Adding the Lipo to Lipopeptides: Do More with Less

Yit-Heng Chooi[1] and Yi Tang[1,2,*]
[1]Department of Chemical and Biomolecular Engineering
[2]Department of Chemistry and Biochemistry
University of California, Los Angeles, Los Angeles, CA 90095, USA
*Correspondence: yitang@ucla.edu
DOI 10.1016/j.chembiol.2010.08.001

Surfactin is a member of the lipopeptide family of antibiotics, which includes the clinical drug daptomycin (Cubicin). The potency of these antibiotics is affected by the attached lipid chain, which is incorporated into the nonribosomally assembled peptidyl backbone via a process known as lipoinitiation. Kraas et al. (2010) have provided valuable insights into the lipoinitiation mechanism, which will be useful for future biosynthetic modifications of lipopeptide antibiotics.

Molecules with common structural features often share similar biosynthetic routes or strategies. However, there are instances where organisms have evolved more than one way to assemble a molecule, and in some cases, fewer than the expected number of enzymes to perform the same function. In this issue, Kraas et al. (2010) illustrated one such example in the biosynthesis of the cyclic lipopeptide surfactin, and the study has provided valuable insights for future investigation and modification of lipopeptide pathways.

The cyclic lipopeptides are a new class of antibiotics that includes the recently approved drug daptomycin (Cubicin, Cubist Pharmaceuticals) (Baltz et al., 2005). Some other well-known examples are the calcium-dependent antibiotics (CDA), mycosubtilin, A54145, friulimicin, fengycin, and others (for recent reviews, see Baltz et al.[2005], and Strieker and Marahiel [2009]). The lipopeptides are assembled through the successive additions of both proteinogenic and nonproteinogenic amino acids by nonribosomal peptide synthetases (NRPSs). They are also characterized by having a medium to long fatty acyl chain attached to the macrocyclic peptide core. The fatty acyl chains are important for the biological properties of lipopeptides and are also the first substructure incorporated into the peptidyl backbone via a process known as lipoinitiation.

A majority of the known lipopeptides are isolated as complex mixtures in which the lipid chain lengths and structures can vary. For example, daptomycin is in fact a minor component of the A21978C complex, a set of lipopeptides containing different fatty acyl chains (Baltz et al., 2005). Structural activity relationship (SAR) studies have shown that variations in the lipid side chains have significant affects on the bioactivities. For A2198C and A54145, it has been shown that the antibacterial activity increases with acyl chain length, but in vivo toxicity was observed when the lipid chain is longer than C_{11} (Baltz et al., 2005). The decanoyl derivative of A21978C, daptomycin, was chosen based on its high therapeutic index and can be produced in sufficient quantities by feeding the producing organism *Streptomyces roseosporus* with decanoic acid during fermentation (Baltz et al., 2005). Consequently, there are considerable interests in studying and engineering the lipoinitiation mechanism.

The lipid side chains of lipopeptides can be incorporated in a number of ways. In a rather complex mechanism, the β-amino fatty acid moiety of mycosubtilin is incorporated via the actions of a polyketide synthase (PKS)-NRPS hybrid with an N-terminal didomain that consists of a fatty acyl ligase (AL) and an acyl carrier protein (ACP). Hansen et al. (2007) showed that decanoic acid, which likely populates from the primary metabolism, is activated by the AL domain of MycA and is directly transferred to the first ACP domain without going through a decanoyl-CoA thioester intermediate (Figure 1A). The adjoining PKS module further extends and aminates the decanoyl chain, which can then initiate the functions of the NRPS portion of MycA. For daptomycin biosynthesis, the AL and ACP are present as discrete proteins (DptE and DptF). The DptE preferentially activates and transfers branched mid- to long-chain fatty acids to DptF, which is incorporated by the N-terminal condensation (C) domain of the initiating NRPS module (also called starter C domain) to prime the synthesis of the peptide portion (Wittmann et al., 2008) (Figure 1B). In both scenarios above, the choice of the fatty acid starter unit is likely controlled by the selectivity of the AL domain as well as the donor site of the starter C domain.

