J Ind Microbiol Biotechnol (2011) 38:1747–1760
DOI 10.1007/s10295-011-1022-8

REVIEW

Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery

Richard H. Baltz

Received: 20 May 2011 / Accepted: 20 July 2011 / Published online: 9 August 2011
© Society for Industrial Microbiology 2011

Abstract *Mycobacterium tuberculosis* encodes mycobactin, a peptide siderophore that is biosynthesized by a nonribosomal peptide synthetase (NRPS) mechanism. Within the mycobactin biosynthetic gene cluster is a gene that encodes a 71-amino-acid protein MbtH. Many other NRPS gene clusters harbor *mbtH* homologs, and recent genetic, biochemical, and structural studies have begun to shed light on the function(s) of these proteins. In some cases, MbtH-like proteins are required for biosynthesis of their cognate peptides, and non-cognate MbtH-like proteins have been shown to be partially complementary. Biochemical studies revealed that certain MbtH-like proteins participate in tight binding to NRPS proteins containing adenylation (A) domains where they stimulate adenylation reactions. Expression of MbtH-like proteins is important for a number of applications, including optimal production of native and genetically engineered secondary metabolites produced by mechanisms that employ NRPS enzymes. They also may serve as beacons to identify gifted actinomycetes and possibly other bacteria that encode multiple functional NRPS pathways for discovery of novel secondary metabolites by genome mining.

Keywords Actinomycetes · Aminocoumarin · Daptomycin · DptG · Genome mining · Glycopeptide · Lipopeptide · MbtH · NRPS · *Streptomyces*

R. H. Baltz (✉)
CognoGen Biotechnology Consulting, 6438 North Olney Street, Indianapolis, IN 46220, USA
e-mail: rbaltz923@aol.com

Introduction

The *mbtH* gene in *M. tuberculosis* encodes a small protein of unknown function; it is located in an NRPS gene cluster encoding the virulence-conferring siderophore mycobactin [46]. MbtH is the founding member of a protein superfamily, and subsequent studies have shown that *mbtH*-like genes are often found in NRPS gene clusters encoding antibiotics or siderophores. With the advent of rapid and inexpensive DNA sequencing, many actinomycete genomes can now be surveyed for known and cryptic secondary metabolite biosynthetic gene clusters, and novel pathways can be expressed using a variety of approaches [4, 5, 7, 10, 14, 60]. Many cryptic NRPS gene clusters have been identified, and they often include *mbtH*-like genes (see below).

Recently, a number of biological, biochemical, and structural studies have begun to shed light on the function(s) of MbtH-like proteins. In this review, I discuss these studies along with new information gleaned from surveying multiple microbial genomes for *mbtH*-like genes. The understanding of the function(s) of MbtH-like proteins will undoubtedly have useful biotechnological applications for secondary metabolite yield optimization, genome mining, heterologous expression, and combinatorial biosynthesis.

Prevalence of *mbtH* homologs in bacteria

Many *mbtH* homologs are deposited in GenBank. An *mbtH* homolog, *dptG*, is associated with the daptomycin gene cluster [34], and is required for high-level production of daptomycin [38]. As *dptG* (DptG) is associated with the production of an important antibiotic by a streptomycete. It was chosen as a surrogate probe for BLAST analyses

Table 1 Distribution of MbtH homologs among bacterial taxonomic groups

Taxonomic group	Sequenced genomes	DptG homologs	DptG homologs/ genome
Actinobacteria	168	188	1.1
Streptomyces sp.	25	65	2.6
Mycobacterium sp.	17	35	2.1
All others	126	88	0.7
Bacteroidites/Chlorobi	128	1	<0.01
Chlamydiae	9	0	<0.2
Cyanobacteria	49	7	0.14
Firmacutes	346	31	0.09
Bacillales	73	31	0.4
Clostridia	138	0	<0.01
Lactobacillales	72	0	<0.01
Molecutes	29	0	<0.03
Others	34	0	<0.03
Proteobacteria	627	188	0.3
Alpha	180	16	0.09
Beta	99	56	0.6
Gamma	247	101	0.4
Delta	45	6	0.1
Epsilon	34	0	<0.03
Spirochaetales	22	0	<0.05

Protein and nucleotide similarity searches for this and subsequent tables were carried out using BLASTP and BLASTN [3] (http://blast.ncbi.nlm.nih.gov/BLAST.cgi)

against microbial genomes deposited in GenBank to identify MbtH homologs. BLASTP with DptG was carried out in February, 2011, and homologs were encoded by members of Actinobacteria, Firmacutes, and Proteobacteria, but not by Archaea. In addition, there are no published reports of *mbtH*-like genes in eucaryotic organisms. Table 1 shows the distribution of MbtH homologs encoded by subgroups of the major bacterial taxa. MbtH homologs are most prevalent within the Actinobacteria, where 188 are encoded by 168 sequenced genomes; 65 of these are encoded by 25 *Streptomyces* sp. Only one MbtH homolog was encoded by a strain within the Bacteroidites/Chlorobi group, none by Chlamydiae, and seven by Cyanobacteria. Within the 346 sequenced Firmacute genomes, only 31 MbtH homologs were observed; all were encoded by strains within the 73 sequenced *Bacillales*, primarily by *Bacillus* sp. Within the Proteobacteria, the vast majority of the 188 MbtH homologs were encoded by the *Beta* and *Gamma* subgroups. The majority of the MbtH homologs in the *Beta* group were encoded by *Burkholderia* sp., and a number of those in the Gamma group were encoded by *Pseudomonas* sp. The general pattern from the analysis is that *mbtH* homologs tend to be more prevalent in free-living soil microbes with relatively large genomes. Not surprisingly, these tend to be the microbes that have the capability to produce complex secondary metabolites.

Prevalence and location of *mbtH* homologs in actinomycetes

The *mbtH* homologs are generally located in secondary metabolite gene clusters encoding nonribosomal peptides (NRPs) in actinomycetes. Table 2 shows some examples, including those encoding the clinically important antibiotics daptomycin, pristinamycin I, vancomycin, teicoplanin, and capreomycin. In most cases, a single *mbtH*-like gene is clustered with more than one NRPS gene. An exception is the teicoplanin gene cluster, which has two *mbtH* homologs. The biosynthetic products have multiple amino acids ranging from five to 13, and include D-amino acids, non-proteinogenic L-amino acids, and other modified amino acids. Many of the products are cyclized, and some are glycosylated, lipidated, or both. The MbtH-like proteins from these antibiotic pathways range in size from 62 to 80 amino acids.

In stark contrast to the peptide antibiotics discussed above, the aminocoumarin antibiotics coumermycin and clorobiocin and the peptidyl nucleoside antibiotic nikkomycin have complex structures derived from different types of building blocks. The common feature among these compounds is that each has precursors derived from single hydroxylated amino acids. In the case of aminocoumarins, the only NRPS-like genes in the clorobiocin and coumermycin pathways are adenylation-thiolation (A-T) didomains *cloH* and *couH* just downstream of the *mbtH* homologs *cloY* and *couY*, respectively [45]. These A-T didomains are likely to be involved in the binding, activation, and covalent tethering of L-Tyr to the phosphopantetheinylated T domain to facilitate β-hydroxylation (e.g., see [11] and below). Similarly, the MbtH-A-T tridomain involved in nikkomycin biosynthesis binds, activates and tethers L-His for β-hydroxylation [12]. These observations suggest possible roles for the MbtH homologs in one or more of the A-T functions, in protein–protein interactions needed for β-hydroxylation of protein tethered L-amino acids, or both.

With many new actinomycete secondary metabolite gene clusters and complete genome sequences becoming available, it is now possible to explore the prevalence and linkage relationships of *mbtH*-like genes. BLASTP analysis with DptG was carried out to survey the number of homologs encoded by sequenced actinomycete genomes. Of the *Streptomyces* genomes surveyed, all but one encode at least one MbtH homolog and the average is about three (Table 3). Although the genome sizes of the streptomycetes range from 6.6 to 11.0 Mb, there is no apparent

Table 2 Examples of proteins encoded by *mbtH*-like genes located in antibiotic biosynthetic (NRPS) gene clusters in actinomycetes

Microorganism	Product (aa)	Product class	MbtH-like protein	Size (aa)	Reference
Streptomyces roseosporus	Daptomycin (13)	Cyclic lipodepsipeptide	AAX31560 (DptG)	75	[34]
Saccharomonospora viridis	Daptomycin-like (13)	Cyclic lipodepsipeptide	YP_003133693 (DptG)	68	[6]
Streptomyces fradiae	A54145 (13)	Cyclic lipodepsipeptide	AAZ23079 (LptG)	80	[35]
Streptomyces coelicolor	CDA (11)	Cyclic lipodepsipeptide	AAD18046 (OrfX)	71	[22]
Actinoplanes friuliensis	Friulimicin (11)	Cyclic lipopeptide	CAM56772 (ORF21)	71	[37]
Streptomyces fungicidicus	Enduracidin (17)	Cyclic lipodepsipeptide	ABD65966 (ORF46)	71	[57]
Streptomyces pristinaespiralis	Pristinamycin I (6)	Cyclic depsipeptide	CBH31049 (MbtY)	72	[32]
Amycolatopsis orientalis	Vancomycin (7)	Glycopeptide	AAL90876	69	[59]
Amycolatopsis balhimycina	Balhimycin (7)	Glycopeptide	CAC48363 (Orf1)	69	[52]
Actinoplanes teichomyceticus	Teicoplanin (7)	Lipoglycopeptide	CAE53354 (Tcp13)	69	[51]
			CAE53358 (Tcp17)	69	[51]
Nonomuraea sp. ATCC 39727	Dalbavancin (7)	Lipoglycopeptide	CAD91210 (Dvb15)	69	[50]
Saccharothrix mutabilis	Capreomycin (5)	Cyclic peptide	ABR67757 (CmnN)	62	[19]
Streptomyces vinaceus	Viomycin (6)	Cyclic peptide	AAP92504 (VioN)	63	[53]
Streptomyces hygroscopicus	Mannopeptimycin (6)	Cyclic lipoglycopeptide	AAU34213 (MppT)	73	[31]
Streptomyces roseochromogenes	Clorobiocin (1)[a]	Aminocoumarin	AAN65223 (CloY)	71	[45]
Streptomyces rishiriensis	Coumermycin (1)[a]	Aminocoumarin	AAG29779 (CouY)	71	[55]
Streptomyces tendae	Nikkomycin (1)[a]	Peptidyl nucleoside	CAC11137 (NikP1)	677[b]	[12, 27]
Streptomyces ansochromogenes	Nikkomycin (1)[a]	Peptidyl nucleoside	AAO73548 (SanO)	677[b]	[54]

