Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Nonribosomal Peptide Biosynthesis: Point Mutations and Module Skipping Lead to Chemical Diversity

Silke C. Wenzel, Peter Meiser, Tina M. Binz, Taifo Mahmud and Rolf Müller*

Experimental Section

Isolation and structure elucidation of myxochromides A from *M. xanthus* DK1050

Myxochromides A were isolated from *M. xanthus* DK1050, a stable yellow derivative of DK1622 (A. Martínez-Laborda et al., *Mol. Gen. Genet.* **1986**, *205*, (1), 107-114). For separation of myxochromides, 100 g cells from *M. xanthus* DK1050 were extracted with 1 l methanol, evaporated to dryness and partitioned between 100 ml methanol and 100ml heptane. The methanolic phase was separated by silica gel chromatography (solvent: ethyl acetate:MeOH:H_2O 5:4:1 + 0.1 % acetic acid) and sephadex LH-20 chromatography (solvent: MeOH). The fraction containing myxochromides was purified by RP-HPLC (Zorbax Eclipse XDB-C8, Agilent Technologies, solvent: MeOH:H_2O 70:30). Myxochromide A_3 was obtained in pure form (10 mg) and the structure was elucidated by one- and two-dimensional NMR analysis (HMBC correlations see Figure 1, NMR data see Table 1).

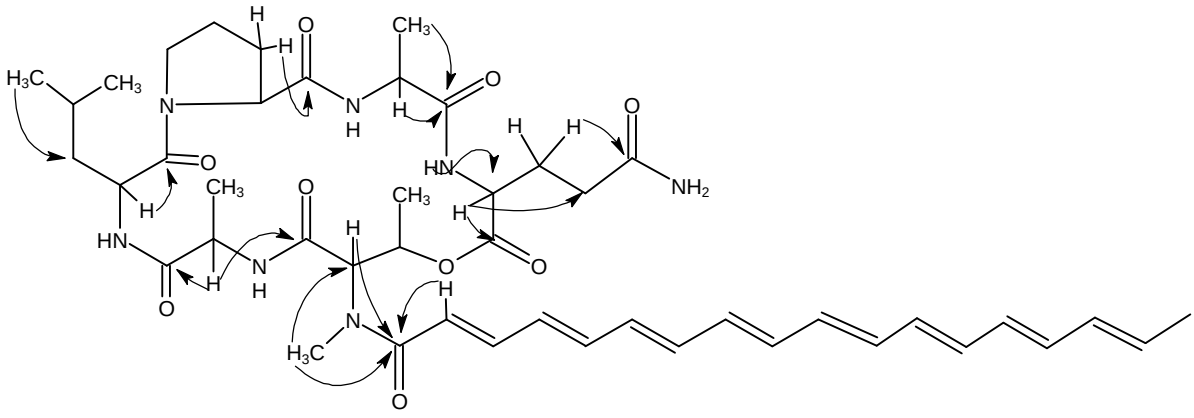


Figure 1: HMBC correlations in myxochromide A_3.

Table 1: ^{13}C NMR and ^{1}H NMR spectral data of myxochromide A_3 in CD_3OD at 500 MHz

