# Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation†

Annabel C. Murphy,[a] Hui Hong,[a] Steve Vance, [a,b] R. William Broadhurst,[a] and Peter F. Leadlay*[a]

[a] *Department of Biochemistry, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom*

[b] *Current address: Crescendo Biologics Ltd, Meditrina Building 260, Babraham Research Campus, Cambridge CB22 3AT, United Kingdom*

† Electronic supplementary information (ESI) available: Figs. S1–S4; Tables S1–S4, full experimental details and procedures.

## Abstract

Type I modular polyketide synthase multienzymes, which catalyze the assembly-line biosynthesis of an impressive array of bioactive natural products, use a different β-ketoacyl-acyl carrier protein synthase domain for each cycle of chain elongation. Despite intensive study, the extent to which the KS domains determine the selectivity of chain extension remains poorly understood. The KS domains of the PKS for biosynthesis of the toxin mycolactone produced by *Mycobacterium ulcerans*

appear uniquely permissive because they link diverse substrates yet they have essentially identical amino acid sequences. Using an *in vitro* model system based on the KS domain of module 3 of the erythromycin PKS, we have examined the effect of altering individual active site residues to their counterparts in the mycolactone PKS. While other mutations had little effect on the native selectivity of carbon-carbon bond formation, alteration of a specific alanine residue to tryptophan led to an emphatic increase in the turnover of a range of substrates. This residue lies at the boundary between the buried active site and a surface loop whose sequence is notably variable among PKS KS domains, and which has previously been implicated in the recognition of the acyl carrier protein substrates. These results encourage the view that the efficiency of engineered PKSs can be improved by targeted mutagenesis.

## Introduction

The manipulation of biosynthetic pathways to provide access to valuable chemicals is an important strand of the emerging field of synthetic biology.[1,2] Manipulating biosynthesis provides a useful source of novel analogues of pharmaceutically-important, complex natural products,[3-8] as well as cost-effective and sustainable routes to known compounds.[9,10] Modular type I assembly-line polyketide biosynthesis in particular provides an attractive platform for producing rationally engineered biocatalysts that can generate organic molecules of specified shape and size.[11]

In the biosynthesis of complex reduced polyketides, such as the antibacterial erythromycin A and the Buruli ulcer toxin mycolactone (Fig. 1), by actinomycete bacteria, carbon chains of specific length are produced from small monomers on giant assembly-line modular polyketide synthase (PKS) multienzymes, each successive module catalysing a different cycle of chain extension.[12,13] The key step of carbon-carbon bond formation involves a thioester-templated Claisen condensation reaction, catalyzed by a ketosynthase (KS), between the growing chain tethered to the KS active site, and a chain-extending monomer borne on an acyl carrier protein (ACP) domain. Selection and loading of the chain-extending monomer is carried out by an acyltransferase (AT) domain. The β-ketoacyl-ACP intermediate resulting from condensation may undergo reduction before transfer to the next module, the degree of reduction depending on the presence of ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) enzymes in each respective module. Many examples of domain or module replacement, insertion and deletion to give functional chimæric PKSs have been described.[14-16] Unfortunately these chimæric or hybrid modular PKSs are often much less efficient than the parent native PKS.[17] Early model studies showed that KSs have some intrinsic tolerance for different substrates.[18,19] However, the inherent substrate specificity of the KS in the adjacent downstream module may often limit the activity of hybrid modular PKS systems.[20-22] The role of KS specificity in assembly-line polyketide biosynthesis has also been previously explored in so-called *trans*-AT PKSs, which lack in-built AT domains in each module. These PKSs frequently have unusual domain order, and they appear to have

followed a different evolutionary path to the canonical *cis*-AT PKSs. Bio-informatic analysis of their KS sequences has shown that they are significantly more variable (typically showing mutual sequence identity of 40-45%) than KS domains from *cis*-AT PKSs and they form distinct clades corresponding to the chemistry of their substrates.[23] This structure-selectivity correlation, which has been confirmed by *in vitro* functional analysis[24-27] of several KS domains from *trans*-AT systems, implies that native *trans*-AT KS domains may be generally poor catalysts for extension of non-natural chains. In contrast, the sequences of KS domains from a given *cis*-AT PKS tend to form a single clade sharing up to 65% sequence identity, irrespective of the chemical nature of the substrate at each stage of elongation.[23] Although apparently more promising as catalysts operating within hybrid PKS assemblies, the determinants of KS active site specificity in the canonical *cis*-AT systems remain rather poorly understood.[16,17] For both *cis*-AT (canonical) and *trans*-AT PKSs, a better understanding of individual structural features that determine specificity would greatly assist the re-engineering or directed evolution of chimæric systems to improve function.

A valuable framework for detailed analysis of KS specificity has been provided by X-ray crystal structures determined for several individual KS domains from PKS extension modules. For *cis*-AT PKSs, the structures of the KS5AT5[28] and KS3AT3[29] didomains from the erythromycin-producing PKS (DEBS) are available, while in total nine KS domain structures from *trans*-AT PKSs have been reported, including a KS involved in vinylogous chain-branching on the rhizoxin PKS,[30] the KS domain from the

second module of the bacillaene PKS,[25] and seven KS domains from the *trans*-AT PKSs for *iso*-migrastatin and oxazolomycin.[31] Both *cis*-AT and *trans*-AT KS domains crystallise as homodimers, each subunit adopting the fold of the thiolase superfamily[32] in which two mixed five-stranded β sheets are surrounded by α helices. The core of the domain is highly conserved between all polyketide synthase KS domains, and each active site lies near the dimer interface, and is contributed to by residues from the second subunit. The sequence and structural differences between KS domains are largely concentrated in a few surface loops,[25,28-31] as illustrated in the sequence alignment of Fig. 2. From their large-scale structure-based sequence alignment of both *cis*-AT and *trans*-AT KS domains Lohman et al.[31] identified a "clasping loop" located between β-sheet 2 and β-sheet 3; and a "dimer interface loop" located between β-sheet 5 and α-helix 4; and they also highlighted the variability of a surface "active site cap" region between β-sheet 7 and β-sheet 8, and discussed the possible impact of these differences on the access of acyl-ACP substrates to the KS active site. In an earlier comparison of the DEBS KS5AT5 and KS3AT3[29] structures, the difference in conformation adopted by one of these loops in the two structures (residues 153-158 in KS3, Fig. 2) was similarly suggested to contribute to the different specificity of KS3 and KS5 for their acyl-ACP substrates. However, the cryo-EM-based model[32] of multienzyme PikAIII, housing intact extension module 5 of the pikromycin PKS, has provided compelling evidence that the "upstream" acyl-ACP4 accesses the KS5 active site for intermodular transfer of the growing polyketide chain by a *different* access point than is used by the

