Organic & Biomolecular Chemistry

PERSPECTIVE

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2017, **15**, 4036

Evolution inspired engineering of antibiotic biosynthesis enzymes

M. Metsä-Ketelä

Received 24th January 2017,
Accepted 21st April 2017

DOI: 10.1039/c7ob00189d

rsc.li/obc

Streptomyces soil bacteria are competent chemists that are able to produce thousands of chemically complex natural products. Key to the development of this rich source of metabolites appears to be an evolutionary pressure that promotes chemical diversity; new biosynthetic pathways are continuously being formed in these bacteria, which may result in the appearance of a novel bioactive compound that provides significant competitive advantage to the producing organism. In recent years, our work has focused on understanding how minor changes in the biosynthetic enzymes has led to drastically altered catalytic properties. We have generated chimeric proteins from functionally distinct homologous enzymes involved in the tailoring steps of related anthracycline and angucycline pathways, with the aim of creating novel catalysts. The work provides an opportunity for further protein engineering efforts for production of improved bioactive natural products.

Introduction

Natural products isolated from micro-organisms have been a cornerstone of modern medicine even since the discovery of penicillin. In particular, Gram-positive *Streptomyces* bacteria have provided numerous drugs with antimicrobial, anticancer or antifungal activities.[1] These secondary metabolites typically harbour highly complex chemical structures, which have provided inspiration for medicinal chemists. In parallel, investigations into their biosynthesis have revealed remarkable details on how simple soil bacteria are able to generate complex chemical libraries with relative ease.[2]

The explosion of microbial genome sequencing data has confirmed the existence of numerous unexplored biosynthetic pathways,[3] while the development of synthetic biology yields promise that novel molecules associated with these pathways could be designed.[4] However, the efficiency of synthetic biology depends on the quality of the tools or "biological parts",[5] but cataloguing natural product biosynthesis pathways *de novo* remains a challenge. In other fields (*e.g.* biofuel), catalysts for a given chemical transformation may be selected from a number of well-characterised enzymes based on their desired properties,[6] but in natural products research even prediction of the function of a given protein based on sequence information alone is challenging. This is because these biosynthetic pathways and enzymes residing on those are continuously evolving for generation of chemical diversity,[7] whereas engineering projects involving proteins from primary metabolism (*e.g.* hydrocarbon metabolism in biofuels) dwell with much more conserved functions.[6] As a consequence of the divergent evolution of secondary metabolism, the catalytic properties of even highly related proteins may greatly differ. Selected examples include mono-oxygenases acting as cyclases, and *vice versa*, or methyl transferases catalysing mono-oxygenation reactions.[8]

In our work, we have attempted to engineer the biosynthetic enzymes *themselves* in order to obtain novel functionalities and products. We have taken advantage of years of structural studies and the functional diversity of proteins residing on related pathways. In an approach inspired by evolution, we have identified short segments (2–13 amino acids) in homolo-

Department of Biochemistry, University of Turku, Vatselankatu 2, FIN-20014, Turku, Finland. E-mail: mianme@utu.fi

M. Metsä-Ketelä

Dr Metsä-Ketelä obtained his Ph.D. in biochemistry at the University of Turku, Finland, in 2003. He conducted his post-doctoral studies at the Leibniz Institute for Natural Product Research and Infection Biology, Jena, Germany and the Karolinska Institutet, Stockholm, Sweden. He returned to Finland in 2009 to establish his own research group. He currently holds a research oriented Senior Lecturer position at the University of Turku.

gous protein pairs that might account for their functional differentiation. We have then exchanged these regions individually and in different combinations to generate chimeric enzymes with altered properties. Once key segments have been identified, the underlying causes for the emergence of novel enzyme functions may be pinpointed more precisely in a facile manner by narrowing down the originally selected region.

The benefit of the method has been that we have experienced very few solubility issues, which may be due to the fact that we have been using as templates two naturally evolved proteins that are soluble, but also because of the way we have selected the chimeragenesis regions. In order to determine suitable boundaries we have extended the initial interchangeable region in both N- and C-terminal directions until sites where both the sequences and spatial positions of equivalent amino acids, as estimated by superposition of the paired protein structures, are conserved. In most cases, the segments have composed of loop regions folding over the active sites, but typically our protocol has extended the sequences to cover adjacent and more conserved secondary structures.

