Biosynthesis Mechanisms

Tandem Acyl Carrier Proteins in the Curacin Biosynthetic Pathway Promote Consecutive Multienzyme Reactions with a Synergistic Effect**

*Liangcai Gu, Eli B. Eisman, Somnath Dutta, Titus M. Franzmann, Stefan Walter, William H. Gerwick, Georgios Skiniotis, and David H. Sherman**

Modular polyketide synthases (PKSs) are large multifunctional biosynthetic enzyme systems that assemble a remarkable array of secondary metabolites with a broad spectrum of biological activities. They are typically comprised of a chain initiation and termination module and multiple chain elongation modules. Chain elongation intermediates are covalently attached to a flexible phosphopantetheine (Ppant) arm of acyl carrier protein (ACP) domains embedded in the PKS modules. They are readily combined with other biosynthetic elements (e.g., non-ribosomal peptide synthetases (NRPSs) and 3-hydroxy-3-methylglutaryl (HMG) β-branching enzyme cassettes) for metabolic diversification.[1–3] These additional enzymes can alter the assembly and tailoring schemes of the original PKS pathways and result in unexpected structural modifications.[4] Many of these hybrid systems remain poorly understood, and biochemical studies have yet to reveal how new features are incorporated and refined in the pathways. However, knowledge gained from these versatile assemblies will guide future engineering of novel secondary metabolic systems.

The biosynthetic pathway for curacin A (Figure 1a), a mixed-polyketide nonribosomal-peptide marine natural product with potent anticancer activities,[5] is a prominent example of a hybrid PKS/NRPS system. In the curacin A chain initiation module, a bifunctional decarboxylase/ S-acetyltransferase (GNAT$_L$) domain, instead of a "canonical" acyltransferase (AT) domain, catalyzes chain initiation.[6] A sulfotransferase domain inserted in the chain termination module was found to catalyze decarboxylative chain termination to form an unusual terminal olefin in the final product.[7] In addition, a 10-enzyme assembly, including a halogenase (Hal) domain, a HMG enzyme cassette, and an enoyl reductase (ER) domain, is incorporated into the first two PKS chain elongation modules, and catalyzes formation of a cyclopropane moiety (Figure 1a).[4]

One key component in the curacin A HMG enzyme cassette is a tandem ACP$_3$ (ACP$_I$–ACP$_{II}$–ACP$_{III}$) tridomain with three almost identical ACPs. All of the other enzymes in the 10-enzyme assembly, except CurB ACP$_{IV}$ and CurC ketosynthase (KS), catalyze modifications only on substrates linked to the tandem ACP$_3$. Moreover, the ACP$_3$ carries substrates for the first chain elongation module and is proposed to mediate substrate transfer to a peptidyl carrier protein (PCP) domain in a downstream CurF NRPS module, instead of an inactive CurF PKS module (Figure 1a).[8] Therefore, the tandem ACP$_3$ might be involved in complex and sequential protein interactions with at least nine enzymes, including CurA AT, KS and Hal domains, CurD HMG-ACP synthase (HCS), CurE dehydratase (ECH$_1$), and CurF decarboxylase (ECH$_2$), ER, condensation/cyclization (Cy) and PCP domains (Figure 1).

Our current understanding of tandem ACP domains in polyketide or fatty acid pathways is that they enable sequential substrate modifications to occur in parallel. Genetic studies based on in vivo assays of truncations and active site mutants of tandem ACP domains in the mupirocin and polyunsaturated fatty acid pathways suggested that they can improve product yields.[9,10] Biochemical studies on a tandem di-ACP in the bacillaene pathway showed that each domain can function in parallel.[3] Interestingly, tandem tri- or di-ACP domains are frequently associated with the HMG enzyme cassettes in hybrid PKS pathways (e.g., jamaicamide, bacillaene, mupriocin, and pederin),[11–15] which catalyze consecutive modifications on tandem ACP-linked substrates (see the Supporting Information, Figure S1). Sequence comparison of the curacin A and jamaicamide pathways shows that these three ACP domains are the most conserved components in the 10-enzyme assembly (ca. 95% sequence identity, except for linker regions between ACPs).[4] Thus, the ACP$_3$ might serve as a special "relay station" to support