intramolecular ACP5 bearing the methylmalonyl extender unit for KS5-catalysed condensation. These authors showed that acyl-ACP4, covalently tethered in a catalytically-competent arrangement to the N-terminus of PikAIII, interacts with KS α-helices 7 and 8 (residues 284-293 and 316-322, PikAIII numbering) and with the C-terminus of α-helix 3 (residues 140-150). In contrast, methylmalonyl-ACP5 interacts with part of the "clasping loop" (residues 83-91, PikAIII numbering), and with the "dimer-interface loop" (residues 163-174). The location of these residues is also highlighted in Fig. 2. Roles in determining substrate specificity in *trans*-AT KS domains have been plausibly assigned to the amino acid residue found immediately N-terminal of the essential active site Cys.[24,26,27] but there have been no equivalent successes for the KS domains of canonical *cis*-AT modular PKSs. Also, at the outset of our work there was no evidence that engineered sequence changes in cis-AT KSs might improve tolerance for non-natural acyl-ACP substrates.

A new approach to this question is suggested by the existence of a natural *cis*-AT PKS showing an exceptionally high level of sequence identity (>98%) between KS domains in different extension modules, even though the KSs act on substrates that vary significantly both in chain length and chemical functionality.[29] This is the PKS responsible for generating the virulence factor mycolactone (**1**) (Fig. 1) in *Mycobacterium ulcerans*, the causative agent of Buruli ulcer.[34-36] This multienzyme appears to offer a unique natural example of broad KS substrate tolerance. In principle, mycolactone modules might therefore serve as universal building blocks in combinatorial polyketide biosynthesis. Unfortunately, the slow-growing mycolactone

producer is intractable for genetic manipulation,[37] and the PKS is not active upon heterologous expression.[38] Further, insertion of an entire *myc* KS domain into a module of a heterologous PKS assembly line to replace the resident KS is not expected to be effective, because of deleterious effects on KS:AT and KS:ACP protein:protein interfaces.

In order to learn from the remarkable tolerance of *myc* KS domains, we instead have analysed *in silico* the active site differences between modelled mycolactone KS domains and the experimentally-determined crystal structure of the KS3AT3 didomain from the erythromycin PKS (6-deoxyerythronolide B synthase, DEBS).[29] Guided by this comparison, we then replaced specific amino acids in the EryKS3 active site by their mycolactone KS counterparts, and determined the condensation activity of each mutant KS3 enzyme towards a panel of surrogate thioester substrates *in vitro*. We report here that although most of the eight mutants conserved the catalytic properties of the parent enzyme, the replacement of a specific alanine residue by tryptophan markedly improved both catalytic turnover and the ability of the enzyme to act on non-natural substrates, encouraging the view that the performance of chimæric *cis*-AT PKS multienzymes may be improved by active-site engineering. Interestingly, the critical KS residue lies within an active site loop already implicated in forming part of the binding site for the extender unit acyl-ACP substrate.[33]

## Results and discussion

***In silico* comparison between the KS domains of the mycolactone PKS and the EryKS3AT3 crystal structure**

An *in silico* model of the KS domain from MlsA2, which houses a single extension module of the mycolactone PKS, was generated using Phyre2[39] and compared with the EryKS3AT3 didomain crystal structure,[29] to identify amino acid residues likely to define the extended substrate binding pocket of the KS domain. The catalytically essential residues Cys202, His337 and His377 (EryKS3 domain numbering) occupy exactly the same positions as their MlsA2 counterparts. However, as shown in Fig. 3, seven EryKS3 residues (Ala154, Lys155, Phe156, Val173, Ala230, Phe263 and Phe265) are replaced by other amino acids in MlsA2 KS. Except for Ala154 and Phe265, the substitutions are the same in all 16 mycolactone KS domains (K155A, F156Q, V173M, A230T, and F263T). Ala154 is replaced either by Gly or Trp, and Phe265 is either conserved or replaced by Trp. These seven residues were therefore selected for mutagenesis of the EryKS3 domain, to determine their potential role in promoting broader substrate specificity.

Given the previous evidence that certain surface loops influence access of the substrates to the KS active site, we used a sequence alignment of 199 exclusively *cis*-AT PKS KS domains (for details see Supplementary Information), to compute the sequence variability at a given position of a KS domain, expressed as the percentage of sequences where the consensus residue is present (Fig. 4A). This showed clearly that for *cis*-AT KS domains the "dimer-interface loop" (residues 154-164, EryKS3 numbering)[29] is the most prominently variable. We then compared this (Fig. 4B) with

a sequence logo featuring all residues identified in the active site, including those surrounding the two active site entrances. The residues we selected for mutagenesis are shown with red triangles.

**Mass spectrometric assay for *in vitro* ketosynthase activity**

The catalytic competence of recombinant EryKS3AT3 has previously been assessed by pre-incubating the enzyme with a radiolabelled surrogate thioester substrate such as (2*S*,3*R*)-2-methyl-3-hydroxypentanoic acid (as its *N*-acetylcysteamine (NAC) thioester **15**) (Scheme 1) to acylate the KS, and then incubating the acyl-enzyme with methylmalonyl-ACP.[40] The methylmalonyl-ACP condenses with the acyl-KS to form a triketide, which is chemically released with hot alkali at the end of the reaction. Unfortunately such radiochemical assays[18,19] are limited in the range of substrates that can be tested, demand chemical synthesis of radiolabelled products as reference standards, and do not readily permit the identification of side-products. More recent studies on *trans*-AT KS domains have successfully used mass spectrometry to assay both self-acylation of the KS[24,25] and subsequent Claisen condensation.[26] We also chose a mass spectrometric method to directly monitor the formation of ketide-ACP condensation products, starting from an acyl-NAC thioester, recombinant EryACP3 and either methylmalonyl- or malonyl-CoA. Acyl-NAC thioesters are convenient surrogate substrates for EryKS3AT3 even though where $K_m$ values have been reported they are 2-3 orders of magnitude higher than those of the

corresponding acyl-ACP substrates.[19] First, we confirmed that when diketide NAC-thioester **15** was incubated with methylmalonyl-EryACP3 in the presence of EryKS3AT3, as described in the Supplementary Information, a new acyl-EryACP3 species m/z (10646) was formed whose mass corresponded to that of the expected Claisen condensation product (10646). A panel of acyl-NAC thioesters was then synthesised and assayed as substrates for Claisen condensation by KS3. This panel included acyl-thioesters of varying chain length; acyl-thioesters containing each of the functional groups routinely encountered during polyketide biosynthesis (β-keto-, β-hydroxy- and 2-enoyl-thioesters); and acyl-thioesters bearing an alkyl branch. For each substrate, the percentage of ACP bearing the respective ketide condensation product after 1h of incubation was measured (Table 1). Acyltransfer from SNAC thioesters to holo-EryACP3 was an observable background reaction, but this was not accelerated by the presence of KS3AT3, and is presumably the result of direct chemical thioester-thiol exchange. Substrates featuring α,β-unsubstituted thioesters were also susceptible to side reactions involving 1,4-conjugate addition of phosphopantetheine. Of the substrates tested, as well as the diketide thioester **15**, the alkanoyl thioesters **3** and **4,** the C-2 branched thioesters **7** and **8**, and the 2,3-unsaturated thioester **10** yielded modest amounts (2-6%) of β-ketoacyl-ACP species, produced by acylation followed by condensation.