All of the CDAs isolated from *Streptomyces* species possess a 2,3-epoxy-hexanoyl chain. Unlike the examples above, the CDA pathways do not use AL for selective fatty acid activation. Rather, CDA biosynthesis uses a dedicated pathway to synthesize the starter unit (Hojati et al., 2002). The *fab* operon in the CDA gene cluster, which encode a β-ketoacyl-ACP synthase (KAS)-II and KAS-III, have been proposed to synthesize the hexanoyl chain (Figure 1C). Two FAD-dependent oxidases (HxcO and HcmO) in the *fab* operon have been shown to desaturate and epoxidize the hexanoyl chain tethered on an ACP (SCO3249) to afford the 2,3-epoxy-hexanoyl-ACP starter (Kopp et al., 2008).

In the study by Kraas et al. (2010), the authors show that the lipoinitiation mechanism of surfactin biosynthesis is much simpler. The starter C domain at the N-terminal of SrfAA synthetase is unique compared with other C domains. First, the SrfAA starter C domain can catalyze the direct formation of an amide

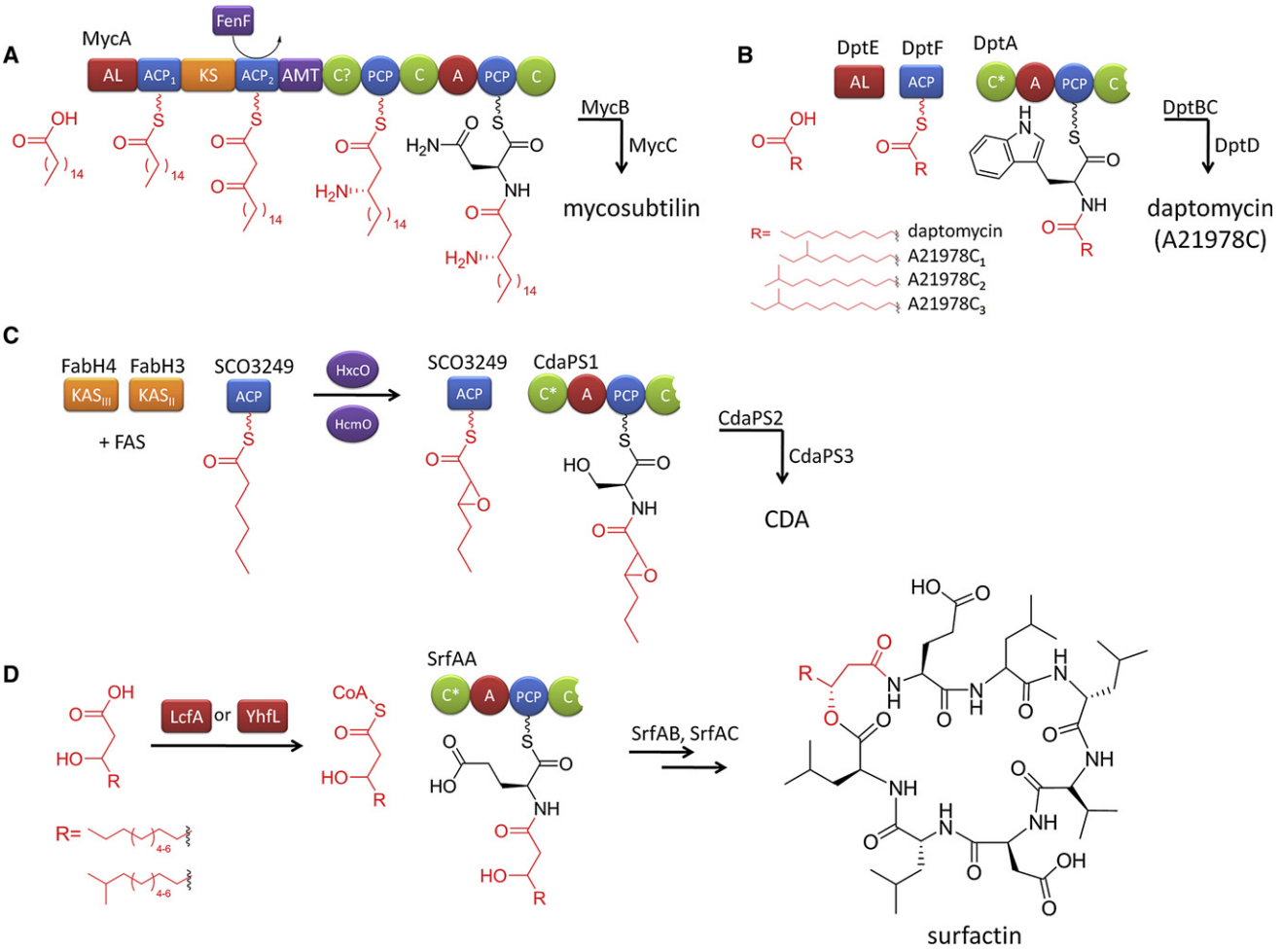

Figure 1. Lipoinitiation Strategies in Biosynthesis of Various Lipopeptides
(A) Mycosubtilin.
(B) Daptomycin.
(C) CDA.
(D) Surfactin.

bond between a fatty acyl-CoA and an amino acid moiety. Second, unlike in previous examples, the surfactin gene cluster does not utilize an AL for acyl activation, and the lipid chain does not need to be first tethered to a dedicated ACP (Figure 1D). Instead, the 3-hydroxy fatty acyl starter is directly derived from free fatty acyl-CoAs; hence, it was reasoned that the starter C domain must have a considerably higher specificity to afford the relatively few variations in the fatty acyl chains of surfactin complex.