In order to test the feasibility of the methodology we have focused on late stage tailoring reactions, where common natural products carbon skeletons are modified in various ways to generate an additional layer of diversity. To date we have modified three protein pairs involved the biosynthesis of anthracycline and angucycline metabolites to surprising effects.

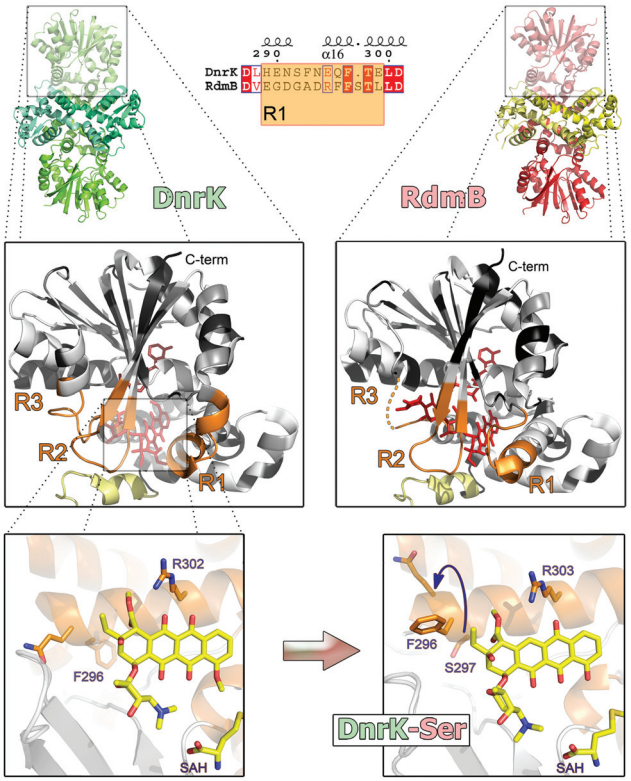

Fig. 1 Crystal structures of DnrK and RdmB with chimeragenesis regions highlighted. Views of the active sites of DnrK and DnrK-Ser reveal that F296 has rotated towards the active site in the mutant.

Results and discussion

Conversion of the SAM-dependent 4-*O*-methyl transferase DnrK into the 10-hydroxylase RdmB

The biosynthetic pathway of the important anticancer agent daunorubicin contains DnrK (Fig. 1), which catalyses the 4-*O*-methylation of various anthracyclines.[9] In contrast, the related rhodomycin pathway harbours RdmB (Fig. 1), with 52% sequence identity to DnrK, that surprisingly has no methyl transferase activity. Rather, RdmB catalyses the 10-hydroxylation of anthracyclines such as **1** (Fig. 2), but requires compounds with a free carboxyl group at C-10 as substrates.[8] We have additionally shown that both enzymes harbour hidden, moonlighting, 10-decarboxylation activity (**2**, Fig. 2), which is a facile reaction that also proceeds non-enzymatically in the presence of light.[10] Another difference between the proteins is that while DnrK requires SAM as a co-substrate for the methylation reaction (**3**, Fig. 2), RdmB only utilizes it as a co-factor for the mono-oxygenation (**4**, Fig. 2).

Structural analysis revealed that the functional differences might be accounted by three regions (R1 to R3, orange, Fig. 1). When the R1 region from RdmB was exchanged to the DnrK scaffold, the chimera was able to catalyse both methylation and mono-oxygenation reactions leading to the formation of a novel double reaction product (**5**, Fig. 2). Further dissection of the R1 area revealed that RdmB contains an additional serine residue in helix α16 (Fig. 1) and insertion of S297 to DnrK indeed led to the appearance of the 10-hydroxylation activity.[10]

