Characterization of the thioesterase domain of the Bacillaene pathway

Mark Horsman

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the degree of Doctor of Philosophy in Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Table of Contents

List of Figures

Abstract

The goals of this work are to advance the understanding of the mechanisms of thioesterase loading-substrate and release-nucleophile selectivity and to develop an analytical methodology and modelling theory as tools for quantitative measures of thioesterase activity. The first aspect of this work is a bioinformatic examination of examples of thioesterases that catalyze different forms of release at different rates of reaction dependent on the type of thioester substrate presented to it. In addition, the type I thioesterase domain from the Type I modular polyketide synthase pathway responsible for the production of bacillaene was cloned, expressed, and purified for kinetic characterization as a model *trans*-AT type thioesterase. N-Acetylcysteamine and acyl carrier protein-bound acyl substrates were synthesized as substrates for hydrolysis catalysed by the bacillaene thioesterase. The rate of reaction was indirectly observed with a widely-adopted visual spectroscopy assay facilitated by the thiol indicator 5,5'-dithio-bis-[2-nitrobenzoic acid]. Variants of the bacillaene thioesterase coded from two different bacterial isolates with 57 % amino acid identity (73 % similarity) both exhibited rapid declining activity under the conditions used for Ellman's, indicative of inhibition. This matches rare literature examples of other thioesterase domains under the same spectroscopic assay conditions. We demonstrate that initial rate of reaction data is insufficient to describe substrate selectivity in these systems and develop a time-dependent model capable of accounting for irreversible substrate inhibition during catalysis. This model was sufficient to reproduce the thioesterase activity observed between the bacillaene thioesterase and methyl valeric-SNAC but is insufficient to model the responses observed during the hydrolysis of *trans*-2-methyl-2-pentenoyl-SNAC or *cis*-3-methyl-2-pentenoyl-SNAC. Finally, we detailed synthetic and assay conditions for the production, cleavage, and detection of the ACP-tethered substrate equivalents. This contributes towards the development of complex thioesterase assays that better estimate type I bacterial thioesterase activity during polyketide synthesis.

Acknowledgements

Graduate funding from NSERC and the University of Ottawa provided to Mark Horsman.

Discovery Grants Program funding to Dr. Christopher Boddy.

Chapter 2 is adapted and updated from the published literature review "Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate?". Mark Horsman and Dr. Christopher Boddy conceived the premise of this article. Mark Horsman performed the literature review. Mark Horsman, Dr. Taylor Hari, and Dr. Christopher Boddy wrote this article.

Technical training and guidance by Dr. Sharon Curtis, Karl Wasslen, Dr. Joern Piel, Dr. Jeffrey Keillor, Dr. Roberto Chica, and Dr. Christopher Boddy.

Extensive and time-intensive feedback from the thesis committee of Dr. Sylvie Garneau-Tsodikova, Dr. Peter Buist, Dr. Adam Shuhendler, Dr. Jeffrey Keillor, and Dr. Christopher Boddy is gratefully acknowledged.

Chapter 1

1.1 Introduction

Polyketides (PKs) and non-ribosomal peptides (NRPs) are important sources of medically relevant natural products that play diverse roles[1] in bacterial and fungal symbiosis,[2, 3] regulation,[4, 5] competition,[6] and ecology.[7]

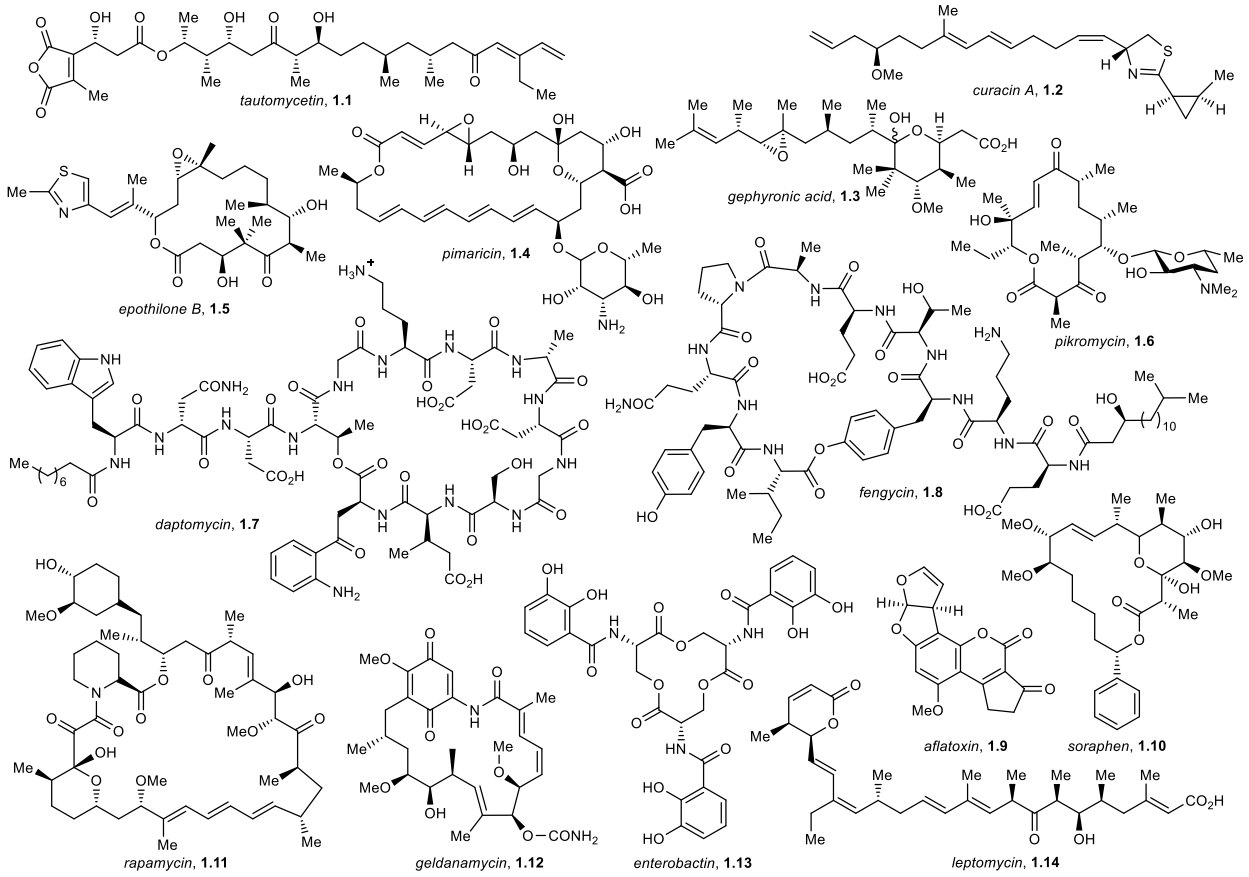


Figure 1-1 Diversity of polyketide and NRP natural products released by thioesterases. Figure adapted from the supplemental information of Horsman *et al.* 2015[8] presented in **Chapter 2** with permission from The Royal Society of Chemistry.

PKs include erythromycin, oxytetracycline, candidicin, amphotericin, and avermectin. NRPs include daptomycin, baxitracin, polymyxin B, gramicidin, vancomycin, and yersiniabactin, and examples of products from hybrid pathways, such as epothilone, are also observed.

PK assembly works by incremental addition/extension and monomer modification. It is built around the action of a limited set of catalytic domains which draw in metabolites from the cellular milieu.

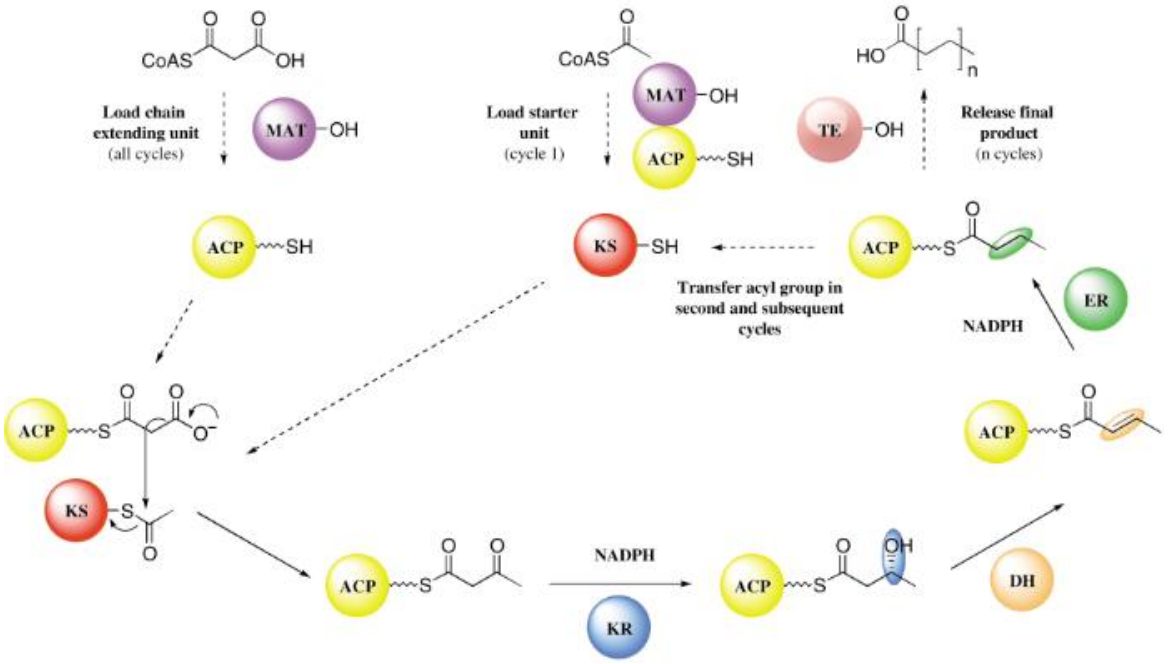


Figure 1-2 Assembly of polyketides through extension. In fatty acid synthesis and interative PKSs, multiple iterations take place on a single set of domains. However, in modular PKSs, the growing acyl chain is passed to a new extending-unit ACP downstream after each extension. MAT: malonyl-acetyl transferase; ACP: acyl carrier protein; KS: ketosynthase; KR: ketoreductase; DH: dehydratase; ER: enoyl reductase. Adapted from Staunton, J. and Weissman, K. J. 2001[9] with permission from The Royal Society of Chemistry.

AT, KS, KR, DH and ER domains draw in metabolites from the cellular milieu and catalyse elongation of an intermediate bound to ACP tethers. Each modification is relatively standard and adds a small element of diversification. The acyl chain thioester tether to the ACP replaces conventional substrate recognition and the chemical structure of the growing chain does not limit the types of reactions the domains perform. As a result of the method of repeated incremental diversification and ACP tethered control, significant diversity can be installed in PKs by catalytic

domains with limited substrate selectivity. Ultimately the complete PK must be cleaved from the ACP it is covalently linked to in order to release the product.

Type I TEs perform the essential role of release of the covalently-bound completed product. Similar to KSs, they are domains that directly capture the metabolite from the succession of ACPs. Release is important for the productivity of the synthase and essential for natural products to diffuse to their site of action. However, the final release stage of polyketide assembly is often also important for the function of the natural product. In a recent sampling of 76 TE-containing secondary metabolite pathways with known final products, 53 % were released as linear free acids, 39 % were released as circular macrolactones, and 8 % were released as macrolactams.[10] Conformational constraint is important to the biological activity of some secondary metabolites,[11] and the choice between different potential internal nucleophiles gives the TE the capacity to sample different constitutional isomers from the same PKS input.

Type I TEs are alpha/beta hydrolases that catalyze these reactions via the catalytic triad named for its most frequent combination as the Ser His Asp catalytic triad. Unlike lipases such as Bacillus lipase EstA,[12, 13] type I TEs are found at the end of PKS polypeptides, and only act on substrates that are transferred to the carrier protein positioned proximal to the TE. Since these enzymes have little implication as off-target hydrolases, stringent substrate recognition and specificity is less important. These pathways have evolved in parallel with type I and II fatty acid synthases and have screened through chemical space through the inactivation of domains (region deletions, point mutations) in the module, module duplication, and domain replacement.[14] It would be evolutionarily advantageous if the TE was capable of releasing a range of substrates as the pathway evolves.[10] TEs have diverged to perform various chemical transformations including hydrolysis, macrocyclization,[15, 16] transacylation,[17] oligomerization,[18] and in unnatural *in vitro* conditions, have been shown to simultaneously perform trace amounts of these different reactions.[10]

However, built into the way TEs catalyze reactions, TEs have some substrate specificities that change which type of release takes place based on which substrate is presented as an input,[8, 19-21]

and in some cases, allows for competing catalysis from other proteins to take place while tethered to the final CP before the TE invokes release.[22, 23] This thesis focuses on the study of TE kinetic selectivity in the thioester substrates it loads and the ways in which the acyl groups are released.

Chapter 2 details the diversity, structure, and mechanism of PKS and NRPS TEs seen in the literature as of 2017. It is structured to include both a general introduction to the type I TE family as well as a critical examination of examples of TEs that catalyze different forms of release at different rates of reaction dependent on the type of thioester substrate presented to it. In particular, it discusses how this selectivity occurs in the TE studied, and how this can be put towards predicting the chemical output of newly discovered or metabolically engineered pathway TEs.

Chapter 3 outlines SNAC substrate reactivity of the TE from the bacillaene pathway. During the synthesis of bacillaene, an unusual conjugated polyene is formed through isomerization and dehydrobacillaene is released through hydrolysis. Bacillaene is a widely-adopted model pathway for *trans*-AT PKSs, and there is little in the way of literature to validate the assumption that TEs in *trans*-AT PKS systems behave like the *cis*-AT equivalents. Additionally, it is unclear whether the bacillaene TE (BaeRTE) is tolerant to bacillaene analogues without the beta, gamma polyene configuration and whether BaeRTE could be used as a generalized hydrolytic off-loading enzyme in engineered pathways. Type I TEs are commonly heterologously expressed as soluble isolated domains and screened against a range of SNAC thioester substrates. However, in the assessment assays presented in **Chapter 3**, the hydrolysis rate decreased before a substantial change in substrate concentrations, and the data reflected examples of unusual TE responses in literature (summarize some of the outcomes). As a result, this assay assessing steady-state properties cannot be directly evaluated to determine the activity and specificity of the TE domain.

Chapter 4 reviews time-dependent models of acyl-SNAC hydrolysis. This starts with a test of the components of the standard Ellman's hydrolysis assay used in **Chapter 3**. The initial substrate concentration is highlighted as a direct factor affecting the rate of hydrolysis inactivation, and fits with the possibility of irreversible substrate inhibition. A time-resolved rate law is developed to

understand the mechanism of inactivation and attempt to delineate between substrate specificity during catalysis and inactivation. The constants from the resulting time-dependent model are fit to the experimental results from **Chapter 3** both through linear regression and non-linear regression directly to the exponential decay. Both methods rendered roughly the same kinetic constants. Finally, the goodness of fit for this model is assessed by modelling individual assay concentrations and comparing the observed and modelled amounts of thiol produced. This model fits the behaviour observed with the saturated 2-methylvaleric-SNAC thioester, but not the two unsaturated substrates studied in parallel. Further work is needed to understand the activity of BaeRTE.

Chapter 5 describes tests of the BaeRTE with more complicated ACP-bound substrates to test whether the presence of the ACP counterpart eliminates inactivation, and whether specific interactions between the BaeRACPIII and TE contribute to substrate recognition, or whether acyl substrates are equally reactive while bound to other ACPs within the *trans*-AT family of pathways. ESI-LCMS detection of intact ACP species was successfully demonstrated (via BaeLACPIII acyl substrates), but BaeRTE was not shown to hydrolyse the BaeRACPIII substrates under the conditions tested.

1.2 References

1. C. P. Ridley, H. Y. Lee and C. Khosla, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 4595-4600.
2. J. Piel, I. Hofer and D. Q. Hui, *Journal of Bacteriology*, 2004, **186**, 1280-1286.
3. E. Freinkman, D. C. Oh, J. J. Scott, C. R. Currie and J. Clardy, *Tetrahedron Letters*, 2009, **50**, 6834-6837.
4. Y. Ohnishi, S. Kameyama, H. Onaka and S. Horinouchi, *Molecular Microbiology*, 1999, **34**, 102-111.
5. T. A. Black and C. P. Wolk, *Journal of Bacteriology*, 1994, **176**, 2282-2292.
6. S. Mueller, S. N. Strack, B. C. Hoefler, P. D. Straight, D. B. Kearns and J. R. Kirby, *Applied and Environmental Microbiology*, 2014, **80**, 5603-5610.
7. J. M. Pang, E. Layre, L. Sweet, A. Sherrid, D. B. Moody, A. Ojha and D. R. Sherman, *Journal of Bacteriology*, 2012, **194**, 715-721.
8. M. E. Horsman, T. P. Hari and C. N. Boddy, *Natural Product Reports*, 2016, **33**, 183-202.
9. J. Staunton and K. J. Weissman, *Natural Product Reports*, 2001, **18**, 380-416.
10. T. P. A. Hari, P. Labana, M. Boileau and C. N. Boddy, *Chembiochem : a European journal of chemical biology*, 2014, **15**, 2656-2661.
11. E. M. Larsen, M. R. Wilson and R. E. Taylor, *Natural Product Reports*, 2015, **32**, 1183-1206.
12. V. Barbe, S. Cruveiller, F. Kunst, P. Lenoble, G. Meurice, A. Sekowska, D. Vallenet, T. Z. Wang, I. Moszer, C. Medigue and A. Danchin, *Microbiology-Sgm*, 2009, **155**, 1758-1775.
13. N. Lenfant, T. Hotelier, E. Velluet, Y. Bourne, P. Marchot and A. Chatonnet, *Nucleic Acids Research*, 2013, **41**, D423-D429.
14. H. Jenke-Kodama, T. Borner and E. Dittmann, *PLOS Computational Biology*, 2006, **2**, 1210-1218.
15. G. L. Luo, R. Pieper, A. Rosa, C. Khosla and D. E. Cane, *Bioorganic & Medicinal Chemistry*, 1996, **4**, 995-999.
16. C. N. Boddy, T. L. Schneider, K. Hotta, C. T. Walsh and C. Khosla, *Journal of the American Chemical Society*, 2003, **125**, 3428-3429.
17. L. Ray, K. Yamanaka and B. S. Moore, *Angewandte Chemie - International Edition*, 2016, **55**, 364-367.
18. C. A. Shaw-Reid, N. L. Kelleher, H. C. Losey, A. M. Gehring, C. Berg and C. T. Walsh, *Chemistry & Biology*, 1999, **6**, 385-400.
19. R. M. Kohli, J. W. Trauger, D. Schwarzer, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2001, **40**, 7099-7108.
20. A. Pinto, M. Wang, M. Horsman and C. N. Boddy, *Organic Letters*, 2012, **14**, 2278-2281.
21. Y. Q. Xu, T. Zhou, S. W. Zhang, L. J. Xuan, J. X. Zhan and I. Molnar, *Journal of the American Chemical Society*, 2013, **135**, 10783-10791.
22. M. Peschke, C. Brieke, M. Heimes and M. J. Cryle, *ACS Chemical Biology*, 2017, **13**, 110-120.
23. R. A. Oliver, R. F. Li and C. A. Townsend, *Nature Chemical Biology*, 2018, **14**, 5-7.

Chapter 2 Polyketide Synthase and Non-Ribosomal Peptide Synthetase Thioesterase Selectivity: Logic Gate or a Victim of Fate?

This chapter is adapted from and updated on Horsman, Hari, and Boddy, 2016.[1]

Type 1, α/β hydrolase-like thioesterase (TE) domains are essential offloading enzymes, releasing covalently bound products from fatty acid, type I polyketide, and non-ribosomal peptide biosynthetic complexes. The release step can occur by attack of an exogenous nucleophile effecting hydrolysis or transesterification or by an intramolecular O-, N-, or C-nucleophile, effecting macrolactonization, macrolactamization, or Claisen-like condensation of the product. Thus, in addition to ensuring turnover of the pathway, TEs provide access to increased chemical diversity. In this chapter I will review the diversity, structure, and mechanism of PKS and NRPS TEs and discuss recent works that highlight the role of TEs as potential arbitrators in offloading. In particular, I examine cases where TEs act as logic gates that ask a particular question about the substrate and use this information to determine the substrate's fate. As the TE mechanism occurs via two steps, I will analyze both the loading and release steps independently as logic gates. The use of logic gates provides an important perspective when evaluating the evolution of TEs within a pathway, as well as highlighting work towards the goal of predicting TE function in unknown and engineered pathways.

2.1 Introduction

Type 1, α/β hydrolase-like thioesterase (TE) domains are essential offloading enzymes in fatty acid, polyketide, and non-ribosomal peptide biosynthesis. Common to the biosynthesis of these three families of compounds is the covalent attachment of the growing compound through a thioester bond to a carrier protein (CP). Expedient offloading of these acyl- and peptidyl-CP intermediates is required to prevent stalling and to enable turnover of the biosynthetic pathways. Cleavage of completed acyl or peptidyl chain from the terminal CP is catalyzed by a two-step TE-mediated mechanism (**Figure 2-1**). The release step can occur

by attack of an exogenous nucleophile effecting hydrolysis or transesterification or by an intramolecular O-, N-, or C-nucleophile, effecting macrolactonization, macrolactamization, or Claisen-like condensation. Thus in addition to ensuring turnover of the pathway, TEs provide access to increased chemical diversity.

Because of the importance of TE-mediated offloading in polyketide and non-ribosomal peptide biosynthesis, TE function has been reviewed in a number of different contexts. At the primary sequence level, a number of reviews have catalogued TEs and examined their alignment[2, 3] with the most recent being the ThYme database from 2010,[4] which sorts the TE clade into 23 families. In the context of all offloading mechanisms for polyketides and non-ribosomal peptides, Du and Lou[5] provided a comprehensive review that included TE-mediated offloading and effectively highlighted its breadth and importance in offloading. As macrocyclization is both chemically challenging and essential for the biological activity of many polyketides and non-ribosomal peptides, TE-mediated macrocyclization was reviewed by Kohli and Walsh[6] in 2003 and non-ribosomal peptide TE-mediated macrolactamization by Kopp and Marahiel[7] in 2007. Because the CP is required to deliver the substrate to the TE, recent reviews have focused on molecular level interactions between CPs and TEs.[8, 9] These reviews provide excellent companions to this chapter.

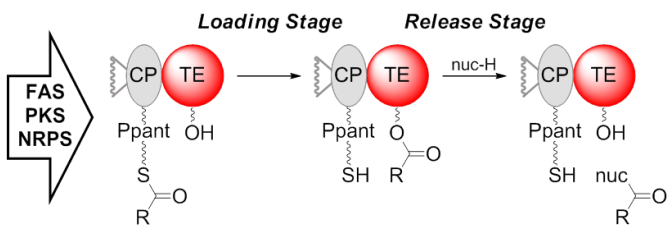


Figure 2-1 TEs catalyze substrate offloading from CPs of fatty acid synthases, polyketide synthases and non-ribosomal peptide synthetases. Acyl and peptidyl chains bound to the Ppant arm of CPs are loaded onto the active site serine of a TE and then released via nucleophilic attack.

Type 1 TEs can be found at the terminus of *cis* acyltransferase (AT) type I polyketide synthase (T1 PKS)[10] modules, *trans*-AT T1 PKS[11] modules, non-ribosomal peptide synthases (NRPS)[9] modules and fungal PKS[12] pathways. The cited reviews cover key points on the function and organisation of each of these pathway types. A second class of TE is also found in some pathways. These type II TEs,[13, 14] along with proof-reading ATs,[15] catalyze release of incomplete acyl and peptidyl intermediates from CP and will not be

discussed here. This chapter summarize the literature on type 1 TE structure, diversity and mechanism, and explores the role of type 1 TEs as logic gates that test for key properties at the loading and release steps.

2.1.1 TE structure

High-resolution structural data have been obtained for a number of TEs (Table 1). In general TEs consist of an N-terminal two helix dimerization region and a repeating β/α/β motif that forms a seven- or eight-stranded parallel β-sheet with a left-handed helical twist (**Figure 2-2**).

Table 2-1 High resolution x-ray structures of TEs

TE	PDB code
Aflatoxin pksA	3ILS[16]
Curacin	3QIT[17]
Deoxyerythronolide B	1KEZ[18], 1MO2[19]
Enterobactin	2ROQ[20], 3TEJ[21]
Fengycin	2CB9[22], 2CBG[22]
Pikromycin	1MN6[19], 1MNA[19], 1MNQ[19], 2H7X[23], 2H7Y[23], 2HFJ[24], 2HFK[24]
Prodiginine	3QMV[25], 3QMW[25]
Surfactin	1JMK[26],2VSQ[27], 2K2Q[28]
Tautomycetin	3LCR[29]

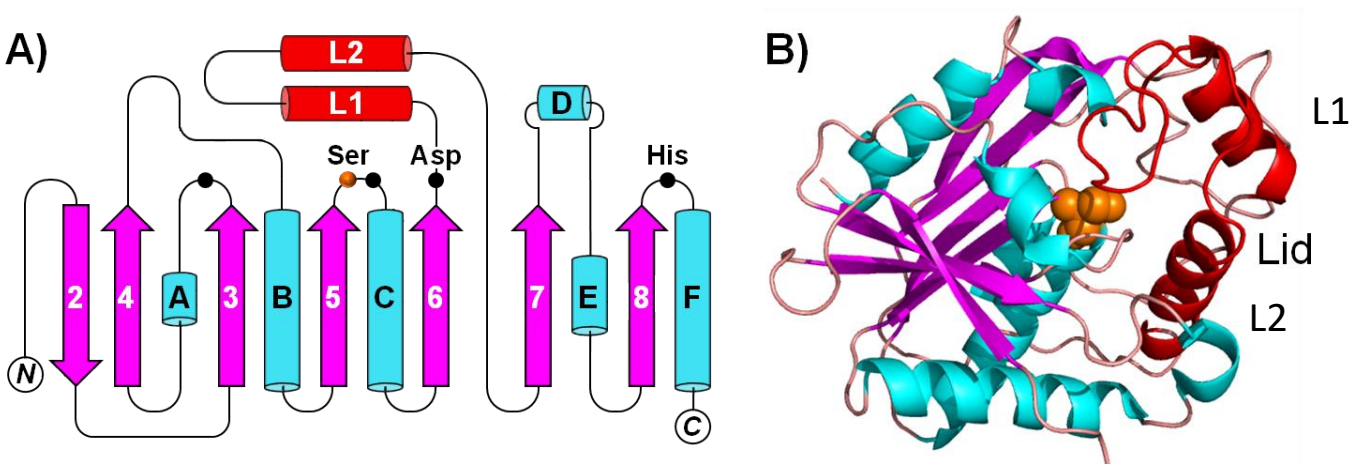


Figure 2-2 Representative structure of a Type I TE (aflatoxin). Topology diagram highlighting the active site triad (labelled residues) and oxyanion participants (shown in black), the lid domain (shown in red) and the overall β-α-β repeating pattern (pink and cyan). This adaptation of the aflatoxin TE is from Korman *et al.*[16] (B) 3D structure of the aflatoxin TE (PDB 3ILS) highlighting the left-handed twisted β-sheet core (pink), the lid region in red, and the active-site serine (orange) at the end of a β-strand and the start of the next α-helix. Substrates are predicted to access the active site by traversing from above the plane of the page.

The protein topology diagram (**Figure 2-2A**) depicts the TE from PksA involved in aflatoxin biosynthesis.[22] In this case the first N-terminal strand of the β-sheet, typical of most α/β hydrolases, is absent and the β-sheet has a single anti-parallel strand. The active-site serine is positioned between the C-terminus

of β-strand 5 and the N-terminus of the subsequent α-helix. In general, the elements between β-strand 6 and 7 are often referred to as the lid region[6, 7] and consist of two α-helices (L1 and L2). These helices line the perimeter of the substrate channel. In dimerizing TEs, this lid has been shown inserted between the core sheet and the N-terminal dimerization region.[18] Some TEs such as the one from curacin biosynthesis,[17] have been crystallized with the lid in an open conformation, which forms an open-faced trough for the substrate and nucleophile. In others, including the TE from erythromycin biosynthesis (6-deoxyerythronolide B synthase TE, DEBS TE),[18, 19] crystal structures show the lid helices spaced to form a channel through the enzyme that restricts substrate and nucleophile entry along the axis of the channel.

Because the upstream CP domain delivers the substrate to the TE active site, its contact with its cognate TE is clearly of importance. CP domains are highly mobile by nature, as they must also deliver substrates to a number of other catalytic domains. They are small domains, typically ~100 residues or less, and are composed of two helices followed by a smaller helical sequence at the C-terminus. To be functional, all CPs are post-translationally modified by addition of a phosphopantetheinyl (Ppant) group that is covalently attached to a serine residue at the N-terminus of helix two.

Work by Liu, Zheng, and Bruner in 2011 captured a high-resolution snapshot of an NRPS CP-TE didomain[21] by cross-linking the CP and TE using a modified Ppant analogue that reacted with the TE active site serine. The crystal structure showed helix three of the CP resting between the first β-strand of the TE's twisted β sheet (β2 in above **Figure 2-2A**) and the first interstrand α-helix (αA), supporting a 2001 docking study prediction.[18] While the docking study proposed an interaction between N-terminal arginines of the TE (coined the Arg patch) and an electron-rich patch on the CP, the crystal structure showed these regions were separated by 6 Å and that substantial hydrophobic contacts dominated the CP-TE binding interaction (**Figure 2-3**). Interestingly the CP appears to use helix two rather than helix three to interact with other catalytic domains such as the phosphopantetheinyl transferase and upstream PKS modules.[8, 20, 27] Discrete hydrogen bonding interactions and hydrophobic packing between the CP Ppant group and the TE were also observed, presumably facilitating penetration of the substrate loaded Ppant arm into the active site binding cavity.[30] Both biophysical characterization of CP-TE interactions and kinetic characterization of substrate transfer from CPs to TEs[30] support a common model. In this model, once the substrate is tethered to the CP upstream

of the TE, essential CP-TE contact points are required for the delivery of the substrate into the TE channel.[30] Once transfer is initiated, little CP-TE contact is detected via NMR;[31, 32] however, discrete Ppant-TE contacts are required for the transfer of the substrate from Ppant thiol to TE active site serine.[21]

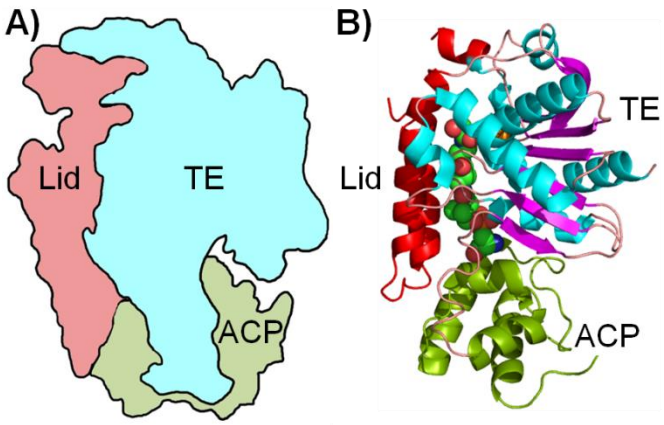


Figure 2-3 Protein-protein interactions between terminal ACPs and their cognate TEs. (A) crystal cartoon model of the structure of cross-linked enterobactin CP-TE (B) A ribbon diagram of the enterobactin CP-TE. Liu et al.[21] linked enterobactin's ACP (green) and TE (lid red, outer helices teal, and core beta sheet pink) via a reactive phosphopantetheine linker (green and red), locking the didomain in a putative substrate transfer conformation

2.1.2 TE sequence diversity

TE sequences are highly diverse. Unlike many other catalytic domains involved in the biosynthesis of polyketides and non-ribosomal peptides, such as ketosynthase (KS),[33] condensation,[34] adenylation,[35-37] AT,[38] ketoreductase (KR),[39] and methyltransferases[40] for which hidden Markov models (HMMs) or conserved motifs can effectively predict selectivity, no model exists for predicting TE loading or release selectivity.[41]

Phylogenetic analysis of TE sequences show that they do not cluster based on type of offloading chemistry they catalyze. This observation is consistent with a convergent evolution model for reactivity types such as hydrolysis, macrolactone formation, macrolactam formation, and macrodiolide formation.[41] It appears the largest factor in their diversity is tied to the nature of the module directly upstream of the TE. For example, T1 PKS[10] and NRPS[9] associated TEs each generate major branches on a TE dendrogram (**Figure 2-4**). In addition, the *trans*-AT PKS[11] associated TEs also appear to cluster away from the bulk of PKS and NRPS TEs. While pathway type is an important determinant of sequence diversity, in some cases TE clustering is based on phylogeny. For example, fungal PKS TEs form a well-defined, tight branch on the dendrogram that clusters well away from other PKS TEs (**Figure 2-4**).

Examination of the alignment of Type 1 TEs highlights how diverse their sequences are. For example, a clustalW alignment of 300 predicted TEs extracted from clustermine360[42] shows there is no 50% consensus region larger than 5 amino acids throughout the entire sequence. As a result, regions of moderate diversity are not immediately isolable to establish sequence property correlations. The largest source of sequence diversity is found in the inter-β6/β7 lid region, which ranges in length from approximately 80 amino acids in a TE from a *Myxococcus xanthus* isolate (|gb|NC_008095.1) to 25 amino acids from the coronatine TE in *Pseudomonas syringae* (PSPTO_4699|gb|NC_004578.1|). Curiously, both the open-pore DEBS TE and the shielded curacin TE have shorter than average lid regions. This suggests that lid-size alone cannot serve as a predictor of reactivity since DEBS TE generates a macrolactone and curacin TE produces a carboxylate that undergoes decarboxylative elimination to generate a terminal olefin.

PKS and NRPS TE diversity is not evenly distributed through the larger Pfam[43] TE family. When these TE sequences are compared with the Pfam training set (PF00975, shown as green circles in **Figure 2-4**) used to generate the HMM for TE domain detection, five of the training set sequences cluster with the major PKS TE branch and only two of the training set sequences cluster with the major NRPS TE branch. The majority of the training set form new branches, many of which are sparsely populated by PKS and NRPS TEs. This suggests that a further refined HMM may better discriminate between pathway origins of TEs and an expanded HMM may better capture NRPS TEs. Bioinformatic detection of NRPS TE can be problematic. For example, a recent study used the powerful bioinformatics tool antiSMASH,[44] which relies on the Pfam TE HMM, to perform an *in silico* screen for NRPS pathways without annotated release mechanisms. This screen was unable to find TE domains in 15 NRPS pathways that after more rigorous analysis were confirmed to possess TEs.[45] Adapting the TE HMM to better detect NRPS TEs may thus improve the false negative rate of essential biosynthetic pathway analysis tools like antiSMASH.

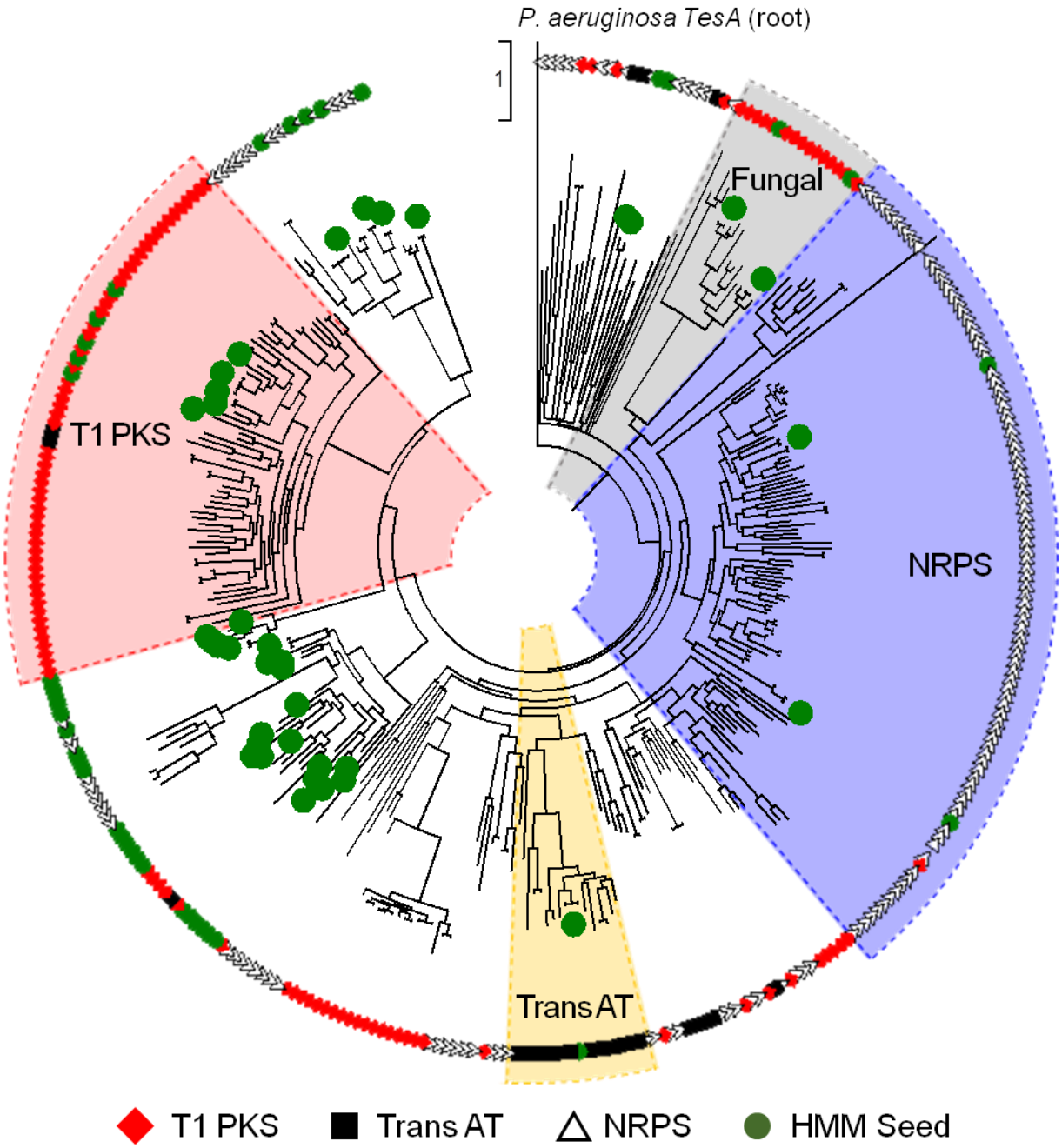


Figure 2-4 Representative structure of a type I TE (aflatoxin). Diversity of sequenced NRPS and PKS TEs at the amino acid level. 295 TEs were downloaded from clustermine360.ca,[42] combined with the 38 sequence TE seed (http://pfam.xfam.org/family/PF00975/alignment/seed) for the current TE HMM, truncated to a clustalW[46]-aligned 268-amino acid core structure, phylogenetically analyzed using WAG[47], and assessed for branch support using approximate likelihood ratio test (aLRT) via phylogeny.fr.[48] For sequences and branch strength see Appendix Figure I-4, page I-8. Seed sequences are highlighted with green circles.

2.1.3 Mechanism

TEs operate via a two-step mechanism, loading followed by release, which is consistent with known serine hydrolase mechanisms (**Figure 2-5**).[49] The active site Ser side chain alcohol is activated by the conserved His-Asp dyad, increasing its nucleophilicity. The thioester substrate approaches the activated Ser from the N-terminal side of the substrate channel. The approach is mediated by the Ppant tether that links the substrate to the CP. It has been hypothesized that the lid region opens to accommodate the presentation of thioester substrates.[18, 20] Once in the active site, the substrate thioester is attacked as the Ser is deprotonated by the His-Asp dyad, and the charged tetrahedral intermediate is stabilized in the oxyanion hole by hydrogen bonding from two backbone amide NH groups typically found at *i*+1 from the active site Ser and in the β-turn after β strand 3 (**Figure 2-2A**).[16] This intermediate is resolved by loss of the Ppant thiolate, generating the acyl-TE intermediate.

The second step of the mechanism involves release of the acyl group and is often referred to as offloading. This step begins with the approach of an intramolecular or intermolecular nucleophile. Townsend has suggested that the active-site histidinium ion is deprotonated by the departing thiolate,[16, 50] and thus capable of activating the incoming nucleophile. Alternatively, the histidinium may be retained since the thiolate leaving group is a weak base (as contrasted with ester hydrolysis) and must be otherwise deprotonated before the release phase. The nucleophile adds into the carbonyl of the acyl-TE intermediate and the tetrahedral intermediate is once again stabilized by the oxyanion hole. Finally the seryl alkoxide is released with concerted protonation and the product leaves the active site.

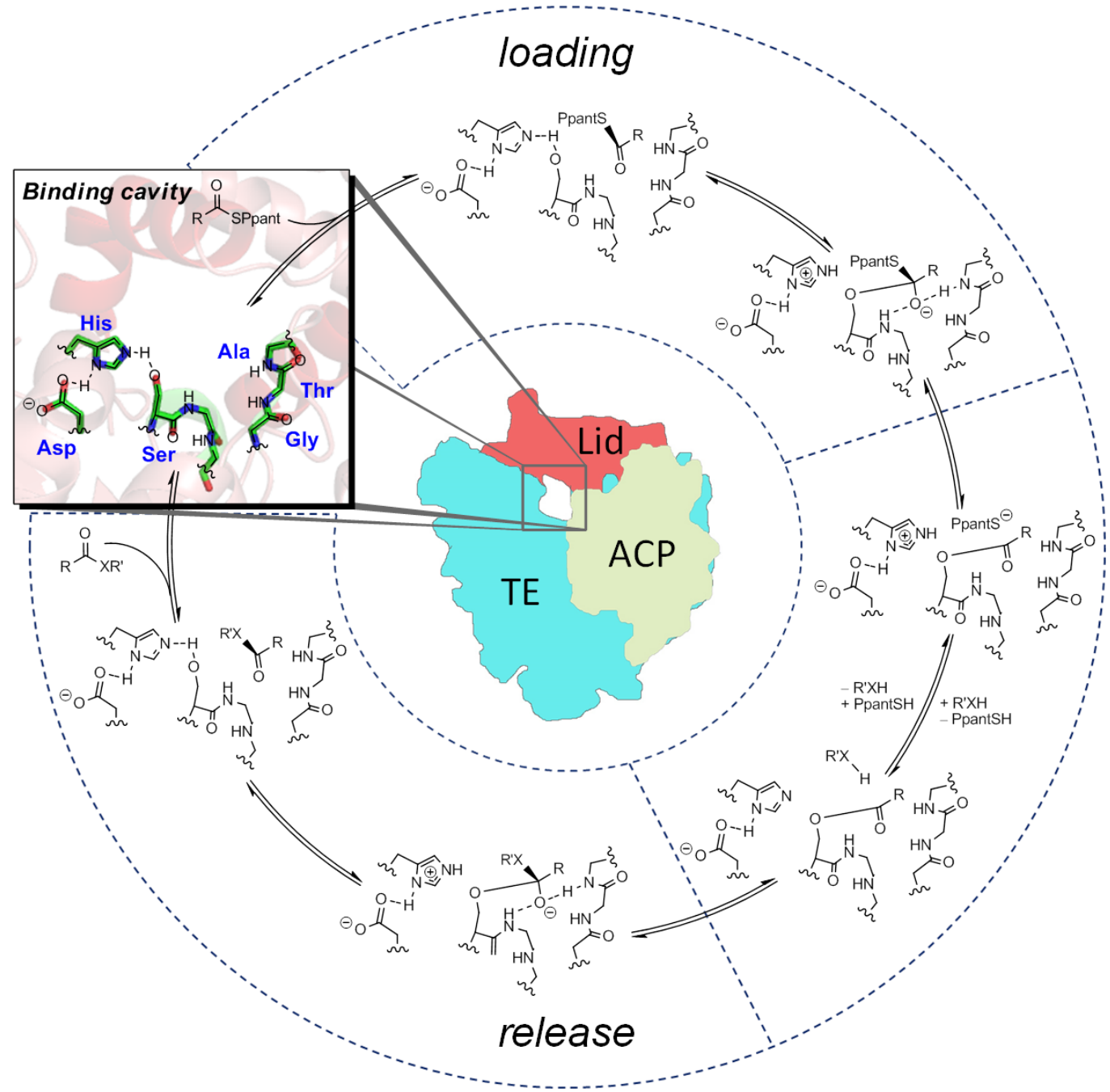


Figure 2-5. Detailed TE mechanism for substrate loading and release. DEBS TE active site as viewed from a perspective where the substrate would enter the cavity from above the plane of the page. The Ppant-conjugated substrate enters the active site with the R-group leading and the Ppant lagging. In the loading stage, serine replaces the thioester bond, and in the release stage, a nucleophile replaces the seryl moiety and the active site is recovered.

Some TEs deviate from the canonical Ser-His-Asp triad (see alignment in Appendix Figure I-3, page I-3) and contain a Cys in place of the Ser. For example, the TE domain from pyochelin biosynthesis in

Pseudomonas aeruginosa[51] and *Streptomyces venezuelae*[52] both contain a GXCXG motif rather than the canonical GXSXG motif. Mutation of the native Ser nucleophile to Cys in both myristoyl-ACP TE from *Vibrio harveyi*[53] and rat mammary gland TE II[54] generates catalytically active TEs, suggesting that native Cys containing TEs are likely active. Deacylation of the active site Cys in Ser to Cys mutants has been shown *in vitro* to be rate limiting.[53, 54] The Ser to Cys and His to Arg double mutant has even slower rates of deacylation and has been used to trap the acyl-TE intermediate for the TE from the tautomycin biosynthetic pathway.[29] The catalytic His appears to be essential for activity of TEs and is invariant among the TEs with the exception of the TE from mycolactone biosynthesis. In this TE the catalytic His has been proposed to be replaced with a Pro, however there is no experimental evidence to support Pro in this role as of yet.[55] In the third position of the catalytic triad, the sample alignment demonstrates six examples of Glu replacing the Asp and ten examples of Ser replacing Asp. This follows the trends seen for general hydrolase active site as discussed by Heikinheimo.[2] Holmquist covered variants of the catalytic triad in greater detail in 2000[3] and a comprehensive update to this work incorporating recent sequences data would be a welcome addition to the literature.

2.1.4 Logic gates in natural product biosynthesis

Enzymatic domains in natural product biosynthesis often intrinsically select for a substrate and stall until these selection criteria can be met. Enzymatic domains of this type have been termed gate-keepers,[56] decision gates,[57] and most recently, logic gates.[58] While these three terms are synonymous, I will use logic gate throughout as it highlights the programming and decision tree inherent in PKS and NRPS biosynthetic pathways. In the context of TE catalyzed chemistry, logic gates describe the test for key properties at the loading and release steps. Enzyme-substrate interactions probe shape, size, rigidity, and electronics of the substrate and a subset of these features can be used at either stage; however, these are only some of the conditions placed on the substrate. Factors external to the TE from the upstream pathway can also place restrictions on loading or release. Collectively, these TE-based and external factors are the logic gates that impact the decision to load and release the substrate. Understanding what question each logic gate "asks" of the substrate will enable us to predict substrate loading and release in natural and engineered *in vivo* pathways as well as *in vitro* pathways and chemoenzymatic situations.

Thus, herein, I examine to what extent TEs are logic gates responsible for discriminating between thioester substrates and directing the route of release. I will highlight factors involved in the selectivity for the loading and release steps, discuss the implication of the differing levels of selectivity, and provide biosynthetic predictions that result from placing TEs with these patterns in selectivity in different metabolic environments.

2.2 The Loading Stage

The location of TEs at the end of PKS and NRPS biosynthetic pathways has important implications for the selectivity of these catalytic domains. The screening hypothesis suggests that there is an evolutionary cost associated with tying stringent substrate selectivity to late-stage components of multistep syntheses.[59, 60] For T1 PKS and NRPS pathways, the substrate presented to the TE is tightly controlled by the modular configuration of the pathway. Thus, pathway evolution, which is required to access new chemical diversity and maintain long term evolutionary fitness, should be facilitated by low intrinsic selectivity of the TE. In certain cases however, the TE must be more stringent in its substrate selection. This includes instances of substantial processing of the terminal CP bound intermediate such as in pyochelin biosynthesis[61] and in iterative extension pathways such as the radicicol biosynthetic pathway.[57]

2.2.1 Typical type I PKS and linear NRPS TEs have low selectivity for the loading stage

There are substantial *in vivo* and *in vitro* data on the ability of NRPS and PKS TEs to load diverse substrates. Reductive domain inactivation in full native T1 PKS pathways provided some of the first examples of TE loading substrate tolerance. For example, inactivation of individual KR domains from the complete biosynthetic pathways for erythromycin[62, 63] and amphotericin[64] produced the corresponding keto polyketide products *in vivo* and inactivation of an ER domain from the erythromycin pathway produced the alkene containing natural product *in vivo*.[65] Both examples demonstrate the flexibility of the TE loading toward oxidation state changes. Similarly precursor directed biosynthesis experiments using synthetic substrates and blocked mutants of native pathways, such as the erythromycin pathway, also showed that analogues of the native polyketide natural product could be produced, confirming that TE loading was substrate tolerant.[66] Lastly, truncation of native pathways to produce fragments of natural products has also

been successful, further demonstrating the flexibility of TE loading for substrates that are smaller and less complex than their native substrates.[67, 68]

The construction of *de novo* designed PKS pathways has enabled diverse substrates to be presented to type I PKS TEs *in vivo*.[69-73] For example, Menzella *et al.* generated a small library of 6-member lactones by combining two different modules from the erythromycin, soraphen, epothilone, geldanamycin, leptomycin, rifamycin, rapamycin, and pikromycin biosynthetic pathways.[69, 70] Fourteen of these combinations produced over 10 mg/L of the expected products. While this suggests the TE can load these diverse substrates, it is not possible to rule out spontaneous non-TE mediated offloading, particularly since non-TE mediated cyclization of 6-member lactones is kinetically and thermodynamically favoured[74] and known to occur.[75] Subsequent work where a third module was added to produce tetraketide 6-member lactones showed similar results, with over twenty systems generating at least 1 mg/L of δ-lactone product when cultured.[71]

McDaniel *et al.* provided unambiguous data on the ability of a PKS TE to load diverse substrates.[70] In this study, over 60 different 14-member macrolactones were produced by swapping the AT domains and the reductive domains from modules 2, 5, and 6 of the erythromycin biosynthetic pathway with AT and reductive domains from the rapamycin biosynthetic pathway. As these 14-member rings cannot be generated from spontaneous non-TE mediated chemistry, their formation confirms that the TE used in this study (DEBS TE) must be able to load all of the macrocyclized substrates.

While the detection of a final product that cannot be offloaded spontaneously from *in vivo* experiments informs our understanding of the substrate tolerance of TEs, little can be deduced from *in vivo* systems that do not generate any observable products. McDaniel *et al.* indicate non-native intermediates may stall partway through the assembly line, fail to transfer downstream, or be cleaved by pathway editors.[70] Alternatively, the TE may fail to load or release the chain. In a state of protein cycling, cell growth, and asynchrony of substrate turn-over, it is impossible to establish which of these possibilities is responsible for the failure of the system to generate a product.

Much effort has thus been extended to isolate individual catalytic domains, including the TE, to be able to deconvolute their activities and selectivities separate from the full PKS pathway. These types of *in vitro*

assays are typically performed at elevated substrate and enzyme concentrations, often with synthetic non-native substrates. Typically, an excised TE domain is used; thus, the loading step is occurring intermolecularly, rather than intramolecularly, through transfer from the upstream, acylated CP. Pulse-chase experiments indicate that the rate limiting step *in vitro* for *Vibrio fisheri* TE LuxD is the loading of myristoyl-CoA.[53] This suggests that steady-state kinetic assays provide a quantitative rate constant that in cases of saturatable TEs with a rapid pre-equilibrium of substrate binding can define the disassociation equilibrium constant and turnover number for formation of the acyl-enzyme intermediate from the enzyme substrate complex.[76] This observation is useful for characterizing substrate loading and has major implications for characterization of the release step.

Kinetic characterization of DEBS TE,[76-79] and the TEs from pikromycin (PikTE)[80] and pimaricin biosynthesis[77] with simple substrates has demonstrated that all substrates are hydrolyzed at rates well above background. This suggests that all small substrates presented to the type I PKS TEs in the absence of other external selectivity-filters load into the active site and the loading gate selectivity is limited. The substrates that match with the shape of the TE active-site react the fastest, while the poorly fitting substrates require extended exposure to the TE to be loaded. In these examples, the selectivity of the TE towards the simple substrates could not be predicted from comparison of the substitution pattern and oxidation state of the small substrate with that of the C-terminal portion of the native substrate. Often non-native substrates were much more efficiently processed.[79]

Much like type I PKS TEs, NRPS TEs studied to date have shown remarkable tolerance for substrate loading. Work from 2000 suggested that the isolated tyrocidine TE was capable of loading activated esters of a diverse array of over 100 tyrocidine A analogues. Replacement of each of the ten residues, except the N-terminal D-Phe and penultimate residue, L-Orn, generated substrates that could be readily processed by the TE.[81, 82] Of particular note is that the tyrocidine TE was able to release a large diversity of peptidyl substrates attached via an ester linkage to PEGA solid phase peptide synthesis resin, impressively demonstrating its application as a biocatalyst.[82, 83] Thus, like type I PKS TEs, the NRPS TEs typically appear to have a limited role as a logic gate in substrate loading.

2.2.2 Stringent logic gates can occur if an acyl-CP intermediate must be processed

While NRPS TEs appear to have limited roles as logic gates in most cases, several exceptions have been documented in NRPS pathways with on-pathway tailoring of the extended peptide chain. For example, teicoplanin is an NRPS-derived linear heptapeptide that is extensively cross-linked through oxidation. Evidence has recently emerged that the extended heptapeptide remains at the end of the NRPS on the CP next to the TE as cytochrome P450 monooxygenases are recruited to the module to process the nascent peptide.[84] In another case, tetrabromopyrrole is biosynthesized using a minimalist *in-cis* NRPS. It consists of a CP that acts as tether for L-proline, and a TE that removes compounds. The tetrabromination is driven entirely by a halogenase acting *in trans*. Remarkably, the compounds are positioned by the TE, and only halogenase isolation of the CP or TE selectivity for the brominated product prevent premature release of the compound.

Finally, the biosynthesis of pyochelin demonstrates a confirmed example of substrate-selective loading to the TE. Pyochelin includes two thiozole-based rings (**Figure 2-6**). The second ring is installed after the full-length peptide is condensed onto PCP_5 and before the TE releases the final compound (**Figure 2-6**). The cyclization is known to be catalyzed by the heterocyclization/condensation domain of PchF.

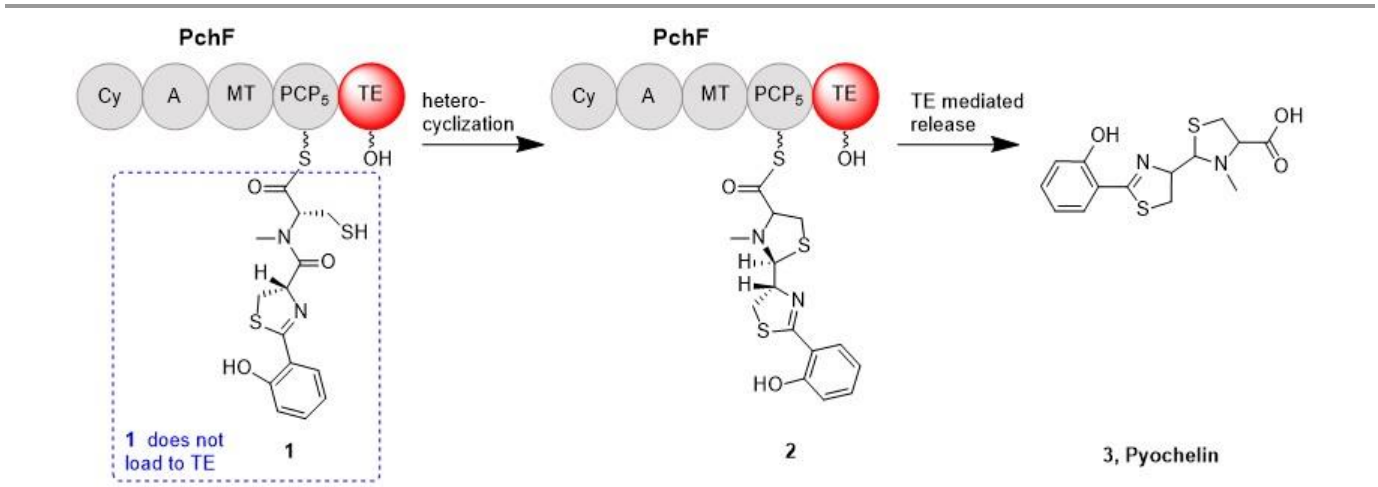


Figure 2-6. The terminal ring-specific release of pyochelin precursor is supported by PchF TE's substrate specificity.[61] MT methyltransferase.

Schnieder *et al.* described this tetrahydrothiazole formation and the subsequent N-methylation as "prerequisites for pyochelin chain maturation and hydrolytic disconnection from PchF [via the terminal

TE]"[61]. The authors did not explicitly discuss PchFTE's selectivity for ring-containing substrates in contrast to Tyr TE's ring-tolerance[85]; however, they showed that oxidation of the first thiazoline ring to a thiazole disrupted release of the dithiazole product. "…[F]urther processing beyond the amide bond formation … was not observed"[61]. While they conceded that the hydrolysis conditions used to release and detect stalled intermediates may have hydrolyzed the second thiazoline ring, they believed the synthesis had stalled at the cyclization stage. In addition, they reasoned that no linear product was detected because the TE gating requirements for the ring and methylation had stalled the synthesis at a check-point. Here, detection of dehydrated (-18) protein-substrate complex or screening for N-methylated hydrolysis product may validate this assumption.

Rapid processing of compound **1** by the TE would release a non-cyclized product. Selective release of **2** however, ensures that the cysteine containing intermediate remains on the CP long enough to be processed to a cyclized product. This TE-mediated substrate pause may thus be a control element in pyochelin pathway fidelity.

2.2.3 TEs from iterative pathways are likely to be stringent logic gates

Unlike in modular polyketide biosynthesis, during iterative polyketide biosynthesis an ACP is acylated with a variety of different substrates. The majority of these substrates must be transferred back to the KS domain for additional elongation. If the pathway possesses an offloading TE, as is the case for some fungal polyketide biosynthetic pathways, the TE competes with the elongation process for each of these substrates. In order to ensure complete elongation, the TE must exhibit selective offloading of the substrate. Based on this concept, a TE from these pathways is expected to display higher chain-length related loading selectivity than their modular equivalents and has the capacity to act as a stringent logic gate, potentially controlling the fidelity of the pathway. The iterative pathways for the biosynthesis of radicicol and other related macrocyclic fungal natural products provide *in vivo* and *in vitro* evidence for the TE acting as a strict logic gate for substrate loading.

Radicicol biosynthesis is catalyzed by a highly-reducing PKS (hrPKS) and a non-reducing PKS (nrPKS). The hrPKS synthesizes the initial pentaketide acyl chain containing an alcohol and an olefin.[86] This residual oxygenation is a result of cryptically encoded incomplete reduction by the embedded reductive domains in

the hrPKS. Interestingly in the related fungal polyketides dehydrocurvularin[87] and zearalenone,[88] the hrPKS generates a tetraketide and hexaketide chain respectively. The chain length control mechanism in this first round of iterative polyketide biosynthesis is still poorly understood.

In contrast, much is known about how the nrPKS regulates exposure of the growing acyl chain to the C-terminal TE. The nrPKS accepts the substrate from the hrPKS and elongates it into a poly-β-ketothioester substrate. During this elongation the ACP is loaded with β-ketothioesters of various lengths. For additional elongation to occur the substrate must be passed back to the KS rather than loaded onto the TE domain. Loading onto the TE would terminate chain elongation leading to alternative products. *In vitro* kinetic characterization of the TE from the zearalenone hrPKS shows that the TE has over a 100-fold lower specificity constant for hydrolysis of a β-ketothioester containing substrate over a benzoate thioester,[56] which is generated by cyclization and aromatization of the poly-β-ketothioester once elongation is complete.[87] This result suggests that the loading selectivity of the TE is matched to ensure that the KS can outcompete the TE for β-ketothioester substrates. Once the substrate is cyclized to generate a benzoate, the complete chain is rapidly offloaded.

Aromatic ring formation from the ACP-tethered poly-β-ketothioester substrate is catalyzed by the product-template (PT) domain, a domain unique to nrPKSs.[87, 89] Because of the high reactivity and dense functionalization of the poly-β-ketothioester intermediate, a variety of aromatic rings can be accessed. In zearalenone and radicicol biosynthesis the intermediate adopts an "F-type" (for fungal)[90] fold that leads to formation of an aromatic ring between C2 and C7 (the resorcylate ring system) (**Figure 2-7A**). In dehydrocurvularin biosynthesis the substrates adopts an "S-type" (for *Streptomyces*)[90] fold leading to aromatic ring formation between C3 and C8 (phenylacteate ring system) (**Figure 2-7B**).

Recent *in vivo* work suggests that these TEs are selective for their native aromatic functional group, either resocylate or phenylacetate. For example, the radicicol pathway contains an F-type PT domain that generates an ACP-bound resorcylate thioester (**Figure 2-7A**, **5**). In the native systems, this is loaded onto the TE and subsequently released as a 14-member ring macrolactone **6**. Replacement of the native TE with the TE from dehydrocurvularin biosynthesis leads to formation of pyrone **7**. The dehydrocurvularin TE is presented, in its native pathway, with a phenylacteate-ACP intermediate **9**. In this experiment, it is presented with a

resorcylate-ACP intermediate **5**, which is not processed into a macrolactone. Instead the intermediate undergoes hydrolysis to the pyrone. This chemistry is observed in trace amounts in the absence of a TE[57] and pyrones are known spontaneous release product of other systems lacking TEs.[75] Interestingly, pyrone formation increases substantially when the dehydrocurvularin TE is added to the radicicol nrPKS versus the TE deletion. Whether the role of the dehydrocurvularin TE is structural and no formation of an acyl-TE intermediate is occurring or catalytic where TE-mediated hydrolysis releases the pyrone is not known. When the same linear radicicol precursor **4** is presented to the dehydrocurvularin PT, it generates a phenylacetate-ACP intermediate which is easily loaded onto the dehydrocurvularin TE and released as a 14-member ring macrocycle (**Figure 2-12**, **27**).[57]

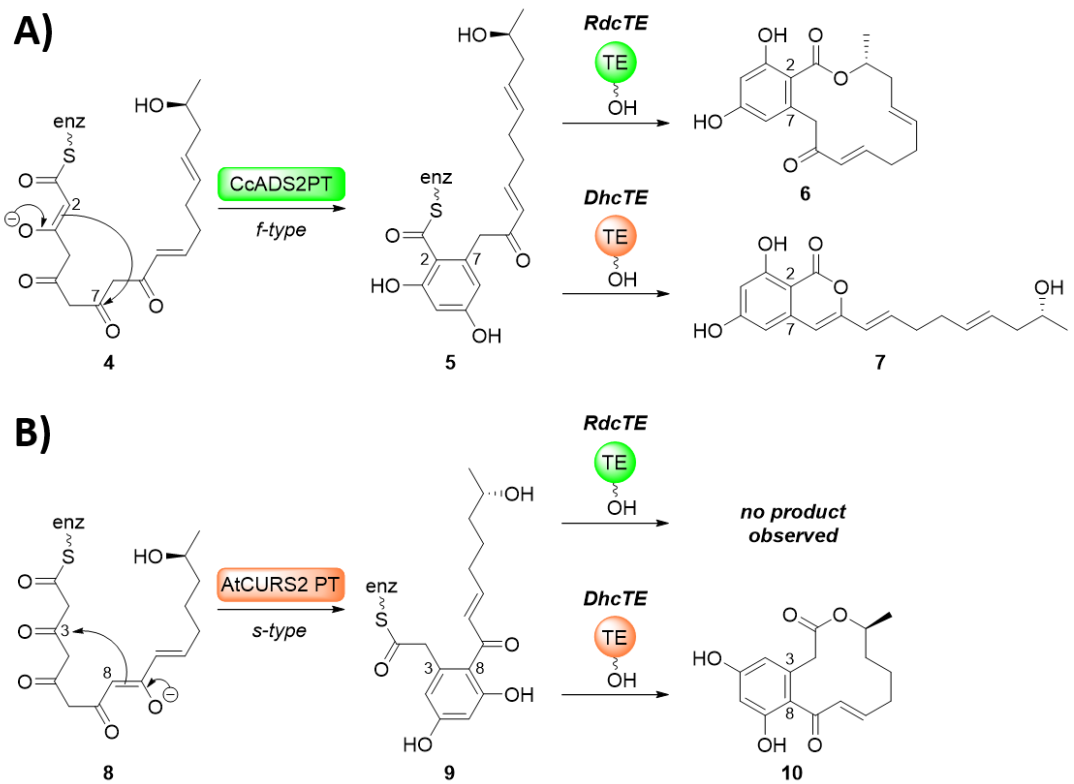


Figure 2-7. Fungal macrocyclizing TEs selectively load substrates with aromatic rings generated by their cognate PT domains. (A) In radicicol biosynthesis the F-type PT domain (CcADS2, green) generates a resorcylate intermediate **5** from the poly-β-keto intermediate **4**. The native radicicol TE (Rdc TE, green) loads the resoryclate substrate and macrocyclizes it whereas the dehdyrocurvularin TE (Dhc TE, orange) is unable to load the substrate, leading to spontaneous pyrone formation. (B) In dehydrocurvularin biosynthesis the poly-β-keto intermediate **8** is cyclized by the S-type PT domain (AtCURS2, orange) producing a phenylacetate intermediate, **9**. The native dehydrocurvularin TE (Dhc TE, orange) effectively loads this substrate producing **10** whereas the radicicol TE (Rdc TE, green) cannot and produces no detectable product. Figure adapted from Xu *et al.*[87]

The native dehydrocurvularin pathway produces a phenylacetate-ACP intermediate (**Figure 2-7B**, **9**). The dehydrocurvularin TE effectively loads this substrate and macrocyclizes it to generate the 12-member macrolactone natural product **10**. Replacing the TE with the radiciol TE does not yield any product from intermediate **9**. Presumably the radiciol TE is unable to load the phenylacetate substrate and the pathway stalls. Only trace products form misprimed dehydrocurvularin nrPKS are released, presumably by background hydrolysis of the acyl-ACP intermediate.[57]

Taken together, these data suggest that matching the PT and TE domains is required to enable efficient offloading. These data support a model where loading of the TE is slow in cases where the PT domain and TE are mismatched, and faster in cases where the PT domain and TE are matched. However, confirmation of this hypothesis will require kinetic characterization of TE loading by both resorcylate and phenylacetate substrates.

2.3 The Release Step

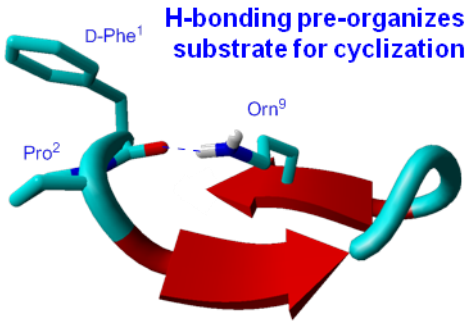
Cleavage of the acyl-TE intermediate releases the untailored polyketide or nonribosomal peptide from the core biosynthetic machinery. A stall in processing of the intermediate at this stage due to enzyme selectivity would impede flux through the pathway. Thus, the selectivity of the release step focuses on the type of release-chemistry catalyzed by the TE. This ranges from hydrolysis, macrocyclization, transesterification,[91, 92] dimerization, and trimerization among others. This diverse reactivity enables the TE to dramatically impact the chemical and functional diversity of the metabolites produced by a pathway.[6] It is in this context that the TE can act as a logic gate during release. The chemistry of release can be controlled by extrinsic factors, such as substrate pre-organization, which has been shown to facilitate macrocyclization in favour of hydrolysis in some NRPS TEs. Intrinsic TE selectivity for macrocyclization versus hydrolysis has been observed for PKS TEs when the acyl chains meet stringent functional group and stereochemical requirements. Fungal RAL TEs, on the other hand, have been shown to favour macrocyclization over hydrolysis for a much wider range of substrates. Fused-ring C-C bond forming Claisen-like condensation (CLC) fungal TEs shield their reactive intermediates from spontaneous 6-member ester formation through

substrate restriction. The ability to toggle between these various release reactions is likely an important evolutionary adaptation and thus, intrinsic release selectivity is predicted to be malleable and unpredictable.

