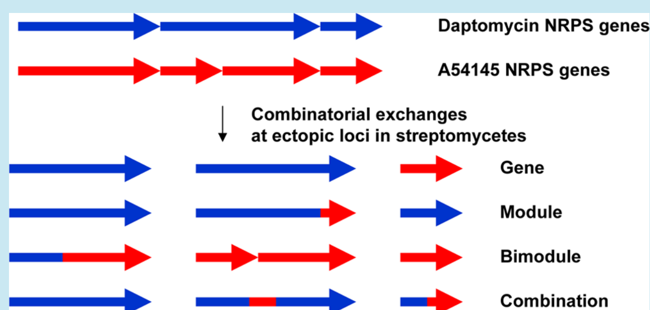
ACS Synthetic Biology
Review
pubs.acs.org/synthbio

Combinatorial Biosynthesis of Cyclic Lipopeptide Antibiotics: A Model for Synthetic Biology To Accelerate the Evolution of Secondary Metabolite Biosynthetic Pathways

Richard H. Baltz*

CognoGen Biotechnology Consulting, 6438 North Olney Street, Indianapolis, Indiana 46220, United States



ABSTRACT: Nonribosomal peptide synthetases (NRPSs) are giant multi-enzymes that carry out sequencial assembly line couplings of amino acids to generate linear or cyclic peptides. NRPSs are composed of repeating enzyme domains with modular organization to activate and couple specific amino acids in a particular order. From a synthetic biology perspective, they can be considered as peptide assembly machines composed of devices to couple fatty acids to L-amino acids, L-amino acids to L-amino acids, and D-amino acids to L-amino acids. The coupling devices are composed of specific parts that contain two or more enzyme domains that can be exchanged combinatorially to generate novel peptide assembly machines to produce novel peptides. The potent lipopeptide antibiotics daptomycin and A54145E have identical cyclic depsipeptide ring structures and stereochemistry but have divergent amino acid sequences. As their biosynthetic gene clusters are derived from an ancient ancestral lipopetide pathway, these lipopeptides provided an attractive model to develop combinatorial biosynthesis to generate antibiotics superior to daptomycin. These studies on combinatorial biosynthesis have helped generate guidelines for the successful assembly of NRPS parts and devices that can be used to generate novel lipopeptide structures and have established a basis for future synthetic biology studies to further develop combinatorial biosynthesis as a robust approach to natural product drug discovery.

KEYWORDS: *A54145, BAC cloning, combinatorial biosynthesis, daptomycin, genetic engineering, ϕBT1 att/int, ϕC31 att/int, Streptomyces cloning hosts*

The goal of synthetic biology is to “improve the process of genetic engineering” and to progress to a future “where the design of genetic systems and the ideosyncrasies of DNA are decoupled, and one can compose living systems by mixing-and-matching genetic parts”.[1] Nature has been mixing and matching genetic parts for at least a billion years while evolving complex secondary metabolite biosynthetic pathways that utilize modular polyketide synthase (PKS) and NRPS multi-enzymes. The products of these pathways have impacted the well-being of mankind as antibiotics, antitumor agents, immune modulators, anthelmintics, and insect control agents. Establishing the methodology for rapid evolution of novel, functional PKS- and NRPS-based biosynthetic pathways could have many important future applications. The goal is to develop synthetic biology processes to direct the evolution of PKS and NRPS multi-enzyme complexes, coupled with accessory processes (e.g., lipidations, amino acid modifications, glycosylations, methylations, and hydroxylations) in highly restricted time and space (a few years in a few scientific laboratories) rather than in natural evolutioary time and space (~billion years and 150 million Km^2 of earth surface land). In other words, the goal is to speed the evolutionary process by about 10^{17}-fold. As natural evolution uses mutation, recombination, and selection for fitness as driving forces, a successful synthetic biology process must undoubtedly incorporate and vastly accelerate these fundamental processes. For accelerated evolution, selection for fitness can be split into two stages: selection for targeted biological activity (fitness I) and selection for robust fermentation production (fitness II). Likewise, mutagenesis can be applied at both stages of selection for fitness: deletion mutagenesis in the first stage and focused base-pair substitution mutagenesis to optimize production titer in the second stage (Figure 1). This sounds like a tall order, but it may be achievable based on current understanding of the “rules” for engineering secondary metabolic pathways that utilize modular NRPS multi-enzyme assembly lines and on the current state of the art on expression of engineered NRPS pathways in streptomycete hosts.

From a synthetic biology perspective (see refs 2 and 3 for terminology), NRPS modular multi-enzymes can be considered to be assembly line peptide coupling machines, composed of parts, devices, and accessory components to optimize the process in microbial expression hosts (chassis). Giessen and

Special Issue: Natural Products

Received: July 26, 2012
Published: August 9, 2012

 © 2012 American Chemical Society

1. Deletion Mutagenesis
2. Multi-domain Recombinations
3. Combinatorial Recombination

Phase I

4. Fermentation Production
5. Product Isolation
6. Selection for Biological Activity

Phase II

7. Focused Mutagenesis
8. Selection for Robust Production

Figure 1. Evolutionary process for synthetic biology to engineer lipopeptide biosynthesis.

Marahiel[4] have pointed out conceptionally how synthetic biology might be used to generate diverse peptides based upon standardization of modular NRPS parts and tailoring enzymes coupled with directed mutagenesis and MS-based screening for enhanced product yield, and Nannemann et al.[5] have reviewed methods for carrying out directed evolution of secondary metabolic biosynthetic genes. Others have reviewed applications of synthetic biology to genome mining and reprogramming natural products biosynthesis.[2,6] In the present review, I focus on what has been learned about the "rules" for successful recombination of NRPS functional domains (parts) to generate functional peptide coupling devices that can be expressed in engineered expression hosts and selected for specific biological activity parameters (fitness I), based upon combinatorial biosynthetic studies on lipopeptide antibiotics related to daptomycin. I also discuss experiments that address focused

A

Daptomycin

A54145 (Factor B1)

B

DptA DptBC DptD

LptA LptB LptC LptD

Figure 2. Structures of daptomycin and A45145 (A) and organization of NRPS subunits (B). Reproduced from Nguyen et al.[23]

dx.doi.org/10.1021/sb3000673 | ACS Synth. Biol. 2014, 3, 748−758

Table 1. Proteins Encoded by A21978C (*dpt*) and A54145 (*lpt*) Gene Clusters

Dpt protein	size (aa)	function	Lpt protein	size (aa)	% identities (aa/aa)
DptA	5830	NRPS	LptA	6292	53 (3175/5920)
DptBC	7338	NRPS	LptB	2148	47 (1014/2156)
		NRPS	LptC	5246	49 (2673/5369)
DptD	2379	NRPS	LptD	2384	53 (1273/2397)
DptE	597	Fatty acyl ligase (AL)	LptEF	732	51 (305/598)
DptF	89	Acyl carrier protein (ACP)		(87)	39 (34/86)
DptG	75	MbtH-like chaperone	LptG	80	58 (40/68)
DptH	271	Editing thioesterase	LptH	264	54 (242/261)
DptI	328	α-ketoglutarate methyltransferase	LptI	352	37 (128/339)
DptJ	245	Tryptophan 2,3-dioxygenase			
		L-Asp9 hydoxylase	LptJ	331	
		L-hAps9 methyltransferase	LptK	262	
		L-Asn3 hydroxylase	LptL	319	
DptM	319	ABC trasporter: ATP-binding	LptM	353	55 (167/300)
DptN	289	ABC transporter: permease	LptN	279	50 (139/276)
DptP	206	Transport/resistance	LptP	206	94 (194/206)
DptR1	341	LuxR family regulatory protein			
DptR2	371	LacI family regulatory protein			
		MerR family regulatory protein	ORF46	312	

mutagenesis to quickly optimize engineered peptide assembly machines for enhanced production (fitness II).

A21978C (Including Daptomycin) and A54145. The cyclic lipodepsipeptide antibiotics A21978C[7] and A54145[8,9] were discovered at Eli Lilly and Company and are biosynthesized by NRPS mechanisms.[10−12] A21978C and A54145 are produced as mixtures of related factors during fermentation by strains of *Streptomyces roseosporus* and *Streptomyces fradiae*, respectively. Both contain 10 amino acids in the cyclic peptide ring and three exocyclic amino acids coupled to long chain fatty acids (Figure 2). The fatty acids can be removed by a deacylase enzyme produced by *Actinoplanes utahensis*,[10,13,14] and the resulting cyclic peptides have been reacylated with different chain length fatty acids. Derivatives with very good antibacterial activities against Gram-positive pathogens were generated,[14−16] leading to the discovery and development of daptomycin.