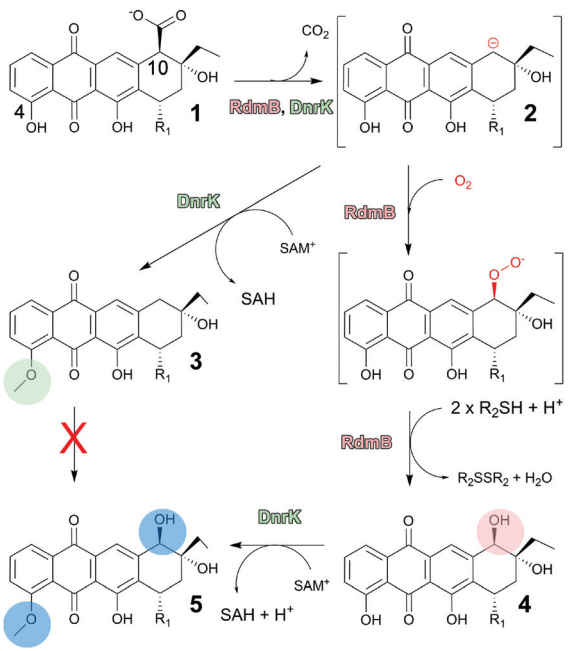

Fig. 2 Mechanistic proposal for the 10-decarboxylation, 4-*O*-methylation and 10-hydroxylation reactions catalysed by DnrK and RdmB. R_1, L-rhodosamine, R_2SH, a thiol reducing agent such as glutathione.

Structure determination of the DnrK-Ser mutant surprisingly uncovered that the inserted serine points away from the active site. However, since the serine was inserted into an α-helical segment, the preceding phenylalanine F296 has rotated towards the active site. In Dnrk-Ser, the bulky F296 blocks a channel to the surface of the protein, which is present in native DnrK. Closure of the active site and protection of the carbanion intermediate **2** from protonation by solvent molecules (Fig. 2) is likely to be crucial for the switch in activity.[10]

A link between substrate inhibition and bifunctionality in the FAD-dependent mono-oxygenase PgaE

Prejadomycin **6** (Fig. 3) is the last common intermediate on many angucycline pathways,[11] from which the metabolites branch off to the various end products *via* complex redox modifications. In gaudimycin/urdamycin biosynthesis, FAD-dependent mono-oxygenases such as Pga/UrdE, respectively, catalyse two consecutive hydroxylation reactions at C12 and C12b (7 and **8**, Fig. 3).[12] An unusual feature of PgaE is the temporal separation of the two reactions due to substrate inhibition; no formation of **8** can be observer prior to depletion of the original substrate **6**. Oxidation of the hydroquinone and 6-ketoreduction by a short chain alcohol dehydrogenase/reductase (SDR) PgaM/UrdMred to form **9** (Fig. 3) finally completes the reaction cascade. The related jadomycin pathway harbours a homologous flavoenzyme JadH, which also catalyses 12-hydroxylation alike PgaE, but differs in the secondary activity where JadH completes 4a,12b-dehydration.[13] The reaction product **10** is then non-enzymatically oxidized to the quinone **11** (Fig. 3). Multiple sequence alignments suggested that the differences between PgaE and JadH might be accounted by one or more of four diverse regions (orange, Fig. 4A). Exchanging the regions from JadH into the PgaE template in different combinations revealed that the chimeras lost the ability to catalyse 12b-hydroxylations.[14] Interestingly, this phenomena was linked to alleviation of substrate inhibition in

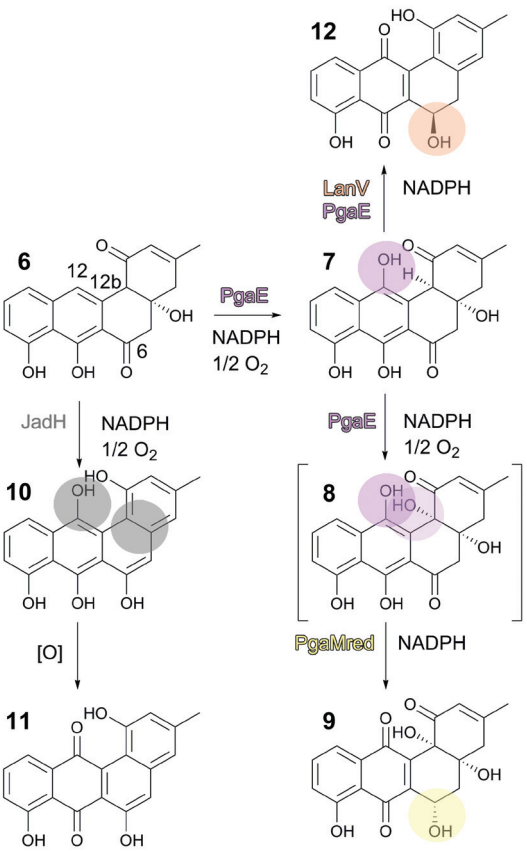

Fig. 3 Mechanistic proposal for the 12- and 12b-hydroxylation and 6-ketoreduction reactions catalysed by PgeE/LanE, PgaMred and LanV. JadH catalyses 12-hydroxylation and 4a,12b-dehydration.

