npg

ORIGINAL ARTICLE

Genomics and the ancient origins of the daptomycin biosynthetic gene cluster

Richard H Baltz

Daptomycin is a clinically useful lipopeptide antibiotic produced by *Streptomyces roseosporus*. The antibiotic is assembled by a nonribosomal peptide synthetase (NRPS) mechanism, and the cyclized tridecapeptide contains three non-proteinogenic L-amino acids: ornithine (Orn), 3-methyl-glutamic acid (3mGlu) and kynurinine (Kyn). Daptomycin also has three D-amino acids. The two genes that encode proteins involved in the formation of 3mGlu and Kyn have no known orthologs, and hence they are good probes to search for daptomycin-like gene clusters among newly sequenced actinomycete genomes. A recent search with these genes revealed a daptomycin-like gene cluster in *Saccharomonospora viridis*, a causative agent for farmer's lung disease. *S. viridis* has a genome of 4.3 Mb, which is about one half the size of *Streptomyces* genomes. Searches for other NRPS and polyketide synthase (PKS) genes revealed only one other NRPS gene and no PKS genes in *S. viridis*. The orthologous lipopeptide biosynthetic genes and gene products in *S. viridis* and *S. roseosporus* have diverged extensively from a last common ancestor dating back to a pathway that had evolved over 1 billion years ago.
The Journal of Antibiotics (2010) **63,** 506–511; doi:10.1038/ja.2010.82; published online 21 July 2010

INTRODUCTION

Lipopeptide and β-lactam antibiotics share a common nonribosomal peptide synthetase (NRPS) mechanism for peptide assembly.[1] My first work on antibiotic biosynthesis at Eli Lilly and Company involved the radioisotopic labeling of cephalosporin C with L-valine-*1*-[14]C.[2] I was also adapting a defined medium devised by Arny Demain and colleagues[3,4] for biosynthetic studies. Some of the first scientific papers that I read were those of Arny Demain, who was working on the biosynthesis of cephalosporin C at Merck. The work that we did on cephalosporin biosynthesis at Lilly stimulated my interest in research, and helped me decide to pursue advanced degrees in Microbiology at the University of Illinois.

The first time that I heard Arny Demain speak was at the first ASM Conference on the Genetics and Molecular Biology of Industrial Microorganisms (GMBIM) in Orlando, Florida, in 1976. I remember thinking at that time that I would like to be like Arny some day. He was and continues to be an influential spokesman for industrial microbiology, and has encouraged international scientific communication and cooperation between academia and industry. He has been one of my role models over the years, and I have had the privilege to work with him on the Genetics of Industrial Microorganisms International Committee, and recently on editing the Third Edition of the Manual of Industrial Microbiology and Biotechnology.

One of the most interesting antibiotic projects that I have worked on at Lilly and at Cubist Pharmaceuticals involves the biosynthesis of the lipopeptide A21978C and its derivative, daptomycin. A21978C (Figure 1) was discovered by Eli Lilly and Company,[5] and daptomycin is a semisynthetic derivative of A21978C containing a decanoic acid side chain rather than the natural lipid side chains of A21978C factors.[6] Daptomycin was licensed from Lilly to Cubist Pharmaceuticals,[7] and has been approved for the treatment of difficult-to-treat skin and skin structure infections caused by Gram-positive pathogens,[8] and for bacteremia and right-sided endocarditis caused by *Staphylococcus aureus*, including strains resistant to methicillin (MRSA (methicillin-resistant *S. aureus*)).[9] Daptomycin failed to meet non-inferiority standards in a clinical trial for community-acquired pneumonia caused by *Streptococcus pneumoniae*,[10] apparently because it becomes sequestered in lung surfactant.[11] Many derivatives of A21978C have been made by chemical modifications of the lipid side chain or by additions to the δ-amino group of ornithine (Orn$_6$), but none have yet proven superior to daptomycin.[12,13]