[a] These pathways have building blocks derived from single hydroxylated amino acids

[b] These proteins have MbtH-like domains fused to A-T didomains as MbtH-A-T

correlation between the number of *mbtH*-like genes and genome size. The largest numbers of MbtH homologs were observed in the clavulanic acid-producing *S. clavuligerus* (7), the streptomycin-producing *Streptomyces griseus* (7), the daptomycin-producing *S. roseosporus* (6), and *Streptomyces* sp. Act-1 (6); their genomes range in size from 7.8 to 8.6 Mb. The latter three strains are closely related to each other based upon BLASTP analysis with the glutamine synthetase (GlnA) protein (Table 3). Both *S. griseus* and *S.* sp. ACT-1 GlnA proteins showed 99.1% amino acid identities with *S. roseosporus* GlnA, and 100% identity with each other. *S.* sp. ACT-1 also encodes a protein that shows 99.7% amino acid identities to the *S. griseus* StrT protein involved in streptomycin biosynthesis, but no such homolog is encoded by *S. roseosporus*. The MbtH top hit homologs in *S. griseus* and *S.* sp. ACT-1 showed only 69.4 amino acid identities with DptG (Table 2), suggesting that they are not orthologs to DptG. However, they are identical to each other, and both are located in apparent siderophore biosynthetic gene clusters near NRPS genes that show 5745/5806 (98.9%) amino acid identities, but less than 50% amino acid identities to other NRPS proteins. Therefore, this specific pair of *mbtH*-like and NRPS genes may be predictive of a siderophore so far unique to *S. griseus*. These combined data also strongly suggest that *S.* sp. ACT-1 is a strain of *S. griseus* closely related to the type strain.

It is also noteworthy that *Streptomyces* sp. AA4, which has a relatively large genome and encodes four MbtH homologs, encodes a GlnA protein that has only 70.5% amino acid identities with *S. roseosporus* GlnA (Table 3) that differs in size (474 amino acids) from the 469 amino acid GlnA proteins encoded by well-characterized *Streptomyces* species. BLASTP analysis was carried out with *S.* sp. AA4 GlnA, and the top hit was the 474-amino-acid GlnA of *Amycolatopsis mediterranei* (448/474 = 94.5%) and the next best hit was to *Saccharomonospora viridis* (406/474 = 85.7%). It is likely that *S.* sp. AA4 is a strain of *Amycolatopsis* or of a closely related genus.

The amino acid identities of the other streptomycete top hits to DptG relative to GlnA identities (Table 3) suggest that they may be paralogs, and likely differ in specific function from DptG. The average amino acid identities of top hits to DptG for all streptomycetes was 64.3%, whereas the average for *S. roseosporus* GlnA was 92.5%.

BLASTP analysis of DptG and GlnA homologs of other representative actinomycetes indicated that the average divergence of MbtH homologs from DptG was similar to that observed within the *Streptomyces* species (63.0 vs. 64.3), whereas the GlnA orthologs from non-streptomycetes showed an average of 71.0% identities compared to 92.5% within the streptomycetes. The divergence of GlnA orthologs is consistent with known phylogenetic relationships of these diverse actinomycetes. The patterns of

Table 3 DptG (MtbH) homologs encoded by *Streptomyces* sp. and other actinomycetes

Strain	Genome size (MB)	DptG homologs	DptG top hit aa identity (%)	GlnA aa identity (%)
Streptomyces sp.				
S. violaceusniger	11	3	51/68 (75.0)	431/469 (91.9)
S. hygroscopicus	10.5	4	47/73 (64.4)	430/469 (91.7)
S. scabiei	10.1	3	43/68 (63.2)	430/469 (91.7)
S. sp AA4	9.2	4	42/70 (60.0)	334/474 (70.5)[a]
S. sviceus	9.1	0	–	426/469 (90.8)
S. avermitilis	9	3	35/69 (50.7)	430/469 (91.7)
S. coelicolor	8.7	2	46/67 (68.7)	419/469 (89.3)
S. sp. Act-1	8.6	6	50/72 (69.4)	465/469 (99.1)
S. viridochromogenes	8.5	2	37/61 (60.7)	433/469 (92.3)
S. griseus	8.5	7	50/72 (69.4)	465/469 (99.1)
S. ghanaensis	8.2	2	47/66 (71.2)	425/468 (90.6)
S. lividans	8.2	2	46/67 (68.7)	419/469 (89.3)
S. sp. C	7.9	1	39/67 (58.2)	422/469 (90.0)
S. roseosporus	7.8	6	75/75 (100)	469/469 (100)
S. pristinaespiralis	7.6	2	42/67 (69.4)	453/469 (96.6)
S. sp. ACTE	7.4	3	39/70 (55.7)	453/469 (96.6)
S. griseoflavus	7.4	3	52/66 (78.8)	427/469 (91.0)
S. sp. E14	7.1	4	41/70 (58.6)	426/469 (90.8)
S. sp. MG1	7.1	1	29/52 (55.8)	424/469 (90.4)
S. sp. SPB78	6.9	2	42/68 (61.8)	432/469 (92.1)
S. clavuligerus	6.8 (+1.8)[b]	7	45/69 (65.2)	446/469 (92.8)
S. albus	6.6	3	49/79 (62.0)	433/469 (92.3)
Streptomyces Ave.	8.1	3.1	64.3	92.5
Other actinomycetes				
Catenulispora acidiphila	10.5	3	34/60 (56.7)	366/474 (77.2)
Streptosporangium roseum	10.3	7	51/67 (76.1)	341/474 (71.9)
Amycolatopsis mediterranei	10.2	6	45/68 (66.2)	332/474 (70.9)
Frankia sp. EAN1pec	9	2	43/68 (63.2)	330/474 (69.6)
Actinosynnema mirum	8.2	7	41/73 (56.2)	342/474 (72.2)
Saccharopolyspora erythraea	8.1	3	45/68 (66.2)	354/474 (74.7)
Rhodococcus jostii	7.8	1	42/73 (57.5)	333/474 (70.3)
Micromonospora aurantiaca	7	1	30/52 (57.7)	348/474 (73.4)
Mycobacterium smegmatis	7	3	42/65 (64.6)	328/475 (69.1)
Rhodococcus erythropolis	6.5	1	43/73 (58.9)	333/476 (70.3)
Nakamurella multipartita	6.1	1	33/61 (54.1)	333/473 (70.4)
Nocardia farcinica	6	2	41/61 (67.2)	340/476 (71.4)
Salinospora arenicola	5.8	3	44/67 (65.7)	345/474 (72.8)
Nocardiopsis dassonvillii	5.7	2	47/70 (67.1)	322/474 (67.9)
Thermomonospora curvata	5.6	2	47/69 (68.1)	334/474 (70.5)
Salinospora tropica	5.2	3	33/69 (47.8)	347/474 (73.2)
Arthrobacter chlorophenolicus	4.4	1	31/63 (49.2)	317/474 (66.9)
Mycobacterium tuberculosis	4.4	2	41/68 (60.3)[c]	330/475 (69.5)
Saccharomonospora viridis	4.3	2	47/68 (69.1)	335/474 (70.7)
Janibacter sp. HTCC2649	4.2	0	–	341/474 (71.9)
Tsukamurella paurometabola	4.2	2	37/65 (56.9)	328/477 (68.8)
Thermobispora bispora	4.2	0	–	339/474 (71.5)
Intrasporangium calvum	4	0	–	337/474 (71.1)

Table 3 continued

Strain	Genome size (MB)	DptG homologs	DptG top hit aa identity (%)	GlnA aa identity (%)
Thermobifida fusca	3.6	1	42/74 (56.8)	323/474 (68.1)
Aeromicrobium marinum	3.1	0	–	337/473 (71.2)
Actinomycete Ave.	6.2	2.2	61.2	71

[a] This low sequence similarity indicates that *Streptomyces* sp. AA4 is not a *Streptomyces* sp., so it was not included in average amino acid identity calculation

[b] *S. clavuligerous* has a 1.8-Mb linear plasmid that carries secondary metabolite biosynthetic genes [33]

[c] Top hit is MbtH

divergence of DptG homologs within streptomycetes and other actinomycetes are consistent with the evolution of paralogous functions and with large-scale horizontal gene transfer (HGT), presumably along with cognate NRPS biosynthetic pathway genes.