Daptomycin entered clinical development in the late 1980s, but the trials were discontinued after muscle toxicity was encountered in a dose escalation study.[17] Lilly licensed daptomycin to Cubist Pharmaceuticals in 1997, and Cubist determined in preclinical and clinical studies that muscle toxicity could be minimized by dosing once per day rather than twice per day.[17] After FDA approval, Cubicin (daptomycin for injection) was launched by Cubist for treatment of skin and skin structure infections caused by Gram-positive pathogens[18] and for bacteremia and right-sided endocarditis caused by *Staphylococcus aureus* strains.[19] Surprisingly, daptomycin failed to show non-inferiority versus standard treatment for *Streptococcus pneumoniae* pneumonia,[20] even though it is very potent at killing *S. pneumoniae* in vitro.[21−23] The failure to adequately treat *S. pneumoniae* appeared to be due to inhibition of antibacterial activity by pulmonary surfactant.[24] To expand the spectrum of daptomycin to include community acquired pneumonia, Cubist carried out medicinal chemistry and genetic engineering studies to alter the lipid side chains and amino acid composition of A21978C and A54145.

Rationale for the Development of Combinatorial Biosynthesis of Acidic Lipopeptide Antibiotic Pathways. The genetic engineering of acidic, cyclic lipodepsipeptide antibiotic biosynthetic pathways has several compelling fundamental and practical features: (i) daptomycin is an important clinical antibiotic, and modified derivatives could have additional clinical indications; (ii) daptomycin is composed of a complex tridecapeptide coupled to a long chain fatty acid and is produced by an NRPS mechanism; (iii) other cyclic lipodepsipeptides and cyclic lipopeptides distantly related to daptomycin have antibacterial activities, and some have been in development; (iv) lipopeptide pathways have been evolving for over a billion years,[25] and a finite set of distantly related peptide scaffolds have already been discovered;[10] and (v) the distantly related lipopeptide biosynthetic genes present attractive starting materials for highly accelerated pathway evolution.

A21978C and A54145 Genes Encode Lipopeptides with Identical Peptide Stereochemistry. The original publication of the A21978C structure by Lilly assigned the stereochemistry of the individual 13 amino acids,[7] whereas the stereochemistry of A54145 was not assigned.[8,9] A21978C had D-amino acids assigned at positions 8 and 11 and L-amino acids or Gly residues at the other positions. However, DNA sequence analysis of the NRPS genes by Cubist suggested that A21978C had an additional D-amino acid at position 2, because of the presence of epimerase (E) domains in modules 2, 8, and 11 (Figure 2). This prediction was verified both chemically and biologically: substitution of L-Asn_2 for D-Asn_2 changed the chemical properties and increased the minimal inhibitory concentration (MIC) by 10-fold over that of daptomycin.[11] The NRPS genes involved in the biosynthesis of A54145 also have E domains in modules 2, 8 and 11,[12] and the predicted stereochemistry was confirmed by amino acid enantiomer labeling.[26] The annotation and localization of E domains in the A21978C and A54145 NRPS genes points out the predictive power of DNA sequence analysis to help assign structure and stereochemistry of complex peptide secondary metabolites, even prior to their isolation.[25,27]

Enzyme Subunit, Module, and Domain Organization of the A21978C and A54145 NRPS Genes. The A21978C and A54145 biosynthetic clusters contain evolutionarily related NRPS and accesory genes, as well as some unrelated genes

Table 2. Fatty Acid and Amino Acid Coupling Devices and Deduced Functional Parts of the A21978C[a] and A54145[b] Lipopeptide Biosynthetic Assembly Line Logic

coupling devise	device structure[c]	no. observed	functionally engineered device	functional parts defined	additional comments	ref
FA to L-AA	AL:ACP:C^{III}-A-T^{I}-	1			ACP:C^{III} important; opportunity to exchange FAs	11
	AL-ACP:C^{III}-A-T^{I}-	1			AL-ACP:C^{III} Important; opportunity to exchange FAs	12
L-AA to D-AA	:C^{III}-A-T^{I}-C^{I}-A-T^{II}E-	2	:C^{III}-A-T^{I}**::C^{I}-A-T^{II}**::E-	**-C^{I}-A-T^{II}-**	T^{I}-C^{I} and T^{II}-E functional and linkers flexible	47
	-C^{I}-A-T^{I}-C^{I}-A-T^{II}-E-	1	-C^{I}-A-T^{I}**::C^{I}-A-T^{II}**::E-	**-C^{I}-A-T^{II}-**	″	22
	-C^{I}-A-T^{I}:C^{I}-A-T^{II}-E-	1	-C^{I}-A-T^{I}**::C^{I}-A-T^{II}**::E-	**-C^{I}-A-T^{II}-**	″	23
	-C^{I}-A-T^{I}-C^{I}-A-T^{II}-E-	1	-C^{I}-A-T^{I}**::C^{I}-A-T^{II}E::**C^{II}	**-C^{I}-A-T^{II}E-**	E: converted to E-; E-C^{II} linker flexible	22
Gly to D-AA	-C^{I}-A-T^{I}-C^{I}-A-T^{II}-E:	2	-C^{I}-A-T^{I}**::C^{I}-A-T^{II}**::E:	**-C^{I}-A-T^{II}-**	T^{I}-C^{I} and T^{II}-E: functional and linkers flexible	22, 23
	-C^{I}-A-T^{I}-C^{I}-A-T^{II}-E:	1	-C^{I}-A-T^{I}**::C^{I}-A-T^{II}-E:**C^{II}	**-C^{I}-A-T^{II}E:**	Heterologous E:C^{II} docking works	22
L-AA to L-AA	-C^{II}-A-T^{I}-C^{I}-A-T^{I}-	2			Opportunity to explore AA substitutions	11
	:C^{II}-A-T^{I}-C^{I}-A-T^{I}-	2			″	12
	:C^{II}-A-T^{I}-C^{I}-A-T^{III}-Te	2	**-E:C^{II}-A-T^{I}-C^{I}-A-T^{III}-Te**	**:C^{II}-A-T^{I}-C^{I}-A-T^{III}-Te**	Heterologous E:C^{II} docking and Te cyclization works	43
			:C^{II}-A-T^{I}**::C^{I}-A-T^{III}-Te**	**:C^{II}-A-T^{I}; -C^{I}-A-T^{III}-Te**	T^{I}-C^{I} linker flexible and module exchanges work	47
			:C^{II}-A-T^{I}**::C^{I}-A::**T^{III}-Te	**:C^{II}-A-T^{I}; -C^{I}-A-; -T^{III}-Te**	T^{I}-C^{I} and A-T^{III} linkers flexible	47
L-AA to Gly/Sar	-C^{I}-A-T^{I}-C^{I}-A-T^{II}:	1			Opportunity to exchange for Thr_4	11
	-C-A-T^{I}-C^{I}-A-M-T^{II}:	1			″	12
	-C^{II}-A-T^{I}-C^{I}-A-T^{I}-	2			Asp_9 required for Ca^{2+} binding[d]	11, 12
Gly/Sar to L-AA	-C^{I}-A-M-T^{II}:C^{II}-A-T^{I}-	1	-C^{I}-A**::**T^{II}:C^{II}-A-T^{I}-	**-M-**	M can be deleted and remaining module is functional	42
	-C^{I}-A-T^{II}:C^{II}-A-T^{I}-	1			Opportunity to insert M in Gly and other modules	11
D-AA to L-AA	-C^{I}-A-T^{II}-E-C^{II}-A-T^{I}-	3			Opportunity to explore AA substitutions	11, 12
	:C^{I}-A-T^{II}-E-C^{II}-A-T^{I}-	1			″	12
	-C^{I}-A-T^{II}-E:C^{II}-A-T^{I}-	2			″	11, 12

[a]A21978C logic: Fatty acid Trp D-Asn Asp Thr Gly Orn Asp D-Ala Asp Gly D-Ser 3mGlu Kyn; AL:ACP:C^{III}-A-T^{I}-C^{I}-A-T^{II}E-C^{II}-A-T^{I}-C^{I}-A-T^{I}-C^{I}-A-T^{II}:C^{II}-A-T^{I}-C^{I}-A-T^{I}-C^{I}-A-T^{II}-E-C^{II}-A-T^{I}-C^{I}-A-T^{I}-C^{I}-A-T^{II}-E:C^{II}-A-T^{I}-C^{I}-A-T^{III}-Te. [b]A54145 logic: Fatty acid Trp D-Glu hAsn Thr Sar Ala Asp D-Lys moAsp Gly D-Asn 3mGlu Ile; AL-ACP:C^{III}-A-T^{I}-C^{I}-A-T^{II}-E-C^{II}-A-T^{I}-C-A-T^{I}-C^{I}-A-M-T^{II}:C^{II}-A-T^{I}-C^{I}-A-T^{I}:C^{I}-A-T^{II}-E-C^{II}-A-T^{I}-C^{I}-A-T^{I}-C^{I}-A-T^{II}-E:C^{II}-A-T^{I}-C^{I}-A-T^{III}-Te. [c]Dashes indicated interpeptide linkers, and colons indicate protein:protein docking. [d]Experiments were not carried out in order to preserve the conserved Ca^{2+} binding site, Asp-Xxx-Asp-Gly.[33]