A21978C and daptomycin are members of a family of acidic cyclic lipopeptide antibiotics with common evolutionary origins.[13] Other members of this family include A54145, calcium-dependent antibiotic and friulimicins, among others.[12,13] Although these antibiotics share some common features, including amino acid chirality and the presence of calcium ion binding motifs in the same positions, they differ in primary amino acid sequence in the ten-member rings and exocyclic amino acid tails, and in lipid side chains.[12] Recently, combinatorial biosynthetic methods were developed to exploit the
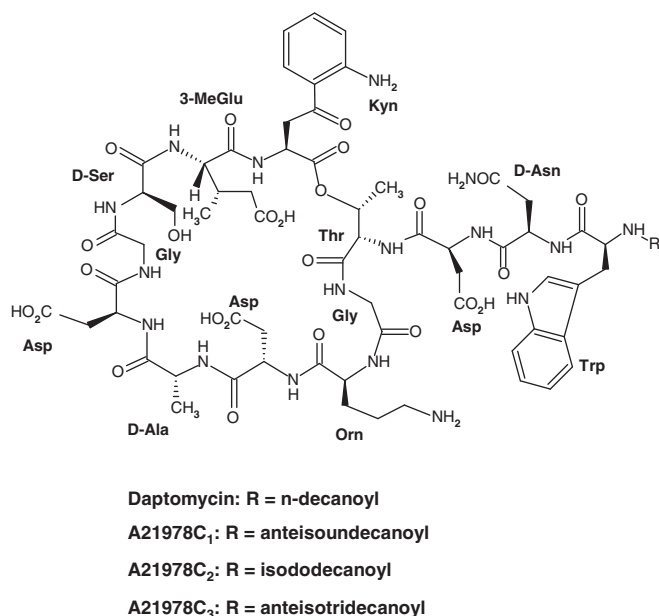
CognoGen Biotechnology Consulting, Indianapolis, IN, USA
Correspondence: Dr RH Baltz, CognoGen Biotechnology Consulting, 6438 North Olney Street, Indianapolis, IN 46220, USA.
E-mail: rbaltz@cubist.com

Daptomycin: R = n-decanoyl

A21978C$_1$: R = anteisoundecanoyl

A21978C$_2$: R = isododecanoyl

A21978C$_3$: R = anteisotridecanoyl

Figure 1 Structures of A21978C factors and daptomycin. Reproduced from Baltz et al.[12]

Table 1 *S. viridis* genes predicted to be involved in lipopeptide biosynthesis

Gene	Protein	Homolog[a]	Microorganism	Function
Svir_18330	YP_003133683	dptP	S. roseosporus	Resistance
Svir_18340	YP_003133684	dptN	S. roseosporus	Transport
Svir_18350	YP_003133685	dptM	S. roseosporus	Transport
Svir_18360	YP_003133686	dptE	S. roseosporus	Acyl-CoA ligase
Svir_18370	YP_003133687	lipB	A. friuliensis	Acyl-CoA DH
Svir_18380	YP_003133688	'acad'	S. ghanaensis	Acyl-CoA DH
Svir_18390	YP_003133689	dptF	S. roseosporus	ACP
Svir_18400	YP_003133690	dptA	S. roseosporus	NRPS
Svir_18410	YP_003133691	dptBC	S. roseosporus	NRPS
Svir_18420	YP_003133692	dptD	S. roseosporus	NRPS
Svir_18430	YP_003133693	dptG	S. roseosporus	Unknown
Svir_18440	YP_003133694	dptI	S. roseosporus	Methyltransferase
Svir_18450	YP_003133695	dptJ	S. roseosporus	TDO
Svir_18460	YP_003133696	'est'	Frankia	Putative esterase
Svir_18470	YP_003133697	'test'	S. ambofaciens	Thioesterase
Svir_18480	YP_003133698	'lysR'	R. opaceus	LysR regulator

Abbreviations: ACP, acyl carrier protein; DH, dehydrogenase; NRPS, nonribosomal peptide synthetase; TDO, tryptophan-2,3-dioxygenase.
[a]The four genes that have not been assigned gene names are listed as 'acad' (acyl-CoA dehydrogenase), 'est' (esterase), 'test' (thioesterase) and 'lysR' (lysR homolog).

differences in primary amino acid sequences in these lipopeptides and they generated large sets of derivatives of daptomycin and A54145.[14–18] Many analogs were very potent, and a number of analogs more closely related to A54145 than to daptomycin had potent antibacterial activity in the presence of bovine surfactant.[17,18].

Another approach to discover antibiotics related to daptomycin (and other highly active antibiotics) has been afforded by the diminished cost for genome sequencing. It is now well established that actinomycetes with large genomes encode multiple secondary metabolite biosynthetic pathways, most of which are cryptic under standard fermentation conditions.[19,20] Genome sequencing provides a means to discover novel chemical scaffolds as well as derivatives of known structures. From an evolutionary point of view, the sequencing of multiple actinomycete genomes will also provide multiple copies of known pathways, thus allowing phylogenic calculations to determine the age of antibiotic biosynthetic and resistance genes.

In this report, a search for specific genes involved in daptomycin biosynthesis revealed a cluster of secondary metabolite biosynthetic genes highly related to those of daptomycin in *Saccharomonospora viridis*, an actinomycete associated with farmer's lung disease.[21] *S. viridis* has a genome of ~4.3 Mb, about one half the size of streptomycete genomes. The daptomycin-like pathway seems to be the only substantial secondary metabolite biosynthetic pathway that uses NRPS or polyketide synthase (PKS) mechanisms in *S. viridis*. Calculation of the last common ancestor of the lipopeptide pathway in *S. viridis* and *S. roseosporus* indicates that the daptomycin-like pathway may have evolved over a billion years ago.