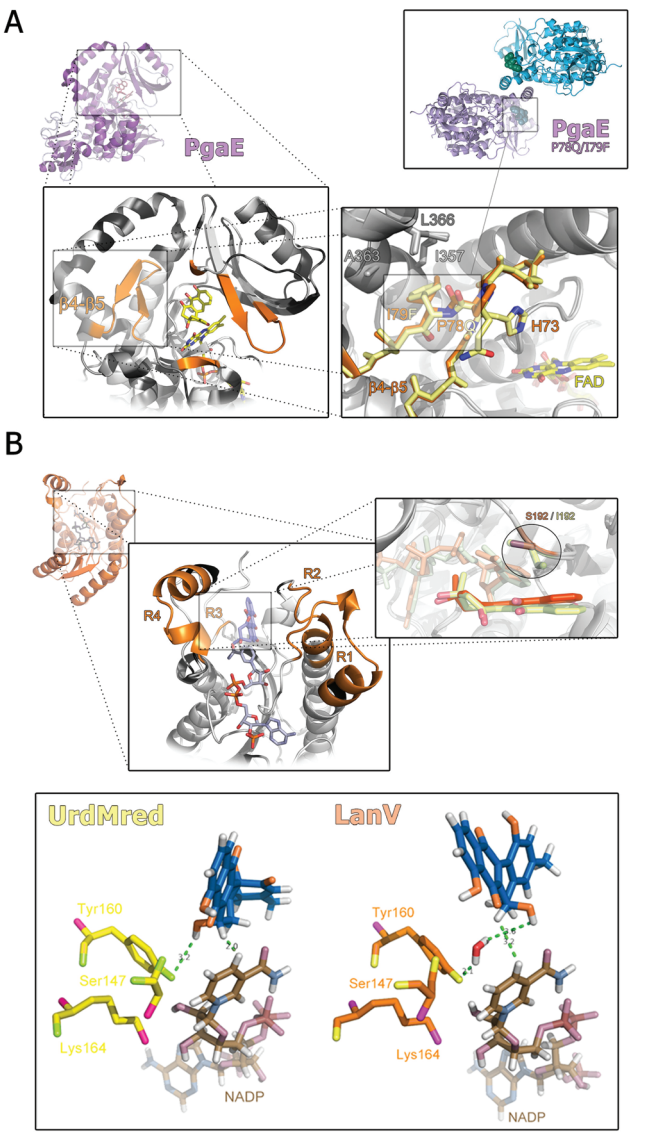

Fig. 4 Engineering angucycline tailoring enzymes. Chimeragenesis of (a) PgaE reveals the importance of P78 and I79, which reside at the dimerization interface, for substrate inhibition and 12b-hydroxylation. (b) LanV suggests the importance of S192 in the switch in activity, but the stereochemical outcome of 6-ketoreduction is likely determined by conformations of the substrates in UrdMred and LanV.

12-hydroxylation. However, none of the engineered proteins gained the ability of native JadH to catalyse 4a,12b-dehydration.

The region containing two short β-sheets, β4 and β5, had the most drastic effect on activity. Dissection of the area suggested that the effect was mediated by two residues, since the double mutant PgaE P78Q/I79F had significantly reduced 12b-hydroxylation activity. Similarly to the DnrK engineering,[10] it was surprising to discover that both P78 and I79 were distal from the active site and in effect the side chains pointed towards the dimerization interface (Fig. 4A).[14] However, this provides an explanation for the simultaneously observed effects in substrate inhibition, where the effect could be mediated to the other subunit of the dimer upon substrate binding. Typically substrate inhibition is associated with regulation of enzymatic activity, but in PgaE we believe it is rather a consequence of the ability of the protein to catalyse C-12b hydroxylation. The elusive 4a,12b-dehydration activity is most likely determined by even more distal secondary-shell interactions from the active site, but discovering such structure/function correlations still present key challenges for protein engineering.