(Table 1). The NRPS genes that encode the biosynthesis of the peptide portions of A21978C and A54145 are composed of modules containing three or four enzyme domains, including condensation (C) domains for coupling fatty acids to amino acids or amino acids to amino acids; adenylation (A) domains for binding and activation of specific amino acids; thiolation (T) or peptidyl carrier protein (PCP) domains for tethering the amino acids and the growing peptide chain and for facilitation of interactions with multiple domains during peptide assembly; epimerase (E) domains for converting L-amino acids to D-amino acids; a methyltransferase (M) domain for *N*-methylation of an amino acid; and thioesterase (Te) domains for cyclization and release of the completed lipopeptides. Typical modules observed in NRPS gene clusters include AT, CAT, CATE, CAMT, and CATTe multidomains,[28] but the daptomycin and A54145 pathways lack AT starter modules because they have special CAT starter modules that couple long chain fatty acids to the first amino acid (Trp_1). Simplistically, the A21978C NRPS genes encode nine CAT, three CATE, and one CATTe modules, whereas the A54145 NRPS genes encode eight CAT, three CATE, one CAMT, and one CATTe modules. However, this level of module description is inadequate for combinatorial biosynthesis and future synthetic biology studies because additional functionally critical information is encoded in the individual domains relevant to amino acid specificity[29,30] and to productive domain−domain interactions.[28,31,32]

The information on amino acid binding and activation specificity is encoded in the A domains. Each A domain has a well-defined 10-amino-acid binding pocket that defines binding specificity.[29,30] The A21978C NRPS genes have 10 different types of A domain, including three Asp, two Gly, and one each for Asn, Glu, Ala, Thr, Trp, ornithine (Orn), and kynurenine (Kyn). The rare (and unique) Orn and Kyn A domain binding pocket codes (DTWDMGYVDK and DAWTTTGVGK, respectively) have been used in conjunction with other analyses to predict a daptomycin-like gene cluster in *Saccharomonospora viridis*.[25] The A54145 NRPS genes have nine types of A domain, including two each for Asp, Gly, Asn, Glu and one each for Ala, Thr, Trp, Lys and Ile (Figure 2). There are only 12 types of amino acid binding pockets in A domains used for A21978C or A54145 biosynthesis, and seven of these are used by both. This emphasizes the phylogenetic relatedness of these two pathways.[33] Some of the amino acids undergo enzymatic modifications (one in A21978C and four in A54145), which further differentiates the two lipopeptide structures. The

dx.doi.org/10.1021/sb3000673 | ACS Synth. Biol. 2014, 3, 748−758

Table 3. Examples of Genetic Changes for Combinatorial Biosynthesis of Lipopeptides in *S. roseosporus* and *S. fradiae*

lipopeptide core	type of change	actual change[a]	amino acid change	relative yield (%)	ref
A21978C (Daptomycin)	none			100	43
	gene deletion	Δ*dptI*	Glu_{12} for $3mGlu_{12}$	50	22
	gene replacement	*dptBC*:***lptD***	Ile_{13} for Kyn_{13}	25	43, 45
	gene replacement	*dptBC*:***cdaPS3***	Trp_{13} for Kyn_{13}	50	43, 45
	dimodule fusion	$C^{II}AT^{I}::\mathbf{C^{I}AT^{III}Te}$	Ile_{13} for Kyn_{13}	86	47
	dimodule fusion	$C^{II}AT^{I}::\mathbf{C^{I}AT^{III}Te}$	Trp_{13} for Kyn_{13}	134	47
	tridomain substitution	$C^{II}AT^{I}::\mathbf{C^{I}AT^{II}}::Te$	Asn_{11} for Kyn_{13}	0	47
	didomain substitution	$C^{II}AT^{I}::\mathbf{C^{I}A}::T^{III}Te$	Asn_{11} for Kyn_{13}	~43	47
	tridomain substitution	$-C^{I}AT^{I}::\mathbf{C^{I}AT^{II}}::E:$	D-Ala_8 for D-Ser_{11}	50	22
	tridomain substitution	$-C^{I}AT^{I}::\mathbf{C^{I}AT^{II}}::E:$	D-Asn_{11} for D-Ser_{11}	17	22
	tridomain substitution	$-C^{I}AT^{I}::\mathbf{C^{I}AT^{II}}::E-$	D-Ser_{11} for D-Ala_8	18	22
	tridomain substitution	$-C^{I}AT^{I}::\mathbf{C^{I}AT^{II}}::E-$	D-Asn_{11} for D-Ala_8	10	22
A54145E	none			100	48
	gene deletion	ΔlptI	Glu_{12} for $3mGlu_{12}$	140	48
	gene deletion	ΔlptJ	Asp_9 for $moAsp_9$	26	49
	gene deletion	ΔlptK	$hAsp_9$ for $moAsp_9$	50	49
	gene deletion	ΔlptL	Asn_3 for $hAsn_3$	27	49
	four gene deletion	ΔlptIJKL	Asn_3-Asp_9-Glu_{12} for $hAsn_3$-$moAsp_9$-$3mGlu_{12}$	53	49
	didomain substitution	$C^{II}AT^{I}::\mathbf{C^{I}A}::T^{III}Te$	Kyn_{13} for Ile_{13}	~40	42
	tridomain substitution	$C^{III}AT^{I}::\mathbf{C^{I}AT^{II}}::E-$	D-Asn_2 for D-Glu_2	5	23
	trimodule substitution	$C^{III}AT^{I}::\mathbf{C^{I}AT^{II}E\text{-}C^{I}AT^{I}\text{-}C^{I}AT^{I}}::C^{I}AMT^{I}$	D-Asn_2-Asp_3 for D-Glu_2-$hAsn_3$	4	23
	tridomain substitution	$-C^{I}AT^{I}::\mathbf{C^{I}AT^{II}}::E:$	D-Ala_8 for D-Asn_{11}	3	23
	tridomain substitution	$-C^{I}AT^{I}::\mathbf{C^{I}AT^{II}}::E:$	D-Ser_{11} for D-Asn_{11}	16	23
	tridomain substitution	$-C^{I}AT^{I}::\mathbf{C^{I}AT^{II}}::E-$	D-Ala_8 for D-Lys_8	3	23
	tridomain substitution	$-C^{I}AT^{I}::\mathbf{C^{I}AT^{II}}::E-$	D-Ser_{11} for D-Lys_8	9	23

[a] :, inter-protein docking site; ::, inter-domain fusion point; -, inter-module linker; **bold** indicates heterologous exchanges.

specificity of A domains is one key design component for combinatorial biosynthesis, and the A21978C and A54145 pathways use only a small subset of the possible A domains observed in NRPS genes.[34]

In addition to A domain specificity, other specificities are encoded in C, T, and Te domains. Three types of C domain were observed in the A21978C and A54145 gene clusters, based initially on phylogeneic analyses: (i) for coupling L-amino acids to L-amino acids (C^{I} or $^{L}C_{L}$); (ii) for coupling D-amino acids to L-amino acids (C^{II} or $^{D}C_{L}$), and (iii) for coupling fatty acids to L-amino acids (C^{III} or $^{F}C_{L}$).[11,12,35] There are no special C domains for coupling L-amino acids to D-amino acids because the epimerization of L-amino acids follows peptide coupling.[36−38]

There are also three types of T domain implied from biochemical studies[31,32,39−41] and supported by BLASTP analyses (R. H. Baltz, unpublished). These T domain subtypes interact with upstream C and A domains and (i) downstream C domains (T^{I} or T_{C}), (ii) downstream E and C domains (T^{II} or T_{E}), and (iii) downstream Te domains (T^{III} or T_{Te}).