MATERIALS AND METHODS
Nucleotide and protein searches
Protein searches were carried out using BLASTp[22] (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Assignment of amino acid binding pocket specifities
The amino acid binding pocket codes[23,24] in the NRPS enzymes were determined as described[25] using NRPS predictor (http://www-ab.informatik.

uni-tuebingen.de/software), and were compared with those used by the daptomycin NRPS enzymes.[26]

Calculation of divergence time from the last common ancestor
The divergence time from the last common ancestors of *S. roseosporus* and *S. viridis*, and of the lipopeptide biosynthetic genes present in these organisms, was determined by the method of Feng et al.[27] by comparing amino acid identities of orthologous proteins.

RESULTS
Search for daptomycin-like biosynthetic genes
There are two genes, *dptI* and *dptJ*, in the daptomycin biosynthetic gene cluster that until very recently did not have any orthologs in GenBank. The *dptI* gene encodes a methyltransferase involved in the formation of 3-methyl-glutamic acid (3mGlu) from α-ketoglutarate.[15,28] The DptI protein has only two distantly related functional homologs, LptI[29,30] from the lipopeptide A54145 pathway and GlmT[28,31] from the calcium-dependent antibiotic pathway. DptI shows only 37 and 36% amino acid identities to LptI and GlmT, respectively. The *dptJ* gene encodes a (TDO) tryptophan-2,3-dioxygenase that is involved in the formation of kynurinine (Kyn) for incorporation at position 13 in the A21978C tridecapeptide. It has no known functional orthologs, and its gene product shares only 30% amino acid identities with KynA, a TDO likely involved in primary metabolism in *S. roseosporus*.[13] We searched GenBank using BLASTp to find protein sequences related to DptI and DptJ, and homologs with amino acid identities of 53 and 47%, respectively, were encoded by the genome of *S. viridis* (Table 1). The *dptI* and *dptJ* genes were also contiguous in *S. viridis* as observed in *S. roseosporus*.[26] The DptI homolog (DptI-*sv*) shared 39% amino acid identities with LptI and GlmT, and hence it is clearly more closely related to DptI.

The discovery of *dptI* and *dptJ* homologs in *S. viridis* prompted further BLASTp analyses with the other daptomycin biosynthetic enzymes. Tables 1 and 2 summarize the results. *S. viridis* encodes three very large NRPS enzymes with molecular masses very similar to those of DptA, DptBC and DptD, which are involved in assembling

Table 2 Amino acid sequence identities of putative daptomycin biosynthetic proteins encoded by *S. roseosporus* and *S. viridis*

dpt protein	Size (aa)	Function	S. viridis homolog[a]	Size (aa)	% Identities (aa/aa)
DptA	5830	NRPS	DptA-sv	5827	51 (2968/5827)
DptBC	7338	NRPS	DptBC-sv	7310	52 (3871/7449)
DptD	2379	NRPS		2365	56 (1325/2375)
DptE	597	Acyl-CoA ligase	DptD-sv	577	48 (259/539)
DptF	89	Acyl carrier protein	DptE-sv	87	38 (33/87)
DptG	75	Unknown	DptF-sv	68	69 (47/68)
DptH	271	Thioesterase	DptG-sv	—	—
DptI	328	Methyltransferase	No hit	349	54 (179/333)
DptJ	245	TDO	DptI-sv	237	47 (105/222)
DptM	319	ABC transporter: ATP-binding	DptJ-sv	313	61 (187/306)
DptN	289	ABC transporter: permease	DptM-sv	282	67 (146/219)
DptP	206	Transport/resistance	DptN-sv	230	52 (91/176)
Ave			DptP-sv		52 (9211/17 601)

Abbreviations: NRPS, nonribosomal peptide synthetase; TDO, tryptophan-2,3-dioxygenase.
[a]The Dpt proteins encoded by *S. viridis* have –sv appended to distinguish them from the Dpt proteins encoded by *S. roseosporus*.
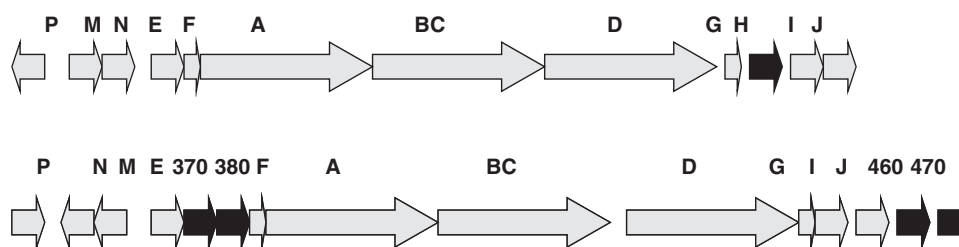