To study the surfactin lipoinitiation mechanism, Kraas et al. (2010) first searched for a fatty acyl CoA ligase (FACL) that may synthesize the 3-hydroxy fatty acyl-CoA. Genetic knockouts of the FACLs in the *Bacillus subtillus* genome and the in vitro results support the hypothesis that surfactin biosynthesis utilizes the fatty acyl CoAs generated from primary metabolism and does not have a dedicated FACL. By using in vitro assays with dissected C and PCP domains, Kraas et al. (2010) were able to demonstrate that the donor site of SrfAA starter C domain indeed shows a high chain length specificity for 3-hydroxy fatty acids, with C14 being the best substrate. Lastly, while the starter C domain acceptor site exhibits specificity toward glutamate as the amino acid substrate, it displayed surprising tolerance toward noncognate PCPs. When incubated with corresponding 3-hydroxy fatty acyl-CoA and various noncognate PCPs loaded with glutamate, the SrfAA starter C domain was able to catalyze the amide bond formation between the two substrates.

The exciting work by Kraas et al. (2010) showed that it takes fewer numbers of enzymes to incorporate the surfactin lipid side chain than the other lipopetide pathways. This was the first time that the proposed acylation function of a starter C domain has been confirmed by in vitro assays using dissected domains. The strategy used by the surfactin starter C domain may be similarly adopted by more lipopeptide pathways. The toleration of the starter C domain toward noncognate PCPs has the significant potential to be exploited for combinatorial biosynthesis. By introducing the starter C domain into foreign NRPSs,

novel lipopeptides with different chain length may be produced for clinical applications.

REFERENCES

Baltz, R.H., Miao, V., and Wrigley, S.K. (2005). Nat. Prod. Rep. *22*, 717–741.

Hansen, D.B., Bumpus, S.B., Aron, Z.D., Kelleher, N.L., and Walsh, C.T. (2007). J. Am. Chem. Soc. *129*, 6366–6367.

Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., Wilkinson, B., Sidebottom, P.J., Rudd, B.A.M., Hayes, M.A., et al. (2002). Chem. Biol. *9*, 1175–1187.

Kopp, F., Linne, U., Oberthur, M., and Marahiel, M.A. (2008). J. Am. Chem. Soc. *130*, 2656–2666.