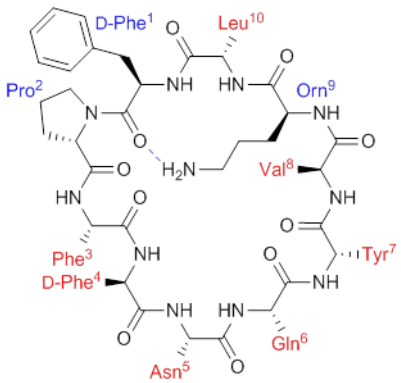
2.3.1 Logic gate for NRPS macrocyclization can be controlled by substrate conformation

Macrocyclization is a key feature of many polyketides and non-ribosomal peptides. In the case of non-ribosomal peptides, it has been proposed to decrease the natural products susceptibility to proteases,[22, 93] improve their membrane solubility, and limit their conformational flexibility.[93, 94] This may constrain a peptide to a target binding favoured conformation, limiting the entropic barrier for binding[93, 94] (Delorbe *et al.* has cautioned that this assumption should be tested[95]). Notably though, most of the known nonribosomal peptides of greater than four residues are produced as macrocycles. It is thus not surprising that NRPS TEs are effective at producing macrocycles from a variety of substrates.[82, 83, 96-100]

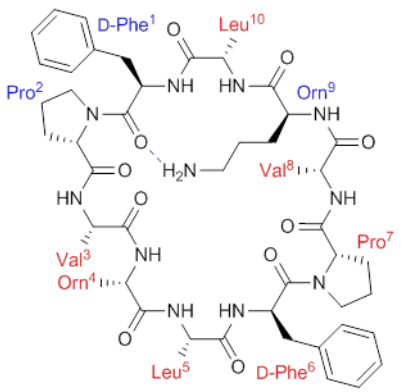
In 2001, Trauger *et al.* established that the tyrocidine TE can macrocylize nona-, deca-, and undecapeptides, and could remarkably dimerize and then macrocyclize the related gramicidin pentapeptide.[101] Through an alanine scan, they established that only the N-terminal and penultimate C-terminal residues were essential for macrocyclization. To account for this substrate tolerance, they proposed a model where the peptidyl substrate backbone can effectively hydrogen bond and in addition form a key hydrogen bonding interaction between the Orn side chain amine and the carbonyl of the N-terminal residue to pre-organize the linear peptide for cyclization (**Error! Reference source not found.**).[101]



tyrocidine A, **11**



tyrocidine A, **12**



gramicidin S, **13**

Figure 2-8. General model of tyrocidine-type substrate-preorganization driven NRPS macrocyclization. Trauger *et al.*'s[101] model for the necessary elements for Tyrocidine TE mediated macrocyclization (shown in blue). Five peptide backbone hydrogen bonding partners are required as well as the L-ornithine side chain (key Orn-backbone carbonyl hydrogen bond is shown). The N-terminal amino acid must be an aromatic D-amino acid.

The TE also exerted intrinsic substrate selectivity for macrocyclization, limiting the N-terminal residue to D-amino acids. When these pre-organized substrates were loaded onto the TE active site, water was excluded from the active site through the bulk of the side chain of the oxyanion-hole residue Phe29 (found at the C-terminus of β3, see **Figure 2-2A**). Hydrolysis of the peptidyl-TE intermediate was favoured by mutation of Phe29 to a small nonpolar amino acid.[101] In 2002 this work was extended to show that a proline residue at the 4th position disrupted the pre-organisation, and macrocyclization did not occur.[82] Furthermore, the N-terminal D-Phe could be replaced with similar aromatic D-amino acids. However, this study did not comment on the rate of processing or macrocycle/free acid product release partition beyond stating that they acted as better substrates for cyclization.

A number of other NRPS TEs have been characterized *in vitro* and shown to be substrate tolerant macrocyclization catalysts. For example the cryptophycin TE was found to cyclize both homo and nor cryptophycin 3 (cryptophycins 51 and 24 respectively).[99, 102] TEs from the lipopeptide antibiotic class, which includes daptomycin and A54145, have been shown to generate macrocycles with specific amino acid substitutions and tolerance for ring size, ranging from 8-11 amino acid residues.[97, 103] However, insufficient data is available to propose a mechanism by which these TEs control macrocyclization over hydrolysis.

2.3.2 Modular type I PKS TEs are capable of ring-size tolerant, stereoselective macrocyclization

In their native pathways, modular type I PKS TEs are selective in the reaction they catalyze. The vast majority of characterized, native pathways produce a metabolite

corresponding either to macrocyclization, such as the epothilone pathway,[104] or hydrolysis, such as the gephyronic acid pathway.[105] There are no characterized examples of both macrocycles and free acids being generated *in vivo* from the same native pathway in appreciable amounts. However, I cannot rule out this absence being due to an isolation bias as most compounds are isolated by an activity guided fractionation approach.[106, 107] In addition to being selective for the type of reaction, macrocyclizing TEs typically are highly selective for a single intramolecular nucleophile. For example, the DEBS TE is highly selective for the C13 alcohol over the C11 and C5 alcohols. Only the native 14-member ring compound (**Figure 2-9**, **17**) is obtained from erythromycin producers,[108,109] even though the 6-member ring is enthalpically and

entropically favoured.[74] Pikromycin biosynthesis is one of the few examples where multiple ring sizes are accessed, though in this case it is due to skipping the terminal polyketide synthase module rather than the TE selecting between two possible nucleophiles.[110]

By modifying native biosynthetic pathways, TEs can be exposed to a range of non-native substrates, however this substrate diversity is limited by the selectivity of the upstream biosynthetic enzymes. Once loaded, these non-native substrates are released through hydrolysis unless they meet the TE's intrinsic requirements for macrocyclization or transesterification. For example, erythronolide analogs lacking methyl substituents (*e.g.*, **Figure 2-9**, **24**)[111] have been generated by taking advantage of *trans*-AT domains with malonyl-CoA selectivity, versus the native methylmalonyl-CoA selectivity. These compounds have been successfully macrocyclized.[111] Similarly, truncated pathways that generate shorter linear polyketide chains have been successfully cyclized to generate macrocycles of reduced ring size (*e.g.*, **Figure 2-9**, **14-16**).[68, 112] Using tools such as precursor directed biosynthesis, larger ring sizes have also been accessed (**Figure 2-9**, **19**).[66] A combinatorial biosynthesis experiment *in vivo* successfully generated over sixty 14-member ring macrocylic analogs of erythronolide by combining catalytic domains from the erythromycin and rapamycin biosynthetic pathways.[70] Conserved amongst all these examples was the D-configuration of the lactone alcohol and the presence of a carbonyl 1,5 to the lactone alcohol.

In vitro characterization of TE activity can provide for even broader probing of the substrate selectivity of TEs since synthetic substrates can be assayed. Isolated TEs can be treated with preloaded CPs or thioester substrates enabling selectivity to be studied independent of the upstream biosynthetic machinery. This dramatically increases the potential diversity that can be screened. In practice however, the challenge of synthesizing complex substrates for these enzymes has limited the probing of the TEs with non-biogenic substrates. In 2005, He *et al.* ring-opened 10-deoxymethynolide produced by *Streptomyces venezuelae* and reduced the ketone to a secondary alcohol.[113] While the ketone containing linear substrate was effectively macrocyclized by the excised recombinant PikTE, which is highly related to the DEBS TE, the reduced compound was only hydrolysed. This is consistent with the trends seen *in vivo*.

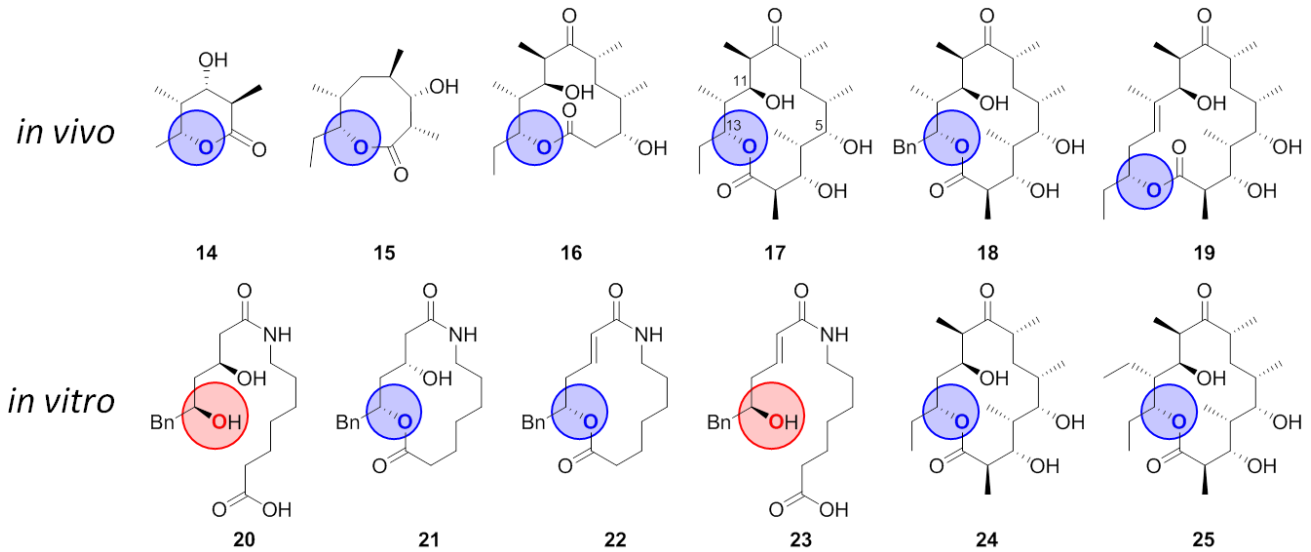


Figure 2-9. Erythronolide analogs macrocyclized via DEBS TE *in vitro* and *in vivo*. 6-,[114] 8-,[112] 12-,[68] 14-, and 16-member[66] rings have been constructed heterologously *in vivo*, and diverse analogues have been constructed synthetically and from *in vitro* reconstituted pathways to incorporate amides,[41, 115] benzyl[66] side chains, remove methyl side-chains,[111] and change the reduction state. Only the D-configuration of the ring-forming nucleophile is tolerated.

Work by Pinto *et al.* has demonstrated that for the DEBS TE the D-configuration of the nucleophilic C13 alcohol is required (**Figure 2-9. Erythronolide analogs macrocyclized via DEBS TE *in vitro* and *in vivo*.** 6-,[114] 8-,[112] 12-,[68] 14-, and 16-member[66] rings have been constructed heterologously *in vivo*, and diverse analogues have been constructed synthetically and from *in vitro* reconstituted pathways to incorporate amides,[41, 115] benzyl[66] side chains, remove methyl side-chains,[111] and change the reduction state. Only the D-configuration of the ring-forming nucleophile is tolerated., **21**) and that L-configured alcohols do not macrocyclize but rather generate the free acid (**Figure 2-9**, **20**).[115] Interestingly, chemical macrocyclization of two diastereomeric substrates – one that undergoes TE-mediated macrocyclization and the other hydrolysis – occur with comparable efficiencies.[115] This result suggests that substrate preorganization is not a major factor in DEBS TE mediated macrocyclization, unlike with the tyrocidine TE.[81]

While TEs are generally selective in the reactions they catalyze such as hydrolysis, macrocyclization, or transesterification, recent work by Hari *et al.* has shown that with some substrates, multiple offloaded products can be obtained.[41] For example, the DEBS TE was shown to generate macrocycle **22** (**Figure 2-9**) from the linear *N*-acetylcystemaine activated thioester, and to catalyze transesterification and dimerizeation,

generating a linear dimer and a macrodiolide product.[41] This result confirms that it is biochemically possible for pathways to produce substantial quantities of a diversity of offloaded products, even though no pathway has been identified with this phenotype.

Because DEBS TE was one of the first PKS TEs characterized, it has served as a model for TE activity in the PKS community. Consequently, there is a substantial amount of *in vivo* and *in vitro* data cataloguing its substrate selectivity.[41, 66, 68, 70, 111, 112, 114, 115] **Figure 2-10** summarizes the current state of characterization for the substrate selectivity for macrocyclization of 14-member rings by DEBS TE. Unlike the tyrocidine TE, where substrate pre-organization appears to drive macrocyclization, it appears that specific enzyme substrate interactions are required for macrocyclization.

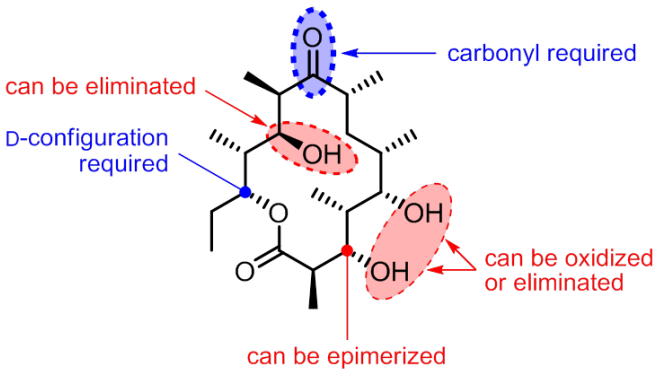


Figure 2-10. DEBS TE substrate tolerance map for modifications to the erythronolide core. Some elements of the erythronolide core are nonessential for DEBS TE mediated macrocyclization. The C2, 4, 6, 8, 10, 12 and 13 alkyl side chains can be omitted, the C11 hydroxyl can be reduced to an alkane, and the C3 and C5 hydroxyls can be oxidized to ketones or reduced to alkanes. The epimer at C3 is tolerated. No experiment has showed ring formation beyond 10-member rings for any analogues lacking the C6 methyl, the C9 carbonyl, or the D-configuration of the C13 hydroxyl.

Recent work has begun to decipher the enzyme-substrate interactions governing loading and release of substrates.[23, 24] Non-hydrolyzable acyl-enzyme intermediate models have been generated by treating the PikTE with diphenylphosphonate-based inhibitors. This leads to the formation of a phosphonate ester on the active site Ser. High-resolution crystal structure data from these covalently modified TEs has shown that the substrate interacts primarily through hydrophobic packing with the enzyme binding cavity, rather than hydrogen-bonding or dipole-dipole interactions.[23, 24] This binding mode has been further supported by kinetic characterization of mutants of the highly-related DEBS TE.[76] However, the use of the phosphonate ester

generates an acyl-enzyme mimic that most closely resembles the tetrahedral intermediate for substrate loading and may not accurately reflect the binding mode for macrocyclization.

Additionally, diphenylphosphonate addition products retain sufficient reactivity to undergo additional phosphoester hydrolysis and have been observed releasing the second phenol moiety or hydrolysing the phosphoserine bond in slow but substantial reversal of inhibition on time-scales relevant to protein crystallization.[116] Alkyl ester fluorophosphonate analogues may improve the stability of the analogues, but the use of active-site mutants such as the cysteine/arginine double mutant discussed in **2.1.3** or the incorporation of an unnatural amino acid to replace the active-site serine with an amine-analogue could permit direct trapping of the native compound in cases when the biosynthetic product is directly available.

High-resolution structural data of TEs with substrates arrested between the loading and release stages, akin to Dias *et al.*'s characterization of the type II TE FIK,[117] would likely give the most accurate measure of how the substrate interacts with the TE for release. NMR-based structural characterization of TEs could also provide exciting data on the detailed mechanism of catalysis.[118, 119] In particular, it offers the exciting prospect of measuring dynamics of the lid region, which is thought to play a role in substrate loading and release.

Prediction of TE offloading reaction-type has proven to be challenging.[41] Sequence-based approaches have not yet been able to distinguish hydrolyzing and macrocyclizing TEs. Using structural information, Jiang *et al.* proposed a model to distinguish α/β hydrolases with esterases activity from those with acyltransferase activity.[120] By examining over forty α/β hydrolase structures, they observed that the β-turn at the C-terminus of β-strand 3 (**Figure 2-2A**), which forms part of the oxyanion hole, adopted a type-II β-turn in hydrolases and a type-I β-turn in acyl transferases. They propose that the type-I β-turn deactivates water for attack by presenting the backbone NH bond, whereas the type-II β-turn presents a carbonyl that could potential activate water for attack (**Figure 2-11**).

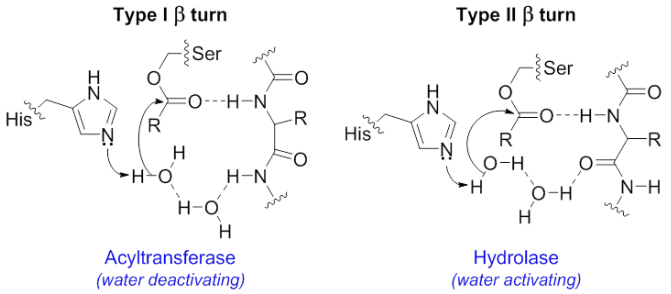


Figure 2-11. Proposed β-turn dependent water activation in α/β hydrolase active-sites. Type II turns orient the *i+1* carbonyl into the active-site, whereas type I turns cannot activate water for hydrolysis.[120] While the model may hold for NRPS TEs, most T1 PKS TEs do not fit this model due to insertion of additional residues into the turn.

While this analysis appears to fit well with NRPS TEs, such as the surfactin[28] and fengycin[22] TEs, PKS TEs, like the DEBS TE[18] and pikTE,[19] contain two residue insertions into this loop. These insertions dramatically change its orientation, rendering them unsuitable for fitting to this model. Potentially, the use of structural conformation predictors such as Phyre2,[121] or β-turn analyzers such as NetTurnP[122] on the current set of TE sequences may enable analysis of the β3 loop to determine if its conformation tracks with the release mechanisms of known TEs. However, as PKS TEs vary substantially in this loop, activity prediction with this model may not be possible.

An alternate structure-based model for prediction of hydrolysis versus macrocyclization, could involve the overall binding pocket size and geometry. Several X-ray crystal structures are available for both hydrolytic and macrocyclic TEs. Assuming macrocyclizing TEs require a larger, more spherical pocket than the rod-like hydrolytic pocket, a reactivity prediction could be based on overall volume of the cavity or the surface area/volume ratio. Using the crystal structure topology program CASTp,[123] 2 hydrolytic (curacin,[17] tautomycenin[29]) and 4 macrocyclizing (2 DEBS TE[18, 19] and 2 pikTE[19, 24]) structures were compared, but no clear difference could be resolved (Fig. I.1, Appendix I). In part, the topology must be refined to better define the pore entrance and eliminate artefacts related to lid repositioning during crystallization. However, the lack of structural data for a substantial number of PKS TEs limits the ability to develop new structural-based models for reactivity prediction.[7]

2.3.3 Macrocylization logic gate is relaxed in some fungal macrocyclizing TEs

Recent work has shown fungal TEs to have a relaxed substrate selectivity for macrocyclization as compared to T1 PKS TEs. For example, while T1 PKS TEs are highly stereoselective, *in vivo* and *in vitro* experiments have shown that TEs from resorcylic acid lactone (RAL) biosynthetic pathways can be highly stereotolerant for macrocyclization. In 2010, Zhou *et al.* used an *in vivo* precursor directed biosynthesis experiment where they fed both enantiomers of a synthetic substrate to the nrPKS module and TE from the radicicol biosynthetic pathway.[86] Both the D- and L-configured substrates were accepted and processed by the PKS domains and cyclized by the TE, generating the two enantiomeric macrocycles. This is in stark contrast to T1 PKS TEs, such as DEBS TE, which has shown very rigid stereoselectivity towards macrocyclization.[41, 115]

In vitro kinetic characterization of the TEs from radicicol and zearalenone biosynthesis has shown that these TEs have little-to-no kinetic selectivity for the configuration of the alcohol involved in macrocyclization. Radicicol and zearalenone are highly related 14-member RALs. Radicicol has a L-configured macrolactone, whereas zearalenone has an D-configured macrolactone. The TE from radicicol biosynthesis showed virtually no kinetic preference for macrocyclization of synthetic substrates with the D- versus L- configured alcohol. The ratio of the specificity constants for the native configuration versus the enantiomer was 0.88:1.[124] The zearalenone TE displayed slightly greater selectivity with the ratio of specificity constants for the D configured substrate over the native zearalenone L configuration of 5.4:1.[124] While this difference in specificity constants is small and insufficient to be used as a catalyst for resolution of racemates, both TEs show slightly enhanced activity for the non-native substrate nucleophile configuration.

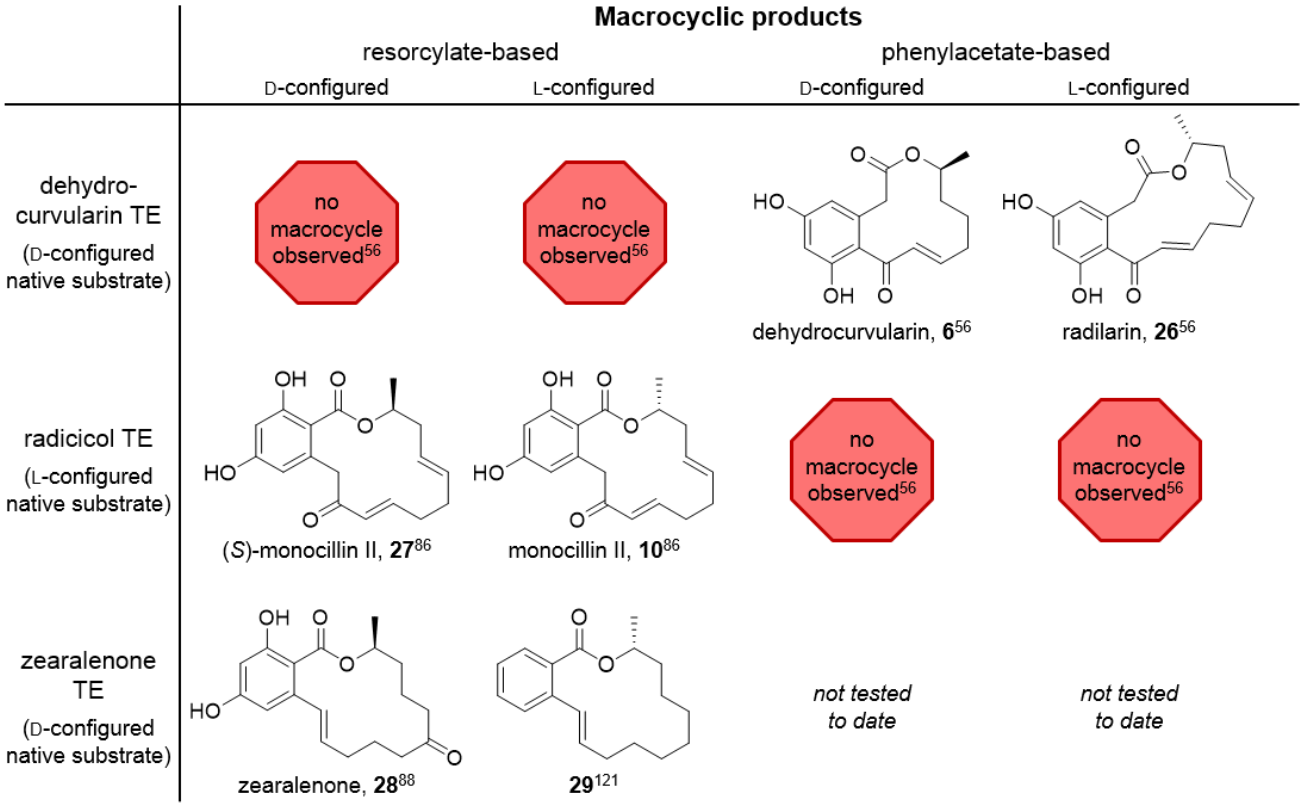


Figure 2-12. Fungal TEs tolerate nucleophile inversion but are PT specific. Examination of macroyclizing TEs from fungal PKS pathways shows them to be stereotolerant for macrocyclization and highly selective for resorcylate or phenylacetate functional groups.

In addition to the TEs from RAL pathways, a TE from the fungal phenylacetic acid lactone pathway for dehydrocurvularin has been characterized. Similar to the RAL TEs, the dehydrocurvularin TE was stereotolerant.[57] Dehydrocurvularin has a native 12-member L-configured lactone, **26** (**Figure 2-12**). By combining the hrPKS from radicicol biosynthesis with the nrPKS from dehdyrocurvularin biosynthesis, it was possible to present the dehydrocurvularin TE with a D-configured substrate which was effectively macrocyclized, generating the 14-member lactone **27**.

The TEs from RAL biosynthesis and phenylacetic acid lactone biosynthesis display similar trends (**Figure 2-12**). Both appear to be stereotolerant, macrocyclizing both D- and L-configured substrates. In addition both RAL TEs and phenylacetic acid lactone TEs appear to be highly selective for the position of the aromatic ring. The RAL TEs macrocyclize compounds that have a resorcylated group, generated by their F-type PT domain, but are unable to macrocyclize phenylacetic acid containing substrates. Similarly, the dehydrocurvularin TE macrocyclizes substrates possessing a phenylacetic acid group, generated by its S-

type PT domains, but is unable to macrocyclize resorcylate containing substrates. Based on these observations I predict that apralactone A, a macrolide isolated from a fungal genus *Curvularia*, which contains a D-configured phenylacetic acid lactone,[125] will be produced by a stereotolerant TE also able to macrocycle both D- and L-configured substrates. In addition, based on the trend seen with the RALs and dehydrocurvularin, I predict that the apralactone TE will not be able to macrocyclize RAL substrates.

The structural similarity between RAL natural products, such as radicicol, zearalenone, and hypothemycin, suggests their biosynthetic pathways diverged from a common genetic ancestor.[124] One of the stages in the divergence of these pathways was the change in stereochemistry of the macrolactone alcohol. This alcohol is introduced by the KR domain of the upstream hrPKS. The cryptic programing of RAL hrPKSs is challenging to decipher.[126] In the case of hypothemycin, the KR domain of the hrPKS reduces one intermediate β-ketothioester to produce a D-configured alcohol and reduces a subsequent β-ketothioester intermediate to produce an L-configured alcohol. Zhou *et al.* demonstrated that this hypothemycin KR D-selective reduction could be inverted by generating a chimeric KR domain with secondary structural elements from the radicicol hrPKS KR domain.[127] This study highlights the ease by which hrPKS systems can switch between D- and L-specific reductions and the capacity for diverse products. As the alcohol generated from this reduction is involved in macrocyclization, it is not unreasonable to assume the relaxed TE stereoselectivity is related to the ability of the KR domain to generate either stereogenic configuration. For example, if the TE was highly stereoselective as is seen in the type I PKS TEs, a change in the stereogenic outcome of the KR domain reduction could lead to a compound that does not cyclize, decreasing the likelihood that the compound would be bioactive. This outcome would likely not increase the producing organism's fitness and would be readily outcompeted by a pathway with a stereotolerant TE able to cyclize structures with either stereogenic configuration.

2.3.4 Some fungal TEs restrict release to Claisen carbon-carbon bond formation

TEs from some non-reducing iterative fungal pathways catalyze Claisen-like carbon-carbon bond formation to generate aromatic rings.[16] Here, the nucleophilic carbon is generated from tautomerization or deprotonation of a 1,3-diketone positioned such that attack on the acyl-enzyme intermediate carbonyl leads to 6-member ring formation. Subsequent tautomerization and release from the enzyme generates an aromatic

ring. In these pathways, a key feature of the TE is to disfavour pyrone formation from the enol or enolate.[16, 128]

One of the best studied of these TEs, often called Claisen-like cyclases (CLC), is from aflatoxin biosynthesis. Based on high-resolution structural data, molecular modelling studies, and mutational analyses, Korman *et al.* propose a detailed model for TE/CLC activity (**Figure 2-13**).[16] They propose that the ACP-bound thioester substrate is loaded on to the active site Ser and the departing thiolate of the Ppant arm deprotonates the histidinium ion from the catalytic triad. Upon departure of the Ppant arm, the hexyl-β-diketo side chain of the substrate can undergo free rotation into the cavity previously occupied by the Ppant arm. The hexyl chain is buried between the lid and the TE core, bringing the α-carbon from the β-diketo group in close proximity to the acyl-enzyme carbonyl and catalytic His. Deprotonation of the α-carbon by the His, promotes substrate offloading via Claisen cyclization. Mutation of the catalytic His to Phe drastically reduced cyclase activity, leading to the vast majority of product being offloaded as the pyrone via C-O bond formation. Presumably pyrone formation is occurring on the acyl-ACP intermediate since in the absence of a TE, the acyl-ACP intermediate is spontaneously offloaded as the corresponding pyrone, suggesting that this is the favoured non-enzymatic pathway.[75]

This last example highlights the TE's role in regulating which release route is selected. This is similar to what is seen for DEBS TE where the TE drives release via macrocyclization in place of the more stable and kinetically favoured 6-member lactone. In aflatoxin biosynthesis, the presence of the TE attenuates the spontaneous release via pyrone formation, and facilitates C-C bond forming release in its place. Unlike the loading stage, substrate gating for elongation or tailoring is not observed at the release stage. It remains to be seen if TEs with permissive loading stages and highly selective release stages become kinetically defined by the rate of offloading.

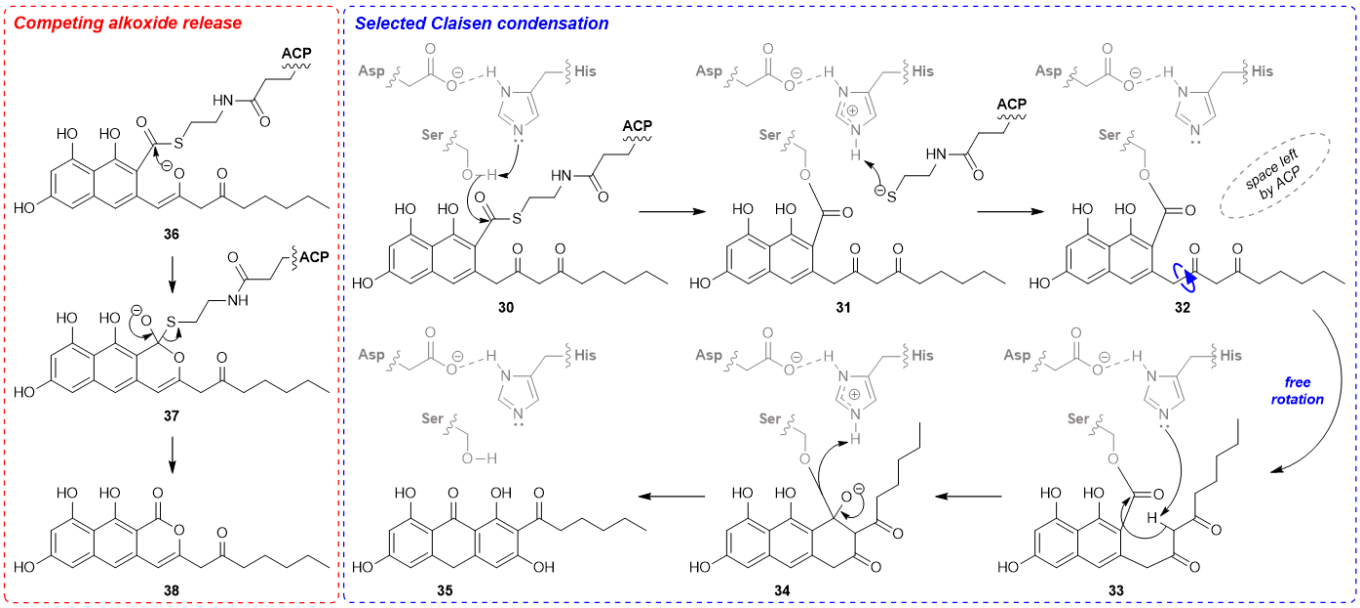


Figure 2-13. Claisen-condensation type fungal TEs control substrate reorganisation during C-C bond forming release. In the presence of the aflatoxin TE[16], the substrate loads via the standard mechanism, but after the ACP Ppant departs, the T promotes attack from C14 rather than the spontaneous attack of the proximal carbonyl oxygen in the alkoxide tautomer (alternate route shown in red).

2.4 Summary and Conclusions

The seven examples presented here highlight the range of loading and release selectivity seen across PKS and NRPS TEs. In a number of cases TEs are seen to faithfully process the majority of substrates, as in the loading of type I PKS TEs or the macrocyclization of gramicidin analogues. In other cases TEs show a high degree of selectivity for specific structural feature of the substrates, as in the case of loading the pyochelin TE or macrocyclization of polyketides by TEs. Understanding and predicting this selectivity is a major challenge. Herein, I summarize some of the key features of TEs and try to correlate these features to selectivity predictions.

TEs are α/β hydrolases and their structures adopt a β sheet core that is twisted in a left-handed helix. The active site is accessible through a trough that is capped with a highly variable lid, the most diverse part of the TE sequence. The channel leading to the catalytic triad terminates at the nucleophilic elbow, the vertex of β-strand 5 and α-helix C (**Figure 2-2**) where the active site Ser is found. The largest factor in sequence diversity associated with TEs appears to be the nature of the

module directly upstream of the TE (**Figure 2-4**), rather than the type of reaction it catalyzes or the substrate selectivity it possesses. While phylogenetic analysis of the full sequence does not predict activity, it remains to be seen if a phylogenetic analysis of TE lids or specific structural units can produce a model capable of predicting channel geometry or release route.

With the large number of TE sequences in the literature, there are well developed HMM for the detection of this family of proteins. Our phylogenetic analysis shows that the seed sequences for the Pfam TE HMM (http://pfam.xfam.org/family/PF00975/alignment/seed) appear to be somewhat biased with few examples from NRPS TEs (**Figure 2-4**). This bias can lead to false negatives in the detection of TEs by bioinformatic tools that rely on this model, such as antiSMASH.[44] Re-examining the cut-off scores for the current TE Pfam model or refining the TE HMM is likely to incrementally improve the effectiveness of bioinformatic TE screens. It should be noted however that this diversity discussion is based on the assumption that newly discovered TEs will be sufficiently similar in sequence to those that are already characterized, such that they will be recognized by the model. Identifying outliers and incorporating them into the model is essential to maintaining a useful predictive tools as well as providing insight in to TE structure-function relationships.

TE reaction prediction substantially lags behind other PKS and NRPS domains.[41] The activity and selectivities of many PKS and NRPS domains can be effectively predicted and are incorporated into bioinformatic tools like anitSMASH. This enables researchers to rapidly and accurately annotate PKS and NRPS biosynthetic pathways. While in many cases these tools can predict meaningful structural aspects of the final product, the lack of ability to predict release reaction type continues to limit overall structure prediction. Advances in this area will greatly impact natural product discovery approaches based on genome mining.[129]

Examples in this chapter show that CPs interact with TEs through helix three of CP, and this helix is not used to mediate other CP-protein interactions. While CP and TE conformations do not appear to change significantly during interaction, there are discrete interaction between the Ppant

arm and the TE that mediate substrate loading. Of note, while non-native substrates activated as *N*-acetylcysteamine derivatives lack some of these key Ppant-TE interactions, they have proven to be extremely useful in characterizing TE activity, providing data that is essential for deciphering the role of TEs in different pathways.

2.4.1 TE logic that impacts loading and release decisions

Several examples in this chapter show that some TEs differentially process substrates during loading and release based on localized substrate properties. Importantly, in each example a different structural property of the substrate was tested for by the TE. In addition, properties tested for by one TE were typically ignored by other TEs. While PKS and NRPS TEs appear to be highly general in their ability to load different substrates some key exceptions emerged. The pyochelin TE shows strong selectivity for thiazole containing substrates. The TE thus appears to be acting as a checkpoint for modification of the ACP-bound substrate, prior to TE loading (**Figure 2-6**). It also appears that iterative fungal TEs are functionally linked to the activity of their cognate PT, which do not offload substrates that are precursors to the PT domain,[56] nor structures generated from alternative ring type processing by the PT (**Figure 2-7**).[57] Current evidence supports this being a loading step selectivity, though additional *in vitro* evidence is required to fully confirm this. Conserved among these examples of TEs acting as logic gates and restricting loading of a substrate is the requirement for ACP-bound intermediates to be extensively processed. I thus hypothesize that in systems where processing of terminal ACP-bound intermediate is required, the TE will exhibit substantial substrate selectivity for loading, acting as logic gates to control the fidelity of the pathway.

Release strategies show some TEs to be highly selective in the reaction they catalyze to offload acyl-TE intermediates. The model bacterial type I PKS macrocyclizing DEBS TE is lenient with regard to a number of structural changes to the substrate, such as removal of the methyl side chains, oxidation or elimination of many of the alcohols, and alterations in ring-size, but as with other characterized type I macrocyclizing TEs, macrocycle forming activity is highly selective for the

absolute configuration of the nucleophilic alcohol (**Figure 2-10**). Fungal TE/CLCs which catalyze carbon-centered nucleophile attack on the acyl-TE intermediate also demonstrated high selectivity for offloading reaction principally however by limiting access of oxygen-centered nucleophiles to the acyl-TE bond (**Figure 2-13**).

2.4.2 TEs as senseless catalysts

In some cases, TEs are victims of fate and will process all substrates presented to them. For example, when exposed to a potential substrate bound to the immediately upstream CP, most type I PKS and NRPS TEs load the acyl or peptidyl substrate without fail. Similarly, the TE from the tyrocidine NRPS pathway was able to effect macrocyclization of all substrates that pre-organized in the absence of the TE to place the intramolecular nucleophile in close proximity to the electrophilic activated ester (**Error! Reference source not found.**). In this case, it is the substrates themselves, rather than the TE that determines if the compounds are release via hydrolysis or macrocyclization thus the TE is exerting no selectivity.

In contrast to the tight regulation of nucleophile stereochemistry seen in macrocyclizing bacterial type I PKS TEs, fungal macrocyclizing TEs are indiscriminate of nucleophile absolute stereochemistry (**Figure 2-12**). This observation suggests that there is something fundamentally different about how the fungal TEs gate the release stage with regards to nucleophile stereochemistry. Understanding the discrete enzyme-substrate interactions that govern the selectivity in bacterial type I PKS TEs and facilitate the lack of selectivity in fungal macrocyclizing TEs is thus a key question moving forward. Insights in this area will be particularly useful for the development of new biocatalysts from TEs.

Historically, TEs as well as many other biosynthetic domains have been characterized by perturbing their pathways though deletion mutants or formation of chimeric pathways and observing the resulting products or by steady-state kinetic characterization of isolated catalytic domains. Higher resolution analysis is becoming needed to answer further questions about these processes. For example, in the context of the TE what is the role of movement of the clearly

dynamic lid region? Are motions of the lid region on the same time scale as the rate constant for TE reactions? Does the motion of the lid facilitate entry of the loaded Ppant arm into the active site cavity? Is it required for release of the Ppant arm after transfer of the substrate or for product release? Insight into these questions will enable the community to better understand offloading and its overall impact on pathway flux and fidelity.

2.4.3 The evolution of loading and release selectivity

Evolution provides a useful lens for interpreting and rationalizing the function and selectivities of biosynthetic pathways. For example, the evolution of two sequentially acting catalytic domains from a pathway can be linked. If variation in one domain varies the structure of the intermediate such that the next domain cannot process it, the pathway yields no product. If the second domain is sufficiently tolerant to accommodate the substrate change, the new pathway can produce a new compound, which may or may not increase the fitness of the producing organism.

In light of the screening hypothesis[59, 60] that starts with the premise that potent biological activity is a rare property of a natural product and predicts that evolution will favour organisms that can generate diversity at a low cost, one can examine the selectivity of TEs. As a late acting enzyme in a pathway, the loading step is predicted to occur with minimal substrate selectivity, ensuring that the intermediates are processed into natural products that can be screened for biological activity. Similarly one can predict that the release step should occur with limited selectivity for the type of reaction catalyzed as this will maximize the chemical diversity accessible from the pathway, increasing the chances of identifying an active compound.

Both of these predictions suggest that the TE should not act as a logic gate for PKS and NRPS biosynthetic pathways but rather, it should behave as a victim of fate, processing all substrates it is supplied with. Herein, however, I have highlighted examples where the TE clearly has a role in controlling fidelity of the pathway through substrate selectivity. I thus hypothesize that these examples of selectivity emerge from selection pressures to retain biological activity for the natural product produced by the pathway or a natural product produced by an ancestral pathway. I propose

the selective pressure for generation of the active natural product led to the evolution of TE selectivity ensuring release of the active compound.

The stereoselectivity of macrocyclizing TEs also provides an intriguing case study for this hypothesis. The RAL TEs show the predicted lack of stereoselectivtiy for late stage enzymes and both D- and L-configured lactone natural products, such as radicicol[130] and zearalenone,[131] exhibit potent though different biological activity. This is consistent with the screening hypothesis. The bacterial type I PKS TE, DEBS TE, which macrocyclizes an erythromycin precursor, shows high stereoselectivity, macrocyclizing only D-configured alcohols. Erythromycin binds to the ribosome[132] inhibiting protein synthesis. Examination of the high-resolution structural model for erythromycin binding shows interaction between the C13 ethyl substituent and the ribosome suggesting that the 13-epi erythromycin A would be unable to bind, and thus inactive. If this is the case, than the stereoselectivity of the DEBS TE may have arisen to ensure that the active diastereomer of the natural product is produced.

Fungal macrocyclizing TEs are selective for type of aromatic system present in the substrate, either resorcylate or phenylacatete. In these systems, the PT domain and the TE need to be matched to effect macrocyclization. The radicicol TE requires a resocylate substrate such as the one generated by its cognate PT. The dehydrocurvularin TE requires a phenylacetate substrate, as generated by its PT, for effective macrocyclization to occur. Based on our above analyses, I predict that the substrate selectivity of these TEs indicates that the resorcylate and phenylacatete structures are required for biological activity. Intriguingly both of these structural scaffolds give rise to active fungal polyktide natural products. Thus why are the TEs selective for one scaffold over the other?

Examining the hypothetical product of a non-selective TE from one of these systems can be insightful in answering this question. The compound radilarin (**26**), which is formed by swapping the PT and TE domains from the radicicol biosynthetic pathway with those from dehydrocurvularin biosynthesis,[57] approximates mutation of the PT domain from F- to S-type in the presence of a TE

that tolerates both rings. Radilarin shows minimal cyctotoxicity activity[133], suggesting that a compound from a pathways with a tolerant TE would still be active and thus be able to impact the fitness of the producing organism fitness. This observation in conjunction with the above analyses suggests that the TE should not be selective for the type of aromatic substrate. However radilarin is much less active than radicicol in HSP90 inhibition, the most potent known activity for radicicol. If HSP90 inhibition is required to improve the fitness of the producing organism, the selectivity of the TE would be expected and consistent with our interpretation of the screening hypothesis. Ultimately resolving this question requires identifying the role polyketide products play in the biology of producing organisms. What is clear is that control of the aromatic ring position is a crucial aspect in determining the fitness of the resorcylate and phenylacate lactone pathways and their producing organisms.

2.5 References

1. M. E. Horsman, T. P. A. Hari and C. N. Boddy, *Natural Product Reports*, 2016, **33**, 183-202.
2. P. Heikinheimo, A. Goldman, C. Jeffries and D. L. Ollis, *Structure (London, England : 1993)*, 1999, **7**, R141-R146.
3. M. Holmquist, *Current Protein & Peptide Science*, 2000, **1**, 209-235.
4. D. C. Cantu, Y. F. Chen and P. J. Reilly, *Protein Science*, 2010, **19**, 1281-1295.
5. L. C. Du and L. L. Lou, *Natural Product Reports*, 2010, **27**, 255-278.
6. R. M. Kohli and C. T. Walsh, *Chemical Communications*, 2003, 297-307.
7. F. Kopp and M. A. Marahiel, *Natural Product Reports*, 2007, **24**, 735-749.
8. W. Xu, K. Qiao and Y. Tang, *Critical Reviews in Biochemistry and Molecular Biology*, 2013, **48**, 98-122.
9. A. Koglin and C. T. Walsh, *Natural Product Reports*, 2009, **26**, 987-1000.
10. A. T. Keatinge-Clay, *Natural Product Reports*, 2012, **29**, 1050-1073.
11. Y. Q. Cheng, J. M. Coughlin, S. K. Lim and B. Shen, in *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates*, ed. D. A. Hopwood, Elsevier Academic Press Inc, San Diego2009, vol. 459, pp. 165-186.
12. N. P. Keller, G. Turner and J. W. Bennett, *Nature Reviews Microbiology*, 2005, **3**, 937-947.
13. M. L. Heathcole, J. Staunton and P. F. Leadlay, *Chemistry & Biology*, 2001, **8**, 207-220.
14. D. Schwarzer, H. D. Mootz, U. Linne and M. A. Marahiel, *Proceedings of the National Academy of Sciences of the United States of America*, 2002, **99**, 14083-14088.
15. K. Jensen, H. Niederkruger, K. Zimmermann, A. L. Vagstad, J. Moldenhauer, N. Brendel, S. Frank, P. Poplau, C. Kohlhaas, C. A. Townsend, M. Oldiges, C. Hertweck and J. Piel, *Chemistry & Biology*, 2012, **19**, 329-339.
16. T. P. Korman, J. M. Crawford, J. W. Labonte, A. G. Newman, J. Wong, C. A. Townsend and S. C. Tsai, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 6246-6251.

17. J. J. Gehret, L. Gu, W. H. Gerwick, P. Wipf, D. H. Sherman and J. L. Smith, *Journal of Biological Chemistry*, 2011, **286**, 14445-14454.
18. S. C. Tsai, L. J. W. Miercke, J. Krucinski, R. Gokhale, J. C. H. Chen, P. G. Foster, D. E. Cane, C. Khosla and R. M. Stroud, *Proceedings of the National Academy of Sciences of the United States of America*, 2001, **98**, 14808-14813.
19. S. C. Tsai, H. Lu, D. E. Cane, C. Khosla and R. M. Stroud, *Biochemistry*, 2002, **41**, 12598-12606.
20. D. P. Frueh, H. Arthanari, A. Koglin, D. A. Vosburg, A. E. Bennett, C. T. Walsh and G. Wagner, *Nature*, 2008, **454**, 903-906.
21. Y. Liu, T. Zheng and S. D. Bruner, *Chemistry & Biology*, 2011, **18**, 1482-1488.
22. S. A. Samel, B. Wagner, M. A. Marahiel and L.-O. Essen, *Journal of Molecular Biology*, 2006, **359**, 876-889.
23. J. W. Giraldes, D. L. Akey, J. D. Kittendorf, D. H. Sherman, J. L. Smith and R. A. Fecik, *Nature Chemical Biology*, 2006, **2**, 531-536.
24. D. L. Akey, J. D. Kittendorf, J. W. Giraldes, R. A. Fecik, D. H. Sherman and J. L. Smith, *Nature Chemical Biology*, 2006, **2**, 537-542.
25. J. R. Whicher, G. Florova, P. K. Sydor, R. Singh, M. Alhamadsheh, G. L. Challis, K. A. Reynolds and J. L. Smith, *Journal of Biological Chemistry*, 2011, **286**, 22558-22569.
26. S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh and M. T. Stubbs, *Structure (London, England : 1993)*, 2002, **10**, 301-310.
27. A. Tanovic, S. A. Samel, L. O. Essen and M. A. Marahiel, *Science (New York, N.Y.)*, 2008, **321**, 659-663.
28. A. Koglin, F. Lohr, F. Bernhard, V. V. Rogov, D. P. Frueh, E. R. Strieter, M. R. Mofid, P. Guntert, G. Wagner, C. T. Walsh, M. A. Marahiel and V. Dotsch, *Nature*, 2008, **454**, 907-911.
29. J. B. Scaglione, D. L. Akey, R. Sullivan, J. D. Kittendorf, C. M. Rath, E. S. Kim, J. L. Smith and D. H. Sherman, *Angewandte Chemie - International Edition*, 2010, **49**, 5726-5730.
30. Z. Zhou, J. R. Lai and C. T. Walsh, *Chemistry & Biology*, 2006, **13**, 869-879.
31. L. Tran, M. Tosin, J. B. Spencer, P. F. Leadlay and K. J. Weissman, *Chembiochem : a European journal of chemical biology*, 2008, **9**, 905-915.
32. L. Tran, R. W. Broadhurst, M. Tosin, A. Cavalli and K. J. Weissman, *Chemistry & Biology*, 2010, **17**, 705-716.
33. G. Yadav, R. S. Gokhale and D. Mohanty, *PLOS Computational Biology*, 2009, **5**.
34. C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, *BMC Evolutionary Biology*, 2007, **7**.
35. M. Roettig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, *Nucleic Acids Research*, 2011, **39**, W362-W367.
36. T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chemistry & Biology*, 1999, **6**, 493-505.
37. G. L. Challis, J. Ravel and C. A. Townsend, *Chemistry & Biology*, 2000, **7**, 211-224.
38. G. Yadav, R. S. Gokhale and D. Mohanty, *Journal of Molecular Biology*, 2003, **328**, 335-363.
39. P. Caffrey, *Chembiochem : a European journal of chemical biology*, 2003, **4**, 654-657.
40. M. Z. Ansari, J. Sharma, R. S. Gokhale and D. Mohanty, *BMC Bioinformatics*, 2008, **9**.
41. T. P. A. Hari, P. Labana, M. Boileau and C. N. Boddy, *Chembiochem : a European journal of chemical biology*, 2014, **15**, 2656-2661.
42. K. R. Conway and C. N. Boddy, *Nucleic Acids Research*, 2013, **41**, D402-D407.
43. R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. Sonnhammer, J. Tate and M. Punta, *Nucleic Acids Research*, 2014, **42**, D222-230.
44. K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano and T. Weber, *Nucleic Acids Research*, 2013, **41**, W204-212.

45. S. Müller, S. Rachid, T. Hoffmann, F. Surup, C. Volz, N. Zaburannyi and R. Müller, *Chemistry & Biology*, **21**, 855-865.
46. J. D. Thompson, D. G. Higgins and T. J. Gibson, *Nucleic Acids Research*, 1994, **22**, 4673-4680.
47. S. Whelan and N. Goldman, *Molecular Biology and Evolution*, 2001, **18**, 691-699.
48. A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie and O. Gascuel, *Nucleic Acids Research*, 2008, **36**, W465-W469.
49. A. Ferscht, *Enzyme structure and mechanism*, W. H. Freeman and Company, New York, 1985.
50. N. M. Gaudelli and C. A. Townsend, *Nature Chemical Biology*, 2014, **10**, 251-258.
51. L. Serino, C. Reimann, P. Visca, M. Beyeler, V. DellaChiesa and D. Haas, *Journal of Bacteriology*, 1997, **179**, 248-257.
52. S. T. Pullan, G. Chandra, M. J. Bibb and M. Merrick, *BMC Genomics*, 2011, **12**.
53. J. Li, R. Szittner, Z. S. Derewenda and E. A. Meighen, *Biochemistry*, 1996, **35**, 9967-9973.
54. A. Witkowski, H. E. Witkowska and S. Smith, *Journal of Biological Chemistry*, 1994, **269**, 379-383.
55. J. L. Meier, T. Barrows-Yano, T. L. Foley, C. L. Wike and M. D. Burkart, *Molecular Biosystems*, 2008, **4**, 663-671.
56. M. Wang, H. Zhou, M. Wirz, Y. Tang and C. N. Boddy, *Biochemistry*, 2009, **48**, 6288-6290.
57. Y. Q. Xu, T. Zhou, S. W. Zhang, L. J. Xuan, J. X. Zhan and I. Molnar, *Journal of the American Chemical Society*, 2013, **135**, 10783-10791.
58. A. O. Zabala, Y.-H. Chooi, M. S. Choi, H.-C. Lin and Y. Tang, *ACS Chemical Biology*, 2014.
59. R. D. Firn and C. G. Jones, in *Phytochemical Diversity and Redundancy in Ecological Interactions*, eds. J. T. Romeo, J. A. Saunders and P. Barbosa1996, vol. 30, pp. 295-312.
60. R. D. Firn and C. G. Jones, *Natural Product Reports*, 2003, **20**, 382-391.
61. T. L. Schneider and C. T. Walsh, *Biochemistry*, 2004, **43**, 15946-15955.
62. R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2003, **42**, 72-79.
63. S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson and L. Katz, *Science (New York, N.Y.)*, 1991, **252**, 675-679.
64. P. Power, T. Dunne, B. Murphy, L. Nic Lochlainn, D. Rai, C. Borissow, B. Rawlings and P. Caffrey, *Chemistry & Biology*, 2008, **15**, 78-86.
65. S. Donadio, J. B. McAlpine, P. J. Sheldon, M. Jackson and L. Katz, *Proceedings of the National Academy of Sciences of the United States of America*, 1993, **90**, 7119-7123.
66. J. R. Jacobsen, C. R. Hutchinson, D. E. Cane and C. Khosla, *Science (New York, N.Y.)*, 1997, **277**, 367-369.
67. J. Cortes, K. E. Wiesmann, G. A. Roberts, M. J. Brown, J. Staunton and P. F. Leadlay, *Science (New York, N.Y.)*, 1995, **268**, 1487-1489.
68. C. M. Kao, G. L. Luo, L. Katz, D. E. Cane and C. Khosla, *Journal of the American Chemical Society*, 1995, **117**, 9105-9106.
69. H. G. Menzella, R. Reid, J. R. Carney, S. S. Chandran, S. J. Reisinger, K. G. Patel, D. A. Hopwood and D. V. Santi, *Nature Biotechnology*, 2005, **23**, 1171-1176.
70. R. McDaniel, A. Thamchaipenet, C. Gustafsson, H. Fu, M. Betlach and G. Ashley, *Proceedings of the National Academy of Sciences of the United States of America*, 1999, **96**, 1846-1851.
71. H. G. Menzella, J. R. Carney and D. V. Santi, *Chemistry & Biology*, 2007, **14**, 143-151.
72. L. Kellenberger, I. S. Galloway, G. Sauter, G. Bohm, U. Hanefeld, J. Cortes, J. Staunton and P. F. Leadlay, *Chembiochem : a European journal of chemical biology*, 2008, **9**, 2740-2749.
73. R. McDaniel, S. Ebertkhosla, D. A. Hopwood and C. Khosla, *Science (New York, N.Y.)*, 1993, **262**, 1546-1550.
74. G. Illuminati and L. Mandolini, *Accounts of Chemical Research*, 1981, **14**, 95-102.

75. A. J. Hughes, J. F. Detelich and A. T. Keatinge-Clay, *Medicinal Chemistry Communications*, 2012, **3**, 956-959.
76. M. Wang and C. N. Boddy, *Biochemistry*, 2008, **47**, 11793-11803.
77. K. K. Sharma and C. N. Boddy, *Bioorganic & Medicinal Chemistry Letters*, 2007, **17**, 3034-3037.
78. R. S. Gokhale, D. Hunziker, D. E. Cane and C. Khosla, *Chemistry & Biology*, 1999, **6**, 117-125.
79. M. Wang, P. Opare and C. N. Boddy, *Bioorganic & Medicinal Chemistry Letters*, 2009, **19**, 1413-1415.
80. H. X. Lu, S. C. Tsai, C. Khosla and D. E. Cane, *Biochemistry*, 2002, **41**, 12590-12597.
81. J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel and C. T. Walsh, *Nature*, 2000, **407**, 215-218.
82. R. M. Kohli, C. T. Walsh and M. D. Burkart, *Nature*, 2002, **418**, 658-661.
83. T. A. Mukhtar, K. P. Koteva and G. D. Wright, *Chemistry & Biology*, 2005, **12**, 229-235.
84. M. Peschke, K. Haslinger, C. Brieke, J. Reinstein and M. J. Cryle, *Journal of the American Chemical Society*, 2016, **138**, 6746-6753.
85. T. Duerfahrt, K. Eppelmann, R. Mueller and M. A. Marahiel, *Chemistry & Biology (Cambridge)*, 2004, **11**, 261-271.
86. H. Zhou, K. J. Qiao, Z. Z. Gao, J. C. Vederas and Y. Tang, *Journal of Biological Chemistry*, 2010, **285**, 41412-41421.
87. Y. Q. Xu, P. Espinosa-Artiles, V. Schubert, Y. M. Xu, W. Zhang, M. Lin, A. A. L. Gunatilaka, R. Sussmuth and I. Molnar, *Applied and Environmental Microbiology*, 2013, **79**, 2038-2047.
88. I. Gaffoor and F. Trail, *Applied and Environmental Microbiology*, 2006, **72**, 1793-1799.
89. D. W. Udwary, M. Merski and C. A. Townsend, *Journal of Molecular Biology*, 2002, **323**, 585-598.
90. R. Thomas, *Chembiochem : a European journal of chemical biology*, 2001, **2**, 612-627.
91. S. Gavalda, F. Bardou, F. Laval, C. Bon, W. Malaga, C. Chalut, C. Guilhot, L. Mourey, M. Daffe and A. Quemard, *Chemistry & Biology*, 2014, **21**, 1660-1669.
92. L. Ray, K. Yamanaka and B. S. Moore, *Angewandte Chemie - International Edition*, 2016, **55**, 364-367.
93. T. W. Giessen and M. A. Marahiel, *FEBS Letters*, 2012, **586**, 2065-2075.
94. J. Rizo and L. M. Gierasch, *Annu. Rev. Biochem.*, 1992, **61**, 387-418.
95. J. E. DeLorbe, J. H. Clements, B. B. Whiddon and S. F. Martin, *ACS Medicinal Chemistry Letters*, 2010, **1**, 448-452.
96. J. Grunewald, S. A. Sieber, C. Mahlert, U. Linne and M. A. Marahiel, *Journal of the American Chemical Society*, 2004, **126**, 17025-17031.
97. V. Miao, M. F. Coeffet-Le Gal, K. Nguyen, P. Brian, J. Penn, A. Whiting, J. Steele, D. Kau, S. Martin, R. Ford, T. Gibson, M. Bouchard, S. K. Wrigley and R. H. Baltz, *Chemistry & Biology*, 2006, **13**, 269-276.
98. G. Christiansen, W. Y. Yoshida, J. F. Blom, C. Portmann, K. Gademann, T. Hemscheidt and R. Kurmayer, *Journal of Natural Products*, 2008, **71**, 1881-1886.
99. N. A. Magarvey, Z. Q. Beck, T. Golakoti, Y. Ding, U. Huber, T. K. Hemscheidt, D. Abelson, R. E. Moore and D. H. Sherman, *ACS Chemical Biology*, 2006, **1**, 766-779.
100. C. Mahlert, S. A. Sieber, J. Gruenewald and M. A. Marahiel, *Journal of the American Chemical Society*, 2005, **127**, 9571-9580.
101. J. W. Trauger, R. M. Kohli and C. T. Walsh, *Biochemistry*, 2001, **40**, 7092-7098.
102. Y. Ding, C. M. Rath, K. L. Bolduc, K. Hakansson and D. H. Sherman, *Journal of the American Chemical Society*, 2011, **133**, 14492-14495.
103. F. Kopp, J. Grunewald, C. Mahlert and M. A. Marahiel, *Biochemistry*, 2006, **45**, 10474-10481.

104. I. Molnar, T. Schupp, M. Ono, R. E. Zirkle, M. Milnamow, B. Nowak-Thompson, N. Engel, C. Toupet, A. Stratmann, D. D. Cyr, J. Gorlach, J. M. Mayo, A. Hu, S. Goff, J. Schmid and J. M. Ligon, *Chemistry & Biology*, 2000, **7**, 97-109.
105. F. Sasse, H. Steinmetz, G. Hofle and H. Reichenbach, *Journal of Antibiotics*, 1995, **48**, 21-25.
106. W. H. Gerwick, P. J. Proteau, D. G. Nagle, E. Hamel, A. Blokhin and D. L. Slate, *Journal of Organic Chemistry*, 1994, **59**, 1243-1245.
107. T. J. Turbyville, E. M. K. Wijeratne, M. X. Liu, A. M. Burns, C. J. Seliga, L. A. Luevano, C. L. David, S. H. Faeth, L. Whitesell and A. A. L. Gunatilaka, *Journal of Natural Products*, 2006, **69**, 178-184.
108. C. M. Kao, L. Katz and C. Khosla, *Science (New York, N.Y.)*, 1994, **265**, 509-512.
109. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science (New York, N.Y.)*, 2001, **291**, 1790-1792.
110. B. J. Beck, Y. J. Yoon, K. A. Reynolds and D. H. Sherman, *Chemistry & Biology*, 2002, **9**, 575-583.
111. B. J. Dunn, K. R. Watts, T. Robbins, D. E. Cane and C. Khosla, *Biochemistry*, 2014, **53**, 3796-3806.
112. C. M. Kao, M. McPherson, R. N. McDaniel, H. Fu, D. E. Cane and C. Khosla, *Journal of the American Chemical Society*, 1997, **119**, 11339-11340.
113. J. He and C. Hertweck, *Chemistry & Biology*, 2003, **10**, 1225-1232.
114. C. M. Kao, G. L. Luo, L. Katz, D. E. Cane and C. Khosla, *Journal of the American Chemical Society*, 1996, **118**, 9184-9185.
115. A. Pinto, M. Wang, M. Horsman and C. N. Boddy, *Organic Letters*, 2012, **14**, 2278-2281.
116. P. Argyropoulos, F. Bergeret, C. Pardin, J. M. Reimer, A. Pinto, C. N. Boddy and T. M. Schmeing, *Biochimica et Biophysica Acta-General Subjects*, 2016, **1860**, 486-497.
117. M. V. B. Dias, F. Huang, D. Y. Chirgadze, M. Tosin, D. Spiteller, E. F. V. Dry, P. F. Leadlay, J. B. Spencer and T. L. Blundell, *Journal of Biological Chemistry*, 2010, **285**, 22495-22504.
118. A. J. Baldwin and L. E. Kay, *Nature Chemical Biology*, 2009, **5**, 808-814.
119. S. Hammes-Schiffer, *Biochemistry*, 2013, **52**, 2012-2020.
120. Y. Jiang, K. L. Morley, J. D. Schrag and R. J. Kazlauskas, *Chembiochem : a European journal of chemical biology*, 2011, **12**, 768-776.
121. L. A. Kelley and M. J. Sternberg, *Nature Protocols*, 2009, **4**, 363-371.
122. B. Petersen, C. Lundegaard and T. N. Petersen, *PLoS One*, 2010, **5**, e15079.
123. J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz and J. Liang, *Nucleic Acids Research*, 2006, **34**, W116-W118.
124. G. W. Heberlig, M. Wirz, M. Wang and C. N. Boddy, *Organic Letters*, 2014.
125. H. Greve, P. J. Schupp, E. Eguereva, S. Kehraus, G. Kelter, A. Maier, H.-H. Fiebig and G. M. Koenig, *European Journal of Organic Chemistry*, 2008, 5085-5092.
126. Y. Xu, T. Zhou, P. Espinosa-Artiles, Y. Tang, J. Zhan and I. Molnar, *ACS Chemical Biology*, 2014, **9**, 1119-1127.
127. H. Zhou, Z. Z. Gao, K. J. Qiao, J. J. Wang, J. C. Vederas and Y. Tang, *Nature Chemical Biology*, 2012, **8**, 331-333.
128. A. L. Vagstad, E. A. Hill, J. W. Labonte and C. A. Townsend, *Chemistry & Biology*, 2012, **19**, 1525-1534.
129. C. N. Boddy, *Journal of Industrial Microbiology and Biotechnology*, 2014, **41**, 443-450.
130. N. Winssinger, J. G. Fontaine and S. Barluenga, *Current Topics in Medicinal Chemistry*, 2009, **9**, 1419-1435.
131. G. G. Kuiper, J. G. Lemmen, B. Carlsson, J. C. Corton, S. H. Safe, P. T. van der Saag, B. van der Burg and J. A. Gustafsson, *Endocrinology*, 1998, **139**, 4252-4263.
132. F. Schlunzen, R. Zarivach, J. Harms, A. Bashan, A. Tocilj, R. Albrecht, A. Yonath and F. Franceschi, *Nature*, 2001, **413**, 814-821.

133. Y. Xu, T. Zhou, S. Zhang, P. Espinosa-Artiles, L. Wang, W. Zhang, M. Lin, A. A. L. Gunatilaka, J. Zhan and I. Molnár, *Proceedings of the National Academy of Sciences of the United States of America*, 2014, **111**, 12354-12359.

Chapter 3 Studies on the PKS/NRPS natural product pathway that produces bacillaene

3.1 Introduction

The bacterial Gram-positive genus *Bacillus* has received sustained research investigation in both academic and industrial settings for a multitude of applications (Bt delta endo toxin, barnase and barstar, *BamHI*, alpha amylase). For example, while traditional natto soybean fermentation has been practiced for at least 500 years, it has been markedly improved with the identification and isolation in the early 1900's of the straw-associated *Bacillus* species essential for the fermentation process.[1, 2] In large part, the fermentation properties are caused by high levels of secreted catalytic enzymes. In the 1930's and 40's, several *Bacillus* species were also identified as producers of secreted peptide antibiotics including polymixcin, bacitracin, gramicidin, and tyrocidine.

In the 1950's and 1960's additional compounds emerged from *Bacillus* isolates including surfactin, iturin, and fengycin. By the 1970's, the *Bacillus subtilis* protease subtilisin had been characterized and became one of the first secreted *Bacillus* enzymes produced industrially, and is produced in bulk as a commodity detergent. Additional proteases, lipases, and glycosidases have since been exploited for their alkaline stability and ease of production (facile culture, secretion-mediated isolation).

By the 1980's pharmaceutical companies actively searched environmental sources for novel *Bacillus*-like strains, and the genome structure of model *Bacillus* strains was studied in earnest. In 1985, Merck patented the production of the polyene difficidin in two strain of *Bacillus subtilis* (ATCC 39320 and ATCC 39374).[3] This was unusual for *Bacillus*, because by and large, polyketide-type compounds had previously been the domain of the actinomycetes whereas *Bacillus* produced ribosomal and NRPS peptides. This set the stage for the discovery of bacillaene.

.

In 1993, Scotti *et al*.[4] described a large open reading frame in *Bacillus subtilis* subspecies *subtilis* var. 168 that they noted resembled the PKS EryA from the deoxyerythromycin B polyketide pathway with the notable exception that no acyltransferase domains were found in the sequence. However no polyketide product was detected in cultures of the strain, and the ORF was termed PksX as an orphan PKS pathway. It has since been shown that the lack of a polyketide product from this strain is due to an inactive PPTase. In 1994, Bristol Myers Squibb patented the production of a polyene-containing secondary metabolite by a *Bacillus subtilis* soil isolate;[5, 6] they dubbed the newly detected compound bacillaene. In 1997, The *Bacillus subtilis* 168 draft genome was released,[7] and support for PksX's role in bacillaene or difficidin production accumulated, but also stirred concern about the ORF's noncanonical PKS architecture as compared to the model Type I modular in-*cis* acyl transferase DEBS.[8]

In the 10 years that followed, heterologous expression of domains of the 168 PksX pathway demonstrated that constituents of the pathway were metabolically active,[9] multiple strains of *Bacillus subtilis* (not including 168) were observed producing both bacillaene and difficidin at the same time in a PPTase-dependent manner,[10] and the use of degenerate primers matching roughly with 168's PksX domains inhibited bacillaene production in *B. subtilis* producer strains.[10] Sequencing of an isolate of *Bacillus amyloliquefaciens* (*B. amy*. FZB42) revealed a genome containing 3 polyketide pathways;[11] as with the *subtilis* strains, the newly sequenced *B. amy* FZB42 produced both bacillaene and difficidin, and when the *B. amy*. FZB42 PksX-ortholog was disrupted, bacillaene production ceased.

The bacillaene biosynthetic pathway is now a well-studied *trans*-AT PKS pathway. **Figure 3-1** outlines the domains of the bacillaene PKS and several of the key biosynthetic intermediates of the bacillaene natural product.