moieties	functional group	d_C [ppm]	d_H [ppm] (mult[a]; $J_{H,H}$ [Hz])
N-Me-L-threonine	-*N*Me	36.0	3.20 (sbr)
	-C=O	173.2	-
	a-CH	60.0	5.55 (d; 3.5)
	ß-CH	72.5	5.72 (m)
	ß-CH_3	16.0	1.08 (dbr; 6.5)
D-alanine[1]	-NH	-	8.46 (d; 6.5)[c]
	-C=O	172.5	-
	a-CH	48.0	3.80 (m)
	a-CH_3	16.0	1.55 (d; 7.5)
L-leucine	-NH	-	8.11 (d; 8.0)[c]
	-C=O	165.0	-
	a-CH	52.1	4.43 (m)
	ß-CH_2	38.5	1.49 (m)
	?-CH	26.4	1.82 (m)
	?-CH_3^a	21.0	0.96 (d; 6.5)
	?-CH_3^b	23.6	0.96 (d; 6.5)
L-proline	-C=O	177.0	-
	a-CH	63.6	4.00 (dd; 7.5, 9.0)
	ß-CH^a	30.0	2.21 (m)
	ß-CH^b	30.0	2.24 (m)
	?-CH^a	26.7	2.14 (m)
	?-CH^b	26.7	2.15 (m)
	d-CH_2	48.1	3.55 (m)
L-alanine[2]	-NH	-	8.25 (d; 6.0)[c]
	-C=O	175.0	-
	a-CH	48.9	4.76 (dd; 7.3, 7.6)
	a-CH_3	16.4	1.25 (d; 7.0)
L-glutamine	-NH	-	7.62 (d; 8.5)[c]
	-C=O	171.7	-
	a-CH	54.0	4.53 (m)
	ß-CH^a	32.4	2.23 (m)
	ß-CH^b	32.4	2.26 (m)
	?-CH_2	29.3	2.37 (m)
	?-C=O	177.1	-
	d-NH^a	-	6.79 (s)[c]
	d-NH^b	-	7.18 (s)[c]
side chain	1`-CO	171.1	-
	2`-CH	120.3	6.59 (d; 14.9)
	3`-CH	145.4	7.34 (dd; 11.9, 14.9)
	4`-CH	131.6	6.47 (dd; 10.9, 14.1)
	5`-CH	142.2	6.73 (dd; 11.3, 14.9)
	6`-11`-CH	-[b]	6.30-6.54 (m)
	12`-CH	133.8	6.38 (m)
	13`-CH	133.7	6.30 (dd; 8.5, 13.7)
	14`-CH	133.5	6.15 (dd; 10.0, 14.0)
	15`-CH	135.5	6.25 (dd; 10.0, 14.0)
	16`-CH	133.4	6.13 (dd; 10.0, 14.5)
	17`-CH	131.6	5.75 (m)
	18`-CH_3	18.3	1.77 (d; 7.0)

[a] s, singlet; sbr, singlet broad; d, dublet; m, multiplet, [b] not assigned: d_C = 131.5, 133.2, 133.5, 135.4, 136.3, 137.3, [c] assigned in $[D_6]$DMSO

Determination of the absolute configurations of the amino acids from myxochromide A

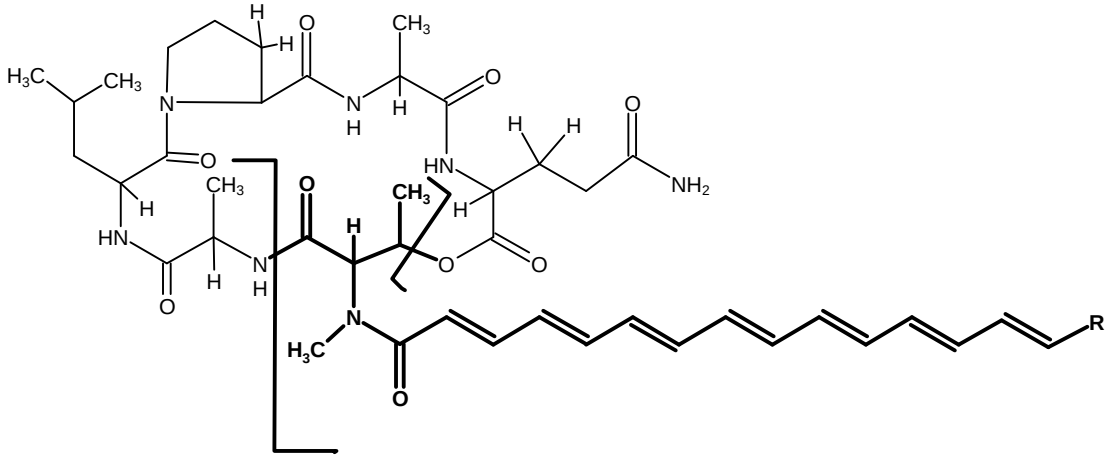
To determine the absolute configurations of the amino acids, 1.5 mg myxochromide A_3 was hydrolyzed with 1.5 ml 6 N HCl, neutralized with borate buffer pH 10, concentrated to 0.5 ml and derivatized with *o*-Phtaldialdehyd and *N*-acetyl-L-cystein (OPA-NAC) (100µl analyte + 150µl freshly prepared derivatisation reagent (each with 1,4 mg/ml o-Phtaldialdehyd and N-acetyl-L-cystein in borate buffer pH 10) (R.H. Buck, K. Krummen *Journal of Chromatography*, 387 (1987) 255-265). HPLC-analysis (Nucleosil RP18, solvent A: 50 mM Na-phoshate buffer pH 6.5, solvent B: 50 mM Na-phoshate buffer pH 6.5:MeOH:THF 350:650:5) could assign L- and D-configuration for alanine.

