E. coli* SURE

The mixtures were incubated for 30 min at 37 °C. 200 µl LB were then added and incubation was continued at 37 °C while shaking 2 h at 165 rpm. The cells were plated out on LB agar containing 60 µg/ml apramycin. The titer of the packaging extract was determined through counting the colonies. Typically, 2 µl of packaging extract yielded 2275 colonies.
2.16.3 Generation of the cosmid library

To obtain a 10-fold coverage from an estimated genome size of 8 Mbp, ca. 2000 cosmid clones were needed. In order to ensure that enough clones were obtained, 5 µl packaging extract were used and prepared as described in 2.16.2. After incubation, the cells were plated on LB agar containing 60 µg/ml apramycin in 22 cm × 22 cm NUNC plates (Nalgene). The plates were incubated at 32 °C overnight and the colonies were picked using the Qbot-robot (Gentrix) in the Department of Genome Analysis at the Helmholtz-Centre for Infection Research in Braunschweig in six 384-well microtiterplates containing 50 µl 2YT medium (60 µg/ml apramycin).

2.16.4 Robotically produced high-density colony arrays

The plates were incubated at 37 °C overnight and the growing colonies were spotted with the Qbot-robot (Gentrix) on fresh NUNC-LB-agar plates (60 µg/ml apramycin) which were covered with a 22 cm × 22 cm nylon membrane. 50 µl freezing solution (Table 2.6) was subsequently added to each well of the microtiter plates were subsequently amended with 50 µl freezing solution (2.4.1) and frozen at -80 °C.

Altogether 12 membranes were spotted with the complete cosmid library. To enable an explicit assignment of the transferred colonies to the six 384er microtiterplates each colony of the microtiterplates was spotted twice in a definitend 4x4 array (Figure 2.1).

Figure 2.1: Spotting array of the high-density colony filters.

The NUNC-LB-agar plates with the nylon membranes were incubated at 32 °C for 16–20h. For the subsequent lysis of the cells, the membranes were processed as follows: Whatman paper was soaked with denaturation solution, neutralization solution and 2× SSC solution (Table 2.4). The membranes (bottom side) were transferred onto the Whatman paper, initially using denaturation solution. After 15 min incubation, the membranes were briefly transferred.
to dry Whatman paper in order to remove remaining denaturation solution. Subsequent incubation in neutralization solution (15 min) and 2× SSC (10 min) was carried out, followed by drying of the membranes on fresh Whatman paper. The colonies were fixed on the membranes through incubation at 80 °C for 2 h. The membranes were treated with 14 ml Proteinase K solution (2 mg/ml) for 1 h at 37 °C. To remove cell debris, the membranes were treated with H₂O soaked tissues. After drying at room temperature, the processed membranes could be used for hybridization.

2.17 DNA Hybridization

2.17.1 Southern Blot
5 µg genomic DNA were digested with different restriction enzymes (2.13.1) and separated by gel electrophoresis (2.11). After gel documentation, the gel was depurinated for 10 min in 0.25 M HCl, with gentle shaking. The gel was subsequently incubated twice for 20 min in denaturation buffer (Table 2.4) and neutralization buffer (Table 2.4) at room temperature. After treatment of the gel, the DNA was blotted onto a membrane by vacuum blot.

2.17.2 Vacuum Blot
The membrane placed onto Whatman paper soaked with 10× SSC (Table 2.4) on the vacuum blotter, in order to prevent the formation of air bubbles. A plastic mask was placed on the membrane leaving a window for placing the gel onto it. The gel was placed onto the membrane and the vacuum pump was switched on to immobilize the gel onto the membrane. The vacuum blot chamber was filled with 10× SSC and blotting was carried out for 90 min, with a vacuum of 5 mm Hg.

2.17.3 Generation of DIG-labeled probes
For the generation of DIG-labeled probes the ‘DIG PCR labeling mix’ from Roche was used. Digoxigenin labeled desoxynucleotides were employed in a standard PCR reaction (2.12). The obtained probe was stored at –20 °C, until it was used for hybridization.

2.17.4 Hybridization and detection with the DIG labeling system
The DIG systems protocol from Roche was used. The membrane was pre-hybridized in hybridization buffer for 2 h at 42 °C. The DIG labeled probe was denaturated for 10 min at 95
°C and directly cooled on ice. The probe was then added to the hybridization buffer to obtain a final concentration of DIG labeled probe of 7.5–15 ng/ml. The pre-hybridization solution was removed and fresh hybridization buffer containing the DIG-labeled probe was added to the membrane. Hybridization was performed at 42 °C overnight. Non stringent hybridization (heterologous) was performed at 38–40 °C. After hybridization, the membrane was washed twice with 2× SSC washing buffer (Table 2.4) at room temperature, followed by stringent washing with 0.5× SSC (Table 2.4) at 68 °C (58 °C for colony hybridization). The membrane was subsequently equilibrated in maleic acid buffer (Table 2.4) for 10 min and then incubated with blocking solution (Table 2.4) for 30 min at room temperature. DIG antibody (Anti digoxigenin-AP conjugate, Fab fragment) was added to fresh blocking solution (1:10000) and the membrane was incubated for 30 min in this solution. The membrane was then washed twice with maleic acid buffer for 15 min. The membrane was then briefly equilibrated with detection buffer (Table 2.4) (5 min) and subsequently covered with CDP-Star solution (Roche) (1 ml/100 cm²). After 1 min of incubation, the membrane was shrink-wrapped and the chemoluminescence was detected by gel documentation (Peqlab).

2.17.5 Stripping of membranes
Membranes were treated with stripping buffer (Table 2.4) for 20 min at room temperature. After a brief equilibration in 2× SSC, the membranes were either used again for hybridization or stored at –20 °C.

2.17.6 Colony Hybridization
Hybridization of the high-density colony filters from S. 14386 was carried out with the DIG system from Roche as described in 2.17.4.

2.17.7 Analysis of DNA- and protein sequences
Sequence data were analyzed with the program Vector NTI (InforMax). Additional programs for sequence analysis were freely accessible via the internet.

Frame plot http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl
PKS/NRPS Analysis http://www.tigr.org/jravel/nrps/
Material and Methods

2.18 Protein and enzymatic analysis

Different proteins from the GE biosynthetic gene cluster were overexpressed as GST- or His-tagged fusion proteins for biochemical characterization. The genes were cloned into the expression vectors pGEX-6P-1 (N-terminal GST-tag) or pET-28b(+) (N-terminal His-tag) (Table 2.13). The genes are expressed under the control of IPTG-inducible promoters (T7 promoter in the pET system, P_{lac} promoter in the pGEX system). The pET-28b(+) plasmid carries kanamycin resistance, the pGEX-6P-1 plasmid ampicillin resistance. The genes were cloned into the expression vectors, sequenced and transformed into *E. coli* BL21(DE3) or *E. coli Rosetta* BL21 (DE3) pLysS/pRARE.

2.18.1 Cultivation and cell disruption

A single colony of *E. coli* BL21 or *E. coli Rosetta* BL21 (DE3) pLysS/pRARE containing the plasmid with the expression vector construct was used for inoculation of a LB culture (with the appropriate antibiotic). This preculture was grown overnight and used to inoculate the expression culture (1:100 dilution). The culture was then grown until an OD_{600} = 0.8 at 30 °C and 180 rpm, and then induced with IPTG to a final concentration of 0.1 M. Incubation was continued at 16 °C overnight, or at 30 °C for additional 2 h. Cells were harvested by centrifugation for 10 min at 10000 rpm and were resuspended in an appropriate buffer (e.g. PBS buffer or binding buffer for Äkta-purification) (Table 2.5)). All further steps were carried out on ice or at 4 °C.

To check for the presence of soluble protein, 1 ml of the cell culture was centrifuged, the pellet was resuspended in 100 µl BugBuster® Protein Extraction Reagent (Novagen) and then the mixture was incubated for 20 min at room temperature. The cell lysate was then centrifuged for 10 min at 4 °C and 10 µl of the supernatant and the remaining pellet were analyzed by SDS-PAGE. If the protein was soluble, the residual culture was used for cell disruption at larger scale. For this, the cells were lysed using a French press at 700 psi (typically two passages). *E. coli Rosetta* BL21 (DE3) pLysS/pRARE strains were induced to
carry out self-lysis by freezing and thawing, which caused the production of lysozyme. In this case, the cell lysate was centrifuged for 45 min at 10000 rpm, and the remaining supernatant was used for further purification.

2.18.2 Protein purification

2.18.2.1 GST-tagged proteins

GST fusion proteins were purified using GST MicroSpin columns (GE Healthcare). The lower caps were removed from the columns, and each column was placed into a clean 2 ml eppendorf tube and spun for 1 min at 3500 rpm. The flow-through was discarded and the lower caps were replaced. 600 µl of the cell lysate (see part 2.18.1) were applied to the column. The columns were recapped at the top and mixed gently by inversion for 20 min at room temperature. The top and bottom caps were removed and each column was placed into a clean 2 ml eppendorf tube. The tube was centrifuged for 1 min at 3500 rpm to collect the flow-through. Each tube was placed into a new eppendorf tube and 600 µl 1× PBS wash buffer was applied to each column. This step was repeated once. 150 µl cleavage buffer containing 10 µl PreScission (GE Healthcare) protease were then added to the column and the top and bottom caps were replaced; the column was then incubated at 4 °C overnight. On the following day, the columns were placed into eppendorf tubes and the eluate containing the protein without GST tag was collected and checked using SDS-PAGE.

2.18.2.2 His-tagged proteins

Prepacked HisTrap™ HP columns were used for preparative purification of histidine-tagged recombinant proteins by immobilized metal ion affinity chromatography (IMAC) on the Äkta prime. 15 ml protein lysate were filtered through a sterile filter and loaded onto the 1 ml HisTrap column. Purification was performed as recommended in the GE Healthcare manual (HisTrap HP, Instructions 71-5027-68 AF). The following gradient was applied:
Material and Methods

Table 2.21: Imidazol gradient applied for Äkta system.

<table>
<thead>
<tr>
<th>Volume [ml]</th>
<th>Eluent A: binding buffer [%]</th>
<th>Eluent B: elution buffer [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>60</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>70</td>
<td>88</td>
<td>12</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>90</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>100</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>110</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Pressure limit: 0.55 MPa; Flow 1 ml/min

To check the fractions for the presence of protein, 150 µl of each fraction were precipitated with an equal volume of acetone at –20 °C for 20 min. After centrifugation, the precipitated protein pellet was resuspended in 10 µl of 4× protein loading buffer and analyzed by SDS-PAGE.

2.18.3 Bradford assay

Bradford assays were carried out according to the manufacturer’s manual (Biorad). In each case, a standard curve was measured with bovine serum albumin (BSA).

2.18.4 Protein concentration by centrifugation

Concentration of protein extracts was performed using Amicon Ultra-30 K centrifugal filters (Millipore), according to the manufacturer’s instructions. The nominal molecular weight limit (NMWL) of the filters was 30,000 Da.

2.18.5 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis, is a technique to separate proteins according to their molecular weight [78]. The following proteins markers were applied for size determination:

Marker PageRuler™, Fermentas: 10, 15 25, 30, 40, 50, 60, 70, 80, 100, 120, 170 kDa
Marker PageRuler™ Prestained, Fermentas: 10, 15 25, 35, 40, 55, 70, 120, 170 kDa
2.18.5.1 Preparing of SDS gels and electrophoresis

For detection of proteins, a 10% separating and a 5% stacking polyacrylamide gel were prepared. Ca. 20 µg protein was loaded into each lane of the gel. The protein samples were mixed with 10 µl of 4× protein loading buffer and boiled for 10 min at 95 °C. After loading the protein sample and a prestained protein marker (Fermentas) onto the gel, electrophoresis was performed in 1× running buffer at 100 V. Electrophoresis was stopped when the bromophenol blue front reached the edge of the gel (approximately 1.5 h).

10% Separating gel:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂O</td>
<td>5.9 ml</td>
</tr>
<tr>
<td>30% Bis-/Acrylamide mix</td>
<td>5 ml</td>
</tr>
<tr>
<td>1.5 M Tris (pH 8.8)</td>
<td>3.8 ml</td>
</tr>
<tr>
<td>10% SDS</td>
<td>150 µl</td>
</tr>
<tr>
<td>10% APS</td>
<td>150 µl</td>
</tr>
<tr>
<td>TEMED</td>
<td>6 µl</td>
</tr>
</tbody>
</table>

The solution was poured between two glass plates, overlaid with isopropanol and allowed to polymerize.

5% Stacking gel:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂O</td>
<td>7 ml</td>
</tr>
<tr>
<td>30% Bis-/Acrylamide mix</td>
<td>1.7 ml</td>
</tr>
<tr>
<td>1.5 M Tris (pH 6.8)</td>
<td>1 ml</td>
</tr>
<tr>
<td>10% SDS</td>
<td>100 µl</td>
</tr>
<tr>
<td>10% APS</td>
<td>100 µl</td>
</tr>
<tr>
<td>TEMED</td>
<td>10 µl</td>
</tr>
</tbody>
</table>

The solution was poured onto the separating gel and allowed to polymerize.

2.18.5.2 Detection of protein bands

Protein gels were stained with Coomassie Brilliant Blue. The staining solution contained methanol/acetic acid/water (10:10:80) and 1% Coomassie Brilliant Blue R250 (Roth). Staining was carried out for 1–2 h, or was accelerated by heating the gel in the microwave for
a couple of seconds. Destaining was performed in the same mixture without Coomassie. Gel-Dry™ Drying Kit (Invitrogen) was used to conserve gels.

2.18.6 Protein identification by MALDI-MS
The desired protein band was excised from the SDS-PAGE gel and washed three times with water. 500 µl of a 50% acetonitrile (ACN) mixture in H₂O were added until the gel piece was destained. The gel piece was then air dried and 500 µl ACN were added for 10 min. The gel piece was air dried again and 20 µl trypsin were added to each sample. After complete rehydration of the gel piece, 20 µl 40 mM NH₄HCO₃ were added and the gel piece was incubated overnight. TFA was added to a final concentration of 0.1%. The sample was then stored at –20 °C prior to analysis by MALDI-MS.

2.18.7 Cyclodeaminase assay
To determine the cyclodeaminase activity one millimole L- or D-Lysine was incubated at 30 °C with 10 µM protein, 5 mM Tris (pH 8.0), 1 mg/ml BSA, and 100 µM NAD⁺. H₂O was added at a total volume of 100 µl and the reaction was incubated for 16 h. The reaction was quenched with two reaction volumes of acetonitrile, stored at -20 °C for at least 10 min and spun at 13000 rpm for 10 min. The supernatant was vacuum-dried and solved in 200 µl borate buffer and used for derivatization. For derivatization, the reactant solution consisted of FMOC-Cl dissolved in acetone (1.5 mg/ml). Two hundred microliters of FMOC solution was added to the 200 µl of the reaction/borate buffer mixture and incubated 10 min at room temperature. Termination of the reaction and removal of excess reagent (Fmoc-Cl), its hydrolysis product FMOC-OH, and acetone were performed by extraction with 600 µl hexane. After two additional extractions, 400 µl 25% acetonitrile in 0.1 M borate buffer was added. A 50 µl aliquot of the diluted sample was injected into the HPLC system. FMOC-derivatized compounds were separated with a standard gradient (see section 2.10.1.). The peak corresponding to FMOC pipecolic acid was detected with a mass of \( m/z = 571.2 \ [\text{M-H}]^- \).

2.18.8 ATP-PP₁ exchange assay
This assay was carried out in order to determine the substrate specificity of adenylation domains. Therefore 2 µM protein, 5 µl dATP (40 mM), 10 µl 10× buffer and 5 µl amino acid (40 mM) were mixed together. The reaction was started by addition of 0.1 µCi \(^{32}\text{P}\)-
pyrophosphate (Perkin Elmer). Incubation was done for 30 min and the reaction was stopped by the addition of 500 µl stop-mix. After centrifugation for 1 min at 13000 rpm the pellet was washed twice with 1 ml 0.1M wash buffer (Table 2.5). The charcoal pellet was then resuspended in 0.5 ml H₂O. 3.5 ml scintillation fluid was then added to the mixture and the radioactivity was determined by scintillation counting.

2.19 Analysis of secondary metabolites in natural producer strains and recombinant *Streptomyces* strains

*Streptomyces* strains used in this work were *Streptomyces* Tü6071 [66], *S. lividans* TK24 [72], *S. coelicolor* A3(2) [11], *S. coelicolor* M512 (ΔredD, ΔactII-ORF4 SCP1–SCP2) [73] and *Streptomyces* 14386 [67] (Table 2.10). For production and heterologous expression of secondary metabolites different cultivation conditions were used for the different strains.

2.19.1 Cultivation in production medium

2.19.1.1 Phenalinolactones

For phenalinolactone production, the cultures were cultivated under standard conditions at 28 °C and 180 rpm on a rotary shaker, and harvested after 6–8 days by centrifugation. NL111 (Table 2.8) liquid medium was used for production. Precultures were grown in the same medium and used to inoculate production cultures (1:100 dilution). Studies on the optimization of PL production were performed under standard cultivation conditions in NL111 medium, but with growth at 37 °C and also in NBG, SM and INA5 medium. Apramycin (60 µg/ml) was used for cultivation of recombinant *Streptomyces* strains.

2.19.1.2 GE81112

Precultures were grown in V6 medium and used to inoculate production cultures (1:100 dilution). INA5 or T6 (Table 2.8) production medium was used for production of GE81112. The cultures were maintained at 30 °C and 180 rpm on a rotary shaker and harvested after 6–8 days. Apramycin (60 µg/ml) was used for cultivation of recombinant *Streptomyces* strains.

2.19.2 Feeding experiments

Precultures were grown in V6 medium at 30 °C for 3 days and then used to inoculate a 50 ml production culture in T6 medium (1:100 dilution). Stock solutions of deuterium-labeled L-
and D-pipecolic acid and uniformly labeled (U\(^{13}\)C) L-histidine (500 mM) were prepared and introduced into the cultures after 72, 96, 120, 144 and 168 h, to a final concentration of 1 mM. Cells were harvested after 196 h. As a control, the same cultivation was performed without feeding of labeled substances.

### 2.19.3 Extraction

After cultivation in production medium, the cells were harvested by centrifugation at 5000 rpm for 10 min. The culture supernatant was extracted twice with ethyl acetate. The solvent was then removed by evaporation, redissolved in methanol (1% of the original culture volume) and centrifuged for 10 min. The supernatant was transferred to a fresh vial. The cell pellet was extracted with a mixture of acetone/methanol (1:1). The solvent was then removed by evaporation redissolved in methanol (1% of the original culture volume) and centrifuged for 10 min. The supernatant was transferred to a fresh vial.

### 2.19.4 Purification of GE compounds

The GE81112 producer *S. 14386* was cultivated at 30 °C in 5 L baffled shake flasks at 180 rpm in GE production medium (see part 2.19.1.2). 13 L were harvested after 6 days of cultivation. Cells were extracted with acetone/methanol (1:1) and the supernatant was extracted with ethyl acetate (1:1), as described in 2.19.3. HPLC-MS analysis revealed that the desired compounds could only be found in the supernatant. 13 L of supernatant were then extracted twice with ethyl acetate (1:1). After evaporation of the solvent, the extract was resuspended in methanol and extracted with heptane (1:1). The extract was then fractionated by preparative RP-HPLC (Dionex P680 HPLC equipped with a PDA-100 Photodiode Array Detector; Waters xBridge™ Prep C18 column (5µm OBD™ 19 × 100 mm)) using a stepwise gradient, (0–2 min, 10% acetonitrile; 2–30 min, 10–95% acetonitrile; 10 ml/min; detection at 280 nm; injection volume, 250 µl) coupled to a mass spectrometer (Bruker HCTplus ion trap, operating in positive-ionization mode at a scan range of m/z 100–1100).

Fractions were checked by HPLC-MS analysis, and combined when they contained the desired compounds. The solvent was evaporated, and the compounds were further enriched by gel permeation chromatography (Sephadex LH-20 in methanol). Pure compound was obtained by sequential RP-HPLC using a stepwise gradient (solvent B: acetonitrile containing 0.01% formic acid, 0–10 min, 30% B; 10–20 min, 30–50%; 20–25 min, 50–95% B; 25–27 min, 95%B, injection volume 200 µl).
2.19.5 Analysis of the secondary metabolites

2.19.5.1 HPLC-UV/Vis, HPLC-MS, HPLC-MS/MS, LTQ-Orbitrap-MS

10 μl of the extracts (see part 2.19.3) were analyzed by HPLC-MS analysis. Analysis was carried out using an Agilent 1100 series solvent delivery system equipped with a photodiode array detector and coupled to a Bruker HCTplus ion trap mass spectrometer. Chromatographic separation was carried out on a Nucleodur C18/3 μm RP column (125 × 2 mm; Macherey & Nagel) equipped with an C18/5 precolumn (8 × 3 mm) using a mobile phase system consisting of H2O (A) and acetonitrile (B), each containing 0.1% formic acid. The following gradient was applied: 0–2 min, 25% B; 2–22 min linear from 25–95 %B; 22–24 min isocratic at 95% B. Detection was carried out in negative or positive ionization mode at a scan range of m/z = 100–1100.

Phenalinolactones were identified by comparison to the retention times (rt) and the MS data of authentic standards (phenalinolactone A: rt = 13.5 min, [M–H]− = 714.5; phenalinolactone D: rt = 18.1 min, [M–H]− = 698.3) in negative ionization mode. Phenalinolactone E was identified by comparison to the retention times and the MS data of phenalinolactone D (phenalinolactone E: rt = 13.0 min, [M–H]− = 570.3). LC-coupled FT-Orbitrap-MS analysis was carried out with an Accella UPLC system (Thermo Electron Corporation) coupled to a LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific) operating in negative or positive ionization mode at a scan range of m/z 100–2000. A Hypersil Gold column (2.1 × 50 mm; Thermo Fisher Scientific) was used for separation with a solvent system consisting of H2O (A) and acetonitrile (B), each containing 0.1% formic acid. A gradient of 5–95 % B was applied over 10 min. Measurements were carried out in single ion mode (SIM). GE8112 compounds were identified by comparison to the retention times (rt) and the MS data of authentic standards (GE factor A: [M+H]+ = 644.21858; GE B: [M+H]+ = 659.22953 and GE factor B1: [M+H]+ = 658.2458760) in the negative ionization mode.

2.19.5.2 NMR

The structure of the isolated compounds 631, 645 and 659 were elucidated by using 1D (1H-, 13C-NMR) and 2D NMR analysis (1H-1H-COSY, HSQC heteronuclear sequential quantum coherence, HMBC heteronuclear multiple bond correlation) on Bruker DRX 500 or Bruker Advance 500 spectrometers using several solvents ([d6]DMSO and CD3OD). Calculation of theoretical NMR-spectra was performed with ACD/Labs 7.00 software.
3 Results

3.1 Heterologous expression and genetic engineering of the phenalinolactone biosynthetic gene cluster using Red/ET recombineering

3.1.1 Goals

The aim of the thesis was the development of heterologous expression techniques for complex secondary metabolite pathways from streptomycetes using Red/ET recombineering. The biosynthetic gene cluster which was chosen for this work was the phenalinolactone biosynthetic gene cluster from *Streptomyces* Tü6071. The gene cluster was already cloned and sequenced in the group of Andreas Bechthold from Freiburg University, from whom the strain was obtained. As the complete gene cluster was located on two different cosmids, we aimed in the first step to reconstitute the complete PL biosynthetic pathway on one expression construct using Red/ET recombineering. After the successful cloning, we planned to express the pathway in different heterologous host strains such as pseudomonads and streptomycetes. The successful heterologous expression would then open the door for generation of novel analogues by genetic engineering.

3.1.2 Reconstitution of the complete *pla* biosynthetic gene cluster

We employed Red/ET recombineering in *E. coli* [36-38] to rebuild the entire *pla* biosynthetic pathway on an integrative *E. coli*-streptomycetes shuttle vector (pOJ436), and to introduce additional genetic elements for heterologous expression in pseudomonads and streptomycetes.

3.1.2.1 Cloning strategy

The biosynthetic gene cluster comprises 11 transcriptional units harboring 35 genes. Two pOJ436-derived cosmids (Cos3-1O12 and Cos10-4D08 [66]) containing overlapping regions of the *pla* gene cluster were used as starting points. The cloning strategy is described in Figure 3.1, and consists of three different cloning steps that are each discussed in detail below.
Figure 3.1: Description of the cloning strategy.

