A cryptic type I polyketide synthase (cpk) gene cluster in *Streptomyces coelicolor* A3(2)

Krzysztof Pawlik · Magdalena Kotowska · Keith F. Chater · Katarzyna Kuczek · Eriko Takano

Received: 2 June 2006 / Revised: 18 August 2006 / Accepted: 1 September 2006 / Published online: 29 September 2006 © Springer-Verlag 2006

**Abstract** The chromosome of *Streptomyces coelicolor* A3(2), a model organism for the genus *Streptomyces*, contains a cryptic type I polyketide synthase (PKS) gene cluster which was revealed when the genome was sequenced. The ca. 54-kb cluster contains three large genes, *cpkA*, *cpkB* and *cpkC*, encoding the PKS subunits. *In silico* analysis showed that the synthase consists of a loading module, five extension modules and a unique reductase as a terminal domain instead of a typical thioesterase. All acyltransferase domains are specific for a malonyl extender, and have a B-type ketoreductase. Tailoring and regulatory genes were also identified within the gene cluster. Surprisingly, some genes show high similarity to primary metabolite genes not commonly identified in any antibiotic biosynthesis cluster. Using western blot analysis with a PKS subunit (CpkC) antibody, CpkC was shown to be expressed in *S. coelicolor* at transition phase. Disruption of *cpkC* gave no obvious phenotype.

**Keywords** *Streptomyces* · Polyketide biosynthesis · Post-polyketide modifications · Antibiotic biosynthesis

**Abbreviations**
- AT Acyltransferase
- ACP Acyl carrier protein
- DH Dehydratase
- DEBS 6-Deoxyerythronolide synthase
- ER Enoyl reductase
- KR Ketoreductase
- KS *afii9826*-Ketoacylsynthase
- MS Mannitol soya flour medium
- NRPS Non-ribosomal peptide synthetase
- PKS Polyketide synthase
- SMM Supplemented minimal medium
- TD Terminal reductase

**Introduction**

Polyketides are a large and structurally diverse group of natural products synthesised by multifunctional or mono- or bi-functional enzymes called polyketide synthases (PKSs) by repetitive condensations of small carboxylic acid units in a manner similar to fatty acid synthesis (Staunton and Weissman 2001). Many of them are of industrial and medical importance (Weber et al. 2003). In bacteria, all the genes required for the biosynthesis of a given secondary metabolite are usually clustered.

There are two main classes of PKSs: types I and II. Type I PKSs consist of multi-functional proteins, whereas type II synthases are complexes of separate enzymes (for review, see Shen 2003). Among type I PKSs, the most studied are enzymes in which catalytic domains are grouped in modules (for review, see Shen...
and the set of modules forms a protein ‘assembly line’ along which the growing carbon chain passes. Each module contains a set of enzymatic domains—β-ketoacylsynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP)—necessary to perform a condensation between an acyl-CoA precursor and the nascent carbon chain for one round of chain extension. Before the next condensation, the β-carbonyl of the incorporated dicarbon unit may be modified to a hydroxyl, a double bond, or a fully saturated carbon depending on the presence of reduction/dehydration domains: ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER). In the case of type I PKS, the order of carbon units incorporated into the polyketide is determined by the order of modules. Successive modules in one protein are separated by linkers whose role is unclear. Correct transfer from one multienzyme to the next is achieved by docking domains which link and maintain the order of proteins (Broadhurst et al. 2003). Finally, a thioesterase domain typically catalyses removal of the completed carbon chain from the PKS and in some cases is involved in cyclisation. Apart from genes for the PKS subunits, most PKS gene clusters also contain a number of other genes, including genes for modifying enzymes, enzymes involved in the synthesis of sugars added to the primary polyketide skeleton during post-polyketide modifications, resistance/transport genes, as well as regulatory proteins.

Numerous DNA sequences of PKS gene clusters along with the structure of the end product have been reported in the past few years. Conservation of the PKS subunits in terms of function and enzyme domains enables in silico analysis of a newly sequenced PKS gene cluster (Yadav et al. 2003).

*Streptomyces coelicolor* is genetically the model organism for the genus *Streptomyces*, with a sequenced genome (Bentley et al. 2002). This organism produces several structurally and genetically characterised polyketide secondary metabolites. Two of these have been extensively studied: actinorhodin (Rudd and Hopwood 1979), made by a type II PKS and undecylprodigiosin and other prodiginines (Feitelson et al. 1985), which are synthesised by a non-ribosomal peptide synthetase (NRPS) and a type I PKS (Cerdeno et al. 2001). Several other gene clusters have also been deduced to encode polyketides though their structures have not been determined. For example, the whiE cluster contains genes encoding the grey spore pigment, which is produced by a type II PKS (Yu and Hopwood 1995); and in silico analysis of the genome sequence has identified two further PKS gene clusters (Bentley et al. 2002; Challis and Hopwood 2003).

Here, we analyse one of these clusters encoding a cryptic modular type I PKS (Cpk cluster) in *S. coelicolor*, part of which had been partially identified by hybridisation using a AT gene as probe (Kuczek et al. 1997) before full characterisation of the cluster by genome sequencing (Bentley et al. 2002). The cluster was recently found to be the principle target of the extracellular regulatory molecule SCB1, a γ-butyrolactone (Takano et al. 2005). However, the product of this gene cluster has not been determined. Using sequence comparison to known modular PKSs and primary structure analysis of the predicted synthase, we attempt to describe features of the first possible intermediate produced by this gene cluster. We have demonstrated by western analysis that the Cpk synthase is expressed at transition phase, and we have also disrupted cpkC.