Determination of the structures of myxochromide A_2 and myxochromide A_4

Myxochromides A_{2-4} have the same peptide core, but differ in the structure of their polyunsaturated side chain. To determine the structures of the myxochromide A_2 and myxochromide A_4, a feeding experiment with *M. xanthus* DK1050 employing [$^{13}C_4$,$^{15}N_1$]threonine was performed. A 100 ml CTT-culture (Casitone 10g/l, 1M Tris pH 7.6 10ml/l, 1M K_2HPO_4 pH 7.6 1ml/l, 0.8M $MgSO_4$ 10ml/l) of DK1050 was supplemented with 30 mg of [$^{13}C_4$,$^{15}N_1$]threonine in three portions (added after 12, 24 and 36 hours of incubation), harvested after three days of growth and extracted with methanol.

The extract was analyzed by LC-MS:HPLC : Agilent 1100 Series, CC125/2 Nucleosil C18 Gravity 3µm-column (Macherey-Nagel), MS : Bruker HCT Plus. At constant flow rate (0.4 ml/min), the following gradient was used (A: H_2O + 0.1% acetic acid, B: acetonitrile + 0.1% acetic acid): 0-2 min 5% B, 2-22 min 5-95% B, 22-26 min 95% B, 26-28 min 95-5% B, 28-31 min 5% B.

A +5 shift, resulting from the incorporation of labeled [$^{13}C_4$,$^{15}N_1$]threonine could be observed in both the molecular ion M^+ and a specific threonine-polyketide chain fragment (shown in bold, see Figure 2), leading to the assignment of the following structures:



		mass of the specific threonine-polyketide fragment
myxochromide A_2	R = C_2H_5,	m/z = 336
A_3	R = CH=CH-CH_3,	m/z = 348
A_4	R = CH=CH-C_2H_5,	m/z = 362

Figure 2: Structures of myxochromides A_{2-4}.

Identification of the *mch*$_A$ biosynthetic gene cluster in *M. xanthus* DK1622

The myxochromide A biosynthetic gene cluster (*mch*$_A$ cluster) was identified in the genome sequence from *M. xanthus* DK1622 (available at the Institute of Genomic Research, Rockville, MD, USA) by comparative analysis with the *mch*$_S$ cluster (Genbank accession number AJ698723) from *S. aurantiaca*. Sequence analysis revealed the presence of three genes *mchABC*$_A$ encoding proteins with a modular organization as shown in Table 2.

Table 2: Deduced functions of genes involved in myxochromide A biosynthesis.