Conformation of substrates determines stereochemistry of 6-ketoreduction in UrdMred and LanV

The landomycin pathway contains an SDR enzyme LanV that has diverged from Pga/UrdMred in that it prefers an earlier pathway intermediate **7** as a substrate and catalyses the ketoreduction with opposite 6*R*-configuration to generate **12** (Fig. 3).[15] The outcome is solely determined by LanV, since the flavoenzyme LanE is capable of performing 12b-hydroxylation alike PgaE.[12] It is curious to note that during formation of **12**, 4a,12b-dehydration is required akin to JadH, but the presence of *any* of the flavoenzymes appears to be sufficient for this.[12]

LanV is a canonical SDR enzyme, which contains a Rossmann-fold core to facilitate NADPH binding and four distinct loop regions that form the active site cleft (Fig. 4b, orange).[15] Exchanging these regions into the scaffolds of either of the two archetypes resulted in the gradual switch in enzymatic activity, with quadruple chimeras reaching 80–90% conversion.[16] The most distinct effect of 30% was cause by a single residue serine *vs.* isoleucine, where ternary crystal structures with the substrate analog rabelomycin established that the longer side-chain of I192 pushes the substrate into a different angle above the NADPH (Fig. 4b).[16]

In all cases, the switch in activity was linked tightly to the stereochemical outcome of ketoreduction and no landomycinones with 6*S*-configuration could be observed. Molecular modelling using density functional theory revealed that the shapes of the reaction products were highly different; in **9** the angular ring is nearly perpendicular to the naphthoquinone, whereas **12** is much more planar due to the aromatization of the angular ring.[16] We concluded, with support from docking calculations (Fig. 4b), that the stereochemistry is effectively determined by the shapes of the substrates and that the only difference between the enzymes is their affinities for the two substrates **7** or **8**.

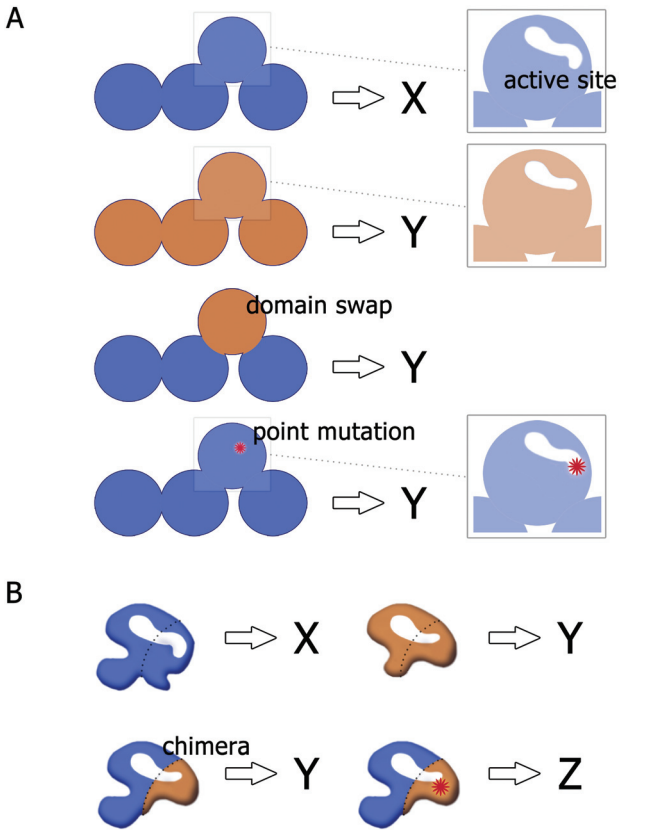

Fig. 5 Strategies for engineering natural product biosynthesis enzymes. (a) Modular polyketide synthases consist of individual domains that provided the required catalytic activities in a conveyer belt fashion. In the example, the blue and brown proteins produce distinct metabolites Y and X due to differences in the active sites that determine substrate specificities. The activities of the proteins have been altered either by exchanging whole domains or probing the active site by point mutations. (b) In chimeragenesis, the shape of the active site is altered by exchanging larger segments of the proteins. In order to obtain novel metabolites such as Z, chimeragenesis may be utilized to identify key regions, which may then be further modified by mutagenesis.