The phenalinolactone biosynthetic gene cluster consists of 35 orfs, and was located on two overlapping cosmids (Cos3-1O12 and Cos10-4D08) carrying the pOJ436 backbone. In the first step, a Scal oriT-tet-trpE cassette was inserted via Red/ET recombineering into Cos10-4D08 at the upstream end of the gene cluster, generating the construct CPhl9. In the next step, introduction of the zeocin resistance gene at the downstream end of the cluster led to the construct CPhl7. After restriction digest of CPhl9 with Scal, the resulting 15 kb fragment was used to reassemble the whole cluster using Red/ET recombineering. The final expression construct containing the whole cluster was designated CPhl8.
3.1.2.2 Cloning step 1: introduction of the Scal-oriT-tet-trpE box at the 5’-end of the pla biosynthetic gene cluster

Cos10-4D08 was modified by insertion of a gene cassette into the upstream end of the pla gene cluster, to generate construct CPhl9. In addition to a Scal restriction site and a tetracycline (tet) resistance gene, the introduced cassette also contained an oriT as well as a portion of the trpE gene from Pseudomonas putida which should in future enable the transfer of the expression construct into pseudomonads. The oriT was needed for conjugation and the trpE gene for integration into the Pseudomonas putida genome. This step also resulted in the deletion of orfs 1–3, which we expected not to be involved in PL biosynthesis. The ca. 4 kb gene cassette was amplified by PCR from template CMch37 [35] which contained the oriT-tet-trpE box. The primers PhLET10 and PhLET11 that were used in the PCR contained three homology arms in total (H1, H2 and H3). Homology arms 1 and 2 were used to enable the integration of the oriT-tet-trpE box into Cos10-4D08, via double homologous recombination (Figure 3.2)

**Figure 3.2:** Red/ET Cloning step 1.

Integration of oriT-tet-trpE box into cosmid 10-4D08 via double homologous recombination resulted in the construct CPhl9. H1, H2 and H3: homology arms 1, 2, and 3

Homology arm 3 and the Scal restriction site were needed for the subsequent cloning steps. Selection for correct recombinant mutants was carried out on LB-agar containing tetracycline (12.5 µg/ml) and apramycin (60 µg/ml). To confirm the individual cloning steps, each of the constructs generated by recombination was digested with a diagnostic set of restriction enzymes (Figure 3.5).
3.1.2.3 Cloning step 2: introduction of the zeocin resistance gene at the 3’ end

In the next cloning step we aimed to reconstruct the whole phenalinolactone biosynthetic gene cluster on one expression construct. Therefore, construct CPhl9 was digested with Scal to generate a linear fragment which could then be recombined with the target construct Cos3-1O12 containing the remaining part of the cluster. As the whole Scal restriction digest was to be used for transformation, it was formally possible that there could be still some undigested construct CPhl9 in the transformation mixture. Selection on tetracycline (oriT-tet-trpE box) and apramycin (target molecule) would then result in mutants carrying the constructs CPhl9 and Cos3-1O12, without any recombination taking place. To avoid this problem, we inserted the zeocin resistance gene in Cos3-1O12 to create CPhl7. The introduction of a new resistance gene enabled selection on tetracycline and zeocin, forcing homologous recombination and avoiding replication of the undigested construct CPhl9 in the mutants. The zeocin resistance gene was inserted into Cos3-1O12 at the opposite end of the gene cluster (downstream of plaZ), an alteration which simultaneously deleted orf6 which was not needed for phenalinolactone biosynthesis.

The ca. 600 kb zeocin cassette was amplified by PCR, again using MCh37 as template. Two homology arms (H4 and H5) were introduced with primers PhiET6 and PhiET7. Double homologous recombination via these homology arms enabled the insertion of the zeocin cassette into Cos3-1O12. Selection for correct recombinant mutants was carried out on LB-agar containing zeocin (25 µg/ml) and apramycin (60 µg/ml). To confirm the cloning step, the
construct CPhl7 generated by recombination was digested with a diagnostic set of restriction enzymes (Figure 3.5).

3.1.2.4 Cloning step 3: stitching of the pla biosynthetic gene cluster

Construct CPhl9 was digested with ScaI to generate a linear 15 kb fragment. A ScaI site had been inserted into the construct with the primer PhlET10 in the first cloning step described in chapter 3.1.2.2. A second ScaI restriction site was already present in the overlapping region of the cosmids (Figure 3.4).

This 15 kb fragment was then transformed into *E. coli* GB2005/pSC101-BAD-\(\gamma\beta\alpha\)-amp in order to achieve recombination with CPhl7. Recombination of the linear fragment with CPhl7 containing the missing end of the gene cluster resulted in the final construct CPhl8 containing the complete phenalinolactone pathway flanked by the oriT-tet-trpE box at the 5′-end and the zeocin resistance gene at the 3′-end of the cluster (Figure 3.4). Selection for correct recombinant mutants was carried out on LB-agar containing zeocin (25 µg/ml) and tetracycline (12.5 µg/ml). The individual cloning steps were confirmed by restriction analysis with the enzymes EcoRI, BglII and PvuII, as shown in Figure 3.5.
Results

Figure 3.5: Restriction analysis of the Red/ET recombination constructs.
Lane 1: Cos3-1O12, 2: CPhl7, 3: CPhl8, 4: CPhl10, 5: CPhl9, 6: Cos10-4D08. Lanes 1 and 6 show the restriction pattern of the starting constructs Cos3-1O12 and Cos10-4D08. Lane 2 shows the restriction fragments of CPhl7 into which the zeocin resistance gene was inserted, and lane 5 shows the restriction pattern of CPhl9 with the introduced oriT-tet-trpE box. Comparing the recombination constructs with the starting constructs enabled us to confirm the identity of all the constructs. Lane 8 shows the final construct CPhl8 containing all the fragments combined from CPhl7 and CPhl9. Lane 4 shows the restriction pattern of the construct CPhl10, which contains a promoter exchange. All restriction patterns showed the expected fragments.

In summary, using this cloning strategy, we were able to reconstitute the complete phenalinolactone biosynthetic pathway on one expression construct containing the pOJ436 backbone. At this point, the expression construct was ready for transformation into various heterologous host strains.

3.1.3 Transformation of the biosynthetic gene cluster into heterologous host strains
Two genera of bacteria, Pseudomonas and Streptomyces, were selected as hosts, as they had already been proven as suitable for heterologous expression of natural products [79].

3.1.3.1 Triparental conjugation of the expression construct CPhl8 into pseudomonads
The final expression construct CPhl8 was transferred into Pseudomonas putida KT2240 by triparental mating. Exconjugants were screened on PMM agar containing tetracycline (12.5 µg/ml) for the selection of P. putida::CPhl8 mutants. The transformation efficiency was very high, yielding several thousand tetracycline-resistant exconjugants. Single colonies were used for inoculation of liquid cultures.
3.1.3.2 Conjugation of the expression construct CPhl8 into streptomycetes

The expression construct CPhl8 carried the vector backbone of the *E. coli/Streptomyces* shuttle vector pOJ436, enabling integration at the chromosomal *attP* site in streptomycetes genomes. Three different *Streptomyces* strains were used for transformation: *Streptomyces lividans* TK24, *Streptomyces coelicolor* A3(2) and *Streptomyces coelicolor* M512. The construct was transformed using biparental conjugation, and selection was carried out at 30 °C on SM agar plates containing apramycin (60 µg/ml) and nalidixic acid (25 µg/ml). In all cases, the transformation efficiency was very high, yielding several thousands apramycin-resistant exconjugants. After 3–5 days, single colonies were transferred to fresh SM agar plates containing apramycin and nalidixic acid. When sporulation was observed, the colonies were used to inoculate liquid cultures.

3.1.3.3 Verification of the mutants

After transformation of the expression construct CPhl8 into pseudomonads and streptomycetes, many antibiotic-resistant colonies were obtained in all cases. To verify the integration of the constructs into the genome of the different strains, primers apra_for and apra_rev were designed to amplify the apramycin resistance gene from the vector backbone of CPhl8. For the verification of the *Pseudomonas* mutants, a colony PCR was performed with 2 randomly chosen colonies. 1 µl of an overnight culture was used as template for the PCR reaction. The *Pseudomonas putida* KT2240 wild type strain was used as a negative control, while DNA from the expression construct CPhl8 was used as a positive control. PCR products with the expected size (777 bp) were obtained for all mutants and for the positive control, whereas the negative control did not yield a PCR product (Figure 3.6). These result indicated the successful integration of the expression construct into the *Pseudomonas putida* genome.

![Figure 3.6: Verification of *P. putida* KT2240::CPhl8 mutants by colony PCR.](image)

Amplification of the apramycin resistance gene. Lane 1: *P. putida* KT2240::CPhl8#1; Lane 2: *P. putida* KT2240::CPhl8#2; Lane 3: *P. putida* WT; Lane 4: CPhl8 DNA.
For the verification of the *Streptomyces* mutants, genomic DNA was isolated from the mutants and from the wild type strain. A PCR was then performed using the apramycin primers to amplify the apramycin resistance gene from the vector backbone. Figure 3.7 exemplifies one PCR reaction using four randomly chosen *S. lividans* TK24 mutants and wild type as template. The expected PCR product with a size of 777 bp was present in all mutants and in the positive control (CPhl8 DNA), but not in the wild type. PCR reactions were performed for all mutants. Integration of the expression construct into the genome of the different *Streptomyces* strains was confirmed in all mutants.

![Figure 3.7: Verification of *S. lividans* TK24::CPhl8 mutants by PCR. Amplification of the apramycin resistance gene. Lane 1: *S. lividans* TK24 WT; Lanes 2–5: *S. lividans* TK24::CPhl8 mutants #1–4; Lane 6: CPhl8 DNA.]

3.1.4 Heterologous expression of the phenalinolactone biosynthetic gene cluster

In order to heterologously express the phenalinolactone pathway in pseudomonads and streptomycetes, cultivation conditions had to be established. Different media and temperatures were tested, as described below.

3.1.4.1 Cultivation of pseudomonads

Two mutants of *P. putida* KT2240::CPhl8 were inoculated into a 50 ml culture LB medium containing 1 ml XAD and tetracycline (12.5 µg/ml). In parallel, the wild type *P. putida* KT2240 was inoculated into a 50 ml culture LB containing 1 ml XAD without antibiotic. Cultivation was carried out at 16 °C and 30 °C for 3 days. The cultures were then harvested by centrifugation and the XAD was extracted with MeOH/acetone and submitted to HPLC-MS analysis.
3.1.4.2 Cultivation of streptomycetes

*S. lividans* TK24::CPhl8, *S. coelicolor* A3(2)::CPhl8 and *S. coelicolor* M512::CPhl8 were grown in 100 ml NL111 production medium containing apramycin (60 µg/ml) at 28 °C for 6–8 days. In parallel, the corresponding wild type strains as well as the natural phenalinolactone producer Tü6071 were grown in production medium without antibiotic. The cultures were harvested by centrifugation and the supernatant was extracted with ethyl acetate and submitted to HPLC-MS analysis.

3.1.4.3 HPLC-MS analysis

3.1.4.3.1 Production profile of the natural producer *Streptomyces Tü6071*

Initially the production profile of an extract of the natural producer Tü6071 was analyzed by HPLC-MS to establish HPLC conditions for the detection of phenalinolactones. Dürr et al. [80] demonstrated previously that production of phenalinolactones in NL111 production medium produced the highest yields. The primary metabolites produced during cultivation in NL111 medium were phenalinolactones (PLs) A and D, whereas PLs B and C were not detected [80] (Figure 3.8).

![Chemical structures of phenalinolactones A, B, C and D.](image)

Figure 3.8: Chemical structures of phenalinolactones A, B, C and D.

Figure 3.9 shows the typical production profile of the natural producer *Streptomyces Tü6071*. Phenalinolactones were identified by comparison to the retention times (rt) and the MS data of authentic standards. Analysis was performed in negative ionization mode enabling detection of the phenalinolactone A and D parent ions \( m/z \ [M–H]^– = 714.3 \) (RT: 13.8 min) and \( m/z \ [M–H]^– = 698.3 \) (18.1 min), respectively).
Figure 3.9: HPLC-MS analysis of the natural producer *Streptomyces* Tü6071.

a) Base peak chromatogram (BPC) in negative ionization mode within a mass range of 100–1100 B) Extracted ion chromatogram (EIC) of PL A with a mass of $m/z \ [M-H]- = 714.3$ c) Extracted ion chromatogram (EIC) of PL D with a mass of $m/z \ [M-H]- = 698.3$.

Fragmentation of the parent ions leads to a characteristic fragmentation pattern (MS/MS fingerprint) that could be used to unambiguously identify PL A and PL D. Figure 3.10 shows the detected parent ions of PL A and PL D, and the corresponding MS/MS fragmentation patterns. The MS/MS fragmentation pattern of PL A exhibited characteristic fragmentation ions with masses $m/z \ [M-H]- = 582.7, 600.5, 642.4, 670.4$ and 686.4. The corresponding fragments for PL D were $m/z \ [M-H]- = 584.7, 626.6, 654.4, 670.4$. On the basis of these fragmentation patterns, it was possible to clearly identify phenalinolactones in the extracts of the heterologous host strains.
3.1.4.3.2 Production profile of the Pseudomonas mutants

The basepeak chromatograms were analyzed for the presence of masses corresponding to PL A and PL D, but neither compound was detected in the extracts. Therefore, a new, higher selectivity and sensitivity HPLC-MS method was used (developed by Daniel Krug). This method detected the CID (collision-induced dissociation) fragmentation of compounds with masses 714.5 and 698.3 eluting between retention times of 12–14 min and 18–20 min, and was validated by analysis of the phenalinolactone A and D standards. However, analysis using this new method still failed to reveal the presence of phenalinolactones A and D.

Figure 3.10: ESI-MS/MS fragmentation patterns of PL A and PL D.

a) Parent ion of PL A with a mass of $m/z \ [M–H]^{-} = 714.3$ detected at a retention time of 13.5 min. b) MS/MS fragmentation pattern of PL A showing characteristic fragmentation ions. c) Parent ion of PL D with a mass of $m/z \ [M–H]^{-} = 698.3$ detected at a retention time of 18.3 min. d) MS/MS fragmentation pattern of PL D showing characteristic fragmentation ions.
3.1.4.3.3 Production profile of the *Streptomyces* mutants

We next investigated the extracts of the *Streptomyces* mutants, using the previously-described method. Phenalinolactones A and D were detected in extracts of the natural producer *Streptomyces* Tü6071, and in the recombinant *S. lividans* and *S. coelicolor* strains harboring the pla gene cluster. As expected no phenalinolactones were present in extracts of the corresponding wild type *S. lividans* and *S. coelicolor* strains. This result clearly demonstrated the successful heterologous production of the terpene glycosides (Figure 3.11).

Figure 3.11: HPLC/MS analysis of streptomycetes mutants carrying the PL pathway.

A) BPC of natural phenalinolactone producer strain *S. sp.* Tü6071, *S. lividans* TK24 WT and *S. lividans::CPhl8*. B) Natural producer strain *S. sp.* Tü6071, *S. coelicolor* A3(2) WT and *S. coelicolor::CPhl8*. C) Natural producer strain *S. sp.* Tü6071, *S. coelicolor* M512 WT and *S. coelicolor::CPhl8*. The chromatograms show the negative ions within a mass range of 100–1100 in auto MS<sup>n</sup> mode. CID fragmentation between retention times of 12–14 min and 16–18 min was monitored for the masses 714.5 and 698.3, corresponding to the phenalinolactone A and D standards. The section in each chromatogram between 12–18 min is shown in expanded form on the right. PL A and/or D were detected in the extracts of the recombinant strains.
3.1.4.4 Optimization of heterologous PL production in *Pseudomonas*

3.1.4.4.1 Promoter exchange

As there was no PL detected in extracts of the *Pseudomonas* mutants harboring the *pla* biosynthetic pathway, we decided to exchange one of the native *Streptomyces* promoters against the inducible Pm-promoter. Importantly, the Pm-promoter had already been validated during the heterologous expression of the myxochromide biosynthetic pathway [35]. As the cluster is composed of 11 different transcription units (operons), we intended to introduce the Pm-promoter at a position which would drive transcription of the genes for the core biosynthetic machinery. Based on biosynthetic considerations, enzymes PlaT1–T6 comprise the machinery to biosynthesize the phenalinolactones scaffold [66]. The first precursor of PL biosynthesis is geranylgeranyl diphosphate (GGDP), which is generated by PlaT4 from 4 monomers of isopentenyl pyrophosphate (IPP) (see Discussion Figure 4.2). PlaT6 and PlaT5 are expected to be involved in IPP production. So these first three enzymes should be able to catalyze the formation of GGDP. PlaT1-PlaT3 then catalyze the formation of PL precursor 1 and 2 (shown in Figure 4.2, Discussion part). Unfortunately, the corresponding genes (*plaT1–3* and *plaT4–6*) are located in different transcription units and exhibit different orientations (Figure 3.12). Therefore, we introduced the promoter in front of the gene *plaM2* where the initial genes for biosynthesis are located (*plaT4–6*), expecting that overexpression of the early genes in the pathway would lead to the initiation of PL production or to the production of GGDP or PL precursors. For this, a cassette containing the Pm-promoter was amplified by PCR using CMch37 [35] as template, the *xylS* gene and the chloramphenicol resistance gene, as well as 50 bp homology arms to enable double homologous recombination. This cassette was introduced by Red/ET recombineering and exconjugants were selected on LB agar containing chloramphenicol (cm) (34 µg/ml) and apramycin (60 µg/ml).

![Figure 3.12: Introduction of the inducible Pm-promoter in front of plaM2 by Red/ET recombineering.](image)
The new construct CPhl10 was verified by restriction analysis (Figure 3.5) and subsequently transformed into the *P. putida* KT2240 by triparental conjugation. Integration of the construct into the chromosome was confirmed by colony PCR analysis, amplifying the apramycin resistance gene as described above (3.1.2.3.). Expression was induced by adding 1 mM toluic acid to the cultures after 2 h of cultivation. We expected at least that GGDP would be expressed when the core biosynthetic genes (*plaT4-6*) were under the control of the *Pm*-promoter. However, HPLC-MS analysis still did not reveal production of PL precursors or PL production in extracts of the *Pseudomonas* mutants.

### 3.1.4.5 Optimization of heterologous PL production in *Streptomyces*

#### 3.1.4.5.1 Changing cultivation conditions

Production of PL in the heterologous *Streptomyces* strains was quantified using authentic PL standards. Production under standard conditions (NL111 medium at 28 °C; underlined in grey, Table 3.1) was much lower in the heterologous host strains (maximum 0.13 µg/L in *S. lividans* TK24) than in the native host *S. Tü6071* (Table 3.1). In an attempt to optimize production in the heterologous hosts, different cultivation conditions (additional media and temperatures) were tested (Table 3.1), with growth at 28 °C. Production was highest under standard conditions in NL111 production medium (Table 3.1). In INA5 and SM medium, PL production was typically lower in the heterologous host strains than in the natural producer, while the same media yielded less phenalinolactones from the native host relative to cultivation in NL111 medium. Growth in NBG medium completely abolished production by all strains. These results indicated that PL production in NL111 medium was optimal relative to cultivation in INA5, SM and NBG media.
Table 3.1: Phenalinolactone production by *Streptomyces* Tü6071, *S. lividans* TK24, *S. coelicolor* A3(2) and *S. coelicolor* M512 during cultivation in different media at 28 °C. Highest production yield underlined in grey.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Production of phenalinolactone (µg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NL111</td>
</tr>
<tr>
<td><em>S. Tü6071</em></td>
<td>500</td>
</tr>
<tr>
<td><em>S. lividans</em> TK24::CPhl8</td>
<td>0.13</td>
</tr>
<tr>
<td><em>S. coelicolor</em> A3(2)::CPhl8</td>
<td>0.01</td>
</tr>
<tr>
<td><em>S. coelicolor</em> M512::CPhl8</td>
<td>0.5</td>
</tr>
</tbody>
</table>

We also evaluated cultivation at 37 °C. Under these conditions, phenalinolactone production by *S. lividans* was increased 100-fold compared to growth at 28 °C (Table 3.2). However PL production by the *S. coelicolor* mutants was completely abolished at this higher temperature.

Table 3.2: Phenalinolactone production by *Streptomyces* Tü6071, *S. lividans* TK24, *S. coelicolor* A3(2) and *S. coelicolor* M512 during cultivation in NL111 medium at 28 °C and 37 °C.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Production of phenalinolactone (µg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>28 °C</td>
</tr>
<tr>
<td><em>S. Tü6071</em></td>
<td>500</td>
</tr>
<tr>
<td><em>S. lividans</em> TK24::CPhl8</td>
<td>0.13</td>
</tr>
<tr>
<td><em>S. coelicolor</em> A3(2)::CPhl8</td>
<td>0.01</td>
</tr>
<tr>
<td><em>S. coelicolor</em> M512::CPhl8</td>
<td>0.5</td>
</tr>
</tbody>
</table>

3.1.4.5.2 Precursor feeding

We also attempted to improve production by supplementation with a putative biosynthetic precursor. Because the core terpenoid backbone in phenalinolactone biosynthesis is derived from isoprenoid building blocks [66], we administered mevalonolactone to the culture medium. Mevalonolactone is the lactonized form of mevalonate, a direct precursor in the mevalonate pathway for the construction of isoprenoid units [74]. Most *Streptomyces* strains employ the non-mevalonate pathway for the formation of isoprenoids, but some strains
additionally utilize the mevalonate pathway to produce terpenoid antibiotics [74]. However, feeding of 1 mM mevalonolactone did not increase phenalinolactone production in *Streptomyces* sp. Tü6071 nor in the heterologous host strains.

3.1.4.5.3 Promoter exchange

We also evaluated the effect of exchanging a number of the native promoters against the strong constitutive *ermE* promoter [81]. We intended to introduce the promoter at the same position (between *plaO2* and *plaM2*) as in the construct for the PL expression in the pseudomonads (3.1.4.4.1), as this transcriptional unit contained the genes for the core biosynthetic gene cluster. We expected that at least GGDP or precursor 1 or 2 should be expressed (Figure 4.2, Discussion part). For this, we constructed a bidirectional promoter so that transcription in both transcription units containing the core genes would be driven (Figure 3.13). We therefore amplified the ampicillin (amp) resistance gene with primers containing 50 bp homology arms to enable homologous recombination. The *ermE* promoter was also introduced using the primers, as it is only ca. 60 bp in size. The amplified cassette was introduced by Red/ET recombineering, and recombinant mutants were selected on LB agar containing amp (100 µg/ml) and apra (60 µg/ml). The new construct CPhl19 was verified by restriction analysis, and subsequently transformed into the three heterologous host strains. In each case, integration of construct CPhl19 into the genome was verified by PCR, as described earlier (3.1.3.3.).

![Figure 3.13: Introduction of two *ermE* promoters between *plaO2* and *plaM2* by Red/ET recombineering.](image)
HPLC-MS analysis of the extracts from the mutants revealed that neither PL nor any PL precursors were produced in the mutants. Instead PL production was completely abolished.

3.1.5 Genetic modification of the *pla* biosynthetic gene cluster

The established heterologous expression system enabled attempts to engineer the PL biosynthetic pathway to generate novel analogs. In theory, using Red/ET recombineering, we could perform single or multiple in-frame deletions at any desired position within the expression construct. As proof-of-principle, we aimed to modify PL biosynthesis by deletion of selected genes from the CPhl8 expression construct.

3.1.5.1 Deletion of *plaA6*

The first aim was to modify the *pla* gene cluster by deletion of the glycosyl transferase-encoding gene *plaA6* from the CPhl8 expression construct, to generate a phenalinolactone derivative lacking the L-amicetose moiety. For this, a kanamycin resistance cassette was amplified by PCR to incorporate flanking flippase (FLP) recombinase target sites (FRT sites), as well as 50 bp homology arms to enable double homologous recombination (Figure 3.14). After replacement of the target gene (*plaA6*) by the selection marker (*kan*) using Red/ET recombineering, the *kan* gene was excised from the expression construct by FLP-recombinase-catalyzed site-specific recombination, to create a markerless in-frame deletion. The modified *pla* gene cluster was again verified by restriction analysis. The construct CPhl18 was subsequently transformed into the three heterologous *Streptomyces* strains by conjugation, and integration of the construct into the chromosome was confirmed by PCR analysis by amplifying the apramycin resistance gene, as described earlier.
Figure 3.14: Deletion of the glycosyl transferase encoding gene plaA6.

Red/ET recombineering was used to replace plaA6 with the kanamycin resistance cassette which was flanked by FRT recognition sites. Incubation with FLP recombinase at 37 °C led to a markerless deletion of plaA6 in the pla cluster.