Gene	Size in bp; size of the deduced protein	Domains (location in the protein sequence)	
mchA$_A$	6459 bp 231.945 Da		KS (11-436) AT (546-863) DH (874-1075) ER (1447-1728) KR (1751-1998) ACP (2028-2097)
mchB$_A$	9147 bp 331.944 Da	module 1: module2: 	C (39-376) A´ [a] (529-922) MT (923-1401) A´´ (1402-1427) T (1430-1534) C (1544-1889) A (2028-2478) T (2480-2584) E (2586-2938)
mchC$_A$	13398 bp 486.826 Da	module3 : module4 : module5 : module6 : 	C (15-354) A (485-943) T (946-1051) C (1055-1388) A (1526-1983) T (1986-2091) C (2099-2435) A (2575-3030) T (3034-3139) C (3145-3480) A (3617-4074) T (4078-4183) TE (4191-4463)

[a] A *N*-MT is integrated into the first module of the NRPS A domain. A1´ represents the first part of this A domain containing the core regions A1-A8, while A1´´ represents the second part of this domain and contains the core regions A9-A10.

Sequence comparison of adenylation domains and complete modules from the myxochromide A and the myxochromide S pathway

To exclude a genetic exchange of complete modules or domains between both myxochromide pathways, the nucleotide and amino acid identities between pairs from the adenylation (A) domains from the same module (A2/A2 and A3/A3) was compared with the identity of sequence pairs from the A domains in charge of incorporating the same amino acids (A2/A3 and A3/A2). Similarily, sequence pairs of corresponding modules (M2/M2 and M3/M3) were compared with sequence pairs of modules in charge of incorporating the same amino acids (M2/M3 and M3/M2).

Table 3: Nucleotide and amino acid identities between domains, modules and complete proteins from the *mch*$_A$ and *mch*$_S$ biosynthetic pathways. The percentage of identical nucleotides or amino acids was determined from an alignment of the corresponding sequence pairs using the program ClustalW (available at http://www.ebi.ac.uk/clustalw/). Comparison of A2 and A3 as well as M2 and M3 is highlighted in grey.

mch$_A$/*mch*$_S$	DNA identity [%]	protein identity [%]
A1/A1[a]	81	85
A2/A2	73	68
A3/A3	71	65
A4/A4	70	62
A5/A5	80	82
A6/A6	80	81
A2/A3	59	37
A3/A2	58	41
M1/M1	78	78
M2[b]/M2[b]	76	73
M3/M3	72	66
M4/M4	71	64
M5/M5	75	72
M6/M6	77	75
M2[b]/M3	43	31
M3/M2[b]	44	30
MchA/MchA	77	76
MchB/MchB	77	76
MchC/MchC	74	70

[a] without the internal MT domain.
[b] without the terminal E domain

Detailed sequence analysis of the adenylation and peptidyl carrier protein domains from module 4 of both myxochromide pathways (encoded by *mch*$_S$ and *mch*$_A$)

	coreA1	coreA2	coreA3	coreA4
	S	V I	T	
consensus	LTYxEL	LKAGxAYLPLD	LAYxxYTSGSTGxPKG	FDxS
mch$_S$	KTYRQL	FKAGATYVPID	LACVLYASGAGGEPKG	TGAS
mch$_A$	RTYRQL	LKAGGAYVPLD	LAYVLYTSGSTGKPKG	FDVS