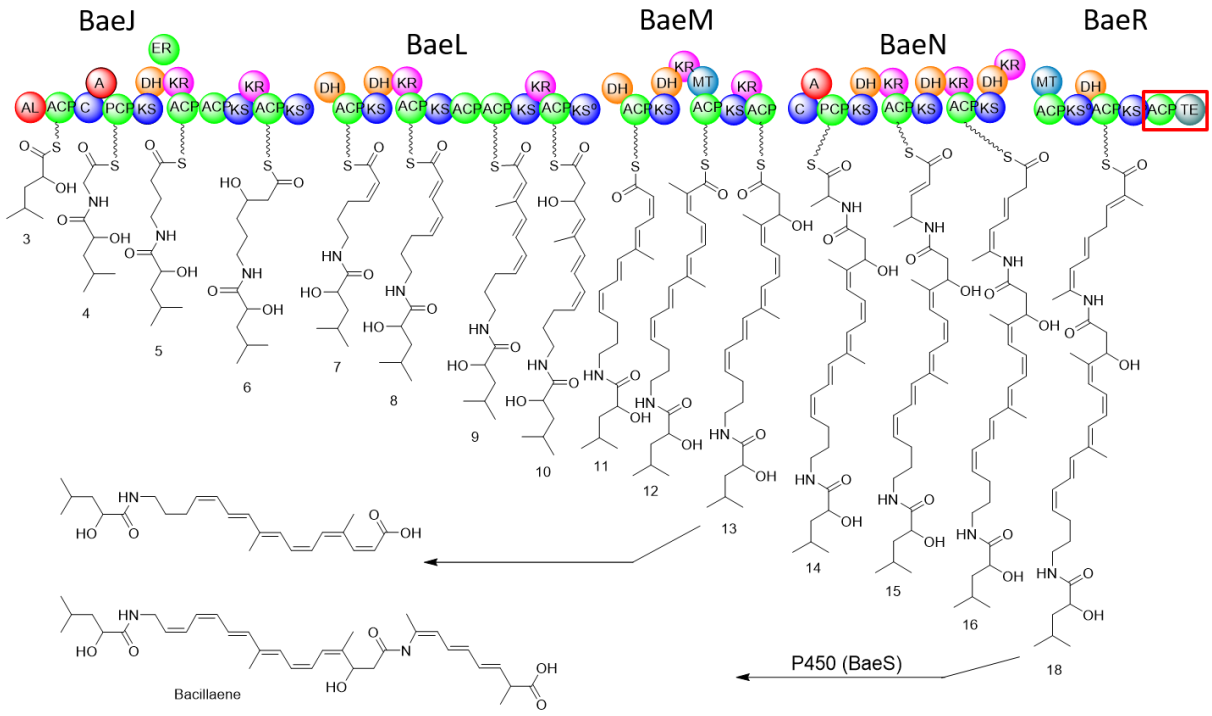


Figure 3-1 Features of the bacillaene biosynthetic pathway. Bacillaene is assembled by the action of NRPS-PKS hybrid polypeptides encoding 14 modules including 2 split modules. The polyene feature is reactive and has complicated bacillaene characterization as a result of oxidation and E/Z isomerization during product isolation.[12] Elements of the biosynthetic route are proposed chronologically from left to right (5' to 3' in a single reading frame in the genome) as the anchored metabolite progresses through the PKS. The *B. amy.* FZB42 bacillaene pathway matches the *B. subtilis.* 168 PksX except for the lack of an additional KS-ACP at the C-terminus of the 168 equivalent of BaeM. The ACP-TE didomain that forms the release module at the C-terminus of BaeR is highlighted in a red box. This region was amplified via PCR to produce pKJ13 (Katja Jensen, Piel Group). Figure adapted from Moldenhauer 2007.[12]

Several of the steps of biosynthesis of bacillaene are unusual compared to the canonical *cis* AT modular PKS biosynthetic method. The extender units of the bacillaene pathway are derived exclusively from malonyl CoA (without use of methyl malonyl CoA), and are units fed from a single in *trans* acyltransferase (*trans*-AT). The installed methyl groups at the α and β positions (based on the underpinning poly keto structure that defines polyketides) are installed post-

elongation via in *cis* SAM-dependent C-methyl transferase (MT) and in trans HMG-CoA synthase homolog domains (HCS) respectively. Second, the action of the first PKS module is completed by complementation by a *trans* ER domain from the stand-alone AT-ER didomain BaeE. Third, a hydroxyl group installed in the second PKS module of BaeJ is dehydrated by a downstream DH from another module. Fourth, unlike most PKS-derived methylene groups, the polyene system proximal to the carboxylate terminus of the molecule has been characterized in the β-γ position in product isolations.[12, 13]

These features were important for furthering the understanding of bacillaene's biosynthesis as well as serving as a model for the discovery, prediction, and compound-pathway matching of future *trans-AT* biosynthetic pathways. Biosynthetic evidence for these steps were derived from feeding experiments,[12] isolated expression of isolated domains,[14] and detection of blocked intermediates.[12] These approaches were based on successful studies carried out for *cis*-AT pathways, but another powerful tool, the repositioning of the TE "upstream" (further 5') to offload metabolic intermediate failed to perform as predicted by previous TE results.

Type 1 PKS follow a collinearity such that the compounds synthesized by the pathway are transferred between domains in the order that the domains appear within a protein, or often even between polypeptides in the same operon.[15-17] Tracking the composition of the tethered metabolite at different modules gives information about the order in which chemical transformations are carried out to give rise to the final molecule, and can help to decipher the biosynthetic route of a polyketide. The genetic fusion of a TE upstream in the pathway within the same module or fused to a preceding module in the pathway has served as an effective metabolic block in pathways including DEBS.[18]

This serves the same block as a domain deletion or point mutation, as well as enhancing the overall productivity of the initial pathway. The TE preferentially offloads intermediates tethered at the proximal ACP based on access to the phosphopantetheine tether as opposed to spontaneous hydrolysis, which is driven by the chemical stability of the acyl-enzyme intermediate bond and the

substrate residence time. This technique has also been adopted to study engineered PKS hybrids[19] and to produce enriched stereotopically dense synthetic intermediates.[20]

However, when Moldenauer *et al.* repositioned the bacillaene TE (BaeTE) upstream,[12] rather than a specific offloading of the proximate intermediate, multiple upstream intermediates were detected (**Figure 3-1**). Work by Jensen *et al.* showed similar release of multiple compounds in *trans-AT* pathway (pederin) and suggested that the released pathway intermediates were produced by the hydrolysis-action of an uncharacterized *trans-AT* domain.[21] Sequence similarity convinced the authors that the same editing was at play for the bacillaene truncants.

Their experience suggests that the approach used to study repositioned and engineered type I TEs from iterative and modular cis AT pathways may not work for TEs from *trans-AT* pathways. In order to establish whether bacillaene TE fusions are possible, and if so, what restrictions must be applied, several questions must be answered:

Does the polypeptide BaeR possess type I TE-mediated hydrolysis activity?

Can this activity be isolated to and maintained in a stand-alone TE construct with domain boundaries akin to cis AT type I modular TEs?

Does the TE display loading stage selectivity towards certain polyketides near the thioester bond?

Does the TE display loading stage selectivity towards certain linkers? In other words, does the TE require specific protein-protein interactions on the ACP structure that must be maintained for successful fusions?

The TE domain reacts at the linkage between substrate and protein tether. As a result, analogue substrates used in the study of TEs must mimic the thioester bond between the polyketide and its protein partner.

In vivo, the TE domain is located at the C-terminus of the final PKS elongation module. At the end of elongation, the final polyketide is bound to an ACP domain found within the same polypeptide as the TE. The polyketide is directly linked to phosphopantetheine via a thioester bond;

in turn, the phosphopantetheine (Ppant) linker is linked to the ACP via a phosphoester bond. Covalent linkage of the compound to the TE through the ACP in the native environment increases the effective concentration of the substrate, and protein positioning may increase favourable interactions as the TE approaches the substrate. Additionally, as discussed in **2.1.1**, some studies have suggested bonding interactions between the Ppant intermediary chain and the TE during the binding of substrate in the active site.[22, 23]

However, the catalytic reloading of synthetic substrates on ACPs is difficult to establish *in vitro* on a scale suitable for detection via visible light spectroscopy. Thus, the study was designed in two phases, with small-molecule thioester (SNAC) analogues to probe the kinetics and single-turnover studies of preloaded ACP tracked via liquid chromatography mass spectrometry (LC-MS). **Figure 3-2** compares ACP, CoA, and SNAC thioester linkages with highlights to the commonalities and differences between SNAC and native ACP linkages:

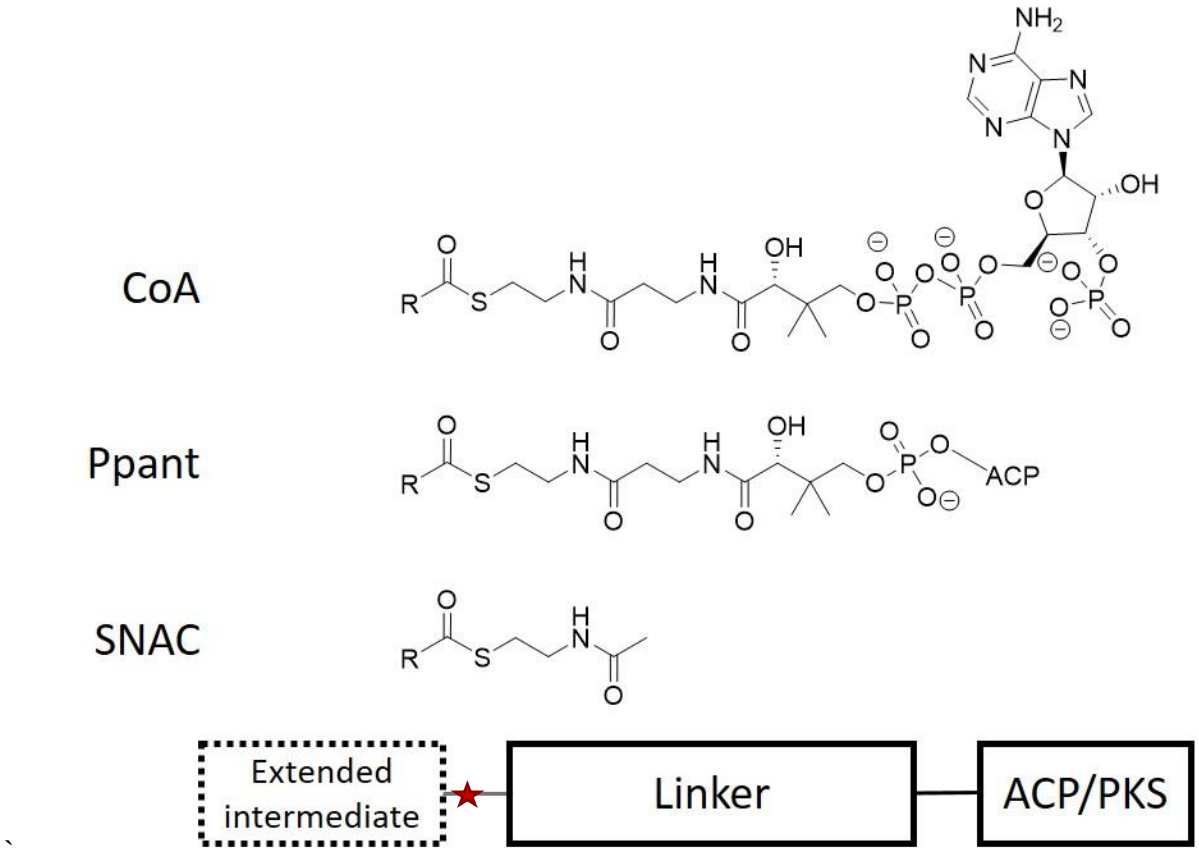


Figure 3-2 Common thioester linkages used *in vivo* and *in vitro* study of PKSs. The extended intermediate chain is presented in the TE active site (**Figure 2.3)** such that the linker region passes down the substrate pore and presents the thioester to the TE active-site Ser. The synthesized polyketide (R) is transferred to the TE in the thioester-cleaving Loading step. At bottom, the thioester bond is denoted with the red star.

In the native setting, the different domains of a type I PKS use different thioester substrates and follow different mechanisms. PPTases such as Sfp transfer an acylated and nonacylated CoA (**Figure 3-2**, top) to an *apo*-ACP active-site serine through the formation of a phosphoester bond and the release of adenine 3,5-diphosphate. The TE cleaves thioesters bond through a mechanism shown in **Figures 2-1** and **2-5** involving an ester intermediate, whereas the KS domain forms a new C-C bond via a cysteine thioester intermediate and a Claisen-type condensation**.** AT domains *trans*thioesterify acyl groups from CoA metabolites onto the free thiol of a Ppant-containing *holo*-ACP. The Ppant linker region is often represented with a wavy bond. Finally, after acyl removal

from an ACP, the Ppant remains as a coenzyme prosthetic group. As such, CoA thioesters are the native substrates of PPTases and ACP-linked thioesters are the native substrates of thioesterases. *N*-acetylcysteamine (SNAC) thioesters do not occur naturally, but PKS domains have exhibited cross-reactivity with these thioester phosphopantetheine analogues *in vitro*.[24] These SNAC analogues are processed by TEs at lower reactivity (higher K_M),[25, 26] but have been used to predict relative tolerance for multiple substrates in side-by-side studies. Their lower structural complexity and synthetic compatibility mean SNAC thioesters serve as useful tools for biosynthetic analysis.

The acyl chain of the biosynthetic intermediates of bacillaene also pose experimental barriers to analysis; the full-length bacillaene is light and oxidation sensitive,[27] demanding to synthesize chemically, and difficult to control isomerization when generating analogues from biosynthetically derived parent compound, so short SNAC acyl analogues of the tethered end of bacillaene were designed to test reactivity and selectivity of the BaeTE:

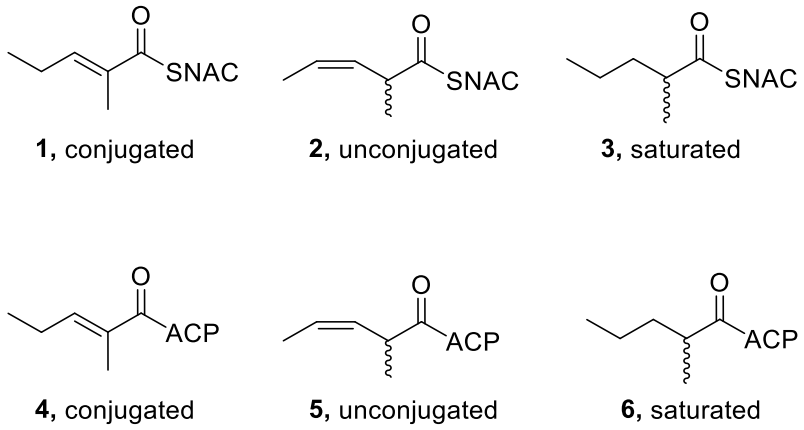


Figure 3-3 SNAC, CoA, and ACP analogues of bacillaene intermediates used as substrates. The substrates with the double bond out of conjugation with the terminal thioester (unconjugated) are hypothesized to be the closest analogues to the bacillaene intermediate that is presented to the TE in the native biosynthetic route. Structures of SNAC, CoA, and ACP thiol linkers are shown in **Figure 3-2**.

The results in **Chapter 3** describe the generation of isolated TE constructs from the bacillaene pathway, and their kinetic characterization *in vitro* against SNAC thioester analogues of bacillaene shown in **Figure 3-3**.

3.2 Results

An expression plasmid coding for the *B. amy* FZB42 ACP-TE didomain from module BaeR (see **Figure 3-1**) was kindly provided by Katja Jensen from Joern Piel's group in ETH Zurich. In order to obtain stand-alone BaeTE, we first constructed a BaeR TE expression vector.

[redacted]Initial BaeTE acyl-SNAC hydrolysis assays

In order to obtain stand-alone BaeTE to evaluate its *in vitro* activity with a panel of substrates shown in **Figure 3-3**, we first constructed a BaeTE expression vector. The BaeR TE plasmid was designed to contain the core TE sequence (outlined in **Figure I-3**) as well as a 29 amino acid leader sequence from the native BaeR module, and N- and C-terminal hexaHis tag sequences. The N-terminus was chosen for the native *SphI* endonuclease site, and this construct is 26 amino acids longer than the common model TE construct of DEBS TE from pRSG33.[28]

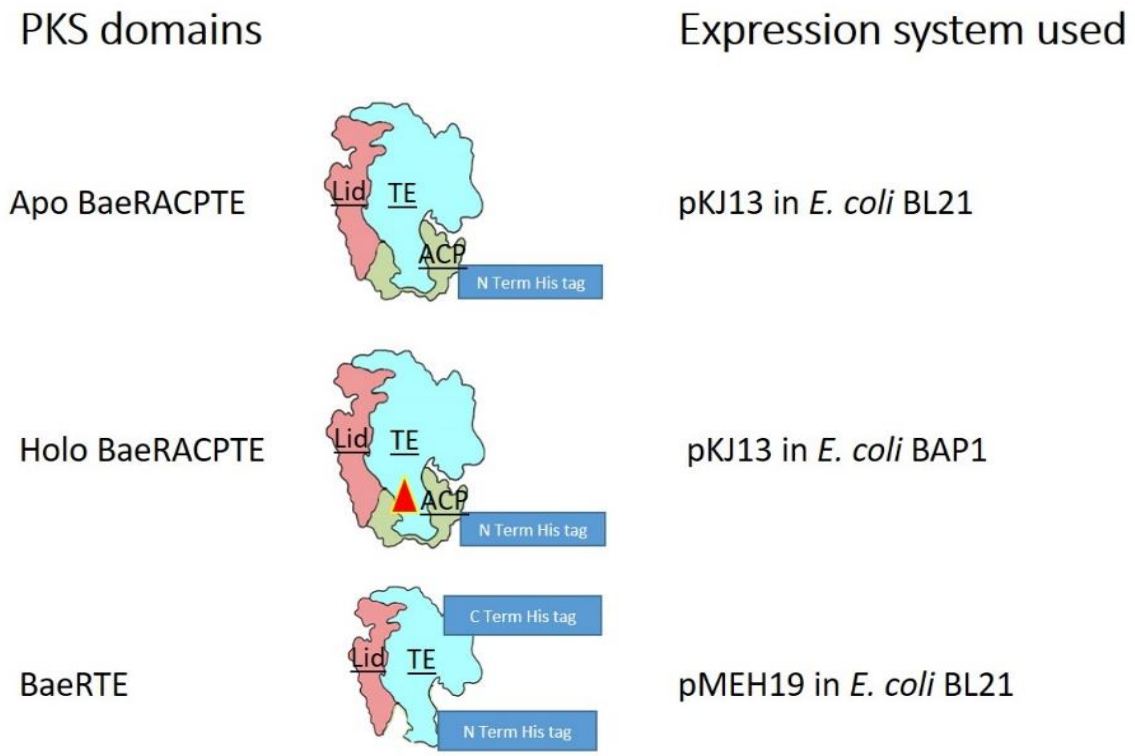


Figure 3-4 Structural summary of principal bacillaene thioesterase constructs. The pKJ13-derived ACP-TE didomain contains a C-terminal 6xHis tag. When pJK13 is expressed in *E. coli* BAP1, the PPTase sfp adds a Ppant on the ACP active-site Ser. The monodomain TE coded in pMEH19 possesses N- and C-terminal 6xHis tags. Full sequence alignment shown in Appendix **Figure III-1**.

This construct represented the simplest enzyme catalyst that was expected (based on modular *cis*-AT TE precedence) to catalyze thioester cleavage. Sequencing confirmed the pET28-based clone, and *E. coli* BL21 (DE3) transformants (NEB) harbouring the resulting plasmid (pMEH19) overexpressed a 41.3 kDa protein matching BaeTE's ORF. Additional details are included in Section **3.6.3**. The four acyl SNAC compounds (**1-3**) used in the initial screen (*trans*-2-methyl-2-pentenoyl-SNAC, ± methyl valeric-SNAC, and *cis*-2-methyl-3-penetnoyl-SNAC) were synthesized through standard acid-activation coupling (described in **3.4.3, 3.4.4,** and **3.4.6**) and prepared as DMSO stocks. The preparation of both components of the assay enabled screening of initial conditions. BaeTE was the first protein to be assayed for reactivity towards SNAC substrates.

The unconjugated SNAC substrate (**2**, **Figure 3-3**) was proposed to most closely resemble the bacillaene intermediate that is off-loaded by the TE, and was used for initial studies to adapt a common colorimetric TE activity assay to kinetic characterization of the BaeTE system.[29] For this assay, the TE construct was incubated with the substrate in the presence of 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, Ellman's reagent) and absorbance at 412 nm was measured every 3 s.

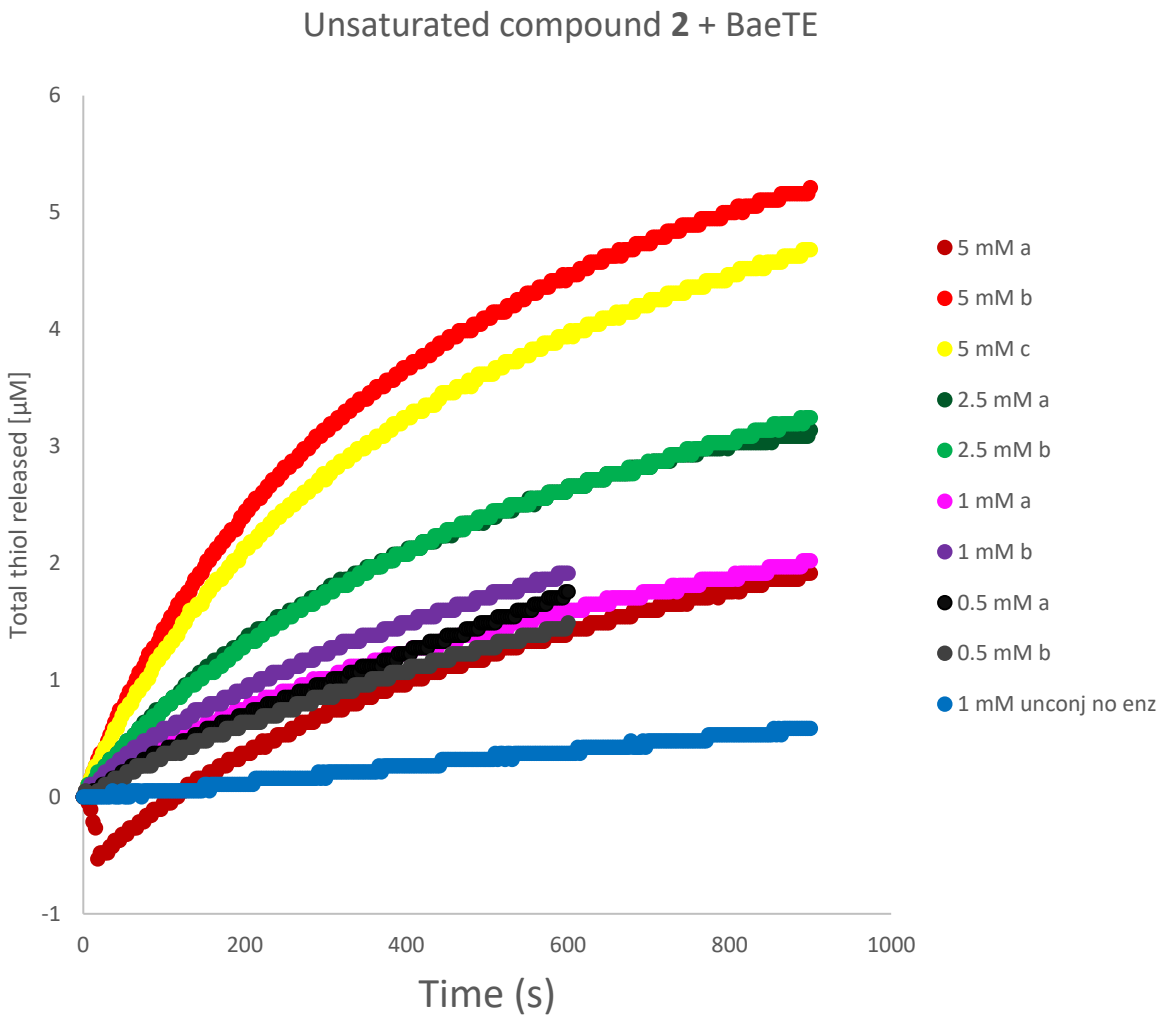


Figure 3-5 Total signal reaction time-course of BaeTE reacted with unconjugated SNAC. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4,(pH 7.42), 10% DMSO (v/v), the concentration of substrate listed in the Figure legend, and 1 µM BaeTE. ΔA_{412} was recorded on an Evolution 300 UV-Vis Spec (ThermoFisher), and converted to thiol produced using a standard thiol curve (**Figure 3-13**). A no enzyme added blank (light blue) estimates the basal hydrolysis of the unconjugated thioester in the assay buffer. "Staircase" curves in the data result from device rounding to the nearest certain digit.

The initial absorbance for the samples in this set ranged from 0.211-0.263. Change in thiol concentration over time was tracked by subtracting the initial absorbance from each raw input and multiplying by the conversion factor from A_{412} to μM thiol released (see **Figure 3-13**).

An initial scout run at 2.5 mM indicated that the assay proceeded in a biphasic non-linear fashion over time. As shown above, the rate of hydrolysis declined initially and reached rates equivalent to spontaneous hydrolysis of substrate in assay buffer after 500 s. These results suggested that frequent sampling would be important and that reaction periods of greater than an hour would produce average rates of reaction matching the rate of background hydrolysis. Subsequent samples were measured every 3 s for 600 s.

The background rate of hydrolysis for the sample containing 1 μM BaeTE without SNAC substrate displayed a background rate of thiol accumulation of 2 nM s^{-1} (**Figure 3.14**). In parallel, samples containing the unconjugated substrate **2** without TE exhibited approximately 1 nM s^{-1} accumulation (**Figure 3.15**), and hydrolysis was greater at higher concentrations of substrate. The increase in hydrolysis was roughly first-order and increased by $1.1 \pm 0.2 \times 10^{-5}$ $M_{substrate}^{-1}$ M_{Thiol} min^{-1}. This exceeds the literature precedent background hydrolytic rate of $4.5 \pm 0.9 \times 10^{-6}$ $M_{substrate}^{-1}$ M_{Thiol} min^{-1} reported for a racemic sample of methyl valeric-SNAC thioester.[29]

The background indicator signals with enzyme or unconjugated substrate alone were subtracted from the total thiol released values as an estimate of the thiol released as a result of the TE's action on the SNAC substrates.

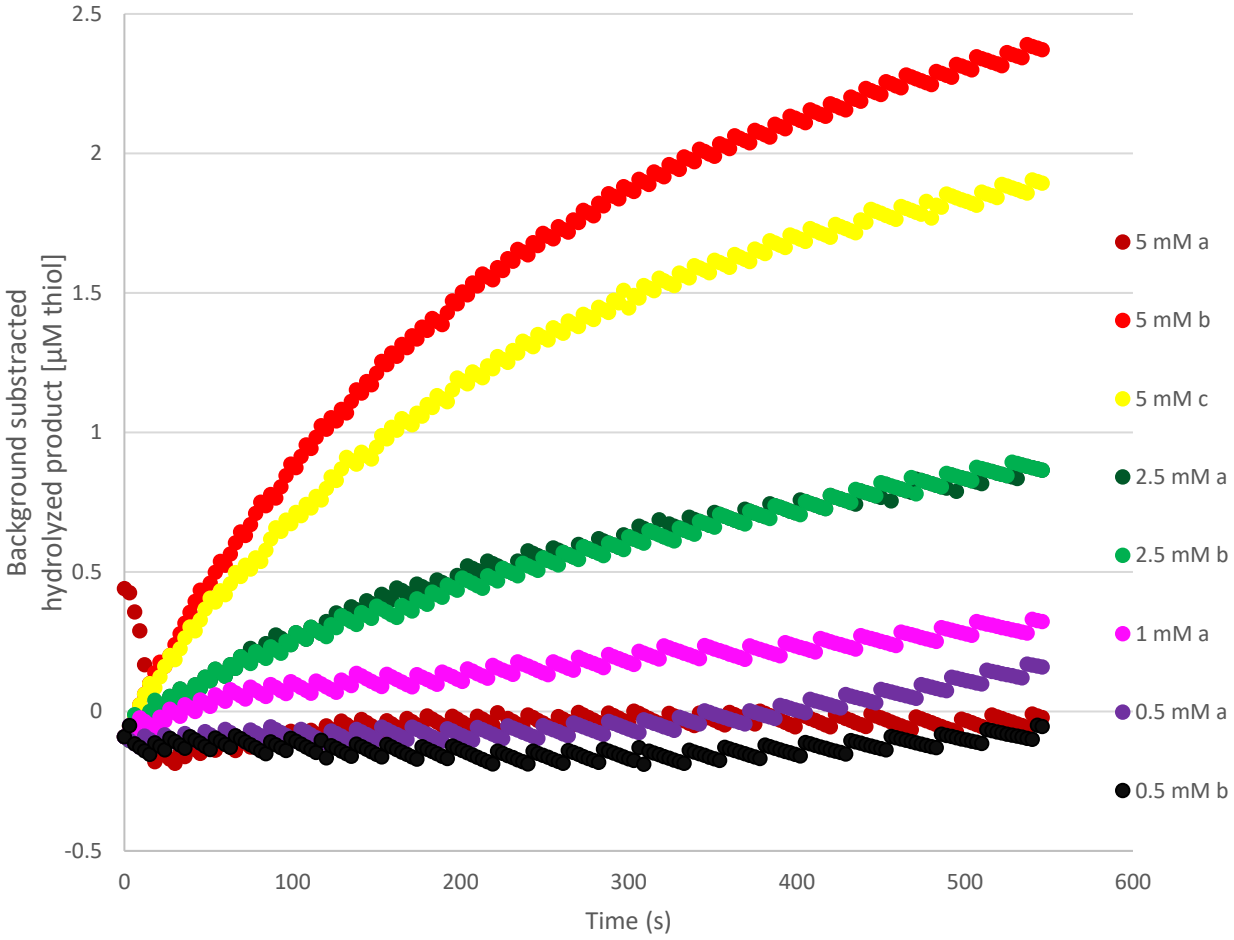


Figure 3-6 Background-subtracted reaction time-course of BaeTE reacted with unconjugated SNAC 2. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4 (pH 7.42), 10% DMSO (v/v), the concentration of substrate listed in the Figure legend, and 1 µM BaeTE. Sample (5 mM a) was adjusted to calculate changes in absorbance relative to the raw absorbance at 12 s to account for the discontinuity. "Sawtooth" curves in the data result from subtracting substrate and protein blank rates of hydrolysis (see **Figure 3-14** and **Figure 3-15**) from a rounded observed absorbance.

The background subtracted rate of thiol production for the unconjugated substrate illustrate 1.5 turnovers with respect to the TE concentration for the reactions with 5 mM of substrate within the first 180 s, whereas 1 mM reactions slowed before 1 equivalent of reaction was observed. The sawtooth peaks in the final curves arises from rounding of the absorbance output by the spectrophotometer, and the subtraction of continuous background hydrolysis as estimated as a continuous function of time. The local range of absorbances represents a measure of spectrophotometer accuracy, and the centre of the range represents the behaviour of the reaction.

Similar assays were conducted with BaeTE combined with the saturated and conjugated substrate equivalents and are shown in **Figure 3.7** and **Figure 3.8**.

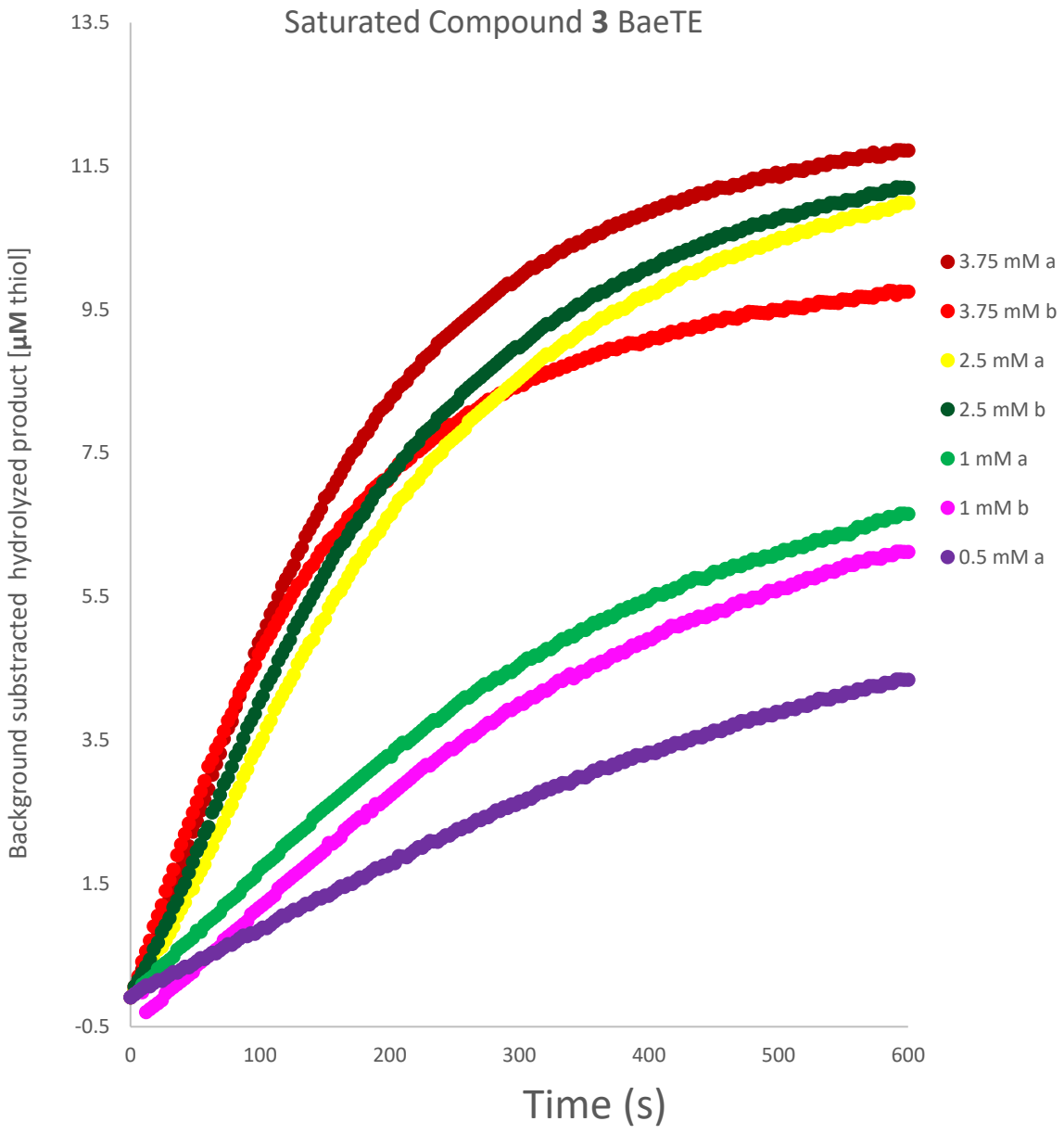


Figure 3-7 Background-subtracted hydrolysis reaction time-course of BaeTE reacted with saturated SNAC 3. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4, (pH 7.42), 10% DMSO (v/v), the concentration of substrate listed in the Figure legend, and 1 µM BaeTE. Absorbance values for 3.75 and 2.5 mM saturated substrate overlapped within the duplicates.

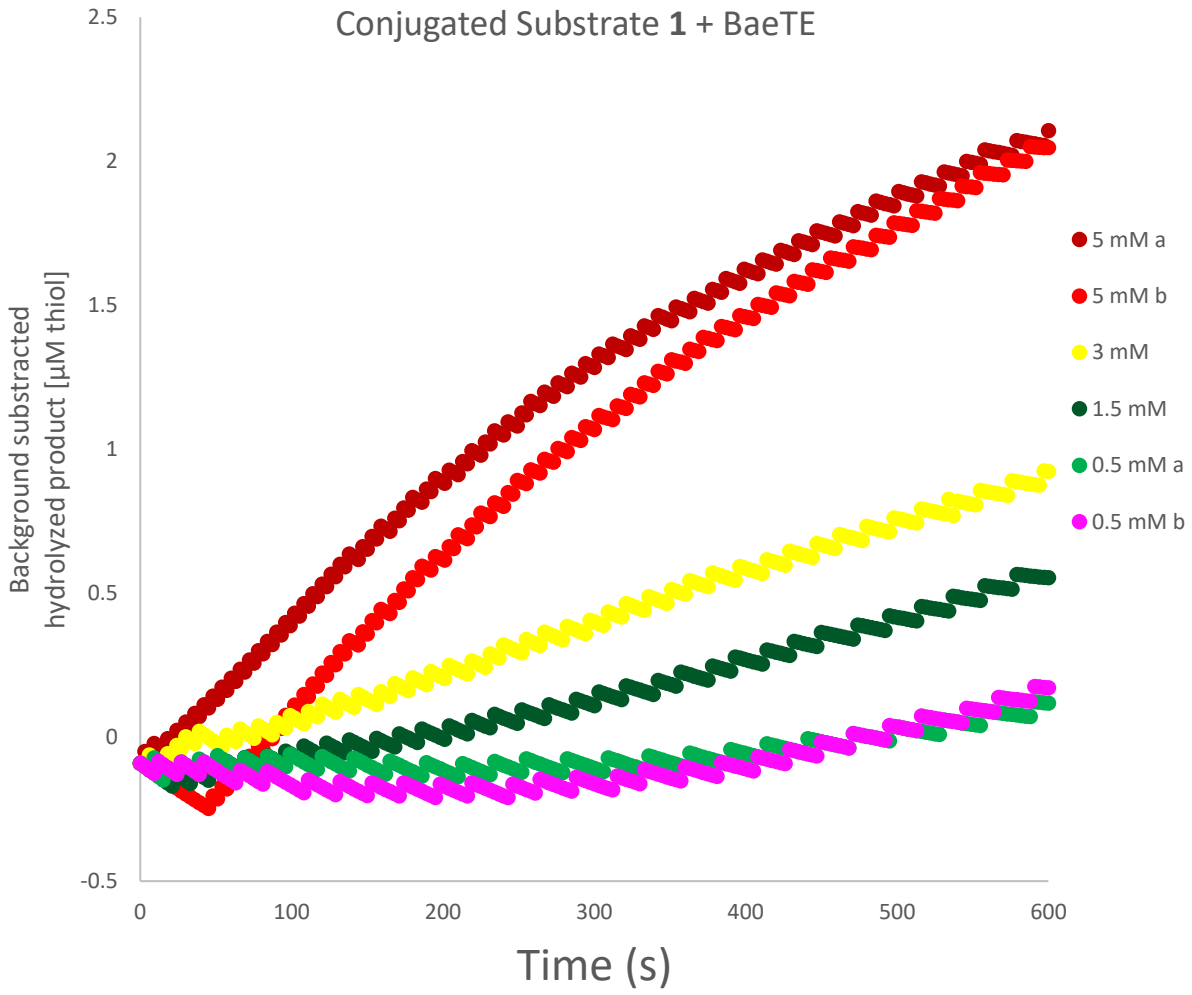


Figure 3-8 Background-subtracted reaction time-course of BaeTE reacted with conjugated SNAC 1. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4 (pH 7.42), 10% DMSO (v/v), the concentration of substrate listed in the Figure legend, and 1 µM BaeTE. The initial rate of hydrolysis for 0.5 mM conjugated substrate matched that of background for the first 250 s and increased after that.

In the assays conducted with saturated SNAC compounds and BaeTE, up to 8 turnovers occurred above background noise, and the reaction rate is described roughly by two linear trends with a transition around 200 s. The reactions with the conjugated substrate and the BaeTE proceeded at rates equivalent to the unconjugated reaction set (~2.1 µM hydrolysed after 600 s as compared with ~2.5 µM for the unconjugated compound).

The reaction conditions from **Figures 3-7-3-9** suggested that quantitative Michaelis-Menten type rate constants could not describe the kinetic behaviour. Additionally, time resolved rate information was considered critical for matching preliminary experiments for the BaeR ACP-TE didomain (BaeACPTE). Overall, initial conditions for substrates with the *holo*-BaeACPTE construct should include 0-5 mM substrate, 1 μM enzyme, and be observed for 600 s in order to capture the full variation of activity over time.

Initial BaeACPTE acyl-SNAC hydrolysis assays

An alternate, larger gene construct of the bacillaene TE, BaeACPTE (**Figure 3-4**), was also constructed and was tested in parallel to the BaeTE monodomain to determine if the presence of the cognate ACP altered the time-dependent loss in activity. The post-translationally phosphopantetheinylated *holo*-BaeACPTE was overexpressed and purified under the same conditions as the initial BaeTE experiments. This construct was then tested in the colorometric assay for thiol production from the three pentane-length acyl-SNAC analogues of bacillaene. **Figures 3-9 - 3-11** summarize the *holo*BaeACPTE results.

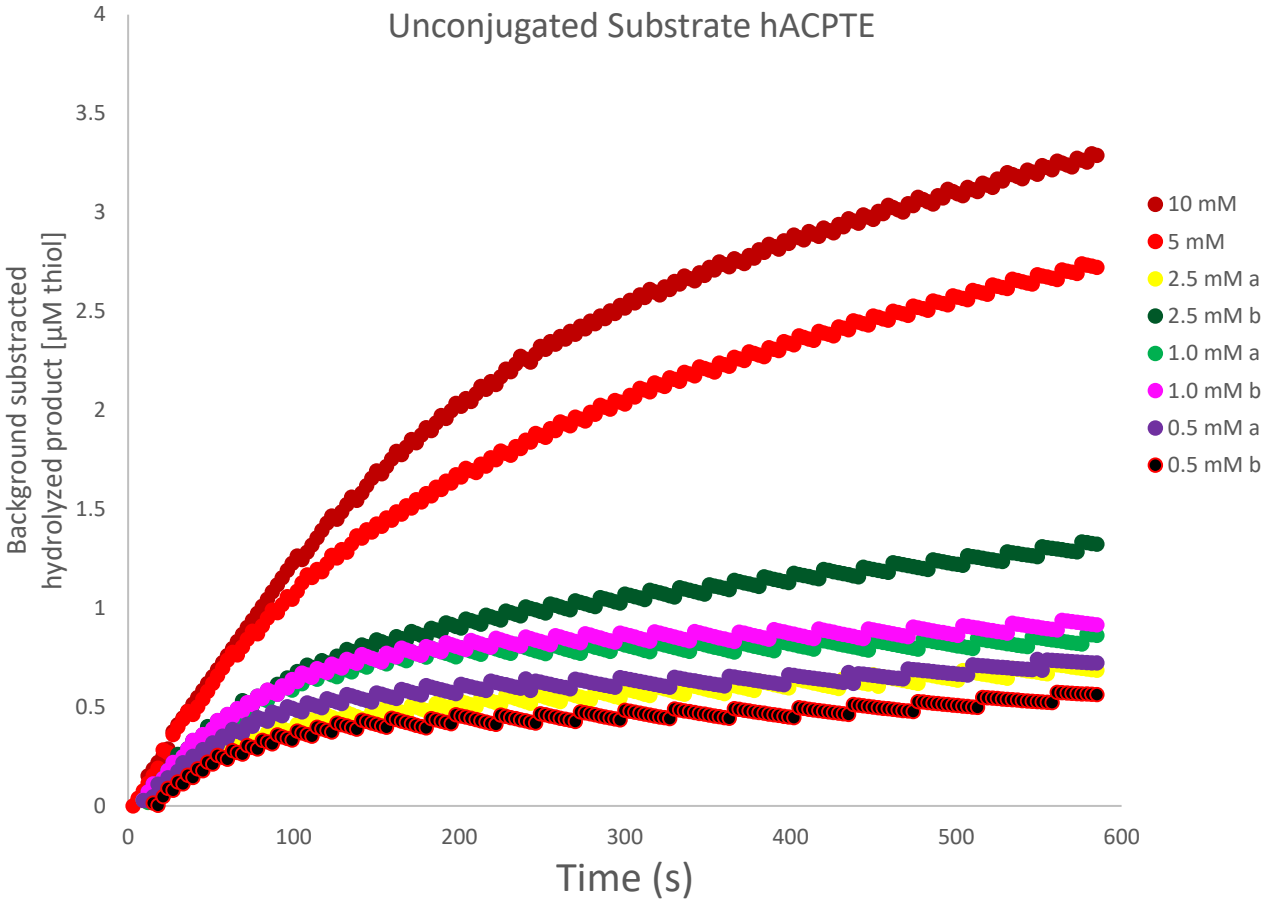


Figure 3-9 Background-subtracted reaction time-course of *holo*-BaeRACPTE with unconjugated SNAC. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4 (pH 7.42), 10% DMSO (v/v), the concentration of substrate listed in the Figure legend, and 1 µM *holo*-BaeACPTE.

The unconjugated substrate assays with the *holo* BaeACPTE (**Figure 3-9**) showed reactivity equivalent to the unconjugated BaeTE assay (**Figure 3-6**) at 5 mM substrate, but the *holo*-BaeACPTE constructs showed rapid initial increase in thiol detection equal to roughly 1 equivalent as compared to the TE concentration; after this (approximately 110 s into each reaction) the rate of reaction dropped to approximately 0.02 µM/min above the background at lower substrate concentrations, as apposed to the slow rate for unconjugated BaeTE.

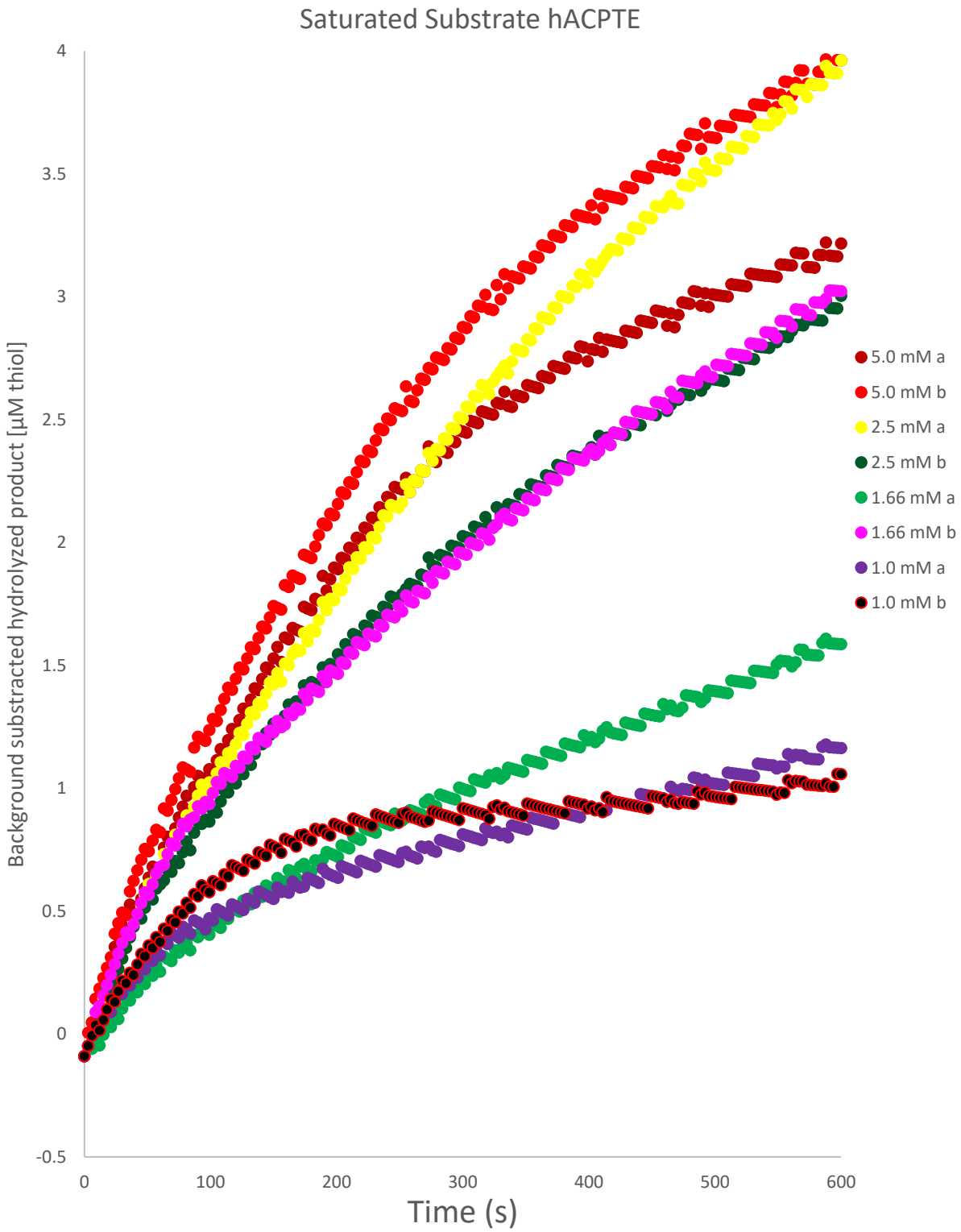


Figure 3.10 Background-subtracted reaction time-course of holoBaeRACPTE reacted with saturated SNAC. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4 (pH 7.42), 10% DMSO (v/v), the concentration of substrate listed in the Figure legend, and 1 µM *holo*-BaeACPTE. Note the large variance between duplicates in the 1.66 and 2.5 mM duplicates.

The reactions with the saturated SNAC compound with the *holo*-BaeACPTE displayed half the initial rate of reactivity of the equivalent saturated BaeTE assays (**Figure 3-7**) and an equivalent rate as compared with unconjugated BaeACPTE. Trial a at 2.5 mM displayed atypical retention of activity beyond 200 s, whereas the majority of the samples displayed at least 66% lower hydrolysis rate after 200 s.

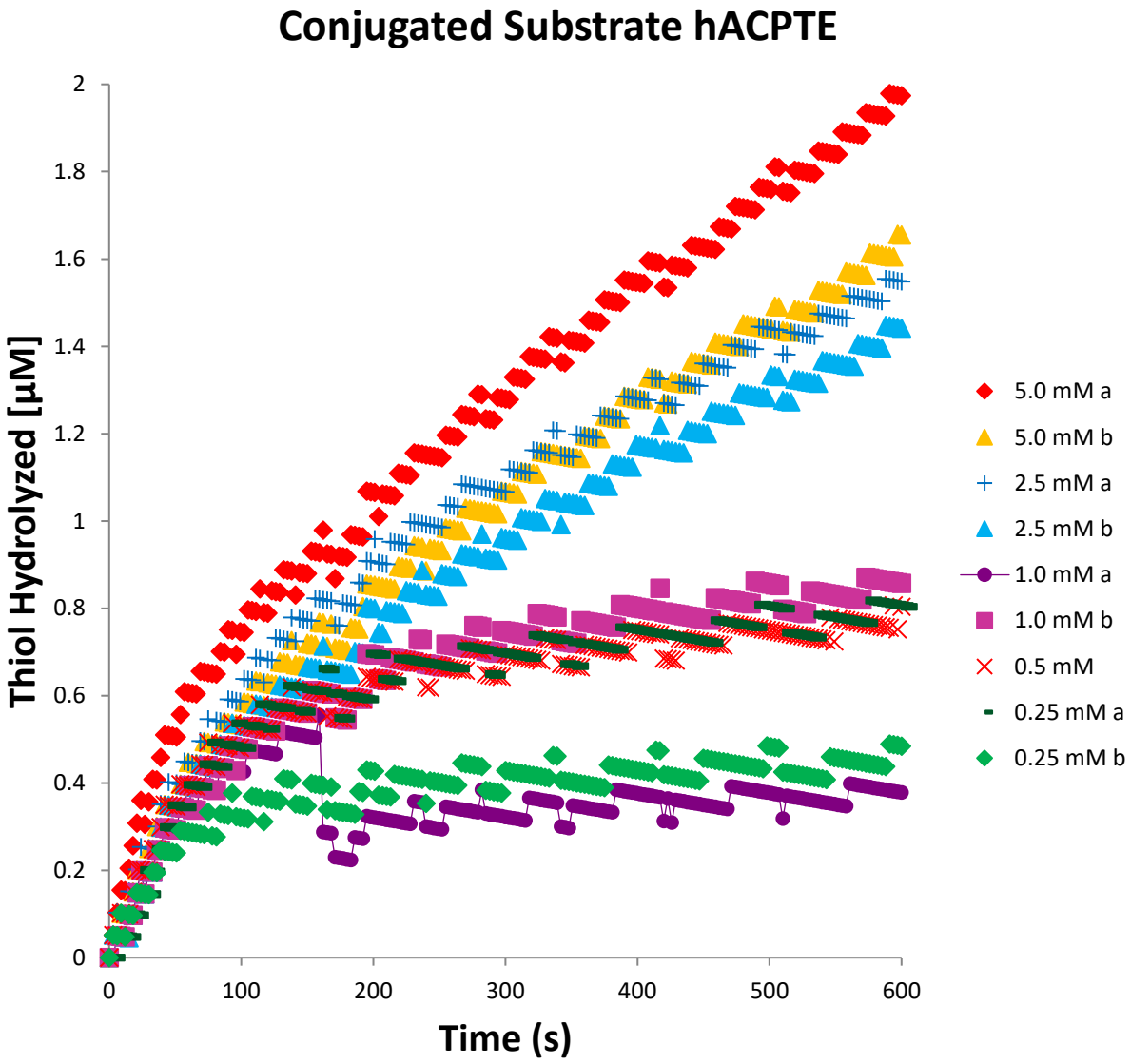


Figure 3.11 Background-subtracted reaction time-course of *holo*BaeRACPTE reacted with conjugated SNAC. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4 (pH 7.42), 10% DMSO (v/v), the concentration of substrate listed in the Figure legend, and 1 µM BaeTE. The recorded drop in absorbance in 1 mM sample a (purple) is highlighted with a line joining subsequent points.

The reaction between the conjugated SNAC and *holo*-BaeACPTE produced amounts of thiol equivalent to those observed in the conjugated SNAC reaction with BaeTE, but as was seen with the unconjugated SNAC *holo*-BaeACPTE reaction, the ACPTE construct generated more thiol in

the first 180 s and after 0.6 enzyme equivalents had been produced, the rate dropped more rapidly than in the BaeTE equivalents.

Overall, the *holo*-BaeACPTE assays demonstrated that the decline in assay activity was not specific to the BaeTE construct. Secondly, the *holo*-BaeACPTE processed two of the three compounds at the same initial rate, while the *holo*-BaeACPTE construct hydrolyzed the saturated substrate approximately 3 times less rapidly over the first 200 s. Finally, it is clear additional analysis is required to characterize the specific enzyme activity in each assay condition such that the rate change is described.

Examination of perceived bacillaene TE inhibition

While engineered overexpression and *in vitro* assays of engineered standalone Type I TEs has frequently produced catalytic activity on a range of substrates, several reports outline means of TE inactivation. These include near-ambient temperature protein-fold melting, substrate inhibition, and activity dependence on other assay components.

For example, one reported DEBS TE construct bore a T_M of 305.5 K,[30] and care was advised during the expression, purification, and assay of the protein around room temperature. Pikromycin TE[31] and Radicicol TE (Rdc TE)[32] both lose enzymatic activity in the presence of DTNB. DTNB is a disulfide indicator that cross-reacts with free thiols in solution and is frequently used to measure the hydrolysis of thioesters in TE assays.

Additionally, disorazole TE (disTE) and DEBS TE exhibited substrate inhibition when hydrolysing the unnatural substrate propionate para-nitrophenol ester.[33] Further reported examples of atypical substrate interactions include Rdc TE[34] and tautomycetin TE.[35] In the Rdc TE report, the enzyme was reacted with an SNAC thioester with one substrate alcohol stereocentre opposite the natural configuration. The Rdc TE exhibited inhibition suggestive of allosteric inhibition by a second molecule of SNAC substrate; conversely, in the tautomycetin report, the authors illustrated that hydrolysis of one SNAC substrate could equally be fit to a cooperative substrate binding model.

Finally, three relevant assay properties are dependent on the pH of the assay solution; namely, the catalytic activity of TEs,[36] the reactivity of DTNB towards SNAC thiolate, and the basal rate of hydrolysis of acyl-SNAC substrates are sensitive to the pH of the assay solution.

In general examples of enzymatic assays, ionic strength of an assay mixture and presence of accumulated products have affected the stability and continued activity of the enzyme *in vitro*.

In one attempted BaeTE protein purification, 10 mL of purified supernatant of cell lysate was loaded onto a HisTrap HP NiNTA column (GE LifeSciences, 1 mL) via an AKTA Purifier 10 (GE LifeSciences) at room temperature with solution A as 100 mM phosphate buffer 300 mM NaCl (pH 7.46), and solution B 100 mM phosphate buffer 300 mM NaCl and 200 mM imidazole (pH 7.48).

During the elution phase (100% B), speckled precipitate was observed in the collection tube. Protein-fold instability was suspected to give rise to both assay activity shifts and this newly observed precipitation. This prompted a follow-up protein expression and purification at pH 7.4, pH 8.0, and in 10% glycerol at pH 7.4 using hand-packed NiNTA using standard salt, buffer, and imidazole concentrations. However, each of these preparations were carried out without observable crash-out and each exhibited the same declining activity shown in 3.2.1 (data not shown).

3.2.3.1 Modified BaeTE constructs

As an effort to characterize the unusual non-linear time-dependent activity for both the TE and ACPTE constructs, the full *B. amyloliquefaciens* FZB42 BaeR C-terminal sequence was further analyzed in comparison to six other *B. amyloliquefaciens* strain sequences in search of unique mutations (Appendix **Figure III-2**), as well as in comparison to the most distantly related bacillaene-producing *Bacillus* species *B. subtilis* subsp. *subtilis* str. 168.[11, 37]

While there is broad homology amongst the *B. amyloliquefaciens* strains in comparison to *B. subtilis* 168, homologue *FZB42* and strain *plantarum* CAU B946[38] alone contain a His 63 AAs upstream from the C-terminus (AA position 289 for BaeTE construct in pMEH19). The other 5 strains and *B. subtilis* 168 all code for an Arg the deviation arises from a single base mutation in position two of the codon.

This mutation aligns with the surface-exposed loop region between β-sheet 7 and α-helix D of the structurally characterized aflatoxin TE (see **Figure 2-2**). In order to test the possibility that this surface-exposed non-conserved histidine reacts non-catalytically with thioesters in an inactivation side-reaction, the codon was altered to code for Arg in the point-mutant construct BaeRTE_H289R.

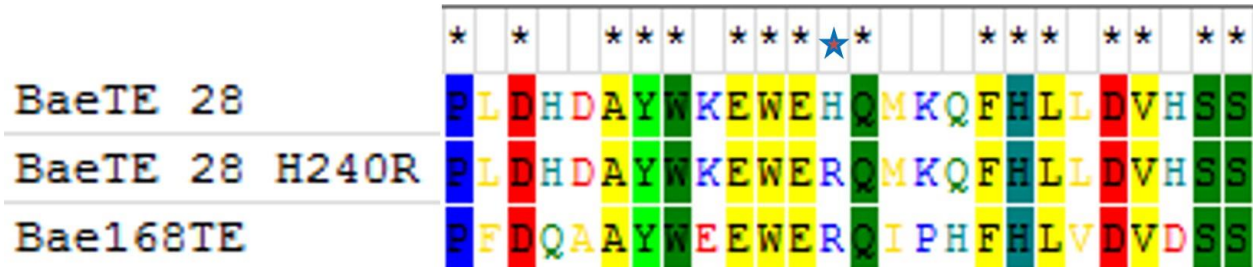


Figure 3-12 Flanking amino acid sequence of BaeTE constructs used to investigate H289R point mutation. The 280-302 region of the AA sequences for the BaeTEH289R point mutant (pMEH20) and the gene synthesis-derived BaeR168TE (pMEH46) are compared with the *FZB42*-derived BaeTE (pMEH19). Full sequence alignment shown in Appendix **Figure III-1**

In an alternate hypothesis, the assay inactivation may be common to all *amyloliquefaciens* isolates, or even common to the *subtilis* 168 TE. To test whether observed differences in the TE coding sequence affected *in vitro* hydrolysis activity, an *E. coli* codon-optimized version of the 168 bacillaene TE was synthesized as well. **Figure 3-12** outlines the 280-302 region of the AA sequences for the BaeTEH289R point mutant (pMEH20) and the gene synthesis-derived BaeR168TE (pMEH46).

Appendix **Figure III-2** shows the full TE domain alignment anchored on the structural elements of a characterized DEBS TE construct. Both the H289R and the 168 TE construct displayed decreasing activity over the first 200 s when reacted with the methyl-valeric-SNAC substrate at 5 mM (**Figure IV-1**).

Additionally, BaeTE (*FZB42*, pMEH19) was fused to an N-terminal maltose-binding protein (MBP) from pMalC5x[39, 40] to form MBP-BaeTE (pMEH48); this was designed to lower the net pI of the protein away from the buffer conditions and improve overall protein solubility. The MBP-BaeTE fusion protein displayed similar kinetics in the kinetic Ellman's assay with the conjugated substrate at 5 mM (**Figure IV-1**).

3.3 Conclusions

In vitro reactions between acyl-SNAC analogues of last-stage bacillaene biosynthetic intermediates and the product-releasing TE domain resulted in time-dependent thiol release. The reactions slowed after 2-10 enzyme equivalents of product was formed and at most 0.25% of substrate had been consumed. While turnover was greater with the saturated substrate, this phenomenon was common to all three substrates and to TE constructs with and without the upstream ACP domain. Natural variance in the TE gene sequence found in different bacillaene-PKS containing bacterial strains did not impact TE activity or assay resilience. **Chapter 4** describes modeling the decline in enzyme activity in an effort to compare activity between substrates quantitatively, and to interpret the mechanism of activity loss.

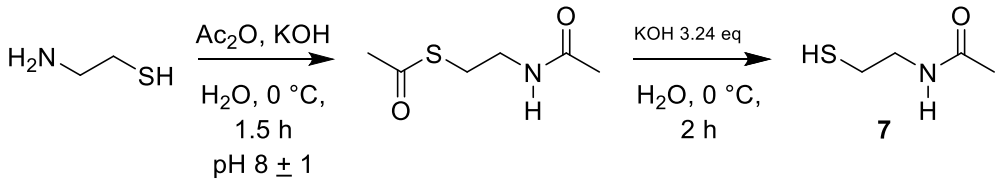
3.4 Compound synthesis

Reagents

Ethyl-2-methyl-3-pentenoate (high *cis*) was purchased from Bedoukian Research (CT, USA). Other synthetic reagents were purchased from Sigma-Aldrich.

Synthesis of *N*-acetylcysteamine (7)

N-acetylcysteamine was prepared according to Ashley *et al.* 2006.[41]



Cysteamine hydrochloride (15.0 g, 0.194 mol, 1 equiv) was added to a 1-L 3-neck round bottom flask fitted with a magnetic stir bar, 2 addition funnels, and a pH electrode. Water (90 mL) was added and the stirred solution was cooled on ice. The pH was adjusted to 8.0 by addition of 8 N KOH. Acetic anhydride (37.5 mL, 0.397 mol, 2.05 equiv) was placed in one addition funnel, and 8

N KOH was placed in the other addition funnel. The acetic anhydride was added dropwise to the cysteamine solution, with 8 N KOH being added so as to keep the reaction pH at 8 $\pm$ 1.

After addition of acetic anhydride, the pH was adjusted to 7.0 using 1 N HCl and the mixture was allowed to stir for 75 min on ice. Solid NaCl was added to saturation, and the solution was extracted with CH_2Cl_2 (4 * 400 mL). The organic extracts were combined, dried over $MgSO_4$, filtered, and concentrated under reduced pressure to yield 21.3 g (68 %) of a colourless oil, which crystallized upon standing at 4 °C.

$C_6H_{11}NO_2S$

^{1}H-NMR ($CDCl_3$, 400 MHz): δ 6.43 (br, 1H) 3.42 (q, *J* = 7 Hz, 2H) 3.03 (t, *J* = 7 Hz, 2H), 2.36 (s, 3H), 1.98 (s, 3H).

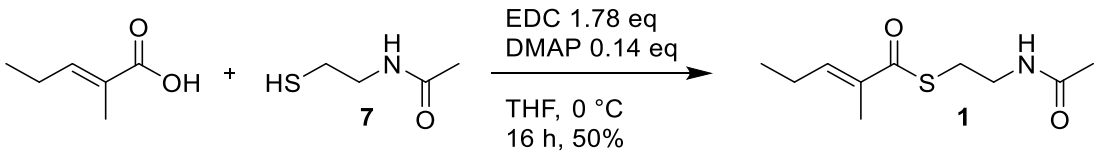
N,S-Diacetylcysteamine (21.3 g, 0.131 mol, 1 equiv) was placed in a 2-L round bottom flask fitted with a magnetic stirrer, and dissolved in 1.4 L of water. The flask was purged with N_2, and the mixture is chilled in an ice bath. KOH (24 g, 0.428 mol, 3.24 equiv) was added, and the mixture was stirred for 2 h on ice under inert atmosphere. The pH was adjusted to 7 using 6 N HCl, and solid NaCl was added to saturation. The mixture was extracted 7 times with 500 mL portions of CH_2Cl_2. The organic extracts were combined, dried over $MgSO_4$, filtered and concentrated under reduced pressure to yield 15.75 g of a colorless oil (89 % yield).

C_4H_9NOS

^{1}H-NMR ($CDCl_3$, 400 MHz): δ 7.03(br, 1H) 3.27-3.19 (m, 2H) 2.53-2.45 (m, 2H), 1.88-1.80 (m, 3H), 1.32 (tt, *J* =8.3, 1.6 Hz, 1H).

The NMR characterization matches the description by Sigma Aldrich.

Synthesis of *trans*-2-methyl-2-pentenoyl-SNAC (1)



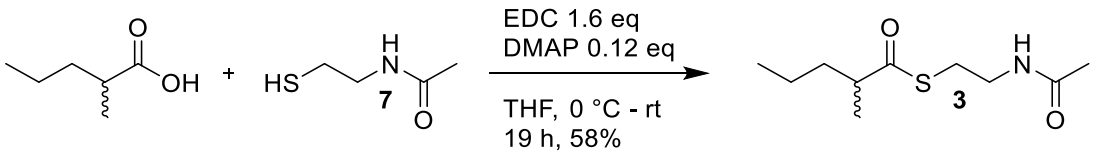
137.8 mg of EDC (0.72 mmol, 1.78 equiv) and 6.7 mg of DMAP (0.055 mmol, 0.14 equiv) were added to an oven-dried 10 mL round bottom flask with a magnetic stirrer. 1.5 mL of dry THF was added. A fraction of white solid remained out of solution. The flask was cooled to 0 °C. 47 μL of *trans*-2-methyl-2-pentenoic acid (0.40 mmol, 1 equiv) was added, and 47.5 μL of *N*-acetyl cysteamine (HSNAC, 0.45 mmol, 1.1 equiv) was added. The reaction was run for 16 h. Product formation was monitored via TLC analysis (product R_f= 0.33 in 50:50 acetone:hexane (v/v)). The crude reaction was diluted in 5 of mL ethyl acetate. The resulting organic fraction was extracted with 10 mL of saturated $NaHCO_3$ solution, and subsequently extracted with 10 mL of brine. The brine fraction was reextracted with 10 mL of ethyl acetate and the organic fractions were pooled and concentrated *in* vacuo. Finally organic concentrate was flash-chromatographically purified with a 50:50 acetone:hexane (v/v) solvent. The reaction yielded 43.8 mg of purified **1** for a yield of 50% (0.20 mmol).

$C_{10}H_{17}NO_2S$

^{1}H NMR (400 MHz, $CDCl_3$):): δ = 6.72 (td, *J* = 7.2, 1.3 Hz, 1H), 6.02 (br, 1H), 3.41 (q, *J* = 6.1 Hz, 2H), 3.01 (t, *J* = 6.3 Hz, 2H), 2.19 (qd, *J* = 7.5, 7.2 Hz, 2H), 1.94 (s, 3H), 1.83 (s, 3H), 1.06 (t, *J* = 7.5 Hz, 3H).

NMR characterization matched previous synthesis by Sharma and Boddy.[29]

Synthesis of methyl valeric SNAC (saturated substrate, 3)



113.5 mg of EDC (0.59 mmol, 1.6 equiv) and 5.3 mg of DMAP (0.044 mmol, .12 equiv) were added to an oven-dried 10 mL round bottom flask with a magnetic stirrer. 1.5 mL of dry THF was added. A fraction of white solid remained out of solution. The flask was cooled to 0 °C. 45 μL of methyl valeric acid (0.36 mmol, 1 equiv) was added, and 47.5 μL of SNAC (0.47 mmol, 1.3 equiv) were added. The reaction was run for 19 h. Product formation was tracked via spot formation below SNAC at an R_f of 0.3 in 50:50 acetone:hexane (v/v). The crude reaction was diluted in 5 mL of ethyl acetate, extracted with ethyl acetate and a saturated $NaHCO_3$ solution, extracted with brine, and finally flash-chromatographically purified with a 50:50 acetate hexane (v/v) solvent. The reaction yielded 45.4 mg (0.21 mmol) of purified **3** for a yield of 58%.

racemic $C_{10}H_{19}NO_2S$

^{1}H NMR (400 MHz:$CDCl_3$): δ = 5.91 (br, 1H), 3.40 (q, *J* = 6.1 Hz, 2H), 2.98 (t, *J* = 7.5 Hz, 2H), 2.64 (tq, *J* = 7.2 Hz, 1H), 1.93 (s, 3H), 1.70-1.59 (m, 1H), 1.41-1.19 (m, 3H), 1.13 (d, *J* = 6.9 Hz, 3H), 0.87 (t, *J* = 7.2 Hz, 3H)

^{13}C NMR (400 MHz, CDCl3): δ = 204.6, 170.5, 48.5, 40.6, 36.6, 28.2, 23.1, 20.4, 17.1, 13.7

Repeated synthesis gave a yield of 66%.

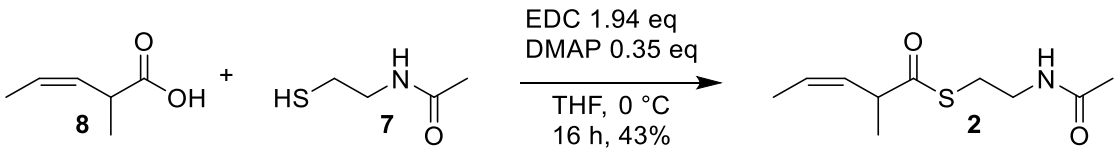
Synthesis of *cis*-2-methyl-3-pentenoic acid (8)

5.27 mL of a 2:1 (v/v) mixture of methanol and water was added to a 25 mL round-bottom flask with stir-bar. 1.6 mL of 1 M NaOH and 0.128 mL of ethyl-*cis*-3-2-methyl-pentenoate (0.8 mmol, 1 equiv) was added to the solution. The mixture was stirred at 700 RPM at room temperature open to the air. After 20 h, 5 mL of 1 M NaOH was added, and the solution was extracted with diethyl ether (3 × 20 mL). The aqueous fraction was acidified with 20 mL saturated NH_4Cl and 15 mL of 1 M HCl. The final pH was ≤ 1 via pH paper. The acidified fraction was extracted with diethyl ether (3

× 15 mL). The ether fractions were pooled, dried with $MgSO_4$, filtered through Whatmann paper and condensed via rotary evaporator to 1 mL.