In addition to the NRPS domains, two other domains participate in the activation and coupling of the fatty acids during biosynthesis of A21978C and A54145. These are fatty acyl ligase (AL) and acyl carrier protein (ACP) domains that have functions similar to those of A and T domains but process fatty acids rather than amino acids. The enzymatic assembly line logic for the biosynthesis of the A21978C linear lipopeptide, fatty acyl-Trp-D-Asn-Asp-Thr-Gly-Orn-Asp-D-Ala-Asp-Gly-D-Ser-3mGlu-Kyn, can be depicted as AL:ACP:C^{III}-A-T^{I}-C^{I}-A-T^{II}E-C^{II}-A-T^{I}-C^{I}-A-T^{I}-C^{I}-A-T^{II}:C^{II}-A-T^{I}-C^{I}-A-T^{I}-C^{I}-A-T^{II}-E-C^{II}-A-T^{I}-C^{I}-A-T^{I}-C^{I}-A-T^{II}-E:C^{II}-A-T^{I}-C^{I}-A-T^{III}-Te, where colons indicate protein:protein docking interactions and dashes indicate interdomain linker peptides. Similarly, the A54145 biosynthetic logic for the assembly of fatty acyl-Trp-D-Glu-hAsn-Thr-Sar-Ala-Asp-D-Lys-moAsp-Gly-D-Asn-3mGlu-Ile is AL-ACP:C^{III}-A-T^{I}-C^{I}-A-T^{II}-E-C^{II}-A-T^{I}-C-A-T^{I}-C^{I}-A-M-T^{II}:C^{II}-A-T^{I}-C^{I}-A-T^{I}:C^{I}-A-T^{II}-E-C^{II}-A-T^{I}-C^{I}-A-T^{I}-C^{I}-A-T^{II}-E:C^{II}-A-T^{I}-C^{I}-A-T^{III}-Te. Modules 5 for A21978C and A54145 pathways are unusual in that they have T^{II} domains but no E domains, as they process the achiral Gly and sarcosine (Sar), respectively. Furthermore, module 6 in both cases begins with a C^{II} domain. The T^{II} domains in modules 5 of DptA and LptA have a ~45 amino acid docking sequence appended to the carboxy terminus to interact with docking sequences in DptBC and LptB, respectively. This T^{II}-dock configuration is unique to the A21978C and A54145 pathways based upon BLASTP analysis

Table 4. Lipopeptide Antibiotics Generated by Combinatorial Biosynthesis

	amino acid at position listed[b]									*S. aureus* MIC (μg/mL)[c]		
compound[a]	2	3	5	6	8	9	11	12	13	− surf	+ surf	ratio (+/−)
daptomycin	D-Asn	Asp	Gly	Orn	D-Ala	Asp	D-Ser	3mGlu	Kyn	0.5	64	128
CB-181,220	D-Asn	Asp	Gly	Orn	D-Ala	Asp	D-Ser	3mGlu	Kyn	0.5	64	128
CB-182,098	D-Asn	Asp	Gly	Orn	D-Ala	Asp	D-Ser	3mGlu	***Trp***	1	32	32
CB-182,107	D-Asn	Asp	Gly	Orn	D-Ala	Asp	D-Ser	3mGlu	***Ile***	2	8	4
CB-182,106	D-Asn	Asp	Gly	Orn	D-Ala	Asp	D-Ser	3mGlu	***Val***	4	8	2
A21978C1(Asn13)	D-Asn	Asp	Gly	Orn	D-Ala	Asp	D-Ser	3mGlu	***Asn***	128	ND	ND
CB-182,130	D-Asn	Asp	Gly	Orn	D-Ala	Asp	D-Ser	***Glu***	Kyn	8	16	2
CB-182,166	D-Asn	Asp	Gly	Orn	D-Ala	Asp	***D-Ala***	3mGlu	Kyn	1	16	16
CB-182,290	D-Asn	Asp	Gly	Orn	D-Ala	Asp	***D-Asn***	3mGlu	Kyn	1	16	16
CB-182,123	D-Asn	Asp	Gly	Orn	***D-Ser***	Asp	D-Ser	3mGlu	Kyn	1	32	32
CB-182,257	D-Asn	Asp	Gly	Orn	***D-Asn***	Asp	D-Ser	3mGlu	Kyn	8	ND	ND
CB-182,286	D-Asn	Asp	Gly	Orn	D-Ala	Asp	***D-Asn***	3mGlu	***Ile***	4	ND	ND
CB-182,251	D-Asn	Asp	Gly	Orn	D-Ala	Asp	***D-Asn***	***Glu***	Kyn	32	ND	ND
CB-182,263	D-Asn	Asp	Gly	Orn	***D-Asn***	Asp	D-Ser	3mGlu	***Ile***	16	ND	ND
CB-182,269	D-Asn	Asp	Gly	Orn	***D-Asn***	Asp	D-Ser	***Glu***	Kyn	128	ND	ND
CB-182,296	D-Asn	Asp	Gly	Orn	***D-Lys***	Asp	***D-Asn***	3mGlu	Kyn	1	32	32
A54145E	D-Glu	hAsn	Sar	Ala	D-Lys	moAsp	D-Asn	3mGlu	Ile	1	32	32
A54145D	D-Glu	hAsn	Sar	Ala	D-Lys	moAsp	D-Asn	***Glu***	Ile	2	4	2
CB-182,548	D-Glu	hAsn	Sar	Ala	D-Lys	moAsp	***D-Ala***	3mGlu	Ile	1	16	16
CB-182,332	D-Glu	hAsn	Sar	Ala	D-Lys	moAsp	***D-Ser***	3mGlu	Ile	2	16	8
CB-182,443	D-Glu	hAsn	Sar	Ala	D-Lys	***Asp***	D-Asn	3mGlu	Ile	2	4	2
CB-182,571	D-Glu	hAsn	Sar	Ala	***D-Ala***	moAsp	D-Asn	3mGlu	Ile	1	32	32
CB-182,549	D-Glu	hAsn	Sar	Ala	***D-Ser***	moAsp	D-Asn	3mGlu	Ile	1	16	16
CB-182,510	D-Glu	hAsn	Sar	Ala	***D-Asn***	moAsp	D-Asn	3mGlu	Ile	8	64	8
CB-182,363	D-Glu	***Asn***	Sar	Ala	D-Lys	moAsp	D-Asn	3mGlu	Ile	2	16	8
CB-182,575	***D-Asn***	hAsn	Sar	Ala	D-Lys	moAsp	D-Asn	3mGlu	Ile	2	4	2
CB-183,296	D-Glu	hAsn	Sar	Ala	D-Lys	moAsp	D-Asn	***Glu***	***Kyn***	1	2	2
CB-182,509	D-Glu	hAsn	Sar	Ala	D-Lys	moAsp	***D-Ala***	***Glu***	Ile	8	16	2
CB-182,336	D-Glu	hAsn	Sar	Ala	D-Lys	moAsp	***D-Ser***	***Glu***	Ile	64	128	2
CB-182,350	D-Glu	hAsn	Sar	Ala	D-Lys	***hAsp***	D-Asn	***Glu***	Ile	8	16	2
CB-182,333	D-Glu	hAsn	Sar	Ala	D-Lys	***Asp***	D-Asn	***Glu***	Ile	32	64	2
CB-182,567	D-Glu	hAsn	Sar	Ala	***D-Ala***	moAsp	D-Asn	***Glu***	Ile	4	8	2
CB-182,532	D-Glu	hAsn	Sar	Ala	***D-Ser***	moAsp	D-Asn	***Glu***	Ile	8	8	1
CB-182,531	D-Glu	hAsn	Sar	Ala	***D-Asn***	moAsp	D-Asn	***Glu***	Ile	16	16	1
CB-182,391	D-Glu	hAsn	***Gly***	Ala	D-Lys	moAsp	D-Asn	***Glu***	Ile	16	32	2
CB-182,325	D-Glu	***Asn***	Sar	Ala	D-Lys	moAsp	D-Asn	***Glu***	Ile	32	32	1
CB-182,444	***D-Asn***	hAsn	Sar	Ala	D-Lys	moAsp	D-Asn	***Glu***	Ile	8	8	1
CB-182,597	D-Glu	***Asn***	Sar	Ala	D-Lys	***hAsp***	D-Asn	3mGlu	Ile	1	16	16
CB-182,390	D-Glu	***Asn***	Sar	Ala	D-Lys	***Asp***	D-Asn	3mGlu	Ile	2	2	1
CB-182,561	***D-Asn***	***Asp***	Sar	Ala	D-Lys	moAsp	D-Asn	3mGlu	Ile	1	2	2
CB-182,349	D-Glu	***Asn***	Sar	Ala	D-Lys	***hAsp***	D-Asn	***Glu***	Ile	32	64	2
CB-182,348	D-Glu	***Asn***	Sar	Ala	D-Lys	***Asp***	D-Asn	***Glu***	Ile	16	32	2
CB-182,560	***D-Asn***	***Asp***	Sar	Ala	D-Lys	moAsp	D-Asn	***Glu***	Ile	8	16	2

[a]Daptomycin has an *N*-decanoyl side chain. All others compounds have *anteiso*-undecanoyl side chains. [b]Amino acid changes are shown in ***bold italics***. [c]MICs against *S. aureus* 42 were determined in the presence and absence of 1% bovine surfactant (surf). Table reproduced from Baltz et al.[42]

and may have evolved from an ancestral module that processed a D-amino acid. Docking sequences are commonly associated with carboxy terminal E domains that interact with docking domains at the amino terminal of C^{II} domains. It is noteworthy that all other cyclic lipodepsipeptides and lipopeptides distantly related to daptomycin (e.g., CDA, friulimicins, amphomycins, and laspartamycins) have D-amino acids at positions comparable to those of Gly_5 and Sar_5.[10] Other subtle differences in biosynthetic logic are discussed below.