### Materials and methods

**Bacterial strains, plasmids, growth conditions and media**

*E. coli* DH5α, XL1-Blue (Stratagene), ET12567 were cultivated and transformed according to Sambrook et al. (1989). *S. coelicolor* strains M145 (Kieser et al. 2000) and P100 (this study) were manipulated as in Kieser et al. (2000). Vectors used were pBluescriptSK(+) (Stratagene), pKC1132, pGEM-T-easy (Promega), and pGEX-KG (Amersham-Pharmacia). Solid media MS (mannitol soya flour medium) and supplemented minimal medium (SMM) were prepared according to Kieser et al. (2000) and Takano et al. (2001), respectively, and medium 79 was prepared as in IMET Catalogue of strains (Jena) 1987.

**Isolation of DNA, PCR and DNA sequencing**

Restriction fragments were cloned into pBluescript SK or pGEM-T-easy (Promega). DNA sequencing, DNA isolation, and Southern hybridisation were performed as described in Kuczek et al. (1997). PCR was performed for 31 amplification cycles (40 s at 95°C, 40 s at 54°C, 40 s at 72°C followed by 3 min extension at 72°C) with PrimeZyme Taq polymerase.

**Nucleotide sequence analysis of ORFs**

Open reading frames were determined using Codon Preference (Wisconsin Package Genetic Computing Group) and Frame Plot (http://www.nih.go.jp/~jun/cgi-bin/frame-plot.pl) (Ishikawa and Hotta 1999). Results were
Antibacterial activity

The parent M145, and P100 were grown on DNA plates for two days at 30°C then flooded with *E. coli* or *Bacillis subtilis* grown overnight which was then diluted to 1/100 in LB. The flooded plate was grown at 37°C for overnight and was assessed for inhibition of growth.

UV absorbance detection

The parent M145 and P100 were grown in 79 liquid medium for 2 days at 30°C. The medium at 24 and 48 h, cell extracts at 48 h were subjected to UV–Vis scanning.

Overexpression, and purification of a fragment of CpkC, and anti-CpkC polyclonal antibody preparation

An N-terminal fragment of CpkC named CpkC-apex (218 aa) was obtained as a fusion protein with glutathione-S-transferase (GST) in the pGEX system (Amer-sham-Pharmacia). A DNA fragment covering the 5′ part of *cpkC* was amplified by PCR with CPKCFw (5′-CACGAATT CAATAACCTCAAGTGCGT-3′) which includes an *Eco*RI site generated in the primer (in bold) and T7 primers using the T7-terminal fragment of cosmid 1G7 cloned in pBluescriptSK(+) as template. The PCR product was cloned in pGEM-T-easy and a 669-bp *Eco*RI–*SaI* restriction fragment was recloned into pGEX-KG which results in an in-frame cloning into CpkC.

Recombinant protein was expressed in *E. coli* DH5α grown at 28°C until the OD550nm reached 0.6 and induced with 0.01 mM IPTG for 1 h. Bacteria collected from a 20-l culture were washed with GST-A buffer [50 mM Tris–HCl pH8.0, 100 mM NaCl, 1 mM EDTA, 1 mM phenylmethylsulphonylfluoride (PMSF)] centrifuged and resuspended in the same buffer. The cells were treated with lysozyme (0.2 mg/ml lysozyme, 30 min on ice), and a crude extract was obtained by ultrasonic disintegration (6 × 2 min Branson Sonifier, power 4) and centrifuged at 20,000g for 40 min. Fusion protein from the supernatant was bound to a glutathione-Sepharose column, and then digested on the column with thrombin (1 μg/ml thrombin in 50 mM Tris–HCl pH 8.0, 2 mM CaCl₂). Free CpkC-apex protein was eluted with PBS. Rabbits were immunised with 0.25 mg CpkC-apex protein (three times at 30-day intervals, for 90 days). Antibodies were purified from serum by salting-out with 20% ammonium sulphate followed by dialysis against 0.5 M Tris–HCl pH8.0. Polishing purification was performed on a 1-ml CNBr-sepharose 4B (Pharmacia) column with 2 mg of bound CpkC-apex protein.

Western blot analysis of CpkC

Wild type *S. coelicolor* M145 and the P100 mutant strain were cultivated in liquid SMM and cells were harvested at different time points from 14 to 25 h. Bacterial growth was monitored by OD450nm measurement. Cells from 10 ml of culture were centrifuged, sonicated in 0.5 ml 0.05 M Tris–HCl pH 8.0 buffer and centrifuged. Proteins from the supernatant were precipitated with three volumes of acetone. The pellet was resuspended in 100 μl 0.05 M Tris–HCl pH 8.0. Fifteen microlitres of crude extracts were used in SDS-PAGE for western blot analysis. Proteins were separated in a 6% SDS-PAGE gel, then transferred to PVDF membrane by wet blotter (2 h, 60 V). The membrane was developed with N-terminal CpkC antibody using a 1:1,000 dilution and NBT-BCIP colour reaction.
Results and discussion

In silico sequence analysis of the cpk gene cluster

We had previously cloned and sequenced a 450 bp S. coelicolor M145 chromosomal DNA fragment encoding a type I PKS AT domain (Kuczek et al. 1997) (accession number: U88833), and accessory genes contained in two BamHI fragments at the T3 terminus of cosmid 1G7 (scoT and a partial scbR2) (Kotowska et al. 2002) (accession number: AF109727). In subsequent work, further fragments identified using the 450 bp fragment were shown to encode an N-terminal docking domain, KS domain and KS–AT linking region (accession number AF202898); AT and DH domains and the N-terminal region of a KR domain (U88833 updated).