3.1.5.2 Deletion of plaP2 and plaP5

By deletion of the acyltransferase encoding gene plaP2, we aimed to generate a derivative lacking the 5-methylpyrrole-2-carboxylic acid moiety. The gene plaP5 is proposed to be responsible for the methylation of the pyrrole carboxylic acid [66], and thus deletion of this gene was expected to give an unmethylated PL derivative. To create these deletions, we exploited the Cre/lox system instead of the FLP/FRT system. Using the Cre system would enable us to do two deletions in parallel using the FRT system too. Therefore a kanamycin resistance cassette was amplified by PCR to incorporate Cre recombinase target sites (loxP sites), as well as 50 bp homology arms to enable double homologous recombination (analogous to Figure 3.14). After replacement of the target gene (plaP2 or plaP5) by the selection marker (kan) using Red/ET recombineering, the kan gene was excised from the expression construct by Cre-recombinase-catalyzed site-specific recombination, to create a markerless in-frame deletion. The modified pla gene cluster was subsequently transformed into three heterologous host strains by conjugation, and integration of the construct (CPhl14 (ΔplaP2) and CPhl16 (ΔplaP5) into the chromosome was confirmed by PCR analysis by amplifying the ampramycin resistance gene, as described earlier.
3.1.5.3 Verification of the new PL derivatives by high-resolution Orbitrap MS analysis

Nine mutant strains were obtained after conjugation with the different constructs listed in Table 3.3. The production profiles of the mutant strains were analyzed by HPLC-MS, and compared to that of the respective wild type strains, and the natural phenalinolactone producer *Streptomyces* Tü6071. The expected masses of the new derivatives are also listed in Table 3.3.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Modification in the biosynthetic pathway</th>
<th>Heterologous host</th>
<th>Sum formula of expected PL derivatives</th>
<th>Expected mass [M–H]–</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPhl14 ΔplaP2</td>
<td>S. lividans TK24</td>
<td>C37H50NO11</td>
<td>684.3</td>
<td></td>
</tr>
<tr>
<td>CPhl14 ΔplaP2</td>
<td>S. coelicolor A3(2)</td>
<td>C37H50NO11</td>
<td>684.3</td>
<td></td>
</tr>
<tr>
<td>CPhl14 ΔplaP2</td>
<td>S. coelicolor M512</td>
<td>C37H50NO11</td>
<td>684.3</td>
<td></td>
</tr>
<tr>
<td>CPhl16 ΔplaP5</td>
<td>S. lividans TK24</td>
<td>C25H35O8</td>
<td>463.3</td>
<td></td>
</tr>
<tr>
<td>CPhl16 ΔplaP5</td>
<td>S. coelicolor A3(2)</td>
<td>C25H35O8</td>
<td>463.3</td>
<td></td>
</tr>
<tr>
<td>CPhl16 ΔplaP5</td>
<td>S. coelicolor M512</td>
<td>C25H35O8</td>
<td>463.3</td>
<td></td>
</tr>
<tr>
<td>CPhl18 ΔplaA6</td>
<td>S. lividans TK24</td>
<td>C31H46NO9</td>
<td>570.3</td>
<td></td>
</tr>
<tr>
<td>CPhl18 ΔplaA6</td>
<td>S. coelicolor A3(2)</td>
<td>C31H46NO9</td>
<td>570.3</td>
<td></td>
</tr>
<tr>
<td>CPhl18 ΔplaA6</td>
<td>S. coelicolor M512</td>
<td>C31H46NO9</td>
<td>570.3</td>
<td></td>
</tr>
</tbody>
</table>

As expected, phenalinolactones A and D (m/z [M–H]– = 714.5 at retention time 13.8 min and m/z [M–H]– = 698.5 at 18.1 min) were detected in the extract of the natural producer, but not in extracts of the wild type heterologous hosts. In the extract of the strain *S. coelicolor* M512::CPhl18, we detected a new compound designated as PL E (m/z [M–H]– = 570.3 at 13.0 min) which differed from PL D by 128 atomic mass units, consistent with the anticipated loss of the 4-O-methyl-L-amicetose moiety (Figure 3.15). MS² analysis confirmed that the new compound had a similar fragmentation pattern to the PLs (shown in Figure 3.16).
A) BPC of parental strain *S. coelicolor* M512 and B) *S. coelicolor* M512::CPhl18 (plaA6 deletion). A new compound, designated PL E, was detected in the extract of *S. coelicolor* M512::CPhl18 at a retention time of 13.0 min. C) Fragmentation pattern of PL E (molecular ion, $m/z [M–H]– = 570.3$).

The compound was not detected in extracts of the *S. lividans*:CPhl18 and *S. coelicolor*:CPhl18 mutants, nor in the extracts of the other heterologous host strains carrying the constructs CPhl14 and CPhl16. No expected derivatives could be detected in the extracts of the mutants carrying constructs CPhl14 and CPhl16 instead PL production was completely abolished in these mutants.

To verify the identity of PL E from *S. coelicolor* M512::CPhl18, we analyzed the compound by high resolution LC-coupled Fourier transform-Orbitrap MS. This analysis confirmed the expected elemental composition of C$_{31}$H$_{40}$O$_9$N$_1$ (measured $m/z [M–H]– = 570.26733$; calculated $m/z [M–H]– = 570.27086$; $\Delta = –3.52166$ amu). Additional structural information was obtained using tandem-MS and subsequent comparison of the MS$^2$ fingerprints with those obtained for the glycosylated compound PL D. The principal collision induced dissociation (CID)-fragmentations for the standard substance PL D and the new derivative PL E were assigned as summarized in Table 3.4. Fragmentation of the terpenoid backbone as well as of the sugar and the pyrrole-carboxylic acid groups was not observed in the MS$^2$
spectra of PL D, indicating that despite the loss of the sugar group, the fragmentation pattern of phenalinolactones D and E should be very similar. Indeed, comparison of the fragmentation pattern of both molecules showed that they exhibited the same set of peaks, but offset by a constant mass difference corresponding to L-amicetose (Table 3.5 and Figure 3.16). The peaks are derived from the parent molecules by expulsion of CO, CO₂ and H₂O, although the precise locations of these losses remain to be elucidated. The expulsion of C₄H₂O₄, which occurred from both compounds, can be attributed to the γ-butyrolactone. These data strongly suggest that the PL E is the desired unglycosylated derivative of PL D.

Figure 3.16: Verification of the unglycosylated derivative PL E by LTQ-Orbitrap MS analysis.

Accurate mass determination of the standard substance PL D (A) and the unglycosylated compound PL E (B) with high resolution LC-coupled Orbitrap-MS yielded molecular ions of $m/z$ [M–H]⁻ = 698.35504 and $m/z$ [M–H]⁻ = 570.26733, respectively, suggestive of the elemental formulas C₃₈H₅₂O₁₁N₁ and C₃₁H₄₀O₉N₁. Additional support for structure assignment was provided by MS² studies, which showed that PL D and PL E have the same fragmentation patterns, but with the peaks offset by a constant mass corresponding to the L-amicetose moiety missing in PL E.
Table 3.4:  Principal fragments of the standard substance PL D and the unglycosylated derivative PL E

<table>
<thead>
<tr>
<th>Accurate mass of fragments</th>
<th>Formula proposal and calculated mass</th>
<th>deviation ΔmDa</th>
<th>Neutral loss</th>
<th>Δm/z</th>
<th>Group</th>
<th>Theor.</th>
<th>ΔmDa</th>
</tr>
</thead>
<tbody>
<tr>
<td>PL D</td>
<td>670.35986 C₆H₁₂O₅N₇ = 670.35967</td>
<td>0.19</td>
<td>698.35504</td>
<td>27.995 8</td>
<td>CO</td>
<td>27.99437</td>
<td>0.001</td>
</tr>
<tr>
<td>698.35504</td>
<td>654.36444 C₅H₁₀O₄N₇ = 654.36476</td>
<td>0.32</td>
<td>698.35504</td>
<td>43.9966</td>
<td>CO₂</td>
<td>43.98928</td>
<td>0.0013</td>
</tr>
<tr>
<td>C₆H₁₂O₅N₇</td>
<td>626.36981 C₆H₁₂O₅N₇ = 626.36984</td>
<td>0.3</td>
<td>670.35986</td>
<td>43.9905</td>
<td>CO₂</td>
<td>43.98928</td>
<td>0.0008</td>
</tr>
<tr>
<td>698.35459</td>
<td>584.35931 C₅H₁₀O₄N₇ = 584.35928</td>
<td>0.34</td>
<td>698.35504</td>
<td>113.99573</td>
<td>C₇H₅O₃</td>
<td>113.99476</td>
<td>0.001</td>
</tr>
<tr>
<td>0.45613 mDa</td>
<td>566.34863 C₅H₁₀O₄N₇ = 566.34871</td>
<td>0.38</td>
<td>584.35931</td>
<td>18.01068</td>
<td>H₂O</td>
<td>18.01002</td>
<td>0.0007</td>
</tr>
<tr>
<td>PL E</td>
<td>542.27325 C₅H₁₀O₅N₇ = 542.27594</td>
<td>2.59</td>
<td>570.26733</td>
<td>27.99408</td>
<td>CO</td>
<td>27.99437</td>
<td>0.0003</td>
</tr>
<tr>
<td>570.26733</td>
<td>526.27832 C₅H₁₀O₅N₇ = 526.28103</td>
<td>2.71</td>
<td>570.26733</td>
<td>43.98901</td>
<td>CO₂</td>
<td>43.98928</td>
<td>0.0003</td>
</tr>
<tr>
<td>C₅H₁₀O₅N₇</td>
<td>498.28391 C₅H₁₀O₅N₇ = 498.28611</td>
<td>2.21</td>
<td>542.27325</td>
<td>43.98934</td>
<td>CO₂</td>
<td>43.98928</td>
<td>0.0001</td>
</tr>
<tr>
<td>570.27086</td>
<td>456.27301 C₅H₁₀O₅N₇ = 456.27555</td>
<td>2.54</td>
<td>570.26733</td>
<td>113.99432</td>
<td>C₇H₅O₃</td>
<td>113.99476</td>
<td>0.0004</td>
</tr>
<tr>
<td>-3.52166 mDa</td>
<td>438.26291 C₅H₁₀O₅N₇ = 438.26428</td>
<td>1.37</td>
<td>456.27301</td>
<td>18.0161</td>
<td>H₂O</td>
<td>18.01002</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
3.2 Identification of the GE81112 biosynthetic gene cluster

3.2.1 Goals

As the GE81112 biosynthetic gene cluster had not been identified, the first objective was to clone the cluster in order to allow studies of the underlying biosynthesis. (For this, we aimed to generate a cosmid library based on the E. coli/Streptomyces shuttle vector pOJ436 [10]). We then hoped to use in vitro protein expression and targeted gene inactivation to obtain deeper insights into the biosynthesis. To support this work, we aimed to establish genetic manipulation methods for the strain, starting with optimization of production and cultivation conditions for the GE81112 compounds.

3.2.2 Production profile of Streptomyces 14386

The GE81112 producing Streptomyces strain 14386 was provided by the company KtedoGen (Italy). Initial experiments for cultivation optimization to produce GE compounds in shake flasks were carried out by Margherita Sosio (KtedoGen). Sosio tested two different media, INA5 and T6, for both cultivation of the strain and GE production. Precultures were grown in V6 medium for 96 h, and then used to inoculate INA 5 and T6 production medium (1:100). The GE81112 production results are shown in Table 3.5.

<table>
<thead>
<tr>
<th>Medium/cultivation time</th>
<th>LC/MS [M+H]^+</th>
</tr>
</thead>
<tbody>
<tr>
<td>INA 5 (96 h)</td>
<td>ND</td>
</tr>
<tr>
<td>INA 5 (140 h)</td>
<td>ND</td>
</tr>
<tr>
<td>T6 (96 h)</td>
<td>ND</td>
</tr>
<tr>
<td>T6 (140 h)</td>
<td>644, 659</td>
</tr>
</tbody>
</table>

ND: Not Detected

Detection of GE81112 factor A and B by LC/MS was only possible after cultivation in T6 medium for 140 h. These results showed that T6 (cultivation time: 140 h) is the best medium identified to date for production of GE compounds. To reproduce these results in our lab, we cultivated the strain under the same conditions and analyzed the extracts for the presence of the GE metabolites. Overall, our results were nearly identical to those obtained by KtedoGen. For example, during cultivation in INA5 medium, GE compounds could never detected by
MS. During cultivation in T6 medium for 140 h, low amounts of the main GE compound B (\(m/z\) [M+H]+ = 659.22953) were detected by high resolution MS analysis on an Orbitrap LTQ mass spectrometer with the sum formula C24H36O10N10Cl1 (see Figure 3.17).

Despite trying several different cultivation conditions, production of the GE compounds remained low or even undetectable by MS. Due to the poor yields, it was not possible to obtain an MS/MS spectrum of the substances from the culture extract. These findings are in full agreement with the results obtained by KdetoGen. Although production was low in T6 medium detection of the GE81112 B compound was reliable so that we used these cultivation conditions for future experiments.

During analysis of the culture extracts from S. 14386, it became apparent that the strain was producing additional metabolites. These substances first attracted attention because one of them (659) had nearly the same mass by HPLC-MS analysis as the corresponding GE81112 compound B and also incorporated chlorine as shown by the isotopic pattern (Figure 3.18).
Figure 3.18: HPLC/MS analysis of an extract of strain S. 14386.

A) Base peak chromatogram (BPC) with extracted ion chromatograms (EIC) showing molecular ions of \( m/z \ [M+H]^+ = 631.3, 645.3, \) and \( 659.3 \). B) Mass spectrum of the three substances showing isotopic patterns characteristic of chlorinated compounds.

A detailed analysis of the extract revealed three likely related substances with masses of \( m/z \ [M+H]^+ = 631.3, 645.3, \) and \( 659.3 \), which all showed an isotopic pattern which is typical for chlorinated substances (Figure 3.18 B)). To obtain accurate mass information, the compounds were analyzed at high resolution using an Orbitrap LTQ mass spectrometer. Determination of the exact masses and the corresponding elemental formulae (Figure 3.19) confirmed that the compounds differed by 14 mass units from each other. Correspondingly, they share the same overall fragmentation patterns, but with peaks offset from each other by 14 mass units.
Figure 3.19: Analysis by high resolution MS of the substances with molecular masses 631, 645, 659.

The determined accurate masses and molecular formulae are shown. A) Pseudo-molecular ion of \([\text{M}+\text{H}]^+=631.3\), B) fragmentation pattern of molecular ion \([\text{M}+\text{H}]^+=631.3\), C) Pseudo-molecular ion of \([\text{M}+\text{H}]^+=645.3\), D) fragmentation pattern of molecular ion \([\text{M}+\text{H}]^+=645.3\), E) Pseudo-molecular ion of \([\text{M}+\text{H}]^+=659.3\) F) fragmentation pattern of molecular ion \([\text{M}+\text{H}]^+=659.3\)

This analysis suggested that the substances are related to the known GE series of compounds, as they are chlorinated and have similar elemental formulas; however, as the fragmentation
patterns differed, the compounds are not identical but remain interesting to investigate. Finally the MS analysis of the production profile revealed that we could now detect the main GE81112 factor B compound. In the next step we aimed to generate a cosmid library to allow us to identify the corresponding biosynthetic gene cluster.

3.2.3 **Generation of a cosmid DNA library**

One strategy for cloning and identifying a gene cluster is to generate a gene library in *E. coli.* A gene library is a population of organisms, each of which carries a DNA molecule that was inserted into a cloning (e.g. cosmid) vector. Ideally, all of the cloned DNA molecules represent the entire genome of the organism meaning that the desired biosynthetic gene cluster should be also present there. Knowing the size of the genome of the organisms from which one would generate the library one can calculate the approximately number of cosmids containing the gene cluster. When we assume that the average size of a *Streptomyces* genome is 9 Mbp and a cosmid vector can carry ca. 40 kb of insert DNA then we need to generate ca. 225 cosmid clones to clone the whole genome. To get a 10-fold coverage of the genome 2250 cosmid clones need to be obtained. As biosynthetic gene clusters are normally between 30 and 60 kb approximately 30-60 cosmids are expected to harbor parts of the gene cluster. The cosmids containing the desired gene cluster can then be identified by screening of the cosmid library with labeled probes (Figure 3.20).
Figure 3.20: General overview of the strategy for generation of a cosmid library.

Genomic DNA of the desired organism is isolated and cloned into cosmid vectors. The desired cosmid(s) containing the whole biosynthetic gene cluster or portions of it can then be identified with a labeled probe. A common cosmid vector is the *E. coli/Streptomyces* shuttle vector pOJ436 [10].

As a cosmid vector, we used the *E. coli/Streptomyces* shuttle vector pOJ436 [10]. Genomic DNA from the GE81112 producing strain was isolated using the salting out procedure (described in section 2.10.1). The isolated DNA was then partially digested with *Sau*3A to obtain fragments 30–35 kb in size. Fragments of the correct size were pooled and used for ligation into the *Pvu*II/*Bam*HI-digested vector pOJ436. Figure 3.21 shows partially-digested DNA generated by using different amounts of enzymes. Fractions 6 and 7 were chosen for the ligation, as they contained the greatest proportion of fragments between 35–40 kb. The two fractions were pooled together and ligation was carried out as described in section 2.16.1. To determine the phage titer different amounts of infected *E. coli* were plated out yielding different amounts of cosmid clones (see section 2.16.1).
We investigated whether 19 randomly chosen cosmid clones contained the correct insert size by digestion with BamHI (Figure 3.22). In Figure 3.22 one can see that most of the clones showed different fragments larger than 10 kb indicating that most of the analyzed clones contained an insert with an average size of 30–40 kb. Based on this information, we created the cosmid library containing 2304 cosmid clones in order to reach a 10-fold coverage of the genome of S. 14386 (section 2.16.2).

The cosmid library was then ready for hybridization with a suitable probe to identify the GE81112 biosynthetic gene cluster.

3.2.3.1 Retrobiosynthetic analysis of GE81112 to design suitable probes
To find suitable probes for the hybridization of the cosmid library, we applied a “retrobiosynthetic” strategy which allowed us to predict some aspects of the genetic composition of the gene cluster from the structure of the natural product. As described previously, GE81112s were assumed to derive from a non-ribosomal peptide synthetase. It is known that for these systems there is a strong correlation between compound structure and biosynthetic organization (even for some domains in non-linear systems) [64]. For example, it
is possible to predict with some degree of confidence, the building blocks incorporated into
the structure and therefore the number of modules and the complement of domains in the
NRPS, as well as which functionalities are likely to arise by post assembly-line processing.
Thus, according to the molecular structure of GE81112, we were able to propose a hypothesis
for the complement of genes that was likely to be present in the GE81112 gene cluster.
Assuming that the pathway will be co-linear (see section 1.4), the biosynthetic gene cluster
should incorporate four NRPS chain extension modules which are responsible for activating
and introducing the four observed amino acids (Figure 3.23). Each of the modules was
expected to incorporate the three standard NRPS domains, C, A and PCP. We also anticipated
that a halogenase would be involved in the chlorination of all three GE81112 compounds.

\[
\begin{align*}
\text{GE81112} & \\
\text{GE factor A} & R=0 \quad R_1=H \\
\text{GE factor B} & R=0 \quad R_1=\text{NH}_2 \\
\text{GE factor B1} & R=\text{NH} \quad R_1=\text{NH}_2
\end{align*}
\]

Figure 3.23: Structures of GE81112 metabolites.
The different building blocks were expected to be derived from different amino acids and are
marked in different colors. Red: piperolic acid, green: ornithine/glutamine, blue: histidine,
purple: histidine

We also hypothesized that the starter unit for GE81112 biosynthesis would be piperolic acid,
or its hydroxylated derivative (shown in red in Figure 3.23). As the stereochemistry of the
GE81112 compounds has not been solved yet, we did not know if L- or D-piperolic acid would
be incorporated. On the basis of published precedent [82], however, we assumed that the
piperolic acid would be formed from lysine via the action of a lysine cyclodeaminase (Figure
3.24).
This proposal is in accordance with the *in vitro* characterization of RapL, a cyclodeaminase, which has been found in the rapamycin biosynthetic gene cluster [82]. Other cyclodeaminases have been identified in the friulimicin [83], the virginiamycin S [84] and in the tubulysin biosynthetic gene clusters [85]. The *in vitro* reaction has also been reconstituted for LipE from the friulimicin gene cluster [83] and TubZ from the tubulysin biosynthetic gene cluster (Y. Chai, A. Sandmann, unpublished results). The pipecolate moiety in GE81112 differs from those in rapamycin and tubulysin, in that it is hydroxylated at position 3. This step was anticipated to be catalyzed by a $\beta$-hydroxylase, but the timing (i.e. before or after incorporation) could not be predicted. Another interesting feature of the structures is the amino acid which is incorporated second – aminohydroxypentanoic acid (building block 2, marked in green) – which we expected to derive from glutamine/glutamate or ornithine. We also hypothesized that the third and fourth amino acid building blocks (indicated in blue and purple) are likely to derive from histidine.

### 3.2.4 Design of the probes

On the basis of these predictions, we were able to design suitable probes for the screening of the cosmid library. For this, it is desirable to design a probe which is specific for an unusual feature of the cluster; as such a probe is likely to improve the chances of finding the correct cosmid. As described above, we predicted that a cyclodeaminase should be present in the genome. As such cyclodeaminases are relatively rare in biosynthetic pathways [86], this gene was judged to be a good candidate for trying to locate the GE81112 biosynthesis gene cluster in the genome of *S.* 14386. We therefore designed a probe based on cyclodeaminase sequences. To create the probe, we used specific primers to amplify two known cyclodeaminase genes – *tubZ* [85] (primers: TubZ_up and TubZ_down (section 2.7.2.2.)) from the tubulysin cluster and *rapL* [87] from the rapamycin cluster (primers: RapL_up and...
RapL_down (2.7.2.2). The amplificates were pooled and used to probe the cosmid library at low stringency. To check if the probe worked, we hybridized the genomic DNA of S. 14386 (digested with different enzymes) with the probe which yielded several distinct hybridization signals, indicating that the cosmids containing cyclodeaminase genes were likely present, and therefore that the library was likely to contain the desired cluster.

As an alternative approach, we designed an additional probe to identify NRPS genes. For this, we used degenerate NRPS primers (RevA3 and PSLGG) (2.7.2.2.) which had already been employed successfully to identify NRPS genes in myxobacteria [88-90]. These primers amplify A domains between the structural regions A3 and A6 [91], where the amino acid binding pocket is located, giving a 1.3 kb PCR fragment. The PCR was carried out using a step-wise gradient with different annealing temperatures (46–52 °C) using the S. 14386 DNA as template. Several of the obtained PCR fragments were cloned into PCR 2.1 TOPO and sequenced. Different A domain binding pockets were obtained after sequencing (presumably specific for the incorporation of valine, threonine, glutamine) indicating that the probes had successfully amplified various A domains. As it was not possible to correlate the substrate specificities of the A domains to the GE81112 structure, we were unable to select a single A domain as a probe, and therefore decided to use the whole PCR reaction containing all of the amplified A domains. This approach was expected to ensure a high diversity of A domain hybridization so that as many as possible A domains would hybridize.

3.2.5 Identification of an unknown NRPS biosynthetic gene cluster

3.2.5.1 Identification of two cosmids containing NRPS genes

Using both the cyclodeaminase and A domain probes at low stringency, we identified several cosmids which hybridized with one or both of the probes (Figure 3.25). Ten of these cosmids were then chosen for a more detailed analysis. A PCR with the degenerate NRPS primers was performed to verify the hybridization results, yielding the expected 1.3 kb fragment in each case. PCR with the cyclodeaminase primers was not successful; this result was expected, however, as the primers used were not degenerate, and so would not have been expected to target the particular cyclodeaminase gene(s) present in the S. 14386 genome. The DNA of the ten cosmids was isolated, and digestion with different enzymes was performed to evaluate the restriction patterns and to check if similar restriction fragments for different cosmids were obtained. This result would indicate that the same DNA sequence (biosynthetic gene cluster) was present on the cosmids.
Two of the cosmids showed similar restriction patterns, indicating that they contained overlapping sequence. End sequencing (Primers T4 and T7 (Table 2.12)) revealed that one of the cosmids (AI6) contained NRPS sequences (Table 3.6), and so this cosmid and the overlapping one were sequenced in their entirety.

<table>
<thead>
<tr>
<th>Cosmid</th>
<th>Homology (Name, Strain, Database no.)</th>
<th>Identity</th>
<th>Score</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD10_T4</td>
<td>Polyketide synthase type I, <em>Streptomyces aizunensis</em>, gi/62737770/AAX98186186.1</td>
<td>40%</td>
<td>37.4 bits</td>
<td>0.26</td>
</tr>
<tr>
<td>FD10_T7</td>
<td>Glutamate synthase, small subunit, <em>Streptomyces avermitilis</em>, gi/29832732/ref/NP827366.1</td>
<td>89%</td>
<td>228 bits</td>
<td>2e-79</td>
</tr>
<tr>
<td>AI6_T4</td>
<td>Hypothetical protein SAV2451, <em>Streptomyces avermitilis</em> MA-4680, gi/29828993/ref/NP823627.1</td>
<td>83%</td>
<td>268 bits</td>
<td>3e-70</td>
</tr>
<tr>
<td>AI6_T7</td>
<td>Peptide synthetase 2, <em>Streptomyces filamentosus</em> gi/60650533/gb/AAX31558.1</td>
<td>51%</td>
<td>304 bits</td>
<td>6e-81</td>
</tr>
</tbody>
</table>

### 3.2.5.2 Annotation of the unknown biosynthetic gene cluster

The two cosmids were sequenced at the HZI (Helmholtz Center for Infection Research) in Braunschweig, in the Department of Genome Analysis. The obtained sequence was then analyzed for the presence of putative open reading frames (ORFs) with FramePlot 2.3.2 [92], and preliminary functional assignments of individual ORFs were made by comparison of the deduced gene products with proteins of known function in the BLAST database. Sequence analysis of cosmids AI6 and FD10 revealed three ORFs for NRPS biosynthesis each containing four modules, which are all transcribed in the same direction (Figure 3.26 and Table 3.7). The modules include typical NRPS domains C, A and PCP domains, as well as E domains.
Figure 3.26: Sequencing results from cosmids FD10 and AI6.

Key: orange, NRPS and PKS genes; brown, regions which are shared between cosmids AI6 and FD10.