Peaks at 5.39 and 5.57 ppm on a ^1H spectrum of the crude reaction product in $CDCl_3$ and the appearance of a new spot with a R_f of 0.7 on a TLC in 50:50 hexane:ethyl acetate indicated the presence of free acid.

Synthesis of cis-3-methyl-2-pentenoyl-SNAC (2)



469 mg of (EDC) and 50.2 mg of DMAP were added to an oven-dried 10 mL round bottom flask with a magnetic stirrer. 5 mL of dry THF was added. A fraction of white solid remained out of solution. The flask was cooled to 0 °C. 45.0 mg of *cis*-3-methyl-2-pentenoic acid was added, and 127.6 µL of SNAC was added. The reaction was run for 16 h. Product formation was tracked via spot formation below SNAC at an Rf of 0.25 in 40:60 acetone:hexane. The crude reaction was diluted in 10 mL of ethyl acetate, extracted with ethyl acetate and a saturated $NaHCO_3$ solution, and finally flash-chromatographically purified with a 50:50 acetate hexane (v/v) solvent. The reaction yielded 118.7 mg of purified **2** for a yield of 43%.

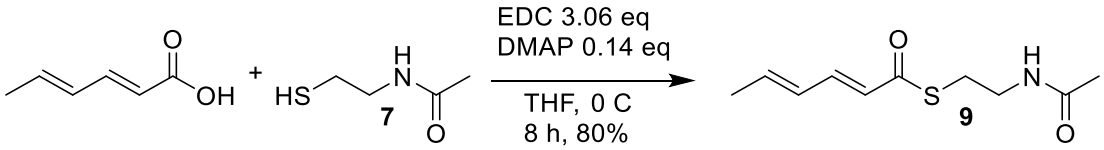
$C_{10}H_{17}NO_2S$

^1H NMR (400 MHz:d_6-DMSO): δ = 7.99 (t, *J* = 5.1 Hz 1H), 5.59 (dqd, *J* = 10.6, 6.9, 1.0 Hz, 1H), 5.29 (ddq, *J* = 10.1, 9.4, 1.8 Hz, 1H), 3.61 (dqd, *J* = 9.4, 7.0, 0.9 Hz, 1H), 3.10 (q, *J* = 6.4 Hz, 2H), 2.84 (td, *J* = 6.7, 1.5 Hz, 2H), 1.75 (s, 3H), 1.63 (dd, *J* = 6.9, 1.8 Hz, 3H), 1.12 (d, *J* = 6.7 Hz, 3H)

^1H NMR (400 MHz:$CDCl_3$): δ= 5.83 (br, 1H), 5.63 (dqd, *J* = 10.7, 6.9, 1.3 Hz, 1H), 5.35 (tq, *J* = 10.1, 1.7 Hz, 1H), 3.61 (dqd, *J* = 9.4, 7.0, 0.9 Hz, 1H), 3.43 (q, *J* = 6.1 Hz, 2H), 2.98 (t, *J* = 6.3 Hz, 2H), 1.93 (s, 3H), 1.68 (dd, 7.0, 1.7, 3H), 1.23 (d, *J* = 7.2 Hz, 3H)

Product dissolved at 50 mM in d_6-DMSO was shown by NMR to be stable to hydrolysis over 4 months at -20 °C.

Synthesis of (2*E*,4*E*)-Hexadienoyl-*N*-Acetylcysteamine (9)



235.1 mg of EDC (1.23 mmol, 3.06 equiv) and 7 mg of DMAP (0.06 mmol, 0.14 equiv) were added to an oven-dried 10 mL round bottom flask with a magnetic stirrer. 1.5 mL of dry THF was added. A fraction of white solid remained out of solution. The flask was cooled to 0 °C. 45.0 mg of sorbic acid (0.401 mmol, 1 equiv) was added, and after 10 min 51.8 μL of SNAC (0.487 mmol, 1.21 equiv) was added. The reaction was run for 8 h. Product formation was tracked via spot formation below SNAC at an R_f of 0.3 in 40:60 acetone:hexane. Crude reaction was diluted in 5 mL of ethyl acetate, extracted with ethyl acetate and a saturated Na_2CO_3 solution, and finally flash-chromatographically purified with a 40:60 acetate hexane (v/v) solvent. The reaction yielded 68.7 mg of purified product for a yield of 80%.

$C_{10}H_{15}NO_2S$

^{1}H NMR (400 MHz;CDCl3 + 0.05% TMS): δ = 7.15 (dd, *J* = 15.2, 10.5 Hz, 1H), 6.20 (dq, *J* = 15.0, 6.6 Hz, 1H), 6.15 (br, 1 H), 6.11 (ddq, *J* = 15.2, 10.5, 0.5 Hz, 1H), 6.03 (dd, *J* = 15.1, 0.4 Hz, 1H), 3.47 (dd, *J* = 6.2, 6.0 Hz, 2H), 3.06 (t, *J* = 6.2 Hz, 2H), 1.91 (s, 3H), 1.83 (dd, *J* = 6.4, 0.4 Hz, 3H);

^{13}C NMR (400 MHz, $CDCl_3$): δ = 190.4, 170.4, 141.9, 141.8, 129.5, 125.7, 39.8, 28.3, 23.2, 18.9;

NMR characterization matched previous synthesis by Ames *et al.*[42]

3.5 Cloning

General plasmid construction

Chemically competent *Escherichia coli* XL1 Blue (Stratagene) were used for plasmid stock-up and sticky-end ligations. Chemically competent *E. coli* TOP10 (Invitrogen) cells were used for blunt end ligations. T4 ligase (NEB) was used for all enzymatic ligations. All restriction endonucleases were purchased from NEB. Pfu Ultra (Agilent) was used for PCR reactions. DNA primers were purchased from Integrated DNA Technologies (IDT), and synthetic gene synthesis was performed by IDT and Life Technologies (Thermo Fisher). DNA isolation was carried out via isopropanol precipitation or using Omega Bio Tek DNA preparation kits (VWR).

DNA sequencing for plasmids was performed by Stemcore Laboratories (Ottawa Hospital Research Institute, Ottawa) for pMEH01-pMEH28, by Eurofins DNA Sequencing (Kentucky) for pMEH29-pMEH42, and by the Genome Quebec Innovation Centre (Montréal, Québec) for pMEH 43-pMEH50.

All plasmids have a plasmid map included in **Appendix II**. Their cloning and contents are summarized in Table II 1.

pKJ13, containing the C-terminal ACP-TE didomain of BaeR of the bacillaene pathway from *Bacillus amyloliquefaciens* FZB42,[11] was provided by Katja Jensen from the lab group of Joern Piel. The ACP and TE domains were amplified off this plasmid using the primer pairs BaciACPf (GATC CATATG CCTCAGCTCTGCCGCG) BaciACPr (GATC GAATTC TTAGCCCGTCTCCTGT), and BacTE_*NheI*_f (GATC GCTAGC CGGAATTTTCCCGAAC) BacTEr_no_P (GATC GAATTC CGTTTTACCAACT), respectively.

PCR reactions were performed in a Mastercycler Personal (Eppendorf) with 0.2 mM DNTP and 2.5 U of Pfu Ultra unless specified otherwise. The general “touchdown” thermocycler method followed 30 s denaturing at 98 °C, 30 s 68 °C annealing, and 72 °C elongation for 15 cycles, with the annealing temperature lowered 1 °C per cycle. In the second stage, the annealing temperature was held at 53 °C for 15 cycles. The runs were finished with 10 minutes of elongation at 72 °C.

Cloning of terminal bacillaene ACP and TE (pMEH08 and pMEH19)

Ligation of the PCR products for the ACP and the TE into pCR-Blunt Topo II (Thermo Fisher) produced pMEH01 and pMEH18 respectively. Restriction digest of pMEH01 and pET28 with *NdeI* and *EcoRI*, followed by ligation produced the ACP expression plasmid pMEH08. Restriction digest of pMEH18 and pET28 with *NheI* and *EcoRI*, followed by ligation produced the TE expression plasmid pMEH19.

QuikChange II Point Mutants

For all point mutations, the standard QuikChange II protocol was followed: 125 ng of each overlap primer was combined with 10 ng of plasmid template, 6 μg dNTP (final conc. 0.25 mM), and 2.5 U of PfuUltra DNA Polymerase for a final volume of 50 μL in reaction buffer (Agilent). The PCR method followed

30 s denaturing at 95 °C

60 s 55 °C annealing

60 s $kb^{-1}_{plasmid}$ 68 °C elongation

12 cycles

Following the PCR, samples were transferred to ice for 2 min before addition of temperature-sensitive *DpnI*. 10 U of *DpnI* (1 μL) was added to half of the solution, and the tubes were incubated at 37 °C for 10 min. PCR product was nonviable at longer incubation time-points. Finally, 1 μL of the mixture was combined with chemically competent *E. coli* XL1blue and transformed via heat-shock.

3.5.3.1 Point Mutation of BaeTE H289R

The CAC codon was mutated at position two to introduce the Arg CGC codon.

BaeRTE H289R was generated using the QuikChange II (Agilent)-type primers BaeTE H240R f (ATTGGAAAGAATGGGAACGCCAGATG) and BaeTE H240 r (AGGTGAAACTGTTTCATCTGGCGTTCC).

3.5.3.2 Active-Site Point Mutation of BaeTE S88A

The TCC codon was mutated at position one to introduce the Ala GCC codon.

BaeRTE S88A was generated using the QuikChange II (Agil ent)-type primers BaeRTE S88A f (CCTTATGATCTCGGCGGATATGCCTTGGGCGGAATG) and BaeRTE S88A r (CATATGCAAGCATTCCGCCCAAGGCATATCCGCCGAGATC)

Gene synthesis of *Bacillus subtilis* 168 BaeRTE

The gene location of the BaeRTE analogue in *Bacillus subtilis* 168 was retrieved via the annotation listed in the database Clustermine360.[43] The region spanning 1850890-1858521 of Genbank accession gi|225184640 was verified through alignment with pMEH19, and the gene sequence was codon-optimized for *E. coli* expression with flanking *NdeI* and *EcoRI* sites. Ligation of the synthetic gene fragment from IDT into pCR-Blunt Topo II (Thermo Fisher) produced pMEH45.

Restriction digest of pMEH45 and pET21 with *NdeI* and *EcoRI*, followed by ligation produced the TE expression plasmid pMEH46.

3.6 Protein Expression

All His_6-Tagged proteins described below are purified with the following general protocol. At the end of the protein expression period, the cell culture was centrifuged at 5000 g for 15 minutes at 4 °C and the supernatant was decanted. The cell pellet was resuspended in 10 mL of lysis buffer (100 mM sodium phosphate, 300 mM NaCl, 10% (v/v) glycerol, 1 mg/mLlysozyme, 1 μg/mL pepstatin A, 1 μg/mL leupeptin, pH 8.0). The mixture was sonicated (30*1 s on 1 s off, 10%, in 2 min periods). The lysed cell extract was centrifuged at 9000 g for 60 min, and the resulting supernatant was incubated with 0.5 mL of Superflow NiNTA resin (Qiagen) for 1 h at 4 °C with gentle agitation. The resin slurry was loaded into a column and rinsed with Elution buffer (100 mM Tris, 300 mM NaCl, X mM imidiazole (X = 0, 20, 100, 250), pH 7.4) in the following wash order: (1 * 5 mL 0 mM, 1 * 5 mL 20 mM, 2 * 2.5 mL 100 mM, 2 * 250 mM). The protein target eluted in

the 100 mM fractions, which were pooled, buffer exchanged into 50 mM KH_2PO_4, 150 mM NaCl (pH 7.42), concentrated to 0.5 mL, and combined with buffered glycerol for a final 1 mL solution of 50 mM KH_2PO_4, 150 mM NaCl, 30% glycerol (v/v) with less than 0.5 mM imidazole.

Expression of *apo*-ACPTE (pKJ13)

pKJ13 was transformed into *E. coli* BL21 and plated on LB + Kan. A colony was grown in 4 mL of LB + Kan liquid medium overnight at 37 °C. 2 mL of the overnight seed was used as inoculum for 400 mL LB + Kan in a preheated 1 L Erlenmeyer flask. After 3.5 h the OD_{600} reached 0.63. The flask was chilled on ice for 10 minutes and IPTG was added to a final concentration of 0.5 mM. The flask was incubated at at 16 °C for 19 h.

The culture yielded 61.7 mg/L of over-expressed protein after immobilized metal affinity column chromatography. The final protein solution was concentrated to 1 mL (514 uM) in 50 mM KH_2PO_4, 150 mM NaCl, 30% glycerol (v/v) with less than 0.5 mM imidazole.

Expression of *holo*-ACPTE (pKJ13)

pKJ13 was transformed into *E. coli* BAP1[44] and plated on LB + Kan. A colony was grown in 4 mL of LB + Kan liquid medium overnight at 37 °C. 2 mL of the overnight seed was used as inoculum for 400 mL LB + Kan in a preheated 1 L Erlenmeyer flask. After 3.0 h the OD_{600} reached 0.71. The flask was chilled on ice for 10 minutes and IPTG was added to a final concentration of 0.5 mM. The flask was incubated at at 16 °C for 18 h.

The culture yielded 39.7 mg/L of over-expressed protein after immobilized metal affinity column chromatography. The final protein solution was concentrated to 1 mL (333 uM) in 50 mM Tris, 150 mM NaCl, 30% glycerol (v/v) with less than 0.5 mM imidazole.

Expression of BaeRTE (pMEH19)

E. coli BL21 failed to produce more than 0.4 mg/L of overexpressed BaeRTE during expression runs at 30 °C, and the protein was observed to precipitate from solution during purification (elution from FPLC, during centrifugation over dialysis membrane). Low temperature cultivation was selected for stabilized protein expression.

pMEH19 was transformed into *E. coli* BL21 and plated on LB + Kan. A colony was grown in 4 mL of LB + Kan liquid medium overnight at 37 °C. 2 mL of the overnight seed was used as inoculum for 400 mL LB + Kan in a preheated 1 L Erlenmeyer flask. After 4 h the OD_{600} reached 0.55 and IPTG was added to a final concentration of 0.5 mM. The flask was incubated at at 16 °C for 21 h. 2 mL of 7 μM protein was produced in 50 mM KH_2PO_4, 150 mM NaCl, 30% glycerol (v/v) with less than 0.5 mM imidazole.

Expression of ElanTE/MBP (pMEH42)

pMEH42 was transformed into *E. coli* BL21 and plated on LB + Amp. A colony was grown in 4 mL of LB, Amp, and 0.2 % (m/v) glucose (added post autoclaving) liquid medium overnight at 37 °C. 2 mL of the overnight seed was used as inoculum for 400 mL LB, Amp, and 0.2% glucose (m/v) in a preheated 1 L Erlenmeyer flask. After 3.5 h the OD_{600} reached 0.4. The flask was chilled on ice for 10 minutes and IPTG was added to a final concentration of 0.5 mM. The flask was incubated at at 16 °C for 16 h.

The protein expression yielded 2 mg/L of protein expression.

3.7 Kinetics

Spectrophotometry was recorded on a dual path Thermo Scientific™ Evolution™ 300 UV-Vis spectrophotometer blanked on water and water was used in the second cuvette path. Unless specified, all Ellman's assay solutions contain 4% saturated solution DTNB (v/v, ~1 mM), 50 mM KH_2PO_4 (pH 7.42), and 10% DMSO (v/v). DTNB stock solutions were prepared from powder on the day of the assay. Spectra were recorded at 412 nm every 3 s.

SNAC Standard Curve

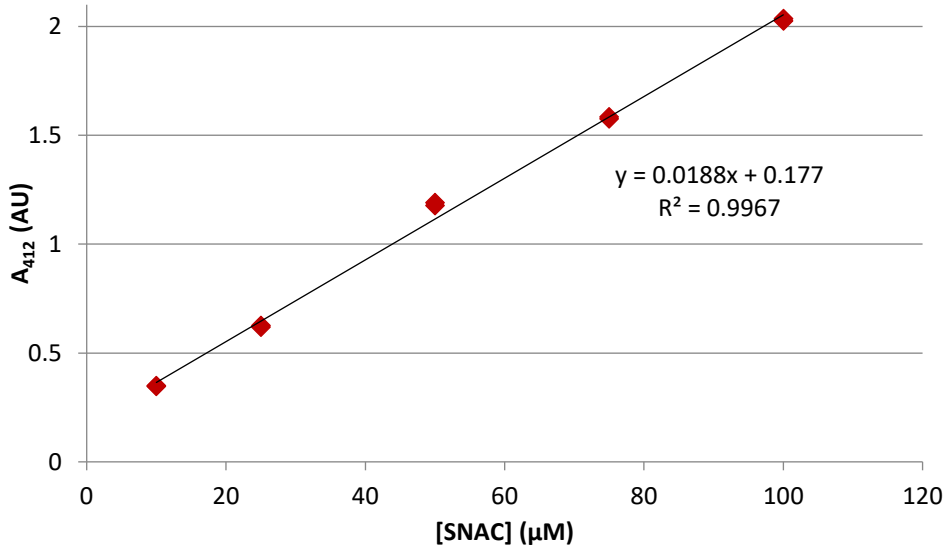


Figure 3-13 Standard curve of SNAC thiol concentrations detected in assay buffer. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4 (pH 7.42), 10% DMSO (v/v), and 0-2 mM SNAC thiol. The SNAC thiol forms a mixed dithiol with Ellman's reagent and the released 3-carboxyl-4-nitrothiophenolate absorbs at 412 nm.

Background Absorbance Signals

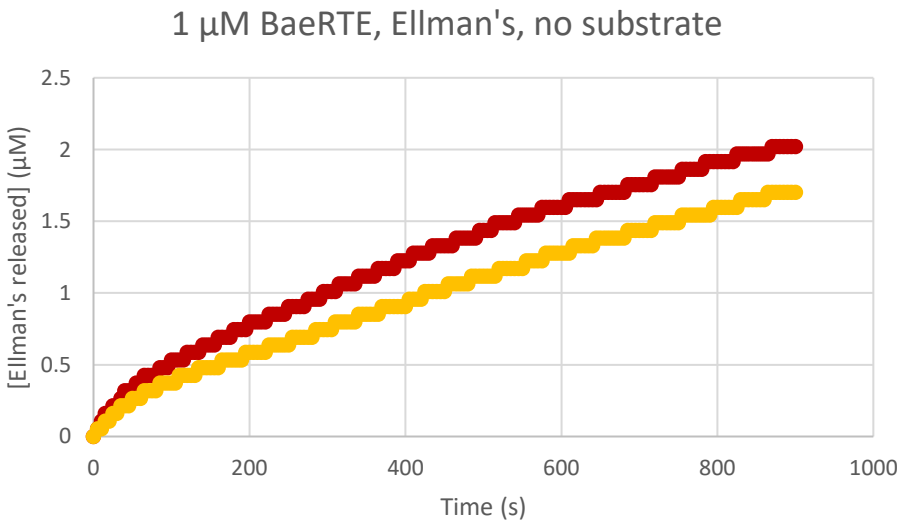


Figure 3-14 Background A_{412} signal produced between BaeTE and Ellman's reagent in assay buffer. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4 (pH 7.42), 10% DMSO (v/v), and 1 µM BaeTE enzyme.

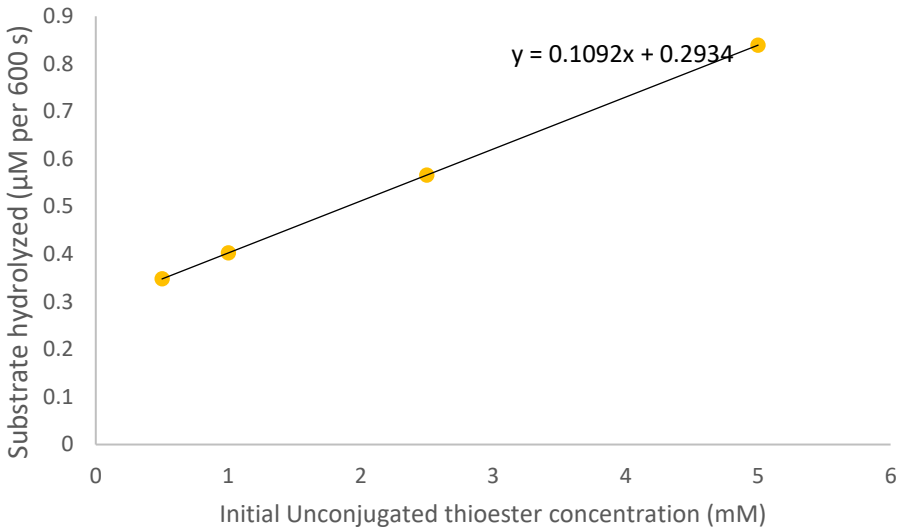


Figure 3.15 Background unconjugated substrate hydrolysis rate in assay buffer. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4 (pH 7.42), 10% DMSO (v/v), and 0.5-5 mM unconjugated substrate. Values reported as full-period totals as a comparison to enzyme equivalents. The 5 mM sample showed 0.84 µM of hydrolysis in 600 s, which was a 0.0168% change from initial substrate concentration, and 0.84 equivalents of the amount of enzyme added to normal assays.

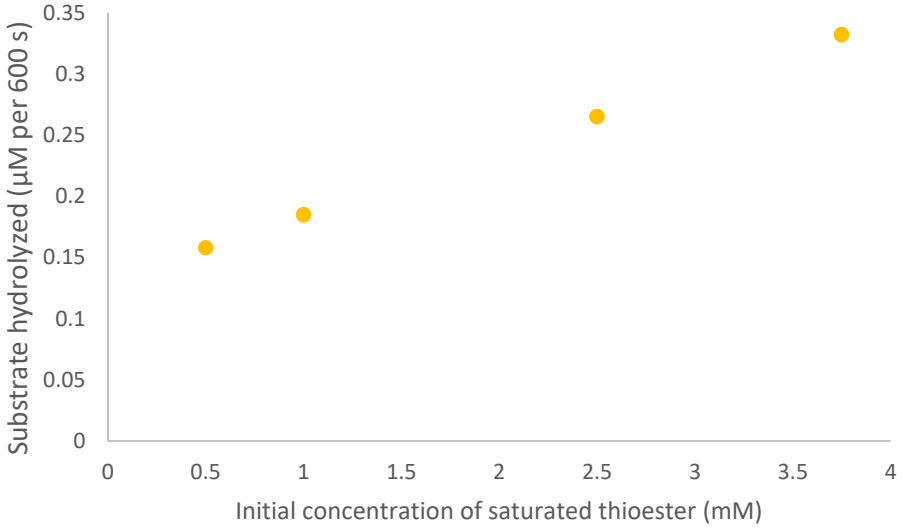


Figure 3-16 Background saturated substrate hydrolysis rate in assay buffer. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4 (pH 7.42), 10% DMSO (v/v), and 0.5-5 mM saturated substrate.

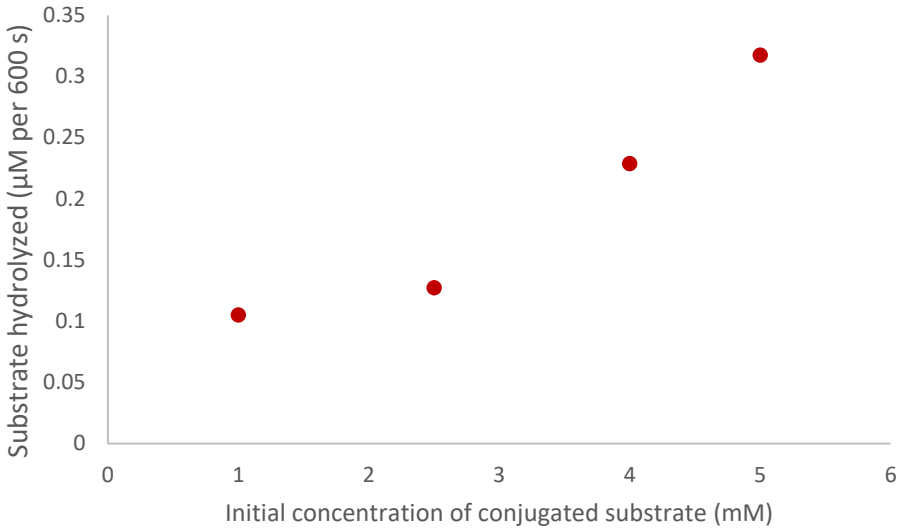


Figure 3-17 Background conjugated substrate hydrolysis rate in assay buffer. Reaction mixtures contained 4% saturated DTNB (v/v), 50 mM KH_2PO_4 (pH 7.42), 10% DMSO (v/v), and 0.5-5 mM conjugated substrate.

3.8 References

1. K. Yabe, *Bulletin of the College of Agriculture*, 1894, **2**, 68-72.
2. S. Sawamura, *Bulletin of the College of Agriculture*, 1906, **7**, 107-110.
3. S. B. Zimmerman, C. D. Schwartz, R. L. Monaghan, B. A. Pelak, B. Weissberger, E. C. Gilfillan, S. Mochales, S. Hernandez, S. A. Currie, E. Tejera and et al., *Journal of Antibiotics*, 1987, **40**, 1677-1681.
4. C. Scotti, M. Piatti, A. Cuzzoni, P. Perani, A. Tognoni, G. Grandi, A. Galizzi and A. M. Albertini, *Gene*, 1993, **130**, 65-71.
5. *USA Patent and Trademark Office, Patent Number US5364623A, Patel, P.S. Mayerl, F., Meyers, E.*, 1994.
6. P. S. Patel, S. Huang, S. Fisher, D. Pirnik, C. Aklonis, L. Dean, E. Meyers, P. Fernandes and F. Mayerl, *Journal of Antibiotics*, 1995, **48**, 997-1003.
7. F. Kunst, N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Dusterhoft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S. Y. Ghim, P. Glaser, A. Goffeau, E. J. Golightly, G. Grandi, G. Guiseppi, B. J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Henaut, H. Hilbert, S. Holsappel, S. Hosono, M. F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. KlaerrBlanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S. M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauel, C. Medigue, N. Medina, R. P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. Oreilly, K. Ogawa, A. Ogiwara, B. Oudega, S. H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, A. M. Prescott, E. Presecan, P. Pujic, B. Purnelle, G. Rapoport, M. Rey, S. Reynolds, M. Rieger, C. Rivolta, E. Rocha, B. Roche, M.

Rose, Y. Sadaie, T. Sato, E. Scanlan, S. Schleich, R. Schroeter, F. Scoffone, J. Sekiguchi, A. Sekowska, S. J. Seror, P. Serror, B. S. Shin, B. Soldo, A. Sorokin, E. Tacconi, T. Takagi, H. Takahashi, K. Takemaru, M. Takeuchi, A. Tamakoshi, T. Tanaka, P. Terpstra, A. Tognoni, V. Tosato, S. Uchiyama, M. Vandenbol, F. Vannier, A. Vassarotti, A. Viari, R. Wambutt, E. Wedler, H. Wedler, T. Weitzenegger, P. Winters, A. Wipat, H. Yamamoto, K. Yamane, K. Yasumoto, K. Yata, K. Yoshida, H. F. Yoshikawa, E. Zumstein, H. Yoshikawa and A. Danchin, *Nature*, 1997, **390**, 249-256.

8. J. Piel, *Natural Product Reports*, 2010, **27**, 996-1047.
9. C. T. Calderone, W. E. Kowtoniuk, N. L. Kelleher, C. T. Walsh and P. C. Dorrestein, *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 8977-8982.
10. J. Hofemeister, B. Conrad, B. Adler, B. Hofemeister, J. Feesche, N. Kucheryava, G. Steinborn, P. Franke, N. Grammel, A. Zwintscher, F. Leenders, G. Hitzeroth and J. Vater, *Molecular Genetics & Genomics*, 2004, **272**, 363-378.
11. X. H. Chen, A. Koumoutsi, R. Scholz, A. Eisenreich, K. Schneider, I. Heinemeyer, B. Morgenstern, B. Voss, W. R. Hess, O. Reva, H. Junge, B. Voigt, P. R. Jungblut, J. Vater, R. Sussmuth, H. Liesegang, A. Strittmatter, G. Gottschalk and R. Borriss, *Nature Biotechnology*, 2007, **25**, 1007-1014.
12. J. Moldenhauer, X.-H. Chen, R. Borriss and J. Piel, *Angewandte Chemie - International Edition*, 2007, **46**, 8195-8197.
13. J. Moldenhauer, D. C. G. Goetz, C. R. Albert, S. K. Bischof, K. Schneider, R. D. Suessmuth, M. Engeser, H. Gross, G. Bringmann and J. Piel, *Angewandte Chemie - International Edition*, 2010, **49**, 1465-1467.
14. D. C. Gay, P. J. Spear and A. T. Keatinge-Clay, *ACS Chemical Biology*, 2014, **9**, 2374-2381.
15. S. Donadio, J. B. McAlpine, P. J. Sheldon, M. Jackson and L. Katz, *Proceedings of the National Academy of Sciences of the United States of America*, 1993, **90**, 7119-7123.
16. T. W. Yu, Y. Shen, Y. Doi-Katayama, L. Tang, C. Park, B. S. Moore, C. Richard Hutchinson and H. G. Floss, *Proceedings of the National Academy of Sciences of the United States of America*, 1999, **96**, 9051-9056.
17. J. Staunton and K. J. Weissman, *Natural Product Reports*, 2001, **18**, 380-416.
18. J. Cortes, K. E. Wiesmann, G. A. Roberts, M. J. Brown, J. Staunton and P. F. Leadlay, *Science (New York, N.Y.)*, 1995, **268**, 1487-1489.
19. J. A. Chemler, A. Tripathi, D. A. Hansen, M. O'Neil-Johnson, R. B. Williams, C. Starks, S. R. Park and D. H. Sherman, *Journal of the American Chemical Society*, 2015, **137**, 10603-10609.
20. A. J. Hughes, J. F. Detelich and A. T. Keatinge-Clay, *Medicinal Chemistry Communications*, 2012, **3**, 956-959.
21. K. Jensen, H. Niederkruger, K. Zimmermann, A. L. Vagstad, J. Moldenhauer, N. Brendel, S. Frank, P. Poplau, C. Kohlhaas, C. A. Townsend, M. Oldiges, C. Hertweck and J. Piel, *Chemistry & Biology*, 2012, **19**, 329-339.
22. Z. Zhou, J. R. Lai and C. T. Walsh, *Chemistry & Biology*, 2006, **13**, 869-879.
23. Y. Liu, T. Zheng and S. D. Bruner, *Chemistry & Biology*, 2011, **18**, 1482-1488.
24. J. Franke and C. Hertweck, *Cell Chemical Biology*, 2016, **23**, 1179-1192.
25. S. Y. Tsuji, D. E. Cane and C. Khosla, *Biochemistry*, 2001, **40**, 2326-2331.
26. T. Liu, X. Lin, X. Zhou, Z. Deng and D. E. Cane, *Chemistry & Biology*, 2008, **15**, 449-458.
27. S. Mueller, S. N. Strack, B. C. Hoefler, P. D. Straight, D. B. Kearns and J. R. Kirby, *Applied and Environmental Microbiology*, 2014, **80**, 5603-5610.
28. R. S. Gokhale, D. Hunziker, D. E. Cane and C. Khosla, *Chemistry & Biology*, 1999, **6**, 117-125.
29. K. K. Sharma and C. N. Boddy, *Bioorganic & Medicinal Chemistry Letters*, 2007, **17**, 3034-3037.
30. M. Wang, Doctorate of Philosophy in Chemistry Monograph, Syracuse University, 2010.
31. H. X. Lu, S. C. Tsai, C. Khosla and D. E. Cane, *Biochemistry*, 2002, **41**, 12590-12597.

32. M. Wirz, 2012.
33. K. Buntin, K. J. Weissman and R. Muller, *Chembiochem : a European journal of chemical biology*, 2010, **11**, 1137-1146.
34. G. W. Heberlig, M. Wirz, M. Wang and C. N. Boddy, *Organic Letters*, 2014, **16**, 5858-5861.
35. J. B. Scaglione, D. L. Akey, R. Sullivan, J. D. Kittendorf, C. M. Rath, E. S. Kim, J. L. Smith and D. H. Sherman, *Angewandte Chemie - International Edition*, 2010, **49**, 5726-5730.
36. S. C. Tsai, H. Lu, D. E. Cane, C. Khosla and R. M. Stroud, *Biochemistry*, 2002, **41**, 12598-12606.
37. V. Barbe, S. Cruveiller, F. Kunst, P. Lenoble, G. Meurice, A. Sekowska, D. Vallenet, T. Wang, I. Moszer, C. Medigue and A. Danchin, *Microbiology*, 2009, **155**, 1758-1775.
38. J. Blom, C. Rueckert, B. Niu, Q. Wang and R. Borriss, *Journal of Bacteriology*, 2012, **194**, 1845-1846.
39. C. Diguan, P. Li, P. D. Riggs and H. Inouye, *Gene*, 1988, **67**, 21-30.
40. C. V. Maina, P. D. Riggs, A. G. Grandea, B. E. Slatko, L. S. Moran, J. A. Tagliamonte, L. A. McReynolds and C. Diguan, *Gene*, 1988, **74**, 365-373.
41. G. Ashley, I. C. Chan-Kai and M. A. Burlingame, Google Patents2006.
42. B. D. Ames, N. Chi, J. Bruegger, P. Smith, W. Xu, S. Ma, E. Wong, S. Wong, X. Xie, J. W. H. Li, J. C. Vederas, Y. Tang and S.-C. Tsai, *Proceedings of the National Academy of Sciences of the United States of America*, 2012, **109**, 11144-11149.
43. K. R. Conway and C. N. Boddy, *Nucleic Acids Research*, 2013, **41**, D402-D407.
44. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science (New York, N.Y.)*, 2001, **291**, 1790-1792.

Chapter 4 Deconvolution of Kinetic Analysis

4.1 Introduction

Kinetic properties allow us to engineer catalysts, give quantifiable comparisons on substrate selectivity and likelihood of *in vivo* processing, and give us insights into the mechanism of action and energetics of an enzyme. The most common kinetic properties determined for enzymatic reactions are the Michaelis-Menten parameters, k_{cat} and K_M.

k_{cat} is defined as the catalytic constant and represents the number of molecules a saturated enzyme would turn over in a unit time.

K_M is defined as the apparent equilibrium constant that represents all substrate-enzyme complexes during catalysis. Empirically, this constant is equal to the substrate concentration at which the reaction velocity reaches half the maximal saturated rate.

For an interconnected series of reactions, the steady-state reaction is determined by the rate limiting step. Type I TEs convert ACP-tethered polyketide acyl compounds into ester and acid terminated free compounds in two stages. In the Loading Stage, the acyl chain is "loaded" onto the TE active-site Ser through displacement of the ACP's Ppant thiol. In the Release Stage, the departing Ppant is replaced with another nucleophile; the nucleophile attacks the activated acyl-enzyme Ser-ester and the acyl chain departs into solution (See **2.1.3** and **Figure 2-5** for further details). This nucleophilic exchange is common to the alpha/beta hydrolase enzyme superfamily that contains type I TEs.[1]

There is a wealth of kinetic characterization of the rate-limiting step during catalysis in serine proteases, which like TEs are α/β hydrolases. For example, kinetic characterization of serine proteases with para nitrophenyl ethyl carbonate ester[2] demonstrates rapid burst kinetics at the start of their experiments. This suggests that the release of the pNP occurs rapidly during the initial enzyme Loading Stage and is subsequently delayed as the process becomes limited by the rate of

deacylation (release stage). In the context of TEs, Li *et al.*[3] showed similar results with luxB-type TE.[1] They showed that radiolabelled acyl groups remain bound to the TE in the protein fraction for a period of time after being chased by cold substrate equivalents.

While both these examples support the generalization that α/β hydrolases are deacylation limited, disambiguation between the rates at both the loading and release stages of the bacillaene TE is needed before kinetic parameters can be ascribed to one of these two stages.

Without differentiating between the steps, literature reports have demonstrated that TEs normally produce product in a pseudo first-order manner that is saturable. This gives rise to a traditional Michaelis-Menten type substrate response profile. In these cases, k_{cat} and K_M constants reflective of the catalysis process can be isolated.

In some instances these constants cannot be obtained. For example some substrates are sparingly soluble, preventing kinetic characterization of the TEs at substrate concentrations well above the K_M. In these cases the specificity constant k_{cat}/K_M cannot be resolved into its individual components.

In other cases, inhibition interferes with the enzyme function. Serine hydrolases have been studied in depth,[1, 4-6] and many cases of inhibition have been observed.[7-9] In order to account for the effects of inhibition, slow-on reversible and irreversible inhibitors have been modelled through pair-wise comparisons of rate of substrate catalysis with and without the inhibitor present, as well as direct analysis of a reaction "progress curve" to directly measure enzyme inactivation as a function of time.[10, 11] However, in cases when TEs have shown substrate inhibition with substrate analogues *in vitro*,[12-14] a pair-wise comparison is not possible as the substrate acts as the inhibitor as well.

In the assay results presented in **Chapter 3**, the kinetic constants of the bacillaene TE cannot be resolved using a direct fit. There, the rate is complicated by factors beyond the concentration of the substrates and products, and the occupancy of the enzyme, as evidenced by change in thiol production rate without significant change in concentration of SNAC substrate or free-thiol product. A quantitative description of the activity and inhibition requires a predictive model for bacillaene

TE inactivation. This would give an indirect measure of the enzyme activity while active, and shed light on the cause of inactivation. This can be approached by treating the enzyme inactivation as a standard chemical reaction, and characterizing the partial order of reaction with respect to enzyme (autocatalytic inactivation similar to proteases), substrate (substrate inhibition), and other assay substituents (*e.g.*, H^+ or OH^- catalyzed).

In this chapter, the conditions and constituents of the colorometric assay, including the concentration of thioester substrate, and, indirectly through loss of activity,enzyme concentration, are investigated for their impact on the rate of enzyme inactivation over time. The partial first-order impacts of substrate and enzyme concentration are incorporated into an empirical rate law, which can be combined with the normal enzyme function of catalysis of thioester. The constants from the resulting time-dependent model are fit to the experimental results from **Chapter 3.** Finally, the goodness of fit for this model is assessed by modeling individual assay concentrations and comparing the observed and modelled amounts of thiol produced.

4.2 Results

4.2.1 Rate of enzyme inactivation for the saturated SNAC 3 - BaeRTE system

The BaeRTE/ saturated 2-methyl valeric-SNAC **3** (**Figure 3-3**) system was selected for in-depth inactivation characterization because it presented the greatest decrease in observed reaction rate (**Figure 3-7, Figure 4-1**) over time, showing 3-10 enzyme turnovers before the hydrolysis rate slowed to background rates.

Estimation of the enzyme activity over time can be determined through the local slope of the progress curve (as an approximation of instantaneous rate of reaction). **Figures 4-1-4-3** examine the dependence of the inactivation process on the concentration of active enzyme over time. In an idealized assay, the functions displayed on the dependent axis will vary linearly with respect to time if the partial order of the inactivation process with respect to enzyme is 0^{th}, 1^{st}, or 2^{nd} order in **Figure 4-1**, **Figure 4-2**, or **Figure 4-3**, respectively.

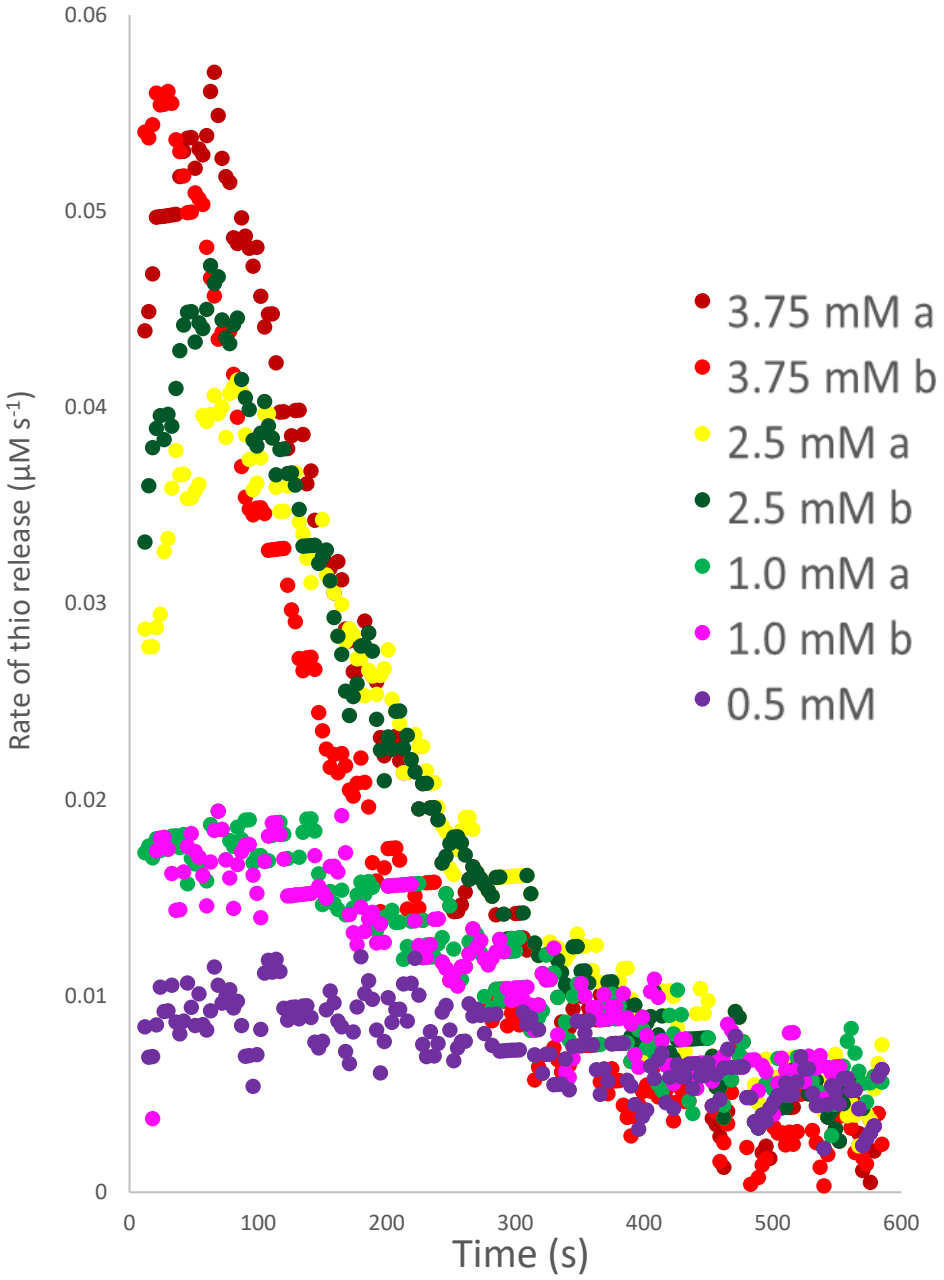


Figure 4-1 Rate of BaeRTE catalyzed thiol release from saturated acyl-SNAC substrate over time. Instantaneous rate estimated as change in thiol concentration over an 18 second window centered at the listed time point. The rate of thiol released dropped non-linearly over time in a substrate-dependent fashion.

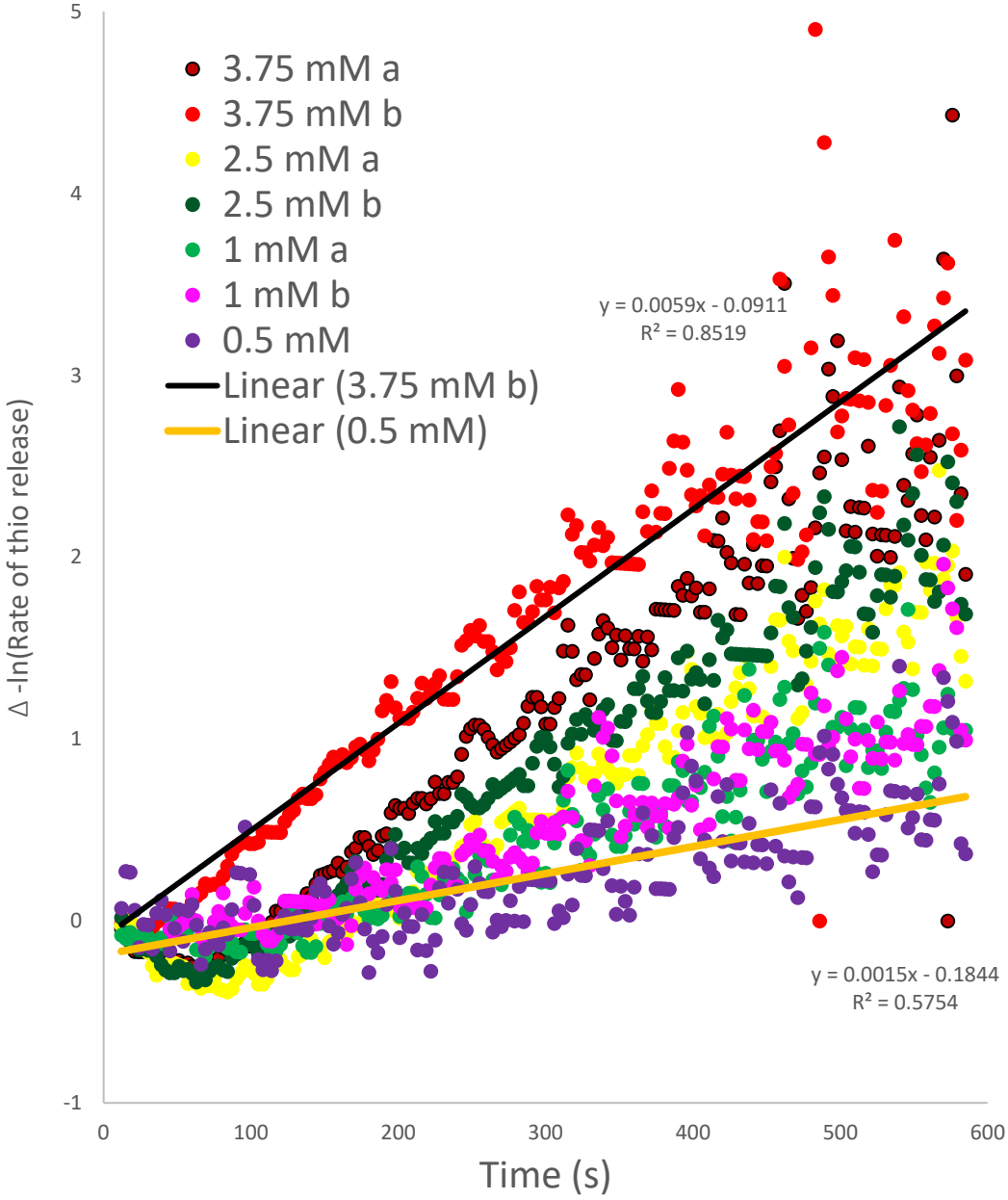


Figure 4-2 Examination of first-order dependence of BaeRTE rate decline based on BaeRTE and saturated acyl-SNAC concentration. Data traces were aligned at zero by subtracting the initial rate of reaction from each data point. Note that the linear trend for 3.75 mM is a linear interpolation of only the first 300 s and estimates the initial inactivation rate constant.

Figure 4-3 Examination of second-order dependence of rate decline on substrate concentration. Rate dropped non-linearly over time in a substrate-sensitive fashion. For enzyme concentrations lower than 1 M, this function diverges upward from linear in truly first-order rates, and downward away from linear in truly 0^{th} order functions.

The linear trend in **Figure 4-2** supports models of enzyme inactivation with first-order dependence of inactivation rate on initial concentration of enzyme. Additionally, **Figure 4-1** and **Figure 4-2** demonstrate that the inactivation rate shows a dependence on substrate concentration. In assays with 3.75 mM saturated acyl-SNAC, the slope of **Figure 4-2** is 0.0068 s^{-1} as compared to 0.0015 μM s^{-1} in assays with 0.5 mM saturated acyl-SNAC. This corresponds to a 4.53 fold change in inactivation for a 7.5-fold change in concentration. If the decrease in BaeRTE activity was proportional to both enzyme and acyl substrate concentration, then the general rate law would take the form below:

$$\frac{d[BaeTE]}{dt} = -k[BaeTE][Acyl\text{-}SNAC]$$

and as shown with **Figure 3-7**, at most 2% of the substrate is hydrolyzed to thiol over the course of the standard 600 s assay (for example, ~ 10 μM for 0.5 mM substrate). Thus,

$$[Acyl\ SNAC] \cong [Acyl\text{-}SNAC]_0$$

and the reaction would behave as pseudo-first order on enzyme concentration alone. As a simplification, the second order rate constant k could be combined with the initial substrate concentration to form a pseudo-first order rate constant

$$k' = k * \ [Acyl\text{-}SNAC]_0$$

and the rate equation can be rearranged and integrated to resemble the form shown in **Figure 4-2**

$$\frac{1}{[BaeTE]}\frac{d[BaeTE]}{dt} = -k'$$

$$\ln([BaeTE]_t) = -k' * t + \ln([BaeTE]_0)$$

$$-(\ln([BaeTE]) - \ln([BaeTE]_0)) = k' * t$$

$$-\ln\left(\frac{[BaeTE]_t}{[BaeTE]_0}\right) = k' * t$$

$$rate\ of\ thiol\ release_t = k * [BaeTE]_t{}^x * [SNAC]_t{}^y$$

Based on **Figure 3-7**, only a maximum of 0.7 % of the substrate was converted in 600 s for any sample.

Thus $[\mathrm{SNAC}] \cong [\mathrm{SNAC}]_0$

Assuming $[SNAC]_t{}^{y} \approx [SNAC]_0{}^{y}$ and $x = 1$

$$[BaeTE]_t = \frac{rate\ of\ thiol\ release_t}{k * [SNAC]_0{}^{y}}$$

$$-\ln\left(\frac{\frac{rate\ of\ thiol\ release_t}{k*[SNAC]_0{}^{y}}}{\frac{rate\ of\ thiol\ release_0}{k*[SNAC]_0{}^{y}}}\right) = k' * t$$

$$-\ln\left(\frac{rate\ of\ thiol\ release_t}{rate\ of\ thiol\ release_0}\right) = k' * t$$

$$-\ln\left(\frac{rate\ of\ thiol\ release_t}{rate\ of\ thiol\ release_0}\right) = k' * t \quad -\ln(rate\ of\ thiol\ release_t) + \ln(rate\ of\ thiol\ release_0) = k' * t$$

$$-\Delta(\ln(rate\ of\ thiol\ release)) = k' * t$$

Overall, the average assay reaction rate decreased by 66% over 600 s and continued to follow a pseudo-first order of decline. If the reaction decline was associated with a significant degree of reversible reactivation or recovery of inhibited TE species, there would be evidence of a significant deviation from linearity after 200 s. Additionally, if the decline in reaction rate was saturable with respect to SNAC substrate within the range of studied concentrations, a combined saturable 1st order term resembling the Michaelis Menten equation could be implemented:

$$k' = \frac{k_{inh} * [Acyl\text{-}SNAC]_0}{K_{I+}[Acyl\text{-}SNAC]_0}$$

However, in the absence of a significant contribution to thiol hydrolysis from a reactivated enzyme species, and in the absence of saturated inactivation, the simplified proportional irreversible model is proposed.

4.2.2 Irreversible substrate inhibition model

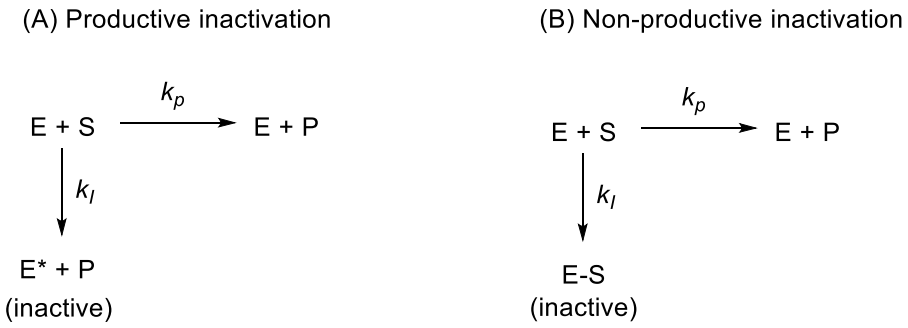


Figure 4-4 Possible routes to irreversible substrate mediated enzyme-inactivation and product-formation. Reversible binding steps are collapsed into single-step irreversible processes for tractable model derivation. Each mechanism gives rise to inactive enzyme as well as hydrolyzed thiol in productive catalysis, but the appropriate k_I and k_P values can only be calculated with the appropriate model; the selection of appropriate model requires verification of the identity of the products and inactive species. Route (A) is assumed based on the likelihood of substrate-thioester cleavage during protein inactivation. Uncompetative and noncompetative allostery were excluded from consideration based on the first-order tendency of **Figure 4-2**.

Assuming that the rate of inactivation is proportional to both the current concentration of free TE [E] and the concentration of substrate [S] (insaturable), then the instantaneous change in inhibited enzyme is

$$\frac{d[EI]}{dt} = (k_I) * [\text{substrate}] * [E] \qquad [\ \mathbf{4\text{-}1}\ ]$$

Based on **Figure 3-7**, only a maximum of 0.7 % of the substrate was converted in 600 s for any sample. Thus

$$[\text{substrate}] \cong [\text{substrate}]_0$$

and based on a mole balance of the TE

$[E] = [E]_0 - [EI]$.

As a result, [**4-1**] can be shown to be a function of [EI]:

$$\frac{d[EI]}{dt} = (k_I) * [\text{substrate}]_0 * ([E]_0 - [EI])$$

This equation can be rearranged to integrate over time

$$\frac{d[EI]}{([E]_0 - [EI])} = (k_I) * [\text{substrate}]_0 * dt$$

$$\frac{d[EI]}{([E]_0 - [EI])} = (k_I) * [\text{substrate}]_0$$

assume $[\text{EI}]_0 = 0$, then the integral becomes

$$\int_0^{EI_t} \frac{d[EI]}{[E]_0 - [EI]} = (k_I) * [\text{substrate}]_0 * t$$

Since $\int_o^x \frac{dx}{-x} = -\ln(x)$

$$(-\ln([E]_0 - [EI]) - (-\ln([E]_0))) = (k_I) * [\text{substrate}]_0 * t$$

$$\ln\frac{([E]_0 - [EI])}{([E]_0)} = -(k_I) * [\text{substrate}]_0 * t$$

Putting both sides to the power of e and returning to a function in a form tied to enzyme activity ([E]) yields:

$$\text{E(t)} = E_0 * e^{-k_I * [\text{substrate}]_0 * t}$$

Here, the active TE concentration is modelled as a first-order decay function similar to Tian and Tsou's derivation for irreversible inhibition of proteases by molecules other than substrate.[10]
This estimated decay model for enzyme activity can be applied to estimate the amount of product produced if the "Productive Inhibition" model of **Figure 4-4A** is correct

$$\frac{dP}{dt} = (k_p + k_I) * [\text{substrate}]_0 * [E]_0 * \left(\text{e}^{-\text{k}_I * [\text{substrate}]_0 * \text{t}}\right)$$

Here, the rate of product produced ($\frac{dP}{dt}$) is proportional to the amount of enzyme remaining (estimated using the decay), and increases as either the catalytic turnover pseudo-first order rate constant (k_P) or the inactivation rate constant (k_I) are enhanced. The additional simplification added here assumes the catalytic turnover of the TE is also not saturated ([S]<<K_M) and proportional to $[\text{substrate}]_0$.

The product rate equation can be integrated over the assay period to estimate the amount of product produced (and the amount of indicator present) at any time (t) in the assay. If we assume

that the product resulting from enzyme activity is zero before the enzyme is added at t=0, then the definite integral becomes

$$P(t) = \left(k_p + k_I\right) * [\text{substrate}]_0 * [E]_0 * \frac{\left(e^{-k_I * [\text{substrate}]_0 * 0} - \left(e^{-k_I * [\text{substrate}]_0 * t}\right)\right)}{k_I * [\text{substrate}]_0}$$

This simplifies to

$$P(t) = \frac{\left(k_p + k_I\right)}{k_I} * [E]_0 * \left(1 - \left(e^{-k_I * [\text{substrate}]_0 * t}\right)\right) \quad [\ \textbf{4-2}\ ]$$

The form above is useful for applying known k_P and k_I to estimate P(t), but this can be rearranged to solve for the rate constants.

$$\frac{\left(k_p + k_I\right)}{k_I} * [E]_0 * e^{-k_I * [\text{substrate}]_0 * t} = \frac{\left(k_p + k_I\right)}{k_I} * [E]_0 - P(t)$$

$$k_I * [\text{substrate}]_0 * t + \ln\left(\frac{\left(k_p + k_I\right)}{k_I} * [E]_0\right) = -\ln\left(\frac{\left(k_p + k_I\right)}{k_I} * [E]_0 - P(t)\right)$$

Let $C = \frac{\left(k_p + k_I\right)}{k_I} * [E]_0$.

$$k_I * [\text{substrate}]_0 * t + \ln(C) = -\ln(C - P(t))$$

$$k_I * [\text{substrate}]_0 * t = -\ln\left(\frac{C - P(t)}{C}\right)$$

Further simplification gives

$$k_I * [\text{substrate}]_0 * t = -\ln\left(1 - \frac{P(t)}{C}\right). \quad [\ \textbf{4-3}\ ]$$

This equation implies that for some value of C, a plot of $-\ln\left(1 - \frac{P(t)}{C}\right)$ vs t should produce a linear relationship for all substrate concentrations, and whose slope, when plotted against substrate concentration, should give k_I.

The thiol release data from **Chapter** 3 was expressed in the form of equation [**4-3**] to solve for the rate constants. If some value of C is found to linearize the function, then the change in enzyme activity may be appropriately fit by the model. First, the function was calculated for some C over the entire time range (for example 0 to 600 s) for all of the substrate concentrations studied. Then C was optimized by a numerical method (Solver, Excel) by minimizing the residual sum of squares

between values of $-\ln(1-\frac{P(t)}{C})$ and the fitting function y = mx + b. Values of C must be bounded for the numerical method, because the natural logarithm of negative numbers is undefined. Using the definition of C:

$$C = \frac{(k_p+k_I)}{k_I} * [E]_0 \qquad C \geq [E]_0 \qquad C = \lim_{t\to\infty}(P(t))$$

Experimentally, C was restricted to larger than the highest observed P(t) for a given enzyme concentration. Tian and Tsou[10] used a similar function to linearize rate equations for non-substrate irreversible inhibitors, and Figure 3 of their work shows the behaviour of $\ln([P]_\infty - [P])$ in response to application of an inhibitor.

Graphically, the shape of the product formation function [**4-2**] is defined piecewise. The value of C represents the horizontal line that defines the maximum product achievable, k_P is proportional to the initial slope of the rate, and k_I is proportional to the initial slope as well as the second-order decline (deceleration). For the same C, larger values of k_I and k_P produce more severe, rapid hyperbolic curves, whereas small rate constants produce a slower progression curve that approaches C in a more linear appearance.

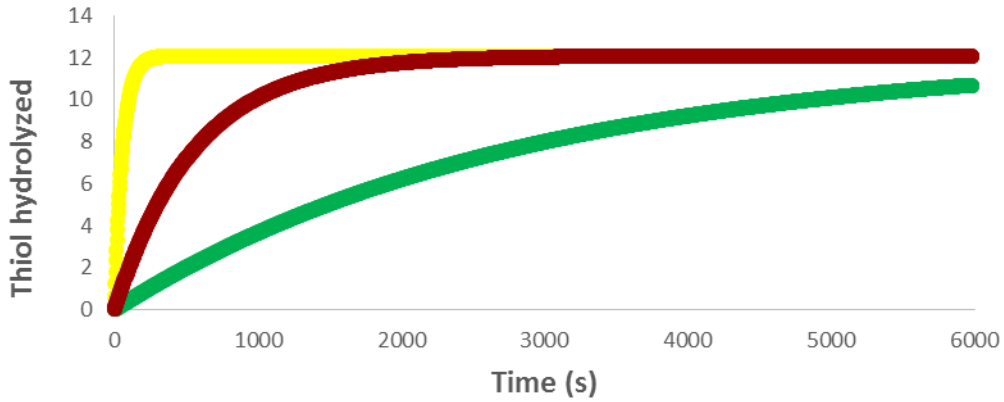


Figure 4-5 Product model dependence on magnitude of rate constants. For the same value of C=12 and a substrate concentration of 1 mM, a five-fold decrease (green) and ten-fold increase (yellow) in the magnitudes of k_I and k_P are shown to alter the function curvature.

Using the solver numerical minimization, C converged to 12.15 for the BaeRTE/saturated SNAC combination. **Figure 4-6**, shown below, illustrates the behaviour of each saturated substrate and BaeRTE reaction fit into [**4-3**] with a C of 12.15 using the experimental P(t) data from **Figure 3-7**.

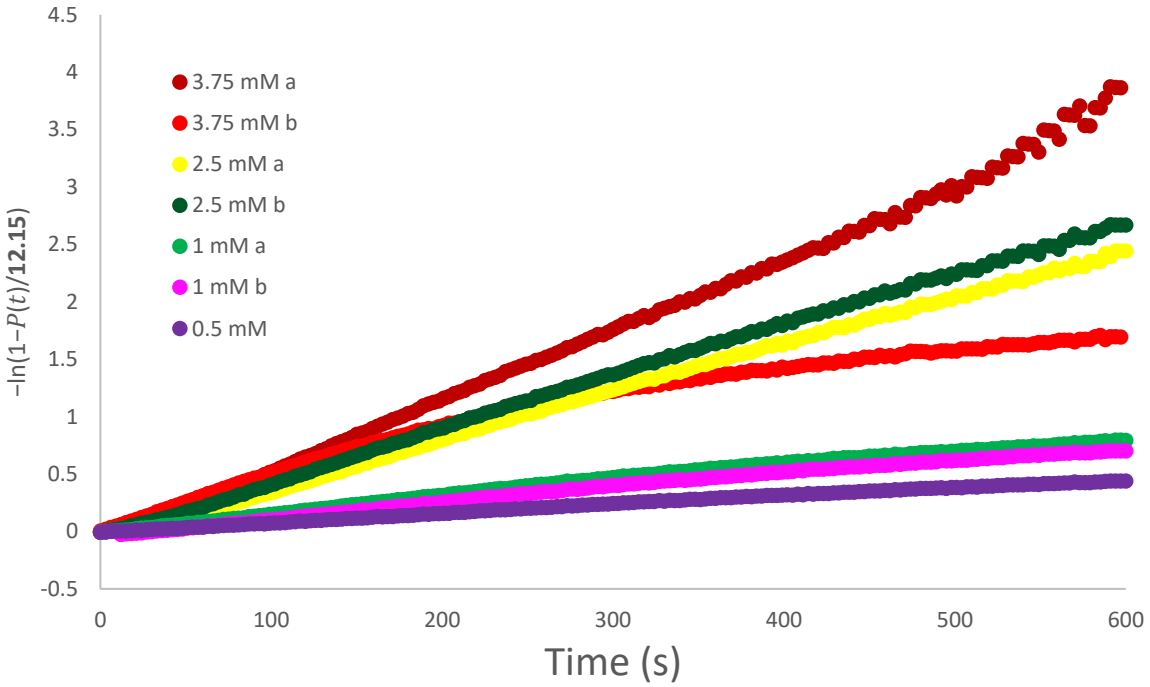


Figure 4-6 Linear estimate of combined rate constant C for BaeRTE and saturated substrate. Rate dropped non-linearly over time in a substrate-sensitive fashion. A value of 12.15 μM minimized the total deviation from linear. Notably, sample 3.75 mM b deviated from the common C and linearized with an alternate value of C = 10.15.

Recall from [**4-3**] that the slope from **Figure 4-6** is $k_I * [\text{substrate}]_0$. k_I can then be estimated graphically by plotting the slopes against substrate concentration:

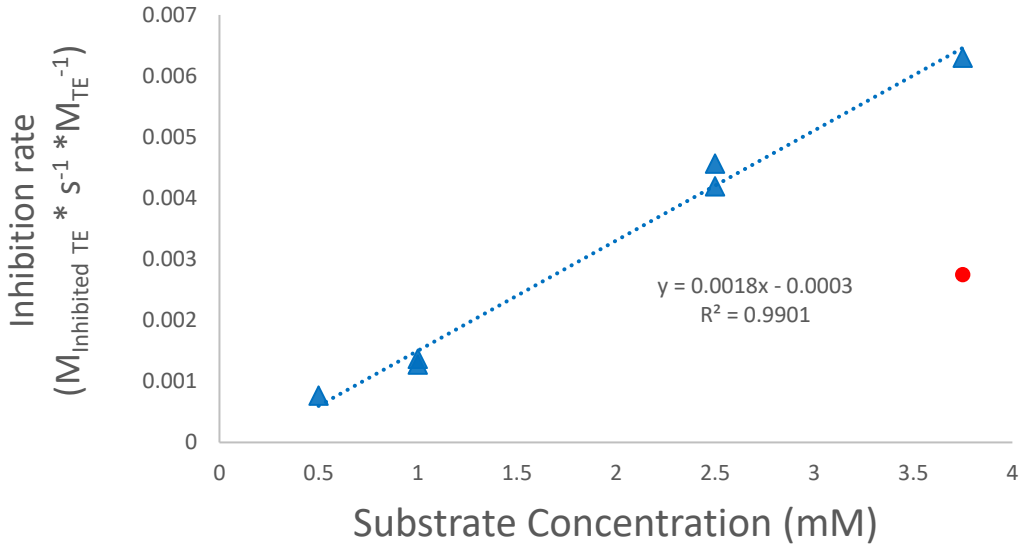


Figure 4-7 Approximation of saturated substrate k_I for BaeRTE. R The value for the non-linearized 3.75 mM b (red circle) was excluded in the estimate of k_I.

With the value of k_I= 1.734 $M_{Product}$ * s^{-1} *M_{TE}^{-1} * $M_{Saturated\ Substrate}^{-1}$, k_P may be calculated from the definition of C

$$k_P = k_I * (\frac{C}{[E]_0} - 1)$$

and with the saturated substrate with BaeRTE, k_p = 19 $M_{Product}$ * s^{-1} *M_{TE}^{-1} * $M_{Saturated\ Substrate}^{-1}$.

If the productive inactivation model from **Figure 4-4A** is appropriate for the observed phenomenon and **[4-2]** reasonably models the accumulated free thiol over the course of an assay, then $P(t)_{modelled} \sim P(t)_{Observed}$. **Figure 4-8** illustrates the model fit to the observed hydrolysis of saturated SNAC **3** by BaeRTE.

4.2.2.1 Test of first-order decay model fit using BaeRTE saturated substrate assay data

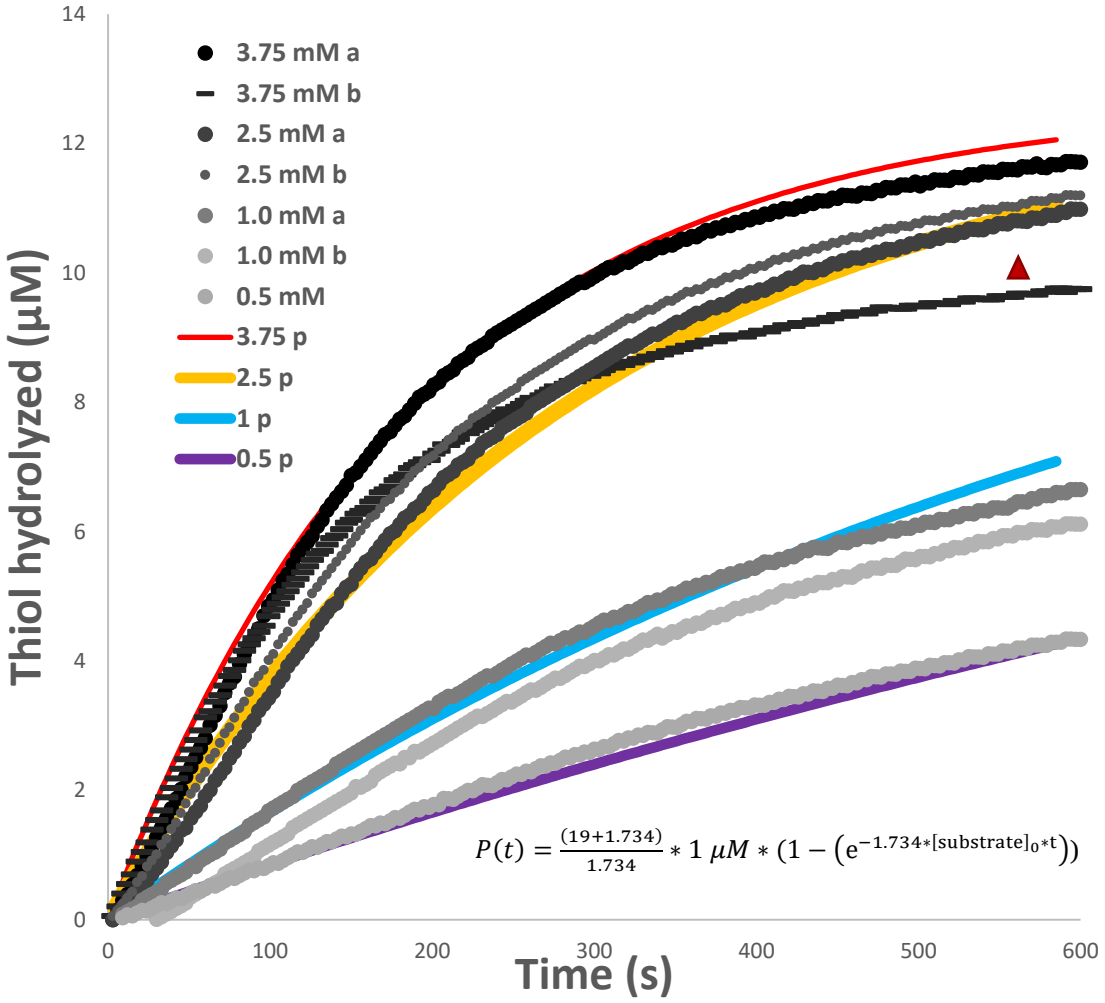


Figure 4-8 Modelled reaction of saturated substrate with BaeRTE. The value for the non-linearized 3.75 b (red triangle) deviated significantly from the collective model. The model underestimates the observed decay of activity for lower concentrations of substrate (400 s onward).