There is one interesting exception in which a DptG homolog did not diverge more rapidly than the GlnA ortholog. In *Sac. viridis*, the DptG homolog showed 69.1% amino acid identity to DptG, and the GlnA protein showed 70.7% amino acid identity to *S. roseosporus* GlnA. This suggests that the DptG homologs in *S. roseosporus* and *S. viridis* have diverged at about the same rate as the orthologous GlnA sequences from a common ancestor, implying that they may also be orthologous. The *dptG* homolog in *Sac. viridis* is located in a daptomycin-like gene cluster just downstream of the NRPS genes just as authentic *dptG* is located in *S. roseosporus* [6].

Table 4 shows a more detailed analysis of MbtH homologs from several actinomycete genomes to assess their relationships to NRPS gene clusters. In addition to DptG, *S. roseosporus* encodes four other MbtH homologs and one MbtH-like fusion to the N-terminus of an NRPS A domain. BLASTP analysis revealed seven MbtH homologs in *S. griseus*: six are located individually in six of the seven annotated NRPS gene clusters, and the other in a PKS-NRPS hybrid pathway. *Amycolatopsis mediterranei* encodes six MbtH homologs that map to different NRPS clusters. *S. avermitilis* encodes three MbtH homologs, and each is located in an NRPS gene cluster. *Saccharopolyspora erythraea* encodes three MbtH homologs and the genes map to the three largest of seven NRPS gene clusters, ranging from about 25–58 kb [43]. *Sac. viridis*, a thermophilic actinomycete with a small genome (Table 3), encodes two MbtH homologs. One *mbtH* homolog is located just downstream of three large NRPS genes that encode a daptomycin-like cyclic lipotridecapeptide [6], and the other does not appear to be linked to an NRPS gene, but is linked to Fe^{3+} siderophore transport genes. Perhaps the latter is a remnant of a defunct pathway lost in genome downsizing.

Gene expression and biological effects of *mbtH*-like gene deletions

In *S. roseosporus* and *Streptomyces fradiae*, the producers or the cyclic lipotridecapeptide antibiotics daptomycin and A54145, respectively, single *mbtH*-like genes, *dptG* and *lptG*, are located just downstream of the NRPS genes [34, 35]. The *dptG* gene is transcribed from a giant mRNA from which the three upstream NRPS genes, *dptA*, *dptBC*, and *dptD* are translated, and RT-PCR indicated that *dptG* transcript levels were similar to those of the NRPS genes [13]. Transcription studies with low and high daptomycin-producing *S. roseosporus* strains indicated that the *dptG* transcript levels were elevated along with *dptA*, *dptBC*, and *dptD* transcripts in high-producing strains [48, 58]. The gene organization and expression suggests that DptG may be required in stoichiometric quantities relative to the NRPS proteins, but the levels of translation have not been measured.

The *dptG* gene is also clustered with downstream *dptH*, *dptI*, and *dptJ* genes which encode an editing thioesterase, a methyltransferase involved in the formation of 3-methyl-glutamic acid (3mGlu), and a tryptophan 2,3-dioxygenase involved in kynurenine (Kyn) biosynthesis, respectively. 3mGlu and Kyn are incorporated at positions 12 and 13 in the daptomycin cyclic peptide [34, 38]. Deletion studies indicated that the presence of all four of these genes (*dptGHIJ*) is required for maximum daptomycin production, and different combinations of deletions ascribed about 50% reduction in yield to the absence of *dptG* [38]. It is not known if the residual daptomycin production in the mutants lacking *dptG* is due to cross-pathway complementation of DptG function by one or more of the MbtH homologs in *S. roseosporus*. It is also not known DptG has a functional interaction with any of the products of the *dptH*, *dptI* or *dptJ* genes.

In *S. fradiae*, the *lptG* gene is located just downstream of the *lptA*, *lptB*, *lptC, and lptD* NRPS genes, and is also clustered with downstream genes encoding an editing thioesterase and enzymes involved in the modification of Glu,

Table 4 Linkage of *dptG* (*mbtH*) homologs to NRPS gene clusters in select actinomycetes

Microorganism	*dptG* homolog (gene or protein)	Protein size (amino acids)	Amino acid identities (%)	Location	Reference
Streptomyces roseosporus[a]	AAX31560 (DptG)	75	75/75 (100)	Dpt gene cluster	[34]
	ZP_04707179	75	49/70 (70)	NRPS-1	This report
	ZP_04713195	77	47/72 (65.3)	Non-NRPS	This report
	ZP_04712039	73	42/65 (64.6)	NRPS-2	This report
	ZP_04709435	74	28/66 (42.4)	NRPS-3 (pacidamycin)	[61, this report]
	ZP_04709625	585	24/51 (47.1)	NRPS-4 (fusion)	This report
Streptomyces griseus	SGR452	72	50/72 (69.4)	NRPS (SGR443-455)	[42, this report]
	SGR3256	72	45/67 (67.2)	NRPS (SGR3239-3288)	[42, this report]
	SGR912	75	34/67 (50.7)	NRPS (SGR895-901)	[42, this report]
	SGR654	68	32/62 (51.6)	NRPS (SGR653-656)	[42, this report]
	SGR6738	67	29/54 (53.7)	NRPS (SGR6730-6742)	[42, this report]
	SGR576	74	30/65 (46.2)	NRPS (SGR574-593)	[42, this report]
	SGR2590	72	27/63 (42.8)	NRPS (SGR2586-2590)	[42, this report]
Streptomyces avermitilis	SAV_851	72	35/69 (50.7)	NRPS7	[25, this report]
	SAV_3155	88	34/54 (63.0)	NRPS2	[25, this report]
	SAV_3644	55	28/49 (57.1)	NRPS3	[25, this report]
Streptomyces coelicolor	SCO3218	71	46/67 (68.7)	CDA (NRPS) cluster	[29, this report]
	SCO0489	70	40/66 (60.6)	Coelichelin (NRPS) cluster	[29, this report]
Amycolatopsis mediterranei	YP_003767198	69	45/68 (66.2)	NRPS-6	[63, this report]
	YP_003765296	66	38/64 (59.4)	NRPS-1	[63, this report]
	YP_003766215	67	34/63 (54.0)	NRPS-3	[63, this report]
	YP_003766936	75	36/71 (50.7)	NRPS-4	[63, this report]
	YP_003768373	72	32/67 (47.8)	NRPS-10	[63, this report]
	YP_003766961	67	32/63 (50.8)	NRPS-5	[63, this report]
Saccharopolyspora erythreae	ZP_06566005	72	45/68 (66.2)	NRPS7	[43, this report]
	ZP_06565122	78	40/67 (59.7)	NRPS3	[43, this report]
	ZP_06566809	67	38/63 (60.3)	NRPS5	[43, this report]
Saccharomonospora viridis	Svir_18430	72	47/68 (69.1)	Dpt-like gene cluster	[6, this report]
	Svir_02560	68	39/63 (61.9)	Non-NRPS	This report
Streptomyces fradiae	LptG	80	40/68 (58.8)	A54145 (NRPS) cluster	[35, this report]

[a] The NRPS clusters in *S. roseosporus* are arbitrarily named NRPS-1, NRPS-2, etc., for this review

Asn and Asp residues in the lipopeptide [1, 2, 35]. In ectopic transcomplementation studies, duplication of *lptG* and downstream genes under the expression of the strong constitutive *ermE** promoter did not have a beneficial effect on A54145 production [1]. None of the constructs resulted in deletion of *lptG*, so it is not known if it is required for maximum A54145 production yield.

Genes encoding MbtH-like proteins are located just downstream of the last NRPS module in actinomycetes that produce glycopeptide antibiotics related to vancomycin or teicoplanin [52]. The MbtH-like proteins in glycopeptide producers form a separate phylogenetic clade, suggesting that they perform similar functions in the biosynthesis of glycopeptides. However, disruption of the *mbtH*-like *orf1* gene in *Amycolatopsis balhimycina* had no apparent effect on balhimycin production, indicating that it was not required for balhimycin biosynthesis [52]. The authors noted that *A. balhimycina* has two other *mbtH* homologs, so they could not rule out the possible participation of heterologous MbtH-like proteins in balhimycin biosynthesis in the absence of Orf1 expression.

S. coelicolor has two *mbtH* homologs associated with NRPS gene clusters: *cdaX* in the CDA cluster; and *cchK* in the coelichelin cluster [29]. Disruption of *cchK* caused reduced coelichelin production, and disruption of *cdaX* under normal growth conditions abolished CDA production [29]. However, when the strain with the *cdaX* mutation was grown under low-Fe^{3+} conditions conducive to the expression of the coelichelin (siderophore) biosynthetic pathway, CDA was produced. On the other hand, a *cchK*,

cdaX double mutant produced no coelichelin or CDA under any growth condition. These data indicate the CchK and CdaX are required for coelichelin and CDA biosynthesis, respectively, and that each can cross complement the other pathway to some extent. The cross-pathway complementation of coelichelin production in the *cchK* mutant by *cdaX* expressed from the strong constitutive *ermE** promoter was substantially higher than that observed with the *cdaX* gene expressed from its native promoter, an observation that suggests possible utility in other heterologous cross-complementation studies.