Table 3.7: List of NRPS genes from cosmids FD10 and AI6

<table>
<thead>
<tr>
<th>Orf</th>
<th>Proposed function</th>
<th>Identity/similarity to protein/origin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cosmid FD10</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nrps1_FD10</td>
<td>NRPS: A1 PCP1 C2 A2 PCP2 E2</td>
<td>52%, 64%; SACE4288, Saccharopolyspora erythraea</td>
</tr>
<tr>
<td>nrps2_FD10</td>
<td>NRPS: C3 A3 PCP3 C4 A4 PCP4 E4</td>
<td>47%, 58%; CDA peptide synthetase I, Streptomyces coelicolor A3(2)</td>
</tr>
<tr>
<td>nrps3_FD10</td>
<td>NRPS: C5</td>
<td>54%, 70%; pstC, Actinoplanes friuliensis</td>
</tr>
<tr>
<td><strong>Cosmid AI6</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nrps1_AI6</td>
<td>NRPS: A1 PCP1 E1</td>
<td>51%, 63%; SACE4288, Saccharopolyspora erythraea</td>
</tr>
<tr>
<td>nrps2_AI6</td>
<td>NRPS: C2 A2 PCP2 C1 A3 PCP3 E3</td>
<td>47%, 58%; CDA peptide synthetase I, Streptomyces coelicolor A3(2)</td>
</tr>
<tr>
<td>nrps3_AI6</td>
<td>NRPS: C4 A4</td>
<td>50%, 65%; pstC, Actinoplanes friuliensis</td>
</tr>
</tbody>
</table>

The *nrps1_FD10* gene from cosmid FD10 most likely starts with an ATG which is located 9 bp downstream of a putative ribosome binding site (RBS) GAGG. *nrps1_FD10* appears to be part of an operon with *nrps_FD10_2-3*. *nrps2_FD10* represents another NRPS with a putative RBS (GGAG) upstream from the start codon ATG. The subsequent gene *nrps3_FD10*
encodes a NRPS protein, which only contains a C-domain. \textit{nrps3\_FD10} starts with GTG, and a RBS (GGAG) is located 8 bp upstream. On the basis of sequence analysis and restriction mapping, it was expected that the two cosmids would contain portions of the same biosynthetic gene cluster, and would therefore share an overlapping region. However, annotation of cosmid AI6 revealed that only the \textit{nrps2\_AI6} gene showed exactly the same sequence as \textit{nrps2\_FD10}. The other two genes, \textit{nrps1\_AI6} and \textit{nrps3\_AI6}, exhibited some sequence identity to the corresponding FD10 genes, but were not identical along their full lengths. Specifically, \textit{nrps1\_AI6} is only half the size of \textit{nrps1\_FD10}, while only a portion of \textit{nrps3\_AI6} is present in \textit{nrps3\_FD10}. Additionally gene \textit{nrps3\_AI6} is interrupted at the end of the cosmid, so it is not clear how large this gene cluster actually is. Taken together, these observations indicate that the clusters might have arisen from gene duplication, as they are overall highly similar, though not identical.

At this point, it remained to establish which of the two clusters, if either, was responsible for the biosynthesis of GE81112. The presence of four modules in each gene cluster was consistent with the incorporation of four amino acids into the corresponding products, so no decision could be made on that basis. We therefore performed a detailed sequence analysis of the NRPS domains from each module.

\subsection*{3.2.5.3 Analysis of the NRPS domains}

\subsubsection{A domains}

NRPS include a specific adenylation domain to select each amino to be incorporated into the peptide product; the sequence of A domains in the NRPS dictates the primary structure of the peptide product. A domains contain 10 highly-conserved core motifs (A1—A10), where the region between motifs A3 and A6 forms the amino acid binding pocket \cite{91}. To determine the substrate specificity of the A domains in the putative GE81112 clusters, each of the A domains was aligned with A domains from the gramicidin, rapamycin and tubulysin synthetases. All ten core motifs were identified in each A domain from both cosmids, indicating that all A domains are functional.
The substrate specificity of A domains can be predicted using the specificity-conferring code of adenylation domains [93;94]. The code was developed by two different groups [93;94] using the crystal structure of the phenylalanine-activating A domain PheA from the gramicidin synthetase (GrsA) from *Bacillus brevis*. The crystal structure led to the identification of the 8 residues that form the substrate binding pocket. By comparing the sequence of PheA, the structure of which was known, with the sequence of 160 other adenylation domains, eight residues that (putatively) form the amino acid binding pockets were identified. The signature sequences for different substrate specificities were then determined from groups of domains activating the same substrate. This code now allows the
substrates of uncharacterized or unknown A domains to be predicted from the primary sequences. Nevertheless some limitations of the code should be noted. First, there is only a limited amount of sequence information available for A domains, and thus it is currently impossible to discern with confidence consensus specificity sequences for all amino acids. Furthermore, NRPS are not restricted to proteinogenic substrates and more than 300 are known to be incorporated but the codes currently available are mostly restricted to proteinogenic amino acids. A clear shortcoming of the available biochemical data is that it is limited mostly to A domains from *E. coli*, *Bacillus* sp. and *Pseudomonas* sp. Actinomycetes, including notably *Streptomyces*, are largely missing from the essential database.

In order to predict the substrate specificity of the FD10 and AI6 A domains from the sequenced cosmids, the A domains were aligned against the A domain from the gramicidin A synthetase (GrsA), and the 8 residues that form the amino acid binding pocket were identified (Figure 3.28; shown for A domains from cosmid FD10).

![Figure 3.28](http://www.tigr.org/jravel/nrps/ and http://www-ab.informatik.uni-tuebingen.de/toolbox/index.php?view=domainpred) which also allow substrate specificities to be predicted. Sequence analysis of the code residues from all of the A domains from both clusters allowed the prediction of substrate specificity in several cases (Table 3.8).
Table 3.8: Putative A domain 'code residues' within A domains from both clusters

<table>
<thead>
<tr>
<th>Position of the amino acid within the A domain</th>
<th>Activated amino acid prediction</th>
<th>Identity to protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>235 236 239 278 299 301 322 330</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD10_A1 D A V D F G T I</td>
<td>No hit</td>
<td></td>
</tr>
<tr>
<td>FD10_A2 D A L L I G A V</td>
<td>Phe/Trp 75%</td>
<td></td>
</tr>
<tr>
<td>FD10_A3 D F W S V G M V</td>
<td>Thr 87%</td>
<td></td>
</tr>
<tr>
<td>FD10_A4 D A A Q L G V I</td>
<td>No hit</td>
<td></td>
</tr>
<tr>
<td>AI6_A1 D A L L I G A V</td>
<td>Phe/Trp 75%</td>
<td></td>
</tr>
<tr>
<td>AI6_A2 D F W S V G M V</td>
<td>Thr 87%</td>
<td></td>
</tr>
<tr>
<td>AI6_A3 D A A Q L G V I</td>
<td>No hit</td>
<td></td>
</tr>
<tr>
<td>AI6_A4 D V F S V A V V</td>
<td>No hit</td>
<td></td>
</tr>
</tbody>
</table>

When identity to known A domains could not be found in the database for the 8 amino acid residues, no hit was predicted from the bioinformatics program. We also compared the A domains manually with the known A domains which were published by Challis et al. [94]. Despite the use of multiple methods no substrate specificity could be predicted for FD10_A1, FD10_A4, AI6_A3 and AI6_A4.

Consistent with the observation that both clusters exhibited an overall high level of sequence similarity, three of the A domains from each cluster showed the same apparent substrate specificity (FD10_A2 and AI6_A1, FD10_A3 and AI6_A2 and FD10_A4 and AI6A3). FD10_A2 (which showed the same specificity as AI6_A1) exhibited 75% identity to the specificity pocket of the TycB-M2 A domain in Bacillus brevis, which selects and activates L-phenylalanine or L-tryptophan. Sequence analysis of FD10_A3 (same specificity as AI6_A2) revealed 87% identity to the specificity pocket of the SyrE-M7 A domain in Pseudomonas syringae, which is involved in selection and activation of L-threonine. Analysis of FD10_A1, FD10_A4 and AI6_A3 and AI6_A4 A-domains revealed that the identified code residues did not allow a prediction of the specificity.
3.2.5.3.2 PCP, C and E domains

The PCP, C and E-domains were analyzed to determine whether or not they were likely to be functional, which would support the hypothesis that the clusters act to biosynthesize a tetrapeptide. The peptidyl carrier protein (PCP) (or thiolation (T) domain) is the site of substrate attachment during chain elongation [95]. These proteins show a conserved signature sequence LGx(HD)S(LI), where the central S is the point of covalent attachment of the Ppant prosthetic group. The signature sequence including the active serine was found in all PCP domains from both cosmids FD10 and AI6, indicating that the domains should be activated to participate in chain extension.

![PCP consensus motif](image)

**Figure 3.29: PCP consensus motif.**

The PCP domains from cosmid FD10 were compared with known PCP domains from the gramicidin (GrsA), surfactin (SrfA), tyrocidine (TycA) and calcium-dependent antibiotic (CDA) NRPSs. The amino acids marked in grey match the consensus motif. “x” can be any amino acid. The highly conserved serine residue (site of Ppant attachment) is indicated in black.

C domains are responsible for the condensation of two amino acids activated on adjacent modules. These domains incorporate seven highly-conserved core regions, all of which were found in the C domains in both cosmids FD10 and AI6. We also analyzed the epimerase domains, which are responsible for substrate epimerization (formation of D-amino acids). Again, the seven characteristic conserved motifs for all E domains could be identified in the E domains on both cosmids. These results demonstrate that, in principle, all modules are active and so each system should be able to generate a tetrapeptide. Thus, these results did not allow us to prove or exclude that one or both of the clusters are responsible for the biosynthesis of GE81112 and further experiments/analysis had to be done. Therefore we continued our analysis, by investigating the cyclodeaminase from FD10.
3.2.6 Analysis of the cyclodeaminase from cosmid FD10

We also expected that the GE81112 cluster would include a cyclodeaminase, providing L-pipeolic acid from L-lysine. BLAST sequence analysis of both cosmids revealed that only cosmid FD10 contained a cyclodeaminase sequence (The protein showed 62% identity to the TubZ protein from *Angiococcus disciformis*). To analyse the conserved regions of the protein which would indicate whether it was functional, we performed an alignment with three different cyclodeaminases TubZ [85], LipE [83] and RapL [87] (Figure 3.30).

![Alignment of the cyclodeaminase from cosmid FD10 with LipE, RapL and TubZ.](image)

No color: Non-similar amino acid residues, blue: conservative, green: block of similar, yellow: identical

Cyclodeaminases are typically 1000–1100 kb in size (corresponding to approximately 350 amino acid residues). However, the cyclodeaminase present on cosmid FD10 was only 465 kb
in size, corresponding to 155 amino acid residues. Thus, more than half of the protein was apparently missing, relative to previously-characterized cyclodeaminases.

3.2.6.1 Heterologous expression of cosmids FD10 and AI6

In principle, the most straightforward method to verify the identity of the gene clusters would have been to knockout one of the NRPS genes, and then to determine if production of GE81112 was abolished. We attempted to generate this knockout by amplifying an internal region from the nrps_2 gene, which was then cloned into the knockout vector pKC1132 [10], and introduced the vector into strain S. 14386 by conjugation. However, to date, the knockout has not been successful.

We therefore decided to use an alternative strategy to determine the identities of the cluster – heterologous expression of the biosynthetic gene clusters from both cosmids. The two cosmids were transformed into S. lividans TK24 and S. coelicolor A3(2) by conjugation. Exconjugants were verified by PCR (amplification of the apramycin resistance gene) and extracts were analyzed by high resolution Orbitrap-MS analysis (data not shown). We did not observe production of GE81112 in the extracts of the mutants harboring either cosmid AI6 or FD10. Instead, three new peaks with masses 631.2242 (retention time 3.86 min), 645.2389 (retention time 4.70 min) and 659.2545 (retention time 4.94 min) were detected in the extract of S. lividans TK24::FD10 (Figure 3.31), although the compounds were not present in S. lividans TK24 wild type or in S. lividans TK24::AI6.
Results

Figure 3.31: Analysis of *S. lividans*::FD10 extract by high-resolution MS.

A) Base peak chromatogram (BPC) of *S. lividans*::FD10 extract and extracted ion chromatograms (EICs), showing molecular ions and elemental formulas of B) m/z = 631.2242 [M+H]+, C) m/z = 645.2389 [M+H]+ and D) m/z = 659.2545 [M+H]+. E), F) and G) show the exact masses and elemental formulas of GE81112 factor A, B1 and B for comparison.

Detailed MS/MS analysis revealed that all 3 of the substances exhibited exactly the same fragmentation pattern as the new chlorinated substances found during analysis of the GE producer strain (section 3.2.2). This result led us to conclude that the three substances 631, 645 and 659, had been heterologously expressed in *S. lividans* from cosmid FD10. No production of the same metabolites was observed from cosmid AI6, which is likely explained by the fact that the cluster is incomplete. In addition, no other new metabolites were detected in the extract of *S. lividans* TK24::AI6.

These results indicated that the biosynthetic gene cluster from cosmid FD10 is not responsible for biosynthesis of GE81112, but for the new substances which might be related to the GE compounds. Therefore we aimed to purify the novel substances and to elucidate their structures.
3.2.7 Isolation and attempted structure elucidation of three chlorinated substances

In previous experiments, we observed that the desired compounds were mostly localized to the culture supernatant, while only traces of the metabolites were present in the cell pellets. Cultivation with XAD resin did not lead to an increase of production, and only small amounts of the compounds were detected in the extract of the XAD. Therefore the strain was grown without XAD, and the compounds were isolated from the supernatant. S. 14386 was grown in GE production medium for 6 days and harvested by centrifugation.

![Diagram of purification procedure]

**Figure 3.32:** Purification procedure for substances 631, 645 and 659.

The purification scheme is outlined in Figure 3.32. First, 13 L of the supernatant were extracted twice with the same volume of ethyl acetate. After the solvent was removed by evaporation, the extract was resuspended in 50 ml methanol and extracted with heptane for degreasing. The heptane and MeOH phases were subsequently analyzed by HPLC-MS.
Figure 3.33: HPLC-MS analysis of heptane and MeOH fractions.

Base peak chromatogram of heptane A) and MeOH B) phases, and extracted ion chromatograms of MeOH fraction showing molecular ions of C) \( m/z \ [M+H]^+ = 631.3 \), D) 645.3 and E) 659.3.

Only traces of the compounds could be detected in the heptane phase, while all three compounds were present in the MeOH phase (Figure 3.33). The methanol phase was therefore evaporated and resuspended in a smaller volume of methanol (2 ml). This sample was then fractionated by semi-preparative HPLC using a stepwise gradient (solvent A: H2O containing 0.01 % formic acid; solvent B: methanol containing 0.01% formic acid, 25%–90% B over 35 min, injection volume 250 µl). As the substances did not exhibit a significant UV signal, the HPLC was coupled to a mass spectrometer (Bruker HCT equipped with an ion trap mass spectrometer) for online monitoring of the masses. Using this method, the exact retention times of the compounds could be measured, allowing straightforward collection of the substances. Five fractions were collected and analyzed for the presence of the compounds (representative results are shown in Figure 3.34).
Figure 3.34: HPLC/MS analysis of different purified fractions after preparative HPLC.

Base peak chromatogram showing molecular ions with the masses $m/z$ [M+H]$^+$ = 631.3, 645.3 and 659.3.

Fractions 2, 4 and 5 contained the compounds 631, 645 and 659, respectively. However, contaminants remained in all of the fractions, so it was necessary to carry out a further purification step. MS analysis revealed that only substances with relatively low masses ($< m/z = 300$) were present in fraction 2 (containing 631). Fractions 4 (645) and 5 (659) also contained substances with masses similar to the desired compounds (in a range of $m/z = 600$). These observations prompted us to use size-exclusion chromatography for fraction 2 (631), followed by purification by isocratic preparative HPLC, while fractions 4 (645) and 5 (659) were directly submitted to separation by isocratic preparative HPLC. All 30 samples from each fraction were combined and the solvent evaporated. The extracts of fractions 4 (645) and 5 (659) were resuspended in 500 µl methanol.

The extract of fraction 2 (631) was resuspended in 1 ml of methanol and subsequently separated on a Sephadex LH20 column (28 h), yielding 48 fractions. Every second fraction was analyzed by HPLC/MS. Fractions 9–11 contained the compound and were combined. The solvent was evaporated and the extract was resuspended in 500 µl methanol. All three extracts were then purified by isocratic preparative HPLC (solvent A: H$_2$O containing 0.01 % formic acid; solvent B: acetonitrile containing 0.01% formic acid, 0–10 min 30% B isocratic, 10–20 min 30%–50% B, 20–25 min 50%–95% B, 25–27min 95% B, injection volume 200
μl). Monitoring was again carried out by mass spectrometry. After this separation step, essentially pure compound was obtained in each case (Figure 3.35)

![Figure 3.35: HPLC-MS analysis of different fractions after preparative HPLC.](image)

Base peak chromatograms and UV chromatograms after preparative HPLC showing molecular ions of A) \(m/z [M+H]^+ = 631.3\), B) \(m/z [M+H]^+ = 645.3\), and C) \(m/z [M+H]^+ = 659.3\).

The extracts were vacuum-dried by lyophilization and weighed. Compound 631 was obtained at a yield of 4.8 mg, compound 645 at 2.3 mg and compound 659, 1.2 mg. Structure elucidation of all three substances was then performed by NMR.

### 3.2.8 NMR structure elucidation

NMR structure elucidation was started with the compound 631 because purification yielded the highest amount (4.8 mg) of this compound. NMR was carried out using 1D (\(^1\)H, \(^13\)C) and 2D (\(^1\)H-\(^1\)H-COSY, HMBC and HSQC) NMR analysis in \([D_6]\)DMSO as well as in CD\(_3\)OD. Although several structural elements have been elucidated to date for compound 631 (Figure 3.36), it has not yet been possible to solve the complete structure.
Figure 3.36: Postulated structural elements of compound 631 identified by NMR spectroscopy. N/O means that N or O might be present at that specific position. R indicates that the functionality could not be identified.

The results (NMR interpretation was done by H. B. Bode) and is not discussed in detail here. The structure is likely to contain four amino acid-derived structural elements (nos. 1–4) including two ornithine moieties (3 and 4) which are probably acylated (H. B. Bode, personal communication). However, these structural elements could not be connected together due to missing correlations in the 2D-NMR experiments (HMBC, COSY). The three compounds 631, 645 and 659 differ in 14 atomic mass units each. To date, it is not clear at which positions these variations occur. Therefore, further experiments will be necessary to complete the structure elucidation.

3.2.9 Identification of the GE81112 biosynthetic gene cluster

3.2.9.1 Identification of a cosmid containing GE81112 biosynthetic genes

As the results described in section 3.2.5 led us to conclude that the identified gene clusters were not involved in the biosynthesis of GE81112, we developed a new strategy to identify the GE81112 gene cluster. The previous cyclodeaminase probe successfully identified a cyclodeaminase, but it was truncated, and therefore presumed to be inactive. However, as such pseudo-genomes often arise by gene duplication of an original active copy; we reasoned that there might be a second, complete cyclodeaminase gene in the genome, which should be functional. We therefore decided to generate a new specific cyclodeaminase probe based on the cyclodeaminase sequence found on cosmid FD10. In order to amplify the cyclodeaminase
gene, a PCR reaction was carried out using specific primers (Table 2.12, Cyclo Probe_for and Cyclo Probe_rev). The resulting 400 bp PCR product was labeled with digoxigenin and used to re-probe the cosmid library. To evaluate the likely success of this strategy, we first carried out a Southern Blot to check if the probe hybridized to more than one gene in the genome. Digestion of cosmid FD10 with \textit{PvuII} produced a 2.5 kb fragment which encompasses the cyclodeaminase sequence. Thus, if the FD10 cyclodeaminase were the only cyclodeaminase in the genome, we would have observed a single 2.5 kb band by Southern Blot.

For the Southern, genomic DNA of \textit{S. 14386} was isolated and digested with \textit{PvuII}. Hybridization was carried out at 42 °C under stringent conditions using the new cyclodeaminase probe. Southern Blot analysis showed the expected 2.5 kb band (Figure 3.37). However, two additional signals were detected at approximately 1 kb and 13 kb (Figure 3.37), suggesting that the probe hybridized a total of three times with genes in the genome (1 kb, 2.5 kb and 13 kb band). We assumed that the hybridization had been successful, as non-specific hybridization would likely have resulted in more signals. Thus, we concluded that the genome was likely to contain a second or even a third copy of the cyclodeaminase gene. As the new probe seemed to function well, we aimed to screen the cosmid library again with this probe to identify cosmids that had not hybridized with the old cyclodeaminase probe. Hybridization was again carried out at 42 °C (Figure 3.38).
The hybridization was successful, as only nine distinct signals were obtained, in addition to that with cosmid FD 10 (positive control). To verify the presence of cyclodeaminase sequence on these nine cosmids, we performed a PCR analysis on the cosmids using the primers employed to generate the probe. PCR products of ca. 400 bp were obtained for all cosmids and the control FD10, confirming in each case that the cyclodeaminase sequence was present. As we expected that NRPS sequence must also be present on the cosmid, we additionally performed a PCR using the degenerate NRPS primers RevA3 and PSLGG [88]. Unfortunately, no PCR products were obtained with these primers. One possible explanation for this result, however, is that the NRPS degenerate primers RevA3 and PSLGG were not optimal. Indeed, attempts by other members of the laboratory to use these primers to amplify A domains from Streptomyces genomes have not been successful, perhaps because they were originally designed for use with myxobacterial DNA.

We therefore designed new primers based on Streptomyces A domains using the CODEHOP software [96]; priming was targeted against the core A3 and A6 motifs. The new primers were then used to amplify A domains from the nine cosmids that produced products of the correct size in the cyclodeaminase PCR. This reaction resulted in PCR products from four of the nine cosmids. These fragments were cloned into pCR2.1-TOPO and sequenced. BLAST analysis revealed that all of the sequences showed high homology to A domains. We then analyzed the A domain sequences as described earlier, to predict the substrate specificity of each A domain (Table 3.9).
Table 3.9: Putative A domain ‘code residues’ within A domains from cosmids

<table>
<thead>
<tr>
<th>Position of the amino acid within the A domain</th>
<th>Activate amino acid</th>
<th>Identity to protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>235 236 239 278 299 301 322 330</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topo BG23</td>
<td>D A S Q V G E V</td>
<td>Asp/Asn 75%</td>
</tr>
<tr>
<td>Topo DB20</td>
<td>D A V D F G T I</td>
<td>No hit</td>
</tr>
<tr>
<td>Topo DK18</td>
<td>D A V D F G T I</td>
<td>No hit</td>
</tr>
<tr>
<td>Topo BI11</td>
<td>D V Q D I A H M</td>
<td>Pro/Pip 70%</td>
</tr>
</tbody>
</table>

This analysis led to a prediction that the A domain in cosmid BI11 would be specific for proline/pipecolic acid. This result was supported by BLAST analysis of the A domain which predicted homology to a “NRPS for pipecolate incorporation, *Streptomyces* sp. NRRL 30748”. We also directly compared the nonribosomal code of this A domain to those of known pipecolic acid-incorporating A domains from the rapamycin and tubulysin systems (Table 3.10). The code residues in each case were similar, supporting the probable specificity for pipecolic acid.

Table 3.10: Substrate specificity pocket of pipecolic acid incorporating A domains

<table>
<thead>
<tr>
<th>Position of the amino acid within the A domain</th>
<th>Activated amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>235 236 239 278 299 301 322 330</td>
<td></td>
</tr>
<tr>
<td>A domain RapP</td>
<td>D Y Q Y C G H L</td>
</tr>
<tr>
<td>A domain BI11</td>
<td>D V Q D I A H M</td>
</tr>
<tr>
<td>A domain TubB</td>
<td>D I Q Y I A Q V</td>
</tr>
</tbody>
</table>

Taken together, these data strongly suggested that cosmid BI11 contained the biosynthetic pathway for GE81112. Furthermore, as we anticipated that the NRPS would contain four modules (30–40 kb), it was possible that entire gene cluster would be located on a single cosmid. To verify this, we end sequenced the cosmid using primers T4 and T7 (Table 2.12.).
Table 3.11: End sequencing results of cosmid BI11 with T7 and T4 primers

<table>
<thead>
<tr>
<th>Cosmid</th>
<th>Homology</th>
<th>Identity</th>
<th>Score</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI11_T4</td>
<td>DNA binding protein, <em>Streptomyces coelicolor</em></td>
<td>36%</td>
<td>143</td>
<td>1e-32</td>
</tr>
<tr>
<td>BI11_T7</td>
<td>L-arginine beta-hydroxylase, <em>Streptomyces vinaceus</em></td>
<td>44%</td>
<td>331</td>
<td>4e-89</td>
</tr>
</tbody>
</table>

The T4 end exhibited homology to a DNA binding protein, while sequence at the T7 showed similarity to an arginine β-hydroxylase (Table 3.11). The DNA binding protein was not expected to have any function in the biosynthesis, and thus cosmid BI11 contained at least one end of the gene cluster. However, it was possible that the β-hydroxylase was involved in the biosynthesis, so we could not conclude that the entire cluster was located on the cosmid. Nonetheless, we sequenced BI11, as at least a significant portion of the cluster was likely to be present.

3.2.9.2 Identification of overlapping cosmids

To find a cosmid which overlapped with the 3’ (T7) end of the GE81112 biosynthetic gene cluster, we amplified a 1 kb fragment from the T7 end of cosmid BI11 to serve as a probe. The sequenced PCR product was then used as a template to carry out a PCR reaction using digoxigenin-labeled nucleotides. The cosmid library was then screened with the new probe, with hybridization at 42 °C.