The difference in function **[4-2]** when using $P(t)_{modelled}$ or $P(t)_{Observed}$ should tend around zero for all substrate concentrations examined. **Figure 4-9** outlines equation **[4-2]**'s residuals when fit with k_I = 1.7 and k_p = 19.34 $M_{Product} * s^{-1} * M_{TE}^{-1} * M_{Saturated\ Substrate}^{-1}$.

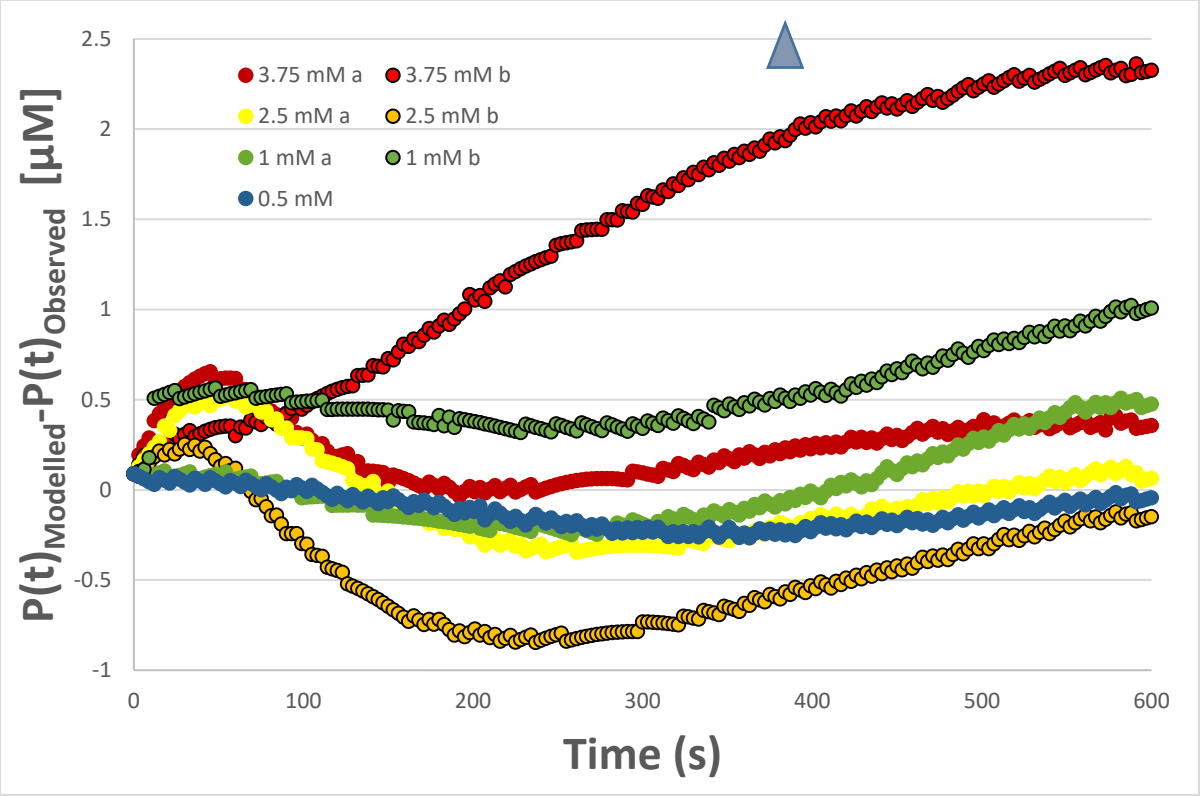


Figure 4-9 Residual of modelled reaction of saturated substrate with BaeRTE. The value for the nonlinearized 3.75 b (blue triangle) displayed a large deviation from the model. The model overestimates hydrolysis in the first 100 s, with the error increasing for larger initial substrate concentrations. After 450 s, the model overestimates the remaining hydrolysis activity in the 1 mM replicates.

Overall, the fitted model from **Figure 4-8** deviates from the observed hydrolysis by $\pm$ 5-14%.

Table 4-1 Fitted kinetic constants for all substrate/enzyme combinations.

Values for k_I, k_P, and C were fitted for each data set from **Chapter 3** using the same numerical method as for the saturated substrate with BaeRTE without independent checks for first-order decay in enzyme activity.

Protein	Constant	Substrate		
		Saturated	Conjugated	Unconjugated
BaeRTE	k_I ($M^{-1}s^{-1}$)	1.7	0.3	0.8
	k_P ($M^{-1}s^{-1}$)	19	0.9	1.2
	C ($10^{-6} * M$)	12.15	3.72	2.59
BaeACPTE (holo)	k_I ($M^{-1}s^{-1}$)	0.8	0.3	0.4
	k_P ($M^{-1}s^{-1}$)	2.9	0.2	1.0
	C ($10^{-6} * M$)	4.43	1.66	3.42

Table 4-1 shows an average lifetime-turnover number of roughly 3.5 for all constructs. This exceeds the one-fold turnover predicted in distinct deacylation-limited burst kinetics. Except for the saturated substrate with BaeRTE, rates of both inactivation and productive turnover are insufficiently distinct to distinguish one substrate from another. Compared to type I *cis*-AT TEs, k_p is = k_{cat}/K_M.

In Liu *et al.*'s 2012 work on the *cis*-AT piericidin TE,[15] three SNAC-diketides were reacted with pieTE in an Ellman's colorometric assay, and each showed a turnover number of 3-16 above the background controls. In order to compare their data against **Table 4-1**, the decay function [**4-2]** was fit to Liu *et al.* Figure 5; pieTE exhibited k_I and k_P values that ranged from 1-2.5 and 5-35 $M^{-1}s^{-1}$ respectively. The authors did not comment on the loss of activity observed in the assay.

The pieTE results constitute both a qualitative observation of TE inactivation in an Ellman's based assay, and a quantitative measure of TE activity towards a range of SNAC analogues, but without an experiment using the same substrate on both enzymes, there is no direct way of commenting on the relative activity of these two constructs without accounting for the inherent reactivity of the substrates.

Figures 4-12 - 4-16 demonstrate the model fit for the five other substrate enzyme combinations. The fitted constants from **Table 4-1** generate a reasonable model of the saturated SNAC **3**-*holo*BaeACPTE hydrolysis that overestimates the rate of hydrolysis at 1 mM **3** after 450 s. However, the kinetic constants derived from **[4-3]** for **1** and **2** hydrolyzed in the presence of either BaeRTE or *holo*BaeACPTE do not accurately model the observed progression curves. This poor modeling may arise either from the inapplicability of the irreversible "Productive inactivation" model proposed in **Figure 4-4** (discussed in 4.2.2.2), or from inadequate model fitting.

4.2.2.2 Test of Equation **[4-2]** fit to literature reports of substrate inhibition

As a test of equation **[4-2]**, an alternate example of suspected substrate inhibition in guinea pig liver transglutaminase (gTGase) was evaluated.[16] In this example, the reaction mechanism follows a two-step Loading and Release similar to a TE, but with gTGase, the hydrolysis step is rate limiting. The enzyme activity in this assay was ~ 1000 fold more rapid than the TE SNAC assay, and data was collected via stop flow at time intervals beginning at 5/100ths of a ms and gradually widening to every 0.46 s over a total observation period of 50 s. The raw data from the progression curve in **Figure 6B** of Gravel *et al.*[16] was fit with **[4-3]** linearization and nonlinear regression. **Figure 4-10** outlines the relative fits.

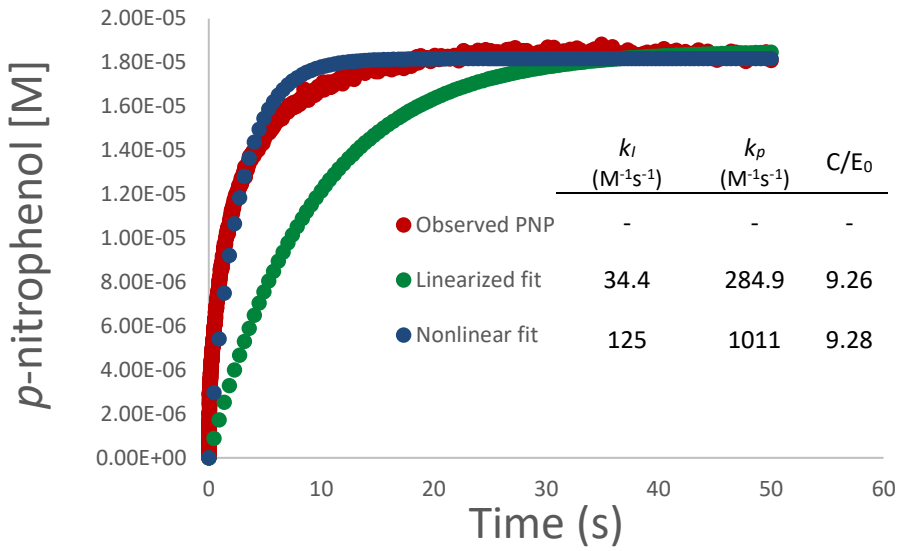


Figure 4-10 Test of Equation [4-2] against literature data of a transglutaminase exhibiting potential substrate inhibition.[16] Fits using the logarithmic linearization **[4-3]** underestimated the observed release of *p*-nitrophenol from 3.1 mM carbonate substrate by gTGase reported by Gravel *et al.*[16] (red circle). Modeling constraints limit **[4-3]** to model a product plateau equal to or greater than the largest observed product concentration. The nonlinear regression of evenly spaced data points (blue circle) generated the closest fit.

The fit linearized to **[4-3]** underestimate the initial reaction rate over the first 15 s, and assigned excess hydrolysis over the 10-25 s range as it converged at the same total product levels. The nonlinear fit of **[4-2]** underestimated the hydrolysis rate for the first 2 s and over estimated it over the period of 5-15 s. Importantly, both fitting methods solve for a value of C of 9.26-9.28. However, the nonlinear fit distinguishes between the different pairs of k_i and k_p that could give rise to a C of 9.28, whereas the linearization requires a second trial with a different substrate concentration in order to uniquely define k_i. The assumption that **[4-3]** passes through the origin was made in order to proceed from C to k_i with a single trial, but as demonstrated with the saturated SNAC **3** + TE assay in **Figure 4-7**, the different substrate concentrations could be equally weighted to find a common C, but the resulting best-fit for k_i function had a non-zero y-intercept.

The deviations of the nonlinear fit of **[4-2]** from the observed data are further described in **Figure 4-11**.

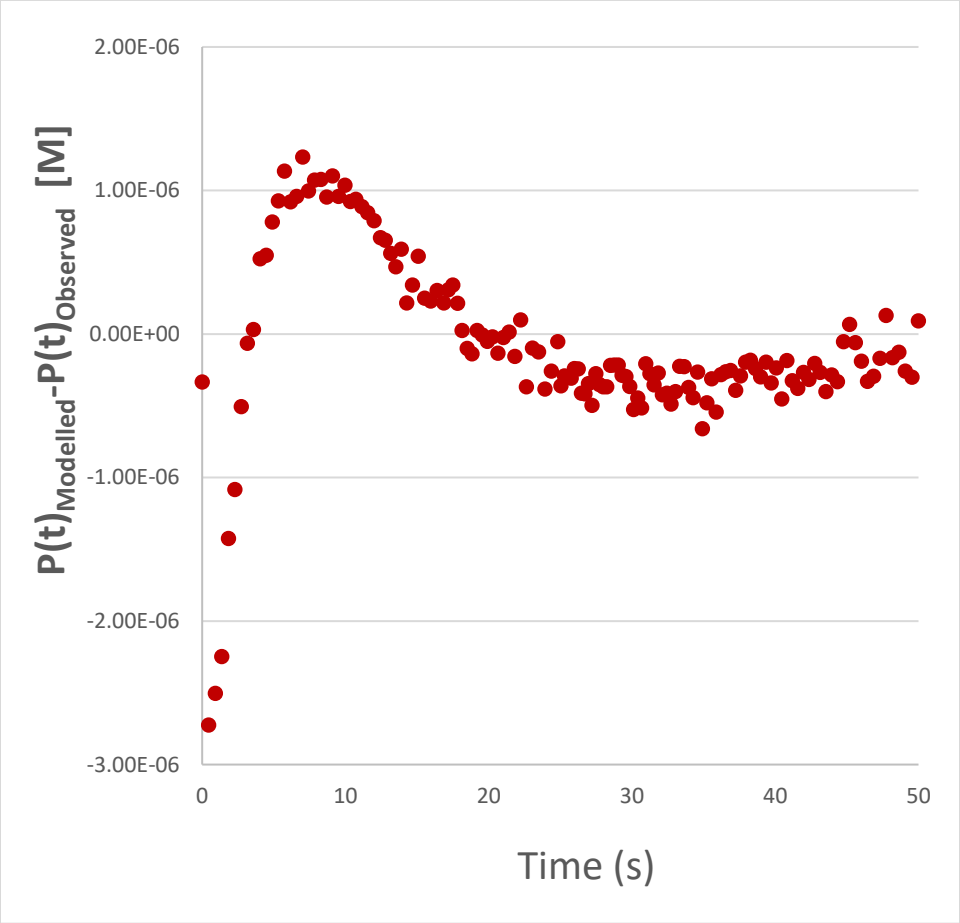


Figure 4-11 Model residuals for the nonlinear regression fit of [4-2] to gTGase hydrolysis data from Gravel et al.[16] The initial acylating burst reported from 0-0.5 s was not accounted for by **[4-2]**.

As shown in **Figure 4-11**, the model does not incorporate the observed initial burst-phase kinetics. After 3.5 s, the model residuals indicate an overestimate in production that compensates for underestimated production rate later in the assay. This shows similarities to some of the **3** + TE assay combinations assessed in **Figure 4-9**.

In the case of gTGase, if the substrate inhibition has slow-off reversible inhibition effect rather than the simplified irreversible **[4-2]**,[16] then a higher order deceleration of reaction rate could be incorporated in a higher order model of inactivation. However the nonlinear regression fit of **[4-2]** 10% of the observed value after 2 s; the extracted rate constants of this parsimonious model may

offer useful quantitative measures of the relative rates of inactivation and hydrolytic turnover of different carbonate and thiocarbonate substrates.

4.2.2.3 Direct non-linear regression

To test the fit of the constant C, the linear fitting algorithm **[4-3]** was compared against nonlinear least squares regression of the non-linearized reaction data using both Excel's Solver rootfinder and GraphPad Prism's nonlinear regression suite.

Both nonlinear methods converged on similar k_i and k_p values for the BaeRTE and BaeACPTE protein constructs reacted with SNAC substrate **3**:

Table 4-2 Nonlinear fitted kinetic constants generated from Solver for hydrolysis of SNAC 3.

	k_p ($M^{-1}s^{-1}$) (% diff from linear **[4-3]** fit)	k_i ($M^{-1}s^{-1}$) (% diff from linear **[4-3]** fit)	C/E_0 (% diff from linear **[4-3]** fit)
BaeRTE	16.37 (-14 %)	1.409 (-17 %)	12.64 (+4 %)
BaeACPTE	2.688 (-7.3 %)	0.9753 (+21 %)	3.76 (-15 %)

Table 4-3 Nonlinear fitted kinetic constants generated from Solver for hydrolysis of 1, 2

Protein	**Constant**	**Substrate**	
		Conjugated	Unconjugated
BaeRTE	k_I ($M^{-1}s^{-1}$)	$4.7 * 10^{-2}$ (-85 %)	0.08 (-90 %)
	k_P ($M^{-1}s^{-1}$)	0.60 (-30%)	0.89 (-26 %)
BaeACPTE (*holo*)	k_I ($M^{-1}s^{-1}$)	3.6 (+110 %)	0.85 (112 %)
	k_P ($M^{-1}s^{-1}$)	$8.7 * 10^{-2}$ (-56 %)	1.6 (+58 %)

For both unsaturated substrates (**1, 2**), nonlinear regression generated constants of the same order of magnitude (**Table 4-3**), but each did not adequately account for a significant portion of the data variance. As a result, none of the resulting kinetic constant pairs (k_i and k_p) served as accurate

quantitative measures of substrate specificity during the TE assays for unsaturated SNAC compounds **1** and **2**, and use of non-linear fitting did not rectify the lack of fit.

Taken together, 4.2.2.2 and 4.2.2.3 suggest that both linear and nonlinear regression fits of **[4-2]** can fit the progression curves for both SNAC substrate **3** with BaeRTE and BaeACPTE as well as a literature report of substrate inhibition,[16] but that the model fails to capture the behaviour of **1** or **2** reacted with BaeTE or BaeACPTE.

4.3 Discussion

The rate law for the rate of enzyme inactivation proposed in this chapter is proportional to both the concentration of the TE enzyme and the concentration of the SNAC substrate. The reaction progression is defined as the consumption of SNAC by the inactivation process, the decrease in active enzyme, and the production of inactivated enzyme. In the Ellman's colorometric assay, detection of inactive enzyme or decrease in substrate specifically from an enzyme-inhibition interaction is not possible. The concentration of remaining active enzyme, while not directly measurable, may be inferred through the local rate of total thiol production. As a result, the partial order of the inactivation reaction with respect to enzyme could be empirically screened. Reaction conditions with SNAC substrate **3** in 500-5000 fold excess of BaeRTE (keep in mind productive reaction diminishes this excess over the course of the reaction) shown in **Figures 4-1** - **4-3** demonstrate that enzyme inactivation fit a first order decay with respect to enzyme concentration, and that the pseudo rate constant was a function of substrate concentration.

The converse analysis of holding the initial substrate concentration low and constant between trials and using a great excess of BaeRTE that is varied between trials was nonviable. At nM concentrations of SNAC, the assay is not sensitive enough to detect hydrolysis. At M concentrations of the enzyme, the limited solubility of the protein and the challenges associated with purifying such as significant amount of protein limit the assay. As a result, the partial order of the reaction with respect to substrate concentration could not be determined in the same manner.

However, **Figure 4-1** and **4-2** demonstrate that both the intial rate and the decline in rate of thiol production are larger in assays with larger concentrations of SNAC **3**. In this example, the substrate **3** concentration did not decrease substantially, and as a result, any effect of substrate concentration on the relationship between enzyme concentration and rate of thiol release is cancelled out in the form seen in **Figure 4-2**. The only effect intial substrate concentration should exert should be on the rate of enzyme inactivation itself. As a result, for instances of low substrate consumption, the analysis of substrate impact on the enzyme activity change does not require proper reaction model selection (options seen in **Figure 4-4**). However, all possible partial orders with respect to substrate concentration will appear as linear functions in graphs of the form of **Figure 4-2**. This includes the possibility of uncompetative binding of a second molecule of substrate to the catalytic [ES] complex. In the case of BaeTE and **3**, the true partial order with respect to substrate appears to be between 0 and 1. Increasing the initial substrate concentration by 7.5-fold (from 0.5 mM to 3.75 mM) increased the rate of enzyme inactivation by 4.53-fold, whereas smaller increases in substrate concentration resulted in larger fractional increases (1.73-fold increase in activity for a two-fold increase in concentration). This may reflect a partially saturated k_i' term (page 4.9), where the half-maximal inhibition constant $K_{Inhibition}$ of 4.53 mM might account for the changing response with respect to substrate.

Recall that the k_i term used in fitting in this chapter is proposed as a second-order rate constant for the insaturable "productive" inactivation of the TE enzyme. Similarly, k_p signifies a second-order rate constant for the conversion of SNAC thioester into free thiol. Both are equivalent to k_{cat}/K_M in the equivalent Michaelis-Menten saturable rate equation under the special case that $[S]<K_M$. The magnitudes of k_p reported in **Table 4-1** fall within the wide range 0.01-1000 s^{-1} M_{TE}^{-1} reported for isolated *cis*-AT Type I TEs[12, 17-20], but the significant change in activity over time in the BaeTE assays indicate substantial deviation from steady-state enzyme turnover. In two cases of reported *in vitro* substrate inhibition of TE domains, the authors reported initial reaction velocities without progression curve data. Importantly, in both cases, the initial rate of reaction reached a maximum at an intermediate concentration and decreases at higher substrate concentrations.[12, 14] In both cases, the behaviour was fit to a reversible allosteric binding model of substrate

inhibition, and $K_I > K_M$. In the example with Radicicol TE, K_M = 0.14 ± 0.03 mM and Ki = 4.5 ± 0.9 mM. Neither group proposes elements of time-dependent slow-binding or irreversible inhibition. However, Buntin *et al.* qualify their analysis with the statement "The rates of reaction were then calculated from the initial linear portions of the curves…",[12] and imply that some of the 10 min observation windows recorded a change in rate of product release.

In contrast, data from the piercicidin TE study was reported as a time-domain progress curve at a single substrate concentration used. After 3-16 turnovers and 20 minutes, the observed rates dropped by 75%. The fit of equation [**4-2**] matched the observed trends for piercicidin TE, and the derived k_I and k_p values of 1-2.5 and 5-35 $M^{-1}s^{-1}$ were similar to those found for BaeTE in this chapter.

For the kinetic assays reported here, all increases in initial substrate concentration corresponded to increases in the instantaneous rate of reaction and does not match the lower activity levels seen at high substrate concentrations with unnatural substrates with isolated Radicicol TE[14] and Disorazole and DebsTE[12]. Additionally, the BaeTE SNAC assay data cannot be presented as a function of initial rate of reaction because of the time-dependent variation within the first 10% of substrate consumption.

The fitted model parameters k_p and k_i give quantitative measures of the progression of hydrolysis and the deceleration of reaction rate for the saturated substrate SNAC **3**. The BaeACPTE construct hydrolyzed **3** (k_i+k_p) at an initial rate 17.9% of that of the BaeTE construct, and the specific rate of inactivation of the BaeACPTE construct was 47% that of BaeTE. However, these comparisons cannot be extended to **1** and **2**.

For **1** and **2**, the observed assays at 1-3 mM SNAC substrate exhibited substantially less hydrolysis than that of the 5 mM samples, and as compared to the distribution of rates between concentrations displayed in **3** and BaeTE. This represents the largest deviation from the modelled data, which predicts the initial rate to be linearly proportional to substrate concentration. Alternately, for **1** and **2** treated with holoBaeACPTE, the reaction progression appears significantly biphasic, with a large increase in free thiol within the first 50-70 s, and a second, much lower rate

of thiol release step. Fitting of [**4-2**] assumes that substantial activity decay (large k_i) leads to rapid suppression of all non-background hydrolysis, and does not appropriately capture the residual activity observed after this burst of 0.5 equivalents. One possible complicating factor in the holoBaeACPTE constructs is the acylation of the ACP domain before steady-state turnover by the TE domain. In order to capture the full properties in this case, a second model term with a slow-on ACP binding site would be required.

While the model does not allow for direct comparison between **1**, **2**, and **3**, it does effectively simulate the time-dependent substrate inhibition-response of both Piercicidin TE and gTGase.

4.4 Conclusions

Initial examination of experimental conditions with SNAC thioester substrate **3** demonstrated that initial thioester substrate concentration correlated with the rate of decline of the hydrolysis rate. The correlation fit a function with a partial first order with respect to substrate concentration. **[4-2]** was derived from a simplified reaction/inhibition model driven by the pseudo first-order constants of reaction (k_p) and enzyme inhibition (k_I).

Chapter 5 describes steps taken to characterize transAT TE domains isolated from related bacterial PKS pathways with respect to protein fold and enzyme activity.

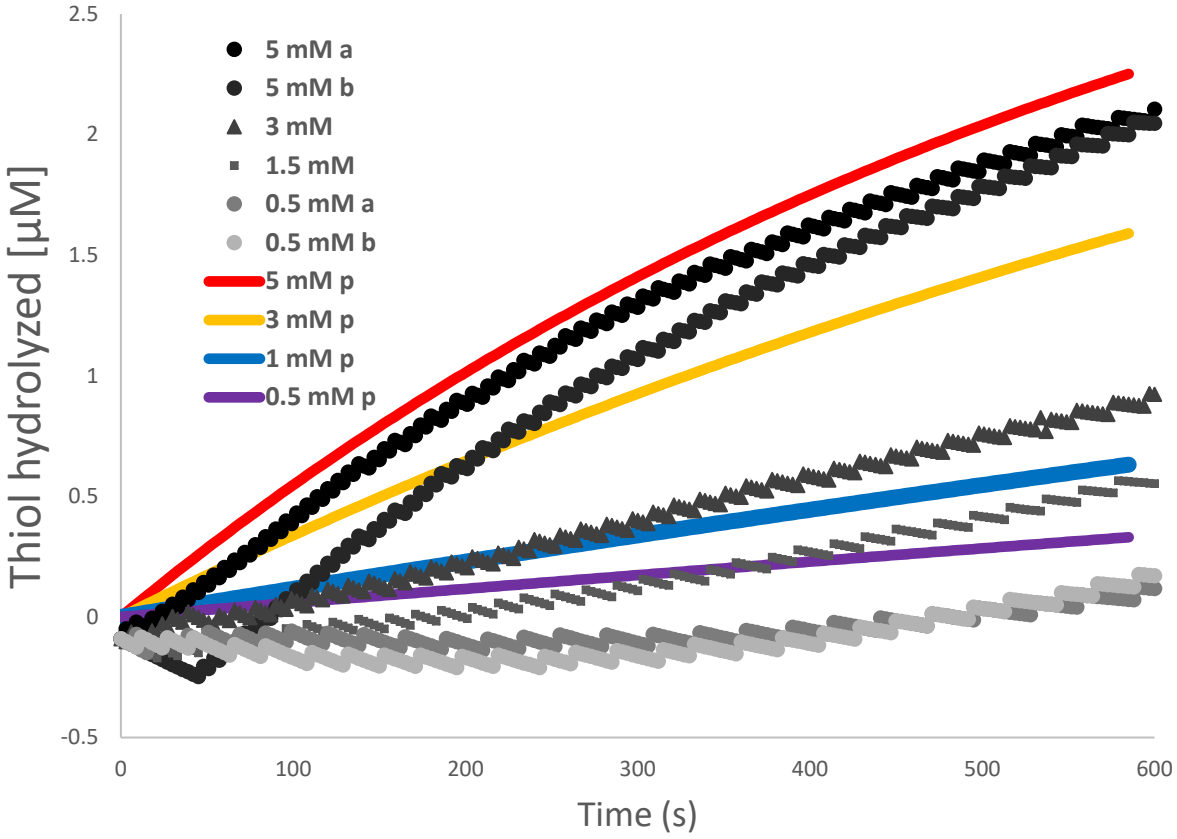


Figure 4-12 Modelled reaction of conjugated SNAC substrate 1 with BaeRTE. With **[4-2]** set with the fitted constants k_I = 0.3 $M^{-1}s^{-1}$ and k_P = 0.9 $M^{-1}s^{-1}$, the model (coloured traces) overestimates the reaction rate for the lower reaction rates.

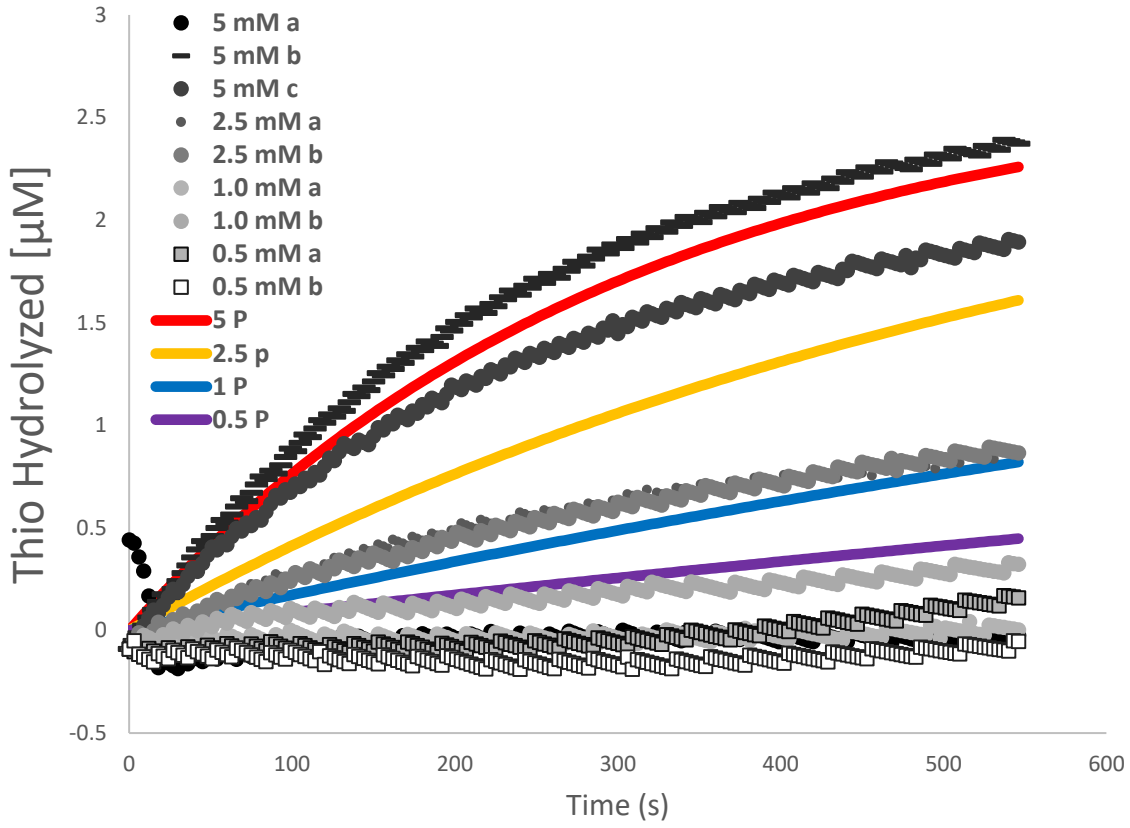


Figure 4-13 Modelled reaction of unconjugated SNAC substrate 2 with BaeRTE. With **[4-2]** set with the fitted constants $k_I = 0.8\ M^{-1}s^{-1}$ and $k_P = 1.2\ M^{-1}s^{-1}$, the model (coloured traces) overestimates the reaction rate for the lower reaction rates. Note: reaction 5 mM a (black circle) did not display hydrolysis above the level of substrate alone + enzyme/Ellman's indicator controls and was excluded from the fitting.

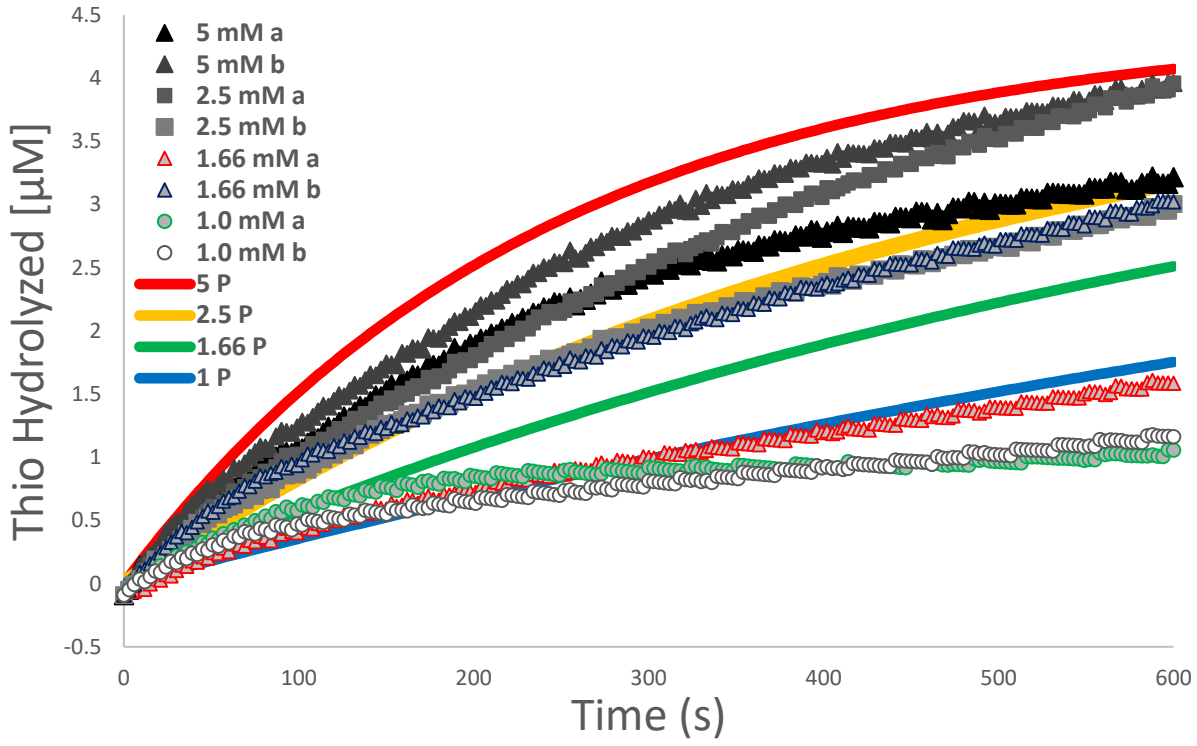


Figure 4-14 Modelled reaction of saturated SNAC substrate 3 with BaeACPTE. [4-2] was fit with the constants k_I = 0.8 $M^{-1}s^{-1}$ and k_P = 2.9 $M^{-1}s^{-1}$; here, the function (coloured traces) reflects the general deceleration of hydrolysis, but significant variations in hydrolysis between replicates complicate the relationship between hydrolysis and initial substrate concentration. The observed progression curves at 2.5 mM (grey squares) overlap with reaction curves at both 1.66 and 5 mM **3**.

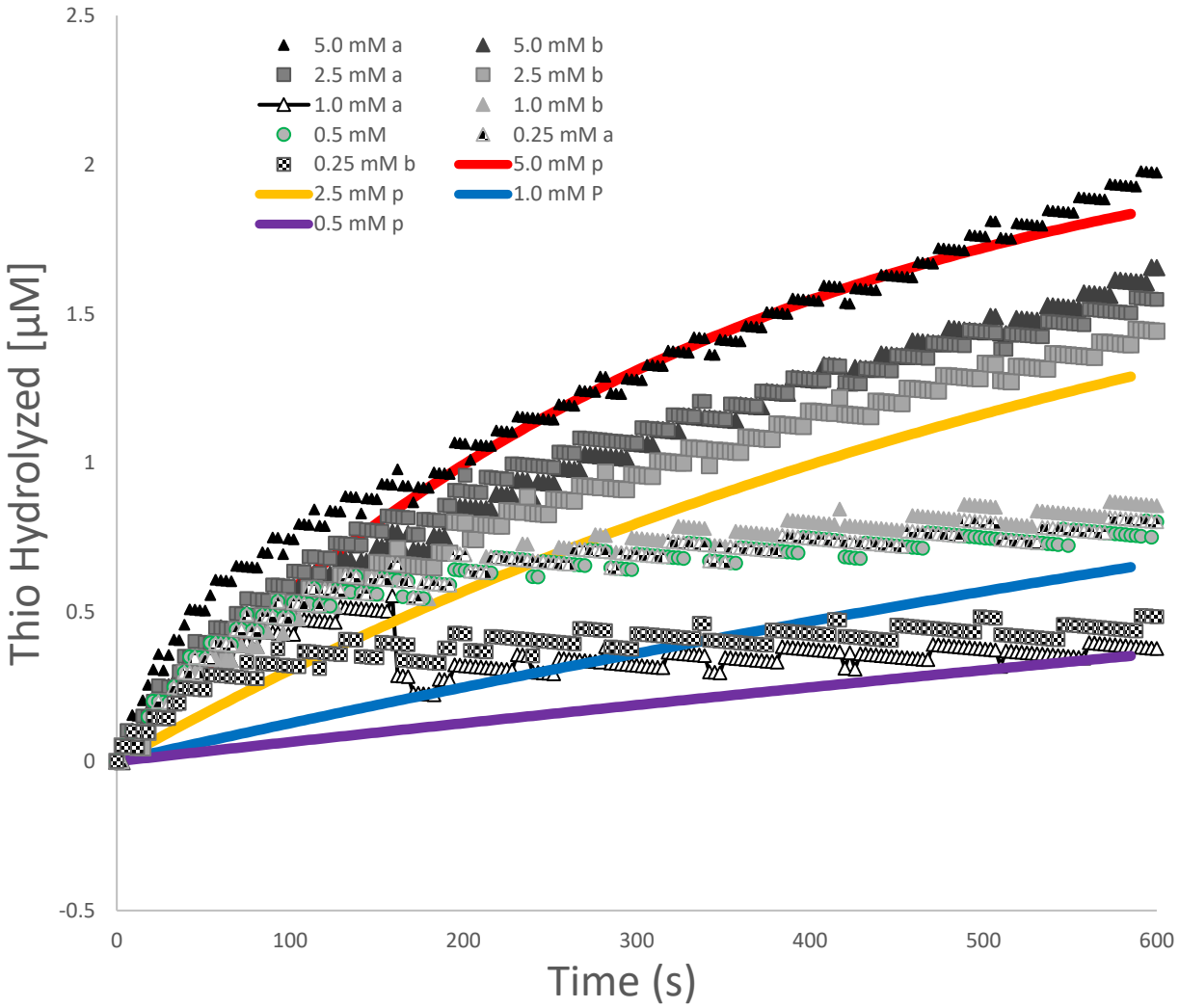


Figure 4-15 Modelled reaction of conjugated SNAC substrate 1 with BaeACPTE. [4-2] was fit with the constants set at $k_I = 0.3\ M^{-1}s^{-1}$ and $k_P = 0.2\ M^{-1}s^{-1}$; this reaction set is the only case where the fit k_I is greater than the associated k_P. However, the model fails to capture the rate decrease at around 80 s and almost complete turn to background hydrolysis at lower concentrations.

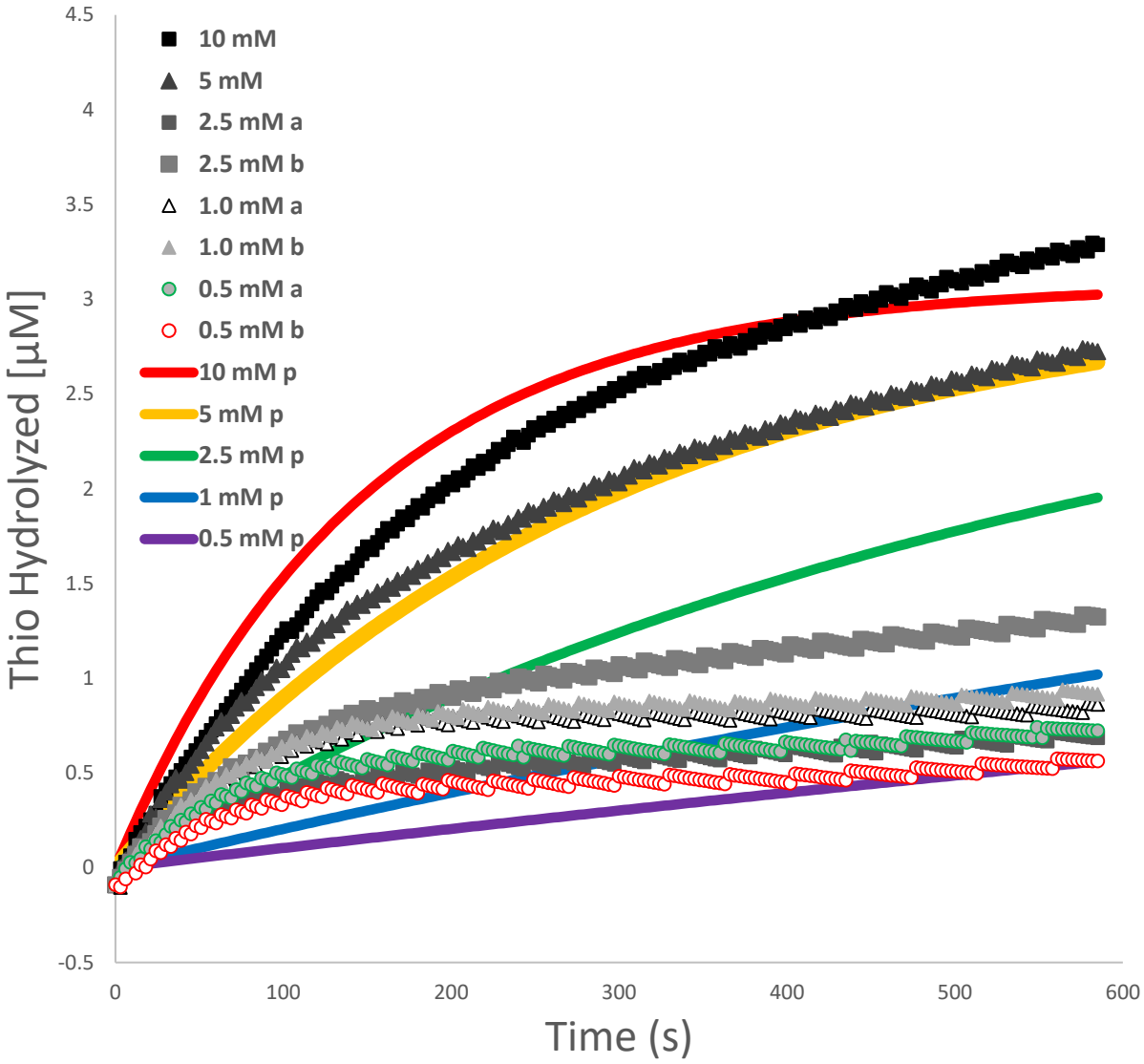


Figure 4-16 Modelled reaction of unconjugated SNAC substrate 2 with BaeACPTE. With **[4-2]** set with the fitted constants $k_I = 0.4\ M^{-1}s^{-1}$ and $k_P = 1.0\ M^{-1}s^{-1}$, the model (coloured traces) overestimates the reaction rate for the lower reaction rates.

4.5 References

1. M. Holmquist, *Current Protein & Peptide Science*, 2000, **1**, 209-235.
2. B. S. Hartley and B. A. Kilby, *Biochemical Journal*, 1954, **56**, 288-297.
3. J. Li, R. Szittner, Z. S. Derewenda and E. A. Meighen, *Biochemistry*, 1996, **35**, 9967-9973.
4. Y. Jiang, K. L. Morley, J. D. Schrag and R. J. Kazlauskas, *Chembiochem : a European journal of chemical biology*, 2011, **12**, 768-776.
5. P. Heikinheimo, A. Goldman, C. Jeffries and D. L. Ollis, *Structure (London, England : 1993)*, 1999, **7**, R141-R146.
6. A. Rauwerdink and R. J. Kazlauskas, *ACS Catalysis*, 2015, **5**, 6153-6176.
7. D. Fegley, S. Gaetani, A. Duranti, A. Tontini, M. Mor, G. Tarzia and D. Piomelli, *Journal of Pharmacology and Experimental Therapeutics*, 2005, **313**, 352-358.
8. M. J. Evans and B. F. Cravatt, *Chemical Reviews*, 2006, **106**, 3279-3301.
9. K. Ahn, D. S. Johnson, M. Mileni, D. Beidler, J. Z. Long, M. K. McKinney, E. Weerapana, N. Sadagopan, M. Liimatta, S. E. Smith, S. Lazerwith, C. Stiff, S. Kamtekar, K. Bhattacharya, Y. H. Zhang, S. Swaney, K. Van Becelaere, R. C. Stevens and B. F. Cravatt, *Chemistry & Biology*, 2009, **16**, 411-420.
10. W. X. Tian and C. L. Tsou, *Biochemistry*, 1982, **21**, 1028-1032.
11. J. G. Bieth, *Proteolytic Enzymes: Aspartic and Metallo Peptidases*, 1995, **248**, 59-84.
12. K. Buntin, K. J. Weissman and R. Muller, *Chembiochem : a European journal of chemical biology*, 2010, **11**, 1137-1146.
13. J. B. Scaglione, D. L. Akey, R. Sullivan, J. D. Kittendorf, C. M. Rath, E. S. Kim, J. L. Smith and D. H. Sherman, *Angewandte Chemie - International Edition*, 2010, **49**, 5726-5730.
14. G. W. Heberlig, M. Wirz, M. Wang and C. N. Boddy, *Organic Letters*, 2014, **16**, 5858-5861.
15. Q. Liu, F. Yao, Y. H. Chooi, Q. J. Kang, W. Xu, Y. R. Li, Y. C. Shao, Y. F. Shi, Z. X. Deng, Y. Tang and D. L. You, *Chemistry & Biology*, 2012, **19**, 243-253.
16. C. Gravel, D. Lapierre, J. Labelle and J. W. Keillor, *Canadian Journal of Chemistry-Revue Canadienne De Chimie*, 2007, **85**, 164-174.
17. H. X. Lu, S. C. Tsai, C. Khosla and D. E. Cane, *Biochemistry*, 2002, **41**, 12590-12597.
18. C. N. Boddy, T. L. Schneider, K. Hotta, C. T. Walsh and C. Khosla, *Journal of the American Chemical Society*, 2003, **125**, 3428-3429.
19. K. K. Sharma and C. N. Boddy, *Bioorganic & Medicinal Chemistry Letters*, 2007, **17**, 3034-3037.
20. A. A. Koch, D. A. Hansen, V. V. Shende, L. R. Furan, K. N. Houk, G. Jimenez-Oses and D. H. Sherman, *Journal of the American Chemical Society*, 2017, **139**, 13456-13465.

Chapter 5

5.1 Introduction

There is strong evidence from *in vivo* work that the BaeTE is active in the native pathway and essential for product release. *In vitro* characterization of BaeTE with SNAC-activated substrates, presented in Chapter 3, showed substrate-based inhibition of BaeTE activity. The goal of the work presented in this chapter is to determine if having ACP-bound substrates can overcome the substrate-based inhibitions of the BaeTE seen with SNAC-activated substrates, recapitulating the *in vivo* hydrolytic activity observed in the native pathway.[1]

In this chapter, I subject intact ACP species to LC-ESI-MS analysis to determine if the BaeRTE construct effectively releases simple model acyl chains from ACPs. The ACP interaction partners include the native partner BaeR module ACPIII (BaeRACPII) found at the N-terminus of BaeRTE, as well as the unnatural substrate carrier BaeLACPIII from the middle of the PKS pathway which is physically shielded from interactions with the TE *in vivo*.

In a broader sense, analysis of covalently bound substrates is a significant challenge in metabolism. Metabolites within the cell exist either as free solutes or as sequestered intermediates. Determining the concentration of the possible metabolites in a pathway as well as the selectivity and kinetics of the converting enzymes is critical to understanding any metabolic pathway. Global metabolomic mass-profiling tools can readily identify the 80-100 most abundant cellular metabolites,[2-4] but these cellular extractions-based approaches are unable "to differentiate free and macromolecule-bound metabolites, as both could be released via organic extraction."[4] As a result, measuring the metabolic flux in macromolecule-bound biosynthetic pathways such as type I polyketide synthesis is particularly challenging. Several assays have been developed to study the conformation, modification, and transfer of acyl substrates onto and off of the post-translationally modified ACP active-site serine and an LCMS-based approach enabled limited kinetic details on the acylation state of ACPs during polyketide biosynthesis. However no "gold standard" approach currently exists for characterizing acylation states of ACP in polyketide biosynthesis.

5.1.1 The structural role of an acyl carrier protein

Type I polyketide function is anchored by ACPs. The growing polyketide product is covalently linked through a thioester bond to the extendable phosphopantetheinyl arm of the ACP thioester during PKS biosynthesis. This covalent intramolecular link increases the effective concentration of substrate and the PKS domains, thereby shifting the "effective specificity constant" relative to other compounds in free solution, and shuttles the metabolite through rapid sequential biosynthesis.

As described in greater detail in the sections that follow, the ACP is a 10-13 kDa four helix domain with a Ser residue positioned at the beginning (N-terminus) of its second α-helix. The peptide sequence in the immediate vicinity of this Ser is recognized by the PKS family of phosphopantetheinyl transferases (PPTases), and the ACP is posttranslationally modified with a Ppant group from coenzyme A (CoA). The *holo* form of the ACP (with Ppant) is acylated by AT domains *in cis* and *in trans*. The acyl-Ppant arm can adopt several conformations. For example while the buried conformation, the acyl-Ppant arm is sandwiched between helices of the ACP, which stabilizes chemically reactive functional groups on the growing polyketide; in the extended configuration, the ACP presents the acyl chain from the acyl-Ppant arm into the active site of nearby PKS domains.

Typical ACPs are acylated with 1,3-dicarbonyl derivatives, such as malonyl or methylmalonyl, that condense with the growing polyketide intermediate metabolite on an upstream KS domain through a decarboxylative Claisen-condensation. The ACP-bound condensation product is then acted on by modifying domains within the module before being released from the ACP through thioester exchange to the downstream KS, premature hydrolysis through error-checking editing domains, or final chain release through domains such as the TE.

For modular Type I PKSs, the ACP proximal to the TE is loaded via an upstream module, and remains acylated until the TE undertakes the Loading Stage. As with the Ellman's assay, detection of Ppant occupancy reports on the rate of loading of the acyl group onto the TE Ser during the

loading stage through study of the ACP leaving group, and the steady state rate of loading is limited by the reaction's rate determining step.

5.1.2 **ACP post-translational modification by PKS PPTase**

While ACP domains are expressed ribosomally within a PKS, they require the presence of another gene-product to transfer the post-translational Ppant coenzyme from CoA to the ACP active-site serine. As with FAS and NRPS, this is carried out by a PPTase, whose coding region is typically located elsewhere in the genome.

For polyketide-expression bacteria, the genome contains both a FAS- and a PKS-specific PPTase. FAS PPTases are typically called AcpS and in general do not recognize PKS ACPs as substrates. A PKS PPTase labels multiple ACPs in a PKS pathway, and can often label ACPs from multiple native or exogenous pathways simultaneously. Cross-compatibility between PKS and FAS PPTases and their respective ACPs is limited.[5]

As a result, PKS ACPs expressed in *E. coli* require the coexpression of a tolerant PKS PPTase for installation of the Ppant coenzyme and full function *in vivo*.[5] The PPTase from *Bacillus subtilis*, surfactin phosphopantetheinyltransferase (Sfp) has been shown to be extremely effective *in vivo* and *in vitro* for posttranslational modification of PKS and NRPS ACPs. PKS ACPs can be expressed in a homogenous apo-form in *E. coli* strains lacking an Sfp-type PPTase. Because PKS PPTases, like Sfp, are insensitive to the chemical modification of the cysteamine portion of CoA, the PPTase can also serve to install one equivalent of unnatural acyl substrate[6] onto an isolated apo-ACP *in vitro* with verifiable purity and efficiency.

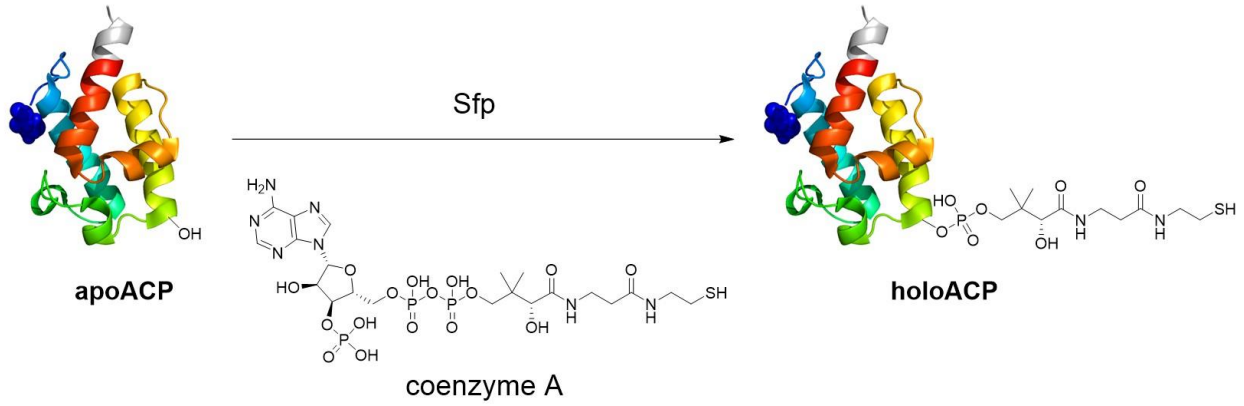


Figure 5-1 Sfp phosphopantetheinyl transfer from CoA to ACP. Sfp transfers Ppant to the Ser positioned at the N-terminus of helix two (green to yellow). The core ACP consists of four α-helices (N-terminus blue to C-terminus in red). ACP structure based on *holo*- curacin ACPI (PDB, 2LIU)[7] trimmed to the design sequence of the ACPs used in this study (see **Figure 5-3** for details). The blue spheres represent the position of the N-terminal his_6 tag, and the red to white transition represents the C-terminal cut-off between the ACP and TE domains.

Synthetic access to acyl-CoAs used in these *in vitro* labelling experiments commonly follows electrophilic activation of the acyl-acid substrate, and nucleophilic addition of CoASH.[8] Here, use of HPLC (or thiol-affinity resins) is essential to remove residual CoASH and imidazole from the acyl-CoA product.

5.1.3 Monitoring an ACP loading-state

Once loaded, monitoring the acylation state of ACPs in PKS pathways is essential for understanding the kinetics of PKS biosynthesis.[9] Characterization of which species are found on the ACP and for what fraction of total ACP occupancy inform on the relative speeds of AT loading, KS condensation, and reductive loop modifications relative to the rate of product off-loading (and premature offloading of intermediates in engineered pathways).[9] In the biosynthesis of bacillaene, the terminal double bond is isomerized by an enoyl-isomerase isolated by condensationally-inactive ketosynthase domains before release by BaeRTE. In this case, the biochemical question remaining is whether this arrangement with the enoyl-isomerase and BaeRTE acting on different, sequential ACP substrates is required in order for the isomerisation to occur without kinetic competition from TE-mediated hydrolytic release. The work in this chapter focuses on the prerequisites of further

BaeRTE characterization; does the BaeRTE construct catalytically hydrolyze ACP-bound acyl substrates and is there selectivity for specific ACP or acyl chain substituents?

Multiple methods have been developed to report on the occupancy of ACP Ppant arms; NMR,[10, 11] X-ray crystallography,[12] ^{3}H, ^{14}C, and ^{32}P radiolabelling,[13] and small molecule and protein mass-spectroscopy[14] (including the Ppant-ejection assay) have been used (primarily discontinuously) to monitor species on the ACP active-site serine. Between 2003 and 2017, variants of LCMS have been the most abundantly used *in vitro* assay tool.[15-20]

Since the thioester linkage is stable to site-specific protease digestion, researchers have been able to digest intact PKS modules and isolate 7-20 amino acid helix II fragments labelled with synthetic intermediates and directly characterize through MS. In some cases, larger mass fragments from the remainder of the module were indirectly observed as multiply charged *m/z* species.[18] A refinement to this assay was the application of multiple reaction monitoring to intact modules.[16] For example, application of infrared multiphoton disassociation (IRMPD) to the 126 kDa GrsA produced detectable phenylalynyl-pantetheine (408.198 Da). Additionally, smaller, stand-alone ACP domains have been studied without fragmentation (**Figure 5-2**).[14, 17, 20]

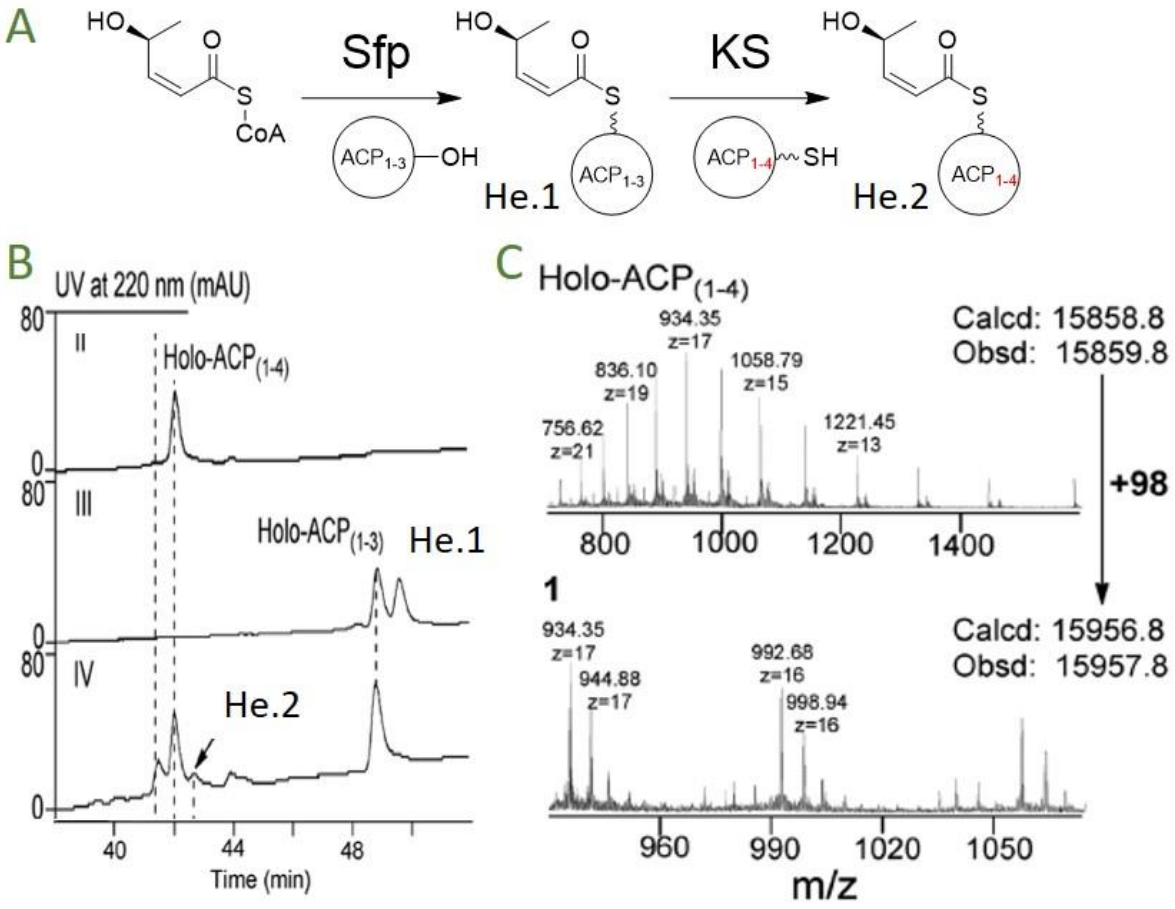


Figure 5-2 Literature examples of intact ACPs detected by LCMS. A-C Study of transfer of dehydrated acyl intermediates in FR901464 biosynthesis. (**A**) ApoACP$_{1-3}$ is acyl-pantetheinylated to He.1 via SFP and deacylated via the KS. *holo*-ACP$_{1-4}$ is transacylated from He.1 via the activity of the KS. (**B**) The individual protein species were resolvable by HPLC. (**C**) The mass spectrum at the corresponding elution times show a characteristic charge envelope. The acylation of ACP$_{1-4}$ can be tracked by a slight shift in retention time (**He.2**) and the appearance of a second, shifted charge envelope corresponding to a deconvoluted mass change of +98. Figure reprinted (adapted) with permission from He *et al.* JACS 2014 DOI 10.1021/ja500942y.

In this chapter the ACP acylation state is primarily characterized by HPLC separation of the ACP species from *in vitro* biochemical assays, followed by detection via UV absorption and ESI-LCMS.

5.2 Results

To test if BaeRTE hydrolyses acyl-ACP substrates without selectivity or substrate inhibition, the work flow was broken up into four component steps. First, a range of ACP monodomains were designed for heterologous overexpression in *E. coli*. Second, CoA substrates were synthesized and purified for enzymatic labelling. Third, LCMS detection of ACP substrates was developed. Finally, in the fourth section, assay conditions for Sfp and BaeRTE labelling and hydrolysis were screened. In the detailed results that follow, each element of this method is described in greater detail.

As the first step for ACP-tethered assays, gene constructs for a range of *B. amy*. FZB42 ACPs were designed. The boundaries of the ACP regions of the bacillaene pathway were delineated through secondary structure prediction (antiSMASH)[21] and confirmed through sequence alignment of bacillaene ACPs against with known, structurally characterized ACPs. **Figure 5-3** shows both the alignment of the primary sequence and secondary structures of the ACPs.

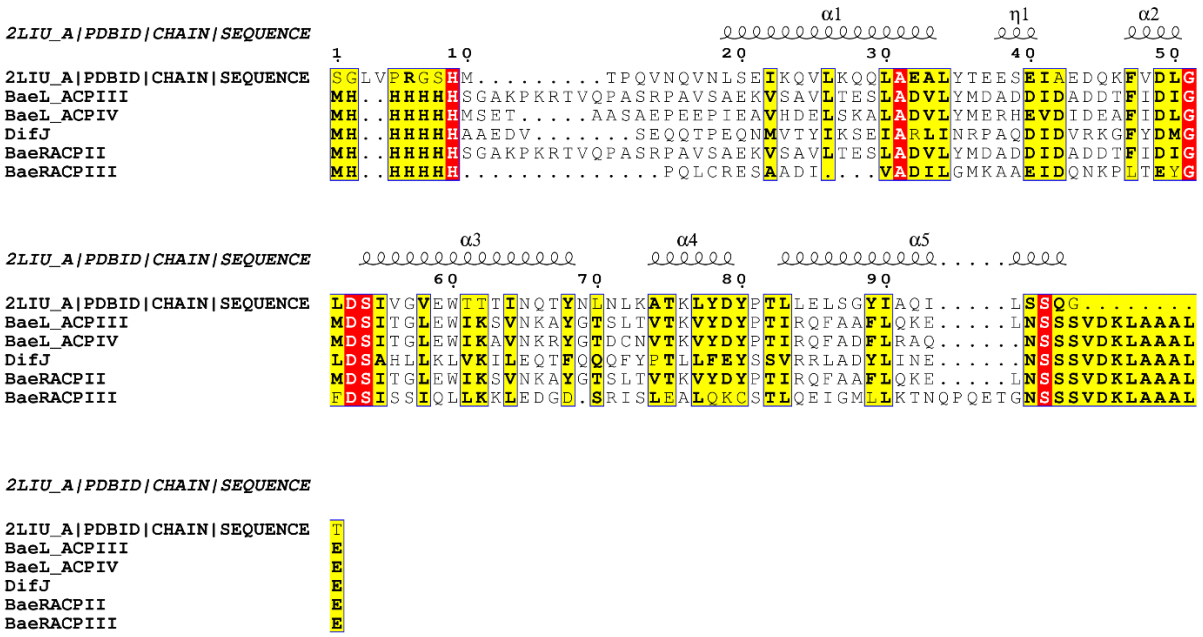
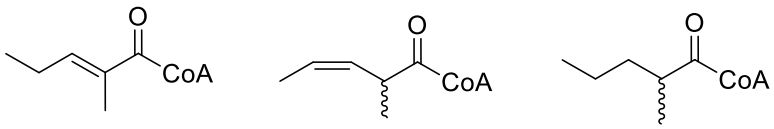


Figure 5-3 Comparison of designed ACP monodomain sequences. The core structure of four ACP domains from the bacillaene pathway (BaeL and BaeR) and one ACP from the difficidin PKS pathway (DifJ) from *B. amyloliquefaciens* FZB42 were predicted. The conserved active-site region including target Ser are highlighted at positions 51-54 based on the reference curacin ACP sequence notation (2LIU, PDB).[7] Positions 1-9 show engineered Met start codon and 6xHis tag. The core ACP sequences range from 20-25% identity and 45-55% similarity to the TE proximal BaeRACPIII. Alignment run through T-COFFEE[22] and secondary structure appended by Espript.[23]

Four of the His tagged ACP constructs were designed as isolated domains of the bacillaene pathway and the fifth was from the difficidin biosynthetic pathway. BaeRACPIII (from pMEH16) is the final ACP found on the terminal BaeR module N-terminal to the BaeRTE. In the native organism this ACP is predicted to transfer the completed polyketide chain to the TE domain prior to cleavage. BaeRACPII is further upstream in of the BaeRACPIII and likely plays a role in the loading of BaeRACPIII. BaeLACPIII and BaeLACPIV are the third and fourth of five ACPs found in the PKS protein BaeL and are not expected to natively interact with a TE. Gene sequences encoding these proteins were optimized for codon usage in *E. coli* and the genes were constructed *de novo* (Integrated DNA Technologies). Constructs encoding BaeRACPIII and BaeLACPIII were cloned into *E. coli* expression vectors generating the expression plamids pMEH16 and pMEH13

respectively. These two ACPs could be readily expressed as apo-ACPs from *E. coli* BL21 (DE3) and purified under standard conditions.

To enable the loading of acyl chains onto these ACPs, acyl CoA derivatives were required. The CoA-coupled equivalents of saturated 2-methylvaleric, *trans*-2- and *cis*-3-pentenoic-acid previously investigated in Chapter 3 (**Figure 3.3**) as SNAC derivatives, were targeted for synthesis. Compounds **10** and **12** could be readily synthesized from the free acid and reduced CoA using established procedures. Unfortunately **11** was not successfully prepared; however, access to **10** and **12** provided the essential reagents necessary to investigate TE hydrolysis from acyl-ACPs.



10, conjugated **11**, unconjugated **12**, saturated

Figure 5-4 Proposed CoA substrates. Acyl-CoA compounds **10-12** were designed as intermediates compatible with the phosphopantetheinyl transferase activity of Sfp for acylation on target apo-ACPs. **10** and **12** were successfully activated from their carboxylic acid equivalents via CDI and COASH and isolated via HPLC.[8] Similar efforts towards **11** failed to isolate the CoA product.

To generate acyl-ACPs, aliquots of expressed apo-ACPs, which lack the Ppant arm, were treated with recombinant purified Sfp from *Bacillus subtilis*[24] and acyl-CoA derivatives. Due to the substrate tolerance of Sfp for the cysteamine portion of CoA, the acyl-Ppant group could be effectively transferred to active site Ser of the ACPs, generating the acylated *holo*-ACPs. These reagents could thus be tested for the ability of the BaeTE to catalyze hydrolysis of the thioester *in vitro*, generating the *holo*-ACP as well as the free acid of the simple substrate. **Figure 5-5** outlines the stages of the acyl-ACP hydrolysis assay and the predicted origin of the different potential ACP species throughout the assay:

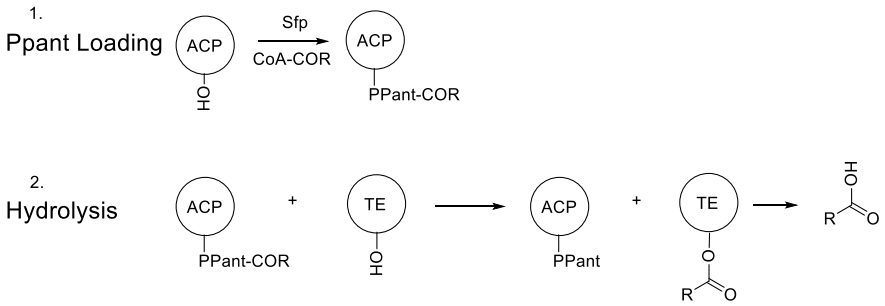


Figure 5-5 Enzymatic loading and acyl hydrolysis of ACPs. The ACP protein is expressed in the apo (Serine OH) form. Ppant loading involves the addition of synthetic acyl-CoA substrate and the purified PPTase Sfp. Hydrolysis of the thioester bond occurs spontaneously at a basal rate in solution, and is accelerated in the presence of an active tolerant TE.

To probe the ACP to evaluate its acylation state, an LC-ESI-MS assay was developed. Literature precedence[17] indicated that the +10 - +14 charge-states of intact ACPs should be detectable via LC-ESI-MS. **Figure 5-6** illustrates that initial samples of malonyl adducts of BaeLACPIII were detected in this multiply-charged fashion.