The biosynthetic gene clusters for the aminocoumarin antibiotics clorobiocin and coumermycin have *mbtH* homologs (*cloY* and *couY*) just upstream of NRPS-like genes *cloH* and *couH*, respectively [45]. The *cloH* and *couH* genes also have a homolog in the novobiocin biosynthetic gene cluster (*novH*), which lacks a *mbtH* homolog. The *cloH*, *couH*, and *novH* genes encode A-T didomains, and NovH has been shown to activate L-Tyr and to form L-Tyr-*S*-NovH at the posttranslationally phosphopantetheinylated T domain [11]. The covalently bound L-Tyr serves as a substrate for oxidation by NovI to form L-β-OH-Tyr-*S*-NovH. This minimalist NRPS-like function in aminocoumarin biosynthesis suggests that the CloY and CouY proteins may have functional interactions with CloH and CouH A or T domains, and might facilitate the hydroxylation of enzyme-bound L-Tyr.

The question of whether CloY is required for chlorobiocin biosynthesis was addressed by Wolpert et al. [56]. It had been previously shown that the complete clorobiocin biosynthetic gene cluster can be expressed in *S. coelicolor* after integration into the ϕC31 *attB* site [18]. More recently, Wolpert et al. [56] deleted the *cloY* gene, and expressed the remaining genes in *S. coelicolor* M512. The recombinant produced 12 mg/l of clorobiocin, or about 22% of the control containing *cloY*. They constructed a *S. coelicolor* strain deleted for *cdaX* and *cchK,* and a transconjugant containing all of the clorobiocin genes except *cloY* produced 0.4 mg/l of clorobiocin, which was <1% of the clorobiocin produced by the strain with all three *mbtH*-like genes. The strain lacking all three *mbtH*-like genes was used to test for functional complementation by the individual *cloY*, *cdaX*, and *cchK* genes expressed from the *ermE** promoter; clorobiocin production was increased to 8.0, 5.1, and 4.1 mg/l, respectively. These combined results demonstrated that CloY is required for clorobiocin production, and that either CdaX or CchK can partially replace CloY function.

Structural features of MbtH-like proteins

Figure 1 shows the alignment of MbtH homologs from several secondary metabolite biosynthetic gene clusters from actinomycetes; unknown pathways from *S. roseosporus* (ZP numbers); MbtH from *M. tuberculosis*; an MbtH homolog from the peptide siderophore pyoverdine biosynthetic pathway in *Pseudomonas aeruginosa*; and the N-terminal end of the fused MbtH-P450 protein from *Lyngbya majuscula*. If we compare the stand-alone MbtH homologs, all have invariant regions, highly conserved regions, and variable regions. The invariant positions include the hallmark Trp residues at positions 25, 35, and 55 (relative to DptG). The MbtH-like protein from *P. aeruginosa* has been crystallized**,** and it is shaped like a thin arrowhead with dimensions of about 25 × 43 × 13 Å [16]. The protein has a three-stranded, anti-parallel β-sheet followed by two α-helices, one of which is packed against the β-sheet and the other forming the point of the arrowhead. Notably, the invariant amino acid residues are mainly exposed on the hydrophobic side of the arrowhead structure. The side chain indoles of W25 and W35 are nearly parallel and form a pocket about 7 Å across. The pocket also includes the invariant N17, S23, P26, and P32 (see Fig. 1). The face of the protein containing the W25–W35 pocket also contains invariant D57 and P60 residues. The D57 residue forms an ionic interaction with an invariant R59.

The solution structure of authentic MbtH has been determined by using nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopy [8]. The protein is monomeric, and contains a three-stranded, anti-parallel β-sheet arranged as $\beta3{:}\beta1{:}\beta2$ followed by a single α-helix, which is packed against the β-sheet. The authors presented evidence for intermediate motion on the millisecond to microsecond timescale between the $\beta1$ and $\beta2$ strands, and that the C-terminal region containing the highly conserved W55, D57, and R59 is disordered in solution. They hypothesized that the dynamics in the $\beta1$ and $\beta2$ regions may be necessary for interactions with other proteins, and that intrinsically disordered regions as in the C-terminus are generally associated with binding to multiple partners or functional diversity. They noted that functional cross-talk between different MbtH-like proteins as described above [29, 56] supports multiple binding partners and functional diversity.

The localization of invariant amino acid residues on one face of the arrowhead from the crystal structure of the MbtH-like protein from the pyoverdine pathway implies that the other face must be variable. This does not imply that it lacks function, but suggests that the mechanism of MbtH-like protein function might (at least in some cases) require interaction with two types of protein, one of which is more variable in structure than the other. This is consistent with amino acid divergence data that suggest that many MbtH homologs have paralogous functions (see below). Also, the primary function of MbtH-like proteins does not appear to be catalytic, as no obvious active site

```
                                     * *  * * **      * **             *       *** **
DptG(Sro            MANPFENNDG SYLVLVNDEG QYSLWPAFAD VPAGWTVTFG ESSRQECLDH INENWTDMRP KSLIRQMEND RTTAA
DptG(Svir)        M KSNPFEDENG VYHVLVNDEG QYSLWPSFIE VPAGWREVLS GKSRQECLDY IEENWTDMRP KSLIKAMEDA S
LptG(Sfr)        MT MTNPFDDTEG VFHVLVNDEN QHSLWPHFVE IPDGWRAVVR ERPRQECLDY IEANWTDMRP QSLIDAMEAH EKSEGAIR
AAD18046(Sco)       MTNPFEDADG RYLVLVNDEG QHSLWPSFVD VPAGWTVALG ESDREACLEY VEKNWTDMRP RSLVEAMSTG N
CAM56772(Afr)     M STNPFEDPEG TYLVLVNDEG QHSLWPAFVE VPAGWETVLP EGPRDAALEY IDTTWTDMRP KSLIAAMDA
ABD65966(Sfu)       MPNPFEDPDA SYLVLVNDEG QHSLWPVFAK VPDGWTSVFG EAGRQDCLDY IEKNWTDMRP KSLIEAMENQ R
CBH31049(Spr)       MSNPFEDAEG TFLVLVNHEG QYSLWPSFAE VPAGWTVALP ATDRESALAH ITDRWTDMRP QSLIDAMNGTAA
CAB53319(Sco)     M STNPFDDADG RFLVLVNDEG QHSLWPAFAA VPGGWTTVFE ENTRDACLAY VEANWTDLRP RSLARTADA
CAC48363(Aba)       MSNPFDNEDG SFFVLVNDEG QHSLWPTFAE VPAGWTRVHG EAGRQECLAY VEENWTDLRP KSLIREASA
AAL90876(Aor)       MTNPFDNEDG SFFVLVNDEG QHSLWPSFAE VPSGWTRVLG ETTRQECLAY VEENWTDLRP KSLIQEAGA
CAD91210(Nsp)       MTNPFENEDG SFLVLVNDEG QHSLWPSFAE VPPGWTRVHG VATRQECLAY VEENWTDIRP KSLIAEAGA
CAE53354(Ate)       MTNPFDNEDG SFLVLVNGEG QHSLWPAFAE VPDGWTGVHG PASRQDCLGY VEQNWTDLRP KSLISQISD
ABR67757(Smu)               MD TYLVVVNHEE QYSVWPADRP LPAGWRAEGT SGDKEQCLAH IETVWTDMRP LSVRRRAEAV
AAP92504(Svi)         MNDTPADT AYQVVLNDEE QYSVWPVGRP LPAGWRAEGT VGGRQACLDH IETVWTDLRP LSARA
AAU34213(Shy)     M GTNPFDDPDG RYLVLVNEED QHSLWPAFAE VPQGWTVALA ETDRQSALDF ITEHWTDMRP RSLVRAMEEA
AAN65223(Src)     M ATNPFEDENG SYLVLINGEG QHSLWPSFAD VPNGWTVIFN EASRQDCLDY VNEHWTDMRP LSLQRAMGGE
AAN29779(Sri)     M ATNPFDDENG VYLLLVNDEG QHSLWPSFAA LPKGWTVILD EASRQECLDY VNEHWTDMRP LSLQPAMGDE
CAC11137(Ste)       MVNPIHDDNG ACLVLVNSEN RHALWADGPD VPPGWRVAHR AAGRRECLEY IGAHWPDIRP ARLPDTGKDA GACLHDL---
ZP_04707179         MSNPFEDADG TYLVLVNDEG QHSLWPAFAD VPAGWTKVFG EAGRAECLEF VEQSWTDMRP KSLISAMQAE VDAKA
ZP_04713195   MTEDA VTNPFEDEDG AYLVLVNDEG QHSLWPAFAE VPAGWTVAHP EDTRRACLDY VEQNWTDLRP KSLMESMAAD PT
ZP_04712039       M TTNPFEDENG TYFVLRNDEG QHSLWPAFAG IPAGWTAVFG EAGRQACLDR IEADWTDMRP AGLVAAMGER AG
ZP_04709435            MSDDEAG EFRIVRNGEE QYSVWPAGQD LPAGWFEVGQ RGTRADCLAR VGRSWTDMRP ASLRRGLSGD EDGASRG
ZP_04709625          MPSPLTTEA TFLVLVNTAG ARCLWPADAS VPAGWSTELD RASRQDCLDF VERAQPHPQP VATPAPAPGP EPVPTLF---
MbtH(Mtu)         M STNPFDDDNG AFFVLVNDED QHSLWPVFAD IPAGWRVVHG EASRAACLDY VEKNWTDLRP KSLRDAMVED
MbtH-like(Pae)   GH MTSVFDRDDI QFQVVVNHEE QYSIWPEYKE IPQGWRAAGK SGLKKDCLAY IEEVWTDMRP LSLRQHMDKA AG
AAT12284(Lmu)       MTNPFADPSR DYWVLCNAEG QYSLWPTSLE IPEGWQTAFG PESWQNCLDN VEKNWTDMRP LSLLGQRKCP VPYPFRE---
                            10         20     W   30     W   40         50     W   60         70
```