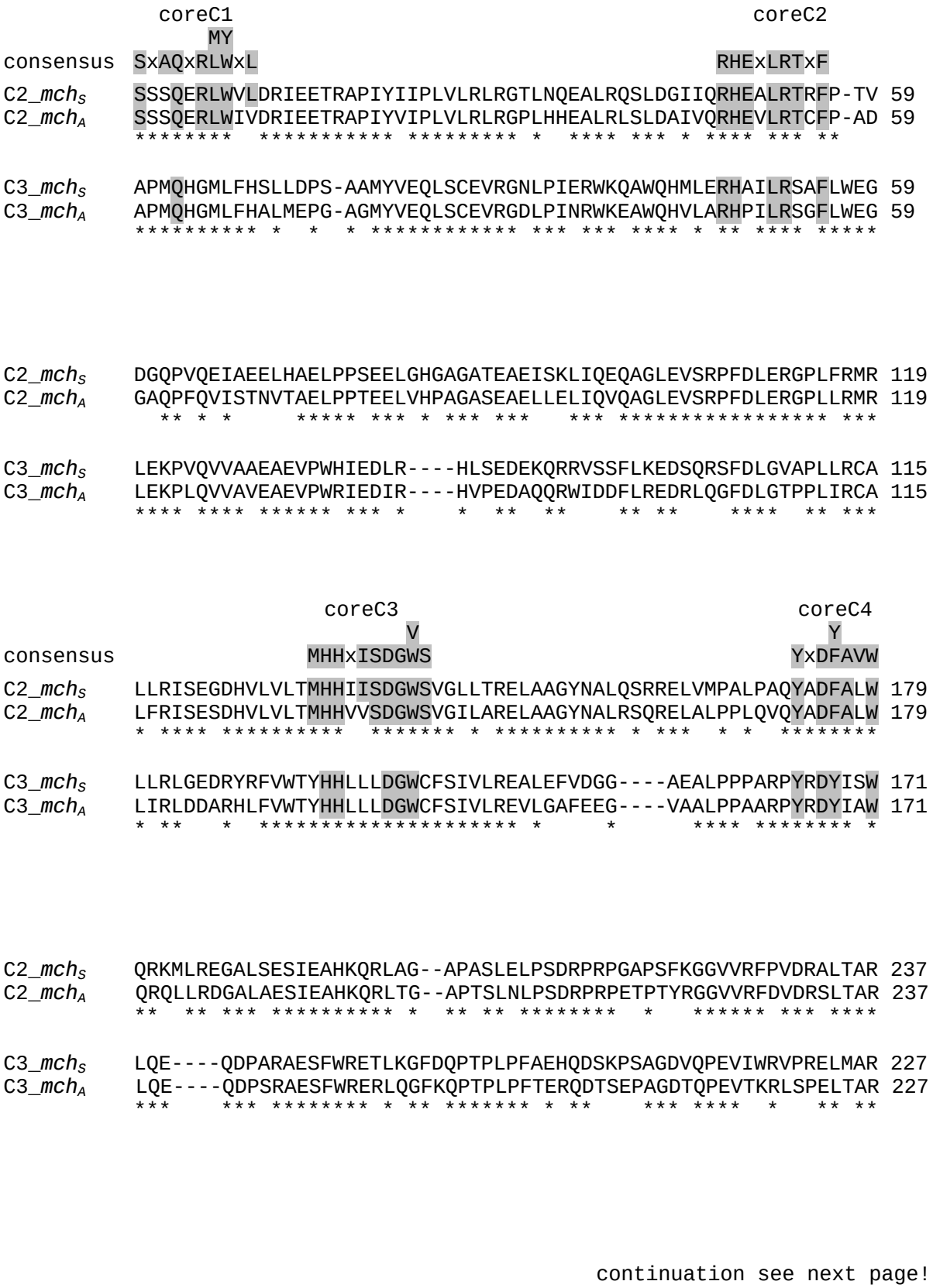
	coreA5	coreA6	coreA7
		L	K
consensus	NxYGPTE	GELxIxGxGVARGYL	YRTGDL
mch$_S$	SQYTLAE	GEIYISGACLADGYL	FQTGDL
mch$_A$	NHYGPTE	GELYLAGTCLADGYL	YRTGDL

	coreA8	coreA9	coreA10
		V	L
consensus	GRxDxQVKIRGxRIELGEIE	LPxYMIP	NGKVDR
mch$_S$	GRADGQILVRGIHVEPGEVE	IPEYMVP	GGKVER
mch$_A$	GRADDQVKIRGFRVEPGEIE	LPEYMVP	SGKVAR

	corePCP
	D I
consensus	DxFFxLGGHSL
mch$_S$	EDFFQMGGNPS
mch$_A$	QDFFQLGGHSL

Figure 3: Alignment of the adenylation (A) and peptidyl carrier protein (PCP) domain core motifs from module 4 of the *mch*$_S$ and *mch*$_A$ biosynthetic gene cluster using the program ClustalW (available at http://www.ebi.ac.uk/clustalw/). The consensus sequences of the core motifs (D. Konz, M. A. Marahiel, *Chem.Biol.* **1999**, 6, R39-R48) are indicated in the upper line. Amino acids differing from the consensus are highlighted in grey.