Different mycolactone KS domains efficiently carry out condensation with either malonyl- or methylmalonyl-ACP as extender units[36] so we also sought to evaluate the effect of the chosen active site mutations on the ability of EryKS3AT3 to

catalyze condensation with malonyl-ACP. In fact, even native EryKS3AT3 has been reported to catalyse condensation with malonyl-ACP as the extender unit *in vitro*, albeit less efficiently than with the natural methylmalonyl-ACP substrate.[40] The malonyl-CoA-specific AT from MLS module 9 was cloned and expressed, and used to prepare malonyl-ACP3 (see Fig. S2) *in situ* from *holo*-EryACP3. EryKS3AT3 and a NAC thioester substrate were added to initiate the condensation reaction. Wild type KS3AT3 accepted the same NAC thioester substrates as before and gave similar yields of condensation product under these conditions from both malonyl- and methylmalonyl-ACP after 1h (see Table S4), showing that the KS3 does not discriminate between these two extender units once they are attached to ACP. These results are consistent with previous work demonstrating an enhanced ability of the KS domain of DEBS module 6 to accept non-natural C-2 substituted malonates as extender units for condensation once they are tethered to the ACP domain.[41]

**The point mutation Ala154Trp dramatically improves *in vitro* ketosynthase condensing enzyme activity against several substrates**

Having established the reactivity of wild type EryKS3AT3 with this range of NAC thioester substrates and with two alternative extender units, the mutant EryKS3AT3 enzymes were tested using the same conditions (Table 1, Table S4 and Fig. 6). Most mutations were found to exert little effect on either substrate (SNAc or extender unit) specificity or the overall yield of the reaction, with two notable exceptions. First, the condensation activity was completely abolished in the

EryKS3AT3 mutant F265W, for all substrates tried, although its acyl transferase activity remained intact (data not shown). This residue is modelled (Fig. 3) to lie close in space to the catalytic triad required for KS-catalyzed condensation, and although Trp is tolerated in this position in certain mycolactone KS domains, it may be that here the increased steric bulk of the tryptophan sidechain interferes with either initial acylation of the KS or the condensation reaction itself. In contrast, mutant Ala154Trp showed a markedly increased substrate tolerance, giving β-ketoacyl-ACP product from nearly all SNAC thioesters tested, with the exception of β-ketoacyl SNAC thioesters **18** and **19**. This mutant also showed a significant increase in turnover of substrates compared to the wild type enzyme. To confirm these preliminary results, side-by-side comparison of Ala154Trp with the wild type was repeated with the inclusion of an internal standard to confirm the stability of the ACP-bound Claisen condensation products.[27] Side reactions proved to be less problematic in these experiments and improved turnover was observed for both wild type and mutant EryKS3AT3. However, Ala154Trp continued to be both significantly more promiscuous and a more effective catalyst. Detectable levels of condensation product could be observed for substrates **5** and **7** even for the wild type, but only the mutant gave condensation products from substrates **6**, **9** and **11** – **14** (Fig. 6). Comparative time courses for the wild type and mutant were carried out using substrate **4**, and initial rates were extracted by fitting the data to an equation that allows derivation of initial rates from reaction progress curves[42] (Fig. 7, see SI for details). This confirmed that the initial rate with Ala154Trp is 4.5-fold greater

than with the wild type. The Ala154Gly mutant behaved similarly to the wild type enzyme, even though this substitution is found in certain mycolactone domains, so this is not an effect simply related to the size of the side-chain at this position. However, the observed increase in both substrate tolerance and catalytic efficiency for Ala154Trp suggests an important role for this region of the active site, with modest improvements in substrate tolerance also observed for mutants Lys155Ala and Phe156Gln. Analysis of the EryKS3AT3 crystal structure using Pymol (www.pymol.org) showed that at least local rearrangement of the active site would be necessary to accommodate a Trp side-chain in place of Ala154. Mapping this residue within a multiple sequence alignment of KS domains (Fig. 4A) reveals that it is located at the start of the "dimer-interface loop",[31] a region that shows a distinctive lack of sequence conservation, indicating a possible role in substrate selection. This region is only partly structured in the crystal structure of EryKS3AT3,[28] and it may be that in the Ala154Trp mutant this loop is significantly re-ordered. Attempts to crystallise this mutant are in progress. Not only does residue Ala154 lie within the active site close to the dimer interface, but it is also between two regions that cryo-EM studies have recently implicated in docking interactions between the KS and both its 'upstream' and intramodular ACP[33] (Fig. 2). The exact mechanism by which this mutation modulates both KS specificity and KS:ACP binding interactions remains to be defined. However, this finding provides clear encouragement for further exploration of targeted mutagenesis of residues predicted to shape, or control access to, KS active sites, with a view to expanding the range of acyl-ACPs

accepted. For an acyltransferase (AT) domain in the erythromycin PKS, such an approach using saturation mutagenesis of active site residues has been successful in uncovering a mutant with only a single amino acid change, that improved the tolerance of the AT in favour of a non-natural extender unit in in vivo experiments.[43,44]

## ACKNOWLEDGMENTS

This work was supported by project grants from the Biotechnology and Biological Sciences Research Council (BBSRC) U. K. to P. F. L. (BB/J007250/1), and from the Wellcome Trust to R. W. B. and P. F. L. (094252/Z/10/Z), and a Herchel Smith Postdoctoral Fellowship from the University of Cambridge to A. C. M.