Fig. 1 Alignment of amino acid sequences of MbtH-like proteins. The proteins are listed by specific gene product designations (e.g., DptG, LptG, MbtH), or by GenBank accession numbers. All proteins beginning with ZP_ are from *S. roseosporus*. Highly conserved amino acids are *shaded*, and amino acids marked with * are invariant in stand-alone MbtH-like proteins. The three proteins ending in–are fusion proteins. Abbreviations: Aba, *A. balhimycina*; Afr, *A. friuliensis*; Aor, *A. orientalis*; Ate, *A. teichomyceticus*; Lma, *L. Majuscula*; Mtu, *M. tuberculosis*; Nsp, *Nonomuraea* sp.; Pae, *P. aeruginosia*; Sco, *S. coelicolor*; Sfr, *S. fradiae*, Sfu, *S. fungicidicus*; Shy, *S. hygroscopicus*; Smu, *Sac. mutabilis*; Spr, *S. pristinaespiralis*; Src, *S. roseochromogenes*; Sri, *S. rishiriensis*; Sro, *S. roseosporus*; Ste, *S. tendae*; Svi, *S. vinaceus*; and Svir, *Sac. viridis*. See Tables 2 and 5 for additional information on many of the individual proteins

with catalytic groups was observed in the crystal structure [16].

The information gleened from the crystal structure of the *P. aeruginosa* MbtH-like protein and the solution structure of MbtH should be useful in predicting functional MbtH homologs from sequenced genomes. One prediction is that functional MbtH-like proteins should contain the invariant amino acids N17, E19, Q21, S23, W25, P26, P32, G34, W35, L48, W55, T56, D57, R59, and P60, or the abbreviated 15 amino acid code, NEQSWPPGWLWTDRP. Alternatively, the signature sequence can be represented as: NXEXQXSXWP-X5-PXGW-X13-L-X7-WTDXRP. These in turn may be useful in targeting potentially functional, but otherwise cryptic NRPS and other NRPS-like pathways for expression studies [5, 7, 60]. I have included the four MbtH homologs and one MbtH-A didomain from *S. roseosporus* in Fig. 1 as examples. DptG and the four stand-alone MbtH homologs have the invariant amino acid code, suggesting that they are functional proteins. The MbtH-A didomain (ZP_04709625) has the conserved W25 and W35, but is missing W55 and six other invariant amino acids. W55 may lie beyond the point of fusion. Inspection of the genomic region in *S. roseosporus* abutting the gene encoding ZP_04709625 indicates that it is located in an NRPS gene cluster, suggesting that ZP_04709625 may be a functional protein. Perhaps MbtH-like fusion proteins do not have the same stringent requirements as stand-alone MbtH-like proteins for all of the otherwise invariant amino acids because of nature of the fusion. For example, the MbtH-A-T fusion protein (CAC11137) from the nikkomycin pathway contains only 11 of 15 invariant amino acids (Fig. 1).

There is an MbtH fusion protein from a cyanobacterium, *L. majuscula*, that is also worth noting. LtxB (AAT12284; Fig. 1) is encoded in the lyngbyatoxin biosynthetic gene cluster, and is comprised of an MbtH-like protein fused to a P450 monooxygenase as MbtH-P450 [17, 23]. The MbtH domain contains all 15 invariant amino acids, suggesting that it carries out some function in lyngbyatoxin biosynthesis. The first steps of lyngbyatoxin biosynthesis are carried out on an NRPS [47], encoded by the di-modular *ltxA* gene located just upstream of *ltxB*, and it was postulated that LtxB carries out an oxygenation of an intermediate after it is released from the NRPS [17]. However, in light of new insights on MbtH-like protein function discussed below, it may be possible that the MbtH-like domain of LtxB assists in the oxygenation of an intermediate bound to the second T domain of the NRPS. So far, LtxB defines a unique example of a MbtH-P450 hybrid protein, as a BLASTP search with AAT11284 picked up only P450 enzymes lacking N-terminal MbtH domains.

Biochemical studies

Several recent biochemical studies with enzymes from actinomycetes and other bacteria have shed light on the

function(s) of MbtH-like proteins. Nikkomycin is a peptidyl nucleoside antibiotic produced by *Streptomyces tendae* Tü901 [27, 28]. The nikkomycin gene cluster has an unusual NRPS gene (*nikP1*) that encodes an A-T didomain fused to a MbtH-like domain as MbtH-A-T. NikP1 functions by binding and activating L-His, then thioesterifies it to the thiol of the phosphopantetheine bound to the T domain. L-His-*S*-NikP1 serves as a substrate for β-hydroxylation by NikQ to produce β-OH-His-*S*-NikP1 [12]. β-OH-His is then released from the NRPS by the NikP2 thioesterase. Although the role of the MbtH domain was not known at the time of publication, in light of new information from other systems reviewed here it is tempting to suggest that the fusion of an MbtH domain to the A-T didomain might play a role in stabilizing the A-T didomain, facilitating A or T functions, facilitating interaction between the A or T domain with the NikQ protein to assist in β-hydroxylation, or a combination of these functions.

The trihydroxamate siderophore vicibactin is produced by species of *Rhizobium* [9, 21]. Early gene disruption studies indicated that the *mbtH* homolog *vbsG*, which is located just upstream of the NRPS gene *vbsS*, is required for vicibactin biosynthesis in *R. leguminosarum* [9]. More recently, it was shown that the *vbsG* gene in *Rhizobium etli* CFN42 is located just upstream of a four domain (C*-A-T-Te) monomodular NRPS gene *vbsS* [21]. (C* indicates that the C domain is probably inactive, and Te is a thioesterase domain). When the *vbsG* and *vbsS* were co-expressed in *E. coli*, the proteins co-purified and the VbsG-VbsS complex was active in an adenylation assay. These combined studies have begun to shed light on the function of MbtH-like proteins.

Capreomycin and viomycin are antituberculosis antibiotics produced by *Sacchrothris mutabilis* subsp. *capreolis* [19] and *Streptomyces* sp. ATCC 11861 [53], respectively. Their related gene clusters have *mbtH*-like genes (*cmnN* and *vioN*) just upstream of NRPS A-T didomains involved in binding and activation of β-lysine (*cmnO* and *vioO*). Felnagle et al. [20] overexpressed and purified N-His$_6$-CmnO and N-His$_6$-VioO, and neither showed adenylation activity with β-lysine or with any other amino acid. However, the addition of purified CmnN or VioN to the individual reaction assays resulted in activation of β-lysine. Titration of the MbtH-like proteins in the adenylation assay indicated that they were needed in stoichiometric amounts to form a 1:1 complex with partner A-T didomains. In addition, the CmnN and VioN proteins were able to stimulate adenylation by the noncognate A-T didomains, and YdbZ from the enterobactin pathway was able to weakly stimulate β-lysine activation by the A-T didomains. When N-His$_6$-CmnO was co-expressed with CmnN, then purified by Ni-affinity chromatography, the two proteins co-purified. Site-directed mutagenesis was carried out to change the invariant Trp47 (equivalent to Trp55 of DptG; Fig. 1) to Ala47 (W47A), and the modified protein neither co-purified with N-His$_6$-CmnO nor stimulated β-lysine activation by N-His$_6$-CmnO. Possible interactions with three other NRPS proteins (CmnA, CmnF, and CmnG) involved in capreomycin biosynthesis were also investigated. Upon expression in *E. coli*, each of the A-T didomains from CmnA were active only when co-expressed with CmnN or when purified CmnN was added to the reaction, whereas CmnF and CmnG were active in the absence of CmnN. These data suggest that some NRPS A domains are dependent on MbtH-like proteins for activity and some are not.

The authors also studied the effects of co-expression of the MbtH homolog YbdZ with two NRPS proteins involved in enterobactin biosynthesis [20]. YbdZ did not co-purify with EntE, but did co-purify with EntF. They showed that the affinity of EntF for L-ser was improved 15-fold in the presence of YbdZ.

The biosynthetic gene cluster for pacidamycin in *Streptomyces coeruleorubidus* contains an *mbtH*-like gene (*pacJ*) and eight dissociated NRPS modules, including four A domains, four T domains, three C domains, and one Te domain [61]. None of the NRPS proteins have more than three domains. Three of the purified proteins containing A domains were active in adenylation assays, but the C*-A-T PacL protein had no adenylation activity unless it was combined with purified PacJ. To further explore the PacJ-PacL interaction, Zhang et al. [62] carried out site-directed mutagenesis of one or two of the invariant Trp residues and generated strains with W22A or W22A/W32A substitutions corresponding to positions W25 and W35 of DptG (Fig. 1). The enzyme with the W22A substitution stimulated adenylation activity by $\sim$50% relative to PacJ, and the enzyme with the W22A/W32A substitutions had no stimulatory activity. The enzyme lacking the two Trp residues also lost the capability to bind tightly to PacL. The authors noted that this is consistent with the hypothesis that the hydrophobic face of MbtH-like proteins containing the conserved Trp residues may be important in moderating protein–protein interactions. The requirement for PacJ to activate PacL adenylation activity provided a robust test for cross-pathway interactions of PacL with heterologous MbtH-like proteins. The PacL adenylation activity was strongly stimulated by KtzJ from kutzneride and GlbE from the glidobactin pathways, and weakly stimulated by VbsG from vicibactin and VdbZ of enterobactin pathways. These in vitro observations are consistent with the in vivo results discussed above that support the notion that some heterologous MbtH-like proteins can partially complement missing pathway-specific MbtH-like proteins. Perhaps more importantly, they demonstrate that not all MbtH have

equivalent activities, indicating that some may have distantly related, paralogous functions.