The enzymatic functions required for the individual couplings (coupling devices) are shown in Table 2. It is noteworthy that A21978C and A54145 use very similar coupling devices. All told, 17 different types of coupling device are used. If we remove three subtle differences between the two pathways, the number of devices is reduced to 12, 11 of which are used once and the other twice in each pathway. The three differences are (i) *dptE* and *dptF* have a fused counterpart *lptEF* encoding AL and ACP or AL-ACP, respectively; (ii) the *dptBC* gene has split gene counterparts *lptB* and *lptC* that have been successfully fused in genetic engineering experiments;[23] and (iii) *lptA* module 5 has an M domain to incorporate Sar that can be deleted to incorporate Gly.[42] Table 2 shows which parts of the coupling devices have been successfully engineered, and Table 3 shows some of the specific changes and product yields from strains generated by combinatorial biosynthesis.[22,23,26,35,42−49] Ten of the 17 devices have been engineered

dx.doi.org/10.1021/sb3000673 | ACS Synth. Biol. 2014, 3, 748−758

to give novel lipopeptides. Nine functional parts were validated in these studies. These parts have maintained homologous CA, CAT, CATE, and TTe functions and have defined flexible A-T, T-C, T-E, and E-C linkers as suitable sites for recombination. It remains to be seen if A domains can be productively exchanged, but biochemical data suggest that C-A linkers will not provide flexible sites for recombination,[31,32,50] and that CA bidomains may be evolutionarily conserved units. Nonetheless, A domain exchanges need to be further tested rigorously, because if they work, they could provide synthetic biology with many additional options for directed evolution.

Genes Involved in Lipoinitiation. In addition to the engineering of NRPS genes, combinatorial biosynthesis encompasses modifications of the lipid side chains of A21978C and A54145. For instance, the A21978C natural factors have three dominant long chain fatty acids, *anteiso*-undecanoyl, *iso*-dodecanyl, and *anteiso*-tridecanoyl,[7] whereas the A54145 factors have predominantly *n*-decanoyl, *iso*-decanoyl, and *anteiso*-undecanoyl side chains that are (on average) two carbon atoms shorter than those for A21978C.[8,9] The fatty acid chain length is not associated with the dominant fatty acids produced by the host organism because the same three dominant fatty acids for A21978C and A54145 were observed when the gene clusters are expressed from BAC vectors in heterologous hosts.[48,51] This suggested that the fatty acid chain length may be associated with the enzyme(s) involved in activating the long chain fatty acids for coupling with the N-terminal Trp_1. In the A21978C gene cluster, there are two genes just upstream of the NRPS genes, *dptE* and *dptF*, that are involved in coupling the long chain fatty acids to Trp_1 to initiate biosynthesis of A21978C. The *dptE* gene encodes an AL, and *dptF* encodes an ACP.[11,52] The A54145 counterpart is a fused AL-ACP encoded by *lptEF* (Table 1).[12] Biochemical studies indicate that the specificity for the fatty acid precursor chain lengths for A21978C and CDA resides in the ALs and the C^{III} domains involved in lipoinitiation.[52,53] It is noteworthy that when A21978C and CDA were expressed in parallel in a *Streptomyces lividans* recombinant containing the A21978C gene cluster inserted in the *ϕ*C31 *attB* site, only the normal A21978C and CDA factors were produced,[51] consistent with the AL and ACP specificities identified in biochemical studies. The distribution of lipid side chains for A21978C and A54145 can be modulated to a certain extent by feeding specific fatty acids or amino acids to enrich for specific factors,[10,54] and this approach has been useful in combinatorial biosynthetic studies to enrich recombinant fermentations for desired factors[23,48,49] Homologues of *dptE* and *dptF* have been observed in the friulimicin and laspartamycin gene clusters in *Actinoplanes friuliensis*[55] and *Streptomyces viridochromogenes*,[56] which use long chain unsaturated fatty acids to initiate lipopeptide biosynthesis, so it is conceivable that the genes encoding ALs and ACPs might be exchanged to add different lipid side chains to the A21978C and A54145 engineered peptides. However, the natural lipid side chains on A21978C and A54145 can be readily removed by incubation with a deacylase from *A. utahensis*,[10,13,14,57] and alternative side chains substituted by chemical acylation.[14,15] As such, genetic manipulation of lipoinitiation has not been a priority for combinatorial biosynthesis directed at drug discovery at Cubist, but it could become important for manufacturing where chemical modification of engineered tridecapeptides may not be economically feasible (e.g., daptomycin is manufactured by feeding decanoic acid during fermentation).[10]

Genes Involved in Amino Acid Modifications. Several amino acids in the assembly of A21978C and A54145 are produced by enzymatic modifications in *S. roseosporus* and *S. fradiae*. A21978C and A54145 have 3-methyl-glutamic acid (3mGlu) at position 12, which is generated by methylation of *α*-ketoglutarate followed by transamination.[58] Disruption of the *dptI* gene caused the production of A21978C derivatives containing Glu_{12} at titers of ∼50% of control,[44] whereas disruption of the *lptI* gene caused the production of A54145 derivatives containing Glu_{12} at titers of ∼140% of control.[48]

A54145 has three additional modified amino acids, hydroxy-Asn ($hAsn_3$), Sar (*N*-methyl-Gly), and methoxy-Asp ($moAsp_9$). The methyltransferase that *N*-methylates Gly_5 is imbedded in module five as CAMT. Deletion of the M domain yielded a functional CAT domain that incorporated Gly instead of Sar, causing a 10-fold increase in MIC (Table 4).[42] The *lptL* gene encodes the L-Asn_3 hydroxylase; *lptJ* encodes the L-Asp_9 hydroxylase; and *lptK* encodes the L-$hAsp_9$ methyltransferase.[49] All combinations of deletions of *lpI*, *lptJ*, *lptK*, and *lptL* were made, and all produced the predicted lipopeptides in yields ranging from 32 to 360 mg/L. All of the compounds displayed antibacterial activities against *S. aureus*, with MICs ranging from 2 to 32 *μ*g/mL in the absence of bovine surfactant and 2 to 64 *μ*g/mL in the presence of 1% surfactant.[49] The best compound (CB-182,390), deleted for *lptJKL*, had an MIC of 2 in the presence of surfactant and was not toxic in mice even though it retained $3mGlu_{12}$ (Table 4).

Protein−Protein Docking Interactions. Functional interactions between lipopeptide NRPS multi-enzymes are facilitated by interpeptide docking sequences.[10,28,33] In principle, conservation of related docking sequences over evolutionary time might facilitate the testing of different combinations of NRPS subunits to generate novel peptide scaffolds. The A21978C, A54145, and CDA NRPS subunit organization afforded a relatively simple test system to determine if heterologous gene exchange involving NRPS subunits could yield functional NRPS multi-enzymes that produce novel lipopeptides. A21978C, A54145, and CDA have related genes that encode the incorporation of the terminal evolutionarily related dipeptides: 3mGlu-Kyn, 3mGlu-Ile, and 3mGlu-Trp, respectively.[10,33] These NRPS proteins (DptD, LptD, and CDA-PSIII, respectively) have N-terminal docking domains to interact with C-terminal docking domains in the upstream DptBC, LptC, and CDA-PSII proteins.[33] It is noteworthy that a *dptD* deletion mutant of *S. roseosporus* could be complemented by *lptD* and *cdaPSIII* driven by the *ermE** promoter, and recombinants produced A21978C derivatives containing Ile_{13} and Trp_{13}, respectively, at 25% and 50% of control yields.[43,45] This indicated that the docking domains had retained similar protein−protein interaction specificities over evolutionary time and that the Te domains from LptD and CDA-PSIII could process heterologous lipopeptides distantly related to A54145 and CDA. The same hybrid molecules were also generated by module exchanges, and the lipopeptide yields were 86−134% of control[47] (Table 3), thus emphasizing the value in maintaining homologous protein−protein docking interactions and further confirming the broad substrate specificity of the Te domains.

MbtH-like Chaperones. The daptomycin and A54145 gene clusters have genes (*dptG* and *lptG*, respectively) just downstream of the NRPS genes that encode MbtH-like proteins.[11,12] *mbtH*-like genes are located in most commercially important and other characterized NRPS gene clusters.[59] Their

gene products are non-enzymatic and function as chaperones. MbtH homologues enhance the solubility and function of some NRPS components when expressed in heterologous hosts, enhance adenylation reactions in some but not all cases, and are required for optimal production of NRPs in some cases,[59−62] but their specific roles in A21978C and A54145 assembly are unknown. Deletion of *dptG* in *S. roseosporus* caused a 50% reduction in daptomycin yield.[44] *S. roseosporus* has five *dptG* paralogs associated with other NRPS pathways,[59] so it is not clear if any of the residual A21978C production in the *dptG* mutant can be attributed to cross-pathway complementation, which has been demonstrated for aminocoumarin biosynthesis.[59,60] These *mbtH*-like genes encode critical chaperone devices that need to be maintained in expression hosts. As a design strategy, *dptG* and *lptG* were left intact in the *S. roseosporus* and *S. fradiae* cloning hosts to facilitate the production of hybrid molecules. In view of recent reports on the biological functions of MbtH-like proteins, inclusion of *mbtH*-like genes from both pathways might facilitate enhanced production of some hybrid lipopeptides derived from A21978C and A54145 parts and devices.