[*] Dr. L. Gu, E. B. Eisman, Prof. Dr. D. H. Sherman
Life Sciences Institute, Departments of Medicinal Chemistry, Chemistry, and Microbiology and Immunology
University of Michigan, Ann Arbor, MI 48109 (USA)
E-mail: davidhs@umich.edu

Dr. S. Dutta, Prof. Dr. G. Skiniotis
Life Sciences Institute and Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109 (USA)

Dr. T. M. Franzmann, Prof. Dr. S. Walter
Department of Biological Chemistry, University of Michigan
Ann Arbor, MI 48109 (USA)

Prof. Dr. W. H. Gerwick
Scripps Institution of Oceanography and
Skaggs School of Pharmacy and Pharmaceutical Sciences
University of California at San Diego, La Jolla, CA 92093 (USA)

Dr. L. Gu
Department of Genetics, Harvard Medical School
Boston, MA 02115 (USA)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201005280.
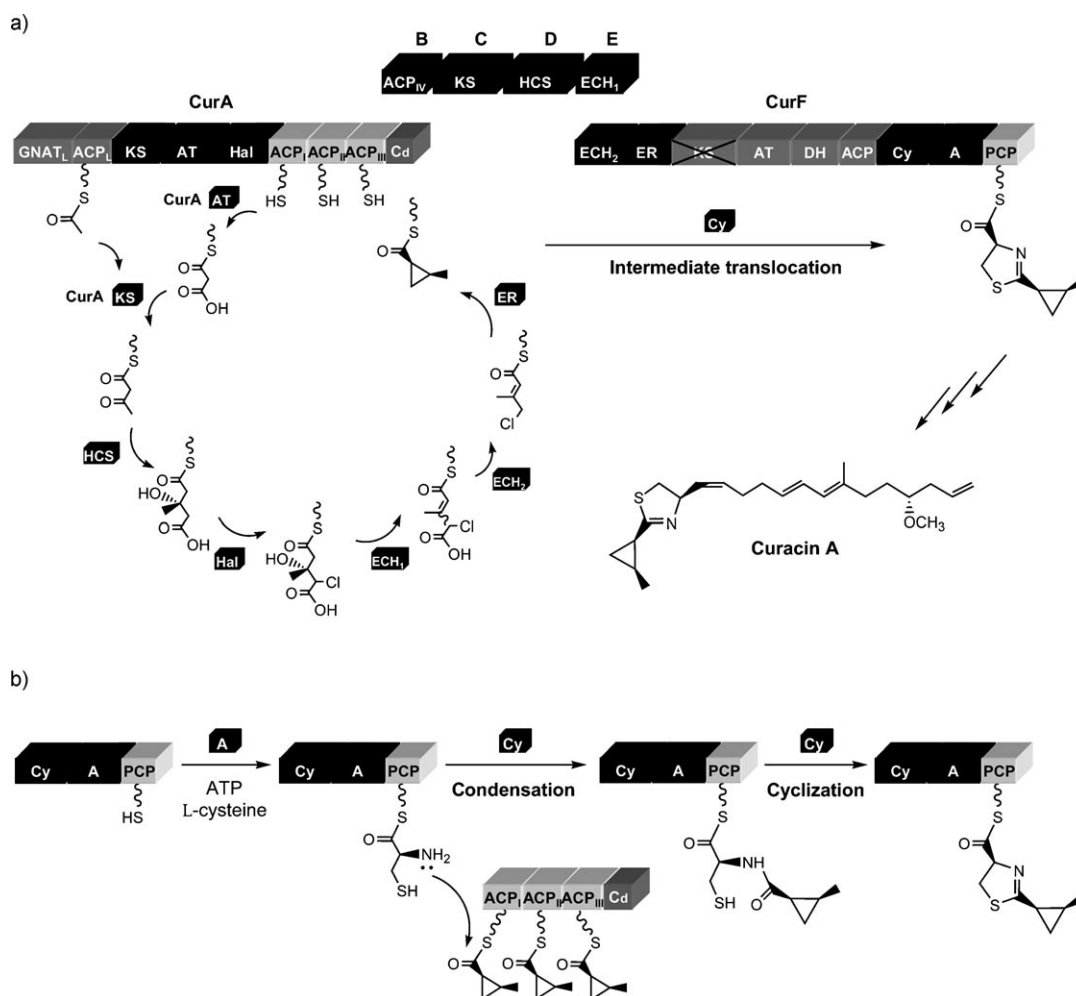
Figure 1. Sequential enzyme reactions for tandem ACP$_3$-linked intermediates. a) The malonyl-ACP substrate undergoes a six-step modification before it is translocated to the CurF PCP. b) Proposed adenylation and condensation/cyclization steps for CurF NRPS module.