## Notes and References

1 D. E. Cameron, C. J. Bashor and J. J. Collins, *Nat. Rev. Microbiol.*, 2014, **12**, 381.

2 J. M. Winter and Y. Tang, *Curr. Opin. Biotechnol.,* 2012, **23**, 736.

3 H. G. Menzella, J. R. Carney and D. V. Santi, *Chem. Biol.*, 2007, **14**, 143.

4 J. T. Fitzgerald, C. P. Ridley, and C. Khosla, *J. Antibiot.*, 2011, **64**, 759.

5 Y. Xu, T. Zhou, S. Zhang, P. Espinosa-Artiles, L. Wang, W. Zhang, M. Lin, A. A. L. Gunatilaka, J. Zhan and I. Molnár, *Proc. Natl. Acad. Sci.*, 2014, **111**, 12354.

6 K. M. Fisch, W. Bakeer, A. A. Yakasai, Z. Song, J. Pedrick, Z. Wasil, A. M. Bailey, C. M. Lazarus, T. J. Simpson and R. J. Cox, *J. Am. Chem. Soc.*, 2011, **133**, 16635.

7 P. G. Arnison, et al., *Nat. Prod. Rep.* **2013**, *30*, 108–160.

8 R. H. Baltz, *ACS Synth. Biol.*, 2014, **3**, 748.

9 C. J. Paddon, *et al*., *Nature*, 2013, **496**, 528.

10 A. C. Murphy, *Nat. Prod. Rep.*, 2011, **28**, 1406.

11 J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380.

12 A. Hagan, S. Poust, T. de Rond, J. L. Fortman, L. Katz, C. J. Petzold and J. D. Keasling, *ACS Synth. Biol.*, 2016, **5**, 21.

13 C. Hertweck, *Angew. Chem. Int. Ed. Engl.*, 2009, **48**, 4688.

14 H. G. Menzella, R. Reid, J. R. Carney, S. S. Chandran, S. J. Reisinger, K. G. Patel, D. A. Hopwood and D. V. Santi, *Nat. Biotechnol. 2005*, **2***3*, 1171–1176.

15 F. Del Vecchio, H. Petković, S. G. Kendrew, L. Low, B. Wilkinson, R. Lill, **J.** Cortés, B. A. M. Rudd, J. Staunton and P. F. Leadlay, *J. Ind. Microbiol. Biotechnol.*, 2003, **30**, 489.

16 K. J. Weissman and P. F. Leadlay, *Nat. Rev. Microbiol.*, 2005, **3**, 925.

17 K. J. Weissman, *Nat. Prod. Rep.*, 2016, **33**, 203.

18 J. A. Chuck, M. McPherson, H. Huang, J. R. Jacobsen, C. Khosla and D. E. Cane, *Chem. Biol.*, 1997, **4**, 757.

19 N. Wu, S. Y. Tsuji, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2001, **123**, 6465.

20 C. J. Rowe, I. U. Böhm, I. P. Thomas, B. Wilkinson, B. A. Rudd, G. Foster, A. P. Blackaby, P. J. Sidebottom, Y. Roddis, A. D. Buss, J. Staunton, and P. F. Leadlay, *Chem. Biol.*, 2001, **8**, 475.

21 K. Watanabe, C. C. C. Wang, C. N. Boddy, D. E. Cane, C. Khosla, *J. Biol. Chem.*, 2003, **278**, 42020.

22 B. Busch, N. Ueberschaar, S. Behnken, Y. Sugimoto, M. Werneburg, N. Traitcheva, J. He and C. Hertweck, *Angew. Chem. Int. Ed. Engl.,* 2013, **52**, 5285.

23 T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck and J. Piel, *Nat. Biotechnol.* 2008, **26**, 225.

24 M. Jenner, S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, G. S. Briggs, J. Piel and N. J. Oldham, *Angew. Chem. Int. Ed Engl.*, 2013, **52**, 1143.

25 D. C. Gay, G. Gay, A. J. Axelrod, M. Jenner, C. Kohlhaas, A. Kampa, N. J. Oldham, J. Piel and A. T. Keatinge-Clay, *Structure*, 2014, **22**, 444.

26 C. Kohlhaas, M. Jenner, A. Kampa, G. S. Briggs, J. P. Afonso, J. Piel and N. J. Oldham, *Chem. Sci.*, 2013, **4**, 3212.

27 M. Jenner, J. P. Afonso, H. R. Bailey, S. Frank, A. Kampa, J. Piel and N. J. Oldham, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 1817.

28 Y. Tang, C.-Y. Kim, I. I. Mathews, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 11124.

29 Y. Tang, A. Y. Chen, C.-Y. Kim, D. E. Cane and C. Khosla, *Chem. Biol.*, 2007, **14**, 931.

30 T. Bretschneider, J. B. Heim, D. Heine, R. Winkler, B. Busch, B. Kusebauch, T. Stehle, G. Zocher and C. Hertweck, *Nature*, 2013, **502**, 124.

31 J. R. Lohman, M. Ma, J. Osipiuk, B. Nocek, Y. Kim, C. Chang, M. Cuff, J. Mack, L. Bigelow, H. Li, M. Endres, G. Babnigg, A. Jochimiak, G. N. Phillips Jr. and B. Shen, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **54**, 6842.

32 A. M. Haapalainen, G. Meriläinen and R. K. Wierenga, *Trends Biochem. Sci.*, 2006, **31**, 64.

33 S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 512.

34 T. P. Stinear, A. Mve-Obiang, P. L. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. Johnson, J. K. Davies, R. E. Lee, S. Adusumilli, T. Garnier, S. F. Haydock, P. F. Leadlay and S. T. Cole, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 1345.

35 H. Hong, J. B. Spencer, J. L. Porter, P. F. Leadlay and T. P. Stinear, *ChemBioChem.,* 2005, **6**, 643.

36 H. Hong, C. Demangel, S. J. Pidot, P. F. Leadlay and T. P. Stinear, *Nat. Prod. Rep.*, 2008, **25**, 447.

37 J. L. Porter, N. J. Tobias, S. J. Pidot, S. Falgner, K. L. Tuck, A. Vettiger, H. Hong, P. F. Leadlay and T. P. Stinear, *PLoS One*, 2013, **8**, e70520.

38 J. L. Porter, N. J. Tobias, H. Hong, K. L. Tuck, G. A. Jenkin and T. P. Stinear, *Microbiology*, 2009, **155**, 1923.

39 L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass and M. J. E. Sternberg, *Nat. Protocols*, 2015, **10**, 845.

40 A. Y. Chen, N. A. Schnarr, C.-Y. Kim, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2006, **128**, 3067

41 I. Koryakina, J. B. McArthur, M. M. Draelos and G. J. Williams, *Org. Biomol. Chem.*, 2013, **11**, 4449.

42 W. Cao, and E. M. De La Cruz, *Sci. Rep.,* 2013, **3**, 2658.

43 U. Sundermann, K. Bravo-Rodriguez, S. Klopries, S. Kushnir, H. Gomez, E. Sanchez-Garcia and F. Schulz, *ACS Chem. Biol.*, 2013, **8**, 443.

