Chemistry & Biology
Previews

Cleaning up Polyketide Synthases

Jason C. Kwan[1] and Eric W. Schmidt[1,*]
[1]Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT 84112, USA
*Correspondence: ews1@utah.edu
DOI 10.1016/j.chembiol.2012.03.003

Complex biosynthetic enzymes such as polyketide synthases make mistakes. In this issue of *Chemistry & Biology*, Jensen et al. report that a discrete family of acyltransferases is responsible for error correction, hydrolyzing key biosynthetic intermediates from a multi-enzyme complex. This activity might find use in understanding polyketide biosynthesis, particularly in uncultivated organisms and in tailoring the synthesis of small molecules.

Many natural products are made by type I polyketide synthase (PKS) pathways, including important drugs such as erythromycin (antibacterial), and rapamycin (immunosuppression). Genome-based prediction of PKS products is also becoming increasingly important in drug discovery. PKSs are large enzymes that harbor multiple discrete enzymatic domains (Hertweck, 2009). A malonate unit or a derivative thereof is added by each functional "module," extending the polyketide chain by two carbon atoms. At each module, several optional tailoring domains control the final reduction state of the unit added by the preceding module, giving rise to structural diversity.

From early examples it was believed that the order of reactions followed a strict co-linearity, both in gene and domain organization, and that the products of gene clusters could be easily predicted. However, as more PKS pathways have been described, it has become clear that the co-linearity rules are not absolute and there remains a great deal to learn about how these pathways work. A particularly intriguing group is the recently described *trans*-acyltransferase (AT) group of PKSs (Piel, 2010), where the ATs are not found in the main PKS but are on separate proteins that act *in trans*. These pathways have many peculiarities that deviate from the common "textbook" conception of PKSs that often make it difficult to annotate gene clusters. For example, they frequently have seemingly superfluous domains and deviations from co-linearity, and their products have unusual chemical features. While *cis*-AT PKS pathways typically incorporate double bonds in an α-β *trans* configuration, *cis* and β-γ double bonds are often seen in *trans*-AT PKS systems. Another hallmark of *trans*-AT PKS pathways is the inclusion of carbon branches at the β position, i.e., the former position of a carbonyl group in a malonate unit. The leinamycin pathway (Cheng et al., 2003) (see Figure 1) is a good example of this, because it features both an unusual order of modules and a "type B" dehydrating dimodule (Piel, 2010) that may introduce two double bonds, one *cis* and one *trans*. It is worth noting that the timing of the two dehydrations by such a dimodule is unknown (Piel, 2010), and they could either both occur at module 4 or sequentially starting at module 3.

Because of the unusual nature of *trans*-AT PKS pathways, there are still many biochemical questions to answer about how they work, with each new pathway raising new questions. This increases the challenges of characterizing new pathways, especially when the structure of the product is not known. A significant number of *trans*-AT pathways are found in uncultured symbionts (see Table 1), identified from metagenomic analysis. For some time, it has been postulated that symbiotic bacteria produce many natural products that were originally attributed to marine invertebrates and other animals (Piel, 2004). Indeed, these compounds may be an important dimension of the symbiotic relationship between animals and associated bacterial communities. When studying pathways from uncultured or unidentified bacteria, biochemical characterization of proteins becomes especially challenging as the native system is not available. Heterologous expression has been difficult, especially for PKSs where the entire pathway can easily be in excess of 120 kbp.

One of the many mysteries surrounding *trans*-AT systems is the inclusion of more than one AT domain in certain pathways. Some previously obtained genetic data provided tantalizing clues about the role of these dual ATs, but much remained uncertain about their biochemical functions. For example, in the case of mupirocin, it has been shown that one of the AT domains is essential to compound production, whereas deletion of the other leads to reduced product titers (Lopanik et al., 2008). In this issue of *Chemistry & Biology*, Jensen et al. (2012) report the role of the second AT domain in *trans*-AT PKS pathways. This work links the second AT to a curious phenomenon: in some pathways, inactivation of the thioesterase (TE) domain in the last module does not simply prevent the final polyketide product from being released from the PKS enzyme, as expected, but rather leads to the release of shorter intermediates from multiple modules. This effect has been reported in three *trans*-AT PKSs, bacillaene (Moldenhauer et al., 2010), rhizoxin (Kusebauch et al., 2010), and mupirocin (Wu et al., 2008), all of which make use of multiple AT domains. Jensen et al. (2012) show that the second AT, which they term "AT2," does not load extender units onto ACP domains, but instead cleaves already loaded intermediates that are longer and bulkier than malonyl-ACP. They therefore propose that the AT2 acts as a "proofreader" for these pathways, and that it cleaves stalled intermediates.