Kraas, F.I., Helmetag, V., Wittman, M., Strieker, M., and Marahiel, M.A. (2010). Chem. Biol. *17*, this issue, 872–880.

Strieker, M., and Marahiel, M.A. (2009). ChemBioChem *10*, 607–616.

Wittmann, M., Linne, U., Pohlmann, V., and Marahiel, M.A. (2008). FEBS J. *275*, 5343–5354.

Expanding the Biological Periodic Table

Javier Seravalli[1] and Stephen W. Ragsdale[2,*]
[1]Department of Biochemistry, University of Nebraska-Lincoln, 1901 Vine Street, Lincoln, NE 68588, USA
[2]Department of Biological Chemistry, University of Michigan Medical School, 1150 W. Medical Center Drive, 5301 MSRB III, Ann Arbor, MI 48109-0606, USA
*Correspondence: sragsdal@umich.edu
DOI 10.1016/j.chembiol.2010.08.004

Metal ions play an indispensable role in biology, enabling enzymes to perform their functions and lending support to the structures of numerous macromolecules. Despite their prevalence and importance, the metalloproteome is still relatively unexplored. Cvetkovic et al. (2010) now describe an approach to identify metalloproteins on a genome-wide scale.

Selected blocks from the iconic Periodic Table that looms over our chemistry lecture halls and classrooms are often highlighted by metallobiochemists and represented as the Biological Periodic Table, which includes the metal ions that play essential roles in the structure and function of macromolecules. We generally consider that about 40% of all enzymes depend on metals for activity because proteins have been isolated whose function depends on the presence of one of the first- (V, Mn, Fe, Co, Ni, Cu, Zn), second- (Mo), or third-row transition metal (W) ions. Yet one wonders how many elements in the Periodic Table might be required for life. As Joni Mitchell sang, "we are stardust," it seems likely that other elements found in the earth's crust and oceans might also be incorporated into proteins.

Genomic and metagenomic sequences are good starting points to predict whether a given organism or a selected ecosystem will utilize a particular metal ion (Andreini et al., 2009). More than 850 genome sequences have been completed. Yet, our current picture of the metalloproteome is quite incomplete; for example, user proteins may be present with no recognizable transporter, and vice versa (Zhang and Gladyshev, 2009). For metalloproteins, complete gene annotation is not possible at present because an experimental method for identifying all metal-binding proteins is been lacking. The study by Cvetkovic et al. (2010) is an important step forward in filling that gap.

While most individual metalloproteins have been identified after purification, characterization of their function, and their metal content, Mike Adams and his colleagues at the University of Georgia used a metal-based approach (Cvetkovic et al., 2010). They cultured the exemplary microbe *Pyrococcus furiosus* in a medium containing 44 metal ions and used a comprehensive method of metal analysis (inductively coupled plasma mass spectrometry [ICP-MS]) coupled with high-throughput mass spectrometry (HT-MS/MS) to determine all of the metals that this organism assimilates and to identify its metalloproteins on a genome-wide scale. Requiring only ~0.2 ml of sample, ICP-MS takes 0.5 s (or a total of ~3 min per sample) to quantify essentially all masses ranging from 6 (Li) to 238 (U) over a linear range of concentration spanning over six orders of magnitude for most elements. Surprisingly, 21 of the 44 elements present in the medium were found to accumulate in the cytoplasm and 343 distinct metal-containing peaks were identified after a single chromatographic separation step. Similar results were found for two other microbes (*Escherichia coli* and *Sulfolobus solfataricus*).

The value of the metal-based experimental approach chosen by Cvetkovic et al. (2010) relative to an in silico prediction of metal content based on genomic analyses is apparent from the finding that nearly half (158) of the 343 peaks did not align with any of the known metalloprotein conserved sequences. The type of thorough whole organism metal-based analysis described by Cvetkovic et al. (2010) promises to enrich the database (such as the InterPro-Metal database) with a wider array of metal coordination sites, eventually enabling researchers to better identify these essentia, yet diverse and poorly recognized components of protein structure. The metal-based