44 K. Bravo-Rodriguez, S. Klopries, K. R. M. Koopmans, U. Sundermann, S. Yahiaoui, J. Arens, S. Kushnir, F. Schulz and E. Sanchez-Garcia, *Chem. Biol*., 2015, **22**, 1425.

45 L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, *Nat. Prot*., 2015, **10**, 845.

## Figure legends

**Fig. 1.** Chemical structures of Mycolactone (1) and Erythromycin (2)

**Fig. 2.** Sequence alignment of KS domains of known structure with mycolactone KS domains. The positions of variable surface loops identified in X-ray crystal structures[25,28-31] are highlighted with the 'outer clasping loop' in yellow, the 'dimer interface loop' in blue and the 'active site cap' in red. Regions previously identified as involved in ACP docking events[33] are outlined in yellow for the "upstream" ACP and green for the intramodular ACP. Residues identified as possible specificity determinants[25] are highlighted with a closed circle. Residues that when mutated obstruct the intramodular ACP docking region[33] are indicated with open triangles. The seven EryKS3 residues mutated in this work are indicated with asterisks.

**Fig. 3.** Active site residues chosen for mutagenesis in the ketosynthase (KS) domain of Ery PKS (DEBS) module 3 (PDB accession number 2QO3) based on comparison with a Phyre2[39] model of a representative mycolactone PKS KS (MlsA2). Conserved active site residues are shown in yellow while those that differ are shown in blue. Residue Val173' lies at the KS-AT dimer interface and is contributed by the other subunit. The catalytic triad is shown in red and electron density observed for the ketosynthase inhibitor cerulenin bound to the catalytic Cys is shown in purple.

**Fig. 4.** Sequence variability in *cis*-AT PKS domains. (A) The percentage of sequences (from 199 aligned PKS domains, see Supplementary Information for details) that have the consensus residue at each position is plotted against residue number. The

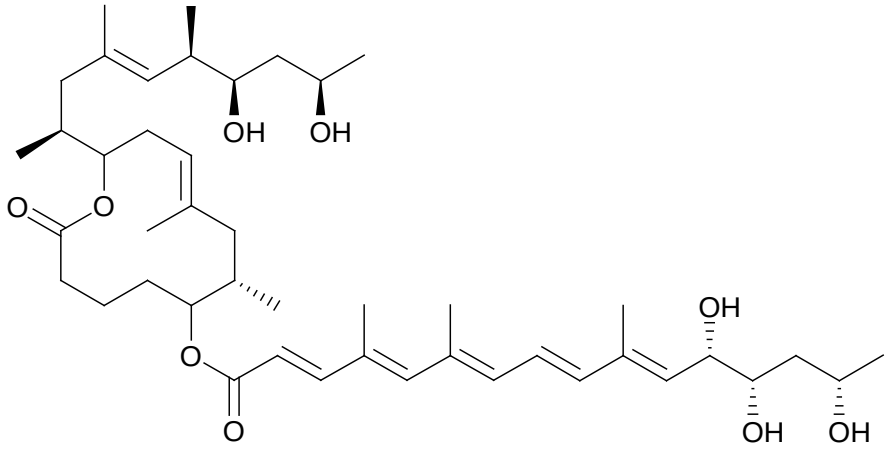
regions shown in colour correspond to "variable surface loops" identified in X-ray crystal structures of both *trans*-AT and *cis*-AT KS domains.[25,28-31] Yellow = outer clasping loop; blue = dimer interface loop; and red = active site cap. (B) A selective sequence logo featuring all residues identified in the active site (shown in green), including those surrounding the two active site entrances (modified from a sequence logo generated by Weblogo3).[45] Residues selected for mutagenesis in this work are indicated with a red triangle (numbering is for EryKS3[29]).

**Fig. 5.** Acyl NAC-thioesters were used as substrates for the condensing enzyme activity of EryKS3AT3. Either methylmalonyl-CoA or malonyl-CoA was used as the source of extender unit. ACP = EryACP3, KS-AT = EryKS3AT3, AT = MlsAT9.

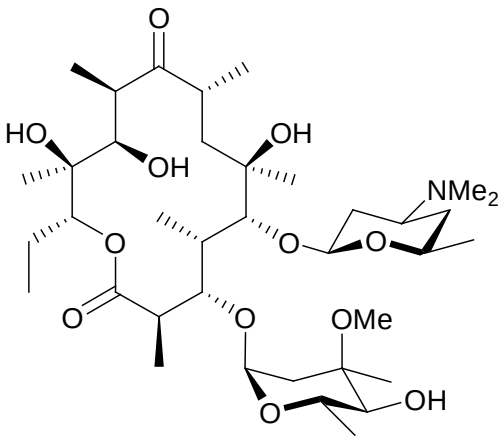
**Fig. 6.** Selected MS analyses of KS-catalysed production of β-ketoacyl-EryACP3. EryACP3 after incubation with methylmalonyl-CoA (MeMal-CoA), substrate **6** and (A) wild type EryKS3AT3 or (B) Ala154Trp; ACP3 after incubation with MeMal-CoA, substrate **11** and (C) wild type EryKS3AT3 or (D) Ala154Trp; ACP3 after incubation with MeMal-CoA, substrate **13** and (E) wild type EryKS3AT3 or (F) Ala154Trp; ACP3 after incubation with MeMal-CoA, substrate **14** and (G) EryKS3AT3 or (H) Ala154Trp. Claisen condensation products for the respective substrates are shown in green, while side products (non-enzyme catalysed) arising from thioester exchange of ACP with SNAc substrate and conjugate addition of ACP onto α,β-unsaturated thioester substrates are shown in purple and orange respectively. An alkylated ACP used as an internal standard is shown in yellow.

**Fig. 7.** Time course of condensation of **4** with methylmalonyl-ACP, catalysed by wild type EryKS3AT3 and mutant Ala154Trp.

**Table 1**. Percentage of EryACP3 bearing Claisen condensation product after incubation with EryKS3AT3 or its mutants, methylmalonyl CoA and synthetic SNAc substrate. T indicates that only trace levels were detected (see SI for details).