Glidobactins are hybrid NRPS-PKS secondary metabolites produced by *Burkholderia* sp. K481-B101 [24]. Biosynthesis of glidobactins is initiated on GlbF, a mono-modular condensation-adenylation-thiolation (C-A-T) NRPS that couples an unsaturated fatty acid to Thr_1. Attempts to express C-His_6-tagged GlbF in *E. coli* were not successful, but co-expression of GlbE, an MbtH homolog encoded by the *glbE* gene just upstream of *glbF*, with C-His_6-GlbF or N-His_6-GlbE with GlbF were successful. Nickle affinity purification of either protein resulted in co-purification of the other untagged protein. The proteins could be separated only by SDS–PAGE gel electrophoresis, so the activity of GlbF could not be tested alone. This study demonstrated that GlbE has a stabilizing or chaperone function for GlbF when the *glbE* and *glbF* genes are expressed in *E. coli* and that GlbE binds tightly to GblF during protein purification. It is noteworthy that the *glbE* and *glbF* genes are adjacent to each other in the glidobactin producing strain.

Orthologs and paralogs

If some MbtH-like proteins differ in function from each other, as suggested from genetic and biochemical studies described above, then their mutational history should help rank them as paralogs rather than orthologs. Orthologous genes experience purifying selection against certain mutations that result in amino acid substitutions, whereas paralogous genes, those that have undergone duplication and divergence of function, experience selection in one of the genes for amino acid substitutions that optimize a new function [26]. Thus with orthologs, the ratio of mutations causing nonsynonymous amino acid substitutions (K_a) to mutations giving synonymous amino acids (K_s) is smaller than the K_a/K_s ratio for paralogs [26]. The K_a/K_s ratio has also been presented as *dN/dS* more recently [41]. In a pairwise comparison of genes not acquired by horizontal gene transfer in two *Salinospora* species, 98.4 of orthologous genes in these marine actinomycetes had *dN/dS* ratios ≤ 1 [44]. Therefore, genes showing *dN/dS* ratios ≥ 1 are likely to be paralogs. Within the actinomyces, it is not known if mutations leading to synonymous amino acid substitutions are always neutral. It is likely that there is some level of selection to maintain ~70% overall G + C content and ~90% G + C at third positions of codons, where a high fraction of otherwise neutral mutations can take place. Also, different orthologous actinomycete genes will undoubtedly show different *dN/dS* ratios, but individual orthologous genes should show relatively constant *dN/dS* ratios in pairwise comparisons of many species of a given genus.

I used glutamine synthetase (*glnA*) genes from *Streptomyces* sp. as an example to estimate the *dN/dS* ratio of a typical ortholog, and explored the variation in *dN/dS* for the *S. roseosporus glnA* gene compared with ten other streptomycete *glnA* genes. The *glnA* gene and GlnA protein sequences diverged at about the same rates, showing an average of 92–93% sequence identities over the ten *Streptomyces* sp. (Table 5). Thus, for every three nucleotide substitutions, about one nonsynonymous amino acid substitution was made, giving a *dN/dS* ratio of 0.5 for each of the three specific comparisons shown in Table 5, and an average of 0.4 for ten comparisons. As another control, I compared several glycosyltransferase genes from the glycopeptide antibiotic vancomycin and the closely related chloroeremomycin pathways. The vancomycin and chloroeremomycin pathways in different strains of *Amycolatopsis orientalis* have one pair of orthologous genes (*gtfE* and *gtfB*) that encode glycosyltransferases that attach glucose to the same position on the vancomycin aglycon [49]. The Walsh group sequenced a *gtfE* gene from another strain of *A. orientalis* that produces vancomycin [30]. Bioinformatic studies indicated that three other apparently paralogous glycosyltransferase genes, *gtfA*, *gtfC*, and *gtfD*, share a common ancestor with *gtfE* and *gtfB*, but encode proteins that carryout different glycosylations during vancomycin or chloroeremomycin biosynthesis [49]. BLAST analysis with the original *gtfE* and GtfE sequences [49] against the orthologous genes and proteins indicated that they diverged at *dN/dS* ratios of 0.4 and 0.6, respectively (Table 5). BLAST analysis of *gtfE* and GtfE against the paralogous *gtfC* and *gtfD* genes and GtfC and GtfD proteins (which have different substrates and NDP-sugar cofactors than those for GtfE), gave *dN/dS* ratios of 1.1.

Having defined a rough range of *dN/dS* ratios of 0.4–0.6 for orthologs, and ~1.1 for paralogs in the glycopeptide glycosyltransferase system, I examined *dN/dS* ratios for MbtH homologs from glycopeptide biosynthetic gene clusters. BLASTP analysis was carried out of the MbtH homolog CAC48363 from the balhimycin gene cluster in *A. balhimycina* against the MbtH homologs from the other glycopeptide pathways (vancomycin, dalbavancin, and teicoplanin; Table 5). The amino acid identities of 77–88% were very similar to the nucleotide identities (78–89%), and the *dN/dS* ratios were 0.5 for each pairwise comparison, suggesting that these MbtH homologs have orthologous functions in glycopeptide biosynthesis.

BLAST analysis was carried out with *dptG* (DptG) against *mbtH* (MbtH) homologs from several other known and unknown NRPS pathways, and *dN/dS* ratios were calculated (Table 5). The lowest *dN/dS* ratio (0.6) was observed with the *dptG*-like gene from a cryptic daptomycin-like pathway from *Sac. viridis*, suggesting that these are true orthologs. The *dptG* homologs from other acidic

Table 5 Amino acid and nucleotide divergence of orthologs and paralogs as determined by BLAST queries

BLAST query	BLAST subject	Subject pathway	Amino acid identities (%)	Nucleotide identities (%)	*dN/dS*[a]
GlnA (Sro)	GlnA (Sco)	GS[b]	89.3	89.1	0.5
GlnA (Sro)	GlnA (Sav)	GS	91.7	91.4	0.5
GlnA (Sro)	GlnA (Sal)	GS	92.3	91.3	0.5
GlnA (Sro)	GlnA 10 strept[c]	GS	93.3	92	0.4
GtfE[d]	GtfE[e]	Vancomycin	90.5	88.8	0.4
GtfE	GtfB	Chloroeremomycin	81.1	84.1	0.6
GtfE	GtfC	Chloroeremomycin	63.4	77.1	1.1
GtfE	GtfD	Vancomycin	65.2	77.8	1.1
CAC48363	AAL90876	Vancomycin	88.4	88.6	0.5
CAC48363	CAD91210	Dalbavancin	85.5	86.2	0.5
CAC48363	CAE53354	Teicoplanin	78.3	78.7	0.5
CAC48363	CAE53358	Teicoplanin	76.8	78.3	0.5
DptG	Svir_02560	Dpt-like	69.1	72.3	0.6
DptG	LptG	A54145	58.9	74.1	1.1
DptG	CAM56772	Friulimicin	62.1	72.6	0.9
DptG	CdaX	CDA	68.7	74.5	0.7
DptG	CchK	Coelichelin	60.6	73.1	1
DptG	Sros_6555	Unknown	76.1	80.8	0.7
DptG	AAL90876	Vancomycin	68.2	75.6	0.8
DptG	CAC48363	Balhimycin	71.2	74.9	0.6
DptG	ZP_04697997	Unknown	65.3	75.8	0.9
DptG	ZP_04696847	Unknown	64.4	76.4	1
DptG	Sgr_452	Unknown	69.4	77.4	0.9
DptG	Sgr_3256	Unknown	67.2	72.1	0.7
DptG	Sgr_912	Unknown	50.7	69.9	1.1

[a] The *dN/dS* ratios were estimated graphically using the approximation: $dN/dS = 0.5\,(\Delta aa/\Delta nt)^2$, where $\Delta aa/\Delta nt$ is the ratio of percent change in amino acid sequence to the percent change in nucleotide sequence, and *dN/dS* is the ratio of the rate of mutations to nonsynonymous amino acids to synonymous amino acids (RH Baltz, unpublished)

[b] GS, glutamine synthetase

[c] The ten streptomycetes are: *S. albus* (Sal), *S. avermitilis* (Sav), *S. coelicolor* (Sco), *S. clavuligerus, S. flavogriseus, S. ghanaensis, S. griseoflavus, S. griseus, S. viridochromogenes, S.* sp. SPB78

[d] GtfE from *A. orientalis* C329.4 [49]

[e] GtfE from *A. orientalis* ATCC19795 [30]

cyclic lipopeptide antibiotics distantly related to daptomycin (CDA, friulimicin, and A54145) gave *dN/dS* ratios of 0.7, 0.9, and 1.1, respectively. The *dN/dS* ratio of 1.1 suggests that DptG and LptG may have different (paralogous) functions in lipopeptide biosynthesis, whereas the ratios of 0.7 and 0.9 suggest that the functions may differ in more subtle ways.

Finally, BLAST analysis with *dptG* (DptG) against homologs from the coelichelin, vancomycin and balhimycin pathways that encode peptides unrelated to daptomycin gave *dN/dS* rations of 1.0, 0.8 and 0.6; and similar analysis with homologs of unknown function from *S. roseosporus* and *S. griseus* gave *dN/dS* ratios averaging 0.9, suggesting divergence of specific functions.