Figure 3.39: Screening of the cosmid library to identify cosmids which overlapped the T7 end of cosmid BI11. Cosmid BI11 is labeled.
Seven new cosmids were identified, in addition to cosmid BI11 (marked in Figure 3.39). We initially investigated the cosmids by PCR analysis by amplifying an 800 bp fragment with the primers used to generate the probe (cyclo probe_for and cyclo probe_rev). All of the cosmids produced a positive signal. However, as some of the cosmids gave a similar restriction pattern to that of BI11, we excluded them from further analysis. We chose one cosmid (BA23) which showed the least similar fragmentation pattern and analyzed it further. To ensure that both cosmids (BI11 and BA23) contained similar portions of the cluster, and to determine whether the cloned DNA was co-linear with the S. 14386 chromosome, we performed a Southern Blot with both the cosmids and the wild type genomic DNA. In theory, the probe that derives from the T7 end of cosmid BI11 should produce identical signals when hybridized to the cosmids and the wild type DNA. As hybridization was observed to the same fragments in all cases (Figure 3.40), the cosmids were verified to be co-linear with the genome.

![Southern Blot analysis of cosmid BI11 and the overlapping cosmid.](image)

**Figure 3.40:** Southern Blot analysis of cosmid BI11 and the overlapping cosmid.

Schematic arrangement of cosmids BI11 and BA23. To verify the colinearity between the cosmids and the genome, the cosmids and the genomic DNA were digested with the same enzymes (NotI, SalI and PvuII). The probe was designed against the T7 end of cosmid BI11, as shown. 1: Cosmid BI11, 2: cosmid BA23, 3: genomic DNA, L: Ladder

### 3.2.9.3 Annotation

The two cosmids BI11 and BA23 were sequenced in the HZI. The obtained sequence was analyzed for the presence of putative open reading frames (*orfs*) with FramePlot 2.3.2 [92], and preliminary functional assignments of individual *orfs* were made by comparison of the deduced gene products with proteins of known function in the BLAST database (Table 3.12).
### Table 3.12: BLAST results for genes on the cosmids BI11 and BA23

<table>
<thead>
<tr>
<th>Gene</th>
<th>Homology</th>
<th>Identity</th>
<th>Score</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf1</td>
<td>putative partitioning protein ParA, <em>Nocardia farcinica</em> IFM 10152</td>
<td>63/222 (29%)</td>
<td>77.0</td>
<td>8e-13</td>
</tr>
<tr>
<td>orf2</td>
<td>hypothetical protein SAV_5443 <em>Streptomyces avermitilis</em> MA-4680</td>
<td>263/478 (55%)</td>
<td>470</td>
<td>1e-130</td>
</tr>
<tr>
<td>orf3</td>
<td>FtsK/SpoIIE family protein <em>Streptomyces avermitilis</em> MA-4680</td>
<td>802/1099 (72%)</td>
<td>1565</td>
<td>0</td>
</tr>
<tr>
<td>orf4</td>
<td>hypothetical protein SAV_6995 <em>Streptomyces avermitilis</em> MA-4680</td>
<td>118/142 (83%)</td>
<td>244</td>
<td>2e-63</td>
</tr>
<tr>
<td>orf5</td>
<td>serine/threonine protein kinase <em>Streptomyces avermitilis</em> MA-4680</td>
<td>249/381 (65%)</td>
<td>395</td>
<td>5e-108</td>
</tr>
<tr>
<td>geA</td>
<td>thioesterase type II <em>Streptomyces atroolivaceus</em></td>
<td>89/230 (38%)</td>
<td>142</td>
<td>4e-32</td>
</tr>
<tr>
<td>geB</td>
<td>putative ABC-type multidrug transport system ATPase and permease component, <em>Streptomyces griseus</em></td>
<td>269/563 (47%)</td>
<td>449</td>
<td>4e-124</td>
</tr>
<tr>
<td>geC</td>
<td>putative ABC transporter ATPase and permease component, <em>Streptomyces griseus</em></td>
<td>291/575 (50%)</td>
<td>533</td>
<td>1e-149</td>
</tr>
<tr>
<td>geD</td>
<td>TubZ protein <em>Angiococcus disciformis</em></td>
<td>182/346 (52%)</td>
<td>337</td>
<td>1e-90</td>
</tr>
<tr>
<td>geE</td>
<td>syringopeptin synthetase <em>Pseudomonas syringae</em></td>
<td>209/522 (40%)</td>
<td>325</td>
<td>7e-87</td>
</tr>
<tr>
<td>geF</td>
<td>putative L-proline 3-hydroxylase protein <em>Ralstonia solanacearum</em> GMI1000</td>
<td>81/253 (32%)</td>
<td>123</td>
<td>2e-6</td>
</tr>
<tr>
<td>geG</td>
<td>peptide synthetase protein, <em>Ralstonia solanacearum</em> GMI1000</td>
<td>765/2401 (31%)</td>
<td>857</td>
<td>0</td>
</tr>
<tr>
<td>geH</td>
<td>non-ribosomal peptide synthetase <em>Mycococcus xanthus</em> DK 1622</td>
<td>168/530 (31%)</td>
<td>194</td>
<td>2e-47</td>
</tr>
<tr>
<td>geI</td>
<td>L-arginine β-hydroxylase <em>Streptomyces vinaceus</em></td>
<td>144/322 (44%)</td>
<td>238</td>
<td>6e-61</td>
</tr>
<tr>
<td>Gene</td>
<td>Homology</td>
<td>Identity</td>
<td>Score</td>
<td>E-Value</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------------</td>
<td>------------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>geJ</td>
<td>bacitracin synthetase 3; BacC, <em>Bacillus licheniformis</em></td>
<td>223/598</td>
<td>368</td>
<td>1e-99</td>
</tr>
<tr>
<td>geK</td>
<td>ribosome-associated GTPase</td>
<td>209/368</td>
<td>395</td>
<td>3e-108</td>
</tr>
<tr>
<td>geL</td>
<td>halogenase, <em>Microcystis aeruginosa</em></td>
<td>199/500</td>
<td>376</td>
<td>2e-102</td>
</tr>
<tr>
<td>geM</td>
<td>bacitracin synthetase 3; BacC, <em>Bacillus licheniformis</em></td>
<td>199/500</td>
<td>397</td>
<td>2e-108</td>
</tr>
<tr>
<td>geN</td>
<td>ChlK, thioesterase type I, <em>Streptomyces antibioticus</em></td>
<td>44/112</td>
<td>80.5</td>
<td>4e-14</td>
</tr>
<tr>
<td>orf6</td>
<td>3-oxoacid CoA-transferase subunit B, <em>Streptomyces avermitilis</em>, MA-4680</td>
<td>177/215</td>
<td>359</td>
<td>1e-97</td>
</tr>
<tr>
<td>orf7</td>
<td>protocatechu 3,4-dioxygenase α subunit, <em>Streptomyces avermitilis</em>, MA-4680</td>
<td>185/253</td>
<td>375</td>
<td>2e-102</td>
</tr>
<tr>
<td>orf8</td>
<td>protocatechu 3,4-dioxygenase β subunit, <em>Streptomyces avermitilis</em>, MA-4680</td>
<td>129/182</td>
<td>253</td>
<td>5e-66</td>
</tr>
<tr>
<td>orf9</td>
<td>3-carboxymuconate cycloisomerase, <em>Streptomyces sviceus</em> ATCC 29083</td>
<td>319/447</td>
<td>570</td>
<td>8e-161</td>
</tr>
<tr>
<td>orf10</td>
<td>deoxyribose-phosphate aldolase, <em>Streptomyces avermitilis</em> MA-4680</td>
<td>150/205</td>
<td>285</td>
<td>4e-75</td>
</tr>
<tr>
<td>orf11</td>
<td>arthrofactin synthetase/syringopeptin synthetase C-related non-ribosomal peptide synthetase module, <em>Bradyrhizobium sp. BTAi1</em></td>
<td>513/1186</td>
<td>777</td>
<td>0.0</td>
</tr>
<tr>
<td>orf12</td>
<td>peptide synthetase protein, <em>Ralstonia solanacearum</em> GMI1000</td>
<td>221/656</td>
<td>244</td>
<td>3e-62</td>
</tr>
<tr>
<td>orf13</td>
<td>putative ATP/GTP binding protein, <em>Streptomyces coelicolor</em> A3(2)</td>
<td>178/193</td>
<td>338</td>
<td>3e-91</td>
</tr>
</tbody>
</table>
Table 3.12: Continuation of Table 3.12

<table>
<thead>
<tr>
<th>Gene</th>
<th>Homology</th>
<th>Identity</th>
<th>Score</th>
<th>E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf14</td>
<td>integrin-like protein, <em>Streptomyces avermitilis</em> MA-4680</td>
<td>200/468 (42%)</td>
<td>299</td>
<td>3e-79</td>
</tr>
<tr>
<td>orf15</td>
<td>two-component system sensor kinase, <em>Streptomyces coelicolor</em> A3(2)</td>
<td>107/218 (49%)</td>
<td>179</td>
<td>2e-43</td>
</tr>
<tr>
<td>orf16</td>
<td>polyketide synthase type I, <em>Streptomyces aizunensis</em></td>
<td>228/435 (52%)</td>
<td>359</td>
<td>4e-97</td>
</tr>
</tbody>
</table>

Based on the chemical structure of GE81112 we expected a NRPS pathway involving a cyclodeaminase and a halogenase. Annotation of the two cosmids revealed 30 *orfs* of which 14, designated *geA–N*, are postulated to be involved in the GE81112 biosynthetic pathway (Figure 3.41). The genes marked in black indicate NRPS encoding enzymes.

The sequenced region starts with the gene *orf1*, which shows homology to a putative partitioning protein ParA from *Nocardia farcinica* IFM 10152. This protein is not expected to be involved in the biosynthesis, as these proteins are known to be involved in bacterial mitosis [97]. Similarly, we excluded *orfs* 2 and 3, as they do not show homology to proteins postulated to be involved in the biosynthesis. *orf2* is a hypothetical protein with no assigned function, and *orf3* encodes for a FtsK/SpoIIIE family protein which belongs to the family of DNA translocases [98]. It is not possible, however, to conclusively identify the boundaries of a gene cluster based solely on the deduced functions of the gene products. Therefore, in future, it will be necessary to carry out gene knockouts to confirm the proposed cluster ends.

![Figure 3.41: GE81112 biosynthetic gene cluster on cosmids BI11 and BA23.](image)

Key: black, NRPS biosynthetic genes; light grey, modification genes; dark grey, *orfs* not involved in the biosynthesis.

The first gene that we predicted to be involved in the biosynthesis is gene *geA*, which encodes a type II thioesterase (TEII). TEIIs are present in many NRPS and PKS systems, where they
perform crucial proof-reading functions by hydrolyzing aberrant substrates from the respective carrier protein domains. In modular PKS, type II TEs are proposed to release from the ACP domains acyl groups that have been produced by aberrant decarboxylation of chain extender units [99]. In NRPS systems, they are proposed to play two functions: removal of acyl groups added to the PCP domains during post-translational priming of the apo proteins, and release of amino acids that have been loaded by mistake [100].

geA appears to be in an operon with genes geB–E. geB and geC encode proteins which show homology to ABC transporter systems. Many antibiotic-producing actinomycetes possess at least one ABC (ATP-binding cassette) transporter which forms part of the antibiotic biosynthetic pathway, and in most cases confers resistance to the drug in a heterologous host. In Type I ABC transporters, two genes are involved, one encoding a hydrophilic ATP-binding protein with one nucleotide-binding domain, and the other encoding a hydrophobic membrane protein. The same type of proteins have been found in the GE81112 cluster [101]. geD encodes the cyclodeaminase gene that was targeted in our second probing strategy. It shows 52% identity to TubZ, the cyclodeaminase from Angiococcus disciformis. This time the protein was not truncated as in cosmid FD10 and was proposed to be functional. The sequence of the truncated cyclodeaminase gene on cosmid FD10 is identical to a region within geD, consistent with the idea that the truncated form arose from gene duplication.

geE is the first gene that encodes for a NRPS protein, and most likely starts with a GTG. A putative RBS (GGAG) was found 7 bp upstream of the gene. The next gene geF is oriented in the opposite direction, and encodes a protein with homology to a putative L-proline 3-hydroxylase. This enzyme is expected to be responsible for the hydroxylation of the pipecolic acid. The following gene, geG, encodes another NRPS protein, and is the likely starting point of a new operon which includes genes geH and geI. geG has an ATG start codon and a putative RBS (GAAGG) is located 19 bp upstream from the start codon. geH encodes a third protein with homology to NRPSs, and again starts with a GTG. The last gene in this operon is geI, which exhibits homology to an L-arginine β-hydroxylase, which might be responsible for the hydroxylation of one of the incorporated histidines in GE81112. The last operon contains 5 genes (geJ–geN). This operon starts with a change in the transcription direction again. The operon begins with the gene geJ which encodes a NRPS enzyme. The next two genes, geK and geL, encode a GTPase protein and a halogenase, respectively; GeL may be responsible for the chlorination of the GE81112 histidine. The NRPS-encoding gene geM starts with a
TTG, and a RBS (AGGG) is located 6 bp upstream of the gene. The last gene of the operon is geN, which encodes a protein with homology to a type I thioesterase. A RBS (GAAA) is located 8 bp upstream of the gene which starts with an ATG. The involvement of orfs 6–16 can not yet be excluded.

**3.2.9.4 Analysis of NRPS domains**

The sequence analysis revealed 14 open reading frames that we predicted to be part of the gene cluster. Five of them showed homology to NRPS genes (Figure 3.41). The constituent domains were assigned using the PKS/NRPS predictor (http://www.tigr.org/jravel/nrps/) and confirmed by manual inspection with BLAST, as the prediction programs can omit some domains.

![Figure 3.42: The GE81112 biosynthetic gene cluster.](image)

In order to incorporate 4 amino acids into the GE81112 structure, we expected the synthetases to contain 4 modules for the biosynthesis of a tetrapeptide including a loading module followed by three condensation module with the standard C-A-PCP arrangement. However, the sequence analysis revealed that the cluster exhibits a highly nonlinear arrangement that did not fit to what we expected. Although the domain complement for the biosynthesis of a tetrapeptide could be identified one extra module consisting of an A-PCP didomain could be identified. Furthermore the modules showed a highly split arrangement (Figure 3.42). Thus, from the domain assignment alone, the overall order of subunits could not be discerned, and so we carried out further analysis of the domains themselves. We started with the analysis of the A domains, as we hoped that the predicted substrate specificities would give us a first hint about the order of the domains.

**3.2.9.4.1 Analysis of the A domains**

To determine the order of the A domains in the GE81112 cluster, we analyzed the substrate specificity of the A domains. The eight conserved motifs were identified for each A domain, and the specificity-conferring code (bioinformatics and manual prediction) was then used to
predict the substrate specificity. The first A domain, encoded by gene \textit{geE} (GeEA1) as a discrete protein, showed specificity for the incorporation of proline/pipecolic acid (see section 3.2.9.1). This observation made it a candidate for selection of the GE81112 starter unit. The second A-domain located on gene \textit{geG} (GeGA2) showed homology to ornithine/glutamine/asparagine-incorporating A domains. The third A domain on gene \textit{geG} (GeGA3) was predicted to be specific for the incorporation of tyrosine/tryptophane, and the two latter domains on \textit{geJ} (GeJA4) and on \textit{geM} (GeMA5) for histidine (Table 3.13).

<table>
<thead>
<tr>
<th>Position of the amino acid within the A domain</th>
<th>Predicted amino acid</th>
<th>Identity to protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>235 236 239 278 299 301 322 330</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GeEA1 D V Q Y I A Q V</td>
<td>Pro/Pip</td>
<td>70%</td>
</tr>
<tr>
<td>GeGA2 D A Y N L G L I</td>
<td>Orn/Gln/Asp</td>
<td>70%</td>
</tr>
<tr>
<td>GeGA3 D A V G V G E V</td>
<td>Tyr/Trp</td>
<td>70%</td>
</tr>
<tr>
<td>GeJA4 D S A S T A E V</td>
<td>His</td>
<td>70%</td>
</tr>
<tr>
<td>GeMA5 D S A L T A E V</td>
<td>His</td>
<td>70%</td>
</tr>
</tbody>
</table>

These results correlated well with our prediction that an ornithine and two histidines are incorporated in the the GE81112 metabolites, in addition to the pipecolic acid. However, it remained unclear why two PCP-A didomains (\textit{geM} and \textit{geJ}) are located within the cluster, although only one is required to give a total of four active modules. To check if any of the domains were inactive in order to account for the presence of the apparently superfluous module, we continued by analyzing the C and PCP domains.

3.2.9.4.2 Analysis of the C and PCP domains

The GE81112 cluster contains three C domains which were aligned with C domains from the rapamycin [102], the gramicidin [103] and the calcium-dependent antibiotic [104] clusters. The seven conserved regions were identified in all of the C domains. The same analysis was carried out for the PCP domains, revealing the signature sequence and active Ser residue, in all cases. These results demonstrate that in principle, all modules are active, so that it is unlikely that one of the modules is skipped.
3.2.9.4.3 Analysis of the TE domains

TE domains are characterized by a conserved signature sequence GxSxG [91]. There are two classes of TE domains known in NRPS systems. TEI domains are normally integrated into the multienzyme subunits, and located at the C-terminal end of the modules which are involved in adding the last amino acid to the linear peptides [91]. These TEs are responsible for cleavage of the linear peptide products. Type II TEs are normally discrete proteins that perform a crucial proof-reading function by hydrolyzing aberrant substrates from the respective carrier proteins [100;105]. We could identify two discrete TE domains (on genes geA and geN) in the GE81112 biosynthetic gene cluster which showed the conserved motif, indicating that they are functional enzymes. As they were both discrete proteins, it appeared as if they were both TEII enzymes. However, while BLAST analysis (see Table 3.12) revealed that the protein GeA showed high homology to a type II TE, protein GeN was similar to type I TEs. This finding was surprising, as GeN is a discrete protein like a type II TE, and not integrated into an NRPS as is typically found in bacterial NRPS systems.

3.2.9.4.4 Docking domain analysis

The analysis of all domains with the NRPS subunits revealed that they all contained the respective conserved regions, indicating that all proteins/domains are probably functional. In the vast majority of NRPS complexes, the modules are distributed over two or more NRPSs which have to interact selectively with each other to bring about the synthesis of a defined product. The molecular basis for the selective interaction between NRPSs are the so called ‘communication-mediating (COM)’ domains [106;107]. A donor COM domain (ComD) located at the C terminus of an aminoacyl- or peptidyl-donating NRPS and an acceptor COM Domain (ComA) located at the N terminus of the accepting partner NRPS form a matching set, required for the proper intermolecular interaction between adjacent modules. In contrast, ComD and ComA domains of nonpartner NRPSs are considered nonmatching, preventing false contact between enzymes [106]. The C-terminal COM domain, which consists of approximately 20–30 amino acids, most often follows a PCP or E domain, while the N-terminal COM domain, which typically precedes a C or HC domain, comprises 15–25 amino acids. Although the overall sequence similarity between COM domains is low, the amino acid composition is relatively uniform. The C-terminal COM domains incorporate a higher than average proportion of acidic amino acids, while N-terminal COM domains are biased toward
polar residues, suggesting that specificity is mediated largely by polar/or electrostatic interactions between key residues.

To obtain insight into how the different NRPS proteins from the GE81112 cluster might interact with each other to form a defined biosynthetic template, we analyzed the GE81112 NRPS proteins for putative COM domains. In contrast to linear NRPS, the GE81112 system is a nonlinear system in which many of the modules are split (Figure 3.43). Thus, we devised an NRPS pathway in which individual multienzymes interact to reconstitute standard C-A-PCP modules (for example GeE and GeH, Figure 3.43). This task was aided by the *in silico* prediction of the substrate specificity of the A domains which gave us some hints as to the ordering of the proteins within the pathway.

![Figure 3.43: GE81112 biosynthetic gene cluster.](image)

We predict that protein GeE (harboring the single A domain GeEA1 which is predicted to incorporate pipecolic acid) interacts with subunit GeH to form the loading module. Interaction of GeH with GeG would then form the next two modules (A domains GeGA2 (Orn/Gln) and GeGA3 (Trp/Tyr)). Protein GeG would then interact with either GeJ (A domain GeGA4 (His)) or GeM (A domain GeGA5 (His)) to form the last module. The protein GeN has a TEI function and is therefore proposed to hydrolyze the final tetrapeptide from the assembly line.

![Figure 3.44: Predicted interaction of NRPS enzymes from the GE81112 biosynthetic gene cluster.](image)

It is not clear if proteins GeG and GeJ or GeG and GeM interact with each other to form the last module.
In contrast to standard COM domains which normally facilitate the interaction between PCP and C domains located on separate multienzymes, intersubunit protein-protein interactions in the GE81112 system must reconstitute interfaces between C and A domains (between proteins GeH/GeG, and between GeG/GeJ or GeG/GeM, see Figure 3.44) to provide a functional pathway. To determine if we could identify COM domains between C and A domains, GeHC₁ and GeGA₂ (module 2), and GeGC₃ and GeJA₄ (module 4) or GeMA (module 5), we constructed a sequence alignment (attached in the appendix, Table 5.1) comprising putative modules 1–5. In this alignment, the sequences of modules 1–5 (Figure 3.44) were compared to standard C-A-PCP modules (modules from NRPS pathway from cosmid FD10). This analysis revealed large putative recognition sequences at the C terminus of the C domains and the N terminus of the A domains in modules 2 and 4 (Figure 3.45).

![Diagram of NRPS interaction and order of subunits in the GE81112 biosynthesis.](image)

Figure 3.45: Proposed NRPS interaction and order of subunits in the GE81112 biosynthesis.

However, these recognition sequences did not match the standard size of COM domains (20–30 amino acid residues). Instead, we found stretches of approximately 100 amino acid residues at each terminus; to reflect this difference, we’ve designated the sequences as ‘recognition elements’ instead of COM domains. Nonetheless, the C-terminal recognition elements incorporate a high proportion of acidic amino acids while the N-terminal linker domains are biased towards polar amino acids, providing the basis for electrostatic interactions as in classical COM domains (the recognition elements contained approximately the same proportion of acidic and polar amino acid residues as COM domains). Such recognition sequences are normally not found between C and A domains, indicating that in this special case, these sequences have a docking function, enabling the interaction between NRPS proteins.

At this stage, it remained unclear why the cluster contains an apparently superfluous fifth module. One hypothesis was that both bidomains GeJ and GeM can interact with the C domain of protein GeG to form the fourth module of the NRPS, incorporating histidine in either case. To evaluate this hypothesis, we aligned the sequences of proteins GeJ and GeM,
in order to evaluate the extent of their sequence similarity (Table 5.2, appendix). We reasoned that a high degree of sequence identity would provide evidence for the evolution of the two didomains by gene duplication, supporting the idea that they might both be able to interact with the C terminal C domain from GeG. However, the alignment showed that the proteins exhibit only 25% mutual sequence identity, consistent instead with a lack of shared partner specificity. Indeed, GeM lacks the putative recognition sequence for GeG which is present at the N-terminus of GeJ. Taken together, these data indicated that the didomain protein GeM was not derived from duplication of GeG, and should therefore have a specific function in the pathway (Figure 3.45). To provide further support for the proposals concerning domain order and functions of GeJ and GeM, we investigated the gene cluster using both gene inactivation and expression analysis of recombinant proteins.

3.2.10 Insertional mutagenesis of the GE81112 biosynthetic gene cluster

To first verify the identity of the gene cluster, we aimed to inactive the gene geE. For this, a knockout construct was designed by amplifying an internal fragment (660 bp) of the gene, which was then cloned into the knockout vector pKC1132 [10] leading to the construct pKC1132_BI11_PipA. Initial attempts to transform S. 14386 with the knockout plasmids pKC1132_FD10 and pKC1132_AI6 however, were unsuccessful, necessitating the development of a revised transformation method. The classical procedures for streptomycetes transformation are protoplast transformation and conjugation [10]. We tried both methods with the knockout vector, using a replicative vector pOJ446 [10] and an integrative vector pSET152 [10], as controls. The protoplast transformation method did not produce any exconjugants using the knockout vector or with the controls. The conjugation was then attempted with varying amounts of E. coli. In the standard procedure, a certain amount (normally 500 µl) of E. coli culture (OD₆₀₀ = 0.6) is mixed with an equal volume of Streptomyces cells. Using this approach, we only obtained a low number of colonies with the replicative and the integrative vector controls, and no colonies with the knockout construct (Table 3.14). We then tried 100-fold and 1000-fold more E. coli cells, but obtained the same result.
Table 3.14: Transformation efficiency of *S. 14386* by biparental conjugation

<table>
<thead>
<tr>
<th>Transformation Method</th>
<th>E. coli amount</th>
<th>E. coli amount</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5×10^8</td>
<td>50×10^9</td>
</tr>
<tr>
<td>Biparental conjugation</td>
<td>pKC1132_BI11_PipA replicative</td>
<td>pKC1132_BI11_PipA replicative</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>50–100</td>
<td>100–200</td>
</tr>
</tbody>
</table>

We next modified the procedure by mixing 5 mL of an overnight culture of *E. coli* which had a very high cell density, with 5 ml of *Streptomyces* cells. This change finally led to several exconjugants containing the knockout vector, indicating that the amount of *E. coli* is crucial for the conjugation efficiency with this strain. The obtained mutants were verified by PCR amplifying the ampramycin resistance gene as well as an internal region from the genomic DNA, proving the correct integration (primers: apra_for and apra_rev, lacZ1 and lacZ2 and BI11A1_pip_for and BI11A1_pip_rev) (section Table 2.12). The mutants were cultivated in production medium, as before. The extracts were then analyzed for the presence of the GE81112 compounds by high-resolution MS (Figure 3.46).
Figure 3.46: High resolution MS analysis of extracts of S. 14386 wild type and S. 14386::pKC1132_BI11_PipA mutant.