Detailed sequence analysis of the condensation domains from module 2 and 3 of both myxochromide pathways (encoded in *mch*$_S$ and *mch*$_A$)



continuation see next page!

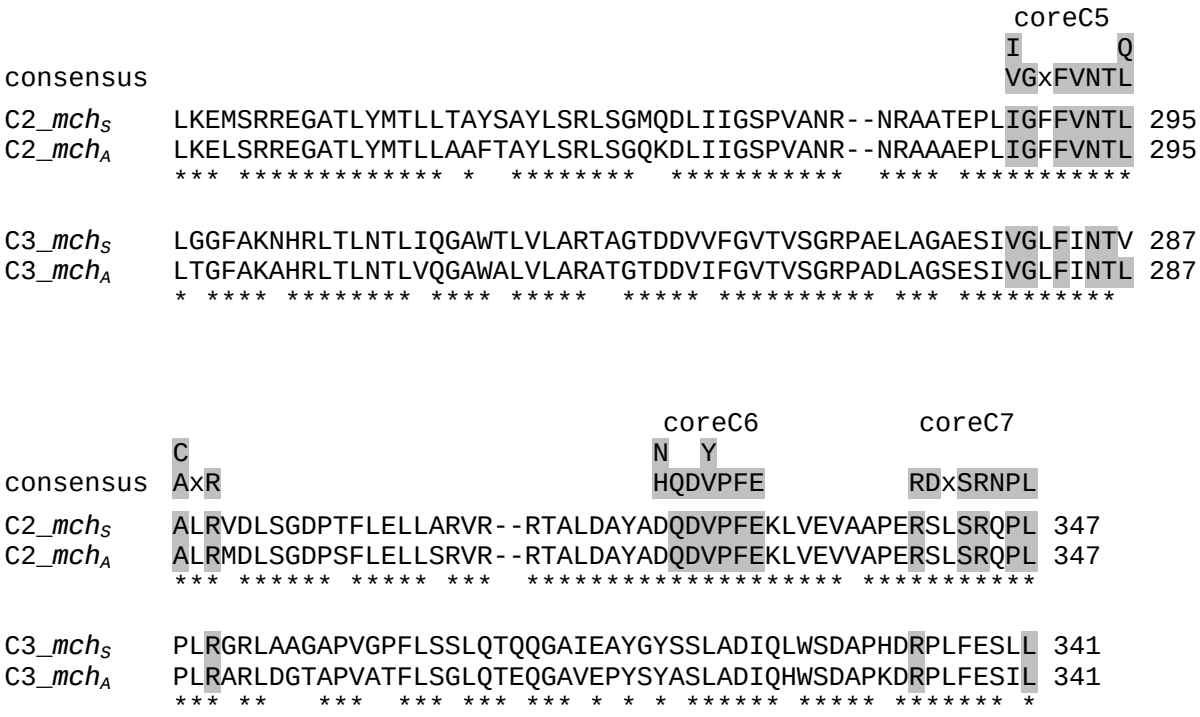


Figure 4: Alignment of the condensation domains from module 2 and 3 of the *mch*$_S$ and *mch*$_A$ biosynthetic gene cluster using the program ClustalW (available at http://www.ebi.ac.uk/clustalw/). The consensus sequences of the core motifs (D. Konz, M. A. Marahiel, *Chem.Biol.* **1999**, 6, R39-R48) are indicated in the upper line Highlighted in grey). Amino acids coinciding with the consensus are highlighted in grey, identical amino acid residues in the C2 and C3 pairs are shown by stars.