Conclusions

The concept of increasing the diversity of natural products by protein engineering came into existence soon after the discovery of the biosynthetic logic of microbial secondary metabolites in the 1990s. Manipulation of the conveyor belt machinery of modular polyketide synthases has been particularly extensive,[17] with two major methods in wide use (Fig. 5a). In module exchanges, entire protein domains, with their active sites intact, are engineered from related systems to alter the products of the pathways. Alternatively, individual amino acids have been targeted inside the active sites to reach the same outcome. Point mutations have also been utilized extensively to alter the cyclization patterns of type III polyketides and terpenes.[18] The chimeragenesis approach lies in between these strategies and aims to mould the active site by fusing together larger elements from two proteins (Fig. 5b). The methodology enables alteration of the substrate specificities, which is in

contrast to domain swaps that relies on the promiscuity of the native enzymes to function in an unnatural environment.[17] Conversely, chimeragenesis is also able to reveal more distal protein–ligand interactions, which may have great importance for catalysis and are difficult to probe by single point mutations. In addition to our work, a related chimeragenesis engineering approach has been utilized to uncover the differences between *O*- and *C*-glycosyl transferases involved in angucycline biosynthesis.[19]

Our chimeragenesis experiments have also yielded detailed insight into the evolution of natural product biosynthesis pathways. Traditional views on molecular evolution emphasize the importance of the gene duplication event, which allows one copy to carry on serving in the ancestral function.[20] However, in secondary metabolism such restriction may not apply, since the original role is not essential and the selective advantage may be a rare event that occurs only under specific conditions. *Streptomyces* typically harbour more than 30 gene clusters for production of natural products, but only a subset of these are active at a given time and many require specific environmental triggers for activation.[3] Therefore the gene clusters may be considered as selfish genes that search for conditions where they can provide a selective advantage, in particular as horizontal gene transfer between species is abundant.[7] The great diversity of natural products may also be explained by rapid evolutionary rates of secondary metabolism genes, since a correlation between high translational activity of a gene and slow evolutionary rate has been inferred.[21]

How have the new protein functions arisen? Our work supports that hidden secondary activities of ancestral proteins are important. For instance, we believe that the 10-decarboxylation activity of DnrK has provided the seed for the 10-hydroxylation activity to evolve in RdmB.[10] The work also provides excellent examples of the serendipitous nature of evolution; in PgaE the shifts in catalysis are mediated by distal effects,[14] while the reversal of stereochemistry in landomycins is not even mediated by the biosynthetic enzymes.[16] This provides a formidable challenge for the protein engineer trying to rationally design novel catalysts.

Finally, it should be noted that our examples support the view that secondary metabolism enzymes can be classified as "generalists", proteins that display slow reaction rates and substrate promiscuity.[22] In primary metabolism, these would be considered intermediates in the evolutionary pathway towards optimization of protein function in purifying selection.[20] However, in secondary metabolism the enzyme may not ever reach such specialization as the selective pressure is towards chemical diversity, which would be more difficult to reach if the enzymes were highly efficient and specific. Our results are also in agreement with the Screening Hypothesis,[23] which essentially dictates that the producing organisms need to be able to generate numerous natural products in order to discover a rare compound harbouring biomolecular activity, which would provide the selective advantage.

The benefit of these observations is that it would indicate that secondary metabolism enzymes should be highly malleable to protein engineering. However, in all of our test cases the evolutionary approach led to direct interconversion of enzymatic activities and the subfunctionalities (*e.g.* substrate and sterospecificity in 6-ketoreductases)[16] were always tightly coupled. Therefore in order to artificially evolve completely novel functions, the much more challenging task of finding new mutations not observed previously in nature is likely to be required. The usefulness of chimeragenesis in this endeavour would be to be able to pinpoint the evolutionary hotspots responsible for functional differentiation (Fig. 5b), which could then be targeted in a more focused manner using either structure-based protein engineering or directed evolution such as iterative saturation mutagenesis.[24]

Acknowledgements

The author would like to acknowledge funding from the Academy of Finland (grant number 285971) to his laboratory.