A) BPC of S. 14386 wild type extract. B) EIC of GE81112 factor B showing a molecular ion of the mass \(m/z = 659.22953\) [M+H]+ from S. 14386 wild type extract. C) EIC of GE81112 B showing the analysis for a molecular ion of the mass \(m/z = 659.22953\) [M+H]+ of S. 14386::pKC1132_BI11_PipA

This analysis clearly showed that production of GE81112 factor B (mass \(m/z [M+H]^+ = 659.22953\)) was abolished in the mutant, verifying the identity of the gene cluster. As an alternative approach we tried the heterologous expression of the cosmids, as described before 3.2.6.1. Heterologous expression couldn’t be observed from the cosmids. This was not surprising as they both didn’t contain the whole cluster.

3.2.11 Feeding experiments

Inspection of the GE81112s metabolite structures suggests that they incorporate modified versions of pipecolate, histidine and ornithine/glutamine. Thus, it was possible that the corresponding proteinogenic amino acids were incorporated and then modified, or alternatively that the modifications occurred prior to activation of the building blocks by the A domains. To investigate this question directly, we designed feeding experiments with labelled substrate. As the stereochemistry of GE81112 compounds has not yet been solved, we decided to feed separately both D- and L-pipecolic acid. In an independent experiment, we also fed \(^{13}\)C-labelled histidine. Unfortunately, we did not detect any incorporation of the
labeled amino acids into GE81112, a result which we attribute to the very low production yield of the GE81112 compounds. We also did not observe any incorporation of the labeled amino acids into the newly-discovered substances 631, 645 and 659. As these substances are produced in reasonable amounts and could be clearly detected in HPLC-MS, we could exclude production yield as the underlying problem. Thus, it is likely that none of the amino acids used in this experiment could be incorporated into the substances. These findings are consistent with the structure elucidation results, as to date, no structural elements similar to pipecolic acid or histidine have been found.

3.2.12 Biochemical characterization of the GE81112 biosynthetic gene cluster

An alternative approach to feeding experiments to evaluate the proposed biosynthesis is to express key enzymes involved in substrate biosynthesis or selection in recombinant form, and assay them in vitro. As we still had no experimental proof that the expected amino acids are incorporated from the A domains proposed in section 3.2.9.4.4 we tried to express three of the A domains in recombinant form in E. coli to assay them in vitro. Additionally the cyclodeaminase was chosen for recombinant expression to confirm its proposed function.

3.2.12.1 Expression and purification of the lysine cyclodeaminase GeD

Lysine cyclodeaminases have been shown in vitro to generate L-pipecolic acid from L-lysine [82;83;108] (and Y. Chai et al. unpublished data). The enzymes were obtained in recombinant form from E. coli [108] and in some cases it could be determined that the L-stereoisomer of pipecolic acid was formed. Based on this precedent, we intended to express the cyclodeaminase geD from the GE81112 cluster in E. coli as a C-terminal fusion to GST, and to test its function in an assay in vitro. For this, we amplified the open reading frame of geD from cosmid BI11 and cloned it as an EcoRI-NorI fragment into the expression vector pGEX-6P-1 (Figure 3.47).
The expression construct was sequenced, and then transformed into *E. coli* BL21 by electroporation. Cultivation was then carried out at both 16 °C and 30 °C. Analysis of 10 μl of the supernatant and the pellet using a 12% SDS polyacrylamide gel showed that most of the protein was present in the pellet, following growth at both 16 °C and 30 °C (Figure 3.48).

To attempt to obtain soluble protein, we tried expression in *E. coli* Rosetta BL21 (DE3) pLysS/RARE. In addition, we evaluated whether two sets of chaperones (Chaperone Plasmid Set, www.takara-bio.eu) would increase the solubility of the protein in *E. coli* BL21. Chaperone set 1 contained chaperones GroE, GroL, trigger factor and chloramphenicol (cm)
resistance [109], while chaperone set 2 contained chaperones GroE, GroL and cm [110] (E. coli Rosetta BL21 (DE3) pLysS/RARE was not suitable as an expression host for the chaperones because it already carries a plasmid with chloramphenicol resistance). Both strains were cultivated as described in section 2.18.1.

E. coli Rosetta BL21 (DE3) pLysS/RARE/pGEX_cyclo was induced with 0.1 mM IPTG, and the E. coli strains containing the chaperones with 1 mM IPTG (according to the manufacturer’s instructions). Cultivations were performed at 16 °C for all three expression cultures. As before, most of the protein was found to be insoluble. However, during expression with chaperone 1 in E. coli BL21 and independently in E. coli Rosetta BL21 (DE3) pLysS/RARE, some soluble protein was present in the supernatant (Figure 3.49).

![Figure 3.49: Expression of the cyclodeaminase GeD.](image)

Expression was performed at 16 °C with chaperone set 1, chaperone set 2 and in E. coli Rosetta BL21 (DE3) pLysS/RARE. Soluble protein is indicated by the arrows. Key: P, pellet; S, supernatant.

The GST-fusion protein was purified from the crude cell extract using GST minicolumns, coupled with on-column digest using PreScission Protease. The final yield of purified GeD from expression in E. coli Rosetta was approximately 0.8 mg/ml protein. Figure 3.50 shows the GST fusion protein (26 kDa) which was eluted with reduced glutathione, and the cleaved GeD (43 kDa) after on-column cleavage with PreScission Protease.
Figure 3.50: **Purified protein GeD.**

GeD expression in *E. coli* Rosetta BL21 (DE3) pLysS/RARE and co-expression with chaperone set 1. The gel shows the protein with the GST tag and following cleavage with PreScission Protease. Cleaved protein is indicated by the arrows.

The identity of the protein (from both samples) was verified by MALDI-MS (data not shown), and frozen at −80 °C until use.

### 3.2.12.2 Cyclodeaminase assay

On the basis of its homology to TubZ and RapL, GeD was presumed to act as a lysine cyclodeaminase, converting L-lysine to L-pipecolic acid [82]. In order to evaluate whether GeD could catalyze the same reaction, we used a method previously established by A. Sandmann and Y. Chai method to detect L-pipecolic acid by HPLC/MS.

In brief, treatment of the pipecolic acid standard (a racemic mixture of L- and D-) with 9-fluorenylmethyl chloroformate (Fmoc-Cl) converted the amine to the corresponding Fmoc-carbamate. The Fmoc-pipecolic acid was then straightforwardly detected by HPLC/MS (*m/z* = 571.2 [M-H]⁻ (Figure 3.51 A)). Additionally pure L- and D-pipecolic acid standards were derivatized and analyzed to establish the retention times of L- and D-pipecolic acid, showing that D-pipecolic acids elutes first. As a positive control, we simultaneously assayed the enzyme TubZ from the tubulysin cluster, which had previously been shown to catalyze the formation of L-pipecolic acid from L-lysine, *in vitro* (Y. Chai, A. Sandmann, unpublished results). As expected, incubation of TubZ with L-lysine resulted in the formation of L-pipecolic acid (Figure 3.51 B)). To verify that exclusively the L-stereoisomer was formed, we
mixed the standard reaction mixture with the product of TubZ. A clear increase of the signal corresponding to the L-stereoisomer was observed (Figure 3.51 C)).

**Figure 3.51: Cyclodeaminase assays.**
A) Standard of D- and L-pipecolic acid. B) L-pipecolic acid formed by TubZ in the presence of L-lysine. C) Mixture of assay reactions A) and B), to verify the stereochemistry of the pipecolic acid generated in B). D) Incubation of GeD with D-lysine. E) Incubated of GeD with L-lysine.

Disappointingly, incubation of GeD (both fractions were used from expression in *E. coli* Rosetta and also from coexpression with the chaperones) with both L- and D-lysine under the same conditions failed to produce any pipecolic acid of either stereochemistry. As the assay worked efficiently with the control protein TubZ, GeD is apparently inactive under the assay conditions, or L-lysine (D-lysine) is not the correct substrate.

### 3.2.12.3 Expression of A domains

In order to directly investigate A domain substrate specificity, we aimed to express the first three A domains from cosmid BI11 as N-terminally His6-tagged proteins from pET28b. DNA
fragments coding for the adenylation domains of geE (1 domain, GeEA1) and geG (2 domains, GeGA2 and GeGA3) were amplified from cosmid BI11, and cloned into pET28b vectors using Ndel and BamHI. Cosmid BA23 was not completely sequenced at this timepoint so that we couldn’t amplify the other two A domains from geJ and geM encoding GeJA4 and GeMA5. The constructs were confirmed by sequencing and transformed into E. coli Rosetta BL21 (DE3) pLysS/RARE. Cultivation was carried out at 16 °C. Expression was induced with 0.2 mM IPTG, at OD600 = 0.8–1.

![Figure 3.52: Expression of A domains in E. coli Rosetta BL21 (DE3) pLysS/RARE at 16 °C.](image)

Protein GeEA1 was found in the soluble fraction after French press lysis of E. coli cells, as judged by SDS-PAGE (Figure 3.52). However, only a small amount of both GeGA2 and GeGA3 were present as soluble proteins. Nonetheless, the observed sizes of the proteins were in good agreement with the calculated molecular masses of 56.6, 59.9 and 61.7 kDa, respectively. To try to improve the solubility of GeGA2 and A3, we co-expressed them with the chaperone set described in section 3.2.12.1. Coexpression with chaperone set 2 led to soluble protein in the case of GeGA2, allowing the enzyme to be purified by Ni²⁺-affinity chromatography using the Äkta prime system (Figure 3.53 and section 2.18.2.2). GeGA3 could not be expressed as soluble protein until now. In the future further attempts to increase solubility need to be done.
Figure 3.53: SDS-PAGE of adenylation domains overexpressed in *E. coli*.

A) Purification of A domain GeEA1. Lanes 1–6: various fractions after Äkta purification; Lanes 4–6 contain the purified A domain used for biochemical analysis. B) Purification of A domain GeGA2. Lane 1–9: various fractions after Äkta purification; Lanes 4–8 contain the purified A domain used for biochemical analysis.

All proteins were analyzed by MALDI-MS, which identified them as the desired proteins. Both the SDS-PAGE (Figure 3.53B) and the MALDI-MS analysis (data not shown) revealed, however, that a proportion of protein GeGA2 was truncated from its C-terminal end. As this was not expected to interfere with the substrate specificities of the intact domains, the mixture was used in subsequent assays.

3.2.12.4 ATP-PP<sub>i</sub> exchange assay

The substrate specificities of the two purified adenylation domains were evaluated using the well-established ATP-PP<sub>i</sub> exchange assay [58;111]. This assay measures the amino acid-dependent exchange of radiolabel from $^{32}$PP<sub>i</sub> into ATP. For this, each protein was incubated with a panel of different amino acids, including the anticipated substrate of each domain. As a control, each protein was incubated in the absence of added amino acid.
Figure 3.54: Relative substrate specificities of the adenylation domains GeEA₁, GeGA₂.

The substrate specificity of the A domains was investigated using the ATP-PPᵢ assay. The highest activities were set to 100%.

GeEA₁ activated L-Pip (100%), D-Pip (82.88%) and L-Pro (80.3%). The background and controls were between 0.85% and 3.79%, confirming that the measured activity reflected the true substrate preference of the A domain (Figure 3.54). GeGA₂ activated L-Orn (100%) as well as L-Gln (62.7%). L-Glu and L-Asp were also activated to a minor extent. So the preferred activation of ornithine and glutamine was clear. Taken together, our results prove that the two internal enzyme fragments investigated exhibit the enzymatic properties of adenylation domains. The determination of the A domains specificity enabled us to place the first two A domains in the correct order within the synthetase. As predicted previously, the A domain GeEA₁ activates pipecolic acid, and thus forms a part of the loading module (module 1). It is followed by module 2, which harbors an A domain (GeGA₂) specific for orithine/glutamine. It remains to test the other three A domains GeGA₃, GeJA₄ and GeMA₅ to determine their function.
4 Discussion

4.1 General scope of this work

This thesis deals with the biosynthesis and heterologous production of complex secondary metabolites from streptomycetes. A method for the heterologous expression of these compounds in streptomycetes was developed, using the phenalinolactone biosynthetic pathway as a model system. The established expression system provided the basis for the modification of the pathway, leading to the production of a new derivative.

Identification and biochemical characterization of the GE81112 biosynthetic gene cluster provided information about the biochemistry of this unique family of secondary metabolites and of nonlinear NRPS megasynthetases more generally. In the course of these studies, genetic manipulation methods for a new streptomycetes strain were developed, leading to the formation of a knockout mutant.

4.2 Phenalinolactone - Heterologous expression of a complex secondary metabolite

4.2.1 Heterologous expression of natural product pathways

Heterologous expression of complex natural product biosynthetic gene clusters in amenable host strains has become an important tool in natural products research and drug discovery. This strategy provides an alternative to (over)producing structurally complex substances that would be difficult or impossible to access by other means, and can enable the generation of novel analogs by combinatorial biosynthesis approaches. However, successful heterologous production of natural products faces significant challenges, deriving both from complicated nature of the secondary metabolites, and the complexity of the biosynthetic enzymes. In addition to laboratory convenience, excellent growth characteristics, and an array of available genetic tools, the heterologous host must be able to express relatively large proteins (e.g. PKS and NRPS proteins) [30;112], post-translationally modify them to their active forms, and produce adequate amounts of activated intracellular building blocks. As the codon usage is usually [113] critical for efficient expression, it is generally important that the GC content of the foreign genes and the host genome are similar. Furthermore, the functionality of regulatory elements and promoter structures as well as mRNA stability need to be considered when choosing the host organism. To date, the most successful heterologous expression
experiments have relied on hosts which are closely-related to the natural producer strain. For example, many clusters derived from *Streptomyces* clusters have been reconstituted in related actinomycete strains [29;33;114-116]. As most of these gene clusters were relatively small (<30 kbp), it was often possible to retrieve the entire gene set from a single cosmid within a genomic library of the natural producer strain [30]. In cases where the libraries were prepared in *E. coli*-*Streptomyces* shuttle vectors (e.g. the medermycin and griseorhodin A biosynthetic gene clusters [31;117]), transfer of the pathways into heterologous strains was straightforward.

However, many natural product biosynthetic gene clusters are larger than the average insert size of common cosmid vectors. One strategy for overcoming this size limitation is to use BAC shuttle vectors for library construction, as they can accommodate inserts in excess of 100 kbp. This approach enabled the successful expression of the 128 kbp daptomycin biosynthetic gene cluster from *Streptomyces roseosporus* in a related *Streptomyces* strain [118]. Another option is to clone subsets of the biosynthetic gene cluster into compatible expression plasmids, followed by their stepwise introduction and co-expression in suitable host strains [34;119]. As the cloning procedure allows the introduction of artificial promoter regions, heterologous production can be performed even in non-related host organisms. Using this strategy, natural products from *Streptomyces* were produced in *E. coli*, and myxobacterial compounds were obtained from *Streptomyces* host strains [120-122].

An alternative to these classical, time-consuming cloning and mutagenesis approaches is the reassembly of large natural product pathways on a single transferable vector system, using Red/ET technology. This recombination approach was developed by the Stewart group in 1998 [37], and is ideal for manipulating large pieces of DNA as it does not depend on the use of restriction enzymes. Phage-derived protein pairs (RecE/RecT from the Rac prophage or Reda/Redβ from the lambda phage) are employed to precisely alter target DNA molecules by homologous recombination within *E. coli* strains [36]. The first notable proof-of-principle for the cloning of a large PKS/NRPS hybrid pathway using Red/ET recombineering was provided by Wenzel *et al.* [35]. The 29.6 kb pathway directing the biosynthesis of myxochromides S was rebuilt on one molecule and the DNA elements for transfer, stable maintenance, and inducible expression were introduced, enabling the production of the natural product in *Pseudomonas putida.*
In *Streptomyces* strains, Red/ET cloning has been used to modify the backbone of cosmids containing complete biosynthetic pathways, enabling the heterologous expression of the aminocoumarin antibiotics novobiocin and clorobiocin [123]. In these examples, all of the genes were oriented in the same direction, consistent with the presence of a single transcriptional unit. In comparison to the aminocoumarin pathways, the phenalinolactone gene cluster of *Streptomyces* Tü6071 is more complex, as it consists of 35 orfs which appear to be organized into 11 individual transcriptional units [66]. This highly divergent architecture would seem to present a significant challenge to the regulatory apparatus of heterologous host strains.

Thus, in attempting to achieve heterologous expression of the phenalinolactone pathway, we developed a Red/ET-based methodology to reconstitute the entire gene cluster on one construct to ensure that all transcriptional units were intact. Whereas similar approaches had already been used to engineer myxobacterial systems [35;42], the ‘stitching together’ of complex actinomycete pathways via Red/ET recombineering had not been previously described when the Ph.D. thesis was started which intended us to start this project. Shortly after our work was published another example for the heterologous expression of a gene cluster using Red/ET was published by the Gust group [124]. Using this method they were able to stitch and express the coumermycin A₁ biosynthetic gene cluster in *S. coelicolor* M512. Using Red/ET recombineering the entire gene cluster of PL was rebuilt from overlapping cosmids based on the integrative *E. coli-Streptomyces* shuttle vector pOJ436, and transformed into different heterologous hosts.

### 4.2.2 Heterologous expression of the PL pathway in pseudomonads

Pseudomonads has been shown to be suitable heterologous hosts for the production of natural products [125]. These microbes possess a high GC content, have a relatively short doubling time, and genetic tools are well established. Therefore we aimed to heterologously express the PL gene cluster in *Pseudomonas putida*. As we had integrated an oriT-tet-trpE cassette into the expression construct, transformation and integration of the vector into the *Pseudomonas* genome was possible. Integration into the genome was proven by PCR and the obtained mutants were cultivated, extracted and the extracts analyzed by HPLC-MS. Unfortunately we did not observe any heterologous production of PLs in *P. putida*. 
One of the major issues for the heterologous expression of biosynthetic pathways in unrelated bacteria is the supply of adequate precursors. In the case of the polyketide myxothiazol, heterologous expression in \textit{P. putida} initially failed. However, integration into \textit{P. putida} of an operon responsible for production of an essential building block, methylmalonyl-CoA, resulted in myxothiazol production \cite{43}. To investigate whether the absence of a precursor could explain the failure to obtain the PLs from \textit{P. putida}, we considered the biosynthetic model for phenalinoalctone biosynthesis published by Dürr \textit{et al.} \cite{66,80} (Figure 4.1). The pathway can be divided into four major parts: (1) formation of the terpene backbone and its extension with phosphoenolpyruvate; (2) modification and rearrangement of the phosphoenolpyruvate-derived side chain toward the $\gamma$-butyrolactone moiety; (3) biosynthesis of L-amicetose and pyrrole-2-carboxylic acid; and (4) group transfer reactions and late-stage tailoring. The precursors for each portion of the biosynthesis are derived from (1) the nonmevalonate pathway (terpene biosynthesis); (2) pyruvate ($\gamma$-butyrolactone biosynthesis); (3) glucose (L-amicetose biosynthesis); and (4) L-proline (pyrrole-2-carboxylic acid biosynthesis).
Figure 4.1: Proposed biosynthetic pathway for phenalinolactone formation.

Figure adapted by Dürr et al. [66].
As the nonmevalonate pathway is present in *P. putida* [126] and the respective genes are also present in the cloned gene cluster, we eliminated this pathway from consideration. The remaining precursors such as pyruvate, glucose and L-proline are derived from the primary metabolism, and were therefore also expected to be present in *P. putida*. Thus, we assumed that *P. putida* is able to produce all precursors which are needed for PL biosynthesis.

As pseudomonads posses a high GC content similar to that of streptomycetes, we also assumed that codon usage was not a problem. This focused attention on the native *Streptomyces* promoters, which may not be recognized in *Pseudomonas putida*. To date, heterologous expression of natural product biosynthetic pathways in pseudomonads has only been successful when the native promoters were exchanged against the inducible $P_m$ promoter [35;43;125]. For this reason, we also aimed to exchange several promoters in the PL biosynthetic gene cluster to achieve heterologous expression in *P. putida*. As the PL biosynthetic gene cluster contains 35 orfs and 11 transcriptional units, we expected that at least one promoter would be present in each operon, for a total of 11 promoters. As a starting point, we decided to exchange the promoter in front of the operon containing the genes responsible for assembling the core structure of the PLs. The proteins PlaT1–PlaT6 (Figure 4.2) comprise the enzymatic machinery to biosynthesize the phenalionalctone terpenoid scaffold. The isoprene building blocks for phenalinalactone biosynthesis are derived from the nonmevalonate pathway which is also present in *P. putida*. So it can be assumed that the isopentenyl pyrophosphate (IPP) should be produced in primary metabolism of *P. putida*. The genes *plaT5* and *plaT6* in the PL cluster comprise additional copies of the *dxp* (1-deoxy-D-xyululose 5-phosphate) and *hmbpp* (1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate) synthase genes from the nonmevalonate pathway. It is speculated that PlaT5 and PlaT6 probably enhance the IPP production [66]. PlaT4 catalyzes the head to tail condensation of four isoprene monomers toward geranylgeranyl diphosphate (GGDP), the phenalinalactones direct terpene precursor. PlaT1 is suggested to catalyze an epoxidation of the C-14/-15 double bond of GGDP prior to its cyclization. PlaT2 is responsible for the cyclization of GGDP to PL precursor 1. PlaT3 might be considered for the phosphoenolpyruvate transfer to the terpene backbone. The next step would be the dehydrogenation of the newly established C-C linkage, giving the C-15/-16 double bond. PlaZ would be a candidate for this dehydrogenation reaction (Figure 4.2).
Unfortunately the corresponding genes \(\text{plaT1-3}\) and \(\text{plaT4-6}\) are located in two different operons, making it impossible to attempt to upregulate transcription with a single promoter switch. We started by targeting the operon containing the first three biosynthetic genes, \(\text{plaT4-6}\). The native promoter was exchanged against a cassette containing the \(P_m\) promoter, the \(\text{xylS}\) gene and a resistance gene (chloramphenicol) using Red/ET cloning, and mutants were verified by PCR. We expected at least the direct precursor GGDP to be produced as \(\text{plaT4-plaT6}\) are present in the operon where the promoter was exchanged. But analysis by HPLC-MS analysis failed to reveal the presence of GGDP, PL precursors or PL in the mutants. This result demonstrated that exchange of this single promoter was insufficient to achieve production of any precursors in \(P.\ putida\).

In principle, a follow-up strategy would be to attempt to exchange all of the native promoters against the same \(P_m\) promoter cassette or other \(Pseudomonas\) promoters. However, this approach is likely to cause difficulties, as the introduction of promoters by Red/ET recombination is based on homologous recombination; the presence of the same promoter sequence at multiple locations within the cluster would probably result in undesired recombination events. Alternatively, it might be possible to use a set of different promoters. However, only a few inducible promoters are currently known for pseudomonads. Taking these considerations into account, we concluded that it would not be feasible to exchange all
of the promoters within the PL cluster. We therefore pursued heterologous expression in an alternative host.

4.2.3 Heterologous expression of the PL pathway in streptomycetes

The Red/ET engineering enabled the efficient transfer and integration of the whole PL pathway into several *Streptomyces* strains, as the construct contained a vector backbone with an *attP* site and the integrase gene. The 62 kbp construct was successfully conjugated with high efficiency into the host strains *S. lividans* TK24 and *S. coelicolor* A3(2), and integration into the chromosomes confirmed by PCR. In contrast to the *Pseudomonas* mutants, the resulting mutant strains of *S. lividans* TK24::CPhl8 and *S. coelicolor* A3(2)::CPhl8 were both shown to produce phenalinolactones A and D. This result strongly suggests that the inserted cluster contains all genes necessary for the biosynthesis of phenalinolactone, although participation of enzymes native to the heterologous hosts cannot be ruled out. Evidently, all of the regulatory elements required for transcription of the *pla* genes were recognized in the heterologous hosts, albeit at reduced efficiency relative to the native host strain (Table 3.2).

One potential complication of using *S. lividans* and *S. coelicolor* as heterologous hosts is their endogenous production of the antibiotics prodiginin (Red), actinorhodin (Act) and calcium-dependent antibiotic (CDA), which may hamper the simultaneous biosynthesis (and detection) of a desired new compound. This problem has been circumvented in several cases by using a mutant strain of *S. coelicolor* M512, in which the Act and Red gene clusters have been deleted [123;127]. Therefore, *S. coelicolor* M512 was also used here as an additional heterologous host for phenalinolactone biosynthesis. Indeed, after successful expression of the cluster in the strain, production of phenalinolactone was 5- to 50-fold higher at 28 °C relative to that in *S. lividans* TK24 or *S. coelicolor* A3(2) (Table 3.2). Interestingly, cultivation at 37 °C in *S. lividans* yielded a further enhancement in productivity (up to 100-fold compared to cultivation at 28 °C) although the optimal growth temperature for *S. lividans* is 30 °C. An increase in phenalinolactone biosynthesis was not observed for the *S. coelicolor* strains at 37 °C. In fact, incubation at 37 °C completely abolished PL biosynthesis, consistent with the fact that production by the native host *Streptomyces* Tü6071 is also significantly lower at this temperature (200 µg/L). Thus, under optimized fermentation conditions, it is clear that *S. lividans* is a superior heterologous host for phenalinolactone biosynthesis than the *S.
coelicolor strains, although the production is still 50-fold lower than from the natural producer.