Construction of the *M. xanthus* DK1622 mutant strain *M. xanthus::pMch22a*

To provide evidence that the mch_A pathway directs the biosynthesis of myxochromides A, the insertional mutant *M. xanthus*::*pMch22a* was constructed. Therefore, an internal fragment of *mchC* was amplified as follows: Using oligonucleotides Mch16 (5`-GTCAGGCCCAGAGGACGCC-3`) and Mch19 (5´-CACCCGGCTCAAGGAGGCG-3`) a DNA fragment 1691-bp in size encoding part of *mch*C was amplified and subsequently inserted into pCR2.1TOPO (according to the manufacturer`s protocol, Invitrogen). The resulting plasmid, pMch22a, was transferred into *M. xanthus* wild type by electroporation (K. Kashefi et al., *Mol Microbiol* **1995**, *15*, 483) and integrated into the genome by homologous recombination. The obtained gene disruption mutant strain, *M. xanthus*::*pMch22a*, was verified by Southern analysis, analyzed for secondary metabolite production and found to be myxochromide A-negative.

Construction of expression plasmids

All recombinant gene fragments were amplified by PCR from chromosomal DNA from *S. aurantiaca* DW4/3-1 or *M. xanthus* DK1622 using the Triple Master Polymerase (Eppendorf) according to the manufacturer`s protocol. The following cycling profile was used for the PCR reaction: 60s at 95°C, 60s at 56°C, 90s (for adenylation (A) domains) or 30s (for peptidyl carrier protein (PCP) domains) at 72°C for a total of 32 cycles. Modified oligonucleotides (Qiagen, Alameda, USA) were used to introduce an *Eco*RI restriction site subsequently used for cloning into the expression vector pGEX-6P-1 (Amersham) and to introduce a stop codon (TGA) into the reverse primers. The A domain PCR products were first cloned into the pCR2.1TOPO vector (according to the manufacturer`s protocol, Invitrogen). The PCR products were verified by sequencing and then digested with *Eco*RI and cloned into pGEX-6P-1 linearized with *Eco*RI. The thiolation domain PCR products were directly digested with *Eco*RI, cloned into pGEX-6P-1 (linearized with *Eco*RI) and sequenced. The correct orientations of the inserts were identified by PCR using the pGEX6P-1_for primer (5`-gggctggcaagccacgtttggtg-3`) or pGEX6P-1_rev primer (5`-ccgggagctgcatgtgtcagagg-3`) in combination with the primers originally used for the amplification of the domains.

The following primers were used for the amplification and cloning of the A domains into pGEX-6P-1 (restriction sites for subsequent cloning are given in bold, modified sequences are shown underlined). Mch57 (5`-C**GAATTC**CTCTTGTCCTCCGAGGAG-3`) and Mch58 (5`-C**GAATTC**<u>TCA</u>CAGCACCTGCTCGAAGGC-3`) for A4 from *S. aurantiaca* to give plasmid pMch60, Mch63 (5`-C**GAATTC**CTCTTGTCCAACGAGGAGC-3`) and Mch64 (5`-C**GAATTC**<u>TCA</u>CACGGTCGCCTCCGTAGG-3`) for A4 from *M. xanthus* to give plasmid pMch65. For the amplification and cloning of the PCP domains into pGEX-6P-1 the following primers were used (restriction sites for subsequent cloning are given in bold, modified sequences are shown underlined). Mch65 (5`-C**GAATTC**ATGGAGGAACTCGTCGCCC-3`) and Mch66 (5`-C**GAATTC**<u>TCA</u>CTCGTTGATCACCCGGCTC-3`) for the PCP4 from *S. aurantiaca* to give plasmid pMch63, Mch67 (5`-C**GAATTC**ACGGAGGCGACCGTGGCAC-3`) and Mch68 (5`- C**GAATTC**<u>TCA</u>CGCATCGAGCACCCGGCTC-3`) for the PCP4 from *M. xanthus* to give plasmid pMch64.