Figure 2 (Top) A21978C (daptomycin) biosynthetic gene cluster in *S. roseosporus*. (Bottom) Lipopeptide biosynthetic gene cluster in *S. viridis*. The genes *Svir_18370*, *Svir_18380*, *Svir_18460* and *Svir_18470* are presented as 370, 380, 460 and 470, respectively. Genes are not drawn to scale, and those shown in black are not conserved between the two pathways.

the tridecapeptide during daptomycin biosynthesis.[12,13,26] The amino acid identities range from 51 to 56% (Table 2). Similarly, *S. viridis* encodes homologs with very similar sizes to DptE and DptF, and with 48 and 39% amino acid identities, respectively. These proteins are involved in coupling long-chain fatty acids to Trp_1 to initiate daptomycin assembly.[12,13,32] It is noteworthy that the DptE and DptF proteins produced in *S. roseosporus* show a slightly higher amino acid conservation (51 and 46% identities) to the fused LptEF protein involved in coupling the lipid side chain to Trp_1 in A54145 biosynthesis.[13] This observation will be further addressed below. *S. viridis* also encodes homologs to DptG, a protein of unknown function required for optimal daptomycin production,[15] and DptM, DptN and DptP, which may be involved in daptomycin export and/or resistance.[13] DptM, DptN and DptP homologs are also encoded by the A54145 biosynthetic gene cluster, whereas only DptM and DptN homologs are encoded by the friulimicin cluster.[13] It is noteworthy that DptP shows 94% amino acid identities with LptP, whereas DptP-sv shows only 50% amino acid identities with LptP. This is a clear indication that LptP is derived from a streptomycete DptP acquired by horizontal gene transfer approximately 100 Myr ago. As the *lptP* gene may have further evolved in function, and may not be a true ortholog to *dptP*, the horizontal transfer could have occurred in more recent evolutionary time (that is, < 100 Myr ago).

S. viridis does not encode a homolog of DptH, a likely editing thioesterase required for optimal daptomycin production,[15] but may use a different gene(s) to encode this function (see below).

Lipopeptide biosynthetic gene organization in *S. viridis*

If the genes identified in this study are in fact involved in biosynthesis of daptomycin or a closely related lipopeptide antibiotic, then the gene cluster in *S. viridis* might resemble the daptomycin gene cluster in *S. roseosporus*. Table 1 indicates that the lipopeptide biosynthetic genes are clustered in *S. viridis*, and Figure 2 shows that they have a similar organization to those in *S. roseosporus*. The differences are that the *dptP* gene and the *dptMN* genes (encoding an ABC transporter) are inverted; two new genes (*Svir_18370* and *Svir_18380*) are inserted between *dptE* and *dptF* genes in *S. viridis*; and *S. viridis* lacks a *dptH* (thioesterase) gene, but has two genes (*Svir_18460* and *Svir_18470*) downstream of the *dptJ* homolog, one or both of which may have role(s) as editing thioesterase(s), based upon their annotations as putative esterase and thioesterase, respectively (Table 1). In addition, downstream of *Svir_18470* is a *lysR* homolog (*Svir_18480*) that has no homolog in the daptoymycin gene cluster.

The presence of *Svir_18370* and *Svir_18380* genes between the *dptE* and *dptF* homologs suggests that they may be involved in processing the lipid side chain. DptE and DptF function as acyl-CoA ligase and acyl carrier protein in the coupling of long-chain fatty acids to the N-terminal Trp_1 to initiate daptomycin biosynthesis.[13,32] Both *Svir_18370* and *Svir_18380* encode acyl-CoA dehydrogenases, and the *Svir_18370* gene product showed 45% amino acid sequence identities with LipB, an acyl-CoA dehydrogenase that introduces the Δ*cis*3 double bond in the lipid side chain of the lipopeptide antibiotic friulimicin in *Actinoplanes friuliensis*.[13,33] Furthermore, the *lipB* gene

in *A. friuliensis* is located between the *lipA* and *lipD* genes, which are homologs of *dptE* and *dptF*.[13,33] Therefore, it would not be surprising if the natural lipopeptide(s) produced by *S. viridis* have lipid side chains with one or two double bonds.

The NRPS genes are aligned between *dptF* and *dptG* in *S. viridis* just as they are in *S. roseosporus*. In *S. roseosporus*, the *dptA*, *dptBC* and *dptD* genes have overlapping stop and start codons,[26] whereas only *dptA* and *dptBC* have overlapping stop and start codons in *S. viridis*. In addition, the *dptD*, *dptG* and *dptI* homologs in *S. viridis* are likely to have overlapping stop and start codons, based upon the annotations in GenBank.