sequential substrate modifications as well as substrate transfers in the hybrid PKS pathway. With this proposal in mind, additional biochemical analysis was needed to elucidate the structures and specific functions of these tandem ACPs, especially in the context of sequential multienzyme reactions.

Given the established sequential steps of Hal chlorination, ECH$_1$ dehydration, ECH$_2$ decarboxylation, and ER cyclopropanation from our previous studies,[2,4,6,7,16] as well as the functional CurF NRPS module developed in this work, we were motivated to assess biochemically how the tandem ACP$_3$ can coordinate a four-step modification process (Figure 1 a) and one-step translocation (Figure 1 b) in an in vitro one-pot reaction.

We initiated a comparative analysis by overexpressing and purifying CurA ACP$_3$ and a series of related proteins (all in apo form) derived from constructs bearing variant truncations and mutations. Our bioinformatic analysis suggested that the ACP$_3$ is flanked by a C-terminal domain (Cd, CurA 2261–2311), which is also observed in the jamaicamide pathway with high sequence similarity (Figure S2). Thus, CurA ACP$_I$–ACP$_{II}$–ACP$_{III}$–Cd tetradomain and ACP$_I$–ACP$_{II}$–ACP$_{III}$ tridomain were prepared as N-terminal His-tagged fusion

proteins (Figure S3a). Five truncation constructs, including ACP$_I$–ACP$_{II}$, ACP$_{II}$–ACP$_{III}$, ACP$_I$, ACP$_{II}$, and ACP$_{III}$, and nine single and double Ser→Ala ACP active-site mutations on the ACP$_I$–ACP$_{II}$–ACP$_{III}$ and ACP$_I$–ACP$_{II}$–ACP$_{III}$–Cd constructs were generated (Figure S3). To prepare the ACP-linked HMG substrate for the one-pot reaction, all the ACP$_3$ constructs, truncations and mutants were loaded with (S)-HMG-CoA by the Sfp phosphopantetheinyltransferase from *Bacillus subtilis* as previously described.[4,17]

Next, we sought to analyze tertiary and quaternary structures of the tandem ACP$_3$ by using analytic size-exclusion chromatography (SEC). All the apo ACP$_3$ constructs and truncations were analyzed by SEC to calculate "estimated molecular weights (MWs)". The estimated MWs of polypeptides containing a single ACP domain were close to their actual MWs (Table 1). In contrast, the estimated MWs of those with two or three ACP domains were approximately 1.8-fold their actual MWs, suggesting flexible didomain or tridomain structures lacking ACP–ACP domain interactions that fold into globular structures. When the Cd domain with a predicted α-helix-dominant secondary structure remains at the ACP$_3$ C-terminus, corresponding to an 8.6 kDa mass

© 2011 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1: Estimated molecular weight of the CurA ACP₃ constructs.