Further sequencing experiments were stopped as the S. coelicolor genome sequencing project began at the Sanger Institute. Completion of the genome sequence (Bentley et al. 2002) confirmed our results and enabled us to identify a new type I modular PKS gene cluster which we have named Cpk (Fig. 1a and Table 1). The exact boundaries for the Cpk cluster are not yet known and further experiments will be needed to define this. However using the microarray analysis of S. coelicolor M145 which showed that the expression of 20 genes in the cluster is co-ordinated (Takano et al. 2005), we have chosen to analyse in detail SCO6269-SCO6288 (6,894,148–6,948,414 bp, numbering from Bentley et al. 2002) which extend over 54.2 kb.

Structure of the PKS subunits

The three largest open reading frames encoding CpkA, CpkB and CpkC (Fig. 1a) have high homology to each other and to well-characterised type I PKS domains (Table 1). The highest homology is in sequences encoding domains engaged in extension of the carbon chain (95% aa identity), while there is lower homology for reducing domains (48% aa identity).

Fig. 1 cpk gene cluster organisation and PKS domains. a The top line denotes the S. coelicolor chromosome. Lines underneath denote the cosmids overlapping the cpk gene cluster. ORFs are shown to scale as arrows. Gene names are given above the arrows with the cpk cluster genes in bold. The SCO numbers and bp numbers correspond to Bentley et al. (2002). Note that cpkO was previously called kasO (Takano et al. 2005). b Modules (boxed on top) and enzymatic domains within the PKS subunits cpkA, cpkB and cpkC are indicated in the grey open arrow. Intermodular linkers are marked with black squares, and docking domains with white spaces. The N-terminus region of CpkC, CpkC-apex, used to prepare antibody is shown as a gridded box and the fragment used to prepare the P100 mutant is shown as a shaded box with the name of the plasmid. The predicted growing carbon chains are shown and the putative non-active dehydratase domain is crossed. “Loading m” denotes loading module
CpkA consists of a loading module and two extension modules 1 and 2, CpkB contains extension modules 3 and 4, and CpkC contains module 5 and a reductase domain (Fig. 1b). No thioesterase domain or ER domain was present in any of the modules. All extension modules and also the loading module contain the minimal set of enzyme domains: AT, KS, and ACP. The fact that among the three proteins, one has a loading module and one an unusual reductase domain in the typical thioesterase position suggests that the process of polyketide chain elongation occurs from CpkA through to CpkC. This is also confirmed by the presence of docking domains of 70 and 74 aa at the C-terminal of CpkA and CpkB, respectively. Similarly, the N-terminal ends of CpkB and CpkC are homologous to N-terminal docking domains of DEBS3 which is the third multienzyme subunit of 6-deoxyerythronolide B synthase which assembles the polyketide core of erythromycin A (Broadhurst et al. 2003). Sequences coding for both junctions between synthase subunits are not so similar as to exclude the possibility of iterative docking of the C-terminal end of CpkB and the N-terminal end of another CpkB.

We detected unusually high nucleotide sequence similarity between parts of CpkABC. In modules 3, 4 and 5, there are fragments of 750 bp within the AT