Amino acid binding pocket specificities

The observation that *S. viridis* encodes three giant NRPS enzymes of nearly identical size relative to DptA, DptBC and DptD (Table 2) predicts that they synthesize a tridecapeptide related to daptomycin. The end-to-end homologies also indicate that they have epimerase domains located at the same three positions as observed in daptomycin NRPSs. If these NRPS enzymes encode a tridecapeptide identical to or highly related to that of daptomycin, then this should be reflected in the amino acid binding pockets of the adenylation domains.[23,24] The amino acid binding pockets of authentic daptomycin and the three NRPS proteins encoded by *S. viridis* are shown in Table 3. Daptomycin has three unusual non-proteinogenic amino acids, Orn_6, $3mGlu_{12}$ and Kyn_{13}, the combination of which has not been observed in any other peptide in nature. The amino acid binding pocket for Kyn has a unique code observed only in DptD. It is therefore noteworthy that the binding pocket for the second module in DptD-*sv* has a binding pocket identical to that of Kyn_{13} in DptD

(Table 3). In addition, the first module of DptBC-*sv* has a binding pocket identical to that of Orn_6 (DptBC module 1), and the binding code in the second module of DptD-*sv* differs from the $3mGlu_{12}$ binding code of DptD by a single conservative substitution of Val for Ile at position eight of the pocket. This strongly suggests that *S. viridis* encodes a tridecapeptide containing Orn_6, $3mGlu_{12}$ and Kyn_{13}.

Seven of the other amino acid binding pockets deviate from those of the daptomycin NRPS by 1 or 2 amino acids. In most cases, the differences are conservative (for example, Val/Ile, Val/Leu and Thr/Ser). It therefore seems likely that the amino acids at these positions will be identical to those observed in daptomycin. The binding pocket for DptA module 2 in *S. viridis* differed from that of *S. roseosporus* by two substitutions and one position that could not be assigned. It is possible that this binding pocket specifies an amino acid other than Asn.

The binding pockets that correspond to the $D-Ala_8$ and $D-Ser_{11}$ in daptomycin differed by 3 and 5 amino acids from those in DptBC modules 3 and 6, respectively, but differed from each other by only 2 amino acids (Table 3). Both of the binding pockets were also more closely related to that of $D-Ala_8$ than $D-Ser_{11}$ in DptBC. It seems possible that the tridecapeptide encoded by the *S. viridis* NRPS may have D-Ala at positions 8 and 11, or some other related amino acid. It has been shown previously in molecular engineering studies that D-Ala can be substituted for D-Ser at position 11 without any loss of antibacterial activity or spectrum.[16] The amino acid binding pocket analysis clearly predicts that the lipopeptide encoded by *S. viridis* is highly related to daptomycin, but may show some subtle structural differences.

Table 3 Amino acid binding pockets in A domains of NRPSs

NRPS subunit	Module	Module structure	Amino acid binding pocket	Amino acid specificity[a]
DptA	SR-1	CAT	DVSSIGAVEK	Trp
	SV-1	CAT	DVSTVGAVEK	(Trp)
	SR-2	CATE	DLTKLGDVNK	Asn
	SV-2	CATE	DLT_MGDVHK	(Asn, other)
	SR-3	CAT	DLTKLGAVNK	Asp
	SV-3	CAT	DLTKVGAVNK	(Asp)
	SR-4	CAT	DFWSVGMVHK	Thr
	SV-4	CAT	DFWNIGMVHK	(Thr)
	SR-5	CAT	DILQLGVIWK	Gly
	SV-5	CAT	DILQLGLIWK	(Gly)
DptBC	SR-1	CAT	DTWDMGYVDK	Orn
	SV-1	CAT	DTWDMGYVDK	(Orn)
	SR-2	CAT	DLTKLGAVNK	Asp
	SV-2	CAT	DLTKVGAVNK	(Asp)
	SR-3	CATE	DVVSAAFVYK	Ala
	SV-3	CATE	DVFSVAAVYK	(Ala, Ser, other)
	SR-4	CAT	DLTKLGAVC*K	Asp
	SV-4	CAT	DLTKVGAVNK	(Asp)
	SR-5	CAT	DILQVGMIWK	Gly
	SV-5	CAT	DILQLGVIWK	(Gly)
	SR-6	CATE	DVWHISLVYK	Ser
	SV-6	CATE	DVFSVALVDK	(Ala, Ser, other)
DptD	SR-1	CAT	DLGKTGVINK	3mGlu
	SV-1	CAT	DLGKTGVVNK	(3mGlu)
	SR-2	CATTe	DAWTTTGVGK	Kyn
	SV-2	CATTe	DAWTTTGVGK	(Kyn)

Abbreviations: 3mGlu; 3-methyl-glutamic acid; Kyn, kynurinine; NRPS, nonribosomal peptide synthetase; Orn, ornithine; SR-1, module 1 in Dpt protein in *S. roseosporus*; SV-1, module 1 in Dpt protein in *S. Viridis*.
[a]The proposed amino acid binding specificity is shown in parenthesis.