A similar process of proofreading is known in some *cis*-AT PKS and NRPS pathways, where type II TEs cleave acyl chains from ACPs where aberrant decarboxylation has occurred or an incorrect unit has been incorporated (Kim et al., 2002; Schwarzer et al., 2002). In contrast to the action of "AT2" domains, type II TEs prefer short acyl chains and therefore

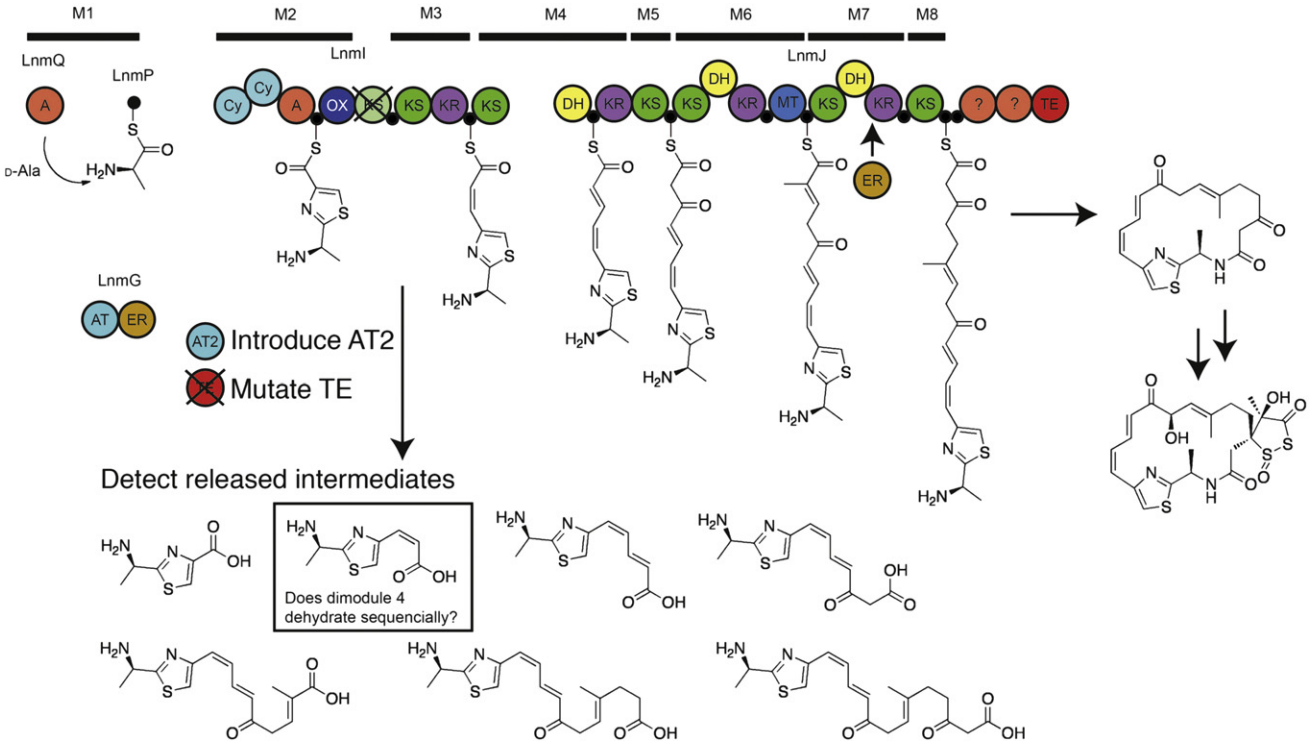

Figure 1. A Future Application for the Proofreader?