### Table 1 Deduced functions of the Cpk cluster

<table>
<thead>
<tr>
<th>Gene product</th>
<th>SCO number</th>
<th>Size&lt;sup&gt;b&lt;/sup&gt;</th>
<th>ID/SIM&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Protein homologue&lt;sup&gt;d&lt;/sup&gt; (Acc. no.)</th>
<th>Domains/putative homologue function&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>CpkPβ</td>
<td>6269</td>
<td>350</td>
<td>95/97</td>
<td>*Orf14 S. carzinostaticus (BAD38883)</td>
<td>Oxidoreductase β-subunit</td>
</tr>
<tr>
<td>CpkPα</td>
<td>6270</td>
<td>630</td>
<td>92/96</td>
<td>Orf13 S. carzinostaticus (BAD38882)</td>
<td>Oxidoreductase α-subunit</td>
</tr>
<tr>
<td>AccA1</td>
<td>6271</td>
<td>590</td>
<td></td>
<td></td>
<td>Acyl-CoA carboxylase α-subunit (Rodriguez and Gramajo 1999)</td>
</tr>
<tr>
<td>ScF</td>
<td>6272</td>
<td>550</td>
<td>53/64</td>
<td>AcoO S. galilaeus (BAB72054.1)</td>
<td>Secrected FAD-binding protein</td>
</tr>
<tr>
<td>CpkC</td>
<td>6273</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module 5</td>
<td>1,718</td>
<td>55/68</td>
<td></td>
<td>S. aizunensis (AY899214)</td>
<td>KS AT DH KR ACP</td>
</tr>
<tr>
<td>End domain</td>
<td>401</td>
<td>39/56</td>
<td></td>
<td>hetM Nostoc sp. PCC 7120 (L22883)</td>
<td>TD</td>
</tr>
<tr>
<td>CpkB</td>
<td>6274</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module 4</td>
<td>1,700</td>
<td>52/64</td>
<td></td>
<td>S. aizunensis (AY899214)</td>
<td>KS AT DH KR ACP</td>
</tr>
<tr>
<td>Module 3</td>
<td>1,740</td>
<td>56/67</td>
<td></td>
<td>S. aizunensis (AY899214)</td>
<td>KS AT DH KR ACP</td>
</tr>
<tr>
<td>CpkA</td>
<td>6275</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module 2</td>
<td>1,726</td>
<td>57/68</td>
<td></td>
<td>S. aizunensis (AY899214)</td>
<td>KS AT DH KR ACP</td>
</tr>
<tr>
<td>Module 1</td>
<td>1,717</td>
<td>51/68</td>
<td></td>
<td>S. aizunensis (AY899214)</td>
<td>KS AT DH KR ACP</td>
</tr>
<tr>
<td>1. module</td>
<td>1,015</td>
<td>54/66</td>
<td></td>
<td>ptaA1 filament gene cluster</td>
<td>KS AT ACP</td>
</tr>
<tr>
<td>CpkD</td>
<td>6276</td>
<td>440</td>
<td>50/67</td>
<td>SgcD2 S. globisporus (AAL06669.1)</td>
<td>Secreted monoxygenase</td>
</tr>
<tr>
<td>CpkE</td>
<td>6277</td>
<td>286</td>
<td>43/54</td>
<td>Nocardoides sp. JS614 (ZP 00658625.1)</td>
<td>Epoxide hydrolase</td>
</tr>
<tr>
<td>CpkF</td>
<td>6278</td>
<td>543</td>
<td>45/61</td>
<td>*SgcB S. globisporus (AAF13999.1)</td>
<td>Transmembrane efflux protein</td>
</tr>
<tr>
<td>CpkG</td>
<td>6279</td>
<td>518</td>
<td>38/55</td>
<td></td>
<td>Aminotransfer</td>
</tr>
<tr>
<td>CpkO</td>
<td>6280</td>
<td>543</td>
<td></td>
<td></td>
<td>SARP regulator, formerly KasO (Takano et al. 2005)</td>
</tr>
<tr>
<td>CpkH</td>
<td>6281</td>
<td>545</td>
<td>49/61</td>
<td>*AcoO S. galilaeus (BAB72054.1)</td>
<td>Secrected FAD-binding protein</td>
</tr>
<tr>
<td>CpkI</td>
<td>6282</td>
<td>265</td>
<td>54/68</td>
<td>JadW3 S. venezuelae (AAL23836.1)</td>
<td>3-Oxoacyl-ACP reductase</td>
</tr>
<tr>
<td>CpkJ</td>
<td>6283</td>
<td>281</td>
<td>33/45</td>
<td>Synchococcus sp WH 5701 (ZP 0183680.1)</td>
<td>Nucleoside-diphosphate-sugar epimerase</td>
</tr>
<tr>
<td>CpkK</td>
<td>6284</td>
<td>530</td>
<td>78/85</td>
<td>Orf1 S. griseus (AAQ08909.1)</td>
<td>Acyl-CoA carboxylase, β-subunit</td>
</tr>
<tr>
<td>CpkL</td>
<td>6285</td>
<td>87</td>
<td>34/50</td>
<td>GrrH Streptomyces sp. J95 (AMM33660.1)</td>
<td>Unknown hypothetical protein</td>
</tr>
<tr>
<td>ScbR2</td>
<td>6286</td>
<td>224</td>
<td>48/56</td>
<td>AplW S. ambofaciens (AAR30167)</td>
<td>Butyrolactone receptor</td>
</tr>
<tr>
<td>ScoT</td>
<td>6287</td>
<td>268</td>
<td></td>
<td></td>
<td>Thioesterase II (Kotowska et al. 2002)</td>
</tr>
<tr>
<td>CpkN</td>
<td>6288</td>
<td>267</td>
<td>49/66</td>
<td>Med-ORF11 Streptomyces sp. AM-7161 (BAC79037</td>
<td>SARP regulator</td>
</tr>
</tbody>
</table>

<sup>a</sup> Numbers according to Bentley et al. (2002)

<sup>b</sup> Size of the protein in aa

<sup>c</sup> Identity/similarity to the homologues using BlastP

<sup>d</sup> Proteins with highest similarity are listed apart from those with asterisk, where the most similar Streptomyces antibiotic biosynthesis genes are listed

<sup>e</sup> Proposed functions based on homology or confirmed experimentally. In the PKS subunits, the domains are listed in their order in the protein
domains with 100% identical nucleotide sequences (in cpkB from 1,811 to 2,561 bp, and 7,173 to 7,833 bp, in cpkC from 1,772 to 2,523 bp, all positions counted from the start of the genes). Other examples of such similarity include the genes involved in mycolatones production by Mycobacterium ulcerans (Ginolhac et al. 2005). However in M. ulcerans, the repeat sequence are insertion sequences which is not the case in the Cpk cluster.

**KS domains**

KS domains generally start with a motif resembling EPIAIV (Donadio and Katz 1992) which is also found in CpkABC (Fig. 2). The active sites are conserved with a QSSS in the loading domain for CpkA and CSSS in the other extension domains (Donadio and Katz 1992). Furthermore, all KS domains end with a consensus of NAHV\(^{\text{V1}}\)\(^{\text{I1}}\)\(^{\text{E}}\).

**ACP domains**

ACP domains from all modules contain standard phosphopantetheine attachment sites with a consensus of GFDS (Donadio and Katz 1992) (Fig. 2). ACPs from the loading module and modules 1 and 3 are connected with the KS domains of subsequent modules by intermodular linkers (Gokhale et al. 1999).