Regulatory, Editing, Resistance, and Transport Devices. Robust production of complex secondary metabolites produced by NRPS mechanisms is generally tuned by regulatory proteins, editing thioesterases, self-resistance mechanisms, and transport systems. The regulatory mechanisms for secondary metabolite bisynthesis can be complex[63] and are important for chassis design. In some cases, complex regulatory circuits have been simplified by overexpressing the ultimate positive regulatory gene or by disrupting a key negative regulatory gene.[64] The A21978C gene cluster has genes that may provide these functions: *dptR1* and *dptR2* for regulation, *dptMNP* for resistance/transport, and *dptH* for editing[11,33] (Table 1). The A54145 gene cluster has homologues to *dptMNP* (*lptMNP*) and *dptH* (*lptH*), but not to *dptR1* and *dptR2*, but has a candidate regulatory gene (ORF46) just downstream of the biosynthetic genes[12,33] (Table 1). The functions of these candidate "device" genes have not been fully characterized, so they were left intact in the *S. roseosporus* and *S. fradiae* expression hosts. It is likely that these expression devices as well as MbtH-like chaperones can be genetically manipulated to increase the production of engineered lipopeptides.

Microbial Factories (Chassis) and Genetic Toolkits (Devises) for Combinatorial Biosynthesis. The daptomycin and A54145 NRPS genes are transcribed as giant single mRNAs, and the individual genes have overlapping translational stop and start sites.[11,12,45] To carry out combinatorial biosynthesis, it was desirable to break up the natural overlapping linkage relationships and to express the individual genes from different sites in the chromosome. By using this approach, the individual NRPS (and other genes) could be engineered separately and then reintroduced into the expression hosts combinatorially. To accomplish this, (i) the A21978C and A54145 gene clusters were cloned in BAC vectors that could be engineered in *Escherichia coli* by λ-Red-mediated recombination; (ii) sets of mutants deleted for one or more A21978C and A54145 biosynthetic genes were generated; (iii) vectors were developed for genetic manipulation in *E. coli*, followed by conjugal transfer into *S. roseosporus* and *S. fradiae* expression hosts and site-specific integration into bacteriophage ϕC31 of ϕBT1 *attB* sites or into the IS*117 att* site; (iv) engineered genes were expressed under the control of the strong, constitutive *ermE** promoter; (v) fermentations were carried out, and novel compounds were confirmed by isolation and mass spectrometry and evaluated for microbiological activities.[11,12,22,23,26,35,42−49] Reconstruction experiments confirmed that the NRPS genes could be expressed from the *ermE** promoter at ectopic loci, giving lipopeptide yields comparable to controls.[45,48] This was significant because it demonstrated that constitutive expression of giant NRPS proteins during the early growth phase was not problematic, that the individual NRPS genes could be expressed from *attB* sites separated from the main NRPS clusters by multiple megabases,[65] and that the NRPS multi-enzymes can assemble efficiently without sequencial translation from a single mRNA in a cellular subspace for antibiotic assembly.

Combinatorial Biosynthesis. In early studies at Cubist on combinatorial biosynthesis, attempts were made to transplant A domains without success (unpublished data). There is now structural evidence suggesting that CA bidomains may have co-evolved as functional units,[31,32,50] which may explain the lack of success. Subsequent studies focused primarily on exchanges that kept the CA or CAT domains intact while making domain fusions exclusively at interdomain linker regions. Table 3 summarizes several examples of successful engineering and one nonproductive example to illustrate what happens if the wrong type of T domain is used. By using combinations of gene deletions, gene exchanges, bi- and tridomain exchanges, whole module exchanges, and multiple module exchanges, over 40 novel lipopeptide antibiotics were produced in sufficient quantities for facile isolation, confirmation of structure, and analysis of antibacterial activities (Table 4). (Because of the natural combinatorial nature of the lipidation reactions, over 120 novel lipopeptides were generated in these studies, but only the most abundant compound from each fermentation was characterized in most cases.) Importantly, some of the compounds displayed the desired properties of antibacterial activities similar to those of daptomycin, minimal or no inhibition by surfactant, and low mammalian toxicity. These studies demonstrated that there are at least three ways to achieve the desired outcome, and each of the molecules retained most of the core peptide structure of A54145E, which has high mammalian toxicity: (i) exchange Glu_{12}-Kyn_{13} for $3mGlu_{12}$-Ile_{13}, exchange Asn_3-Asp_9 for $hAsn_3$-$moAsp_9$, and exchange D-Asn_2-Asp_3 for D-Glu_2-$hAsn_3$. The latter two compounds retain 3mGlu, indicating that 3mGlu is associated with mammalian toxicity in A54145E only in the context of other amino acids (e.g., modified amino acids at positions 3 and 9). Substituting the daptomycin amino acids at positions 2 and 3 also mitigated mammalian toxicity, perhaps in part by eliminating $hAsn_3$. All of these compounds had activity in a mouse model for *S. pneumoniae* pneumonia, but none was as active as vancomycin. Other higher order combinations of the three pairwise combinations that yielded the most active compounds have not been generated, so it may be possible to further improve the activity. Also, it would be interesting to substitute Sar_5 for Gly_5 in A21978C to see if it has an antibacterial enhancing effect as observed in A54145. In addition, only a very minor set of amino acid substitutions relative to the over 500 non-proteinogenic amino acid building blocks used by NRPS enzymes[34] have been sampled, so there is an untapped opportunity to explore much more chemical space by combinatorial biosynthesis around the core structures of daptomycin and A54145.

Prospects for Mutagenesis and Selection for Robust Production. To progress a lipopeptide synthetic biology project to commercial scale, it is imperative that the fermentation yield can be rapidly improved from the usually suboptimal production afforded by engineered biosynthetic pathways (fitness II in the evolutionary scheme; Figure 1). There is one key example that demonstrates that rapid yield enhancement is indeed feasible. Fischback et al.[66] cloned the andrimid biosynthetic genes from *Pantoea agglomerans* in *E. coli*. They deleted the A domain for Val and inserted an A domain for 2-aminobutyrate from *Streptomyces* sp. RK95-74, and the recombinant produced andrimid at 32-fold lower yield than the control. They carried out three rounds of error-prone PCR on the AT bidomain and improved the yield 10-fold.

The mutagenized AT bidomain had nine amino acid substitutions distributed throughout the A domain. They also substituted an A domain for Ile from *Bacillus licheniformis* for the Val A domain, and the recombinant produced 7-fold less product. One round of error-prone PCR caused two amino acid substitutions associated with a 4.5-fold increase in production. These studies demonstrated that a standard technique of error-prone PCR can rapidly improve the productivity of a hybrid NRPS by generating a small number of amino acid substitutions in the A domain. This and other site-directed mutagenesis approaches should be generally applicable to enhance the productivity of hybrid NRPS pathways generated by synthetic biology approaches.

■ DISCUSSION AND CONCLUDING COMMENTS

From the work on the development of molecular genetic tools, expression hosts, and engineering design rules, it is apparent that complex lipopeptide biosynthetic pathways can be redesigned to produce novel molecules related to daptomycin and A54145. The product titers are variable, ranging from 3% to >100% of controls for single gene deletions or for bi- or tridomain substitutions. Product titers are generally reduced in recombinants containing two or more changes. These studies have validated the first part of the synthetic biology paradigm of accelerated evolution: Deletion mutagenesis and recombination followed by selection for clinical utility works well. To commercialize the process, the second half of the accelerated evolution for fitness II (mutagenesis and selection for robust production) needs to be appended to the process, but methods are available to accomplish this.

Some design rules have been deduced from the structures of fatty acid and amino acid coupling devices; from the locations of the different types of C, A, and T domains; from structural studies on domain−domain interactions;[23−25,28] and from successful combinatorial biosynthesis experiments reviewed here: (i) use a C^{III} ($^{F}C_{L}$) domain to couple a fatty acid to the N-terminal L-amino acid; (ii) use a C^{I} ($^{L}C_{L}$) domain to couple an L-amino acid to an L-amino acid; (iii) use a C^{II} ($^{D}C_{L}$) domain to couple a D-amino acid to an L-amino acid; (iv) use a T^{I} (T_{C}) domain to interact with a downstream C domain; (v) use a T^{II} (T_{E}) domain to interact with downstream C and E domains; (vi) use a T^{III} (T_{Te}) domain to interact with a downstream Te domain; and (vii) keep natural CA bidomains and CAT tridomains together whenever possible. In addition, it is important to ensure that the engineered NRPS pathway encodes appropriate accessory devices for amino acid modification, chaperone functions (MbtH-like proteins), editing (stand-alone thioesterase), self-resistance, and transport. The synthetic biology process reviewed here was facilitated by the generation of different streptomycete expression hosts from relatively productive starting strains (chassis), BAC cloning, rapid recombineering in *E. coli*, heterospecific conjugation and site-specific integration of engineered genes, and ectopic expression from the constitutive *ermE** promoter. This system was robust enough to provide levels of novel lipopeptides for isolation, characterization and biological testing.