1




2

Fig. 1

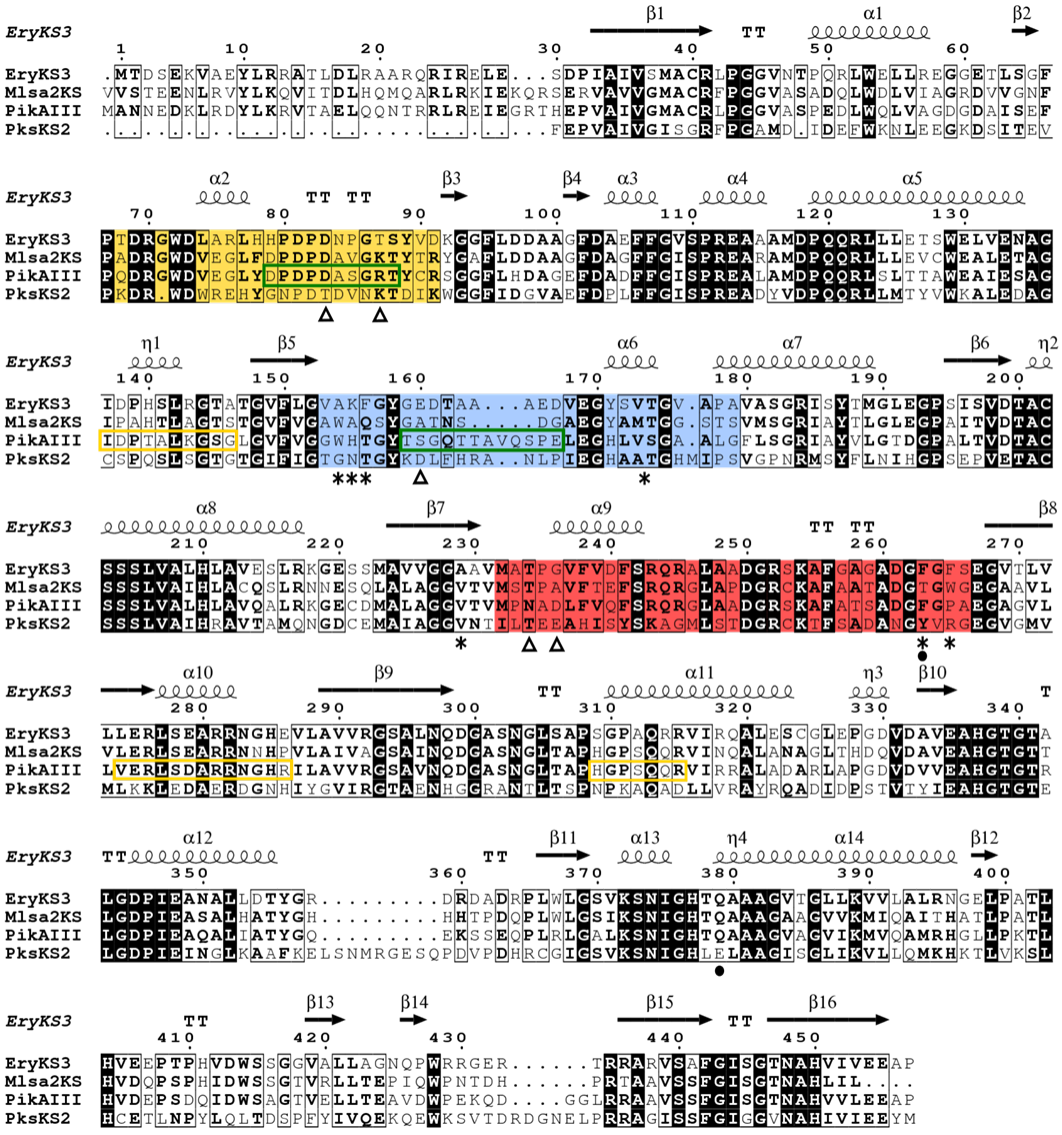



Fig. 2

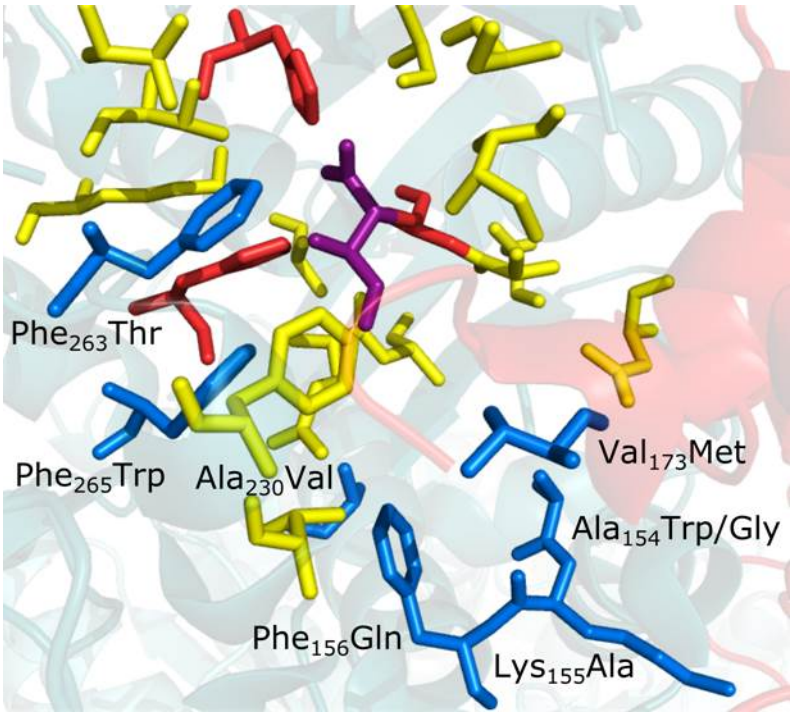

Phe263Thr
Phe265Trp
Ala230Val
Val173Met
Ala154Trp/Gly
Phe156Gln
Lys155Ala

Fig. 3

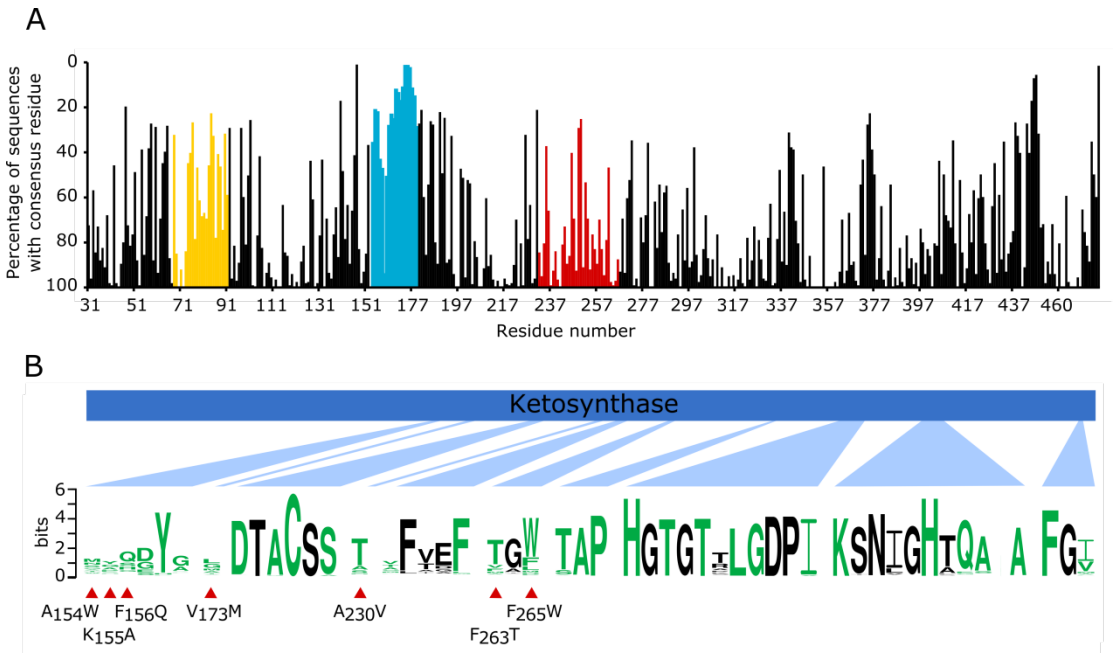

A
Percentage of sequences with consensus residue
0
20
40
60
80
100
31 51 71 91 111 131 151 177 197 217 237 257 277 297 317 337 357 377 397 417 437 460
Residue number
B
Ketosynthase
bits
6
4