ACP construct	Calculated MW [kDa]	Estimated (SEC) MW [kDa]
ACP$_I$	10.9	13.1 ± 0.5
ACP$_{II}$	11.3	14.2 ± 0.5
ACP$_{III}$	11.0	13.2 ± 0.7
ACP$_I$–ACP$_{II}$	23.6	42.3 ± 2.5
ACP$_{II}$–ACP$_{III}$	22.6	40.7 ± 2.0
ACP$_I$–ACP$_{II}$–ACP$_{III}$	36.0	63.1 ± 5.1
ACP$_I$–ACP$_{II}$–ACP$_{III}$–Cd	44.6	106.6 ± 7.3

Figure 2. a) LC-MS detection of **1** cleaved from the Cy–A–PCP. b) Relative yields of **1** from one-pot reactions with variant HMG-linked ACP₃ constructs, truncations, and mutants. The asterisk denotes an unidentified diastereomer of **1**. See the Supporting Information for reaction conditions.

addition, the estimated MW changed from 63.1 kDa to 106.6 kDa (Table 1 and Figure S4). To determine whether the dimerization process is more facile in ACP$_I$–ACP$_{II}$–ACP$_{III}$–Cd, we compared it against ACP$_I$–ACP$_{II}$–ACP$_{III}$ by using analytical ultracentrifugation and negative-stain electron microscopy (EM) analyses. Under analytical ultracentrifugation conditions, the major population of the tetradomain and tridomain polypeptides were captured as monomeric forms (42.0 and 36.8 kDa, respectively). However, the EM analysis shows that both multi-domain proteins exist as a mixture of monomers and dimers (Figures S5 and S6), with about 70% of ACP$_I$–ACP$_{II}$-ACP$_{III}$–Cd, compared to about 25% of ACP$_I$–ACP$_{II}$–ACP$_{III}$, existing as dimers (Figure S7). This data suggests that the C-terminal Cd domain results in a shift of monomer–dimer equilibrium toward dimerization. Moreover, various conformations of protein particles were observed by EM, verifying that the tandem ACP domains adopt a flexible quaternary structure.

Our starting point for biochemical assays was to prepare the soluble forms of the apo and the Ser→Ala PCP active-site mutant of CurF Cy–adenylation (A) domain–PCP tridomain and the excised PCP domain with a C-terminal His-tag (Figure S3a). We next sought to confirm the proposed adenylation and condensation/cyclization activities of the A and Cy domains (Figure 1b), respectively, by detecting a proposed product with a thiazoline ring on the PCP (Figure 1b). The ACP₃-linked cyclopropane substrate was generated by loading the corresponding acyl-CoA substrate onto the apo ACP$_I$–ACP$_{II}$–ACP$_{III}$–Cd, and then incubated with apo CurF Cy–A–PCP, Sfp, CoA, and ʟ-cysteine. To facilitate further analysis, the Cy–A–PCP was purified from the reaction mixtures by reverse-phase high-performance liquid chromatography (HPLC). The product was cleaved from the Cy–A–PCP by a hydrolysis reaction,[18] and then confirmed as **1** by comparison with an authentic standard using liquid chromatography (LC) mass spectrometry (MS; Figure 2a).