If synthetic biology is to mature as a robust scientific discipline, it needs to be able to address the fundamental questions on the evolution of complex structures, biosynthetic pathways, cell-to-cell communication, and living systems in general. To accomplish this, it must harness the driving forces of evolution: mutation, recombination, and selection for fitness. The secondary metabolite biosynthetic pathways that are hallmarks of actinomycetes with large genomes are attractive early targets for synthetic biology advances because of their relative simplicity, their advanced levels of biochemical and genetic understanding, their predictable and measurable Darwinian outcomes, and their practical applications to medical and agricultural advancements. From the work reviewed here, it is clear that standardization of biosynthetic parts is complex because of certain critical domain−domain interactions. Some parts and devices are amenable to standardization for initial testing, as demonstrated in the combinatorial biosynthetic studies reviewed here, but some will need to be further optimized by directed mutagenesis for commercialization (selection for fitness II).

So how can synthetic biology approaches help advance combinatorial biosynthesis of novel peptides, polyketides, and mixed biosynthetic pathways? It is clear that future advances will not come from the traditional pharmaceutical industry, where combinatorial biosynthetic studies on polyketides and peptides were carried out in the 1980s and 1990s at companies such as Abbott Laboratories and Eli Lilly and Company. Further advances were made at biotechnology companies such as Kosan, Biotica, and Cubist in more recent years, but the level of effort afforded by such companies was not sufficient to advance the technology to its full potential. Technology advancements are a means to an end in such companies and are driven by specific product profiles. Thus Cubist focused on improving certain properties of daptomycin, and when it became doubtful that the full targeted activity could be achieved in a relatively short time frame, the project was discontinued. Thus further advances need to be driven by a mechanism that focuses first on technology advancements, including development of computer-assisted design software to predict successful outcomes, and second on immediate practical applications. It is not obvious how this should be optimally funded, but the field needs to have more academic involvement in close collaboration with industry.[67] Industry can also help to advance the field by providing access to highly productive actinomycete strains to build robust expression hosts capable of providing key precursors and appropriate support mechanisms (devices) to ensure high level production of engineered pathways.[68] As robust chassis design is a key strategic component for accelerated evolution, it is unlikely that "model organisms" such as *E. coli* can be retrofitted with the metabolic capabilities needed to express these complex biosynthetic pathways to compete with industrial actinomycetes in a timely manner, based upon the disappointing expression yields so far obtained with heterologously expressed polyketides.[69]

AUTHOR INFORMATION

Corresponding Author

*Tel: + 1-317-251-3667; E-mail: rbaltz923@aol.com.

Notes

The author declares no competing financial interest.

REFERENCES

(1) Voigt, C. A. (2012) Synthetic biology. *ACS Synth. Biol. 1*, 12.

(2) Zotchev, S. B., Sekurova, O. N., and Katz, L. (2012) Genome-based bioprospecting of microbes for new therapeutics. *Curr. Opin. Biotechnol. 23*, May 2 ; Epub ahead of print; DOI: 10.1016/j.copbio.2012.04.002.

(3) Synthetic Biology Engineering Research Center. http://www.synberc.org.

(4) Giessen, T. W., and Marahiel, M. A. (2012) Ribosome-independent biosynthesis of biologically active peptides: application of synthetic biology to generate structural diversity. *FEBS Lett. 586*, 2065−2075.

(5) Nannemann, D. P., Birmingham, W. R., Scism, R. A., and Bachmann, B. O. (2012) Assessing directed evolution methods for the generation of biosynthetic enzymes with potential in drug biosynthesis. *Future Med. Chem. 3*, 808−819.

(6) Winter, J. M., and Tang, Y. (2012) Synthetic biological approaches to natural product biosynthesis. *Curr. Opin. Biotechnol. 23*, Jan 3 ; Epub ahead of print; DOI: 10.1016/j.copbio.2011.12.016.

(7) Debono, M., Barnhart, M., Carrell, C. B, Hoffmann, J. A., Occolowitz, J. L., Abbott, B. J., Fukuda, D. S., Hamill, R. L., Biemann, K., and Herlihy, W. C. (1987) A21978C, a complex of new acidic peptide antibiotics: isolation, chemistry, and mass spectral structure elucidation. *J. Antibiot. 40*, 761−777.

(8) Boeck, L. D., Papiska, H. R., Wetzel, R. W., Mynderse, J. S., Fukuda, D. S., Mertz, F. P., and Berry, D. M. (1990) A54145, a new lipopeptide antibiotic complex: discovery, taxonomy, fermentation and HPLC. *J. Antibiot. 43*, 587−593.

(9) Fukuda, D. S., Du Bus, R. H., Baker, P. J., Berry, D. M., and Mynderse, J. S. (1990) A54145, a new lipopeptide antibiotic complex: isolation and characterization. *J. Antibiot. 43*, 594−600.

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

(11) Miao, V., Coëffet-LeGal, M.-F., Brian, P., Brost, R., Penn, J., Whiting, A., Martin, S., Ford, R., Parr, I., Bouchard, M., Silva, C. J., Wrigley, S. K., and Baltz, R. H. (2005) Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology 151*, 1507−1523.

(12) Miao, V., Brost, R., Chapple, J., She, K., Coëffet-Le Gal, M.-F., and Baltz, R. H. (2006) The lipopeptide antibiotic A54145 biosynthetic gene cluster from *Streptomyces fradiae*. *J. Ind. Microbiol. Biotechnol. 33*, 129−140.

(13) Boeck, L. D., Fukuda, D. S., Abbott, B. J., and Debono, M. (1990) Deacylation of A21978C, an acidic lipopeptide antibiotic complex, by *Actinoplanes utahensis*. *J. Antibiot. 41*, 1085−1092.

(14) Fukuda, D. S., Debono, M., Molloy, R. M., and Mynderse, J. S. (1990) A54145, a new lipopeptide antibiotic complex: microbial and chemical modification. *J. Antibiot. 43*, 601−606.

(15) Debono, M., Abbott, B. J., Molloy, R. M., Fukuda, D. S., Hunt, A. H., Daupert, V. M., Counter, F. T., Ott, J. L., Carrell, C. B., Howard, L. C, Boeck, L. D., and Hamill, R. L. (1988) Enzymatic and chemical modifications of lipopeptide antibiotic A21978C: the synthesis and evaluation of daptomycin (LY146032). *J. Antibiot. 41*, 1093−1105.

(16) Counter, F. T., Allen, N. E., Fukuda, D. S., Hobbs, J. N., Ott, J., Ensminger, P. W., Mynderse, J. S., Preston, D. A., and Wu, C. Y. E. (1990) A54145 a new lipopeptide antibiotic complex: microbial evaluation. *J. Antibiot. 43*, 616−622.

(17) Eisenstein, B. I., Oleson, F. B., Jr, and Baltz, R. H. (2010) Daptomycin: from the mountain to the clinic, with the essential help of Francis Tally, MD. *Clin. Infect. Dis. 50*, S10−S15.

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

(19) Fowler, V. G., Boucher, H. W., Corey, G. R., Abrutyn, E., Karchmer, A. W., Rupp, M. E., Levine, D. P., Chambers, H. F., Tally, F. P., Vigliani, G. A.; Cabell, C. H.; Link, A. S., DeMeyer, I., Filler, S. G., Zervos, M., Cook, P., Parsonnet, J., Bernstein, J. M., Price, C. S., Forrest, G. N., Fatkenheuer, G., Gareca, M., Rehm, S. J., Brodt, H. R., Tice, A., Cosgrove, S. E., *S. aureus* Endocarditis and Bacteremia Study Group (2006) Daptomycin versus standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*. *N. Engl. J. Med. 355*, 653−655.

(20) Pertel, P. E., Bernardo, P., Fogerty, C., Matthews, P., Northland, R., Benvenuto, M., Thorne, G. M., Luperchio, S. A., Arbeit, R. D., and Alder, J. (2008) Effects of prior effective therapy on the efficacy of daptomycin and cephtriaxone for the treatment of community-acquired pneumonia. *Clin. Infect. Dis. 46*, 1142−1151.

(21) Piper, K. E., Steckelberg, J. M., and Patel, R. (2005) In vitro activity of daptomycin against clinical isolates of Gram-positive bacteria. *J. Infect. Chemother. 11*, 207−209.

(22) Nguyen, K., Ritz, D., Gu, J.-Q., Alexander, D., Chu, M., Miao, V., Brian, P., and Baltz, R. H. (2006) Combinatorial biosynthesis of lipopeptide antibiotics related to daptomycin. *Proc. Natl. Acad. Sci. U.S.A. 103*, 17462−17467.

(23) Nguyen, K., He, X., Alexander, D. C., Li, C., Gu, J.-Q., Mascio, C., Van Praagh, A., Morton, L., Chu, M., Silverman, J. A., Brian, P., and Baltz, R. H. (2010) Genetically engineered lipopeptide antibiotics related to A54145 and daptomycin with improved properties. *Antimicrob. Agents Chemother. 54*, 1404−1413.