2
0
A154W
K155A
F156Q
V173M
A230V
F263T
F265W

Fig. 4

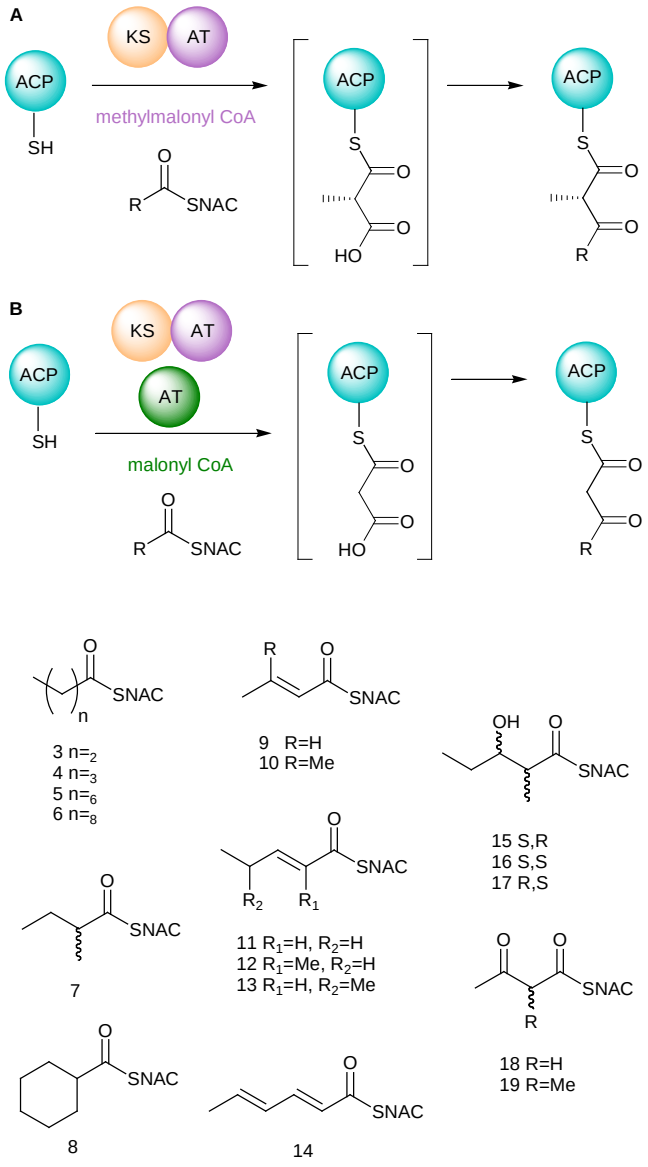

A
KS
AT
ACP
SH
methylmalonyl CoA
O
R
SNAC
ACP
S
O
O
HO
ACP
S
O
O
R
B
KS
AT
AT
ACP
SH
malonyl CoA
O
R
SNAC
ACP
S
O
O
HO
ACP
S
O
O
R
O
SNAC
n
3 n=2
4 n=3
5 n=6
6 n=8
R
O
SNAC
9 R=H
10 R=Me
OH
O
SNAC
15 S,R
16 S,S
17 R,S
O
SNAC
R2
R1
11 R1=H, R2=H
12 R1=Me, R2=H
13 R1=H, R2=Me
O
SNAC
7
O
O
SNAC
R
18 R=H
19 R=Me
O
SNAC
8
O
SNAC
14