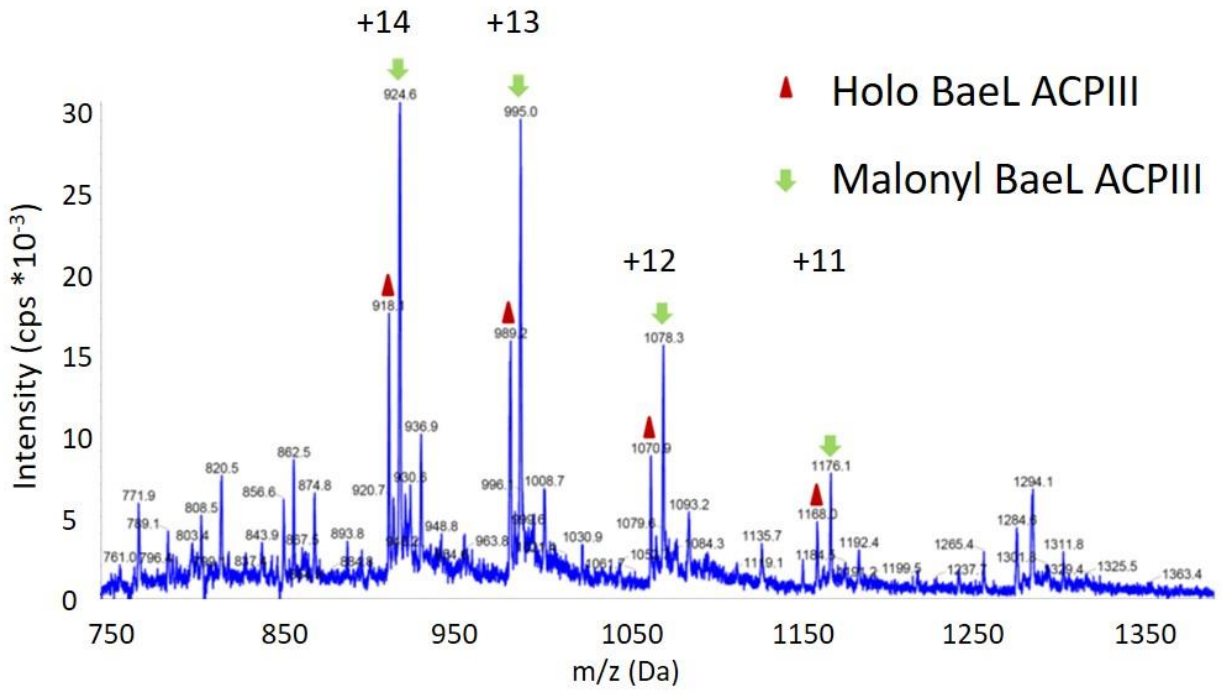


Figure 5-6 Malonyl CoA labelling of apoBaeLACPIII. An aliquot of apoBaeLACPIII was incubated with 10 μM Sfp PPTase and 500 μM malonyl CoA. When fractionated over a Hypersil Gold C8 column, the elution peak between 15 and 17 min displayed a distinctive charge envelope. The spectrum displayed here in baseline subtracted using the 25.8-27.4 min window of the same run. The deconvoluted hypermasses for *holo*-BaeLACPIII (red triangles, exp. 12839 ± 5 Da, calc. 12834.37 Da) and malonyl-BaeLACPIII (green arrows, exp. 12926 ± 4 Da, 12920.38 Da) suggest substantial PPTase labelling of CoA species and hydrolysis of the malonyl cargo.

The *apo*, acyl, and *holo*-ACP species outlined in the assay scheme in **Figure 5-5** could all be detected depending on the treatment regime. **Figure 5-7** shows the apoACP, which is expressed from *E. coli* BL21 (DE3), the *holo*-ACP (labelling with CoA and Sfp), the acyl-ACP (labelling with **12** and Sfp, and mock hydrolysis solution).

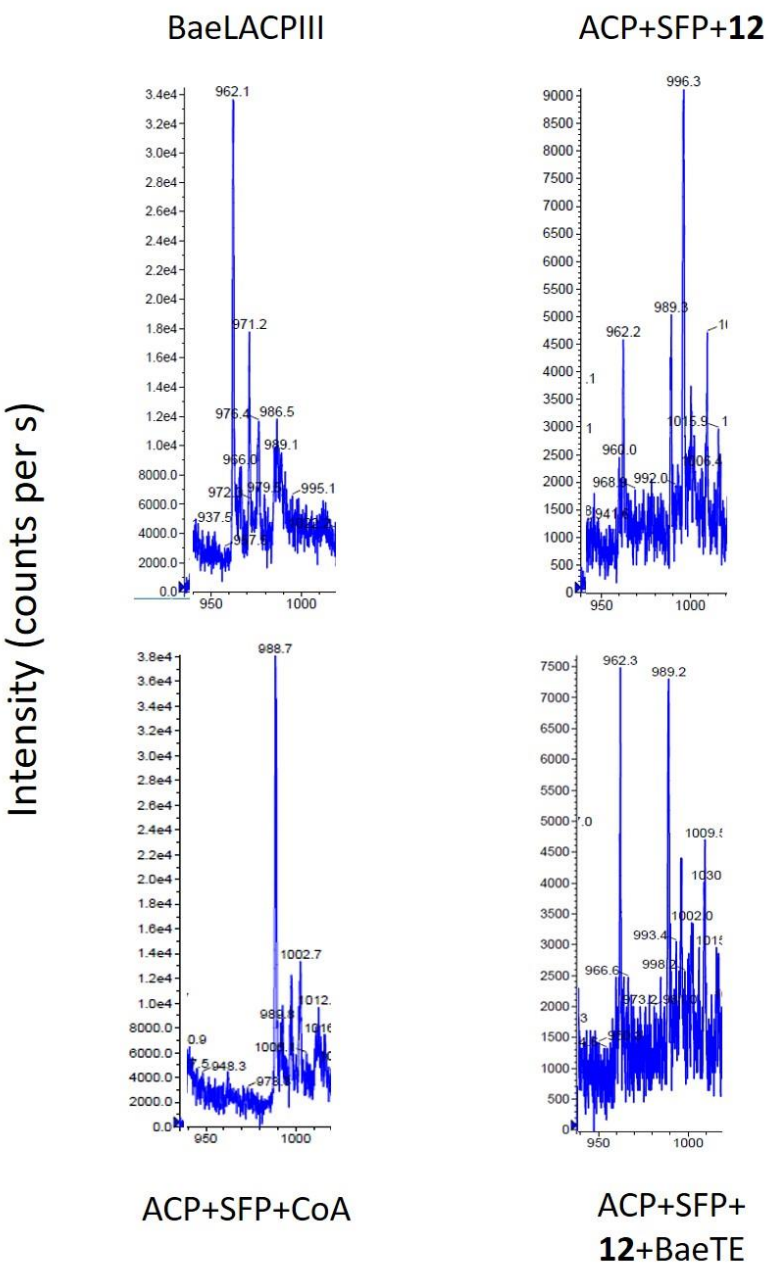


Figure 5-7 Comparison of +13 species region of extracted mass spectra in BaeLACPIII assays with different treatments. The apo (m/z 962), holo (m/z 988), and acylated (**12**, m/z 996) +13 states are present in elutions between 21 and 23 minutes. Treatment with CoA rather than saturated substrate **12** serves as a controlled generation of *holo*-BaeLACPIII. The presence of apoBaeLACPIII in both **12** labelling treatments indicates incomplete labelling of BaeLACPIII. In the presence of BaeRTE, the signal for acylated BaeLACPIII species decreased.

Based on the assumption that the three major peaks in **Figure 5-7** represent ACP species with a difference of mass of Ppant (340 Da) or acylated Ppant (**12**, 439.7 Da), the spacing of the peaks can be used to estimate the charge state (i.e. 996.3-962.1 ≈ 439.7/13). Additionally, the experimental mass of the uncharged species can be calculated using the estimated charge state, and this can be compared against the theoretical mass based on the amino acid sequence of the ACP constructs. This supports the assignment of the peak identity, offers an estimate of experimental precision of mass estimates, and serves as a check for post-translational modification of the ACP. For example, the +13 charge of apoBaeLACPIII (962.1, **Figure 5-7A**) extrapolates to a mass of 12494 Da compared to the theoretical monoisotopic $^{12}C^{14}N$ mass for apoBaeLACPIII ($C_{553}H_{869}N_{149}O_{175}S_3$) of 12493 Da. However, the ESI-quadrupole detection method used in these experiments cannot resolve isotopes within a given charge-state peak, and average value is recorded as a maximum in the mass spectrum. In the LC fractionation method used, the ACP species coelute at approximately the 21st minute of the run. Each of the ACP species resolves as separate peaks in the spectra, but there is variation in each peak centre between LCMS runs. In the case of *holo*-BaeLACPIII (**Figure 5-7**), the peak maximum is estimated as 988.7, but is detected as 989.2 in the equivalent TE-treated trial in panel D. The 0.5 Da trial-to-trial variation is less than 1/5th the difference in masses between species, and does not prevent distinguishing *holo*-ACP signal from acylated CoA species for acyl groups larger than 50 Da, but this does hinder direct estimation of parent mass from a single peak. As shown in **Figure 5-6**, each extracted mass spectrum contains signals from two to five charge states. A quick check for background signals in the assay involves comparing the parent mass derived from each charge state against the average mass-estimate taken from a combination of all peaks. Consistent abundance ratios between each peak in all charge states, and a consistent molar mass estimates support correct ascribing of an ACP title. Finally, the difference in mass estimates between the charge-states acts as an estimator of uncertainty of the calculated neutral mass. This within-sample variance is reported in all data reported in this chapter when more than one charge state is detected. Overall, the spacing between neighbouring peaks as well as matching peaks in

sequential charge states on the spectrum are used to identify the number of charges on the ACP species. The uncertainty of this indirect, semi-quantitative detection method is estimated by the standard deviation of the single-sample neutral mass estimates of each peak in the charge envelope.

The native ACP-pair of BaeRTE, BaeRACPIII was tested for labelling with either conjugated CoA substrate **10** or saturated CoA substrate **11** via Sfp and hydrolysis by the standalone BaeRTE construct. **Table 5.1** outlines the ACP species present in each treatment based on detection of +10-+14 species in the 800-1300 *m/z* range.

Table 5.1 Acylation and hydrolysis of final ACP of the bacillaene pathway BaeRACPIII.

Masses listed are calculated from all components of their observed charge envelope. Standard deviations listed in parentheses are calculated from the variances between individual charge states within a single LCMS run (ESIprot). In the cases where the charge envelope matching to the same parent ion was detected at two different HPLC elution time points, each experimental hypermass is listed individually and separated by a comma. n.d. not detected.

Treatment				**BaeR ACPIII Species Detected**			
ACP	Substrate	SFP	TE	Apo (10984.5 Da)	Conjugated Acyl (11420.5 Da)	Saturated Acyl (11422.5 Da)	Holo (11324.5 Da)
+	**12**	-	-			11427 (N/A)	11324 (6), 11348 (10)
+	**12**	+	-			11430 (1.4)	
+	**12**	+	BaeTE			11430 (2.5)	
+	**12**	+	BaeACPTE			11430 (2.6)	
+	**10**	-	-	11005 (9.7)	11430 (1.6), 11436 (7.9)		11348 (1.2)
+	**10**	+	BaeTE		11429 (1.8)		11348.1 (0.8)
+	**10**	+	BaeACPTE		11434 (14), 11430 (3)		11348 (1.2)

Under the assay incubation conditions, detection of the acylated species in control samples (treatment 1 and 5) lacking Sfp was unanticipated. Similarly, detection of the *holo*-BaeRACPIII with or without BaeRTE renders the possibility of TE-catalyzed hydrolysis inconclusive.

The assay conditions were repeated using the non-cognate BaeLACPIII and the results are summarized in **Table 5.2**.

Table 5.2 Acylation and hydrolysis of mid pathway ACP BaeLACPIII.

Masses listed are calculated from all components of their observed charge envelope. Standard deviations listed in parentheses are calculated from the variances between individual charge states within a single LCMS run (ESIprot).

Treatment				BaeLACPIII Species Detected			
ACP	Substrate	SFP	TE	Apo (12493.28 Da)	Conjugated Acyl (12931.44 Da)	Saturated Acyl (12933.45 Da)	Holo (12834.37 Da)
+	**12**	-	-	12461 (11)			
+	**CoA**	+	-				12838 (1.8)
+	**12**	+	-			12939 (1.4)	
+	**12**	+	BaeTE			12939 (1.4)	12839 (10, faint 2 peaks)
+	**12**	+	BaeACPTE			12939 (1.2)	
+	**10**	-	-	12473 (23)			
+	**10**	+	-			12939 (1.3)	
+	**10**	+	BaeTE	12487 (N/A)	12939 (2.4)		12816 (N/A)
+	**10**	+	BaeACPTE	12481 (9)	12938 (2.2)		

The results of **Table 5.2** follow the principle that Sfp is required to catalyse the non-spontaneous loading of CoA substrates onto generic ACPs, and that acyl ACP hydrolysis is enhanced by the addition of a type I TE domain. The detection of apoBaeLACPIII in TE-treated samples of ACP+**7**+Sfp but not in the sample ACP+**7**+Sfp alone is a discrepancy that suggests there may have been undetected apoBaeLACPIII present in the ACP+**7**+Sfp sample. These results indicate that in an overnight incubation, BaeRTE's non-cognate ACP partner BaeLACPIII did not show hydrolysis in the absence of a TE or in the presence of the BaeRACPTE construct, and that BaeRTE monodomain showed incomplete hydrolysis (producing the *holo*-BaeLACPIII species, with the *holo*-BaeLACPIII peak integrating for 59% of the **12**+BaeLACPIII peak at the +15 charge state). The presumed major species remaining was the acylated ACP. These results suggest that the BaeRTE shows activity above background hydrolysis in this assay, that the non-cognate ACP BaeLACPIII is recognized as a substrate when acylated with **12** and **10**, and that the hydrolysis reaction is not complete after 10 h of incubation. This may have resulted from catalytic inactivation of the BaeRTE construct or uniformly slow hydrolysis kinetics. The assay design set the ACP substrate at 75 molar TE equivalents (1 μM BaeRTE), and at the initial rate of reaction observed

between the BaeRTE and the equivalent SNAC substrates **1-3** in **Chapter 3**, an assay with 75 μM **7**-ACP and 1 μM BaeRTE would perform 51.3 turnovers (second-order k_p 19 $M^{-1}s^{-1}$). In depth analysis of the extent of conversion requires a standard curve calibration of the LCMS detection method, and tests as to whether kinetic data derived from the colorometric Ellman's assay with SNAC substrates is informative for the ACP equivalents.

The ACP species identified in each condition in **Table 5.2** followed literature precedence and suggested optimization of assay conditions could produce pure acylated BaeRACPIII upon treatment of Sfp, and that the BaeRTE construct is enzymatically active in the assay solution. A second batch of CoA substrates **7** and **9** were synthesized and purified. A modified assay was then conducted with a 1 h Sfp incubation followed by an extended 16 h TE hydrolysis period. The incubations were submitted to an HPLC-ESI MS instrument pair that matched the same HPLC column and ionization conditions used in the initial study. Notably, efforts to buffer exchange using Zeba size-exclusion columns resulted in drastic loss in protein signal, and this work-up step was not added into the final modified sample preparation. **Table 5.3** outlines the repeated BaeRACPIII results.

Table 5.3 Repeated Acylation of BaeRACPIII for repeatability

Assays were modified to include a 6 h hydrolysis step and samples were evaluated using a Hypersil Gold C8 column (100 * 2.1 mm) in a Dionex UltiMate 2000 HPLC system coupled to an API 4000 ESI mass spectrometer. Under these conditions, acylated ACP was unambiguously detected as the principle species in samples supplemented with Sfp, and no significant holo-product was detected in the presence of BaeTE. Standard deviations listed in parentheses are calculated from the variances between charge states +9-11 within a single LCMS run (ESIprot). BaeTES118A (position 128 of TE alignment in **Appendix III-1**) bears an active-site mutation and serves as a negative control for catalyzed hydrolysis with the addition of all components of the assay.

Treatment				**BaeRACPIII Species Detected**			
ACP	Substrate	SFP	TE	Apo (10984.5 Da)	Conjugated Acyl (11420.5 Da)	Saturated Acyl (11422.5 Da)	Holo (11324.5 Da)
+	-	-	-	10990 (1.4)			
+	**10**	+	-		11428 (0.8)		
+	**10**	+	BaeTE		11430 (1.4)		
+	**12**	+	-			11430 (1.2)	
+	**12**	+	BaeTE			11430 (2.1)	
+	**12**	+	BaeTES 118A			11432 (3.5)	

In contrast to the initial assay in **Table 5.1**, only the acylated BaeRACPIII species was detected following incubation of Sfp and mock TE treatment. This assay did not result in detectable amounts of *holo*-BaeRACPIII through hydrolysis alone or in the presence of BaeRTE. Taken together with the results in **Table 5.2**, these results suggest that the synthesized CoA mimics of simplified bacillaene intermediates were loaded onto standalone BaeLACPIII and BaeRACPIII; the current HPLC fractionation did not fully resolve the different loading-state ACP species, and downstream ESI-MS resolved 1-6 multi-charge *m/z* peaks. While the BaeRTE construct showed hydrolysis activity towards BaeLACPIII with both the saturated methyl valeric acyl group **12** and the conjugated alkene **10**, the assay has not been optimized with an adequate control substrate to demonstrate BaeRTE activity with the cognate BaeRACPIII.

5.3 Discussion

My work with BaeRTE in **Chapter 4** indicated that the simple SNAC thioester substrates inactivate the enzyme *in vitro*, but *in vivo* there is strong evidence for the TE being a functional component of the bacillaene pathway during biosynthesis.[1, 25] I therefore questioned whether the presence of the ACP domain enables the TE to effectively hydrolyze the acyl chain without TE inactivation. The assay developed in this chapter detected apo, acyl, and *holo*- BaeLACPIII species resulting from hydrolytic turnover by BaeRTE, but BaeRACPIII did not show catalytic hydrolysis under the same conditions. The work described below details the initial hydrolysis assay conditions and proves detection of intact bacillaene ACP monodomain constructs. The results here demonstrate that hydrolysis conditions must be adjusted to establish suitable for BaeRTE-mediated hydrolysis of acyl-chains off of a BaeRACPIII construct. Only then can BaeRTE be tested for the property of chemical discrimination between different ACP-bound acyl thioesters.

Due to frequently observed cross-reactivity of acyl-ACP type TEs with synthetic acyl-SNAC substrates,[26] acyl-SNACs are used abundantly in studies of TE selectivity. In some cases however, TE activity is dependent on the presence of an acyl-carrier protein intermediate. For example, in both fengycin and mycosubtilin, researchers found that peptidyl-SNAC substrates were not loaded onto the TE active site Ser *in vitro*. However, CPs from the end of the native pathways, directly upstream of the TE, loaded with the same peptidyl chains were processed by the TEs to produce the natural product.[27] These observations clearly show that while the ACP-Ppant analogue SNAC dominate *in vitro* PKS studies, several groups now conduct follow-up ACP-bound assays in order to distinguish between selectivity towards specific acyl substrates and unique enzyme-ACP interaction effects.[28, 29]

Structurally complex acyl-ACP intermediates are often required to characterize not only TE domains, but many other PKS and NRPS catalytic domains as well. Two major strategies have been used for their construction. ApoACPs can be loaded with synthetic-CoA substrates[27, 30] or complex

multimodular PKS systems can deliver acyl chains to the ACP. As construction of multimodular systems to access non-native acyl-ACP chains has proven to be highly challenging, the method of choice to access non-native intermediate is via loading of synthetic substrates. However, this approach is not compatible with repeated loading of the ACP, and thus is limited to stoichiometric use of acyl-ACPs via the substrate tolerant PPTase loading by Sfp.

This single-turnover ACP substrate design is technically limited to lower substrate concentrations and inter- rather than intra-molecular interactions. As a result, the reaction kinetics are intrinsically slower and sensitive ACP detection methods were required for micromolar detection of the reaction products in this chapter.

The structural requirements for an ACP-PPTase interaction, required for loading of synthetic substrates, are relatively small. The consensus sequence on the ACP is a ~ 15 amino acid sequence found within helix 1 and helix 2 (**Figure 5-3**). Using the PPtase Sfp, labelling of both the BaeLACPIV and BaeRACPIII constructs with CoA substrates **10**, **12**, acetyl-, and malonyl-CoA occurred despite the presence of an N-terminal His tag. This observation is consistent with previous literature examples[17, 20] and expected as it relies on the native *in trans* ACP-PPTase interaction. However, the presence of *apo*ACP indicated ACP loading was incomplete for Sfp acting on **10** and **12.** Additionally, synthetic **10** and **12** showed evidence of hydrolysis to free thiol CoA, and stringent HPLC purification of the ACP species prior to addition of the TE would be important to minimize the direct formation of *holo*-ACP via Sfp and CoA.

An ESI-MS based assay of the full length ACP domain was used to confirm formation of the acyl or *holo*-ACP from the apo-ACP. ACP *m*/*z* in the +11 to + 14 charge state could be readily obtained and used to evaluate the state of the active site Ser of the ACP. An HPLC front-end separation was used to limit concerns over ion suppression by buffer components and DMSO, as well as to reduce the complexity of the background during the analysis. For all assay arrangements (API 2000 and API 4000), at least one ACP species was selectively retained on the HPLC column beyond the injection peak and was detected above the baseline count detection for the MS in the

assay. For BaeLACPIII, UV spectroscopy indicated overlapping HPLC elution times for acyl and *holo* ACP species. He *et al.* also observed incomplete separation of the monodomain construct of ACP_{1-4} in similar elution conditions.[20] As a result, tracking of ACP species via HPLC-resolved UV spectroscopy[31] was not possible. Observations that BaeRACPIII did not elute in the flow through over size-exclusion based resin (Zeba, Thermo-Fisher) suggest that work-up conditions should be standardized when comparing between acyl-ACP-TE pairs, and that insoluble protein deposition is a risk on the HPLC columns. This can alter the packing of the column, develop blockages, and creates the possibility of sample contamination if protein remobilizes in later runs.

In some BaeRACPIII samples shown in **Table 5.1**, a second elution of ACP occurred at concentrations of acetonitrile higher than the ACP standard. These second peaks eluted at approximately 60% acetonitrile (% V/V) as apposed to approximately 30% acetonitrile for the common elution range. Blank runs between samples were used to detect and prevent carry-over between samples, but these steps do not control partition of a given assay sample into subfractions in the HPLC based on protein stability.

With regards to the hydrolysis conditions, the BaeTE construct was shown to cleave both **7** and **9** off of the non-cognate BaeLACPIII. This demonstrated the TE was active against some ACP-thioester type substrates, and that the BaeLACPIII monodomain was a tolerated substrate. However, unexpected irregular selectivity against the cognate BaeRACPIII with acyl groups **10** and **12** occurred under the current assay conditions. Additionally, the BaeRACPTE didomain construct did not produce detectable amounts of *holo*-ACP in any of the four acyl ACP combinations. This is in contrast to the results in **Chapter 3**, where both the BaeRTE and BaeRACPTE constructs displayed similar amounts of non-background reactivity towards all three acyl-SNAC substrates.

The rates observed in this stand-alone ACP-TE system must be used only in relative terms. The ACP-TE interaction is mediated in-cis *in vivo*, so the relative activity rates are useful for comparison between two substrates or ACPs, but it may not reveal bottleneck limitations of the *in vivo* assembly line. However, the presence of the ACP is important in some chemical reactivity examples such as

substrates for fengycin TE,[27] in some evolutionary contexts where TE selectivity is important (pyochelin[32] and teicoplanin[33] biosynthesis) and may be important in late-state modified trans AT TE systems.

With respect to the assay method, the LCMS detection of intact ACPs used in this work requires additional refinement for quantitative description of acyl-ACP hydrolysis kinetics. The implementation of HPLC fractionation increases signal sensitivity, but limits this technique to end-point analysis of kinetic assays rather than in-line monitoring. Additionally, the 1 h required for each run/blank combination is less flexible in capturing kinetic events over a range of time-scales than the exhaustive sampling available in the Ellman's colorometric approach (there, the sampling was set at every 3 s after initial characterization of the process suggested no added information would be added by sampling every 0.33 s for 15 min). Finally, a standard curve must be established in order to quantify the LCMS' limit of quantification and to quantify the relative amounts of the different ACP species over the course of the hydrolysis assays; however, as the assay stands, LCMS detection of intact ACPs can be used in end-point assays to confirm whether or not BaeRTE cleaves a majority of an acyl-ACP species. This TE observation method has been used for other enzymes acting on the ACP (KS, AT, type II TE). It is more sensitive than MALDI-tof data of ACPs treated with TEs,[17, 34] more amenable to kinetic assessment than radiolabelled TLC,[13, 34, 35] and less demanding than FTMS[17, 35] or proteolysis.[18]

The next steps in the analysis of the BaeRTE selectivity towards different acyl ACP substrates includes generating an equipment-specific standard curves of the apo, *holo*-, and acylACP species and identifying the limit of quantification on these systems. This will inform whether 2-8 turnovers of 1 μM TE are detectible in a 75 μM ACP assay. Secondly, an acyl substrate and hydrolysis condition must be identified for which BaeRTE constructs accept BaeRACPIII as a catalytic substrate consumed to completion. With these conditions in hand, the assay should be extended to include ten other labelled ACPs from other PKS modules. Finally, this should be compared back

against the SNAC hydrolysis assay and the evidence of *in vivo* activity to determine if the ACP plays a role in maintaining the TE's activity during thioester hydrolysis.

5.4 Conclusion

This chapter described the development of an assay to examine the impact of ACP on BaeRTE mediated acyl-thioester hydrolysis. This required the synthesis of precursor CoA substrates, the cloning and revalidation of the PPTase Sfp and 3 ACPs of the bacillaene pathway, the detection of ACP standards via LCMS, and the adaptation of the ESI-LCMS detection.

For BaeLACPIV, BaeRTE hydrolysed the saturated 2-methyl valeroyl acyl chain **7** as well as the conjugated 2-methyl-thiopent-2-enoate **9**; however, under the same conditions, the BaeRTE construct did not react detectably with acyl groups on BaeRACPIII, and the didomain construct did not react with any substrate combination.

ESI-LCMS of intact ACP monodomains is moderately suited for monitoring protein-bound reaction kinetics without the requirement of MSMS functionality; due to the broad success of SNAC-based TE substrates, only six examples of this technique were documented up to 2017 in the course of this literature review. However, reconstituting the ACP in the general assay environment is important for some aspects of the process, including more relevant measures of binding affinity. Finally, in rare examples such as fengycin TE, the presence of the proper for of carrier protein is essential for any observable activity at all.

5.5 Methods

5.5.1 Apparatus

A reverse phase C8 Hypersil Gold (125 mm *10 mm, 3 μM particle diameter, ThermoFisher Scientific) was used for the purification of CoA substrates **10-12** on a Prominence 20A Modular HPLC (Shimadzu). This column was also used for ESI-LCMS detection of TE hydrolysis products detected on the UltiMate 2000 (HPLC, Dionex) API 4000 (MS, Sciex).

A reverse phase BDS Hypersil C18 column (100 mm * (2.1 mm I.D.), 3 μm particle size, ThermoFisher Scientific) was used for the detection of TE hydrolysis products on the Prominence 20A Modular HPLC (Shimadzu) API 2000 (MS, Applied Biosystems).

5.5.2 Synthesis of CoA-*cis*-3-methyl-2-pentenoate (unconjugated CoA, 11)

2.29 g (2.58 mL) of THF was added to crude product of the *cis*-3-methyl-2-pentenoic reaction acid (**8**, see **Chapter 3.4.5**). 35.8% (by mass) of the reaction (0.286 mmol maximum hypothetical) was mixed with 62.9 mg of carbonyl diimidazole. The reaction was vented and then left closed for 1 h. 36 mg of humid CoA substrate was added to 0.1 mL of deionized water and added to the reaction solution. A white precipitate was formed. The reaction was run for 12 hours. The reaction was evaporated via rotary evaporator and dissolved in 0.750 mL 1:1 (v:v) acetonitrile:water. HPLC analysis on Agilent 1260 indicated bands at elution times of 8.50 min and 22.25 min. A second extraction of the evaporated reaction showed bands 100 times less intense and this second extraction was not carried forward.

0.86 g of crude residue from the previous reaction (dissolved in 2.29 g of THF) was added to 56 mg of carbonyl diimidazole in a dried 10 mL round bottom flask. The reaction was mixed by stir bar for 15 minutes with venting to allow produced gas to escape. 13 mg of coenzyme A (white powder) was dissolved into 200 μL of deionized water. The 200 μL was added to the THF mixture, and a 200 μL rinsing of the coenzyme A tube was added. The reaction proceeded for an hour at

room temperature. The reaction was condensed via rotary evaporator. 800 μL of solution was loaded into HPLC vials.

6.7 mg of CoA was dissolved in 200 μL of water and detected via ESI-LCMS, but no equivalent signal was detected in the CoA-*cis*-3-methyl-2-pentenoate isolates.

5.5.3 Synthesis of 2-methyl-pentanoyl-CoA (12)

10.3 mg of carbonyl diimidizole was added to a 10 mL round bottom flask and dissolved in 0.5 mL THF. 5.0 μL trans-2-methyl-2 pentenoic acid was added to the solution. The round bottom was supplied with nitrogen atmosphere and a vent line to permit escape of CO_2. Gas was produced during the reaction. After 2 h, 6.9 mg of CoA was dissolved in 32 μL of deionized water and added to the THF solution.

After 5 h, the reaction was evaporated in a rotary evaporator at 30 degrees and resuspended in 300 μL 1:1 water acetonitrile and transferred to a plastic conical HPLC vial. The vial was stored in a Shimadzu Prominence 20A Modular HPLC with the following settings: Autosampler chamber 4 °C, solvent A H_20 with 0.05 % trifluoroacetic acid, solvent B acetonitrile, column Thermofisher C8 Hypersil Gold 125*10 mM 3 μM particle diameter heated to 40 °C.

The HPLC was set to run 0.2 mL/min of solvent A at a pressure of 100 bar. 4 μL was injected into the following isocratic separation run: 4 min hold at 100% solvent A followed by a 16 minute gradient to 70% A, and a second steeper gradient to 0% solvent A over 20 minutes.

Absorption at 254 nm was detected between 10 and 24 minutes, but the predicted product mass (M+H 866.19, M+Na 888.17) only eluted significantly between 20 and 22 minutes (See **Appendix IV.10**).

A set of preparatory runs set with 50 μL injections was set for the same solvent conditions, and the eluent was collected between 19.5 and 21.2 minutes.

$C_{27}H_{46}N_7O_{17}P_3S$

ESI-LCMS Observed 886.0 (calc. 886.19)

5.5.4 Synthesis of trans-2-methyl-2 pentenoyl CoA (10)

10.1 mg of carbonyl diimidizole was added to a 10 mL round bottom flask and dissolved in 0.5 mL THF. 5.1 µL of trans-2-methyl-2-pentenoic acid was added to the solution. The round bottom was supplied with nitrogen atmosphere and a vent line to permit escape of CO_2. Gas was produced during the reaction. After 2 h, 6.25 mg of CoA was dissolved in 35 µL of deionized water and added to the THF solution.

After 5 h, the reaction was evaporated in a rotary evaporator at 30 °C and resuspended in 300 µL 1:1 water acetonitrile and transferred to a plastic conical HPLC vial. The vial was stored in a Shimadzu Prominence 20A Modular HPLC with the following settings: Autosampler chamber 4 °C, solvent A H_20 with 0.05 % trifluoroacetic acid, solvent B acetonitrile, column ThermoFisher C8 Hypersil Gold 125*10 mM 3 µM particle diameter heated to 40 °C.

The HPLC was set to run 0.5 mL/min of solvent A at a pressure of 100 bar. 5 µL was injected into the following isocratic separation run:

Time (min)	Module	Event
0	Injection	5 µL injection
3	Pump B (Start of 27 minute ramp)	0%
4	Valco Switch Valve	Switch from waste to MS
30	Pump B (End of 27 minute ramp)	100%
38	Pump B	100%
38.5	Pump B	0%
43.0	Pump B	0%

A large UV peak from 9-13 minutes was observed at 254 nm. The MS mass spectrum from 750-2000 g/z showed a prominent peak at 863.9 corresponding to trans-2-methyl-2 pentenoyl CoA. In subsequent runs, the MS line was rerouted to a 15 mL Falcon tube on ice and 50 µL loadings of the CoA reaction were loaded per run for purification. The same HPLC program was applied with the modification that Valve A switch to collect at 9.3 minutes and switched to waste at 13 minutes.

The Falcon tube was flash frozen and left to lyopholize over night. The final isolate was resuspended in 150 µL of deionized H_2O, aliquoted into 50 µL fractions and frozen at -80 °C.

$C_{27}H_{46}N_7O_{17}P_3S$

Low-res ESI-LCMS Observed 863.9 (calc. 863.17)

5.5.5 Gene synthesis of *Bacillus amyloliquefaciens* FZB42 ACPs

Using antiSMASH 2.0,[21] the sequences of five ACP domains from Bacillus amyloliquefaciens FZB42[36] were identified, aligned for secondary structure and gene sequences codon-optimized for E. coli expression were designed. The sequence surrounding the pET21 MCS was appended to these sequences to facilitate sequencing, transcription, and translation. 117 nt from the pET21 5' region of the MCS and 179 of the 3' region were appended to each sequence. The sequences were designed to ligate into IDT's IDTSmart Kan vector. These synthesized plasmids were designated pMEH12 (difficidin J ACP), pMEH13 (BaeLACPIII), pMEH14 (BaeLACPIV), pMEH15 (BaeRACPII), and pMEH16 (BaeRACPIII).

5.5.6 Cloning of *Bacillus subtilis* Sfp (pMEH22)

Colony PCR was performed on a frozen glycerol stock of BAP1[24] using the primers BAP1_*sfp*_f (GATCCAT ATGAAGATTTACGGAATTTATATG) and BAP1_*sfp*_r (GATCGAATTC AAAAGCTCTTCGTACG) following the general touchdown cycle. Ligation of the PCR into pCR-Blunt Topo II (Thermo Fisher) produced pMEH21. Restriction digest of pMEH21 and pET28 with NdeI and EcoRI, followed by ligation produced the expression plasmid pMEH22.

5.5.7 Expression of Sfp (pMEH22)

pMEH22 was transformed into *E. coli* BL21 and plated on LB + kan. A colony was grown in 4 mL of LB+kan liquid medium overnight at 37 °C. 2 mL of the overnight seed was used as inoculum for 400 mL LB+kan in a preheated 1 L Erlenmyer flask. After 4.5 h the OD_{600} reached 0.63 and IPTG was added to a final concentration of 1 mM. The flask was grown at 20 °C for 15 h. 49 μM protein was produced in 50 mM KH_2PO_4, 150 mM NaCl, 30% glycerol (v:v) with less than 0.5 mM imidazole.

5.5.8 Synthesis of ACP substrates

5.5.8.1 Phosphopantetheine labelling of BaeLACP3 with coenzyme A

25 µL of 200 mM KH_2PO_4 (pH 7.38), 8 µL of 125 mM MgSO4, and 26.2 µL of PCR grade H_20 were combined in a 1.5 mL Eppendorf tube and vortexed. 5 µL of 200 mM coenzyme A, 20 µL of 48.7 µM SFP (in 50 mM KH2PO4, 15% glycerol), and 15.8 µL of 475 µM BaeLACP3 (in 50 mM KH2PO4, 15% glycerol) were added to fill to 100 µL and mixed via pipetting. A control reaction with Sfp and CoA fractions replaced with H_20 was run as an unlabelled *apo* control.

5.5.8.2 General phosphopantetheine labelling and hydrolysis of ACP substrates

25 µL of 200 mM KH_2PO_4 (pH 7.38), 20 µL of 125 mM $MgCl_2$, and sufficient filtered deionized water to final to a final volume of 100 µL were combined in a 1.5 mL Eppendorf tube and vortexed. 5 µL of 50 mM methyl valeric coenzyme A (**12**), 5 µL of 48.7 µM Sfp (in 50 mM KH2PO4, 15% glycerol), and 20 µL of BaeRACP3 (in 50 mM KH_2PO_4, 15% glycerol) were added to fill to 100 µL and mixed via pipetting. Control reactions with Sfp and CoA fractions replaced with H_20 was run as an unlabelled *holo*-ACP and a 50:50 *apo:holo* commix control. The reactions proceeded at room temperature.

25 µL aliquots were withdrawn from the reaction tubes after 1 hour and 10 hours and were mixed with 25 µL of acetonitrile to quench the enzyme reaction. After 10 hours of labelling, 16.7 µL of 7 µM BaeRTE (pMEH19) was added to an unsampled 100 µL labelling reaction. After 1 hour, 25 µL samples were withdrawn and quenched.

These samples were analysed with a reverse phase BDS Hypersil C18 column (100 mm * 2.1 mm I.D., 3 µm particle size) on the Prominence 20A Modular HPLC (Shimadzu) API 2000 (MS, Applied Biosystems) ESI-LCMS configuration. The following chromatography regime was applied:

50 uL of quenched sample was injected into the system with a flowrate of 0.5 mL min^{-1}.

Time	Module	Events	Parameter
3.00	Pumps	Pump B Conc.	0.0
30.00	Pumps	Pump B Conc.	100.0
38.00	Pumps	Pump B Conc.	100.0
38.50	Pumps	Pump B Conc.	0.0
43.00	Pumps	Pump B Conc.	0
43.50	System Controller	Stop	

The API 2000 was operated in MS1 (Q1) with the following parameters:

Current Gas:	20.00
Ion Transfer Voltage:	5500.00
Source and drying gas Temp:	250.00
Gas Flow 1:	30.00
Gas Flow 2:	20.00
Interface heater:	ON
Declustering potential	22.00
Focusing potential	400.00
Entrance Potential	5.00

This procedure was also applied for BaeLACPIII and **10**, as well as both **10** and **12** in combination with BaeRACPIII.

5.6 References

1. J. Moldenhauer, X.-H. Chen, R. Borriss and J. Piel, *Angewandte Chemie - International Edition*, 2007, **46**, 8195-8197.
2. R. S. Kombu, H. Brunengraber and M. A. Puchowicz, *Methods in Molecular Biology*, 2011, **708**, 147-157.
3. W. Y. Lu, M. F. Clasquin, E. Melamud, D. Amador-Noguez, A. A. Caudy and J. D. Rabinowitz, *Analytical Chemistry*, 2010, **82**, 3212-3221.
4. B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien and J. D. Rabinowitz, *Nature Chemical Biology*, 2009, **5**, 593-599.
5. J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel and M. D. Burkart, *Natural Product Reports*, 2014, **31**, 61-108.
6. V. Agarwal, S. Diethelm, L. Ray, N. Garg, T. Awakawa, P. C. Dorrestein and B. S. Moore, *Organic Letters*, 2015, **17**, 4452-4455.
7. A. Busche, D. Gottstein, C. Hein, N. Ripin, I. Pader, P. Tufar, E. B. Eisman, L. Gu, C. T. Walsh, D. H. Sherman, F. Löhr, P. Güntert and V. Dötsch, *ACS Chemical Biology*, 2012, **7**, 378-386.
8. A. Kawaguchi, T. YoshlMura and S. Okuda, *The Journal of Biochemistry*, 1981, **89**, 337-339.
9. A. Hagen, S. Poust, T. de Rond, J. L. Fortman, L. Katz, C. J. Petzold and J. D. Keasling, *ACS Synthetic Biology*, 2016, **5**, 21-27.
10. L. Tran, R. W. Broadhurst, M. Tosin, A. Cavalli and K. J. Weissman, *Chemistry & Biology*, 2010, **17**, 705-716.

11. L. Tran, M. Tosin, J. B. Spencer, P. F. Leadlay and K. J. Weissman, *Chembiochem : a European journal of chemical biology*, 2008, **9**, 905-915.
12. Y. Liu, T. Zheng and S. D. Bruner, *Chemistry & Biology*, 2011, **18**, 1482-1488.
13. Z. Zhou, J. R. Lai and C. T. Walsh, *Chemistry & Biology*, 2006, **13**, 869-879.
14. H. Nakamura, H. A. Hamer, G. Sirasani and E. P. Balskus, *Journal of the American Chemical Society*, 2012, **134**, 18518-18521.
15. D. B. Stein, U. Linne, M. Hahn and M. A. Marahiel, *Chembiochem : a European journal of chemical biology*, 2006, **7**, 1807-1814.
16. P. C. Dorrestein, S. B. Bumpus, C. T. Calderone, S. Garneau-Tsodikova, Z. D. Aron, P. D. Straight, R. Kolter, C. T. Walsh and N. L. Kelleher, *Biochemistry*, 2006, **45**, 12756-12766.
17. S. Garneau, P. C. Dorrestein, N. L. Kelleher and C. T. Walsh, *Biochemistry*, 2005, **44**, 2770-2780.
18. C. A. Shaw-Reid, N. L. Kelleher, H. C. Losey, A. M. Gehring, C. Berg and C. T. Walsh, *Chemistry & Biology*, 1999, **6**, 385-400.
19. P. W. Murphy, E. E. Rowland and D. M. Byers, *Journal of the American Society for Mass Spectrometry*, 2007, **18**, 1525-1532.
20. H.-Y. He, M.-C. Tang, F. Zhang and G.-L. Tang, *Journal of the American Chemical Society*, 2014, **136**, 4488-4491.
21. K. Blin, M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano and T. Weber, *Nucleic Acids Research*, 2013, **41**, W204-212.
22. P. Di Tommaso, S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J. M. Chang, J. F. Taly and C. Notredame, *Nucleic Acids Research*, 2011, **39**, W13-17.
23. X. Robert and P. Gouet, *Nucleic Acids Research*, 2014, **42**, W320-W324.
24. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science (New York, N.Y.)*, 2001, **291**, 1790-1792.
25. J. Moldenhauer, D. C. G. Goetz, C. R. Albert, S. K. Bischof, K. Schneider, R. D. Suessmuth, M. Engeser, H. Gross, G. Bringmann and J. Piel, *Angewandte Chemie - International Edition*, 2010, **49**, 1465-1467.
26. J. Franke and C. Hertweck, *Cell Chemical Biology*, 2016, **23**, 1179-1192.
27. S. A. Sieber, C. T. Walsh and M. A. Marahiel, *Journal of the American Chemical Society*, 2003, **125**, 10862-10866.
28. J. W. Sims and E. W. Schmidt, *Journal of the American Chemical Society*, 2008, **130**, 11149-11155.
29. N. M. Gaudelli and C. A. Townsend, *Nature Chemical Biology*, 2014, **10**, 251-258.
30. L. E. Quadri, T. A. Keating, H. M. Patel and C. T. Walsh, *Biochemistry*, 1999, **38**, 14941-14954.
31. A. S. A. Mady, O. E. Zolova, M. San Millan, G. Villamizar, F. de la Calle, F. Lombo and S. Garneau-Tsodikova, *Molecular Biosystems*, 2011, **7**, 1999-2011.
32. H. M. Patel and C. T. Walsh, *Biochemistry*, 2001, **40**, 9023-9031.
33. M. Peschke, C. Brieke, M. Heimes and M. J. Cryle, *ACS Chemical Biology*, 2017, **13**, 110-120.
34. X. Y. Yu, T. G. Liu, F. Y. Zhu and C. Khosla, *Proceedings of the National Academy of Sciences of the United States of America*, 2011, **108**, 18643-18648.
35. L. M. Miller, M. T. Mazur, S. M. McLoughlin and N. L. Kelleher, *Protein Science*, 2005, **14**, 2702-2712.
36. X. H. Chen, A. Koumoutsi, R. Scholz, A. Eisenreich, K. Schneider, I. Heinemeyer, B. Morgenstern, B. Voss, W. R. Hess, O. Reva, H. Junge, B. Voigt, P. R. Jungblut, J. Vater, R. Sussmuth, H. Liesegang, A. Strittmatter, G. Gottschalk and R. Borriss, *Nature Biotechnology*, 2007, **25**, 1007-1014.

Chapter 6 Conclusion

Our understanding of the chemical diversity of polyketide and non-ribosomal peptide natural products in microbial ecology and chemoenzymatic synthesis has accelerated with the development of pathway-function prediction. Unlike many free-standing metabolic protein pathways, modular polyketide synthases hold the metabolite tethered in place during synthesis, and a strong element of collinearity has evolved. The order in which the on-pathway reactions occur is dictated by the order in which the genes occur in the pathway. As a result, the function of each domain can be predicted, and the sequence of these extensions and stereo-specific modifications (keto-reduction, dehydration, enoyl-reduction, isomerization, methylation) can be combined to roughly predict the biosynthetic product produced from the gene sequence alone.[1] With over 119 000 draft bacterial genomes sequenced and 40 000 projected to be sequenced in 2018 (PatricBRC)[2], study of how the sequence-based properties of the individual domains influence the chemistries they perform advances not only our understanding of the mechanism of action, but also enhances the predictive power of algorithms designed to predict the chemical outcome of newly discovered and designed pathways alike.

Understanding of TE loading and release selectivity lags behind other domains in this respect.[3] Kinetic study of TE domains against panels of thioester substrates is a core approach in mapping correlations between elements of TE protein sequence and different classes of PK and NRP product release. The goals of this work are 1) to advance the understanding of the mechanisms of TE loading-substrate and release-nucleophile selectivity and 2) to develop analytical methodology and modelling theory as tools for quantitative measure of TE activity.

The bioinformatic analysis of TE sequence diversity, structure, and mechanism presented in **Chapter 2** demonstrates that overall TE sequence alignment correlates to natural product family and that TEs from *trans*-AT PKSs such as BaeRTE form a distinct clade from the canonical model *cis*-AT TEs. In addition, through PKS and NRPS case-studies of selectivity in the loading and

release stages, this work frames the selectivity of TEs as discrete interactions between subsections of the TE protein sequence and discrete aspects of a substrate's chemical topology. This work has been adopted by the natural product scientific community as a general overview of the role of TEs in release as well as in specific examples focusing on unusual TE selectivity in the biosynthesis of salinamide,[4] sulfazecin,[5] and teicoplanin[6] biosynthesis.

We have presented data that BaeRTE coded from two isolates with 57 % amino acid identity (73 % similarity) both exhibit declining activity under the conditions used for Ellman's spectroscopic kinetic characterization. This matches reports that piericin TE (*cis*-AT),[7] radicicol TE (fungal TE),[8] and disorazole TE (*trans*-AT)[9] display substrate inhibition under these experimental conditions as stand-alone protein domains. We demonstrated that initial rate of reaction data is insufficient to describe substrate selectivity in these systems and developed a time-dependent model capable of accounting for irreversible substrate inhibition during catalysis. This model was sufficient to reproduce the TE activity observed between BaeRTE or BaeRACP and methyl valeric-SNAC **3** as well as PieTE with medium chain diketoacyl-SNAC substrates.[7] We also demonstrated that this model is applicable to other fields of enzymology.[10] However, the model does not fit the data observed for the equivalent unsaturated substrates **1** and **2**. Finally, we detailed synthetic and assay conditions for the production, cleavage, and detection of the ACP-tethered substrate equivalents. This contributes towards the development of complex TE assays that better estimate type I bacterial TE activity during polyketide synthesis.

6.1 References

1. T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Muller, W. Wohlleben, R. Breitling, E. Takano and M. H. Medema, *Nucleic Acids Research*, 2015, **43**, W237-W243.
2. A. R. Wattam, J. J. Davis, R. Assaf, S. Boisvert, T. Brettin, C. Bun, N. Conrad, E. M. Dietrich, T. Disz, J. L. Gabbard, S. Gerdes, C. S. Henry, R. W. Kenyon, D. Machi, C. Mao, E. K. Nordberg, G. J. Olsen, D. E. Murphy-Olson, R. Olson, R. Overbeek, B. Parrello, G. D. Pusch, M. Shukla, V. Vonstein, A. Warren, F. F. Xia, H. Yoo and R. L. Stevens, *Nucleic Acids Research*, 2017, **45**, D535-D542.
3. T. P. A. Hari, P. Labana, M. Boileau and C. N. Boddy, *Chembiochem : a European journal of chemical biology*, 2014, **15**, 2656-2661.
4. L. Ray, K. Yamanaka and B. S. Moore, *Angewandte Chemie - International Edition*, 2016, **55**, 364-367.
5. R. A. Oliver, R. F. Li and C. A. Townsend, *Nature Chemical Biology*, 2018, **14**, 5-7.
6. M. Peschke, C. Brieke, M. Heimes and M. J. Cryle, *ACS Chemical Biology*, 2017, **13**, 110-120.
7. Q. Liu, F. Yao, Y. H. Chooi, Q. J. Kang, W. Xu, Y. R. Li, Y. C. Shao, Y. F. Shi, Z. X. Deng, Y. Tang and D. L. You, *Chemistry & Biology*, 2012, **19**, 243-253.
8. G. W. Heberlig, M. Wirz, M. Wang and C. N. Boddy, *Organic Letters*, 2014, **16**, 5858-5861.
9. K. Buntin, K. J. Weissman and R. Muller, *Chembiochem : a European journal of chemical biology*, 2010, **11**, 1137-1146.
10. C. Gravel, D. Lapierre, J. Labelle and J. W. Keillor, *Canadian Journal of Chemistry-Revue Canadienne De Chimie*, 2007, **85**, 164-174.

Appendix I Supplemental information for Chapter 2

I.1 Figures

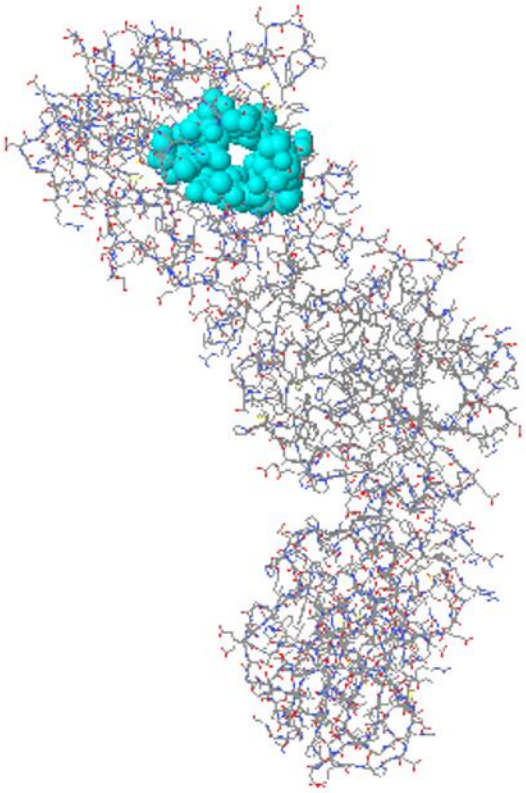


Figure I-1 **CASTp[1] pore estimation of DEBS TE from 1kez[2] (PDB).** The spheres represent the surface atoms CASTp uses to calculate the surface area and volume of the TE pore.

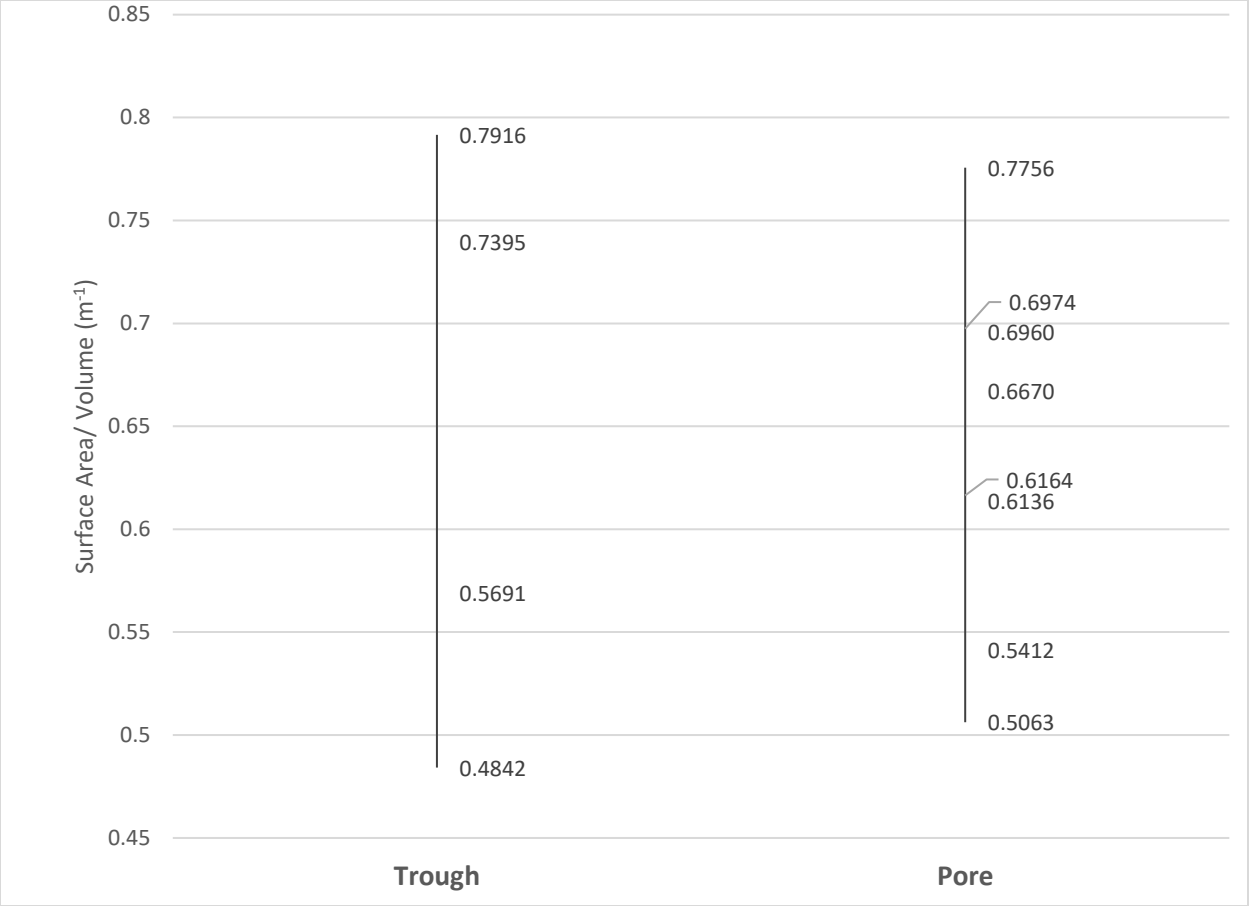
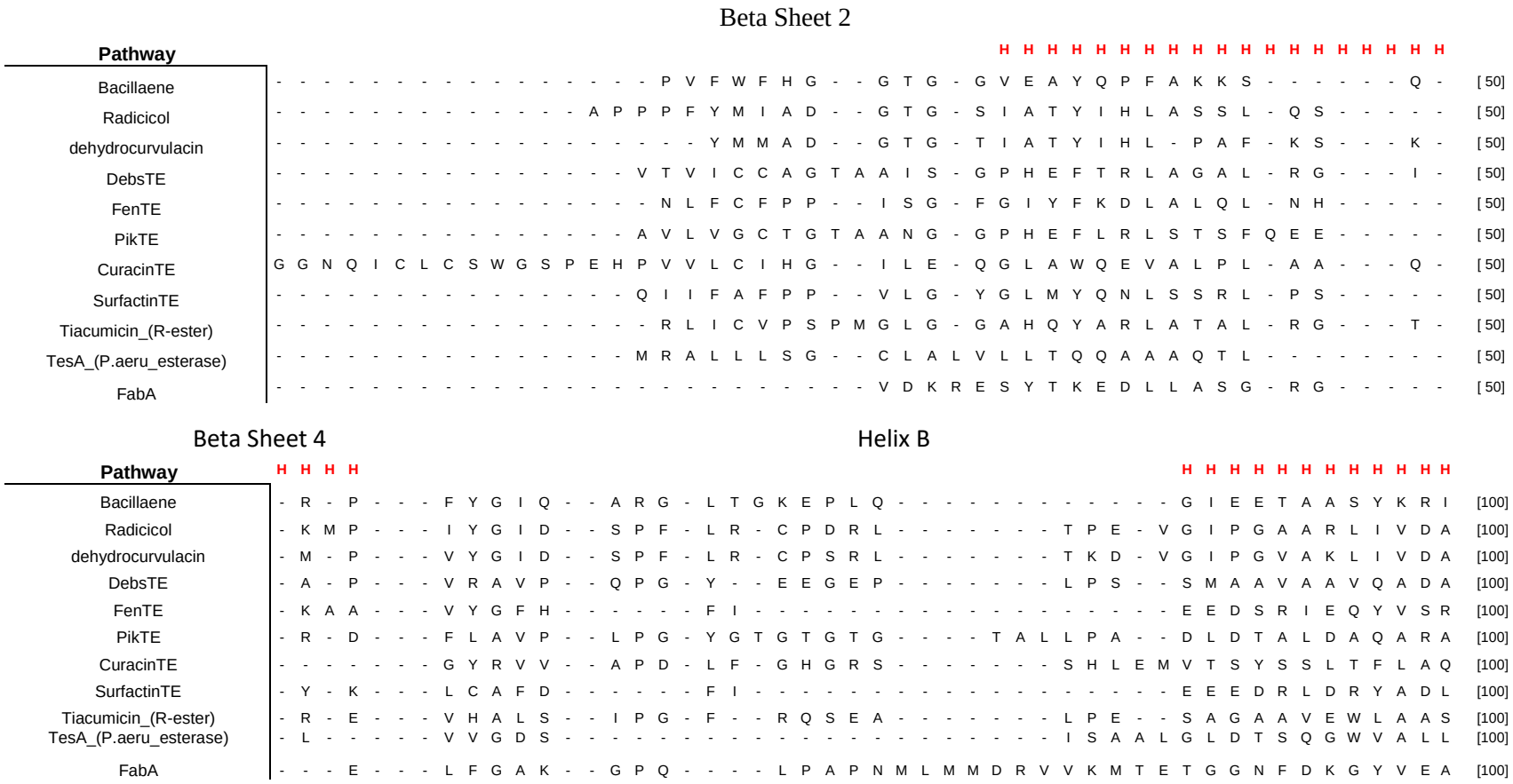
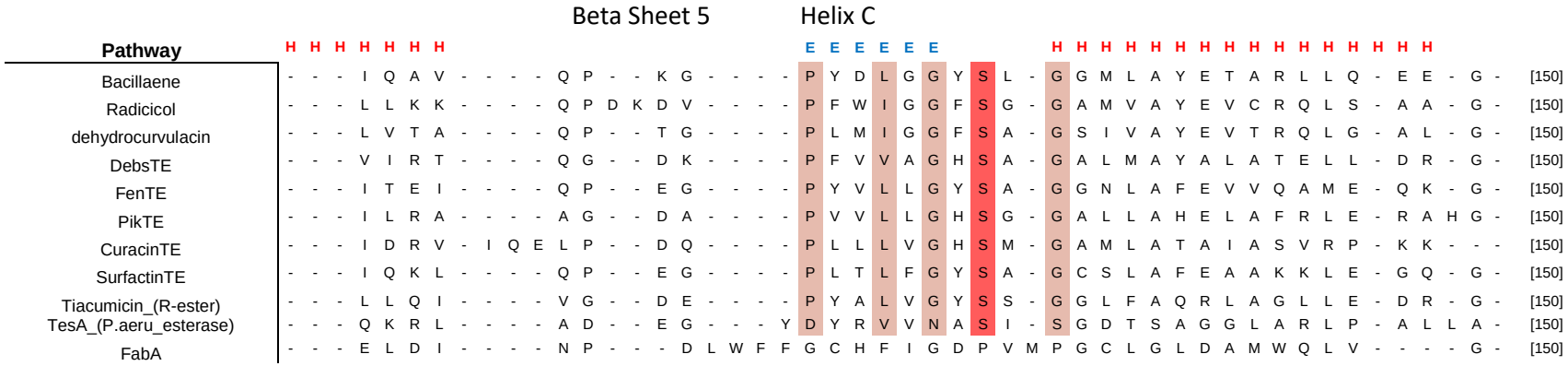
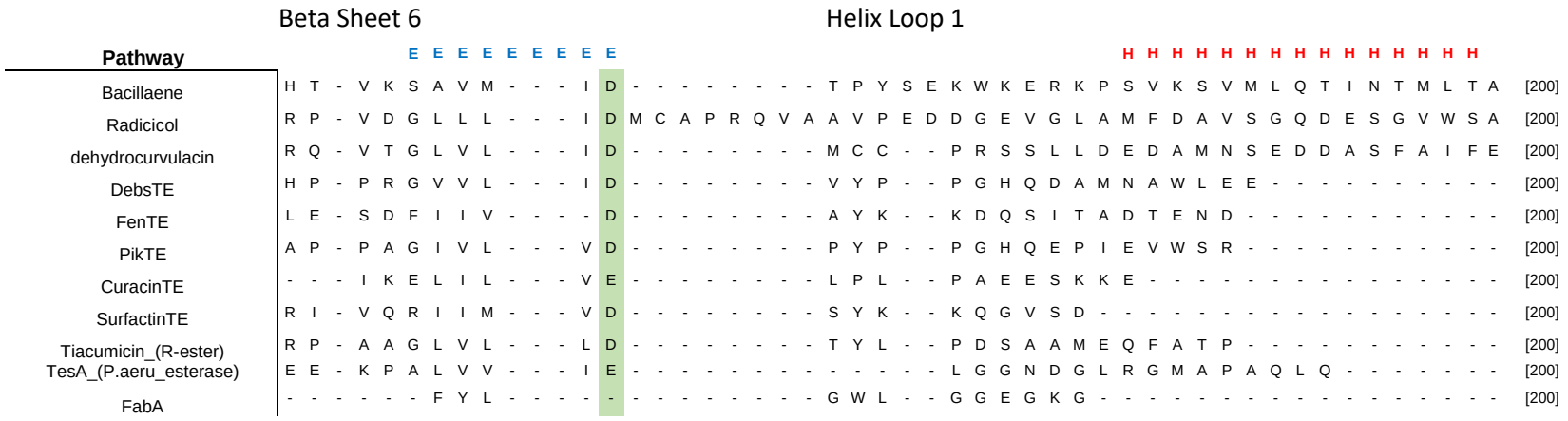


Figure I-2 **Estimated surface to area to volume ratio of active-site cavities for crystallized trough- and pore-type thioesterases.** Ratio ranges of both types overlap. Overlapping pore values include 0.6136 and 0.6164, as well as 0.6960 and 0.6974

Figure I-3 **Sample of thioesterase domains aligned and sorted phylogenetically in Figure 2.4.** Secondary structure naming follows **Figure 2.2A** and secondary structure prediction (by jPred4)[3] is listed in the first line with H for helix and E for extended. The catalytic triad and consensus sequence conserved in 80% of the 340 thioesterases is highlighted.




Beta Sheet 5
Helix C
Pathway
Bacillaene
Radicicol
dehydrocurvulacin
DebsTE
FenTE
PikTE
CuracinTE
SurfactinTE
Tiacumicin_(R-ester)
TesA_(P.aeru_esterase)
FabA

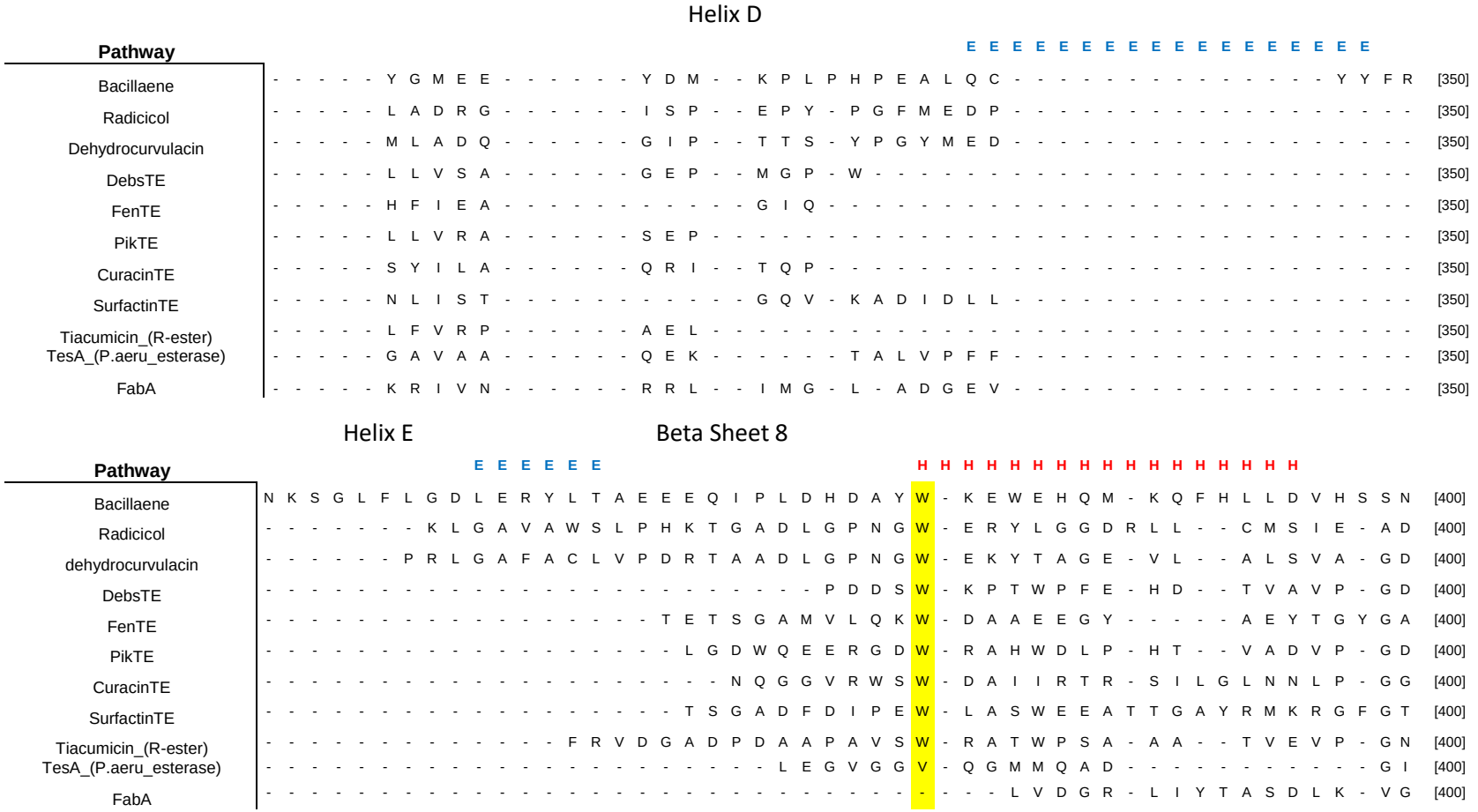

Beta Sheet 6
Helix Loop 1
Pathway
Bacillaene
Radicicol
dehydrocurvulacin
DebsTE
FenTE
PikTE
CuracinTE
SurfactinTE
Tiacumicin_(R-ester)
TesA_(P.aeru_esterase)
FabA

Helix Loop 2

```
Pathway                                                   HHHHHHHHHHHHHHHHHHHHH
Bacillaene                  YLKPENLEEALIS----------REDITVQAE------------------  [250]
Radicicol                   TDGTRRHLGAMF-----------SCVATYNPEP-----------------  [250]
dehydrocurvulacin           NAVSKDGLWSLASTTQDHFRAYHVAMHAYHPP------------------  [250]
DebsTE                      -----------------------LTATLFDRE------------------  [250]
FenTE                       -----------------------DSAAYLPEAV-----------------  [250]
PikTE                       -----------------------QLGEGL---------------------  [250]
CuracinTE                   -----------------------SAVNQLTT-------------------  [250]
SurfactinTE                 -----------------------LDGRTVESD------------------  [250]
Tiacumicin_(R-ester)        -----------------------MLEGLFARE------------------  [250]
TesA_(P.aeru_esterase)      -----------------------QNLASMAQK------------------  [250]
FabA                        --------------------------------------------------  [250]
```

Beta Sheet 7

```
Pathway                     HHHHHHHHH        HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Bacillaene                  EEEVLKELIGLAKKRGLPKSEEAI------------RMQLEQMMKTQLA-  [300]
Radicicol                   LPPRGGGNTPAKRAAMIWARKGMI------------DRCVSSVRFRQM--  [300]
dehydrocurvulacin           ---YMTAQERPSHTAVIWAEKGMV------------NRVVGND-RLMQ--  [300]
DebsTE                      ----TVRMDDTRLTALGAYDRLTG------------QWRPRET-GLPT--  [300]
FenTE                       RETVMQKKRCYQEYWAQLINEGRI--------------------KSNI--  [300]
PikTE                       FAGELEPMSDARLLAMGRYARFLA------------GPRPGRS-SAPV--  [300]
CuracinTE                   CLDYLSSTPQHPIFPDVATAASRL------------RQAIPSL-SEEF--  [300]
SurfactinTE                 VEALMNVNRDNEALNSEAVKHGLK------------QKTHAFY-SYYV--  [300]
Tiacumicin_(R-ester)        --ESFGPFTSARLSAMGRYVRLIA------------ECEAPPV-TTPT--  [300]
TesA_(P.aeru_esterase)      ARAEGAKVLLLGIQLPPNYGPRYI------------EAFSRVY-------  [300]
FabA                        -----RALGVGEVKFTGQVLPTAK---------------KVTY-RIHF--  [300]
```


Helix D
Pathway
Bacillaene
Radicicol
Dehydrocurvulacin
DebsTE
FenTE
PikTE
CuracinTE
SurfactinTE
Tiacumicin_(R-ester)
TesA_(P.aeru_esterase)
FabA
[350]
Helix E
Beta Sheet 8
Pathway
dehydrocurvulacin
[400]

Helix F

Pathway		Sequence	
		H H H H H H H H H (secondary structure row above the columns from P to K in the Bacillaene sequence)	
Bacillaene	H	V T - - - - M L S E P K P Q N A I K A F C E K L Y - - - - - -	[432]
Radicicol	H	L E - - - - M P T P G Y A Q L L G E T M D K A F R - - - - - -	[432]
dehydrocurvulacin	H	L D - L - - - - - - -	[432]
DebsTE	H	F T - - - - - M V Q E H A D A I A R H I D A W L G - - - - - -	[432]
FenTE	H	K D - - - - M L E G E F A E K N A N I I L N I L D - - - - - -	[432]
PikTE	H	F T - - - - - M M R D H A P A V A E A V L S W L D - - - - - -	[432]
CuracinTE	R	S Q - - - - Y L E M L K S I Q V P T T L V Y G D - - - - - - -	[432]
SurfactinTE	H	A E - - - - M L Q G E T L D R N A G I L L E F L N - - - - - -	[432]
Tiacumicin_(R-ester)	H	F T - - - - - M L E S E A A G T A R V I E Q W L - - - - - - -	[432]
TesA_(P.aeru_esterase)	H	P A - - - L A A Q P R L L E N V W P T L K P L L - - - - - - -	[432]
FabA	L	F Q - - - - - - - - - - - - - - - - D T S A F - - - - - - - -	[432]

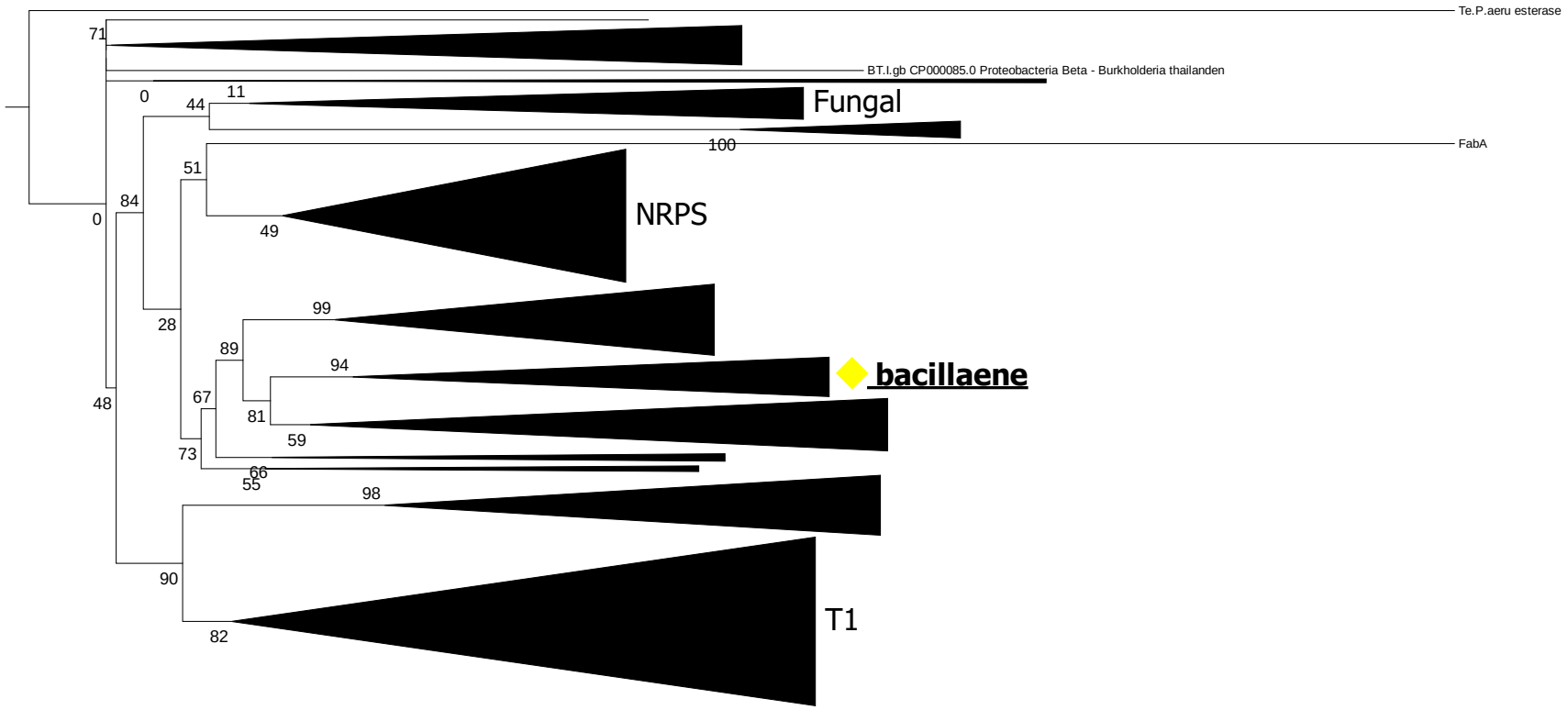


Figure I-4 **Branch support for Figure 2.4.** Branches with label contain only one source of thioesterase (type 1, either PKS, NRPS, or fungal). Unlabeled branches contain a mixture. Bacillaene highlighted with the diamond. Branch support using Shimodaira-Hasegawa (SH) type approximate likelihood ratio test (aLRT) via phylogeny.fr.[4] The aLRT value is a measure of the loss in model accuracy if a given branch is collapsed. Values range from no loss (0), to significant loss (100).