Notes and references

1 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2012, **75**, 311–335.
2 (*a*) A. Das and C. Khosla, *Acc. Chem. Res.*, 2009, **42**, 631–639; (*b*) T. Robbins, Y.-C. Liu, D. E. Cane and C. Khosla, *Curr. Opin. Struct. Biol.*, 2016, **41**, 10–19; (*c*) A. M. Gulick, *Curr. Opin. Chem. Biol.*, 2016, **35**, 89–96.
3 P. J. Rutledge and G. L. Challis, *Nat. Rev. Microbiol.*, 2015, **13**, 509–523.
4 W. Mitchell, *Curr. Opin. Chem. Biol.*, 2011, **15**, 505–515.
5 J. C. Way, J. J. Collins, J. D. Keasling and P. A. Silver, *Cell*, 2014, **157**, 151–161.
6 C. A. Rabinovitch-Deere, J. W. K. Oliver, G. M. Rodriquez and S. Atsumi, *Chem. Rev.*, 2013, **113**, 4611–4632.
7 M. A. Fischbach, C. T. Walsh and J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 4601–4608.
8 (*a*) V. Siitonen, B. Selvaraj, L. Niiranen, Y. Lindqvist, G. Schneider and M. Metsä-Ketelä, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 5251–5256; (*b*) V. Siitonen, B. Blauenburg, P. Kallio, P. Mäntsälä and M. Metsä-Ketelä, *Chem. Biol.*, 2012, **19**, 638–646; (*c*) A. Jansson, J. Niemi, Y. Lindqvist, P. Mäntsälä and G. Schneider, *J. Mol. Biol.*, 2003, **334**, 269–280.
9 A. Jansson, H. Koskiniemi, P. Mäntsälä, J. Niemi and G. Schneider, *J. Biol. Chem.*, 2004, **279**, 41149–41156.
10 T. Grocholski, P. Dinis, L. Niiranen, J. Niemi and M. Metsä-Ketelä, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 9866–9871.
11 M. K. Kharel, P. Pahari, M. D. Shepherd, N. Tibrewal, S. E. Nybo, K. A. Shaaban and J. Rohr, *Nat. Prod. Rep.*, 2012, **29**, 264–325.
12 (*a*) P. Kallio, P. Patrikainen, J.-P. Suomela, P. Mäntsälä, M. Metsä-Ketelä and J. Niemi, *Biochemistry*, 2011, **50**, 5535–5543; (*b*) P. Patrikainen, P. Kallio, K. Fan, K. D. Klika,

K. A. Shaaban, P. Mäntsälä, J. Rohr, K. Yang, J. Niemi and M. Metsä-Ketelä, *Chem. Biol.*, 2012, **19**, 647–655.
13 Y. Chen, K. Fan, Y. He, X. Xu, Y. Peng, T. Yu, C. Jia and K. Yang, *ChemBioChem*, 2010, **11**, 1055–1060.
14 P. Kallio, P. Patrikainen, G. A. Belogurov, P. Mäntsälä, K. Yang, J. Niemi and M. Metsä-Ketelä, *Biochemistry*, 2013, **52**, 4507–4516.
15 P. Paananen, P. Patrikainen, P. Kallio, P. Mäntsälä, J. Niemi, L. Niiranen and M. Metsä-Ketelä, *Biochemistry*, 2013, **52**, 5304–5314.
16 P. Patrikainen, L. Niiranen, K. Thapa, P. Paananen, P. Tähtinen, P. Mäntsälä, J. Niemi and M. Metsä-Ketelä, *Chem. Biol.*, 2014, **21**, 1381–1391.
17 K. J. Weissman, *Nat. Prod. Rep.*, 2016, **33**, 203–230.
18 (*a*) I. Abe, *Curr. Opin. Chem. Biol.*, 2012, **16**, 179–185; (*b*) S. C. Hammer, P. O. Syrén, M. Seitz, B. M. Nestl and B. Hauer, *Curr. Opin. Chem. Biol.*, 2013, **17**, 293–300.
19 (*a*) J. Härle, S. Günther, B. Lauinger, M. Weber, B. Kammerer, D. L. Zechel, A. Luzhetskyy and A. Bechthold, *Chem. Biol.*, 2011, **18**, 520–530; (*b*) H. K. Tam, J. Härle, S. Gerhardt, J. Rohr, G. Wang, J. S. Thorson, A. Bigot, M. Lutterbeck, W. Seiche, B. Breit, A. Bechthold and O. Einsle, *Angew. Chem., Int. Ed.*, 2015, **54**, 2811–2815.
20 G. C. Conant and H. W. Kenneth, *Nat. Rev. Genet.*, 2008, **9**, 938–950.
21 D. A. Drummond, A. Raval and C. O. Wilke, *Mol. Biol. Evol.*, 2006, **23**, 327–337.
22 A. Bar-Even and D. S. Tawfik, *Curr. Opin. Biotechnol.*, 2013, **24**, 310–319.
23 R. D. Firn and C. G. Jones, *Nat. Prod. Rep.*, 2003, **20**, 382–391.
24 C. G. Acevedo-Rocha, S. Hoebenreich and M. T. Reetz, *Methods Mol. Biol.*, 2014, **1179**, 103–128.