Fig. 5

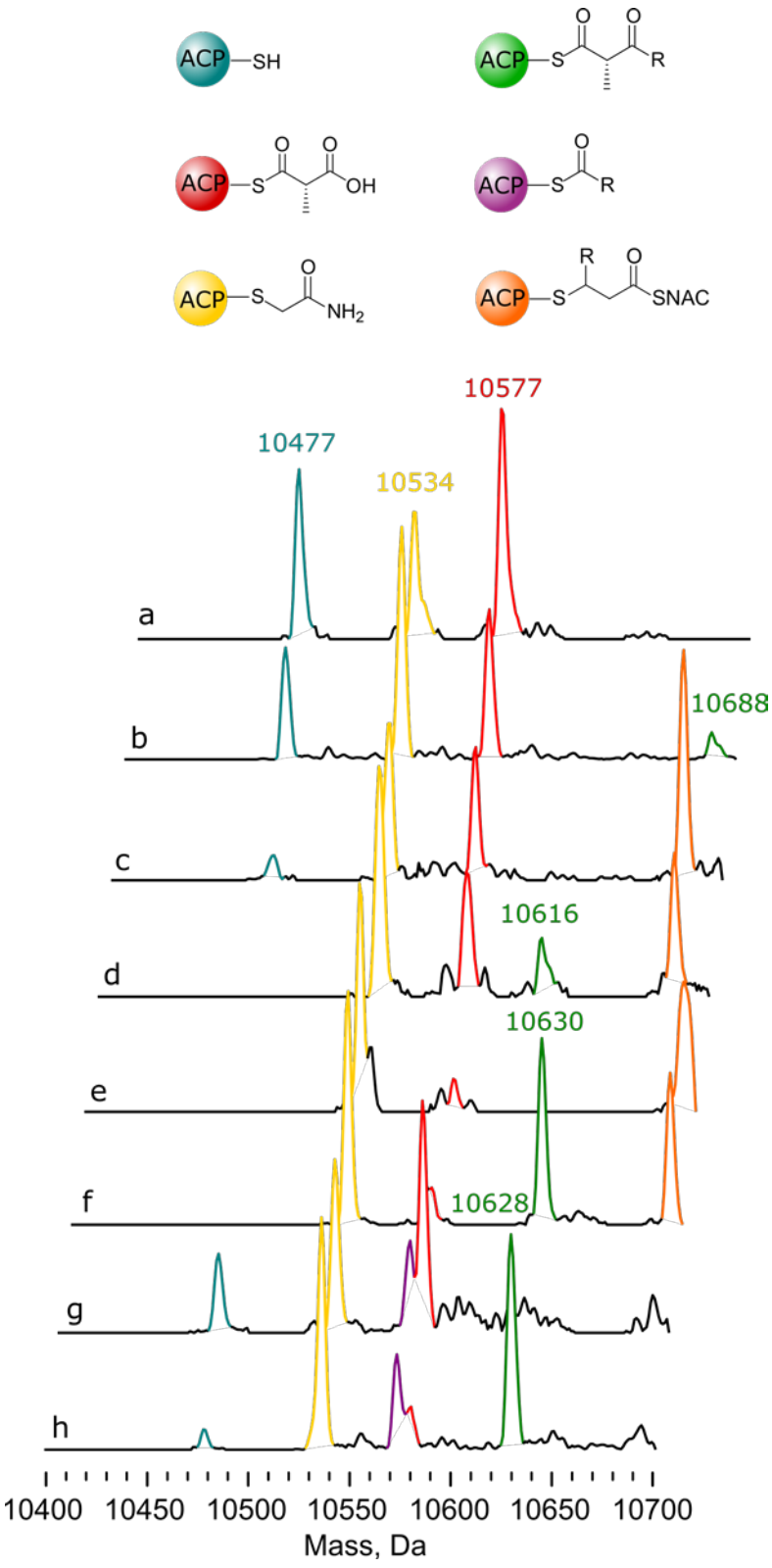

ACP
SH
ACP
S
O
O
R
ACP
S
O
O
OH
ACP
S
O
R
ACP
S
O
NH2
ACP
S
R
O
SNAC
10577
10477
10534
a
10688
b
c
10616
d
10630
e
f
10628
g
h
10400
10450
10500
10550
10600
10650
10700
Mass, Da

Fig. 6.

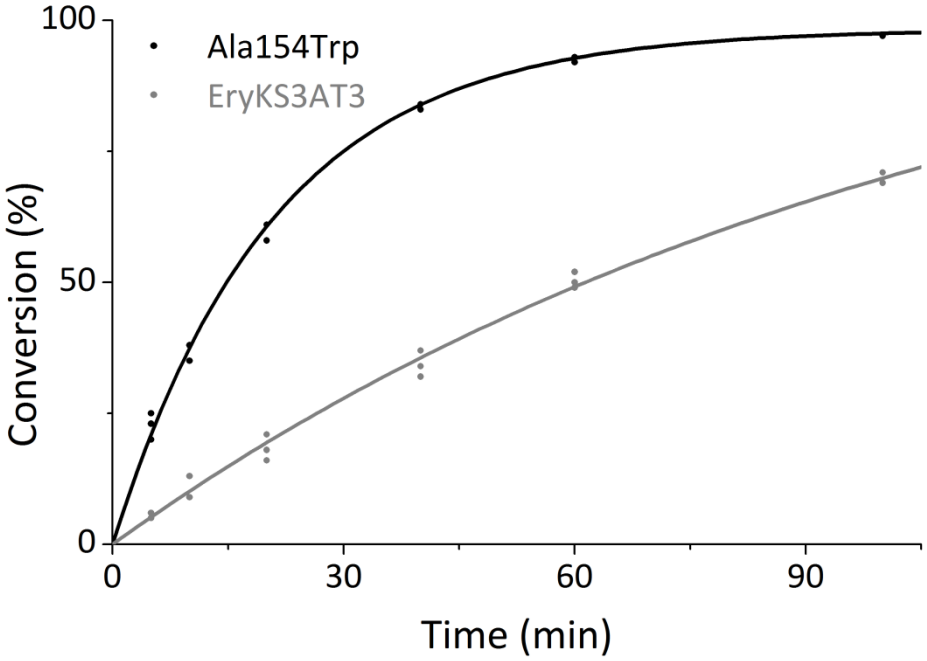



Fig. 7

| Substrate | eryKS3AT3 | $A_{154}G$ | $A_{154}W$ | $K_{155}A$ | $F_{156}Q$ | $V_{173}M$ | $A_{230}T$ | $F_{263}T$ | $F_{265}W$ |
|---|---|---|---|---|---|---|---|---|---|
| **3** | 4 | 4 | 46 | 3 | T | 6 | 4 | T | ND |
| **4** | 3 | 6 | 83 | 6 | 7 | 4 | 2 | 23 | ND |
| **5** | ND | ND | 77 | T | 3 | ND | ND | T | ND |
| **6** | ND | ND | 13 | ND | ND | ND | ND | ND | ND |
| **7** | 1 | 5 | 20 | 2 | T | T | T | T | ND |
| **8** | 2 | 2 | 72 | T | 5 | 2 | 1 | 4 | ND |
| **9** | ND | ND | 2 | ND | ND | ND | ND | ND | ND |
| **10** | 6 | 17 | 91 | 16 | 18 | 11 | 8 | 10 | ND |
| **11** | ND | 1 | 5 | ND | ND | ND | ND | ND | ND |
| **12** | ND | ND | 2 | ND | ND | ND | ND | ND | ND |
| **13** | ND | ND | 24 | ND | ND | ND | ND | ND | ND |
| **14** | ND | ND | 33 | ND | ND | ND | ND | ND | ND |
| **15** | 2 | 3 | 7 | 5 | 2 | 2 | T | 4 | ND |
| **16** | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| **17** | ND | ND | T | T | T | ND | ND | ND | ND |
| **18** | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| **19** | ND | ND | ND | ND | ND | ND | ND | ND | ND |

**Table 1.**