Discussion

Although it has been observed that *mbtH*-like genes are generally associated with secondary metabolite biosynthetic gene clusters that include NRPS genes, the function(s) of MbtH-like proteins have remained obscure until recently. There are now several studies indicating that they play important roles in secondary metabolite biosynthesis, but not directly as catalysts. Their roles seem to be more as facilitators or chaperones. Their unique structures, which can be best described as flat, two-sided arrowheads, imply possible interactions with two proteins, one interaction on each face of the arrowhead. There is conservation of amino acids on one side of the arrowhead that may interact with A

or T domains, as demonstrated in recent biochemical studies, and likely higher variability on the other face that may interact with various other proteins in different systems. The variable function of the second side of the protein is supported by the genetic analysis that indicates that many MbtH homologs are paralogs rather than orthologs. Several recent studies suggest a possible linkage of MbtH-like protein function with steps in amino acid modifications, so this might be one chaperone role, to facilitate the modification of NRPS enzyme-bound amino acids, including but not limited to those bound to simple A-T didomains.

There are a number of mechanistic questions that remain: (1) Why do most NRPS pathways have only one MbtH homolog? (2) Why do some MbtH homologs complement others at varying efficiencies? (3) Do MbtH homologs interact only with A-T didomains, or do they also interact directly with other proteins? (4) What is the range of functions of diverse paralogous MbtH-like proteins?

For practical applications, MbtH homologs might be exploited in the following ways:

Strain improvement. The overexpression of MbtH homologs might be a way to enhance productivity during peptide antibiotic biosynthesis. In the cases of daptomycin and A54145, the *dptG* and *lptG* genes are located just downstream of the NRPS genes and appear to be transcribed along with the NRPS genes from giant multicistronic mRNAs. This transcriptional arrangement may facilitate coordinated expression of DptG and LptG (perhaps in stoichiometric amounts) with the NRPS proteins. Thus, mutations that enhance the expression of the NRPS genes would also enhance the expression of DptG or LptG, and this has been shown in the case of DptG [48, 58]. Overexpression of MbtH-like proteins may be a way to enhance NRPS-based secondary metabolite production during strain development programs.

Combinatorial biosynthesis. In designing strategies to carryout combinatorial biosynthesis, it may be important to include appropriate *mbtH*-like genes to facilitate expression of hybrid pathways. Because of the limited information on specific functions of most MbtH homologs, the design at present is empirical. In combinatorial biosynthetic work carried out at Cubist Pharmaceuticals, two platforms were established for genetic engineering in the daptomycin-producer, *S. roseosporus* [13, 36, 38, 39], and the A54145-producer, *S. fradiae* [1, 2, 40]. In each case, the homologous *dptG* or *lptG* was included in the design strategies, even though hybrid molecules were being generated. In retrospect, it would be interesting to see if yields of hybrid lipopeptides in certain recombinant strains might be enhanced by co-expressing *dptG* and *lptG*.

Heterologous expression. There are a number of strains that may be useful for the heterologous expression of cryptic gene clusters identified in genome sequencing projects [5, 7]. In some of these strains, certain secondary metabolite biosynthetic gene clusters have been deleted to facilitate the channeling of precursors into the heterologous pathway and to simplify the identification of novel metabolites. In light of the examples of cross-pathway stimulation of secondary metabolite production reviewed here, it may be useful to maintain several functional *mbtH*-like genes in hosts used for the expression of cryptic NRPS gene clusters.

Genome mining. It is significant that the vast majority of important antibiotic biosynthetic pathways that employ NRPS mechanisms have *mbtH* homologs in their biosynthetic gene clusters. One notable exception is cephamycin biosynthesis, which employs a relatively small, three-module NRPS for peptide assembly. Since most important NRPS pathways include *mbtH* homologs, the number of *mbtH* homologs per genome might be used to identify strains that have particularly good or poor capacity to produce novel peptides. Thus, *mbtH* homologs can serve as beacons to identify gifted actinomycete or other bacterial strains. Several examples of potentially gifted strains among actinomycetes (e.g., *S. clavuligerus*, *S. griseus*, *S. roseosporus*, *S. roseum*, *A. medeterranei*, and *A. mirum*) were identified by BLAST analysis (Table 3). Other less-gifted, or secondary metabolically challenged strains were also identified (e.g., *S. sviceus, S.* sp. C, and actinomycetes with genomes <5.0 Mb). Also, among the non-actinomycetes, the vast majority of sequenced bacteria and all Archaea are not good candidates for secondary metabolite production based upon the absence of *mbtH* homologs. These assessments are consistent with data from analysis of the presence or absence of NRPS and PKS genes sequences in sequenced bacterial genomes [4, 15]. It may now be possible to use low-pass DNA sequencing or other methods to screen for gifted actinomycetes encoding multiple functional NRPS-based secondary metabolites, which then can become the focus for drug discovery. The potential novelty of the NRPS pathways can be predicted by BLAST analyses to identify novel paralogs to *mbtH*, *dptG* and other *mbtH*-like genes.

References

1. Alexander DC, Rock J, He X, Miao V, Brian P, Baltz RH (2010) Development of genetic system for lipopeptide combinatorial biosynthesis in *Streptomyces fradiae* and heterologous expression of the A54145 biosynthetic gene cluster. Appl Environ Microbiol 76:6877–6887

2. Alexander DC, Rock J, Gu J-Q, Mascio C, Chu M, Brian P, Baltz RH (2011) Production of novel lipopeptide antibiotics related to A54145 by *Streptomyces fradiae* mutants blocked in biosynthesis of modified amino acids and assignment of *lptI*, *lptK* and *lptL* gene functions. J Antibiot 64:79–87
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
4. Baltz RH (2008) Renaissance in antibacterial discovery from actinomycetes. Curr Opin Pharmacol 8:557–563
5. Baltz RH (2010) *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. J Ind Microbiol Biotechnol 37:759–772
6. Baltz RH (2010) Genomics and the ancient origins of the daptomycin biosynthetic gene cluster. J Antibiot 63:506–511
7. Baltz RH (2011) Strain improvement in actinomycetes in the postgenomic era. J Ind Microbiol Biotechnol 38:657–666
8. Buchko GW, Kim C-Y, Terwilliger TC, Myler PJ (2010) Solution structure of Rv2377c-founding member of the MbtH-like protein family. Tuberculosis 90:245–251
9. Carter RA, Worsley PS, Sawers G, Challis GL, Dilworth MJ, Carson KC, Lawrence JA, Wexler M, Johnston AWB, Yeoman KH (2002) The *vbs* genes that direct synthesis of the siderophore vivibactin in *Rhizobium leguimosum*: their expression in other genera requires ECF σ factor RpoI. Mol Microbiol 44:1153–1166
10. Challis GL (2008) Mining microbial genomes for new natural products and biosynthetic pathways. Microbiology 154:1555–1569
11. Chen H, Walsh CT (2001) Coumarin formation in novobiocin biosynthesis: β-hydroxylation of the aminoacyl enzyme tyrosyl-*S*-NovH by cytochrome P450 NovI. Chem Biol 8:301–312
12. Chen H, Hubbard BK, O'Connor SE, Walsh CT (2002) Formation of β-hydroxy histidine in the biosynthesis of nikkomycin antibiotics. Chem Biol 9:103–112
13. Coëffet-Le Gal M-F, Thurson L, Rich P, Miao V, Baltz RH (2006) Complementation of daptomycin *dptA* and *dptD* deletion mutations *in-trans* and production of hybrid lipopeptide antibiotics. Microbiology 152:2993–3001
14. Corre C, Challis GL (2009) New natural product biosynthetic chemistry discovered by genome mining. Nat Prod Rep 26:977–986
15. Donadio S, Monciardini P, Sosio M (2007) Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. Nat Prod Rep 24:1073–1109
16. Drake EJ, Cao J, Quiz J, Shah MB, Straubinger RM, Gulick AM (2007) The 1.8 Å crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of *Pseudomonas aeruginosa*. J Biol Chem 282:20425–20434
17. Edwards DJ, Gerwick WH (2004) Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. J Am Chem Soc 126:11432–11433
18. Eustáquio AS, Gust B, Galm U, Li S-M, Chater KF, Heide L (2005) Heterologous expression of novobiocin and clorobiocin biosynthetic gene clusters. Appl Environ Microbiol 71:2452–2459
19. Felnagle EA, Rondon MR, Berti AD, Crosby HA, Thomas MG (2007) Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin. Appl Environ Microbiol 73:4162–4170
20. Felnagle EA, Barkei JJ, Park H, Podevels AM, McMahon MD, Drott DW, Thomas MG (2010) MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. Biochemistry 49:8815–8817
21. Heemstra JR Jr, Walsh CT, Sattely ES (2009) Enzymatic tailoring of ornithine in the biosynthesis of the *Rhizobium* cyclic trihydroxamate siderophore vicibactin. J Am Chem Soc 131:15317–15329
22. Hojati Z, Milne C, Harvey B, Gordon L, Borg M, Flett F, Wilkinson B, Sidebottom PJ, Rudd BA, Hayes MA, Smith CP, Micklefield J (2002) Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. Chem Biol 9:1175–1187
23. Huynh MU, Elston MC, Hernandez NM, Ball DB, Kajiyama S, Irie K, Gerwick WH, Edwards DJ (2010) Enzymatic production of (−)-indolactam V by LtxB, a cytochrome P450 monooxygenase. J Nat Prod 73:71–74
24. Imker HJ, Krahn D, Clerc J, Kaiser M, Walsh CT (2010) *N*-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. Chem Biol 17:1077–1083
25. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Ōmura S (2003) Complete genome sequence of and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 21:526–531
26. Jordan IK, Rogozin IB, Wolf YI, Koonin EV (2002) Microevolutionary genomics of bacteria. Theor Popul Biol 61:435–447
27. Lauer B, Russwurm R, Bormann C (2000) Molecular characterization of two genes from *Streptomyces tendae* Tü901 required for the formation of the 4-formyl-4-imidazolin-2-one-containing nucleoside moiety of the peptidyl nucleoside antibiotic nikkomycin. Eur J Biochem 267:1698–1706
28. Lauer B, Russwurm R, Schwarz W, Kálmánczhelyi A, Bruntner C, Rosemeier A, Bormann C (2001) Molecular characterization of co-transcribed genes from *Streptomyces tendae* Tü901 involved in the biosynthesis of the peptidyl moiety and assembly of the peptidyl nucleoside antibiotic nikkomycin. Mol Gen Genet 264:662–673
29. Lautru S, Oves-Costales D, Pernodet J-L, Challis GL (2007) MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. Microbiology 153:1405–1412
30. Losey HC, Peczuh MW, Chen Z, Eggert US, Dong SD, Pelczer I, Kahne D, Walsh CT (2001) Tandem action of glycosyltransferases in the maturation of vancomycin and teicoplanin aglycones: novel glycopeptides. Biochemistry 40:4745–4755
31. Magarvey NA, Haltli B, He M, Greenstein M, Hucul JA (2006) Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant Gram-positive pathogens. Antimicrob Agents Chemother 50:2167–2177
32. Mast YJ, Wohlleben W, Schinko E (2011) Identification and functional characterization of phenylglycine biosynthetic genes involved in pristinamycin biosynthesis in *Streptomyces pristinaespiralis*. J Biotechnol (in press)
33. Medema MH, Trefzer A, Kovalchuk A, van den Berg M, Müller U, Heijne W, Wu L, Alam MT, Ronning CM, Nierman WC, Bovenberg RAL, Breitling R, Takano E (2010) The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. Genome Biol Evol 2:212–224
34. Miao V, Coëffet-LeGal M-F, Brian P, Brost R, Penn J, Whiting A, Martin S, Ford R, Parr I, Bouchard M, Silva CJ, Wrigley SK, Baltz RH (2005) Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. Microbiology 151:1507–1523
35. Miao V, Brost R, Chapple J, She K, Coëffet-Le Gal M-F, Baltz RH (2006) The lipopeptide antibiotic A54145 biosynthetic gene cluster from *Streptomyces fradiae*. J Ind Microbiol Biotechnol 33:129–140
36. Miao V, Coëffet-Le Gal M-F, Nguyen K, Brian P, Penn J, Whiting A, Steele J, Kau D, Martin S, Ford R, Gibson T, Bouchard M, Wrigley SK, Baltz RH (2006) Genetic engineering in *Streptomyces roseosporus* to produce hybrid lipopeptide antibiotics. Chem Biol 13:269–276