Search for other NRPS and PKS genes in S. viridis

S. viridis is an actinomycete with a circular genome of ∼4.3 Mb. This is ∼50% smaller than the genomes of streptomycetes. In streptomycetes, many of the secondary metabolite biosynthetic functions are located in the subtelomeric regions at the linear chromosome extremities, and the core primary metabolic functions are located in a central region of ∼4.5 Mb. Thus, it seems that S. viridis has coding capacity mainly for primary metabolic functions, and might not encode multiple secondary metabolic functions as observed from actinomycetes with large genomes.[19,20,34] BLASTp analysis using typical PKS and NRPS amino acid sequences as probes indicated that the S. viridis genome has no genes encoding PKSs, and only one other gene encoding an NRPS. Therefore, the presence of a complete biosynthetic pathway for a lipopeptide related to daptomycin suggests that this pathway may have a critical role in S. viridis growth and survival. It is noteworthy that S. viridis does not encode a primary metabolic TDO, further emphasizing the significance of DptJ for the formation of Kyn for the postulated lipopeptide biosynthesis in S. viridis.

Calculation of time to last common ancestor

If the putative daptomycin biosynthetic genes identified in S. viridis are indeed orthologs of those from the daptomycin gene cluster in S. roseosporus, and if orthologs encoding important secondary metabolites are under the same evolutionary constraints as are orthologs of primary metabolic functions required for viability, then we can use the drift in amino acid sequence identities to calculate the time to the last common ancestor of the daptomycin gene clusters identified in S. roseosporus and S. viridis. We can also compare this with the calculation of last common ancestor of S. roseosporus and S. viridis to address the question of vertical versus horizontal transmission of daptomycin genes. Feng et al.[27] have estimated that the last common ancestor of Escherichia coli and Bacillus subtilis existed approximately 2000 Myr ago. The proteins encoded by both E. coli and B. subtilis show an average of 45% amino acid identities. The drift in amino acid identities over time is an logarithmic function defined by the Grishin formula: $q = \ln(1+2D)/2D$, where q is the fraction of unchanged residues and D is the evolutionary distance. Feng et al.[27] have related D to evolutionary time, and hence it is possible to calculate the evolutionary time from last common ancestor based upon q. As a reference, the time to last common ancestor for S. roseosporus and several other Streptomyces species was calculated using the glutamine synthetase (GlnA) as the protein clock. Twelve Streptomyces species showed 89 to 92% amino acid identities over the full-length GlnA protein. This corresponds to divergence times of 200–280 Myr ago. S. roseosporus and S. viridis showed 71% amino acid identities in GlnA, translating to a divergence time of approximately 860 Myr ago. The lipopeptide biosynthetic proteins from the S. roseosporus daptomycin gene cluster were summed and compared with the apparent orthologous proteins from the S. viridis lipopeptide pathway, and the average amino acid identity was 52%, which corresponds to a divergence time of approximately 1640 Myr ago. However, some of the proteins may not be true orthologs. For instance, the DptE and DptF homologs have likely diverged somewhat to accommodate predicted differences in fatty acid starter units (that is, saturated in S. roseosporus and unsaturated in S. viridis). The amino acids in the proteins showing the least divergence, DptG, DptM and DptN, show an average of 64% identity. This corresponds to a divergence time of 1,100 MY ago. These calculations strongly suggest that these lipopeptide biosynthetic genes evolved before the time of divergence from the last common ancestor of Streptomyces and Saccharomonospora

lineages. This would indicate that the daptomycin gene cluster must have been transferred horizontally to one or both of the S. roseosporus and S. viridis strains, probably from some other actinomycete(s).

DISCUSSION

S. viridis is an actinomycete with a relatively small genome, about one half the size of the genomes of Streptomyces species. S. viridis therefore must not require the large repertoire of metabolic functions common to the free-living Streptomyces to survive in its more selective thermophilic niche. The observation that S. viridis encodes a single antibiotic biosynthetic pathway highly related to that of daptomycin strongly suggests that this pathway is important for the growth and survival of this naturally genome-minimized actinomycete. If so, then we might expect to find this lipopeptide pathway in other strains of S. viridis. This could be easily tested by sequencing several more S. viridis strains from different geographical sources.

S. roseosporus strains that produce daptomycin have been isolated at least twice at Eli Lilly and Company. Two strains have been compared by partially sequencing the three NRPS genes (V Miao and RH Baltz, unpublished), and they showed 98.7% amino acid identities.[35] This corresponds to a divergence time of approximately 20–30 Myr ago. It is noteworthy that 16S rRNA sequence analysis also suggested that these two S. roseosporus strains diverged approximately 20 Myr ago.[35] The relatively recent divergence led to the erroneous suggestion that the daptomycin pathway might have evolved relatively recently.[35] In fact, the data reported in this study refute this notion, and indicate that this pathway predates the establishment of the genus Streptomyces, which seems to have branched off approximately 300 Myr ago. This strongly suggests that S. roseosporus, or a recent ancestor, obtained the daptomycin gene cluster by horizontal gene transfer from some other actinomycete, but probably not from a strain of S. viridis.