**AT domains**

All AT domains from the Cpk synthase analysed here have the conserved active site consensus of GH\text{XG-}\(\text{98aa-AFH}\) and are predicted to incorporate malonyl extender units (Haydock et al. 1995; Yadav et al. 2003) (Fig. 2). The C-terminal fragment (positions 280–300), which in the case of malonyl specificity consists of a \(\beta\)-turn, \(\beta\)-strand and \(\alpha\)-helix (Lau et al. 1999), is also present in all six AT domains (Fig. 2).

**Reducing domains**

All five modules contain sequences homologous to other known KR domains. The KR domains all have a site which binds NADPH with a consensus sequence of GxxGxxGxxA (Aparicio et al. 1996), followed by a conserved aspartate, then 42 aa spacing to a triad-building active site: S, Y, N (Reid et al. 2003) (Fig. 2). All Cpk KR domains except CpkBM4 have a consensus motif of LDD and with either P144 or N148 conserved, placing them in the B-type KR (Fig. 2) (Caffrey 2003, 2005; Siskos et al. 2005).

In the DH domain, motifs similar to the NADPH-binding site consensus HxxxGxxxxP containing the two active residues (H and P) (Donadio and Katz 1992) were identified in all but DH4 (Fig. 2). Though DH4 has a conserved proline in the NADPH-binding site, the other two aa are not conserved, and this domain is 40–51 aa shorter. The NysI M10 DH in the nystatin gene cluster also only has a conserved proline and has a shorter domain. This DH is not functional (Brautaset et al. 2000), which may suggest that DH4 is also not functional.

**Terminal reductase domain**

The C-terminal domain in CpkC contains seven typical consensus “core regions” characteristic of reductases found among some fungal and bacterial NRPSs (Fig. 2, Konz and Marahiel 1999). It also shows high homology in some PKSs from Cryptosporidium parvum (Zhu et al. 2002), Anabaena sp. (Black and Wolk 1994) and S. avermitilis (Ikeda et al. 2003). There is also homology to fungal aminoadipate reductase, which is involved in lysine biosynthesis (Casqueiro et al. 1999) and to a reductase (SCO1273) in S. coelicolor annotated as a part of a type II fatty acid synthase (Challis and Hopwood 2003). This was quite unexpected, as typically a thioesterase domain which catalyses hydrolysis of the carbon chain from the ACP and releases the product as a lactone or carboxyl acid is found after the last condensing module in type I PKSs from Streptomyces (Staunton and Weissman 2001). On the other hand, a reductase can also release the final product from the ACP by possibly producing an aldehyde or an alcohol (Silakowski et al. 2001). Aldehydes are unstable and can be toxic to cells, so the intermediate can be converted into an alcohol, acid or a more complex form (Zhu et al. 2002).

**Other genes of the cpk cluster**

The products of three of the other 17 genes in the Cpk cluster have been described previously: ScoT, AccA1 and CpkO. The protein encoded by scoT is a type II thioesterase with homology to an enzyme encoded by the modular PKS cluster for tylosin (Kotowska et al. 2002). AccA1 is an alpha subunit of a propionyl-CoA carboxylase (Rodriguez and Gramajo 1999). CpkO is a Streptomyces antibiotic regulatory protein (SARP) (Wietzorrek and Bibb 1997) regulated by ScoR (Takano et al. 2005). The cpkO gene was previously called kasO (Takano et al. 2005), but we have changed this to cpkO because the gene designation kas has also been used for the genes of kasugamycin biosynthesis in Streptomyces kasugaensis M388-M1 (Ikeno et al. 2006).
Fig. 2 Characteristic motifs in KS, AT, DH, KR, ACP and TD domain. For KS, ACP, AT, KR and DH domains, the sequences are marked by the module numbers as follows: Cpk_A_LM, loading module from Cpk synthase; Cpk_A_M1–Cpk_C_M5, modules 1–5 from Cpk synthase; Ery_M4, module 4 of DEBS synthase, from *Saccharopolyspora erythraea* (Donadio and Katz 1992); Nys_M10, module 10 of nystatin synthase from *Streptomyces noursei* ATCC 11455 (Brautaset et al. 2000). Consensus sequences proposed by different authors are given: cons KD from Donadio and Katz (1992); cons Y from Yadav et al. (2003); cons H from Haydock et al. (1995). The numbers between the aa sequences stand for the aa spacing between the conserved motifs. N- and C-terminal motifs are denoted by N-ter and C-ter, respectively. The active site is marked with an asterisk. The phosphopantetheine attachment site is marked with a hash. The histidine residue important for NAD(P)H binding is marked with a cross. The \( \beta \)-Turn and \( \beta \)-Strand structures in AT domains are marked with \( \beta \)-T and \( \beta \)-S. KR alignment of KR domains. Active sites S, Y and N are marked with asterisks. The aspartate residue is marked with an arrow and the aa important for polyketide stereochemistry are bold (Reid et al. 2003). Cons A denotes a consensus for the NAD(P)H-binding site (Aparicio et al. 1996). TD alignment of terminal reductase domains. **KSC TD** denotes reductase domain from CpkC; *SaFA* for saframycin Mx1 synthetase A from *Myxococcus xanthus* (T18552); *MsA* for non-ribosomal peptide synthetase from *Stigmatella aurantiaca* (AAK57184); *CpPKSI* for polyketide synthase from *Cryptosporidium parvum* (AAN60755). Cons M for consensus sequence of seven (R1–R7; indicated on the top of the alignments) typical domains for reductases given by Konz and Marahiel (1999).
The possible functions of the remaining 14 genes were assessed by amino acid homology search. Their features are summarised in Table 1.