With the enzymes and substrates in hand, we assessed how the tandem ACP₃ tridomain can affect an overall yield of the consecutive five-step enzyme reactions. To optimize quantification of the final product **1** by LC-MS, the one-pot reaction was performed by incubating stoichiometric amounts of the ACP-linked HMG and PCP-linked ʟ-cysteine substrates (see Supporting Information) with the enzymes to achieve multiple turnovers, including the Hal, ECH₁, ECH₂, ER, and the Ser→Ala mutant of Cy–A–PCP. The final product was subsequently cleaved from the purified PCP for LC-MS quantification. For comparison of all the ACP₃ constructs,
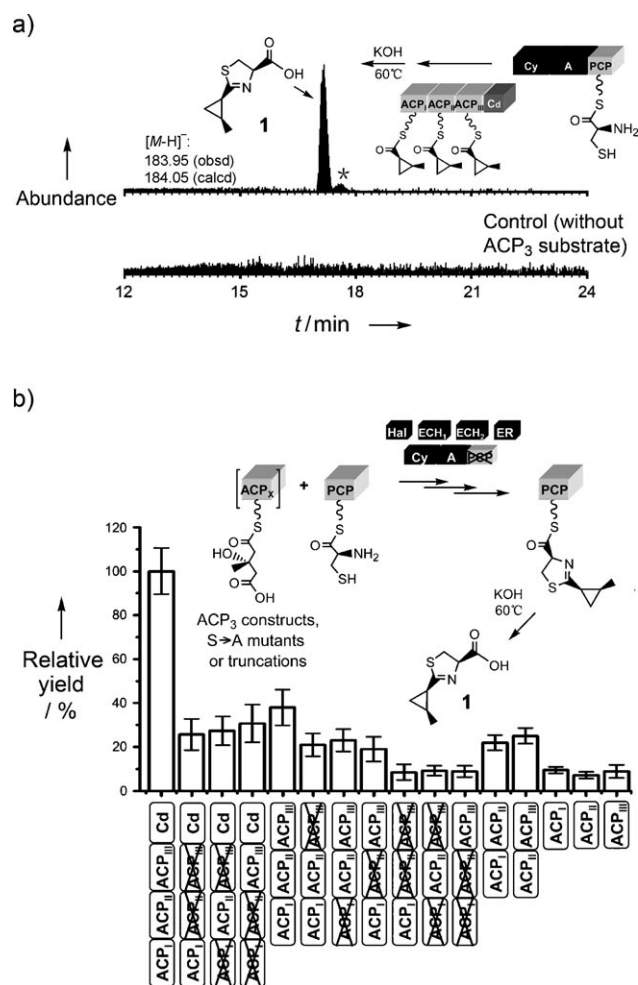
including the truncated and mutant forms, the stoichiometric amount of the HMG substrate remained the same based on numbers of active ACP domains on the substrate polypeptides (e.g., one-fold ACP$_I$–ACP$_{II}$–ACP$_{III}$ corresponds to 1.5-fold ACP$_I$–ACP$_{II}$ or 3-fold double active-site mutant of ACP$_I$–ACP$_{II}$–ACP$_{III}$). The reaction time was adjusted to ensure less than half of the PCP was occupied by the final product.

The analysis above revealed that relative yields of **1** from the one-pot reactions using the polypeptides with only one ACP active site (except those flanked by Cd) are similar (ca. 9%), regardless of the ACP domains excised from or embedded in the ACP₃ tridomain (Figure 2b). This is consistent with the previous finding that tandem ACPs are functionally equivalent.[3,10] For those with two or three active ACP domains (not including ACP$_I$–ACP$_{II}$–ACP$_{III}$–Cd), the product yields were increased to 19–25% and 37%, respectively. Thus, these data reveal that at the same HMG substrate concentration, tandem ACP domains function not only in parallel, but with an unexpected synergistic effect. Notably,

the relative yields of the reaction with ACP_I–ACP_{II}–ACP_{III}–Cd or mutants of ACP_I–ACP_{II}–ACP_{III}–Cd containing a single active ACP were 100% and 26–30%, respectively, which are significantly higher than those lacking the Cd domain. This is likely due to enhanced ACP_I–ACP_{II}–ACP_{III}–Cd dimerization and specific protein–protein interactions mediated by the Cd domain.

In typical PKS pathways, an ACP-linked malonyl or methylmalonyl substrate first undergoes a Claisen condensation catalyzed by KS, and then may be modified by keto reduction, dehydration, and enoyl reduction by ketoreductase, dehydratase, and enoylreductase catalytic domains, respectively, before it is transferred to a downstream KS domain. However, the CurA malonyl-ACP_3 substrates participate in additional modification steps that might compromise catalytic efficiency of the entire pathway. Our study indicates that the efficiency of the five-step enzyme reaction series increases with the level of ACP domain aggregation (e.g., mono-ACP domain vs di-, tri-, or hexa-ACP domains). In principle, the clustering of ACP-linked substrates might result in binding of adjacent enzymes to corresponding substrates through non-specific enzyme–substrate interactions. For example, enzymes associated with ACP-linked substrates usually can bind to the ACP Ppant arm.[16,19,20] In the Cur system this interaction appears to be essential based on the decreased yields of product observed in Ser→Ala ACP active-site mutants devoid of Ppant arms. Thus, the tandem organization of ACP domains can promote formation of multienzyme complexes in sequential reactions, thus facilitating rapid access and pairing of enzymes and substrates. This represents a highly effective strategy of nature to improve sequential multienzyme reactions, and is of significance for de novo design of natural product pathways by metabolic engineering.