I.2 References

1. J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz and J. Liang, *Nucleic Acids Research*, 2006, **34**, W116-W118.
2. S. C. Tsai, L. J. W. Miercke, J. Krucinski, R. Gokhale, J. C. H. Chen, P. G. Foster, D. E. Cane, C. Khosla and R. M. Stroud, *Proceedings of the National Academy of Sciences of the United States of America*, 2001, **98**, 14808-14813.
3. A. Drozdetskiy, C. Cole, J. Procter and G. J. Barton, *Nucleic Acids Research*, 2015, **43**, W389-W394.
4. A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie and O. Gascuel, *Nucleic Acids Research*, 2008, **36**, W465-W469.

Appendix II

Table II.1 Key Plasmid Features

Plasmid Name	Description	5' Resctriction site	3' Resctriction site	Resistance	Origin
pMEH01	BacACP in Blunt	*NdeI*	*EcoRI*	Kan	pKJ13
pMEH02	BacTE in blunt (has internal *NdeI* site)	*NdeI*	*EcoRI*	Kan	pKJ13
pMEH03	pNRC181.1 XbaI *EcoRI* into pKH22	*NdeI*	*EcoRI*	Kan	pNRC181.1
pMEH04	pNRC51.1 *BamHI EcoRI* into pMEH03	*NdeI*	*EcoRI*	Kan	pNRC51.1
pMEH05	pNRC181.1 XbaI *AvrII* into pMEH04 *AvrII*	*NdeI*	*EcoRI*	Kan	pNRC181.1
pMEH06	BacTE from 02 in pET28 (has internal *NdeI* site)	*NdeI*	*EcoRI*	Kan	pMEH02
pMEH07	BaeACP in pET21 (no His tag (proline))	*NdeI*	*EcoRI*	AMP	pMEH07
pMEH08	BacACP in pET28 (observed product)	*NdeI*	*EcoRI*	Kan	pMEH07
pMEH09	BacTE PCR in blunt (*NheI* no c term stop)	*NheI*	*EcoRI*	Kan	pKJ13
pMEH10	BacTE from pMEH09 into pET21	*NheI*	*EcoRI*	AMP	pMEH09
pMEH11	BacTE from pMEH09 into pET28	*NheI*	*EcoRI*	Kan	pMEH09
pMEH12	difJ in IDTSmart	*NdeI*	*EcoRI*	Kan	synthetic
pMEH13	BaeL ACPIII IDTSmart	*NdeI*	*EcoRI*	Kan	synthetic
pMEH14	BaeL ACPIV IDTSmart	*NdeI*	*EcoRI*	Kan	synthetic
pMEH15	BaeR ACPII IDTSmart	*NdeI*	*EcoRI*	Kan	synthetic
pMEH16	BaeR ACPIII IDTSmart	*NdeI*	*EcoRI*	Kan	synthetic
pMEH17	gephJTE in pMalC5x	*NdeI*	*EcoRI*	AMP	
pMEH18	baeTE in pCR blunt	*NheI*	*EcoRI*	Kan	pKJ13
pMEH19	baeTE from pMEH18 in pET28	*NheI*	*EcoRI*	Kan	pMEH18
pMEH20	baeTE H289R from pMEH19	*NheI*	*EcoRI*	Kan	pMEH19
pMEH21	BAP1 sfp in blunt	*NdeI*	*EcoRI*	Kan	BAP1
pMEH22	BAP1 sfp in pET28	*NdeI*	*EcoRI*	Kan	pMEH21
pMEH23	BAP1 sfp in pET21	*NdeI*	*EcoRI*	AMP	pMEH21
pMEH24	StoA mutant BaeRTE in pET28	*NdeI*	*EcoRI*	Kan	pMEH19
pMEH25	Codon optimized disorazol in blunt	*NdeI*	*EcoRI*	Kan	synthetic

pMEH26	Codon opt chivTE in blunt (point mut 3' *EcoRI* region)	*NdeI*	*EcoRI*	Kan	synthetic
pMEH27	Codon opt RhiTE in blunt (point mut: extra T)	*NdeI*	*EcoRI*	Kan	synthetic
pMEH28	dis from pMEH25 in pET28	*NdeI*	*EcoRI*	Kan	pMEH25
pMEH29	dis from pMEH25 in pET21	*NdeI*	*EcoRI*	AMP	pMEH25
pMEH30	ChiTE from pMEH26 in pET28	*NdeI*	*EcoRI*	Kan	pMEH26
pMEH31	ChiTE from pMEH26 in pET21	*NdeI*	*EcoRI*	AMP	pMEH26
pMEH32	BaeTE with *AvrII* and XbaI sites in pJet 1.2	XbaI	*AvrII*	Amp	pMEH19
pMEH33	BaeTE with '5 *BamHI* and 3' XbaI in pHT01	*BamHI*	XbaI	*E. coli* amp, *B. amy* chlor	pMEH32
pMEH34	5' third of BaeTE from JM33 in pGemT	XhoI	*EcoRI*	Amp	pJM33
pMEH35	5' third of BaeTE from JM33 in pJM33	XhoI	*EcoRI*	Amp	pMEH34
pMEH36	Fixed pMEH15 (BaeRACPII) in IDTsmart	*NdeI*	*EcoRI*	Kan	pMEH15
pMEH37	Fixed ChiTE in blunt	*NdeI*	*EcoRI*	Kan	pMEH26
pMEH38	Fixed ChiTE in pET21	*NdeI*	*EcoRI*	Amp	pMEH37
pMEH39	ElaTE in blunt	*NdeI*	*EcoRI*	Kan	synthetic
pMEH40	ElaTE in pET21	*NdeI*	*EcoRI*	Amp	pMEH39
pMEH41	ElaTE in pET28	*NdeI*	*EcoRI*	Kan	pMEH39
pMEH42	ElaTE in pMalc5x (doesn't require T7)	*NdeI*	*EcoRI*	Amp	pMEH40
pMEH43	Fixed RhiTE in blunt	*NdeI*	*EcoRI*	Kan	pMEH27
pMEH44	Fixed RhiTE pET21	*NdeI*	*EcoRI*	Amp	pMEH43
pMEH45	BaeRTE168 in blunt	*NdeI*	*EcoRI*	Kan	synthetic
pMEH46	BaeRTE168 in pET21	*NdeI*	*EcoRI*	Amp	pMEH45
pMEH47	BaeRTE168 in pMEH16	*NdeI*	SphI	Kan	pMEH45
pMEH48	BaeRTEfzb42 in PMalC5x	*NdeI*	*EcoRI*	Amp	pMEH46
pMEH49	BaeR last KS with C-his in blunt	*NheI*	*EcoRI*	Kan	FZB42
pMEH50	BaeR last KS with C-his in pET28	*NheI*	*EcoRI*	Kan	pMEH49

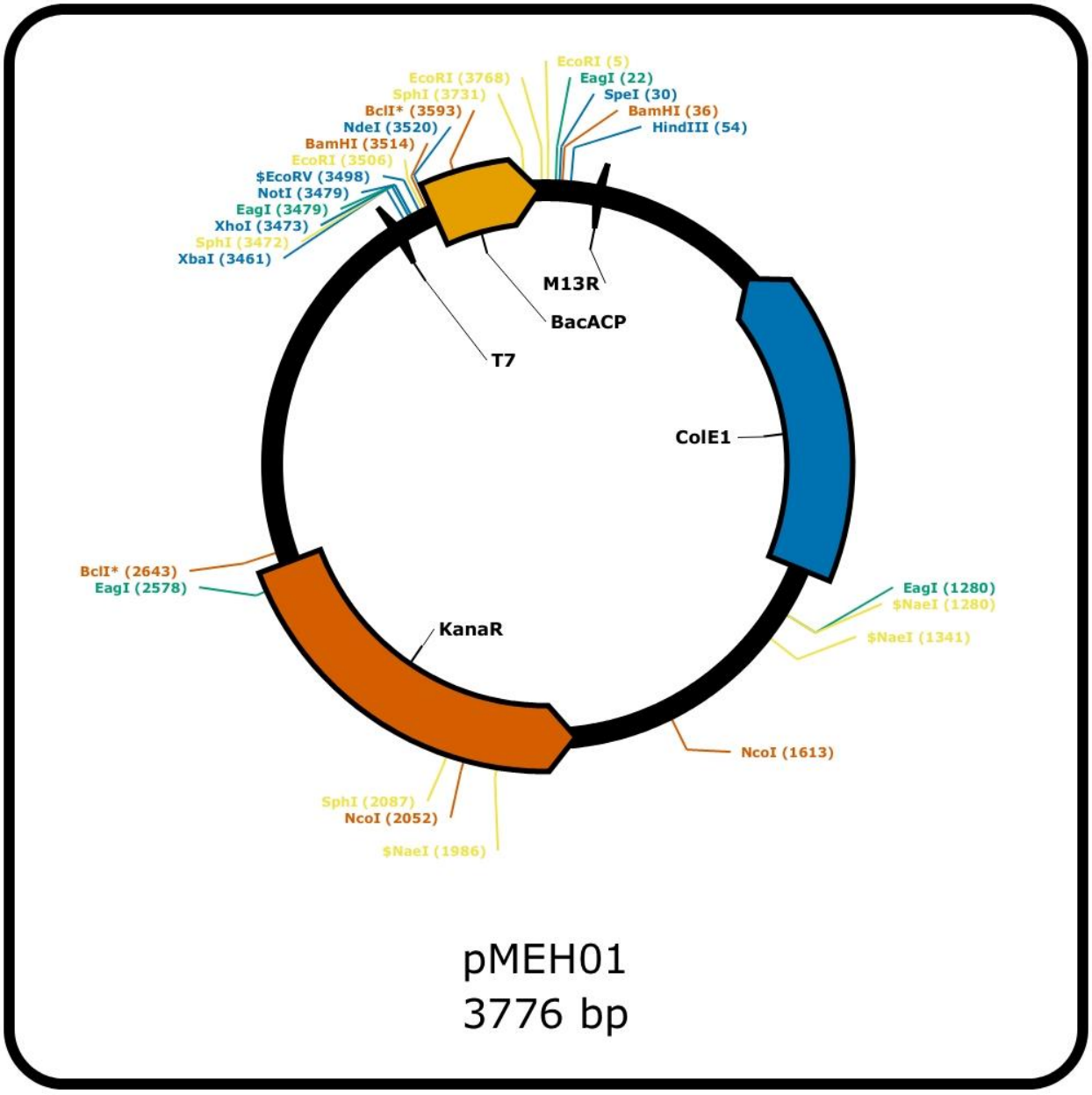


Figure II-1 pMEH01 *B. amyloliquefaciens* FZB42 BaeRACPIII in pTOPOblunt.

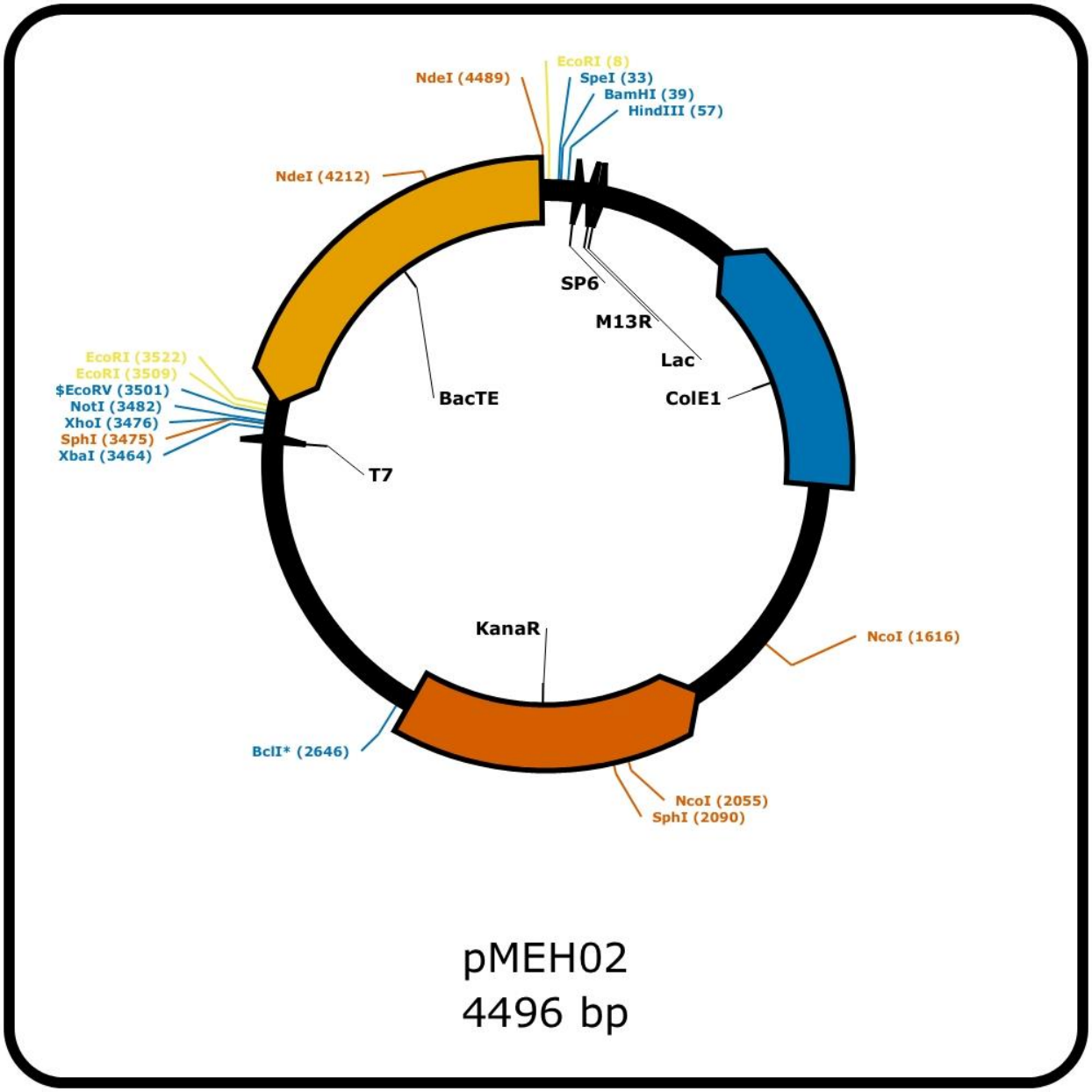


Figure II-2 pMEH02 B. amyloliquefaciens BaeRTE *NdeI* truncant in pTOPOblunt.

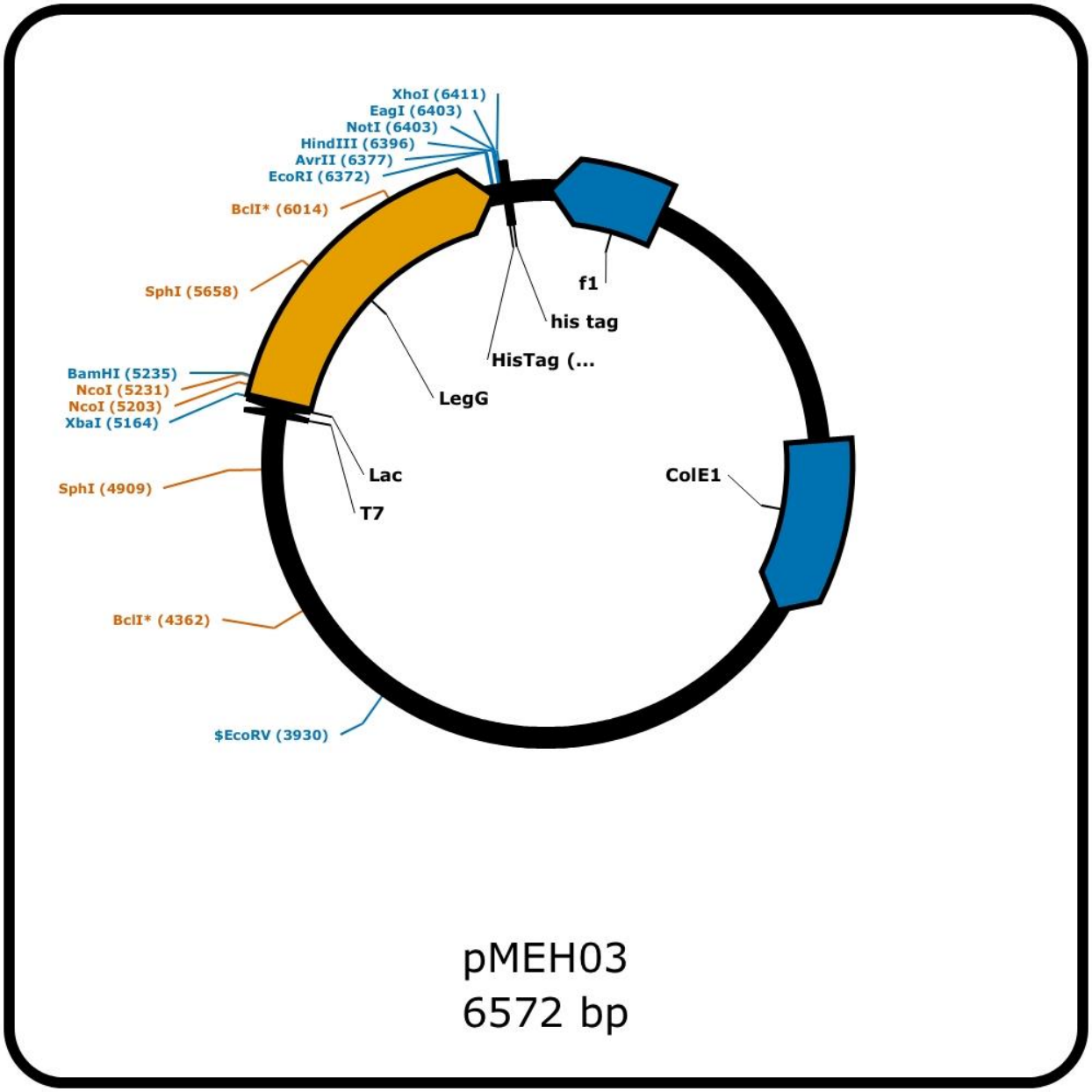


Figure II-3 pMEH03 pNRC181.1 XbaI *EcoRI* into pKH22.

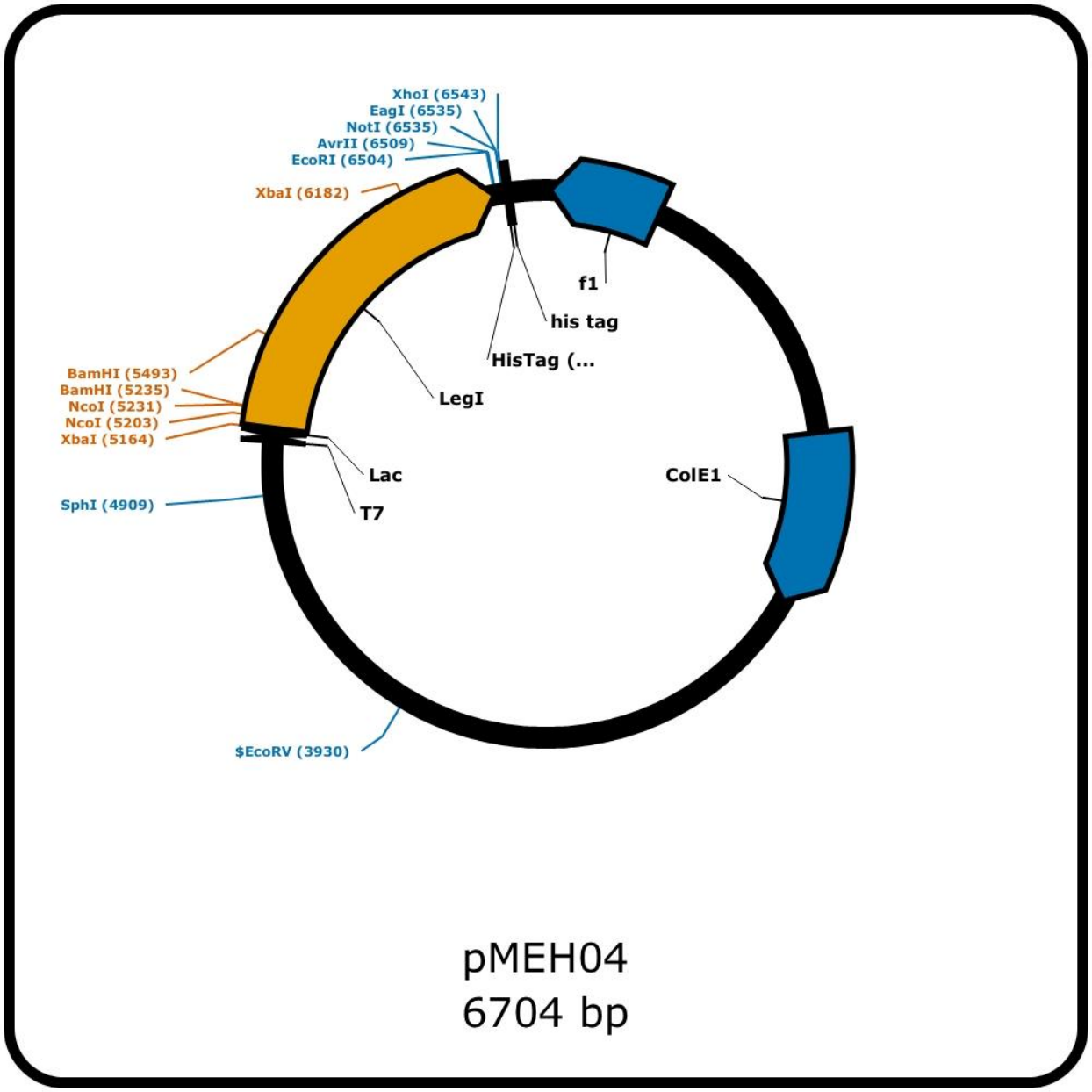


Figure II-4 pMEH04 pNRC51.1 *BamHI EcoRI* into pMEH03.

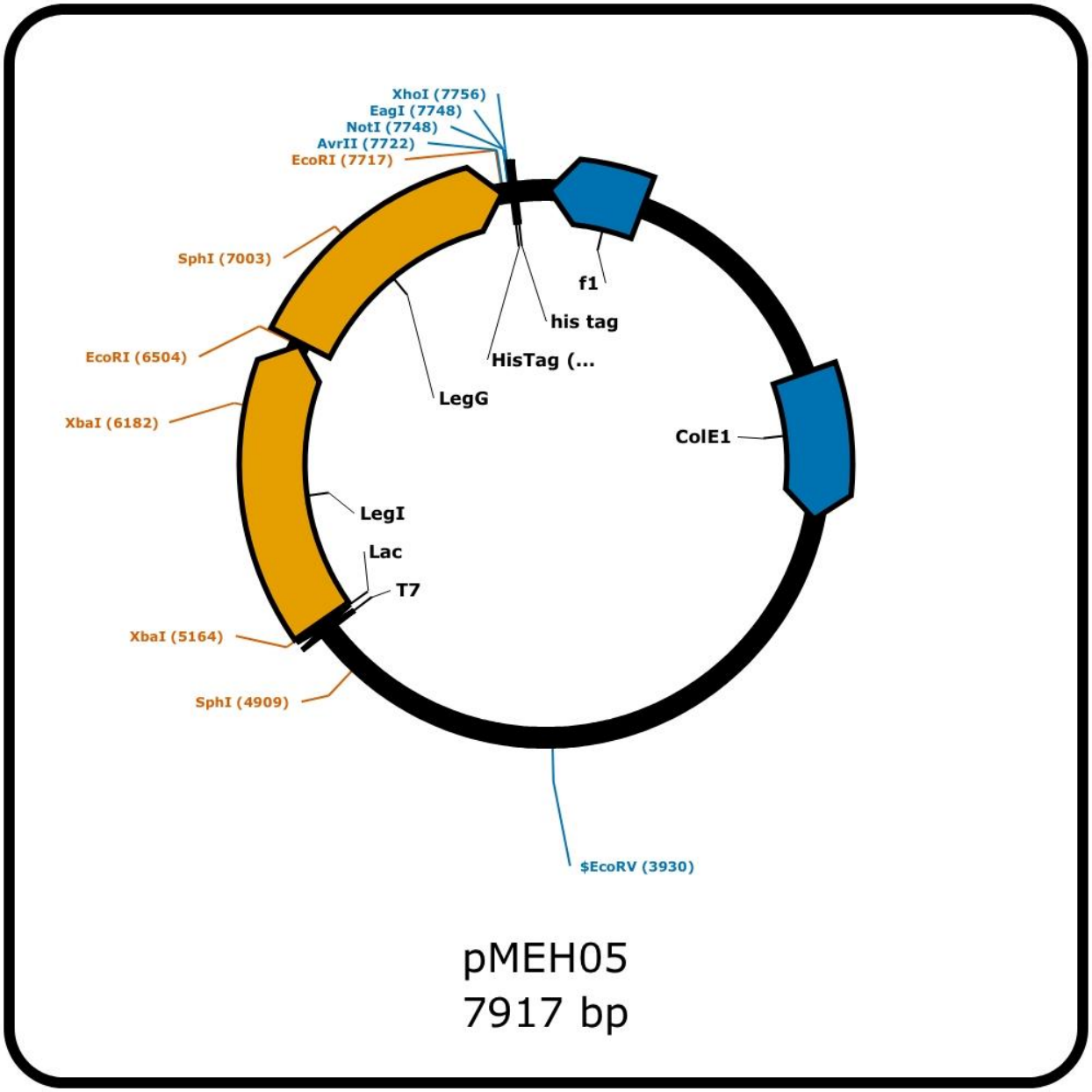


Figure II-5 pMEH05 pNRC181.1 *XbaI AvrII* into pMEH04 *AvrII*.

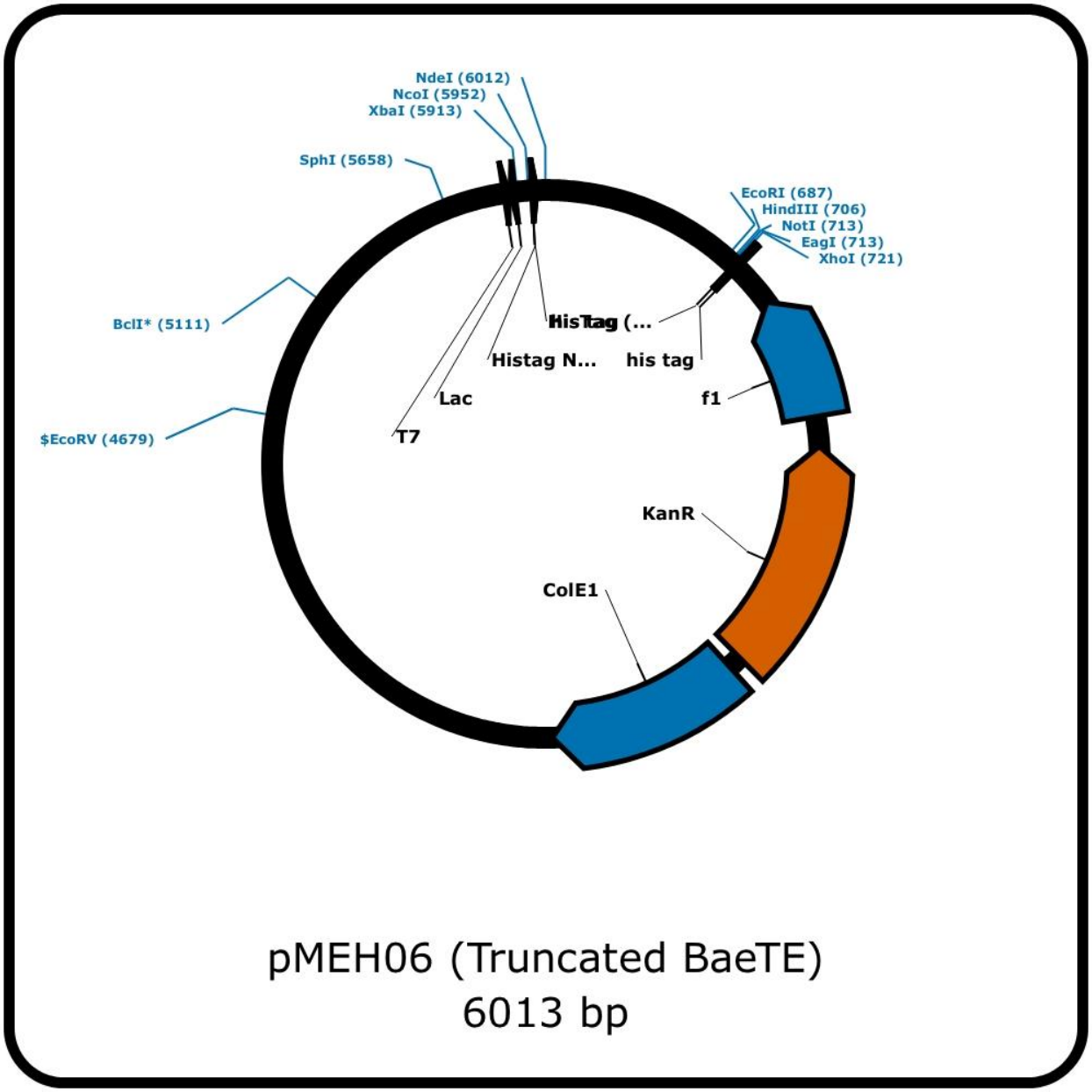


Figure II-6 pMEH06 BaetTE from pMEH02 in pET28. It has a natively encoded unintentional interal *NdeI* site.

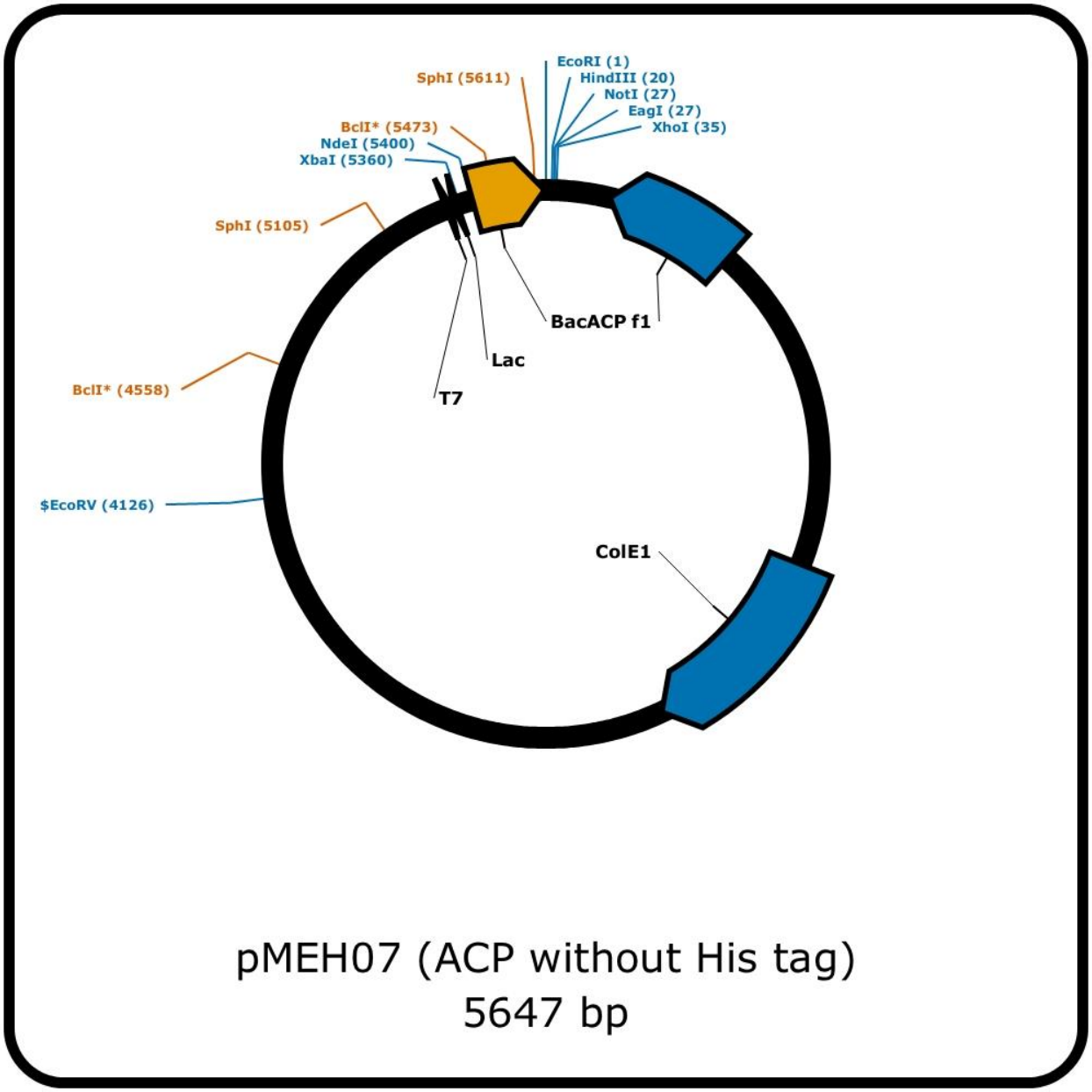


Figure II-7 pMEH07 BaeRACPIII from pMEH01 in pET21 (no His tag).
C-terminal His tag frame-shifted to proline tag.

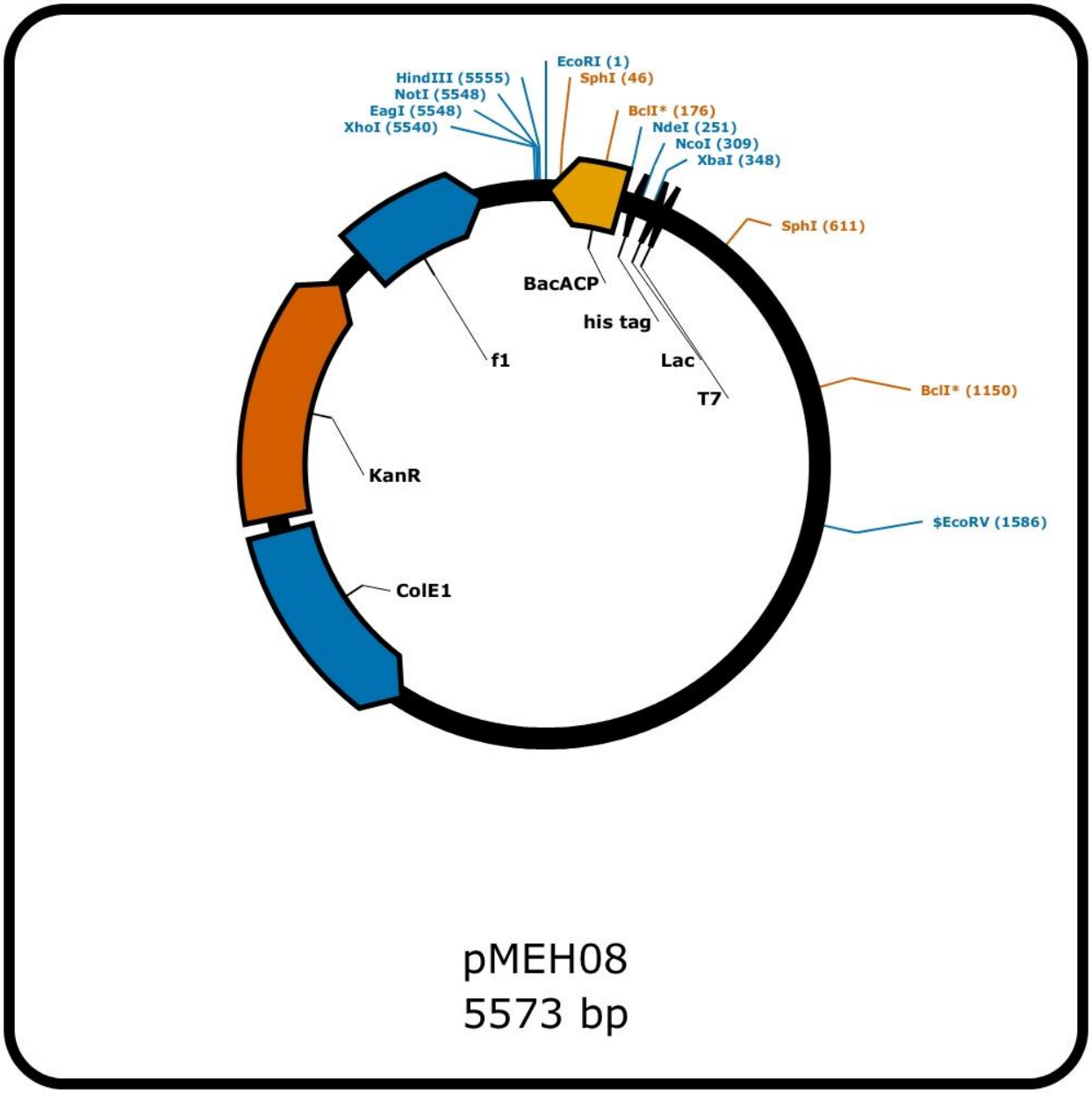


Figure II-8 pMEH08 BaeRACPIII from pMEH01 in pET28.
Construct produces purifiable protein because of N-terminal His tag despite C-terminal proline tag.

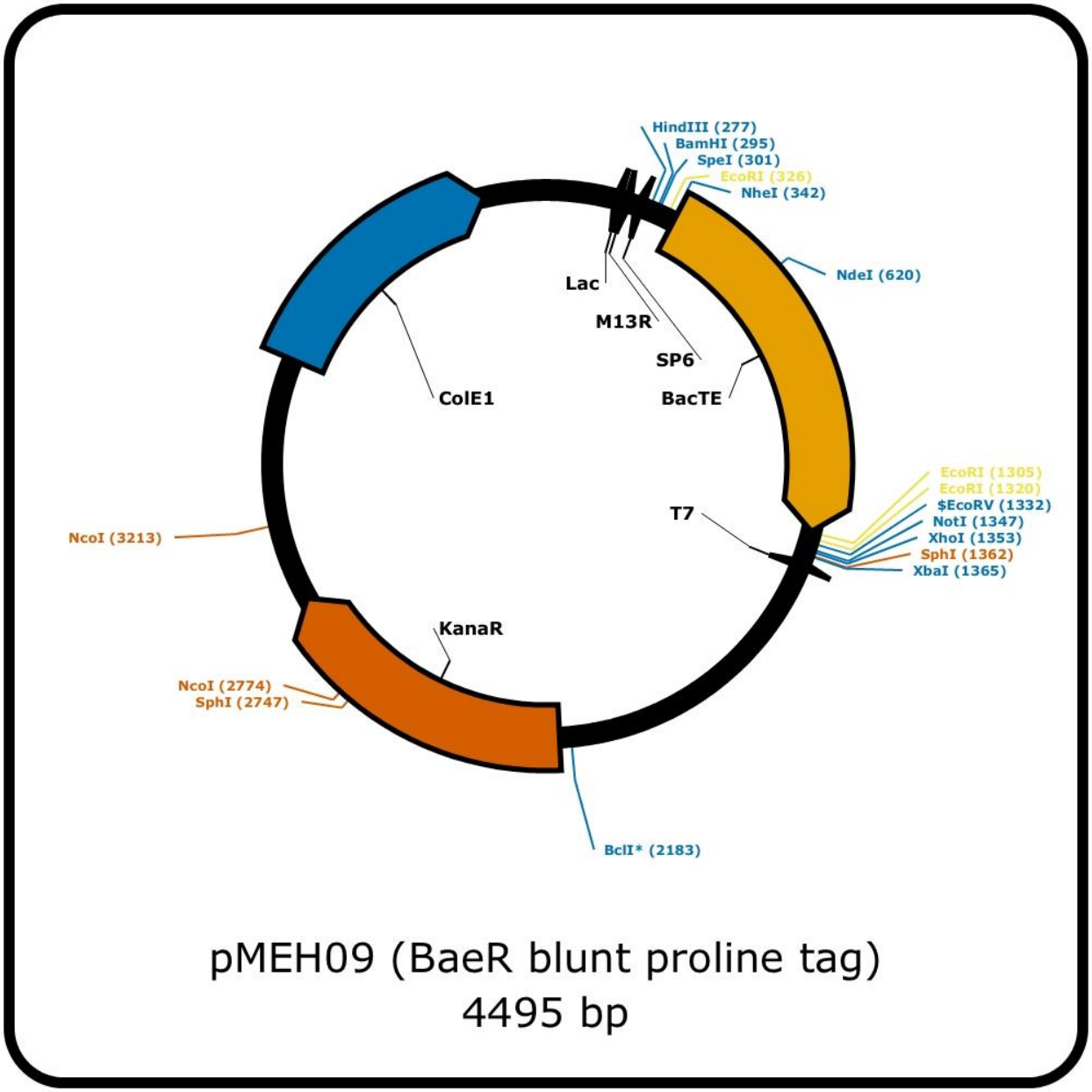


Figure II-9 pMEH09 BaeRTE PCR from pKJ13 into blunt with N-terminal *NheI* and C-terminal *EcoRI*. C-terminal frame leads to a proline tag if cloned into pET vectors.

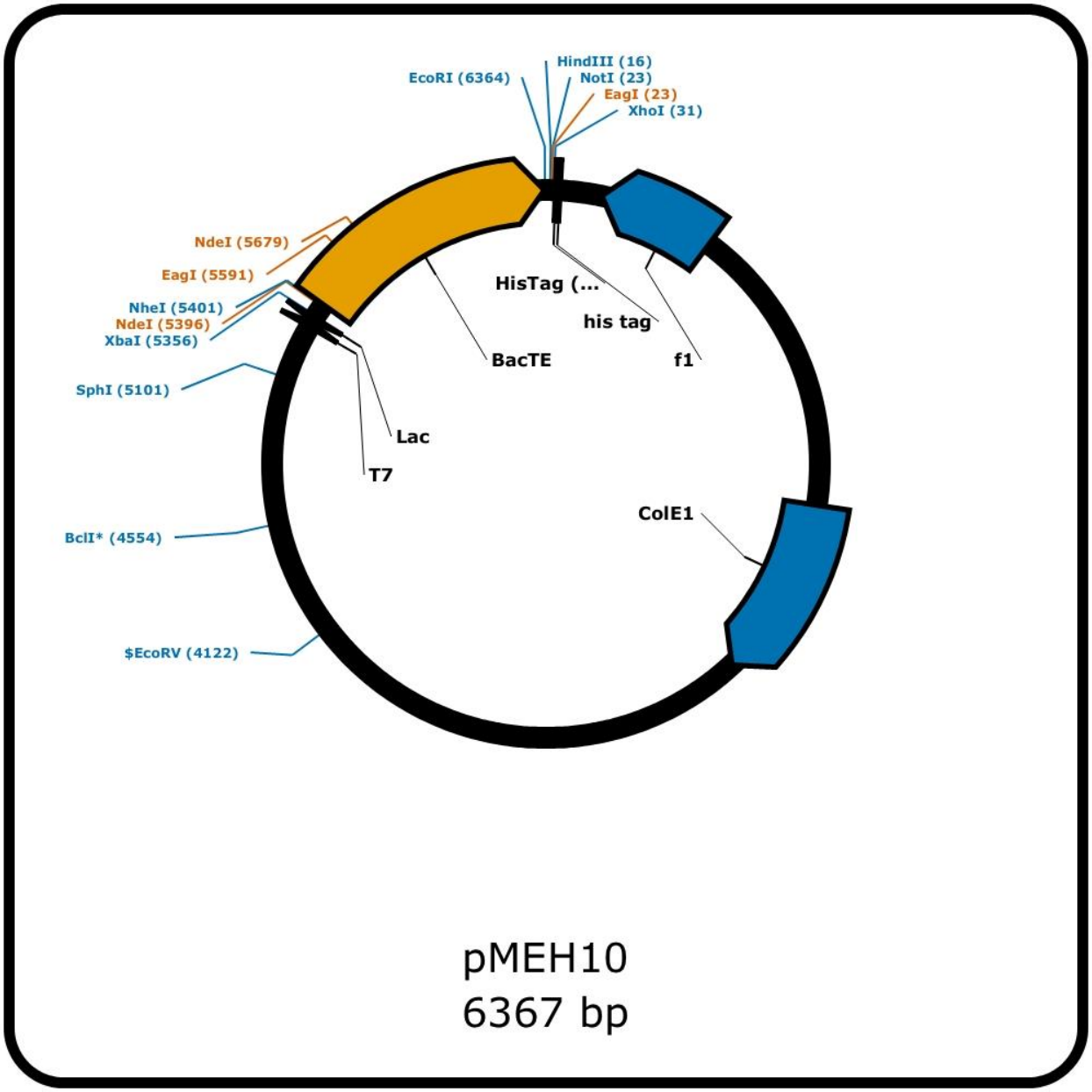


Figure II-10 pMEH10 BaeRTE from pMEH09 in pET21.

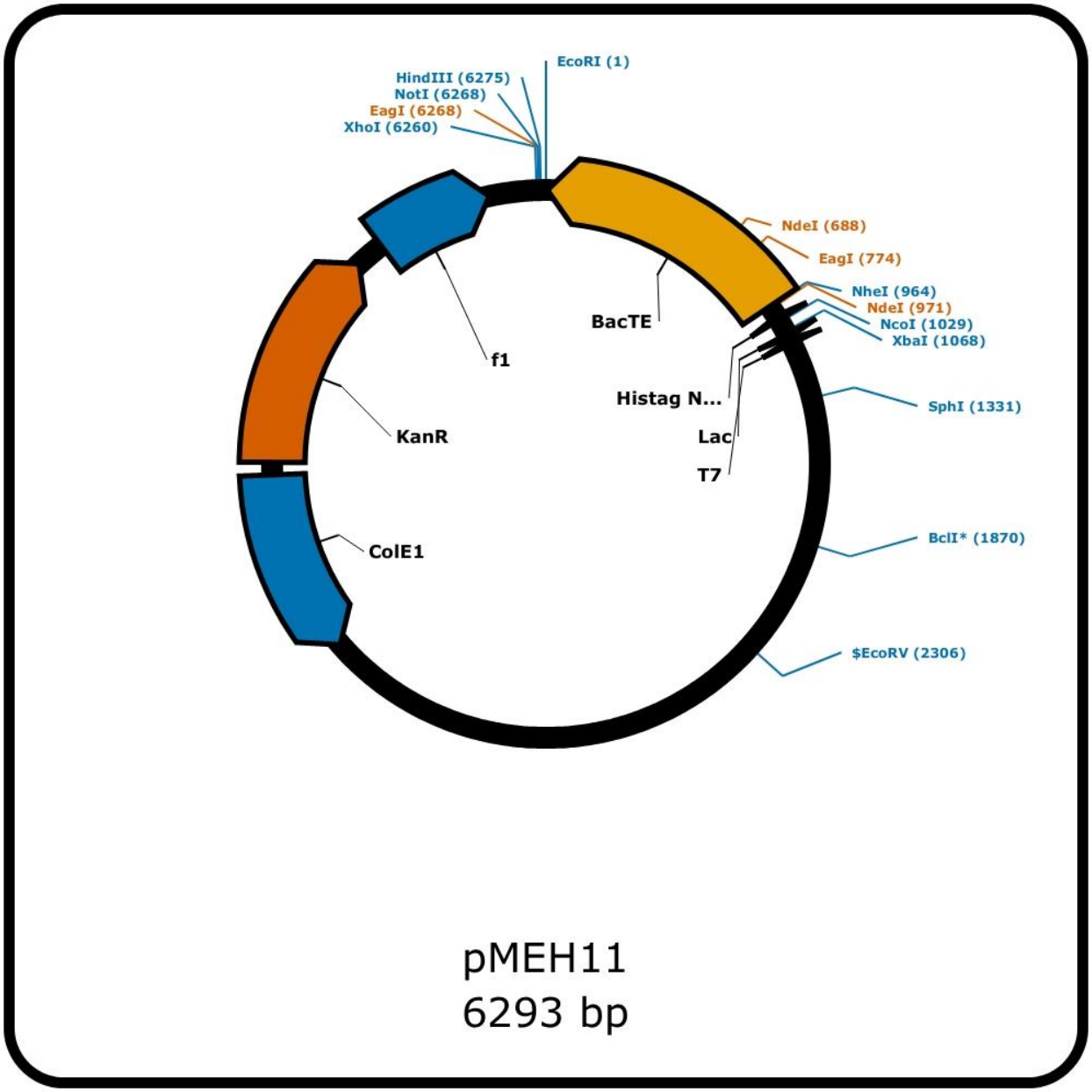


Figure II-11 pMEH11 BaeRTE from pMEH09 NheI EcoRI into pET28.

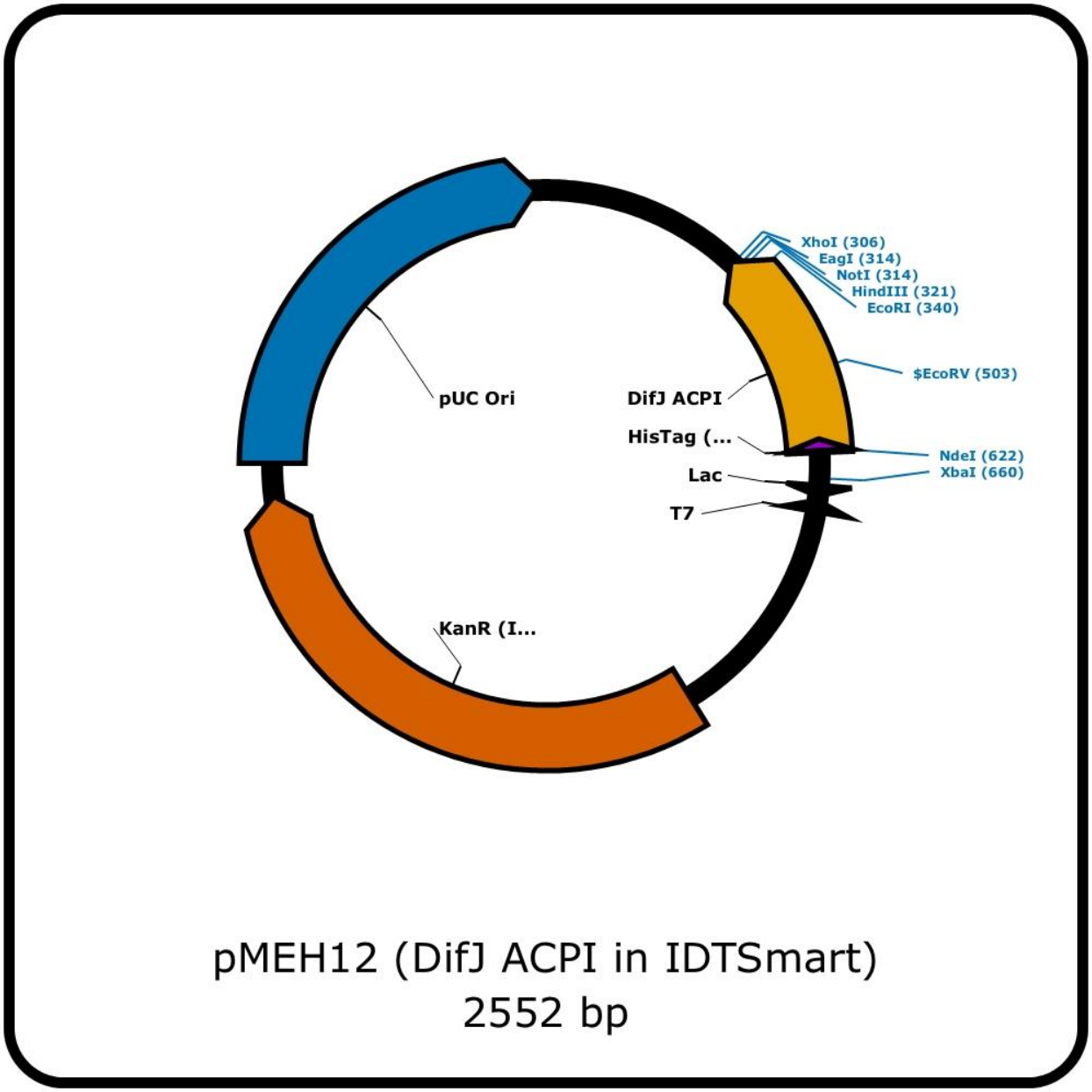


Figure II-12 pMEH12 codon optimized difficidin J ACPI in IDTSmart Kan vector.

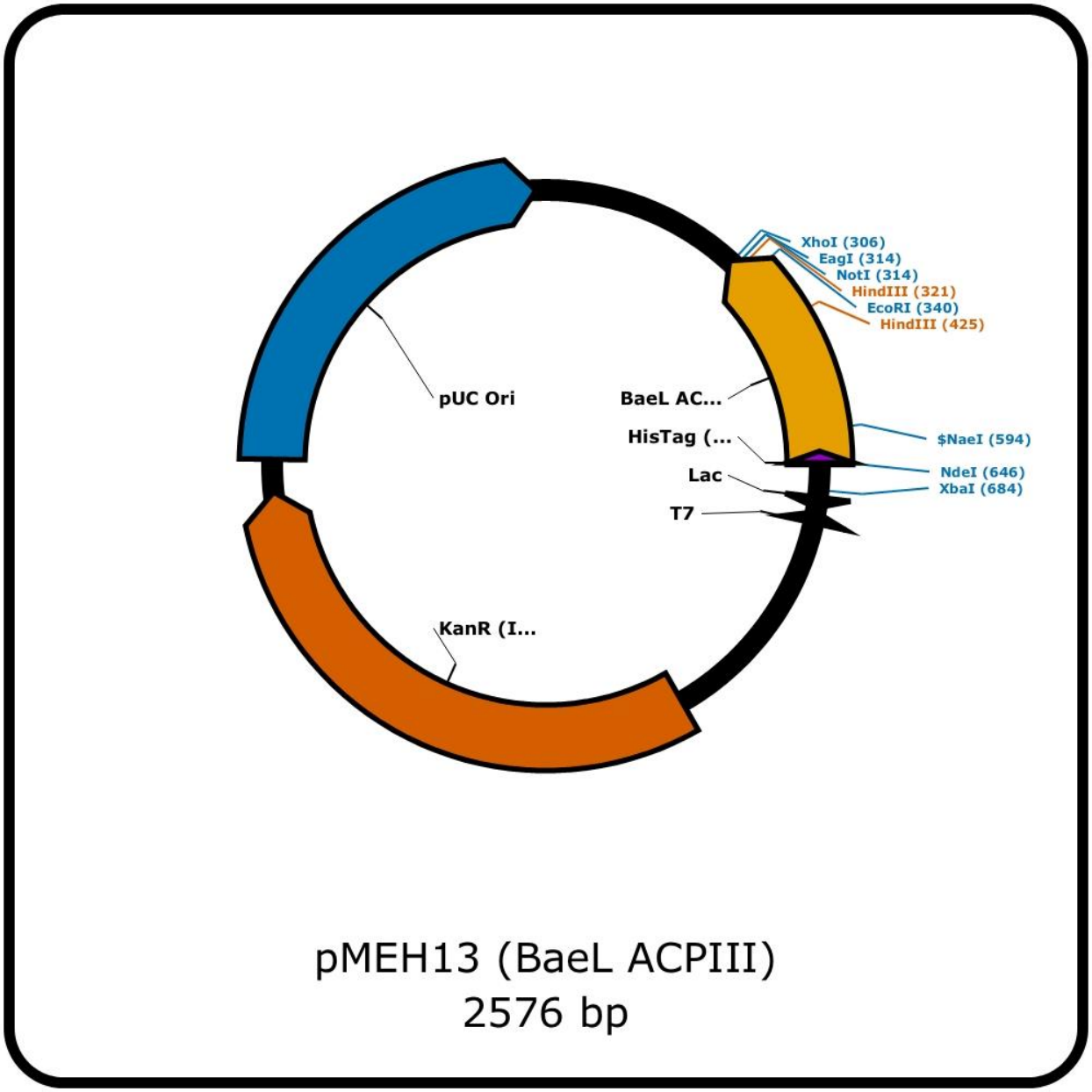


Figure II-13 pMEH13 vodon optimized BaeL ACPIII in IDTSmart Kan vector.

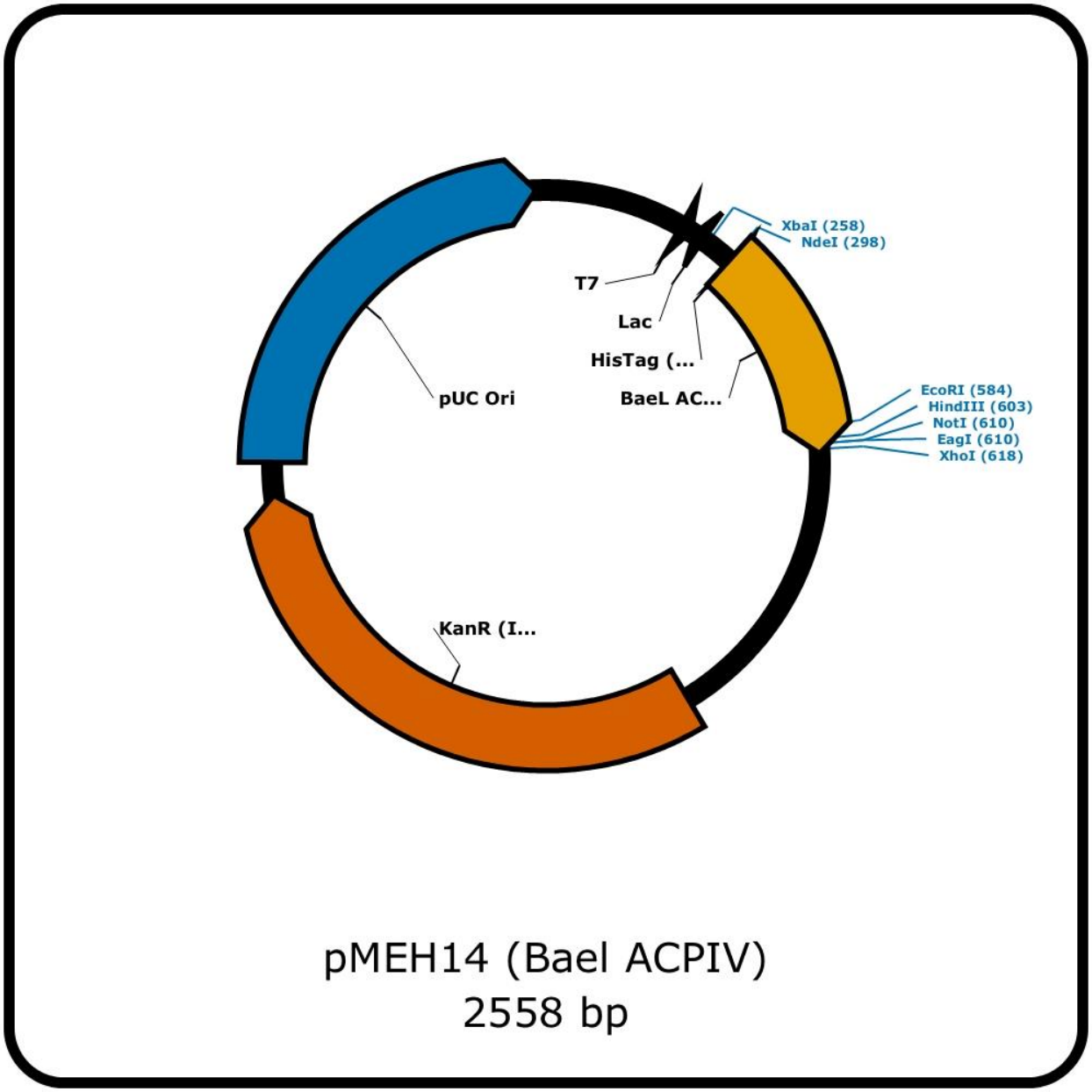


Figure II-14 pMEH14 codon optimized BaeL ACPIV in IDTSmart Kan vector.

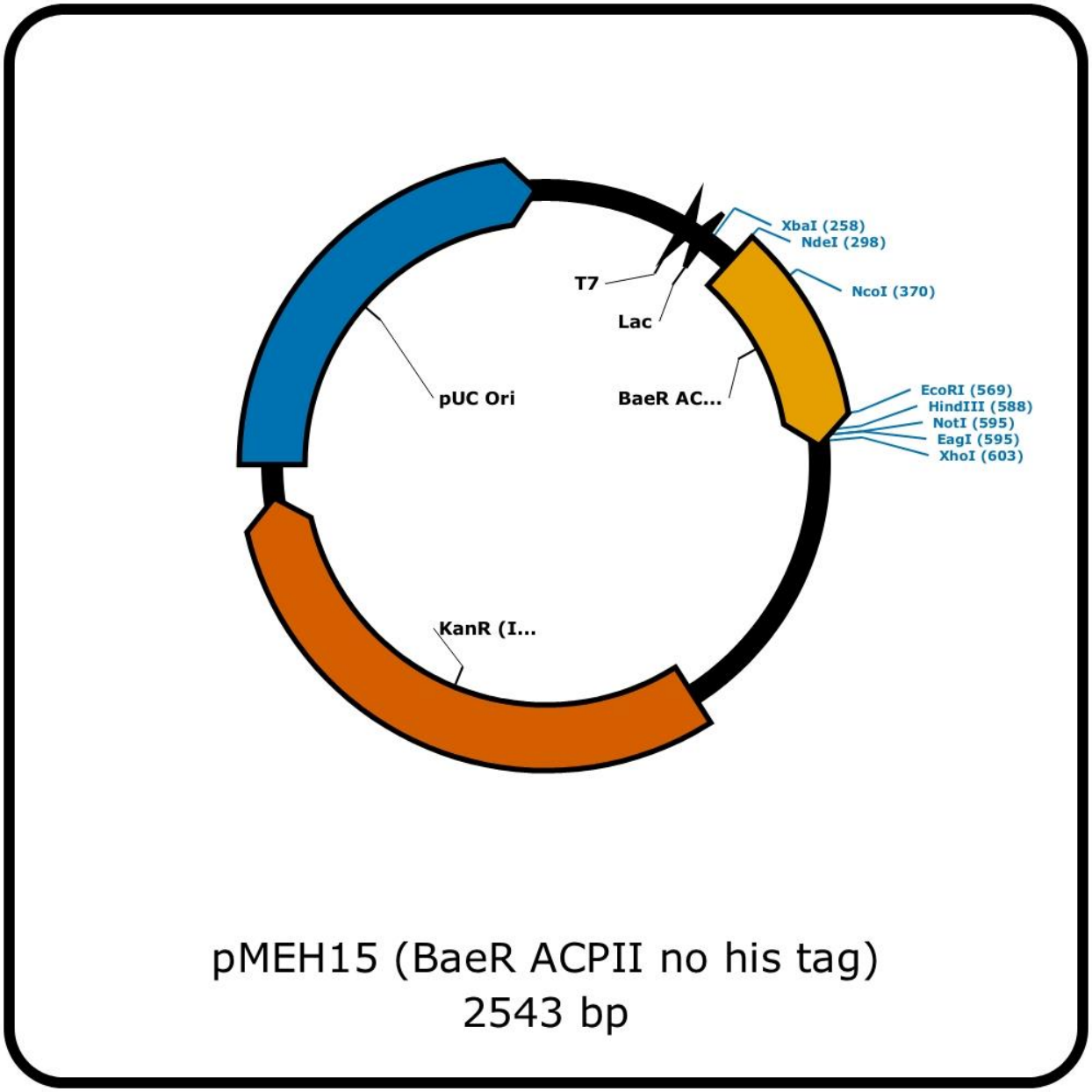


Figure II-15 pMEH15 codon optimized BaeR ACPII in IDTSmart Kan vector.