Overexpression of PCP domains and mass analysis of the fusion proteins

E. coli BL21 cells (Novagene) carrying either of the PCP domain expression plasmids pMch63 or pMch64 alone or in combination with plasmid pSUMtaA (encodes the Ppant transferase MtaA, [26]) were grown in 250 ml LB medium (1% tryptone, 0.5% yeast extract and 0.5% NaCl) amended with 100 µg ml^{-1} ampicillin or 100 µg ml^{-1} ampicillin/25 µg ml^{-1} chloramphenicol, respectively. Cells were cultured at 28 °C to reach an OD_{600} 1, and subsequently the expression was induced with 0.1 mM isopropyl-ß-D-thiogalactoside (IPTG). The cells were harvested by centrifugation after an additional 2 h incubation and washed with 30 ml STE buffer (10 mM Tris, pH 8.0; 150 mM NaCl; 1 mM EDTA). Bacteria were resuspended in 22.5 ml of STE containing 100 µg ml^{-1} of lysozyme (added immediately prior to resuspension). After 15 min incubation on ice, dithiothreitol (DTT, 5 mM final concentration), protease inhibitor (PMSF, 1 mM final concentration) and 1.5% *N*-laurylsarcosine (from a 10% stock in STE, 1.5% final concentration) were added. Cells were lysed by sonication, centrifuged and the supernatant was adjusted to 2% Triton X-100 concentration (added from a 10% stock in STE). The purification of the PCP-GST fusion proteins was performed according to the manufacturer`s protocol (Amersham). Molecular masses of the PCP-GST fusion proteins were determined with MALDI/TOF (data not shown) and HPLC-ESI-ICR-MS (ESI-ICR-MS: electrospray - ion cyclotron resonance - mass spectrometry).

Overexpression of the adenylation domains and ATP-PP_i exchange assay

E. coli BL21 cells (Novagene) carrying either the A domain expression plasmid pMch60 or pMch65 were grown in 500 ml LB medium amended with 100 µg ml^{-1} ampicillin. Cells were cultured at 30 °C to reach an OD_{600} 1, and subsequently the expression was induced with 0.1 mM IPTG. The cells were harvested by centrifugation after an additional 2 h incubation and resuspended in 10 ml PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4). Cell lysis was achieved by french press (2 times at 1300 PSI). After centrifugation, an aliquot of the supernatant was used for the purification of the A-GST fusion proteins with the MicroSpin GST Purification Module (Amersham) according to the manufacturer`s protocol. PreScission protease (Amersham) was used for the on-column cleavage of the GST fusion. The eluates were concentrated employing Microcon YM-10 columns (Millipore) and the protein concentration was determined. The amino acid specificity of the purified A domains was determined with an ATP-PP_i exchange assay.[24] Tetrasodium [^{32}P]pyrophosphate (1

mCi/ml) was purchased from PerkinElmer. In a total volume of 100 µl each assay mix contained: A domain (1 µM final concentration), 5 µl 40 mM dATP, 10 µl 10x buffer (750 mM Tris (pH 8.0), 1 M NaCl, 100 mM $MgCl_2$), 5 µl 40 mM amino acid, 0.1 µCi [^{32}P]pyrophosphate. After 15 min incubation at 30 °C 500 µl stop mix (1.2 % (w/v) activated charcoal, 0.1 M tetrasodium pyrophosphate, 0.35 M perchloric acid) was added and the mixture was centrifuged. The charcoal pellet was washed two times with wash buffer (0.1 M tetrasodium pyrophosphate, 0.35 M perchloric acid) and resuspended in 500 µl water and 3.5 ml scintillation cocktail (ULTIMA GOLD, Perkin Elmer). Counts per minutes from each sample were recorded with a Liquid Scintillation Analyzer (TRI-CARB 2900TR, Packard).