This preliminary analysis of the age of the daptomycin-like gene clusters in S. roseosporus and S.viridis suggests that this complicated NRPS pathway is over a billion years old. However, it is not yet established that orthologous secondary metabolic biosynthetic genes diverge at the same rates as genes encoding highly conserved primary metabolic functions. If they diverge faster, then the divergence times for lipopeptide biosynthetic genes estimated in this study may be inflated. In addition, some of the genes may not be true orthologs (for example, dptE and dptF). As more genomes are sequenced, this question can be addressed directly. As secondary metabolite biosynthetic gene clusters are identified in pairs or groups of strains in which a strong case can be made for the presence of the pathways in a progenitor strain before speciation, thus ruling out more recent horizontal transfers, then sequencing multiple isolates of the producing species will allow direct comparisons of divergence rates of primary and secondary metabolic genes and proteins. For instance, S. coelicolor and S. lividans (which are in fact a single species) encode the same three antibiotics: actinorhodin, undecylprodigiosyn and calcium-dependent antibiotic. However, these strains may be too closely related to give meaningful calculations. Another example is erythromycin production in Saccharopolyspora erythraea. There are several independent isolates of S. erythraea that produce erythromycin,[36] and analysis of 16S rRNA sequences from four strains showed pairwise divergence times ranging from 7 to 73 Myr ago (V Miao and RH Baltz, unpublished). This range of divergence within a single species provides a good system to compare the divergence of erythromycin biosynthetic genes and enzymes relative to primary metabolic functions. Undoubtedly, there will be other secondary metabolic pathways suitable for additional evolutionary analysis as more actinomycete genomes are sequenced.

The notion that antibiotic biosynthetic pathways, including resistance mechanisms, may in fact be ancient functions present in microbes hundreds of millions[35] to over one billion years ago deserves to be analyzed comprehensively. There currently exists within the lay population and the general scientific community a naive understanding of antibiotic resistance, and genomic sequencing studies will help educate us on the ubiquitous nature and ancient origin of antibiotic biosynthesis and resistance, and perhaps help us to embrace it in the context of the vast number of microbes that coexist with us on this planet.

Finally, the observation that *S. viridis* encodes a lipopeptide antibiotic apparently highly related to daptomycin could not have been made at this time if the DOE had not chosen to sequence *S. viridis*, a causative agent of farmer's lung disease. This unpredicted outcome supports the notion that a large number of actinomycetes covering all currently known genera should be sequenced and analyzed. The data from comprehensive genome sequencing will be of enormous value for fundamental studies on the origins of antibiotic biosynthetic and resistance genes, and for practical applications aimed at drug discovery.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

1 Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).

2 Huber, F. M., Baltz, R. H. & Caltrider, P. G. Formation of desacetylcephalosporin C in cephalosporin C fermentation. *Appl. Microbiol.* **16**, 1011–1014 (1968).

3 Demain, A. L. L-valine: a precursor of cephalosporin C. *Biochem. Biophys. Res. Comm.* **18**, 45–48 (1963).

4 Demain, A. L., Newkirk, J. F. & Hendlin, D. Effect of methionine, norleucine, and lysine derivatives on cephalosporin C formation in chemically defined media. *J. Bacteriol.* **85**, 339–344 (1963).

5 Debono, M. *et al.* A21978C, a complex of new acidic peptide antibiotics: isolation, chemistry, and mass spectral structure elucidation. *J. Antibiot.* **40**, 761–777 (1987).

6 Debono, M. *et al.* Enzymatic and chemical modifications of lipopeptide antibiotic A21978C: the synthesis and evaluation of daptomycin (LY146032). *J. Antibiot.* **41**, 1093–1105 (1988).

7 Eisenstein, B. I., Oleson##Jr, F. B. & Baltz,, R. H. Daptomycin: from the mountain to the clinic, with the essential help from Francis Tally, MD. *Clin. Inf. Dis.* **50**, S10–S15 (2010).

8 Arbiet, R. D., Maki, D., Tally, F. P., Campanaro, E. & Eisenstein, B. I. The safety and efficacy of daptomycin for the treatment of complicated skin and skin-structure infections. *Clin. Infect. Dis.* **38**, 1673–1681 (2004).

9 Fowler, V. G. *et al.* Daptomycin versus standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*. *N. Engl. J. Med.* **355**, 653–655 (2006).

10 Pertel, P. E. *et al.* Effects of prior effective therapy on the efficacy of daptomycin and ceftriaxone for the treatment of community-acquired pneumonia. *Clin. Infect. Dis.* **46**, 1142–1151 (2008).