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

(25) Baltz, R. H. (2010) Genomics and the ancient origins of the daptomycin biosynthetic gene cluster. *J. Antibiot. 63*, 506−511.

(26) Gu, J.-Q., Alexander, D. C., Rock, J., Brian, P., Chu, M., and Baltz, R. H. (2011) Structural characterization of a lipopeptide antibiotic A54145(Asn_3Asp_9) produced by a genetically engineered strain of *Streptomyces fradiae*. *J. Antibiot. 64*, 111−116.

(27) Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res. 39*, W339−W346.

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

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

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

(31) Koglin, A., and Walsh, C. T. (2009) Structural insights into nonribosomal peptide enzymatic assembly lines. *Nat. Prod. Rep. 26*, 987−1000.

(32) Strieker, M., Tanović, A., and Marahiel, M. A. (2010) Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol. 20*, 234−240.

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

(34) Walsh, C. T., and Fischbach, M. A. (2010) Natural products version 2.0: connecting genes to molecules. *J. Am. Chem. Soc. 132*, 2469−2493.

(35) Baltz, R. H., Nguyen, K. T., Alexander, D. C. (2010) Reprogramming daptomycin and A54145 biosynthesis to produce novel lipopeptide antibiotics, in *Enzyme Technologies: Metagenomics,*

Evolution, Biocatalysis, and Biosynthesis (Yeh, W.-K., Yang, H.-C., and McCarty, J. R., Eds.) pp 285−307, Wiley, Hoboken, NJ.
(36) Luo, L., Kohli, R. M., Onishi, M., Linne, U., Marahiel, M. A., and Walsh, C. T. (2002) Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry 41*, 9184−9196.
(37) Koglin, A., Mofid, M. R., Löhr, F., Schäffer, B., Rogov, V. V., Blum, M.-M., Mittag, T., Marahiel, M. A., Bernhard, F., and Dötsch, V. (2006) Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science 312*, 273−276.
(38) Rausch, C., Hoof, I., Weber, T., Wohlleben, W., and Huson, D. H. (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol. 7*, 78.
(39) Tanovic, A, Samel, S. A., Essen, L. O., and Marahiel, M. A. (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science 321*, 659−663.
(40) Zhou, Z., Lai, J. R., and Walsh, C. T. (2006) Interdomain communication between the thiolation and thioesterase domains of EntF explored by combinatorial mutagenesis and selection. *Chem. Biol. 13*, 869−879.
(41) Frueh, D. P., Arthanari, H., Koglin, A., Vosburg, D. A., Bennett, A. E., Walsh, C. T., and Wagner, G. (2008) Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature 454*, 903−906.
(42) Baltz, R. H., Nguyen, K. T., Alexander, D. C. (2010) Genetic engineering of acidic lipopeptide antibiotics, in *Manual of Industrial Microbiology and Biotechnology* (Baltz, R. H., Davies, J. E., and Demain, A. L., Eds.) pp 391−410, American Society for Microbiology, Washington, DC.
(43) 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, S. K., and Baltz, R. H. (2006) Genetic engineering in *Streptomyces roseoporus* to produce hybrid lipopeptide antibiotics. *Chem. Biol. 13*, 269−276.
(44) Nguyen, K., Kau, D., Gu, J.-Q., Brian, P., Wrigley, S. K., Baltz, R. H., and 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.
(45) Coëffet-Le Gal, M.-F., Thurson, L., Rich, P., Miao, V., and Baltz, R. H. (2006) Complementation of daptomycin *dptA* and *dptD* deletion mutations *in-trans* and production of hybrid lipopeptide antibiotics. *Microbiology 152*, 2993−3001.
(46) Gu, J.-Q., Nguyen, K. T., Gandhi, C., Rajgarhia, V., Baltz, R. H., Brian, P., and Chu, M. (2007) Structural characterization of daptomycin analogues A21978C$_{1-3}$(D-Asn$_{11}$) produced by a recombinant *Streptomyces roseosporus* strain. *J. Nat. Prod. 70*, 233−240.
(47) Doekel, S., Coëffet-Le Gal, M.-F., Gu, J.-Q., Chu, M., Baltz, R. H., and Brian, P. (2008) Non-ribosomal peptide synthetase module fusions to produce derivatives of daptomycin in *Streptomyces roseosporus*. *Microbiology 154*, 2872−2880.
(48) Alexander, D. C., Rock, J., He, X., Miao, V., Brian, P., and Baltz, R. H. (2010) Development of a genetic system for lipopeptide combinatorial biosynthesis in *Streptomyces fradiae* and heterologous expression of the A54145 biosynthetic gene cluster. *Appl. Environ. Microbiol. 76*, 6877−6887.
(49) Alexander, D. C., Rock, J., Gu, J.-Q., Mascio, C., Chu, M., Brian, P., and Baltz, R. H. (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.
(50) Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A., and Essen, L.-O. (2007) Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure 15*, 781−792.
(51) Penn, J., Li, X., Whiting., A., Latif, M., Gibson, T., Silva, C. J., Brian, P., Davies, J., Miao, V., Wrigley, S. K., and Baltz, R. H. (2006) Heterologous production of daptomycin in *Streptomyces lividans*. *J. Ind. Microbiol. Biotechnol. 33*, 121−128.
(52) Wittmann, M., Linne, U., Pohlmann, V., and Marahiel, M. A. (2008) Role of DptE and DptF in the lipidation reaction of daptomycin. *FEBS Journal 275*, 5343−5354.
(53) Krass, F. I., Giesen, T. W., and Marahiel, M. A. (2012) Exploring the mechanism of lipid transfer during biosynthesis of the acidic lipopeptide antibiotic CDA. *FEBS Lett. 586*, 283−288.
(54) Boeck, L. D., and Wetzel, R. W. (1990) A54145, a new lipopeptide antibiotic complex: factor control through precursor directed biosynthesis. *J. Antibiot. 43*, 607−615.
(55) Müller, C., Gebhardt, P., Heinzelmann, E., Lange, C., Puk, O., Welzel, K., Wohlleben, W., and 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.
(56) Wang, Y., Chen, Y., Shen, Q., and Yin, X. (2011) Molecular cloning and identification of the laspartamycin biosynthetic gene cluster from *Streptomyces viridochromogenes*. *Gene 483*, 11−21.
(57) Kreuzman, A. J., Hodges, R. L., Swartling, J. R., Pohl, T. E., Ghag, S. K., Baker, P. J., McGilvray, D., and Yeh, W. K. (2000) Membrane-associated echinocandin B deacylase of *Actinoplanes utahensis*: purification, characterization, heterologous cloning and enzymatic deacylation reaction. *J. Ind. Microbiol. Biotechnol. 24*, 173−180.
(58) Mahlert, C, Kopp, F, Thirlway, J, Micklefield, J., and Marahiel, M. A. (2007) Stereospecific enzymatic transformation of α-ketoglutarate to (2*S*,3*R*)-3-methyl glutamate during acidic lipopeptide biosynthesis. *J. Am. Chem. Soc. 129*, 12011−12018.
(59) Baltz, R. H. (2011) Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery. *J. Ind. Microbiol. Biotechnol. 38*, 1747−1760.
(60) Boll, B., Taubitz, T., and Heide, L. (2012) Role of MbtH-like proteins in the adenylation of tyrosine during amino coumarin and vancomycin biosynthesis. *J. Biol. Chem. 286*, 36281−36290.
(61) Mitchell, C. A., Shi, C., Aldrich, C. C., and Gulick, A. M. (2012) Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry 51*, 3252−3263.
(62) McMahon, M. D., Rush, J. S., and Thomas, M. G. (2012) Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. *J. Bacteriol. 194*, 2809−2818.
(63) van Wezel, G. P., and McDowall, K. J. (2011) The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Nat. Prod. Rep. 28*, 1311−1333.
(64) Baltz, R. H (2011) Strain improvement in actinomycetes in the postgenomic era. *J. Ind. Microbiol. Biotechnol. 38*, 657−666.
(65) Baltz, R. H (2012) *Streptomyces* temperate bacteriophage integration systems for stable genetic engineering of actinomycetes (and other organisms). *J. Ind. Microbiol. Biotechnol. 39*, 661−672.
(66) Fischbach, M. A., Lai, J. R., Roche, E. D., Walsh, C. T., and Liu, D. R. (2007) Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc. Natl. Acad. Sci. U.S.A. 104*, 11951−11956.
(67) Baltz, R. H. (2008) Renaissance in antibacterial discovery from actinomycetes. *Curr. Opin. Pharmacol. 8*, 557−563.
(68) Baltz, R. H. (2010) *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. *J. Ind. Microbiol. Biotechnol. 37*, 759−772.
(69) Baltz, R. H. (2006) Molecular engineering approaches to peptide, polyketide and other antibiotics. *Nat. Biotechnol. 24*, 1533−1540.