An attempt was also made to increase phenalinolactone production by feeding of the putative precursor mevalonolactone, but no increase in the yield of PL was observed. Analysis of the pla cluster shows that it incorporates genes for producing isoprenoid building blocks via the mevalonate-independent pathway, and indeed this biosynthetic origin has previously been demonstrated [66]. In addition, genes belonging to the mevalonate pathway have not yet been identified in either S. lividans and S. coelicolor [128;129]. Thus, it seems that the heterologous host strains used here lack the enzymes to metabolize mevalonate, which accounts for our failure to increase phenalinolactone yields by supplementation.

In a further attempt to increase productivity, we exchanged two of the native promoters against the strong constitutive ermE promoter using Red/ET recombinaseening again. The advantage in this strategy was the ermE promoters sequence could be introduced with the primers, so that two promoters could be exchanged in a single step. The two promoters were placed in front of the two operons harboring the first biosynthetic genes (plaT1-3 and plaT4-6). It has previously been demonstrated that introduction of the ermE promoter can increase the yields of heterologously-expressed natural products, or even lead to initiation of production. For example, cloning of eight validamycin biosynthetic genes under the control of the ermE promoter led to heterologous production of this metabolite in S. lividans [130]. Another example was the heterologous expression of an intermediate of the kendomycin starter unit. The heterologous expression of a gene set of 7 genes revealed that 3,5-dihydroxybenzoic acid is an intermediate in the kendomycin starter unit biosynthesis [131]. Successful heterologous production was only achieved in S. coelicolor but not in S. lividans. Cloning of the ermE promoter in front of the gene set initiated the heterologous expression in S. lividans. Production under the control of the ermE promoter in S. lividans was even much more higher (150mg/ml) compared to production in the natural producer or the heterologous production in S. coelicolor with and without the ermE promoter (S. C. Wenzel unpublished data). Unfortunately, exchange of the native promoters against the ermE promoter in the pla biosynthetic gene cluster did not yield any increase in productivity. As we introduced two ermE promoters in front of both operons containing the first biosynthetic genes we expected at least a production of GGDP or precursor 1 (Figure 4.2). As PlaZ which is probably needed for the formation of precursor 2 was not part of both operons precursor 2 was not directly
expected to be produced. But no GGDP or any other precursors could be detected; instead, PL production was completely abolished. These results led us to conclude that promoter exchange interrupted transcription or led to problems with regulation or both, probably as a result of the complex operon structure of the *pla* cluster. Again, a solution might be to exchange of all the promoters within the cluster, but this approach would raise the same issues as in the pseudomonads. Thus, overall we achieved the greatest success in improving PL production by changing the growth temperature of *S. lividans* showing that standard microbiological optimization methods can still be effective.

### 4.2.4 Genetic modification

The expression system described here further enabled the generation of an unglycosylated phenalinolactone derivative. For this, Red/ET technology was combined with the FRT/FLP recombinase system to construct a markerless deletion of the glycosyl transferase gene *plaA6*. Heterologous expression of the modified construct resulted in the unglycosylated derivative phenalinolactone E (PL E), demonstrating that the glycosyl transferase PlaA6 catalyzes the attachment of L-amicetose to the hydroxylated terpenoid backbone. Surprisingly, biosynthesis of PL E was only observed in *S. coelicolor* M512, although yields of the parent glycosylated compound from *S. lividans* TK24 were much higher. In this case, the utility of *S. coelicolor* mutant strain M512 as suitable host may stem from its overall lower background of secondary metabolism. The use of the FRT/FLP recombinase system in combination with the Red/ET technology is a powerful tool to create a markerless gene deletion in a single expression construct. This modification is conveniently achieved by introducing the FRT recognition sites into the PCR primers. Recombination with FLP recombinase results in a markerless mutation, in which only a 34 bp scar sequence is left in place of the deleted gene of interest. This scar sequence does not present a problem for further recombination procedures. To construct two mutations simultaneously, the Cre/loxP system can be employed in parallel.

A similar approach has been used by Heide *et al.* [132]. New aminocoumarin antibiotic derivatives were generated by targeted gene replacement using Red/ET recombineering. However, instead of introducing FRT or loxP recognition sites with the primers, they introduced recognition sites for the restriction enzymes *Spe*I and *Xba*I which are rare cutters in the *Streptomyces* DNA. Digestion of the cosmids by *Spe*I and *Xba*I resulted in compatible overhangs, and religation then resulted in a cosmid in which only an 18 bp scar sequence was
left in place of the deleted gene. One drawback of this approach in comparison to the FRT/FLP or Cre/lox system, however, is that a ligation step is always required. Additionally, no other SpeI or XbaI sites can be present in the target sequence, whereas the recombinase system is not dependent on restriction sites but on the FRT/loxP recombination site.

To obtain additional PL derivatives, we generated two constructs containing markerless deletions of the genes *plaP2* and *plaP5*. By removal of the acyltransferase-encoding gene *plaP2*, we aimed to generate a PL derivative lacking the 5-methylpyrrole-2-carboxylic acid moiety (Figure 4.3). The gene *plaP5* is proposed to be responsible for the methylation of the pyrrole carboxylic acid [66], and thus deletion of this gene was expected to give an unmethylated PL derivative (Figure 4.3). In this case, we exploited the Cre/lox system instead of the FLP/FRT system. Using both systems in combination, we aimed to obtain a small set of different PL derivatives (Figure 4.3).

![Figure 4.3: PL derivatives expected after deletion of different biosynthetic genes, in heterologous streptomyces hosts.](image)

Unfortunately, no variant PLs could be detected in extracts of mutants carrying the deletions in *plaP2* and *plaP5*, and instead, PL production in these strains was completely abolished. The failure to detect new derivatives might be explained by very low production levels: as PL
E is already produced in miniscule amounts, it seems reasonable that production of derivatives could fall under the detection limit. However, we cannot rule out the possibility that deletion of the genes interrupted transcription, or that regulatory problems occurred.

4.2.5 Future aspects

Although several *Streptomyces* gene clusters have been relocated to heterologous hosts using Red/ET cloning, transcription in each case was under the control of a single promoter [123]. Thus, the *pla* cluster represents the most complicated gene set yet to be heterologously expressed in any *Streptomyces sp*. In future, efforts to understand the regulatory factors governing phenalinolactone biosynthesis should aid in efforts to optimize heterologous production, for example by overexpressing positive regulators. Also other heterologous actinomycetes hosts could be tried to optimize heterologous production. In addition, the results described here set the stage for further engineering the biosynthetic machinery in *E. coli* towards the generation of additional novel PL analogues.
4.3 GE81112 family compounds - Novel tetrapeptide inhibitors of bacterial protein synthesis

Identification of the biosynthetic gene cluster

The emergence of multi-drug resistant microbial pathogens is driving the search for novel and more effective antibiotics, especially those with new mechanisms of action. The majority of known antibiotics inhibit the fundamental process of protein synthesis. Prominent examples of antibiotics which target this machinery are the macrolides erythromycin and azithromycin, the aminoglycosides kanamycin and tobramycin, and the tetracyclines [133]. Elongation activities such as aminoacyl-tRNA binding and decoding in the ribosomal A-site, transpeptidation, and translocation are frequently found to be affected by antibiotics, albeit by different mechanisms. However, other functions such as translation initiation and termination, are only rarely targeted [133]. Thus, the tetrapptide GE81112 compounds represent a unique class of novel antibiotics, as they selectively and effectively inhibit the formation of the prokaryotic 30S initiation complex [67]. To obtain a better understanding of the biosynthesis of these unique metabolites, we set out to identify and clone the corresponding biosynthetic gene cluster. We hoped that insights into the biosynthesis would enable the directed engineering of novel GE81112 derivatives.

4.3.1 Attempts to identify the GE81112 biosynthetic gene cluster

4.3.1.1 Generation of a cosmid library and hybridization strategy

At the beginning of the work, we obtained the GE81112 producer strain S. 14386 from the company KtedoGen. In order to enable the identification and cloning of the biosynthetic gene cluster, we started by optimizing production conditions for the GE81112 compounds. After establishing cultivation and detection of the GE81112 metabolites, we aimed to locate the corresponding gene cluster by generating a cosmid library, and screening it with suitable probes. Before rapid genome sequencing methods became available, numerous NRPS and/or PKS biosynthetic gene clusters were sucessfully identified using this approach [134-137].

We therefore generated a 2304 clone cosmid library from the genomic DNA of the GE81112 producer strain containing. A suitable probe was obtained by applying a “retrobiosynthetic strategy”, which allowed us to predict some aspects of the composition of the GE81112 cluster from the metabolite structures. On this basis [67], we predicted that the biosynthesis
would be carried out by a non-ribosomal peptide synthetase, and so designed probes to identify NRPS genes. Furthermore, we were able to predict that the starter unit would probably derive from pipecolic acid, which is known to be formed from lysine via the action of a lysine cyclodeaminase [82]. As cyclodeaminase genes are relatively rare in the genome [86] this gene was used to design a second probe to identify the GE81112 cluster.

### 4.3.1.2 Identification of an unknown NRPS biosynthetic gene cluster

Screening of the cosmid library with both probes revealed two cosmids apparently harboring overlapping regions of a gene cluster. Furthermore, sequence analysis revealed that both cosmids harbored NRPS biosynthetic genes. As the modules in both clusters contained comparable domains in the order C-A-PCP [56] (Figure 4.4), we expected the two cosmids to contain the same biosynthetic gene cluster. Surprisingly, however, a more detailed sequence analysis revealed that the gene clusters, although highly similar, were not identical. This finding might indicate that they have arisen from a recent gene duplication event, followed by sequence divergence.

The presence of two highly-similar clusters was unexpected, but might have accounted in principle for the biosynthesis of the GE81112 family of tetrapeptides (e.g. each of the clusters might have given rise to a subset of family members). However, no examples of such ‘in parallel’ biosynthesis of similar metabolites within the same strain have been described to date, although closely-similar derivatives of a single natural product have been isolated from different strains (e.g. myxochromide A and S) [138].

As it was not clear at this stage if the identified clusters were responsible for the biosynthesis of the GE81112 compounds, we analyzed the NRPS domains from each module. For the
incorporation of four amino acids we expected four modules containing the standard C-A-PCP organization and additionally a cyclodeaminase for the generation of the starter unit pipecolic acid. The four module arrangement was present on cosmid FD10 as we predicted, as well as a cyclodeaminase gene. These results supported the hypothesis that the identified assembly line on cosmid FD10 is able to biosynthesize a tetrapeptide or tetrapeptides, as expected in the biosynthesis of the GE81112 family. The gene cluster on cosmid AI6 was not completely present on the cosmid. Three complete modules could be identified on this cosmid indicating that the cluster should at least be able to synthesize a tripeptide. Module four contains a C and an A domain but lacks a PCP where the cluster is interrupted. This could imply that there would still be a PCP in the genome sequence so that the cluster would also be able to synthesize a tetrapeptide or even a longer peptide, as we don’t know what sequence follows. When considering the similarity with FD10 one could also conclude that the cluster on AI6 ends with the C domain on gene \textit{nrps}_3 (same domain arrangement than in cosmid FD10) and that the additional A domain would belong to another cluster. To answer these questions the overlapping cosmid providing the whole pathway for the NRPS on AI6 would need to be identified which failed until now.

To test if all domains are likely to be functional, the conserved regions of all C, A, PCP and E domains \cite{91} were analysed and found to be present. Using the specificity-conferring code of adenyltransferation domains, we were able to predict the substrate specificity of some of the A domains (Table 3.8). As both clusters had an overall high content of sequence similarity, it was not surprising to find three A domains from each cluster which showed the same apparent substrate specificity. The identical A domains are found to be in the same order assuming that module 1 has been deleted/added to the FD10 cluster (Figure 4.4: identical domains are highlighted in the same colors). The A domains FD10\textunderscore A2 (AI6\textunderscore A1) showed substrate specificity for the incorporation of phenylalanine/tryptophan. A domain FD10\textunderscore A3 and AI6\textunderscore A2 were predicted to show substrate specificity for the activation of threonine. No confident predictions could be made for the remaining A domains, possibly indicating that they active uncommon amino acids.

Overall, the two predicted A domain specificities for both clusters did not fit well with the biosynthetic hypothesis for the GE81112 compounds, as it was expected from the “retrobiosynthetic analysis” that pipecolic acid and two histidines would be incorporated.
In addition to the NRPS genes, we found the expected cyclodeaminase gene, but it was only present on cosmid FD10. Furthermore, sequence analysis revealed that more than half of the protein was missing relative to previously-described cyclodeaminases [82;83;85]. We assumed, therefore, that the truncated protein was no longer functional. This result provided an additional strong hint that the identified gene cluster was probably not responsible for GE81112 biosynthesis. As these results were based solely on in silico predictions, however, we aimed to obtain experimental validation for the identity of the gene clusters.

4.3.1.3 Heterologous expression

The most straightforward strategy to verify a gene cluster is to knockout one of the NRPS genes which are normally clustered in the genome and essential for the biosynthesis. We tried to do this knockout but, to date, it has not been successful. We therefore pursued an alternative strategy to determine the identities of the identified gene clusters, using heterologous expression [30;118;139;140]. Both cosmids were transferred into heterologous hosts S. lividans TK24 and S. coelicolor M512 and the extracts were checked for GE81112 production. Unfortunately, the desired compounds could not be detected in the heterologous hosts. Instead three new compounds were detected in the heterologous host carrying cosmid FD10, which were not present in the wild type. These substances exhibited exactly the same fragmentation pattern as three compounds that were found during our initial analysis of the natural producer strain S. 14386, and which attracted our attention because one of them had the same mass (659) as GE81112 factor B. In addition, their fragmentation patterns suggested that they are chlorinated. Thus, the compounds might be structurally-related to the GE81112 metabolites, but were definitely not the desired GE81112 compounds.

Taken together, these results indicated that the biosynthetic gene cluster present on cosmid FD10 is not responsible for biosynthesis of GE81112. The new compounds were not expressed from cosmid AI6, although it contained a highly similar gene cluster. One possible explanation for this result was that the cluster was interrupted in the nrps_3 gene and was not fully present on the cosmid. When we expect the cluster on AI6 to biosynthesize a tripeptide (as discussed earlier) then this tripeptide should at least incorporate the three identical amino acids than cosmid FD10, which could lead to a similar fragmentation pattern of the compound. But no such tripeptides could be identified. This result rather indicates that the cluster was not complete and was therefore not heterologously expressed as essential genes.
are missing. We also concluded that the cluster on AI6 is not responsible for GE81112 biosynthesis, due to its high overall similarity to the gene cluster on FD10 which has been proofen to biosynthesize the new compounds. However, to provide further support for our conclusions, we aimed to purify the novel compounds from the wild type and elucidate their structures. All three compounds could be purified in reasonable amounts for NMR elucidation using standard chromatographic techniques. We focused our initial structure elucidation efforts on compound 631, as it was obtained in the highest yields following purification. The NMR data obtained to date suggest that the structure of compound 631 contains four amino acids, consistent with the four modules found within the NRPS biosynthetic gene cluster on cosmid FD10. However, in silico analysis of the A domain specificity predicted that FD10_A2 would incorporate phenylalanine and FD10_A3 threonine, while it appears that the compound includes a starter most likely derived from proline and a second amino acid derived from alanine. Analysis of the A domains FD10_A1 and FD10_A4 did not yield a clear prediction for their substrate specificity, consistent with the idea that they are uncommon A domains which may activate two unusual, modified ornithines.

The gene cluster found on cosmid FD10 exhibits few additional features that hint as to how to put together the structural elements identified to date by NMR. One element of note is that the cluster ends with a C domain and not with a TE, as is typical for bacterial NRPS. This observation suggests that the C domain might catalyze a head-to-tail condensation to yield a cyclic product, as found for the cyclosporine synthetase [141]. However, evaluating this hypothesis awaits the full structural elucidation of the compounds. Additionally there was also a PKS part found to be present on FD10 but not on AI6. But this gene was also not fully present on the cosmid. So it cannot be excluded that the PKS is somehow involved in the biosynthesis of the new compounds, but this seems rather unlikely as it is not expected to be functional when interrupted. Final evidence would be given by a knockout in the natural producer or even on cosmid FD10 which was heterologously expressed. As the new compounds contained unusual amino acids harboring modifications it is also not clear how they can be biosynthesized from the relatively simple cluster on the cosmid. One explanation would be that the heterologous host provides some of the enzymes which are needed for this most likely complicated biosynthesis. One fact which is supporting this hypothesis is that the new compounds were only expressed in S. lividans and not in S. coelicolor, indicating that S. lividans has probably some advantages (e.g. special enzymes) for the biosynthesis of the new compounds.
4.3.2 New strategy-new probe

Identification of the GE81112 biosynthetic gene cluster

As we had proven by heterologous expression that the gene cluster on cosmFD10 (and probably also on AI6) were not responsible for the biosynthesis of the GE81112 compounds but for a new class of peptides, we developed an alternative strategy to identify the GE81112 gene cluster. The basis for our approach was to screen with a novel probe. We knew from the earlier experiments that our previous probe had successfully identified the cyclodeaminase gene on cosmFD10. However, the protein was truncated and therefore presumed to be inactive. As such pseudo-genes typically arise from gene duplication of an original active gene copy, we assumed that there might be a second, complete cyclodeaminase in the genome. To evaluate this hypothesis, we designed a probe specific for additional copies of the cyclodeaminase gene, on the basis of the sequence of the truncated cyclodeaminase. Screening the genome with this probe by Southern Blot revealed two signals in addition to that corresponding to cosmFD10, indicating that the genome contained a second or even a third copy of the cyclodeaminase gene. Importantly, this gene was likely to be associated with the remainder of the GE81112 gene cluster. We therefore re-probed the cosmid library with the specific cyclodeaminase probe, leading to the identification of new cosmids. We further confirmed that the cosmids contained the cyclodeaminase sequence by PCR. As we strongly expected NRPS sequence to be present on the cosmids, we used degenerate NRPS primers to amplify A domain sequences. Using these primers, we successfully amplified A domains from four of the identified cosmids. Prediction of the A domain substrate specificity revealed one cosmid containing an A domain with specificity for proline/pipelicolic acid, exactly as predicted for the GE81112 starter unit. Sequencing of the cosmid revealed a biosynthetic pathway which harbored different genes (e.g. cyclodeaminase) which we expected for GE81112 biosynthesis. However, as the entire gene cluster was not present on the cosmid, we identified an overlapping cosmid. Furthermore, we verified that the genetic organization on both was co-linear with the genome.

To confirm the identity of the cloned biosynthetic locus, we first aimed to establish a method to genetically manipulate the strain. Although genetic manipulation systems are well established for some streptomycetes strains [10], it turned out that the GE81112 producer was significantly less amenable to genetic manipulation. Nonetheless, using biparental mating, we developed a method for DNA transfer from E. coli via conjugation. We started method
development by conjugating both a replicative and an integrative vector as controls to test transformation efficiency. Using the standard amount of \textit{E. coli} as described in Kieser \textit{et al.} \cite{10}, we only obtained low numbers of colonies. As it was known from studies with other bacteria that the amount of \textit{E. coli} can be crucial for conjugation efficiency \cite{142}, we tested different amounts of \textit{E. coli}. Indeed, we found that transformation efficiency was 10–20-fold higher when large numbers (50×10^9) of \textit{E. coli} were used. With a conjugation method established, we were able to inactivate gene \textit{geE}, which resulted in the complete abolishment of the main GE81112 derivative factor B, which we could detect by high-resolution MS. This result provided clear evidence that the identified cluster was responsible for the production of the GE81112s.

### 4.3.3 Analysis of the GE81112 biosynthetic gene cluster

Sequence analysis of the identified cluster revealed that it harbored only NRPS components. Many clusters have a co-linear genetic and protein organization, following the so-called co-linearity rule, in which each module is responsible for one discrete chain elongation step and the specific order of the modules defines the sequence of the incorporated amino acids \cite{64}. Contrary to our expectation that we would find such a linear NRPS system providing four modules with the standard C-A-PCP domain order, the cluster exhibited a highly nonlinear genetic arrangement (Figure 4.5).

![Figure 4.5: Nonlinear arrangement of the GE81112 biosynthetic gene cluster.](image)

Figure 4.5: Nonlinear arrangement of the GE81112 biosynthetic gene cluster.

Domains that are proposed to interact are highlighted in the same colors.

Several examples of such nonlinear NRPS systems have been reported and initially, these NRPS were thought to be rare exceptions to the co-linearity rule \cite{56}. However, during the past few years it has become clear that nonlinear NRPS are not an exception, but constitute a considerable fraction of the NRPS repertoire present in nature. The characteristic feature of these nonlinear NRPSs is at least one unusual arrangement of the core domains C-A-PCP. The GE81112 system also revealed a highly split genetic and module architecture, but when rearranging the domains they provide a linear assembly line as described in the results part. Although four amino acids are incorporated into the structure which would normally require
four modules, five modules harboring five distinct A domains were identified. Further
unusual features of the cluster include two freestanding A-PCP didomains encoded by genes
g
ej and
g
eM, and a stand-alone A domain, encoded by gene
g
eE. Such stand-alone A
domains have previously been identified in the nonlinear NRPS pathways to myxochelin
[143] and yersiniabactin [144]. Similarly, freestanding A-PCP didomains have been found in
the zorbamycin and syringomycin gene clusters [145;146].

The conserved regions of all domains in the GE81112 cluster were analyzed, and predicted to
be functional. However, from the domain assignment alone, the order of subunits (modules)
could not be predicted, as there were more modules present then expected for the biosynthesis
of a tetrapeptide. We therefore began our analysis of the cluster with the prediction of the
substrate specificity of the A domains, which we anticipated would give us a hint as to the
order of the subunits. Based on analysis of the GE81112 structures, we expected that four
amino acids would be the precursors of the biosynthesis incorporated into the compound in
the order: pipecolic acid-ornithine/glutamine-histidine-histidine (Figure 4.6).

![Figure 4.6: GE81112 tetrapeptide structure and corresponding amino acid sequence Pip-Orn/Gln-His-His.](image)

The predicted specificity for four of the five A domains correlated with our expectations
(Table 3.13). Together with an analysis of putative docking regions, we were able to predict
how the NRPS proteins interact with each other to biosynthesize the GE81112s (Figure 4.7).
Overall, we suggest that the linear assembly line consists of four modules exhibiting the
standard C-A-PCP domain arrangement, and that the pathway incorporates one additional A-
PCP didomain, where the function was not clear, as well as a stand-alone thioesterase domain
for chain release. On both ends of the cluster additional *orfs* have been (Table 3.12) found where it was not clear if they have a function in the biosynthesis. What we suggest here is a first preliminary model of the biosynthesis where it remains to determine the boundaries of the gene cluster by knockout experiments.

Nonetheless, the function of the apparently superfluous module 5 remained unclear. One hypothesis was that the corresponding gene *geM* arose by gene duplication of *geJ* (or the opposite way around), as both genes showed the same domain arrangement and the encoded A domains were both predicted to activate histidine. If true, the protein GeM would also be expected to be able to interact with GeG to form the last module. However, sequence analysis revealed that the two proteins are not highly homologous, and that in addition, GeM was missing a putative recognition sequence for interaction with GeG as was present in GeJ. We therefore concluded that GeM is unlikely to interact with GeG, and therefore assigned the role of the fourth extension module to protein GeJ. However, as the two domains from GeM were predicted to be functional, we assumed that the didomain must have a specific function in the biosynthesis. A detailed discussion about a possible biosynthesis and the function of the single proteins will be given in the next chapter.

### 4.3.3.1 Biosynthetic proposal for the GE81112 compound family

Our data taken together allowed us to propose a pathway to the GE81112 metabolites, albeit incomplete. GE81112 biosynthesis likely starts with the biosynthesis of L-pipecolic acid from L-lysine via the action of the putative cyclodeaminase (Figures 4.8A and 4.14). Unfortunately we were not able to reconstitute this reaction *in vitro* when we expressed the GE81112
cyclodeaminase GeD, as GeD was inactive under our assay conditions. Pipecolic acid would then be activated by the A domain GeEA1 in module 1 and loaded onto the adjacent PCP (Figures 4.8 B and 4.14). To evaluate the function and substrate specificity of the A domain, we performed an ATP-PPi exchange assay in vitro. This experiment clearly showed that L-pipecolic acid was activated, supporting our hypothesis that this is the starter unit for the biosynthesis. The pipecolic acid moiety in GE81112 is hydroxylated at the β-position; a reaction which we predict is catalyzed by the putative L-proline-3-hydroxylase, GeF. However, the timing of hydroxylation (on the free amino acid [147], while the amino acid is thethered to the PCP [148], or following release of the chain from the assembly line [55;90]), remains unclear and several examples of all three reactions have been reported [55].

![Proposed biosynthetic mechanisms](image)

**Figure 4.8:** Proposed biosynthetic mechanisms.

A) Generation of pipecolic acid from lysine via the action of a cyclodeaminase (GeD). B) Activation and loading of pipecolic acid by module 1 followed by subsequent hydroxylation at position 3 by a putative β-hydroxylase (GeF). Alternative timings of the hydroxylation reaction are possible.

It might have been possible to evaluate whether hydroxylation occurs before or after activation by the module 1 A domain, by assaying for the domain’s relative specificity for pipecolic acid and hydroxypipecolic acid. However, as the required hydroxypipecolic acid was not commercially available, it was not possible to carry out this experiment. Nonetheless, we propose, based on literature precedent [148], that GeF most likely hydroxylates the PCP-bound pipecolic acid (Figures 4.8 B and 4.14).
It was not directly obvious from which amino acid the second building block is derived as there are two different structures found in the GE81112 factors A, B and B1 which differ at position 6. In GE81112 factor A and B the second amino acid incorporated into the structure is hydroxypentanoic acid, which is hydroxylated at position 4 and O-carbamoylated at position 6. In factor B1 it is an ornithine which is hydroxylated at position 4 and N-carbamoylated at position 6. To account for the biosynthesis of these two derivatives, we hypothesized that the A domain GeGA2 of module 2 can activate either glutamic acid or glutamine, followed by reduction of the carboxylic groups which would lead to hydroxypentanoic acid (factor A and B) or ornithine (factor B1), respectively. The proposal is shown in Figures 4.9A and B.