CpkD has very high similarity to squalene monooxygenase-like SgcD2 from *Streptomyces globisporus* and to orf34 from *Streptomyces carzinostaticus* and also other secreted oxidoreductases as CpkD also has a signal peptide. SgcD2 belongs to an enediyne PKS cluster and is proposed to catalyse hydroxylation of amino-benzoic acid (Liu et al. 2002), and orf34 is also involved in synthesis of an enediyne antibiotic, neocarzinostatin (Liu et al. 2005).

CpkE has homology to epoxide hydrolases from genome-sequenced bacteria but did not show any homology to known antibiotic biosynthesis genes.

CpkF has high homology to transmembrane multidrug resistance efflux proteins which include SgcB from *S. globisporus* (Liu and Shen 2000) and PqrB from *S. coelicolor* (Cho et al. 2003). PqrB has been shown to be required for paraquat resistance by a probable proton-dependent efflux of paraquat. This suggests that CpkF may also play a role in resistance by exporting the Cpk final product.

CpkG is similar to class III aminotransferases from genome-sequenced bacteria, but did not show any homology to known antibiotic biosynthetic genes. Class III aminotransferase is unrelated to classes I and II in structure and is the only enzyme to directly metabolise the amino group from amino acids (Yonaha et al. 1992).

The closest homolog of CpkH is ScF, an ORF from the same gene cluster. Both proteins are described as FAD-binding proteins and are similar to AclO (Ac No. AB008466.1), an oxidoreductase involved in acalinomycin production from *Streptomyces galilaeus*. These proteins are longer by 56 and 61 aa respectively at the N-terminal end compared to AclO and both contain a signal peptide signature. The exact function of AclO has not been reported.

The closest homologue of CpkI is encoded by orfX (SCO6264) which is located next to *scbR* (Takano et al. 2001) and encodes a protein with similarity to an oxygenase reductase. Both have similarity to the product of *jadW3* from the jadomycin PKS gene cluster of *Streptomyces venezuelae* (Wang and Vining 2003), and *jadW3* is also situated convergent to a gene for a potential \(\gamma\)-butyrolactone receptor. The full length CpkI is also highly similar to the C-terminal aa sequence of UrdM involved in urdamycin A angucycline production by *Streptomyces fradiae* (Faust et al. 2000). By mutational studies, UrdM was shown to be involved in the oxygenation of the 12b position of urdamycin A (Faust et al. 2000).

CpkJ shows similarity to nucleoside-diphosphate-sugar epimerase, also known as UDP-glucose 4-epimerase. The highest similarities are to the products of genes with unknown function from genome sequences. Some of these are involved in antibiotic biosynthesis, among which the most similar is from the oviedomycin producer *S. antibioticus* ATCC 11891 (Lombo et al. 2004), followed by several others annotated as hydro-lases. None of these gene functions are known.

CpkK has extremely high homology to many putative decarboxylases and the \(\beta\)-subunit of the carboxyl transferase found mostly in angucycline and aromatic polyketide antibiotic biosynthesis gene clusters, including Orf1 from the fredericamycin producer *Streptomyces* *fradiae* ATCC 49344 (Wendt-Pienkowski et al. 2005). A \(\beta\)-subunit of a carboxyl transferase, AccB (SCO5535) from *S. coelicolor* (Rodriguez et al. 2001), which is essential for growth, is also highly similar to CpkK. Orf1 mutation in *S. griseus* retained fredericamycin production, and heterologous expression of the cluster excluding this gene produced the antibiotic. The authors concluded from these results that the gene was not part of the fredericamycin gene cluster (Wendt-Pienkowski et al. 2005). However, in the jadomycin producer in *S. venezuelae*, a mutant for *jadI*, encoding the \(\alpha\)-subunit of the carboxyl transferase which is required with the \(\beta\)-subunit for activity, produced 15% of the parental level of the antibiotic (Han et al. 2000). \(\alpha\)- and \(\beta\)-subunits of the carboxylase form a complex that carboxylates acetyl-CoA to form malonyl-CoA. A putative \(\alpha\)-subunit of the carboxylase complex homologue, AccA1, is also found in the *cpk* cluster (Rodriguez and Gramajo 1999) and seems likely that CpkK may form a complex with AccA1. Though these genes may not be essential for the production of the antibiotic, they may provide efficient supply of the polyketide building blocks.

CpkL is a very small protein of 87 aa. However, there seems to be significant similarity to the C-terminal part of GrhG from the griseorhodin A producer *Streptomyces sp.* JP95 (40-110/113aa of GrhG) (Li and Piel 2002). The function of GrhG is not known.

CpkP\(\alpha\) and \(\beta\) are almost identical to Orf13 and 14 from the neocarzinil producer, *S. carzinostaticus* (Otsuka et al. 2004). These two genes encode the 2-oxoacid ferredoxin oxidoreductase complex which is proposed to be involved in the control of the starter supply in neocarzinil production by having a reverse function of the branched-amino \(\alpha\)-keto acid dehydrogenase (BCDH) complex. In neocarzinil production, branched amino acids are proposed to be the starter units and are synthesised by using BCDH subunits encoded in the biosynthetic gene cluster. Orf13 and
14 are postulated to degrade these starter units. As there are no BCDH complex homologues in the Cpk cluster, there is a possibility that CpkR and \( \beta \) may degrade branched amino acids synthesised by primary metabolism.