37. Müller C, Nolden S, Gebhardt P, Heinzelmann E, Lange C, Puk O, Welzel K, Wohlleben W, Schwartz D (2007) Sequencing and analysis of the biosynthetic gene cluster of the lipopeptide antibiotic friulimicin in *Actinoplanes friuliensis*. Antimicrob Agents Chemother 51:1028–1037
38. Nguyen K, Kau D, Gu J-Q, Brian P, Wrigley SK, Baltz RH, Miao V (2006) A glutamic acid 3-methyltransferase encoded by an accessory gene locus important for daptomycin biosynthesis in *Streptomyces roseosporus*. Mol Microbiol 61:1294–1307
39. Nguyen K, Ritz D, Gu J-Q, Alexander D, Chu M, Miao V, Brian P, Baltz RH (2006) Combinatorial biosynthesis of lipopeptide antibiotics related to daptomycin. Proc Natl Acad Sci USA 103:17462–17467
40. Nguyen K, He X, Alexander DC, Li C, Gu J-Q, Mascio C, Van Praagh A, Morton L, Chu M, Silverman JA, Brian P, Baltz RH (2010) Genetically engineered lipopeptide antibiotics related to A54145 and daptomycin with improved properties. Antimicrob Agents Chemother 54:1404–1413
41. Novichkov PS, Wolf YI, Dubchak I, Koonin EV (2009) Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archael genomes. J Bacteriol 191:65–73
42. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. J Bacteriol 190:4050–4060
43. Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, Haydock SF, Leadlay PF (2007) Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. Nat Biotechnol 25:447–453
44. Penn K, Jenkins C, Nett M, Udwary DW, Gontang EA, McGlinchey RP, Foster B, Lapidus A, Podell S, Allen EE, Moore BS, Jensen PR (2009) Genetic islands link secondary metabolism to functional adaptation in marine Actinobacteria. ISME J 3:1193–1203
45. Pojer F, Li S-M, Heide L (2002) Molecular cloning and sequence analysis of the clorobiocin biosynthetic gene cluster: new insights into the biosynthesis of aminocoumarin antibiotics. Microbiology 148:3901–3911
46. Quadri LEN, Sello J, Keating TA, Weinreb PH, Walsh CT (1998) Identification of *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. Chem Biol 5:631–645
47. Reed JA, Walsh CT (2007) The lyngbyatoxin biosynthetic assembly line: chain release by four-electron reduction of a dipeptidyl thioester to the corresponding alcohol. J Am Chem Soc 129:15762–15763
48. Rhee K-H, Davies J (2006) Transcription analysis of daptomycin biosynthesis genes in *Streptomyces roseosporus*. J Microbiol Biotechnol 16:1841–1848
49. Solenberg PJ, Matsushima P, Stack DR, Wilkie SC, Thompson RC, Baltz RH (1997) Production of hybrid glycopeptide antibiotics in vitro and in *Streptomyces toyocaensis*. Chem Biol 4:195–202
50. Sosio M, Stinchi S, Beltrametti F, Lazzarini A, Donadio S (2003) The gene cluster for the biosynthesis of the glycopeptide antibiotic A40926 by *Nonomuraea* species. Chem Biol 10:541–549
51. Sosio M, Kloosterman H, Bianchi A, de Vreugd P, Dijkhuizen L, Donodio S (2004) Organization of the teicoplanin gene cluster in *Actinoplanes teichomyceticus*. Microbiology 150:95–102
52. Stegmann E, Rausch C, Stockert S, Burkert D, Wohlleben W (2006) The small MbtH-like protein encoded by an internal gene of the balhimycin biosynthetic gene cluster is not required for glycopeptide production. FEMS Microbiol Lett 262:85–92
53. Thomas MG, Chan YA, Ozanick SG (2003) Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. Antimicrob Agents Chemother 47:2823–2830
54. Wang G, Nie L, Tan H (2003) Cloning and characterization of *sanO*, a gene involved in nikkomycin biosynthesis in *Streptomyces ansochromogenes*. Lett Appl Microbiol 36:452–457
55. Wang Z-X, Li S-M, Heide L (2000) Identification of the coumermycin A_1 biosynthetic gene cluster of *Streptomyces rishiriensis* DSM 40489. Antimicrob Agents Chemother 44:3040–3048
56. Wolpert M, Gust B, Kammerer B, Heide L (2007) Effects of deletion of *mbtH*-like genes on chlorobiocin biosynthesis in *Streptomyces coelicolor*. Microbiology 153:1413–1423
57. Yin X, Zabriskie M (2006) The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*. Microbiology 152:2969–2983
58. Yu G, Jia X, Wen J, Lu W, Wang G, Caiyin Q, Chen Y (2011) Strain improvement of *Streptomyces roseosporus* for daptomycin production by rational screening of He-Ne laser and NTG induced mutants and kinetic modeling. Appl Biochem Biotechnol 163:729–743
59. Zerbe K, Pylypenko O, Vitali F, Zhang W, Rouset S, Heck M, Vrijbloed JW, Bischoff D, Bister B, Süssmuth RD, Pelzer S, Wohlleben W, Robinson JA, Schlichting I (2002) Crystal structure of OxyB, a cytochrome P450 implicated in the oxidative phenol coupling reaction during vancomycin biosynthesis. J Biol Chem 277:47476–47485
60. Zerikly M, Challis GL (2009) Strategies for the discovery of new natural products by genome mining. ChemBioChem 10:625–633
61. Zhang W, Ostach B, Walsh CT (2010) Identification of the biosynthetic gene cluster for the pacidamycin group of peptidyl nucleoside antibiotics. Proc Natl Acad Sci USA 107:16828–16833
62. Zhang W, Heemstra JR, Walsh CT, Imker HJ (2010) Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. Biochemistry 49:9946–9947
63. Zhao W, Zhong Y, Yuan H, Wang J, Zheng H, Wang Y, Cen X, Xu F, Bai J, Han X, Lu G, Zhu Y, Shao Z, Yan H, Li C, Peng N, Zhang Z, Zhang Y, Lin W, Fan Y, Qin Z, Hu Y, Zhu B, Wang S, Ding X, Zhao G-P (2010) Complete genome sequence of the rifamycin SV-producing *Amycolatopsis mediterranei* U32 revealed its genetic characteristics in phylogeny and metabolism. Cell Res 20:1096–1108