Trans-AT polyketide synthase (PKS) leinamycin pathway (Cheng et al., 2003), showing hypothetical intermediates that would be detected upon introduction of an AT2 protein, such as PedC, and mutation of the terminal thioesterase (TE) on LnmJ. Detection of the boxed intermediate would show that dimodule four dehydrates the intermediate made by module 3, as opposed to carrying out two dehydrations after extension at module 4. Unusual features of this pathway are as follows: (1) LnmQ loads D-Ala rather than L-Ala with later epimerization; (2) unusual module order, for example the ACP domain is in the middle of modules 2 and 4, and there is an extra ACP in module 6; (3) module 4 is split across two PKS proteins and is proposed to introduce two double bonds, one *cis* and one *trans* (Piel, 2010); and (4) unique post-PKS transformations give rise to a 1-oxo-1,2-dithiolan-3-one moiety.

do not give rise to release of the native intermediates (Schwarzer et al., 2002). Intriguingly, in the rifamycin pathway, a *cis*-AT type PKS has been observed to release chain intermediates when the terminal TE is mutated (Yu et al., 1999). This gene cluster contains a type II TE (RifR), whose deletion causes a reduction in compound yield (Kim et al., 2002). However, there is also an unannotated gene in the rifamycin gene cluster, orf20 (AAG52990), that is a standalone AT. In light of the work of Jensen et al. (2012) it would be interesting to determine whether orf20 or another AT is involved in intermediate release.

Importantly, Jensen et al. (2012) found that the standalone AT2 PedC was also able to cleave a hexanoyl starter unit from an ACP taken from a fungal iterative PKS. This opens up the possibility that enzymes such as PedC could be used to induce bacillaene-like intermediate ejection into even distantly related pathways if the TE is disrupted. For example, PedC coexpression with the leinamycin pathway could shed some light on the mechanism of dehydration by dimodule 4 (see Figure 1). If the enzyme proves to be tolerant enough to ACP substrates, this could be an extremely useful biochemical tool to probe the function of cryptic PKS pathways. Enzymes such as PedC, therefore, could prove valuable in the biochemical characterization of *trans*-AT PKSs where expression of the whole pathway or knockout experiments are not feasible, such as when the genes belong to an uncultured symbiont.

Table 1. Polyketides Made by *trans*-AT PKS Pathways Found in Uncultured Symbionts

Compound(s)	Host	Symbiont
Pederin	*Paederus* and *Paederidus* spp.	*Pseudomonas* sp.
Onnamides, theopederins	*Theonella swinhoei, Trachychaedus laevispirulifer, Discodermia* sp.	Unknown
Bryostatins	*Bugula neritina, Amathia convoluta*	*Candidatus* "Endobugula sertula"
Rhizoxin	*Rhizopus microsporus*	*Burkholderia rhizonica, B. endofungorum*
Psymberin	*Psammocinia* aff. *bulbosa*	Unknown

REFERENCES

Cheng, Y.-Q., Tang, G.-L., and Shen, B. (2003). Proc. Natl. Acad. Sci. USA *100*, 3149–3154.

Hertweck, C. (2009). Angew. Chem. Int. Ed. *48*, 4688–4716.

Jensen, K., Niederkrüger, H., Zimmermann, K., Vagstad, A.L., Moldenhauer, J., Brendel, N., Frank, S., Kohlhaas, C., Townsend, C.A., et al. (2012). Chem. Biol. *19*, this issue, 329–339.

Kim, B.S., Cropp, T.A., Beck, B.J., Sherman, D.H., and Reynolds, K.A. (2002). J. Biol. Chem. *277*, 48028–48034.

Kusebauch, B., Busch, B., Scherlach, K., Roth, M., and Hertweck, C. (2010). Angew. Chem. Int. Ed. *49*, 1460–1464.

Lopanik, N.B., Shields, J.A., Buchholz, T.J., Rath, C.M., Hothersall, J., Haygood, M.G., Håkansson, K., Thomas, C.M., and Sherman, D.H. (2008). Chem. Biol. *15*, 1175–1186.

Moldenhauer, J., Götz, D.C.G., Albert, C.R., Bischof, S.K., Schneider, K., Süssmuth, R.D., Engeser, M., Gross, H., Bringmann, G., and Piel, J. (2010). Angew. Chem. Int. Ed. *49*, 1465–1467.