**Regulation**

CpkO and CpkN show homology to the pathway-specific *Streptomyces* antibiotic regulatory proteins, SARPs (Wietzorrek and Bibb 1997), with CpkO having highest homology to Asm18 from the ansamitocin producer *Actinosynnema pretiosum* (Yu et al. 2002). Neither gene contains a TTA codon. Some SARPs have been shown to directly activate the transcription of the biosynthesis genes by binding to direct repeats in these promoter regions (Wietzorrek and Bibb 1997), but we were unable to recognise suitably spaced direct repeats in the likely promoter regions of the *cpk* cluster.

ScbR2 has high similarity to \( \gamma \)-butyrolactone binding proteins including AlpW from *S. ambofaciens* and BarB from *S. virginiae*. \( \gamma \)-butyrolactone binding proteins bind to DNA to repress expression of target genes. When \( \gamma \)-butyrolactones are produced, the small signalling molecules bind to these proteins and the protein no longer binds to the DNA and the expression of the target genes is induced (Takano 2006, a review). \( \gamma \)-butyrolactones are known to regulate antibiotic production in several streptomycetes, and recent analysis of antibiotic biosynthetic clusters has revealed many \( \gamma \)-butyrolactone binding proteins (Takano et al. 2005). A BarB mutant has been shown to produce virginiamycin earlier than the wild type (Matsumo et al. 2004). Further investigation will be needed to prove the activity of ScbR2 as a repressor and identify its target.

It is interesting that the Cpk cluster has at least three regulatory genes. Though the number of regulatory genes is not as high as in the tylosin biosynthetic gene cluster from *S. fradiae*, which contains six such genes (Bate et al. 1999), it is highly likely that the control of the Cpk cluster will be comparably complex. Takano et al. (2005) have previously reported the involvement of the SCBs, \( \gamma \)-butyrolactones and their receptor, ScbR in the regulation of the *cpk* cluster, and found indications that other players apart from the three regulatory genes may contribute to the control of the *cpk* genes. The need for such a complex regulation system may become clearer when the structure of the final compound has been determined. Further investigations are underway to understand this complex regulation.

**Overview**

From *in silico* analysis of the Cpk cluster, we conclude that the putative PKS consists of three subunits: CpkA, CpkB and CpkC, which together consist of a loading module, five extension modules and an additional reductase domain. No ER domains were found. All ATs, including the AT domain of the loading module, are specific for malonyl extenders, and all KRs have motifs characteristic for B-type domains. On the basis of these data, the intermediate produced by CpkA, B and C should be built from six dicarbon units derived from malonyl-CoA to give a 12-carbon chain, without any side chains (Fig 1b).

Terminal reductase (TD) is located in a position typically occupied by a thioesterase domain, suggesting that the carbon chain may be released in the form of an aldehyde or alcohol. Such forms, together with four double bonds, may make the compound chemically active.

The genes for the main PKS subunits are adjacent to genes whose products may well modify the core structure: a transport protein, CpkF which may be involved in resistance, a monoxygenase, a reductase, a thioesterase and three regulatory proteins. However, quite a number of genes are unique or of unknown function in antibiotic biosynthesis gene clusters, the latter including ScF and CpkH, CpkE, CpkG, and CpkL. Furthermore, several other gene products are homologous to primary metabolic enzymes (e.g. CpkR and \( \beta \), AccA1, CpkK, and CpkJ). To deduce the final structure of the Cpk pathway thus becomes very difficult, but the evidence points to the possibility that the structure itself, as well as the enzyme functions, may have unique features. It is also interesting to note that neither the product nor the gene cluster was revealed during the past decades of intensive study of *S. coelicolor*.

ScF, CpkD and CpkH have signal peptides, suggesting that these proteins may be secreted. This is surprising, if ScF and CpkD require an FAD cofactor. To understand why these enzymes possess this domain and how the enzymes function will need further investigation. It is interesting to note that some of the enzymes of the actinorhodin pathway appear to be extracytoplasmic (Hesketh et al. 2002).

The Cpk cluster does not include genes encoding recognisable enzymes either for the synthesis or modification of sugars, or for glycosyl transfer. Although we cannot exclude the possibility of incorporating sugars from other metabolite pathways into the Cpk product, the final product is most probably not glycosylated. There is also no obvious chlorination or cyclase function. However, as the unique intermediate released
from the PKS maybe unstable, cyclisation still may occur.

We have suggested a structure for the intermediate and also the functions of the modifying enzymes, however to understand the exact role of each gene in the Cpk cluster we will need to detect the final compound, determine its structure, and to make further genetic and biochemical studies.

Disruption of the \textit{cpkC} gene

In order to examine the role of Cpk synthase, \textit{cpkC} was disrupted by insertional inactivation. The choice of the insertion site was limited by the availability of the DNA sequence at the time and by high similarity between DNA sequences encoding the synthase modules. Linker regions in modular synthases are relatively less conserved than enzymatic domains, and reductase domains are not as conserved as extension domains. Taking this into consideration, the linker between the DH and KR domains in module 5 was selected for an insertion site. A 3.5-kb plasmid, pK1132, carrying the apramycin resistance gene was inserted into \textit{cpkC} on the \textit{S. coelicolor} M145 chromosome via a short segment internal to \textit{cpkC} giving strain P100. Thus CpkC (229 kDa) was expected to be shortened by 773 aa to give a protein of 148 kDa (Fig. 3), and there might possibly be polar effects of the disruption on the expression of downstream genes. This truncation of CpkC would remove the KR, ACP and reductase domains.