[1] C. T. Walsh, *Acc. Chem. Res.* **2008**, *41*, 4.

[2] L. Gu, J. Jia, H. Liu, K. Hakansson, W. H. Gerwick, D. H. Sherman, *J. Am. Chem. Soc.* **2006**, *128*, 9014.

[3] C. T. Calderone, W. E. Kowtoniuk, N. L. Kelleher, C. T. Walsh, P. C. Dorrestein, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8977.

[4] L. Gu, B. Wang, A. Kulkarni, T. W. Geders, R. V. Grindberg, L. Gerwick, K. Hakansson, P. Wipf, J. L. Smith, W. H. Gerwick, D. H. Sherman, *Nature* **2009**, *459*, 731.

[5] P. Verdier-Pinard, J. Lai, H. Yoo, J. Yu, B. Marquez, D. G. Nagle, M. Nambu, J. D. White, J. R. Falck, W. H. Gerwick, B. W. Day, E. Hamel, *Mol. Pharmacol.* **1998**, *53*, 62.

[6] L. Gu, T. W. Geders, B. Wang, W. H. Gerwick, K. Hakansson, J. L. Smith, D. H. Sherman, *Science* **2007**, *318*, 970.

[7] L. Gu, B. Wang, A. Kulkarni, J. J. Gehret, K. R. Lloyd, L. Gerwick, W. H. Gerwick, P. Wipf, K. Hakansson, J. L. Smith, D. H. Sherman, *J. Am. Chem. Soc.* **2009**, *131*, 16033.

[8] Z. Chang, N. Sitachitta, J. V. Rossi, M. A. Roberts, P. M. Flatt, J. Jia, D. H. Sherman, W. H. Gerwick, *J. Nat. Prod.* **2004**, *67*, 1356.

[9] A. S. Rahman, J. Hothersall, J. Crosby, T. J. Simpson, C. M. Thomas, *J. Biol. Chem.* **2005**, *280*, 6399.

[10] H. Jiang, R. Zirkle, J. G. Metz, L. Braun, L. Richter, S. G. Van Lanen, B. Shen, *J. Am. Chem. Soc.* **2008**, *130*, 6336.

[11] R. A. Butcher, F. C. Schroeder, M. A. Fischbach, P. D. Straightt, R. Kolter, C. T. Walsh, J. Clardy, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1506.

[12] A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson, C. M. Thomas, *Chem. Biol.* **2003**, *10*, 419.

[13] J. Piel, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14002.

[14] J. Piel, D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, S. Matsunaga, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16222.

[15] D. J. Edwards, B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts, W. H. Gerwick, *Chem. Biol.* **2004**, *11*, 817.

[16] T. W. Geders, L. Gu, J. C. Mowers, H. Liu, W. H. Gerwick, K. Hakansson, D. H. Sherman, J. L. Smith, *J. Biol. Chem.* **2007**, *282*, 35954.

[17] L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, C. T. Walsh, *Biochemistry* **1998**, *37*, 1585.

[18] W. L. Kelly, N. J. Hillson, C. T. Walsh, *Biochemistry* **2005**, *44*, 13385.

[19] C. Khosla, Y. Tang, A. Chen, N. A. Schnarr, D. E. Cane, *Annu. Rev. Biochem.* **2007**, *76*, 11.1.

[20] D. Khare, B. Wang, L. Gu, J. Razelun, D. H. Sherman, W. H. Gerwick, K. Håkansson, J. L. Smith, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14099.