Piel, J. (2004). Nat. Prod. Rep. *21*, 519–538.

Piel, J. (2010). Nat. Prod. Rep. *27*, 996–1047.

Schwarzer, D., Mootz, H.D., Linne, U., and Marahiel, M.A. (2002). Proc. Natl. Acad. Sci. USA *99*, 14083–14088.

Wu, J., Hothersall, J., Mazzetti, C., O'Connell, Y., Shields, J.A., Rahman, A.S., Cox, R.J., Crosby, J., Simpson, T.J., Thomas, C.M., and Willis, C.L. (2008). ChemBioChem *9*, 1500–1508.

Yu, T.-W., Shen, Y., Doi-Katayama, Y., Tang, L., Park, C., Moore, B.S., Hutchinson, C.R., and Floss, H.G. (1999). Proc. Natl. Acad. Sci. USA *96*, 9051–9056.

We See the Light: Chemical-Genetic Protein Regulation

Steven A. Farber[1,*] **and Erin M. Zeituni**[1]
[1]Department of Embryology, Carnegie Institution for Science, 3520 San Martin Drive, Baltimore, MD 21218, USA
*Correspondence: farber@ciwemb.edu
DOI 10.1016/j.chembiol.2012.03.001

The challenge of studying complex protein networks in whole animals has driven the development of new methods for manipulating protein function with spatial and temporal precision. A novel combination of chemical and genetic protein regulation (Rodriguez and Wolfgang, in this issue of *Chemistry & Biology*) achieves levels of control that will revolutionize the study of protein function.

Identifying the function of a protein has often relied on characterizing the impact of naturally occurring or directed mutations in the gene encoding that protein. Protein functions were first elucidated in microorganisms through the use of temperature-sensitive alleles, mutations, and deletions. In whole organisms, developmental biologists successfully described a variety of complex protein networks by designing ways to perturb specific proteins in model organisms and characterizing the resulting phenotypes. The past 20 years have seen the advent of homologous recombination techniques in mice that enable protein regulation through mutations, transcriptional control, and RNAi (Thomas and Capecchi, 1990). However, many of these technical advances have serious drawbacks. Attenuated gene expression, gene knockouts, and mutations in mammals frequently have lethal or detrimental outcomes during early development that prevent exploration of protein function on whole animal physiology and at later stages. Although RNAi has proven to be a powerful knockdown technique, designing functional RNAi probes can be problematic, because it is difficult to predict if a given sequence will attenuate mRNA levels. Further, changes in mRNA levels do not always produce a corresponding change in protein levels. Despite these challenges, it is critical to develop approaches that study the function of proteins in whole animals because complicated and overlapping factors such as secreted factors, cell-cell interactions, and metabolic state most certainly impact protein function but are impossible to replicate in less complex systems. Therefore, the future of functional protein studies in whole organisms lies in developing sophisticated tools that reversibly regulate proteins in a specific time and place.

Our understanding of complex biological networks would particularly benefit from improved capabilities in controlling the levels of any given protein with fine-tuned precision. For example, studies of lipid metabolism have been especially challenging because there are often many similar proteins (as defined by enzymatic properties) in the vertebrate genome that can act in concert with functionally-related lipid-modifying proteins. Elucidating the function of these proteins would be facilitated by designing experiments that discretely control expression of a protein. To this end, small molecule-mediated protein regulation has been the most successful approach to temporally regulate proteins due to its broad applicability to almost any protein of study, its specificity in targeting, and its reversibility. The basic principle of this approach is to fuse inherently unstable protein domains to a protein of interest, thereby conferring instability to the entire fusion construct. Introducing a biologically inert small molecule that binds specifically to the unstable protein domain in the fusion construct subsequently stabilizes the fusion protein in a dose-dependent manner, thereby effectively activating the protein of interest. Reversal of this process is achieved by removing the small molecule. Several groups have spent the better part of a decade modifying and perfecting a small protein domain of FK506 binding protein 12 (FKBP12) that can convey instability to any protein to which it is fused (Stankunas et al., 2003; Yang et al., 2000). Improvements in high throughput structure-based