![Fig. 3](image-url)  
---

**Fig. 3** Growth curve, antibiotic production and the expression of CpkC in \textit{S. coelicolor} and P100. \textit{a} Growth of \textit{S. coelicolor} cultures and antibiotic production. \textit{Black diamonds, triangles and circles} denote wild type \textit{S. coelicolor}, and \textit{open diamonds, triangles and circles} denote P100. Bacterial growth shown in diamonds was monitored by measuring \textit{OD}_{450} and is shown in log scale. Undecylprodigiosin concentration (mM) is marked with \textit{triangles}, and actinorhodin concentration (mM) with \textit{circles}. The growth phases are indicated on the top with \textit{arrows}. Arrowheads indicate the crude extract time points used for western analysis in (b). Western analysis of CpkC in crude extracts of \textit{S. coelicolor} A3(2) and P100. Crude extracts after 14, 16, 17, 18, 19, 20, 21, 25 h of growth were examined as indicated in (a). The 16–22 h corresponds to transition phase as indicated. Signals from wild type CpkC and the truncated form of CpkC from the mutant P100 are showed in \textit{arrows}. Molecular weights are indicated in the middle.
Lack of the ACP domain would completely abolish the function of module 5, leading to no polyketide chain or a chain shorter by two carbons.

Growth of wild type and mutant strains was compared on SMMS, MS, 79 medium and liquid SMM (Fig. 3). Mutant strain P100 did not exhibit any morphological, growth or antibiotic production (actinorhodin and undecylprodigiosin) differences compared to the wild type. Antibacterial activity was also assessed but no difference was observed between wt and P100 (data not shown) as well as the UV–Vis absorbance spectra (data not shown). As there was no phenotype of the mutant in the conditions investigated, it became important to assess whether the Cpk cluster was expressed in \textit{S. coelicolor}.

CpkC expression in \textit{S. coelicolor}

To determine whether the Cpk cluster is expressed in \textit{S. coelicolor} M145, crude extracts were analysed by western blot analysis using rabbit anti-CpkC polyclonal antibodies raised from the N-terminal fragment of the CpkC (CpkC-apex).

Wild type \textit{S. coelicolor} and P100 were grown in SMM liquid medium and crude extracts were obtained at various time points from 14 to 40 h. Actinorhodin and undecylprodigiosin production was also assayed at each time point (Fig. 3).

In crude extracts from wild type \textit{S. coelicolor} a protein with a molecular mass of about 220 kDa reacting with anti-CpkC antibodies was identified (Fig. 3). In extracts from the P100 mutant a protein of about 150 kDa was detected at parallel time points. Calculated molecular masses of CpkC and its truncated form are 229.8 and 148 kDa, respectively and correspond well with the masses of the proteins detected in the extracts.

The expression of \textit{cpkC} in both wild type and mutant was detected after 19 h of cultivation, which corresponds to the mid-transition phase of growth. This also correlated with the start of actinorhodin and undecylprodigiosin production. This is consistent with the results obtained for the expression of \textit{scoT} (Kotowska et al. 2002) and of \textit{cpkO, cpkC, cpkD,} and \textit{cpkG} (Takano et al. 2005).

Expression of several others genes from the Cpk cluster was confirmed by 2D gels in the \textit{S. coelicolor} 2D Gel Protein Database (http://www.dbk.ch.umist.ac.uk/StreptoBASE/s_coeli/referencegel/index.php) (Hesketh et al. 2002) and the \textit{S. coelicolor} database located on SWICZ (Swiss-Czech proteomic server http://www.proteom.biomed.cas.cz/). These databases show that proteins coded by genes \textit{cpkPβ, cpkPα, accAI, cpkE, cpkG, cpkI, cpkJ, srbR2,} and \textit{scoT} are detectable in \textit{S. coelicolor} liquid cultures. Expression of ORFs SCO6276–SCO6288 was also detected in the course of microarray experiments reported by Huang et al. (2005). This clearly indicates that the end product of the Cpk cluster may be produced by \textit{S. coelicolor}, so our as yet unsuccessful attempts to purify the Cpk end product and determine its chemical structure will continue. The difficulty of finding a candidate compound has reinforced the hypothesis that the Cpk product may have unique features that make it difficult to isolate.

Acknowledgments This manuscript is dedicated to the late Katarzyna Kuczek. We thank David Hopwood, Wolfgang Wohlleben, Tillman Weber, Jolanta Zakrzewska-Czerwińska and Kira Weissman for critical reading of the manuscript. K. Pawlik, M. Kotowska and K. Kuczek were supported by Polish Ministry of Scientific Research and Information Technology, grant number 3 P04B 004 25 and by the Institute of Immunology and Experimental Therapy. E. Takano was supported from a grant from Human Frontier Science Program RG0330/1998. K. Pawlik was also supported by the John Innes Foundation. Antibodies were raised in the Institute of Immunology and Experimental Therapy in Wroclaw.

References


Rudd BA, Hopwood DA (1979) Genetics of actinorhodin biosynthesis by Streptomyces coelicolor A3(2). J Gen Microbiol 114:35–43