![Diagram](image)

**Figure 4.9: Proposed biosynthetic mechanisms for generation of the second building block.**

A) Reduction of PCP-bound glutamate to hydroxypentanoic acid. B) Reduction of PCP-bound glutamine to ornithine.

The timing of the reduction is not clear, nor is the enzyme which would catalyze this reaction. The next step in either case, however, would be the hydroxylation of the $\gamma$-carbon putatively catalyzed by GeI. Finally, the hydroxypentanoic acid/ornithine would be modified by N-/O-carbamoylation. Again, the timing of the carbamoylation is not clear, but precedent examples suggests that it would happen following chain release [149;150]. Carbamoylation is normally catalyzed by a carbamoyltransferase, which we couldn’t find in the biosynthetic pathway. So we cannot exclude that the carbamoylation is catalyzed by another mechanism. Additionally it might be possible that the carbamoyltransferase gene is located somewhere else in the
genome. Alternatively, ornithine (factor B1) would be activated directly by the A domain, but this model does not explain how factors A and B are generated. To determine directly which amino acid is activated by this A domain, we tested the function of the enzyme \textit{in vitro} using the ATP-PP\textsubscript{i} exchange assay. According to our hypothesis, the A domain must show a broad substrate acceptance, activating at least two different amino acids (glutamate, glutamic acid or ornithine) to generate all three GE81112 derivatives. Indeed, activation of all the three amino acids was observed, although the A domain showed a preference for ornithine. Thus, it appears that ornithine is directly activated by GeGA\textsubscript{2}, which accounts for the structure of factor B1. The reduced degree of activation of the other two substrates may indicate that modified versions of glutamic acid and glutamine are preferred.

The third amino acid incorporated into the chain is either histidine (in GE81112 factor A) or aminohistidine (as in factors B and B1). In the case of GE81112 factor A, histidine is likely to be incorporated directly, but the origin of the aminohistidine is less obvious. One hypothesis is that the aminohistidine might be derived from cyclization of arginine by nucleophilic attack of the \(\gamma\)-carbon (Figure 4.10). This step could be catalyzed by an arginine cyclase operating on a PCP-bound substrate. This proposal implies that the A domain GeGA\textsubscript{3} tolerates different substrates, activating both histidine (factor A) and arginine (factors B and B1). However, an additional possibility is that cyclization of arginine to aminohistidine occurs before the amino acid is tethered to the PCP, making aminohistidine the second substrate of the A domain. Thus, according to this alternative hypothesis, the A domain should be able to activate both histidine and aminohistidine.

![Figure 4.10: Proposed mechanism for the PCP-bound cyclization of arginine by a putative arginine cyclase.](image)

As the cluster lacks an obvious arginine cyclase, it appears likely that aminohistidine is generated by another mechanism. The \textit{in silico} analysis of GeGA\textsubscript{3} appears to support this proposal, as it suggests a specificity for Tyr/Trp and not arginine. This prediction may indicate a general preference for aromatic residues, accounting for activation of both histidine
and aminohistidine, but it is not clear how aminohistidine could be generated from tyrosine or tryptophane. Furthermore, amination of the PCP-bound histidine or after release of the chain cannot be excluded. To evaluate one of the hypotheses the suggested putative substrates must be tested in the ATP assay, which remains to be done.

The last amino acid to be incorporated in GE81112 is a 6-chlorinated histidine. The chlorination reaction is expected to be catalyzed by a halogenase. Halogenation is a common modification of natural products and many halogenated compounds are known [151-153]. The timing of halogenation reactions during natural product biosynthesis has been shown to vary. For example, pre-assembly line chlorination of tryptophane occurs in rebbecamycin assembly [154], a PCP-bound threonine is modified during the biosynthesis of syringomycin [155], while β-tyrosine is chlorinated following release of the peptide chain, in the assembly of chloroereomycin [156]. A candidate for catalyzing the chlorination of the GE81112 compounds is the halogenase GeF. A number of different scenarios can be proposed for how the histidine is chlorinated and incorporated into the metabolites (Figures 4.11 A, B and C). Common to all three is the role of module 4 in incorporating the (chloro)histidine, which is also hydroxylated at the β-position.
Figure 4.11: A linear model for the GE81112 NRPS templated assembly from four amino acids.

In the first proposal (A), biosynthesis starts with the activation of histidine by the A domain of GeM (which is predicted to show specificity for this amino acid) which is then bound to the PCP of GeM (didomain protein). The putative halogenase GeL would then chlorinate the PCP-bound histidine to afford chlorohistidine, as has been proposed for the chlorination of L-threonine in the syringomycin biosynthesis [146]. The hydroxylation could then occur on the Cl-His-S-GeM intermediate; alternatively hydroxylation would occur on the free histidine (prior to PCP attachment), or following release of the tetrapeptide from the synthetase. However, as the majority of hydroxylations have been shown to occur on PCP-bound amino acids [148], we believe that this mechanism is most likely. Nonetheless, it is unclear whether hydroxylation proceeds or follows the chlorination. The 3-hydroxy-6-chlorohistidine would
then be transferred onto the PCP of GeJ. There are several examples of such PCP-to-PCP shuttling in the literature. In the syringomycin pathway, for example, the freestanding A-PCP protein SyrB1 activates and loads threonine. The halogenase SyrB2 then acts on the Thr-S-SyrB1 intermediate to produce 4-Cl-Thr-S-SyrB1. The Cl-Thr is then shuttled \textit{in trans} to the PCP of a C-PCP didomain of the NRPS synthetase SyrE. This reaction is catalyzed by the acyltransferase SyrC, a protein belonging to the hydrolase superfamily [157]. Postulated functions for members of this protein family also include thioesterase and haloperoxidase [158]. A similar mechanism has been elucidated recently for zorbamycin biosynthesis [145], in which L-valine is loaded onto the PCP domain of ZmbVIIb, hydroxylated by ZmbVIIc, and subsequently transferred onto the PCP domain of ZmbVIIa by the predicted type II thioesterase ZmbVIIId (Figure 4.12).

![Figure 4.12: Proposed mechanism for the biosynthesis and incorporation of L-hydroxyvaline in the zorbamycin biosynthetic pathway.](image)

The coronamic acid biosynthetic pathway incorporates a further example of the shuttling of aminoacyl moieties [159]. In this example, the aminoacyltransferase CmaE shuttles an \textit{allo}–Ile moiety between the pantetheinyl arm of the PCP domain of CmaA (A-PCP didomain) and the freestanding PCP domain of CmaD. By mechanistic analogy, the histidine in GE81112 biosynthesis may be loaded onto the PCP domain of GeM, chlorinated by GeL and/or hydroxylated by GeI, and subsequently transferred onto the GeJ PCP domain by the predicted type II thioesterase GeA, acting as acyltransferase (Figure 4.11A). In this model, the A domain in module 4 would not be necessary.

Alternatively (pathway B) histidine is loaded onto the PCP of GeM and chlorinated by GeL leading to chlorohistidine. The chlorohistidine would then be released by GeA acting as a thioesterase, as described for BarC from the barbamide biosynthetic pathway [160]. The free chlorohistidine would then be activated by the A domain GeJA4 and loaded onto the PCP of module 4. In the next step, $\beta$-hydroxylation catalyzed by GeI would take place on the PCP-bound species.
In these proposed pathways, GeA is predicted to function either as an acyltransferase or as a type II thioesterase, both of which are known to release amino acids from PCP domains [100]. To attempt to provide support for one or the other of the pathways, we aligned GeA against other known acyltransferases/thioesterases. The aminoacyltransferases identified to date have been assigned into two groups [145]. The first group, comprising SyrC, CmaE, ZmbVIIId, contains a GxCxG motif harboring an active site cysteine; enzymes in this group are predicted to act as acyltransferases. Indeed, the active-site cysteine of these enzymes has been shown to be directly involved in aminoacyltransfer [159]. The second group contains acyltransferases BarC [160] and CouN7 [161], which both contain an active site serine (GxSxG) and are predicted to function as ordinary thioesterases. For example, in the barbamide biosynthetic pathway, BarC is proposed to release trichloroleucine by hydrolysis [160]. When we aligned the two thioesterase-like proteins GeA and GeN from the GE81112 biosynthetic gene cluster, the presence of an active site serine was clear, suggesting that the proteins both function as normal thioesterases (Figure 4.13). Thus, it seems likely that GeA is acting as a type II TE catalyzing the release of chlorohistidine (pathway B). The second thioesterase, GeN, is expected to catalyze the release of the peptide from the assembly line (Figure 4.14), even though it is a discrete protein (i.e. type II architecture).

![Sequence alignment of acyltransferase/thioesterases found in NRPS clusters.](image)

The final alternative (pathway C) is that free histidine is chlorinated, and then activated and loaded onto the PCP of module 4. The remainder of the pathway would then follow that of proposal B. Thus in pathway C, module 5 does not play a role, but according to sequence analysis it is probably active.
Overall, pathway B appears to be the most reasonable, as the TE is clearly predicted to act as a thioesterase catalyzing the release of the chlorohistidine, rather than its transfer to GeJ. Further supporting pathway B is the fact that in pathways A and C, no function is required of a single domain or a didomain, respectively, when there is no evidence from sequence analysis that any of these domains should be inactive. Indeed, there are a number of examples in the literature in which apparently superfluous modules have been shown to be essential for natural product biosynthesis [157;159-161].
4.3.4 Summary and Future aspects

The cloned GE81112 biosynthetic pathway could be identified and cloned and comprises a complex NRPS biosynthetic machinery. Many unusual features were found in its modular architecture, including A domains which likely incorporate unusual amino acids, such as pipecolic acid and chlorohistidine. Although many questions about the biosynthesis remain, the availability of the GE81112 cluster as well as tools for genetic manipulation now provide a platform for attempts to decipher these issues, and to generate new GE81112 derivatives by genetic engineering. A future goal will be to define the cluster boundaries by further gene inactivation experiments. To obtain further experimental evidence for the proposed biosynthetic pathway selected proteins can be studied in vitro. At first the three remaining A domains should be expressed and tested in the ATP assay for the activation of the suggested substrates. Furthermore it would be interesting to characterize the PCP of module 5 to demonstrate that this module is really needed for the biosynthesis. Another possibility to further investigate the gene cluster and to increase the production of the GE81112 compounds would be the heterologous expression. Therefore the cluster would need to be stitched from the two cosmids as it was already described for the phenalinolactone cluster and transferred into heterologous hosts, such as *S. lividans*.

Additionally to the GE81112 biosynthetic gene cluster another novel NRPS cluster could be identified from *Streptomyces* 14386. Successful heterologous expression of this gene cluster revealed that it was responsible for the biosynthesis of three new peptides, which we speculated to be related to the GE81112 family as they seemed to be also chlorinated tetrapeptides. To get an idea about the structure of this compound family we purified all three compounds and tried to elucidate their structure by NMR. Although structure elucidation was not completely finished, it turned out that these substances incorporate unusual amino acids. To finally determine if these substances are related to the GE81112 family, full structure elucidation needs to be done. Once the structure elucidation is finished investigations on the mode of action of these substances can also be done. However, with the heterologous expression of these substances and the genetic manipulation methods developed for *S. 14386*, we now have the basis for gene deletion experiments to investigate the biosynthesis of this new class of peptides.
## Appendix

### Table 5.1: Sequence alignment of putative modules 1-5 from the GE81112 biosynthetic pathway with two modules of the NRPS pathway from cosmid FD10.

<table>
<thead>
<tr>
<th>C-A-PCP_FD10_2</th>
<th>1</th>
<th>---------------------------</th>
<th>MLFLALYDEQAVDTYTFQLAFELEGDVDAAGLRAAGDALLRRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>44</td>
<td>PN LR---GRHEKLSRPVQV---HRPF</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>C-A-PCP_F D10_3</td>
<td>59</td>
<td>ES L---IPAGEIDGSRQ---GQLQV</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>124</td>
<td>DM LRAQALPRGGPTEAELTVLDRLPDGVPPLRVESG</td>
<td>---FALDRAPLFELRLL</td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>124</td>
<td>DM LRAQALPRGGPTEAELTVLDRLPDGVPPLRVESG</td>
<td>---FALDRAPLFELRLL</td>
</tr>
<tr>
<td>C-A-PCP-3</td>
<td>129</td>
<td>QA LRTVFRTEPEGLRQSIVLTPGEAEPE--DARFAACVDERAVAYHPATGQAHH</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>---</td>
<td>-----------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>216</td>
<td>ET--TAARSSARALRNYTLRVLQTH---</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>302</td>
<td>TDTSGDPAFDDELVCFERVSQEEPRGSMVGLAQQAEARLELPGL</td>
<td>---</td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>293</td>
<td>RPESQ VVEDLLPDDLPATSDLWLAVYARVKGDELHLDLTYDDNLLDEAEAGAYADDVAAVLRGLAAGATAPP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>293</td>
<td>RPESQ VVEDLLPDDLPATSDLWLAVYARVKGDELHLDLTYDDNLLDEAEAGAYADDVAAVLRGLAAGATAPP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP-3</td>
<td>317</td>
<td>GEDTTRA VQGLCRELGTTSFTAMVAAVRVLLLRWSGADDIVLGTVVSGRDQPE--LADSVGLFVNTVALRTT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>---</td>
<td>-----------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Protein</td>
<td>Length</td>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>442</td>
<td>ARLPLTGAAADGLAKS GIRLKLVLKTVTVLVAATVVGSAEHTDEPVVAVSAGSPGVPVLCLIGIAAAPATVAD</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>469</td>
<td>LVRADVAALRSAPAADVKDAALAGAVLVSSAVGTGAGAPAVKRDAAELTLSEPGADGGRDLVALAEEAPERA</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>502</td>
<td>LRRVQGAPLRERIGHDN---SAGGRLSTNRDPPK---</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>501</td>
<td>LLLTVAAARLALAALDEESRHVVLVPAPTNASRADADTAADTLACPACPLDRSGAPASVFLTSINHEL</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_3</td>
<td>484</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>1</td>
<td>APSSPPHAGPVVDTVGSTVLELIDRPAHPGDPETP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>442</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>469</td>
<td>VTLPGLFEAQVVRAPG</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>634</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>502</td>
<td>LRRVQGAPLRERIGHDN---SAGGRLSTNRDPPK---</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_3</td>
<td>484</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>1</td>
<td>APSSPPHAGPVVDTVGSTVLELIDRPAHPGDPETP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>458</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>469</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>634</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>502</td>
<td>LRRVQGAPLRERIGHDN---SAGGRLSTNRDPPK---</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_3</td>
<td>484</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>1</td>
<td>APSSPPHAGPVVDTVGSTVLELIDRPAHPGDPETP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>458</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>469</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>634</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>502</td>
<td>LRRVQGAPLRERIGHDN---SAGGRLSTNRDPPK---</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_3</td>
<td>484</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>1</td>
<td>APSSPPHAGPVVDTVGSTVLELIDRPAHPGDPETP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>458</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>469</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>634</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>502</td>
<td>LRRVQGAPLRERIGHDN---SAGGRLSTNRDPPK---</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_3</td>
<td>484</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>1</td>
<td>APSSPPHAGPVVDTVGSTVLELIDRPAHPGDPETP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>458</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>469</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>634</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>502</td>
<td>LRRVQGAPLRERIGHDN---SAGGRLSTNRDPPK---</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_3</td>
<td>484</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>1</td>
<td>APSSPPHAGPVVDTVGSTVLELIDRPAHPGDPETP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>458</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>469</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>634</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>502</td>
<td>LRRVQGAPLRERIGHDN---SAGGRLSTNRDPPK---</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_3</td>
<td>484</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>1</td>
<td>APSSPPHAGPVVDTVGSTVLELIDRPAHPGDPETP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>458</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>469</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>634</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>502</td>
<td>LRRVQGAPLRERIGHDN---SAGGRLSTNRDPPK---</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_3</td>
<td>484</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>1</td>
<td>APSSPPHAGPVVDTVGSTVLELIDRPAHPGDPETP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>458</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>469</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>634</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>502</td>
<td>LRRVQGAPLRERIGHDN---SAGGRLSTNRDPPK---</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_3</td>
<td>484</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>1</td>
<td>APSSPPHAGPVVDTVGSTVLELIDRPAHPGDPETP</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_2</td>
<td>458</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_FD10_3</td>
<td>469</td>
<td>ST ALVFGGESVSYGELNARANRLARLFIARG-VGPERFVAVSLPRSVELVVAVLAVLKAGGAYVPVDPDY</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_4</td>
<td>634</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_5</td>
<td>502</td>
<td>LRRVQGAPLRERIGHDN---SAGGRLSTNRDPPK---</td>
<td></td>
</tr>
<tr>
<td>C-A-PCP_3</td>
<td>484</td>
<td>EAWFLEVLLRVARVLAGFTDPHDAVADLPRAAGADDVAAATDFGRRFEDGVTITLAGZAVATT</td>
<td></td>
</tr>
<tr>
<td>A-PCP_Loading1</td>
<td>1</td>
<td>APSSPPHAGPVVDTVGSTVLELIDRPAHPGDPETP</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.2: Sequence alignment of proteins GeJ and GeM from the GE81112 biosynthetic pathway

Sequence homology was 29%. Protein GeJ shows an N-terminal linker sequence which is not present in GeM.

<table>
<thead>
<tr>
<th>Section 1</th>
<th>Section 2</th>
<th>Section 3</th>
<th>Section 4</th>
<th>Section 5</th>
<th>Section 6</th>
<th>Section 7</th>
<th>Section 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>gel</strong></td>
<td>(1)</td>
<td>LTPQGRVEMTSDSHAVPATTYVPRRGLPAHSGAPARRALPTGAAAAGLAKVSGHRFL</td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
<td>LTPQGRVEMTSDSHAVPATTYVPRRGLPAHSGAPARRALPTGAAAAGLAKVSGHRFL</td>
<td>(1)</td>
</tr>
<tr>
<td><strong>geM</strong></td>
<td>(1)</td>
<td>---------</td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
<td>---------</td>
<td>(1)</td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td>(1)</td>
<td>---------</td>
<td>(1)</td>
<td>---------</td>
<td>(1)</td>
<td>---------</td>
<td>(1)</td>
</tr>
<tr>
<td>(58)</td>
<td>100</td>
<td>80</td>
<td>100</td>
<td>114</td>
<td>115</td>
<td>120</td>
<td>130</td>
</tr>
<tr>
<td><strong>gel</strong></td>
<td>(58)</td>
<td>---------</td>
<td>(58)</td>
<td>---------</td>
<td>(58)</td>
<td>---------</td>
<td>(58)</td>
</tr>
<tr>
<td><strong>geM</strong></td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td>(58)</td>
<td>---------</td>
<td>(58)</td>
<td>---------</td>
<td>(58)</td>
<td>---------</td>
<td>(58)</td>
</tr>
<tr>
<td>(115)</td>
<td>115</td>
<td>120</td>
<td>130</td>
<td>140</td>
<td>150</td>
<td>160</td>
<td>171</td>
</tr>
<tr>
<td><strong>gel</strong></td>
<td>(115)</td>
<td>---------</td>
<td>(115)</td>
<td>---------</td>
<td>(115)</td>
<td>---------</td>
<td>(115)</td>
</tr>
<tr>
<td><strong>geM</strong></td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td>(115)</td>
<td>---------</td>
<td>(115)</td>
<td>---------</td>
<td>(115)</td>
<td>---------</td>
<td>(115)</td>
</tr>
<tr>
<td>(172)</td>
<td>172</td>
<td>180</td>
<td>190</td>
<td>200</td>
<td>210</td>
<td>220</td>
<td>230</td>
</tr>
<tr>
<td><strong>gel</strong></td>
<td>(172)</td>
<td>---------</td>
<td>(172)</td>
<td>---------</td>
<td>(172)</td>
<td>---------</td>
<td>(172)</td>
</tr>
<tr>
<td><strong>geM</strong></td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(1)</td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td>(172)</td>
<td>---------</td>
<td>(172)</td>
<td>---------</td>
<td>(172)</td>
<td>---------</td>
<td>(172)</td>
</tr>
<tr>
<td>(229)</td>
<td>229</td>
<td>240</td>
<td>250</td>
<td>260</td>
<td>270</td>
<td>280</td>
<td>290</td>
</tr>
<tr>
<td><strong>gel</strong></td>
<td>(229)</td>
<td>---------</td>
<td>(229)</td>
<td>---------</td>
<td>(229)</td>
<td>---------</td>
<td>(229)</td>
</tr>
<tr>
<td><strong>geM</strong></td>
<td>(43)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(43)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(43)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(43)</td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td>(229)</td>
<td>---------</td>
<td>(229)</td>
<td>---------</td>
<td>(229)</td>
<td>---------</td>
<td>(229)</td>
</tr>
<tr>
<td>(266)</td>
<td>266</td>
<td>280</td>
<td>290</td>
<td>300</td>
<td>310</td>
<td>320</td>
<td>330</td>
</tr>
<tr>
<td><strong>gel</strong></td>
<td>(266)</td>
<td>---------</td>
<td>(266)</td>
<td>---------</td>
<td>(266)</td>
<td>---------</td>
<td>(266)</td>
</tr>
<tr>
<td><strong>geM</strong></td>
<td>(98)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(98)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(98)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(98)</td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td>(266)</td>
<td>---------</td>
<td>(266)</td>
<td>---------</td>
<td>(266)</td>
<td>---------</td>
<td>(266)</td>
</tr>
<tr>
<td>(343)</td>
<td>343</td>
<td>350</td>
<td>360</td>
<td>370</td>
<td>380</td>
<td>390</td>
<td>399</td>
</tr>
<tr>
<td><strong>gel</strong></td>
<td>(343)</td>
<td>---------</td>
<td>(343)</td>
<td>---------</td>
<td>(343)</td>
<td>---------</td>
<td>(343)</td>
</tr>
<tr>
<td><strong>geM</strong></td>
<td>(155)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(155)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(155)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(155)</td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td>(343)</td>
<td>---------</td>
<td>(343)</td>
<td>---------</td>
<td>(343)</td>
<td>---------</td>
<td>(343)</td>
</tr>
<tr>
<td>(400)</td>
<td>400</td>
<td>410</td>
<td>420</td>
<td>430</td>
<td>440</td>
<td>450</td>
<td>468</td>
</tr>
<tr>
<td><strong>gel</strong></td>
<td>(400)</td>
<td>---------</td>
<td>(400)</td>
<td>---------</td>
<td>(400)</td>
<td>---------</td>
<td>(400)</td>
</tr>
<tr>
<td><strong>geM</strong></td>
<td>(201)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(201)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(201)</td>
<td>RILVGLTAAATVUGSARHTEPDVUVGSPSGFVULCSGIAMAAAPAVDLRVAADAL</td>
<td>(201)</td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td>(400)</td>
<td>---------</td>
<td>(400)</td>
<td>---------</td>
<td>(400)</td>
<td>---------</td>
<td>(400)</td>
</tr>
</tbody>
</table>

---

**Notes:**
- Sequence homology was 29%.
- Protein GeJ shows an N-terminal linker sequence which is not present in GeM.
6 References


Lebenslauf/Curriculum vitae

Tina Maria Binz
Diplom-Biologin
mit Schwerpunkt Human-und Molekularbiologie
* 27.09.1979 in Saarlouis, Deutschland

Am Homburg 59
66123 Saarbrücken

Schulausbildung:

1986-1990 Grundschule Bachem
1990-1999 Gymnasium am Stefansberg, Merzig, Abitur

Hochschulstudium:

10/1999-12/2004 Studium der Human-und Molekularbiologie an der Universität des Saarlandes, Saarbrücken
10/2001 Diplomvorprüfung
09/2002-10/2002 Forschungspraktikum im Rahmen des Studiums, am Institut für Technische Biochemie unter der Leitung von Prof. Dr. Ing. E. Heinzle, Universität des Saarlandes
Titel: „Analytische Methoden zur Quantifizierung intrazellulärer Metabolite in Bakterien“
08/2003-11/2003 Forschungspraktikum im Rahmen des ERASMUS Programms an der „University of Edinburgh“, Schottland, am Institut für Reproduktive Biologie unter der Leitung von Prof. Dr. Ian Mason
Titel: „Construction and characterization of fusion proteins containing human type2 3β-hydroxysteroid dehydrogenase and human type1 11β-hydroxysteroid dehydrogenase“
02/2002 Diplom Prüfung
04/2004-12/2004 Diplomarbeit unter der Leitung von Prof. Dr. Ing. E. Heinzle am Institut für Technische Biochemie, Universität des Saarlandes
Titel: „Physiologie von Corynebacterium glutamicum- Einfluss des Nährmediums auf die Lysin Produktion“
01/2005 Praktikum unter der Leitung von Prof. Dr. Rolf Müller am Institut für Pharmazeutische Biotechnologie, Universität des Saarlandes
02/2005 Beginn der Promotion unter der Leitung von Prof. Dr. Rolf Müller,
Titel: “Studies on the biosynthesis and heterologous expression of complex secondary metabolites from streptomycetes“