11 Silverman, J. A., Morton, L. I., Vanpraagh, A. D., Li, T. & Alder, J. Inhibition of daptomycin by pulmonary surfactant: *in vitro* modeling and clinical impact. *J. Infect. Dis.* **191**, 2149–2152 (2005).

12 Baltz, R. H., Miao, V. & Wrigley, S. K. Natural products to drugs: daptomycin and related lipopeptide antibiotics. *Nat. Prod. Rep.* **22**, 717–741 (2005).

13 Baltz, R. H. Biosynthesis and genetic engineering of lipopeptide antibiotics related to daptomycin. *Curr. Top. Med. Chem.* **8**, 618–638 (2008).

14 Miao, V. *et al.* Genetic engineering in *Streptomyces roseosporus* to produce hybrid lipopeptide antibiotics. *Chem. Biol.* **13**, 269–276 (2006).

15 Nguyen, K. T. *et al.* Identification of a glutamic acid 3-methyltransferase gene by functional analysis of an accessory gene locus important for daptomycin biosynthesis in *Streptomyces roseosporus*. *Mol. Microbiol.* **61**, 1294–1307 (2006).

16 Nguyen, K. *et al.* Combinatorial biosynthesis of lipopeptide antibiotics related to daptomycin. *Proc. Natl Acad. Sci. USA* **103**, 17462–17467 (2006).

17 Nguyen, K. *et al.* Genetically engineered lipopeptide antibiotics related to A54145 and daptomycin with improved properties. *Antimicrob. Agents Chemother.* **54**, 1404–1413 (2010).

18 Baltz, R. H., Nguyen, K. T. & Alexander, D. C. *in Manual of Industrial Microbiology and Biotechnology* (eds Baltz, R. H., Demain, A. L., & Davies, J. E.) 391–410 (ASM, Washington, D C., 2010).

19 Baltz, R. H. Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharmacol.* **8**, 557–563 (2008).

20 Nett, M., Ikeda, H. & Moore, B. S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26**, 1362–1384 (2009).

21 Amner, W., Edwards, C. & McCarthy, A. J. Improved medium for recovery and enumeration of the farmer's lung disease organism, *Saccharomonospora viridis*. *Appl. Environ. Microbiol.* **55**, 2669–2674 (1989).

22 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

23 Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).

24 Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211–224 (2000).

25 Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide synthetase (NRPS) using transductive support vector machines (TSVMs). *Nucl. Acids Res.* **33**, 5799–5808 (2005).

26 Miao, V. *et al.* Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology* **151**, 1507–1523 (2005).

27 Feng, D.- F., Cho, G. & Doolittle, R. F. Determining divergence times with a protein clock: update and reevaluation. *Proc. Natl Acad. Sci. USA* **94**, 13028–13033 (1997).

28 Mahlert, C., Kopp, F., Thirlway, J., Micklefield, J. & Marahiel, M. A. Stereospecific enzymatic transformation of α-ketoglutarate to (2S,3R)-3-methyl glutamate during acidic lipopeptide biosynthesis. *J. Am. Chem. Soc.* **129**, 12011–12018 (2007).

29 Miao, V., Brost, R., Chapple, J., Coëffet-LeGal, M.- F. & Baltz, R. H. The lipopeptide antibiotic A54145 biosynthetic gene cluster from *Streptomyces fradiae*. *J. Ind. Microbiol. Biotechnol.* **33**, 129–140 (2006).

30 Alexander, D. C. *et al.* Development of a genetic system for lipopeptide combinatorial biosynthesis in *Streptomyces fradiae* and heterologous expression of the A54145 biosynthetic gene cluster. *Appl. Environ. Microbiol.* (submitted).

31 Hojati, Z. *et al.* Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem. Biol.* **9**, 1175–1187 (2002).

32 Wittmann, M., Linne, U., Pohlmann, V. & Marahiel, M. A. Role of DptE and DptF in the lipidation reaction of daptomycin. *FEBS J.* **275**, 5343–5354 (2008).

33 Heinzelmann, E. *et al.* An acyl-CoA dehydrogenase involved in the formation of the Δcis3 double bond in the acyl residue of the lipopeptide antibiotic friulimicin in *Actinoplanes friuliensis*. *Microbiology* **151**, 1963–1974 (2005).

34 Donadio, S., Monchiardini, P. & Socio, M. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat. Prod. Rep.* **24**, 1073–1109 (2007).

35 Baltz, R. H. Antibiotic discovery from actinomycetes: will a renaissance follow the decline and fall? *SIM News* **55**, 186–196 (2005).

36 Stanzak, R., Matsushima, P., Baltz, R. H. & Schoner, B. E. DNA homology relationships among erythromycin-producing actinomycetes. *J. Gen. Microbiol.* **136**, 1899–1904 (1990).