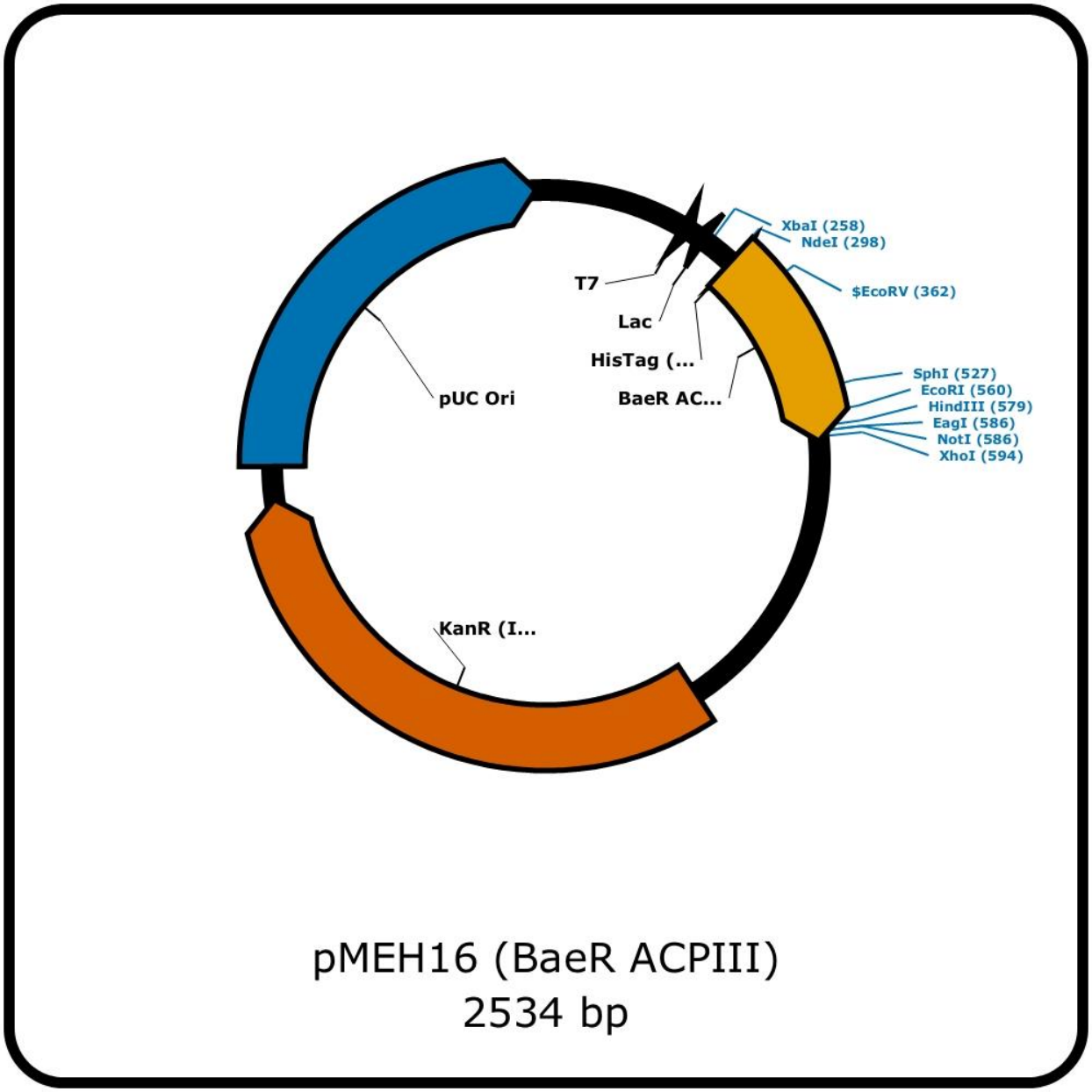


Figure II-16 pMEH16 codon optimized BaeR ACPIII in IDTSmart Kan vector.

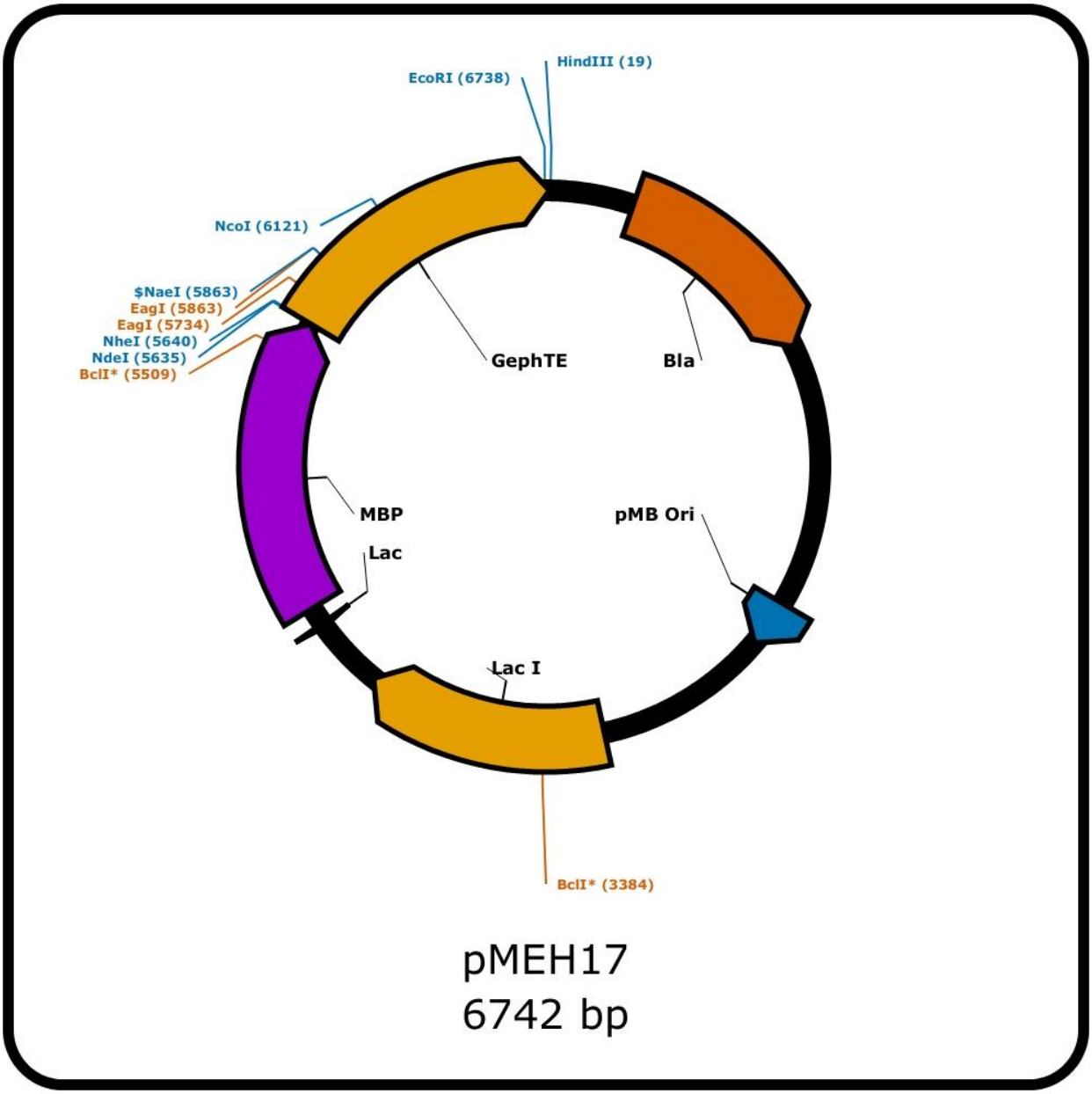


Figure II-17 pMEH17 gepheronic acid module J TE with *NdeI* and *EcoRI* restriction sites in pMalC5x. Gene fused to N-terminal maltose-binding protein. Amp resistance.

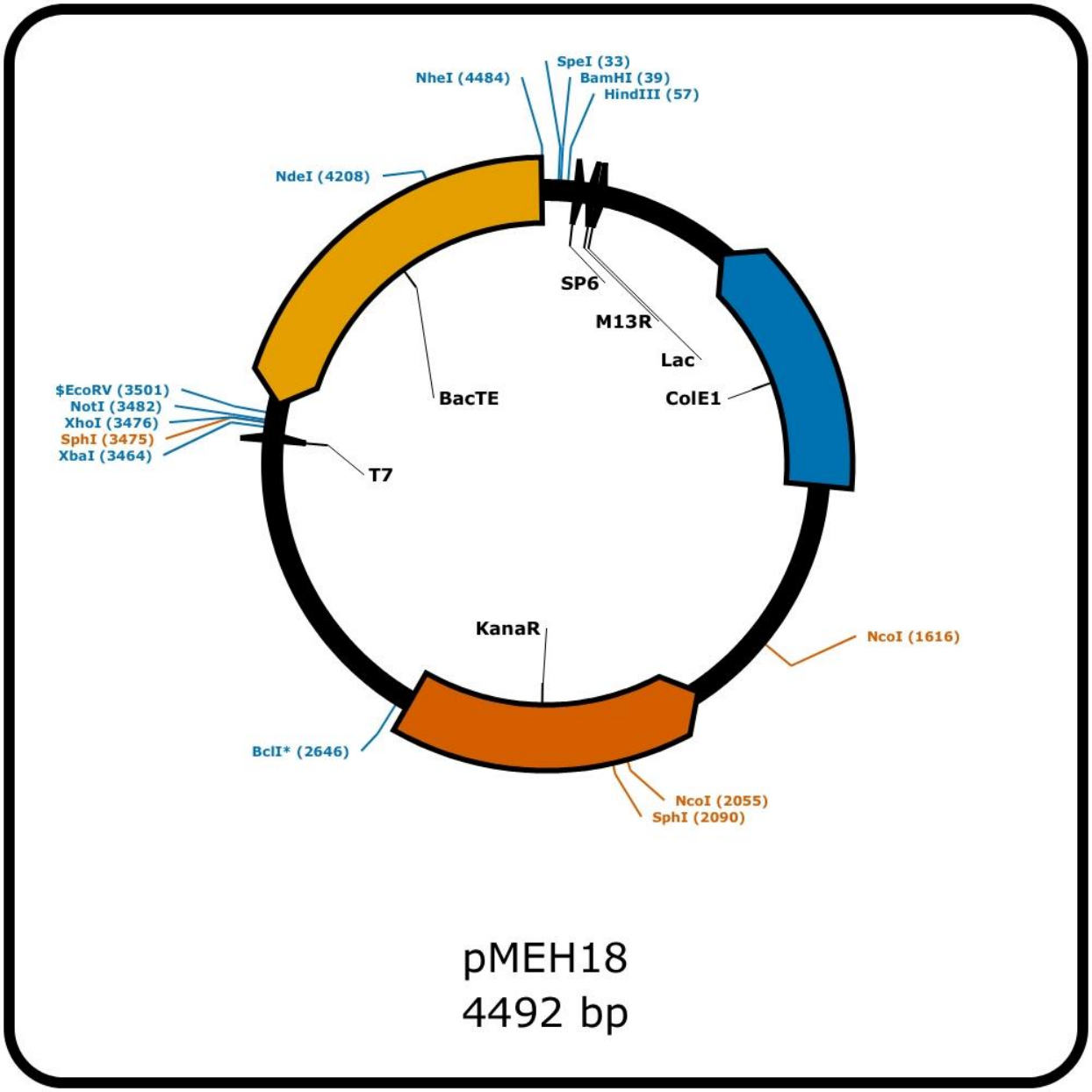


Figure II-18 pMEH18 BaeRTE *NheI EcoRI* (proper frame) PCR product from pKJ13 in blunt.

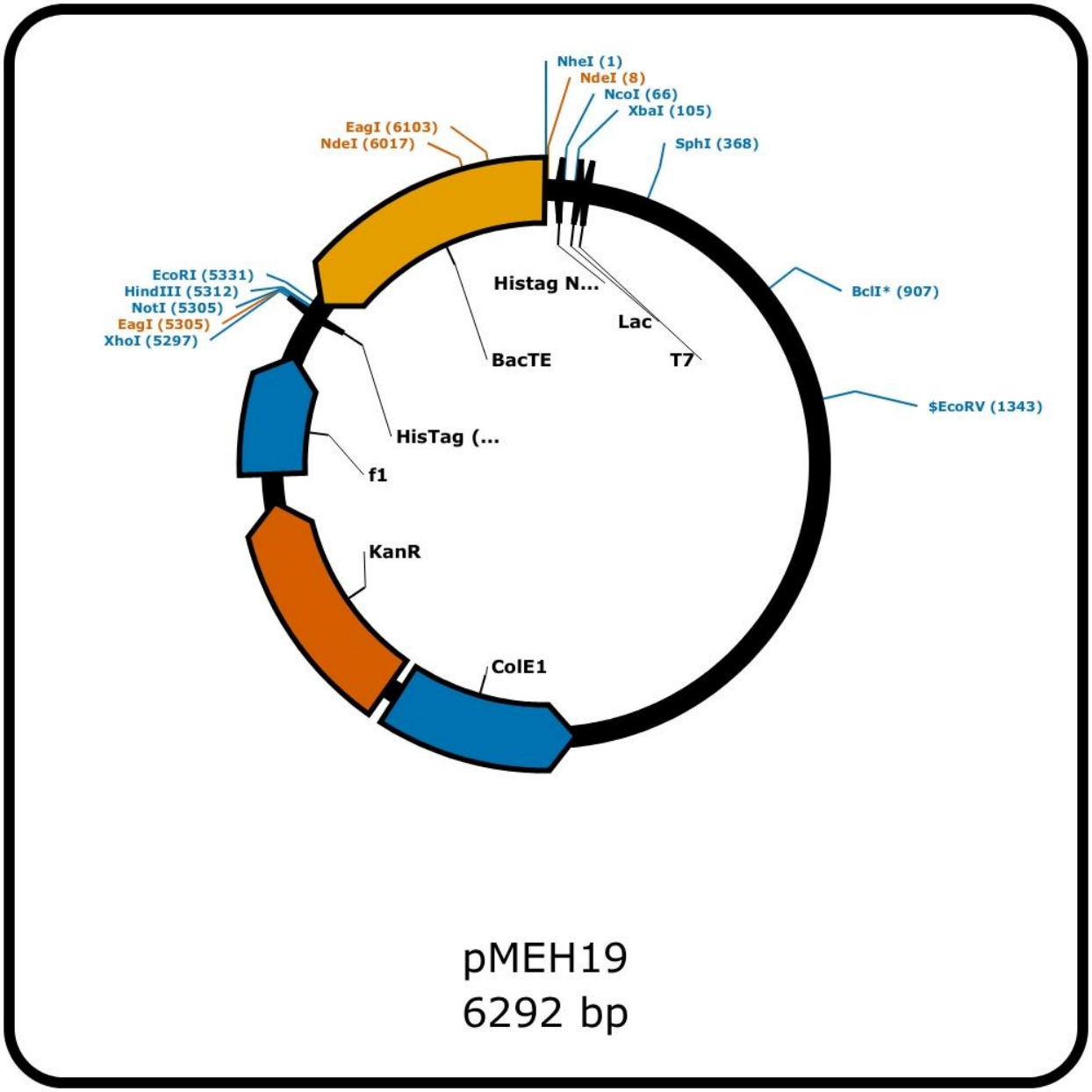


Figure II-19 pMEH19 BaeRTE from pMEH18 in pET28.

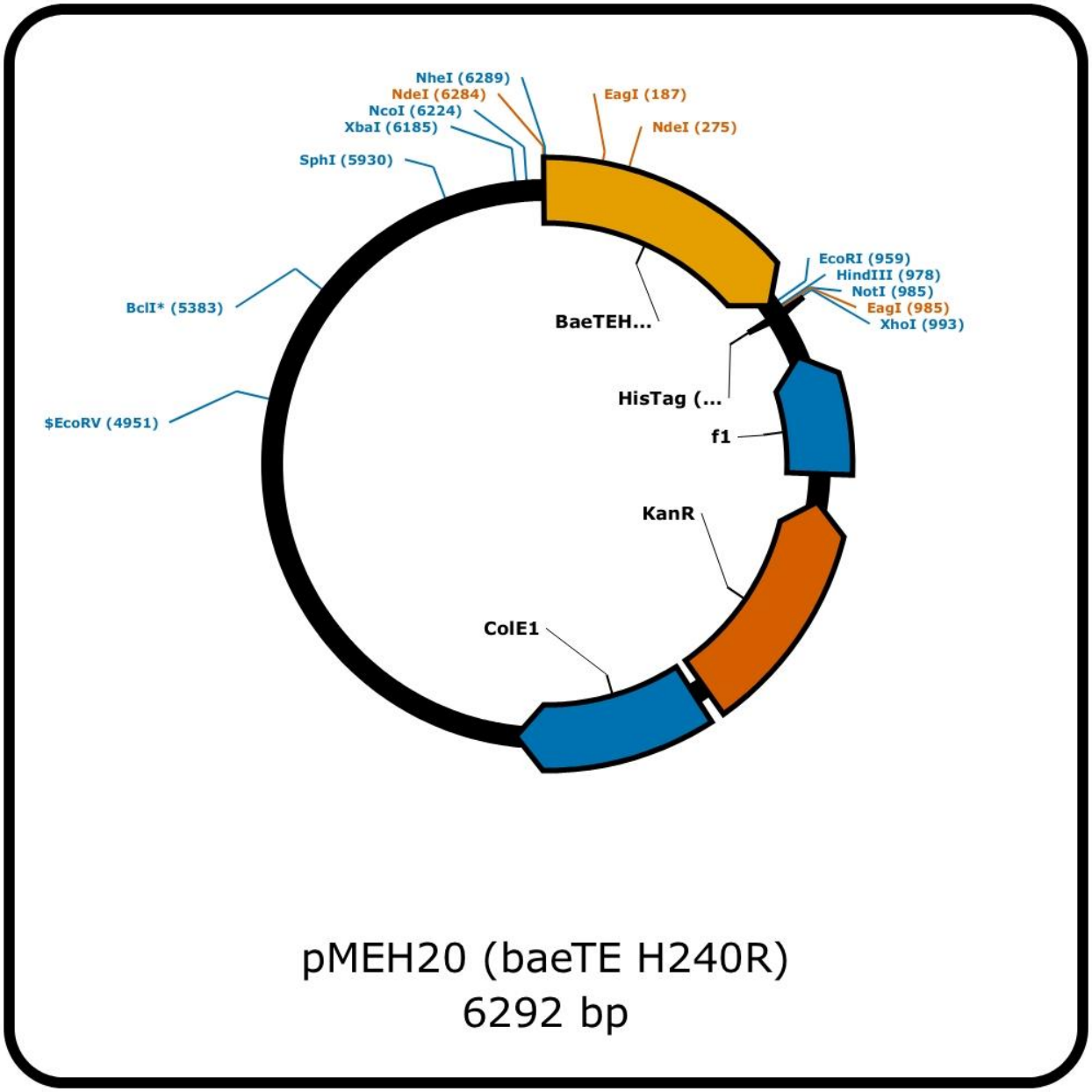


Figure II-20 pMEH20 H289R point mutant of pMEH19 BaeRTE. Quik Change II PCR mutant replacing a possibly surface-facing His with an Arg. The CAC codon was mutated at position two to introduce the Arg CGC codon. Primers BaeTE H240R f (ATTGGAAAGAAT GGGAACGCCAGATG) BaeTE H240 r (AGGTGAAACTGTTTCATCTGGCGTTCC).

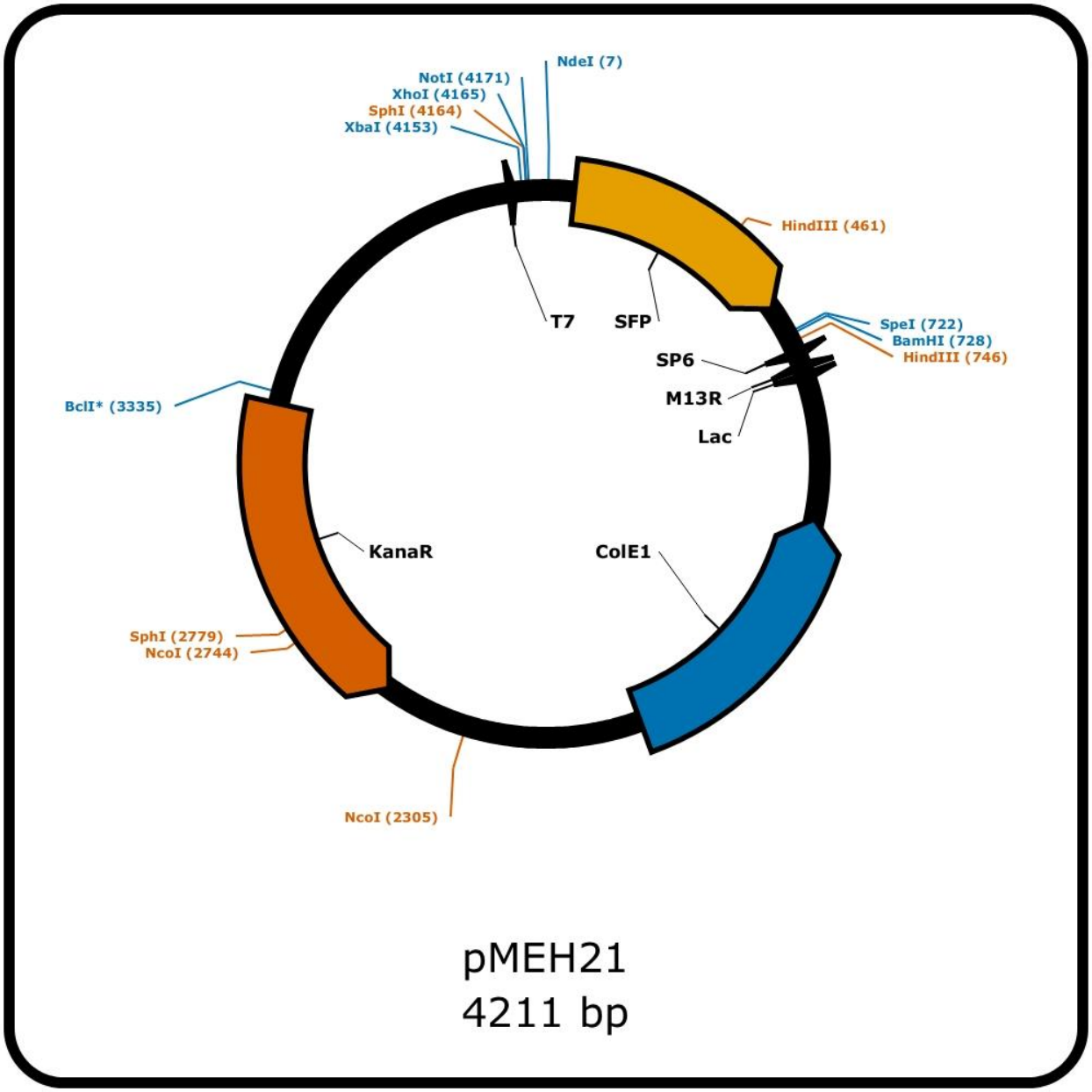


Figure II-21 pMEH21 *B. subtilis* sfp in pTOPObluntII. Colony PCR was performed on a frozen glycerol stock of BAP1 (Pfeifer *et al.* 2001)[1] using the primers BAP1_sfp_f (GATCCAT ATGAAGATTTACGGAATTTATATG) and BAP1_sfp_r (GATCGAATTC AAAAGCTCTTCGTACG), with sequence after space representing the annealing region. Amplified region is roughly 173 to 841 from http://www.ncbi.nlm.nih.gov/nuccore/40138.

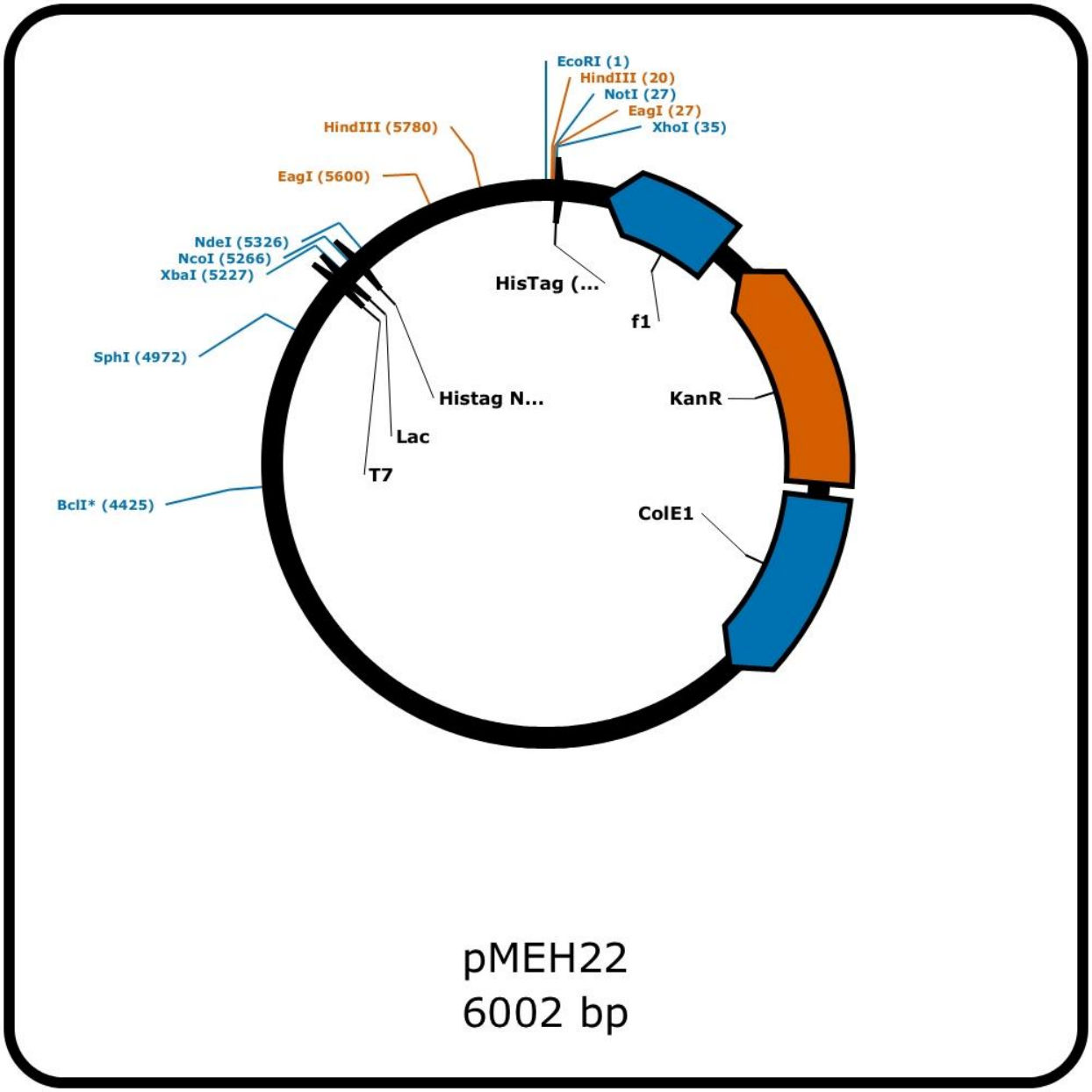


Figure II-22 pMEH22 *B. subtilis* sfp in pET28. Restriction digest with *NdeI* and *EcoRI*. N-terminal His tag only.

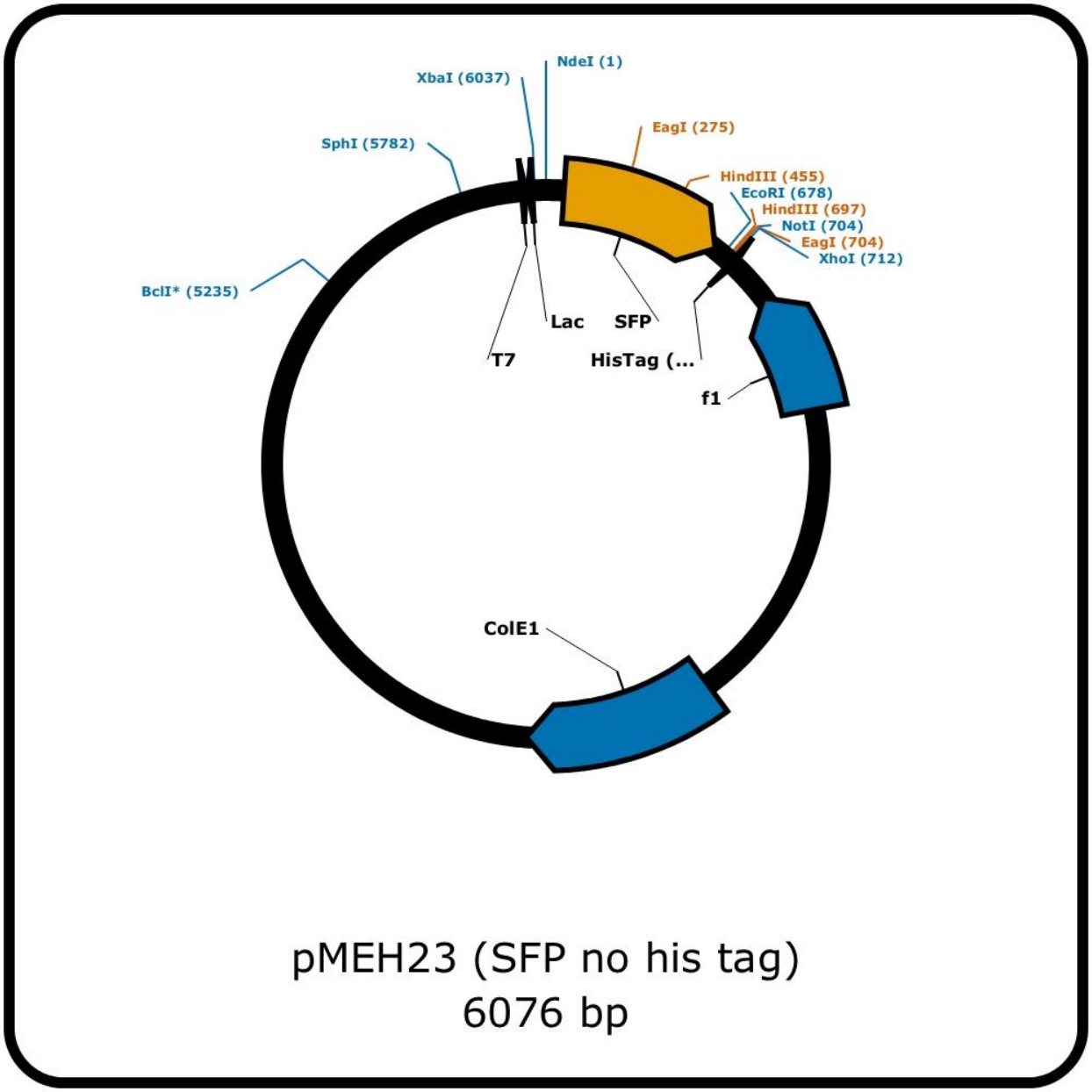


Figure II-23 pMEH23 *B. subtilis* sfp in pET21. Restriction digest with *NdeI* and *EcoRI*. No His tag; C-terminal ends in stop codon.

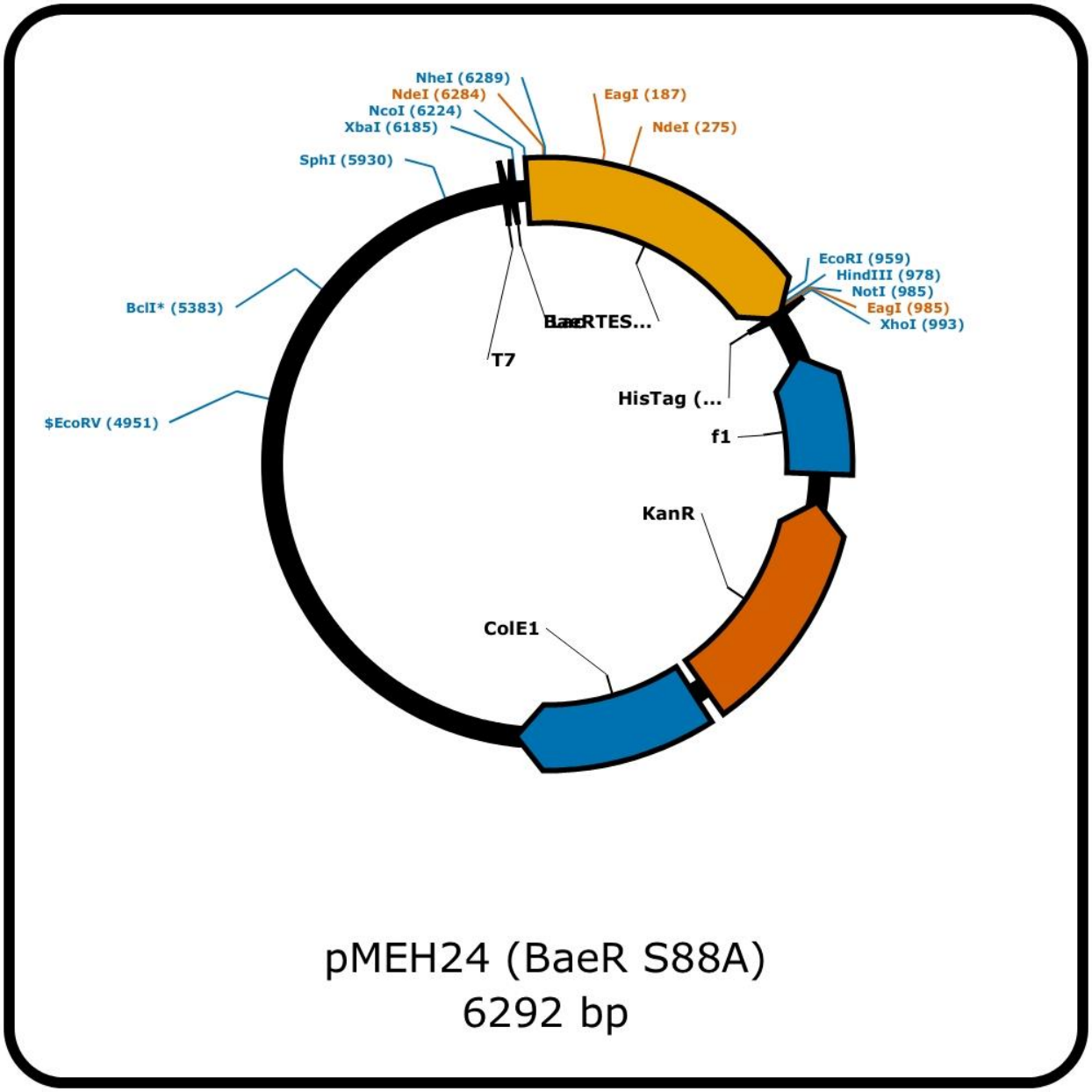


Figure II-24 pMEH24 BaeTE S88A point mutant. Quik Change II PCR mutant of active-site Ser to Ala. Restriction digest with *NdeI* and *EcoRI*. No C-terminal His tag; C-terminal ends in stop codon. The Ser 88 TCC codon was mutated at position one to introduce the Ala GCC codon. BaeRTE S88A was generated using the QuikChange II (Agilent)-type primers BaeRTE S88A f (CCTTATGATCTCGGCGGATATGCCTTGGGCGGAATG) and BaeRTE S88A r (CATATGCAAGCATTCCGCCCAAGGCATATCCGCCGAGATC)

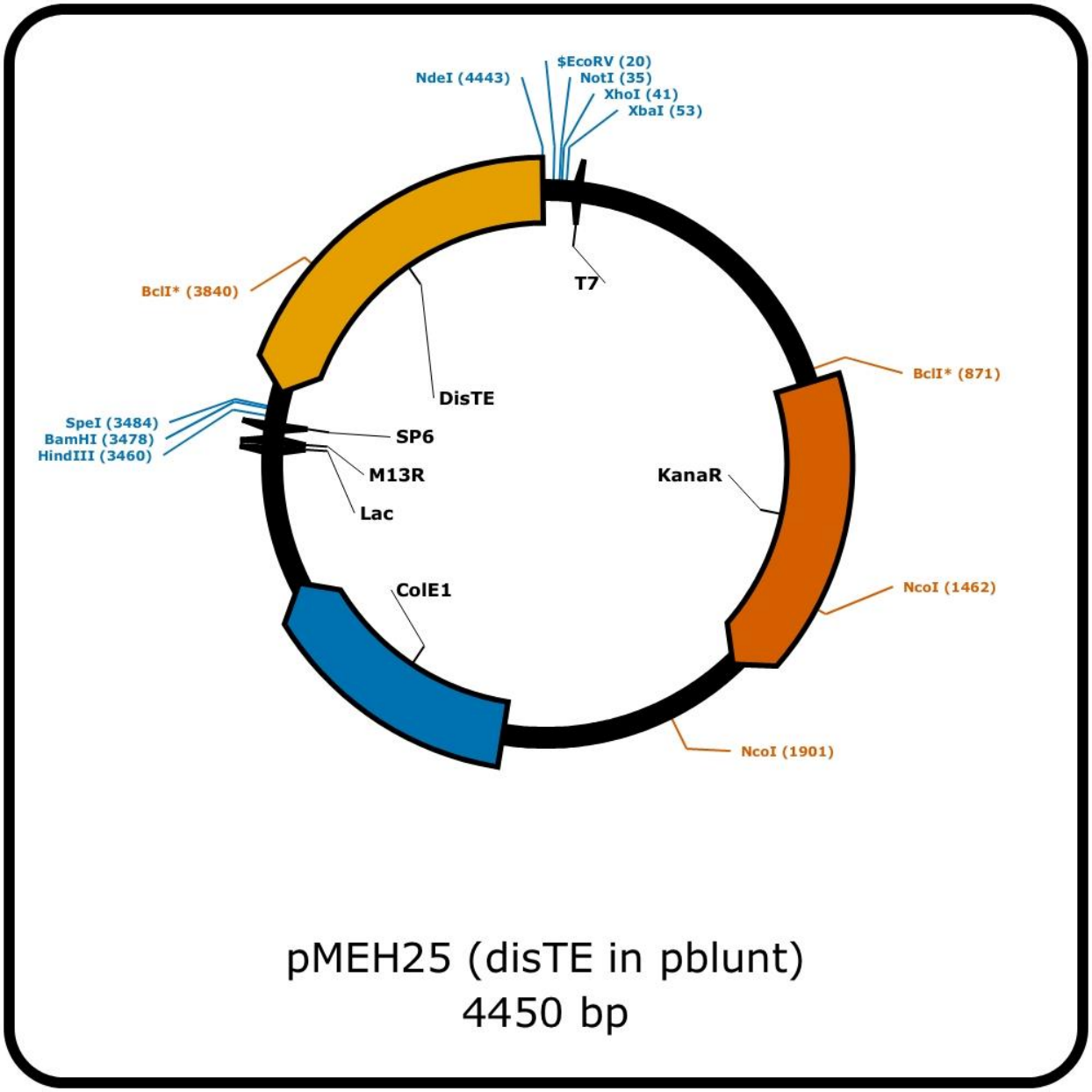


Figure II-25 pMEH25 disorazole TE in blunt. Quik Change II PCR mutant.

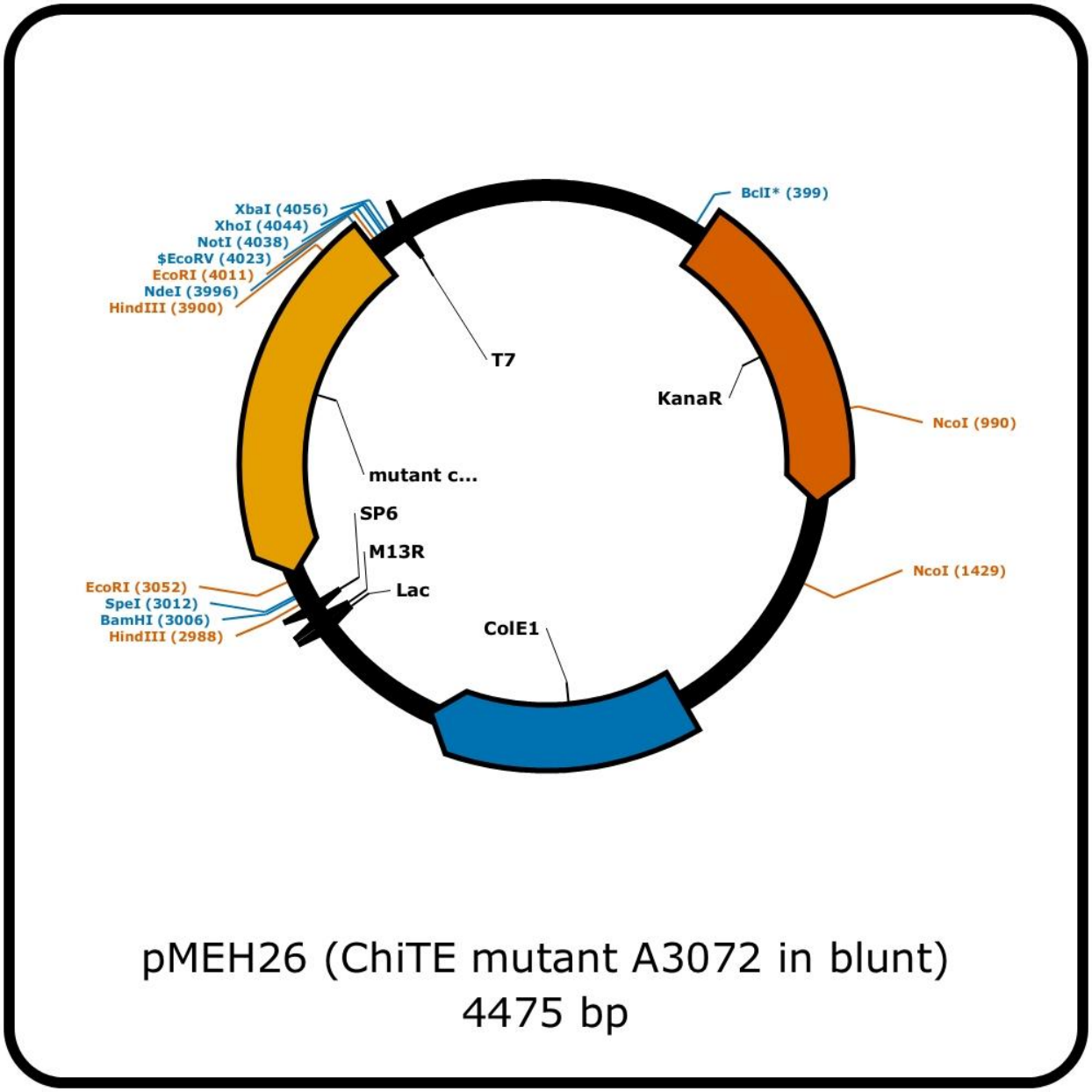


Figure II-26 pMEH26 chivosazole TE3072 mutant in blunt. Invitrogen GeneArt String codon optimized ChiTE ligated into blunt. After sequencing "GTATTCTGGAA" region of design sequenced as "GTA_TCTGGAA", indicating a frame-shift just upstream of *EcoRI* site.

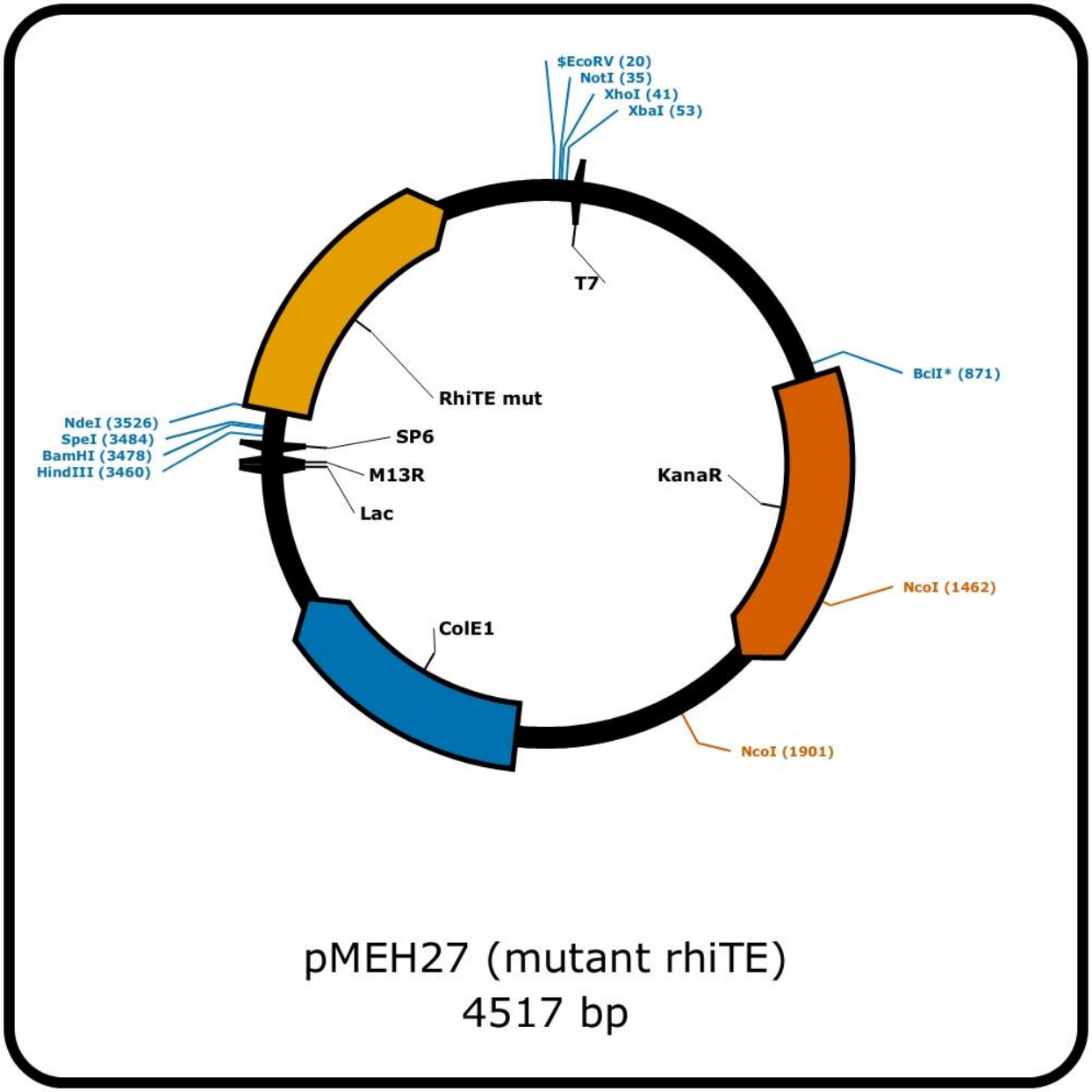


Figure II-27 pMEH27 rhizoxin TE insertion mutant in blunt. Invitrogen GeneArt String codon optimized RhiTE ligated into blunt. After sequencing, extra T detected in clone, causing frameshift.

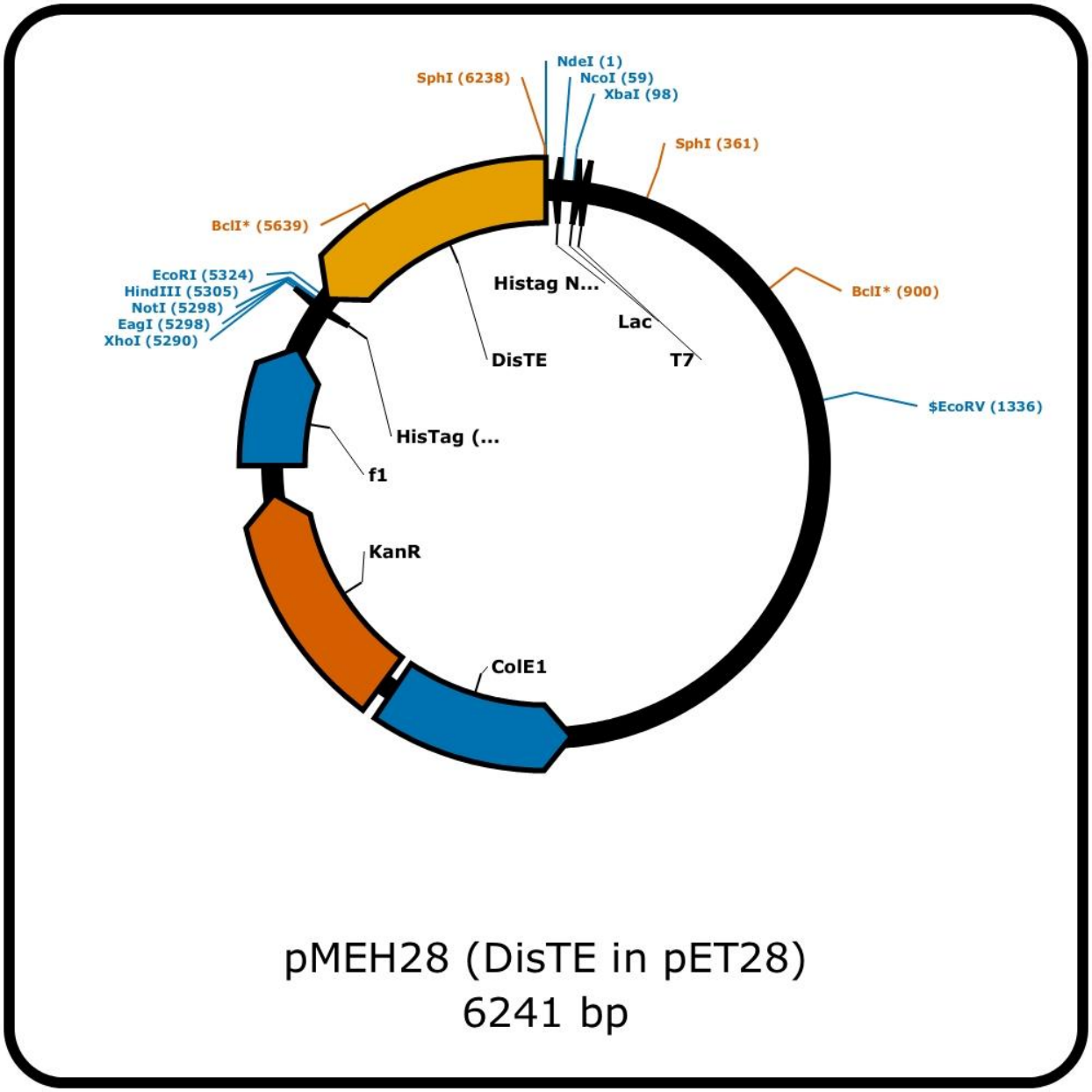


Figure II-28 pMEH28 DisTE in pET28. *NdeI* and *EcoRI* sites used to transfer insert from pMEH25 into pET28.

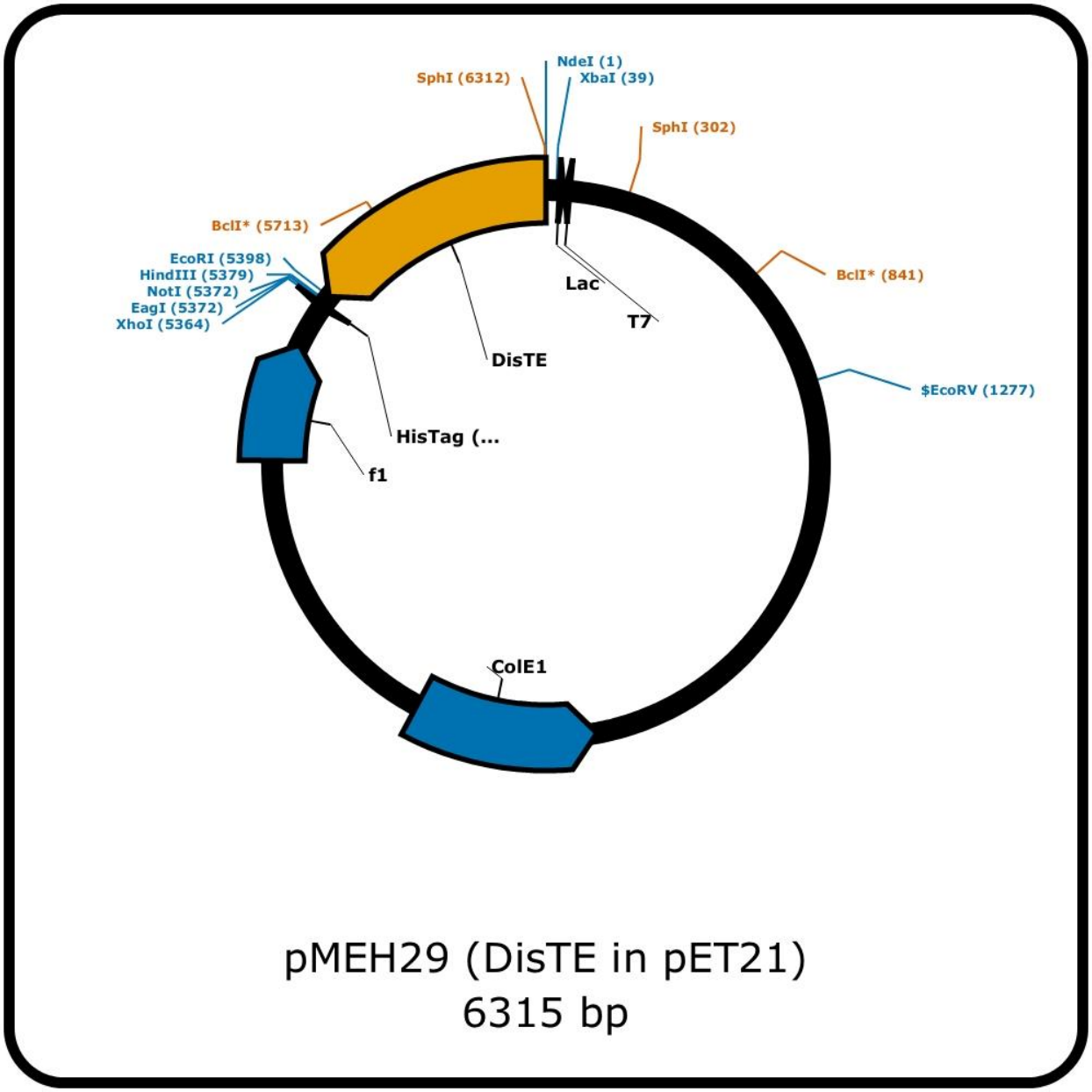


Figure II-29 pMEH29 DisTE in pET21. *NdeI* and *EcoRI* sites used to transfer insert from pMEH25 into pET21.

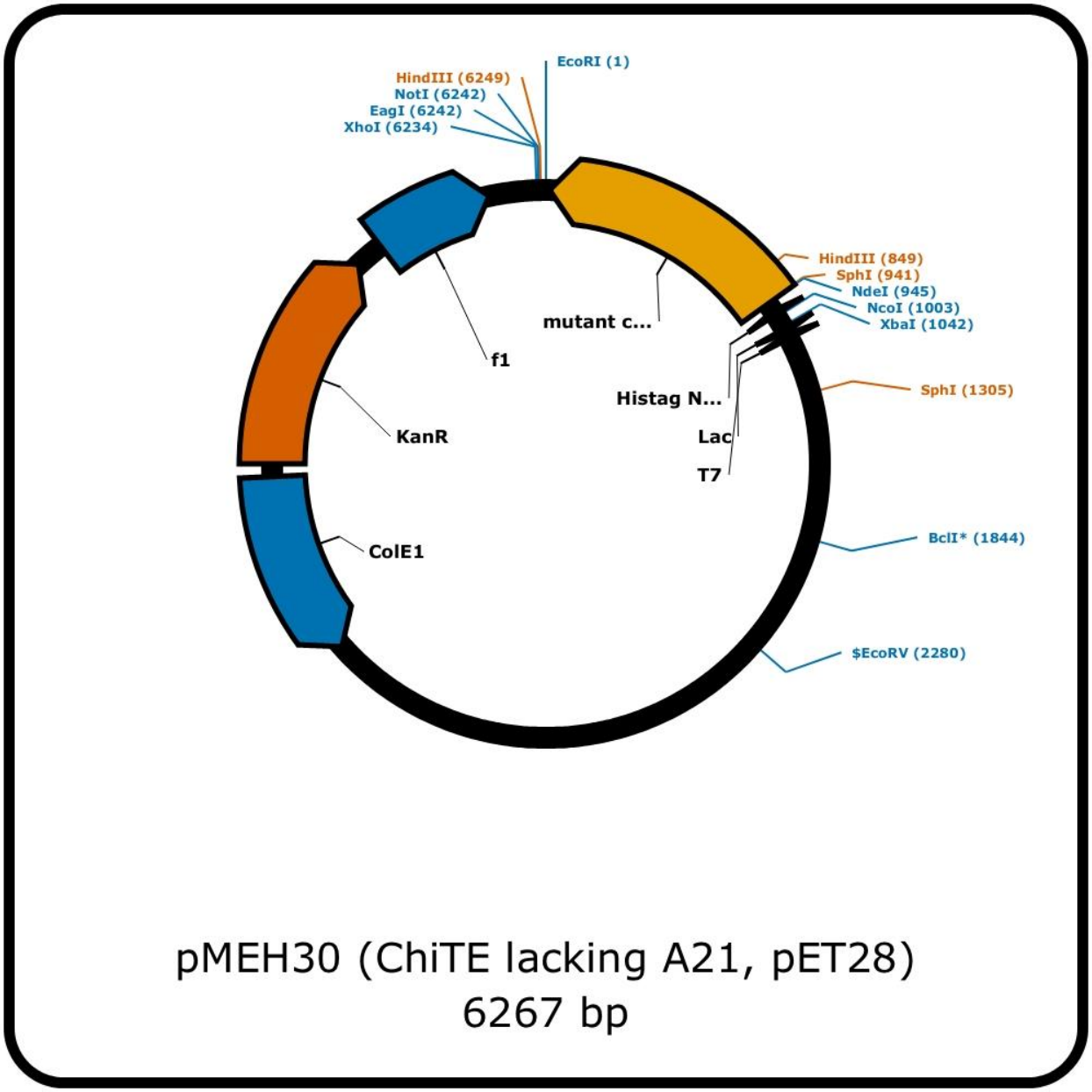


Figure II-30 pMEH30 ChiTE in pET28. *NdeI* and *EcoRI* sites used to transfer insert from pMEH26 into pET28, but frameshift mutant.

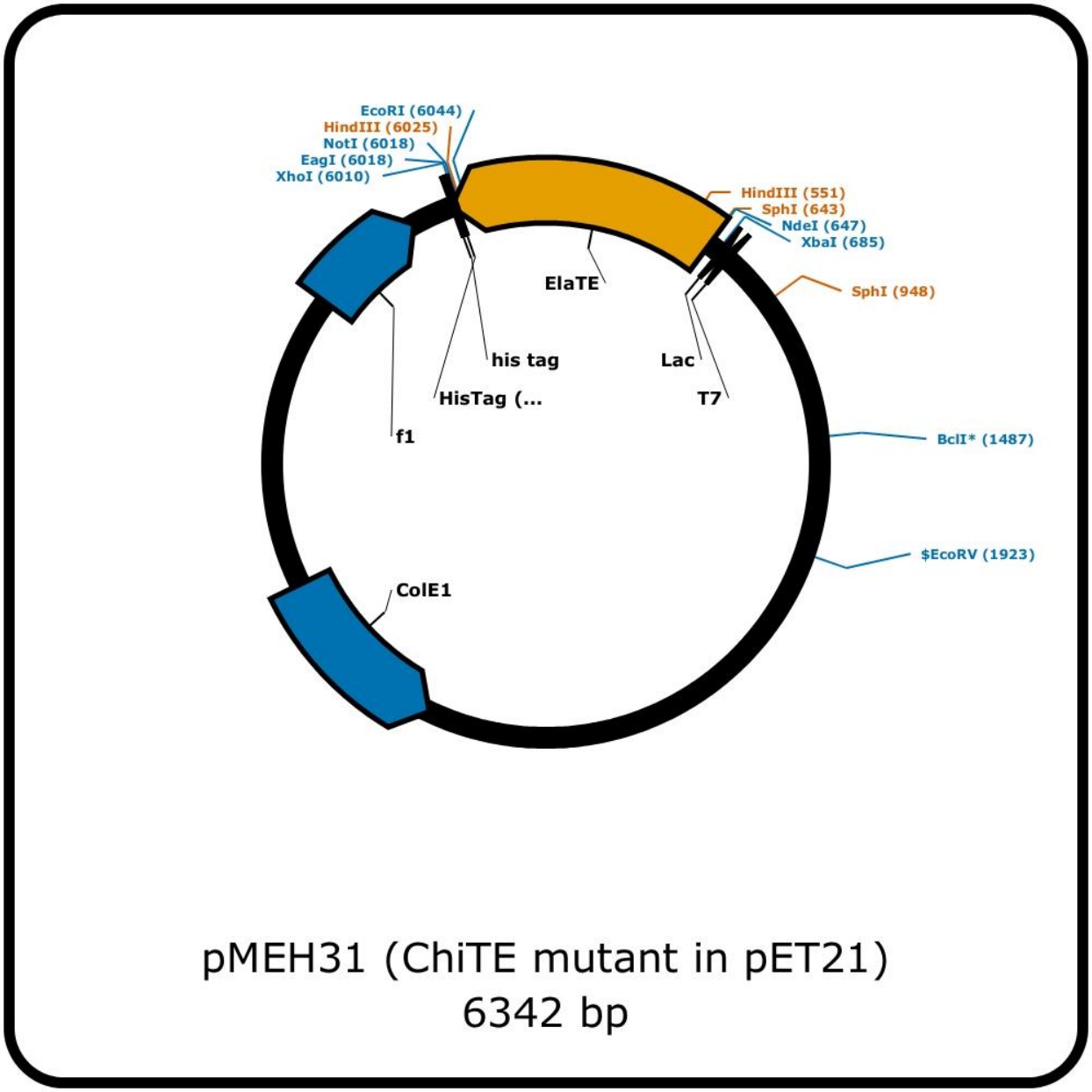


Figure II-31 pMEH31 ChiTE in pET21. *NdeI* and *EcoRI* sites used to transfer insert from pMEH26 into pET21, but frameshift mutant.

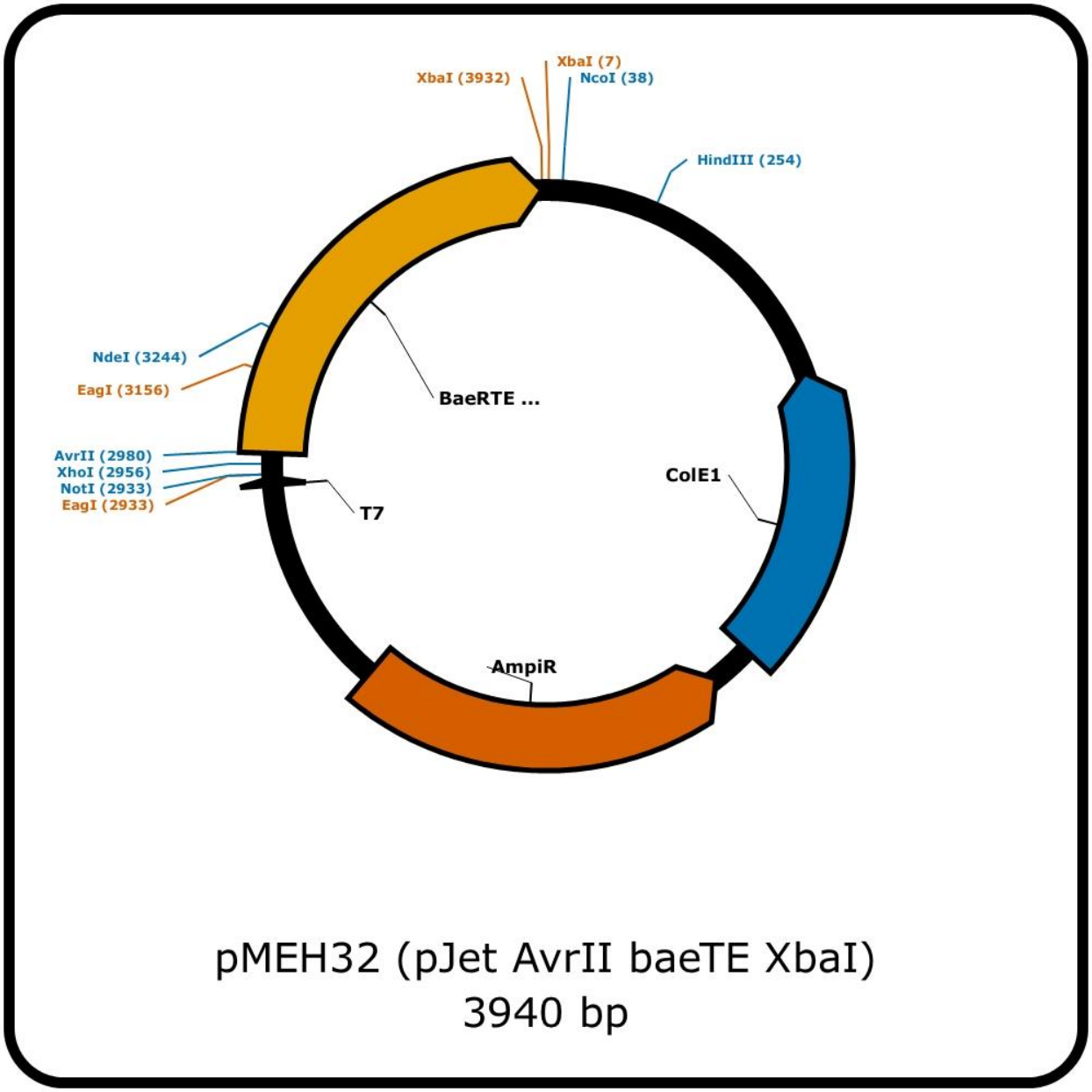


Figure II-32 pMEH32 BaeTE from pMEH19 in pJet. *XbaI* and *AvrII* cut sites inserted via PCR.

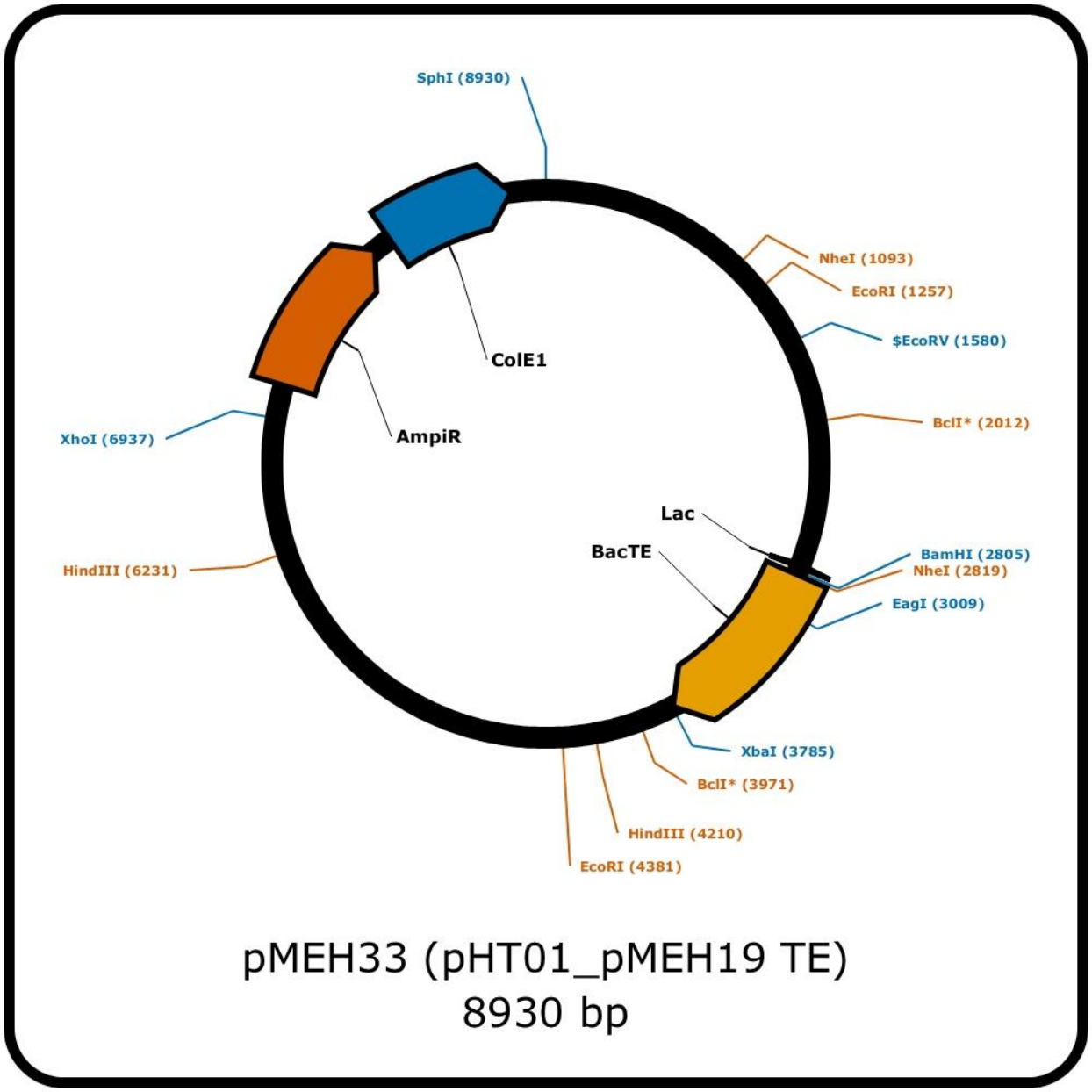


Figure II-33 pMEH33 BaeTE from pMEH32 in pHT01. Cloned into mBio multispecies vector via *BamHI* and XbaI for expression in *Bacillus.*

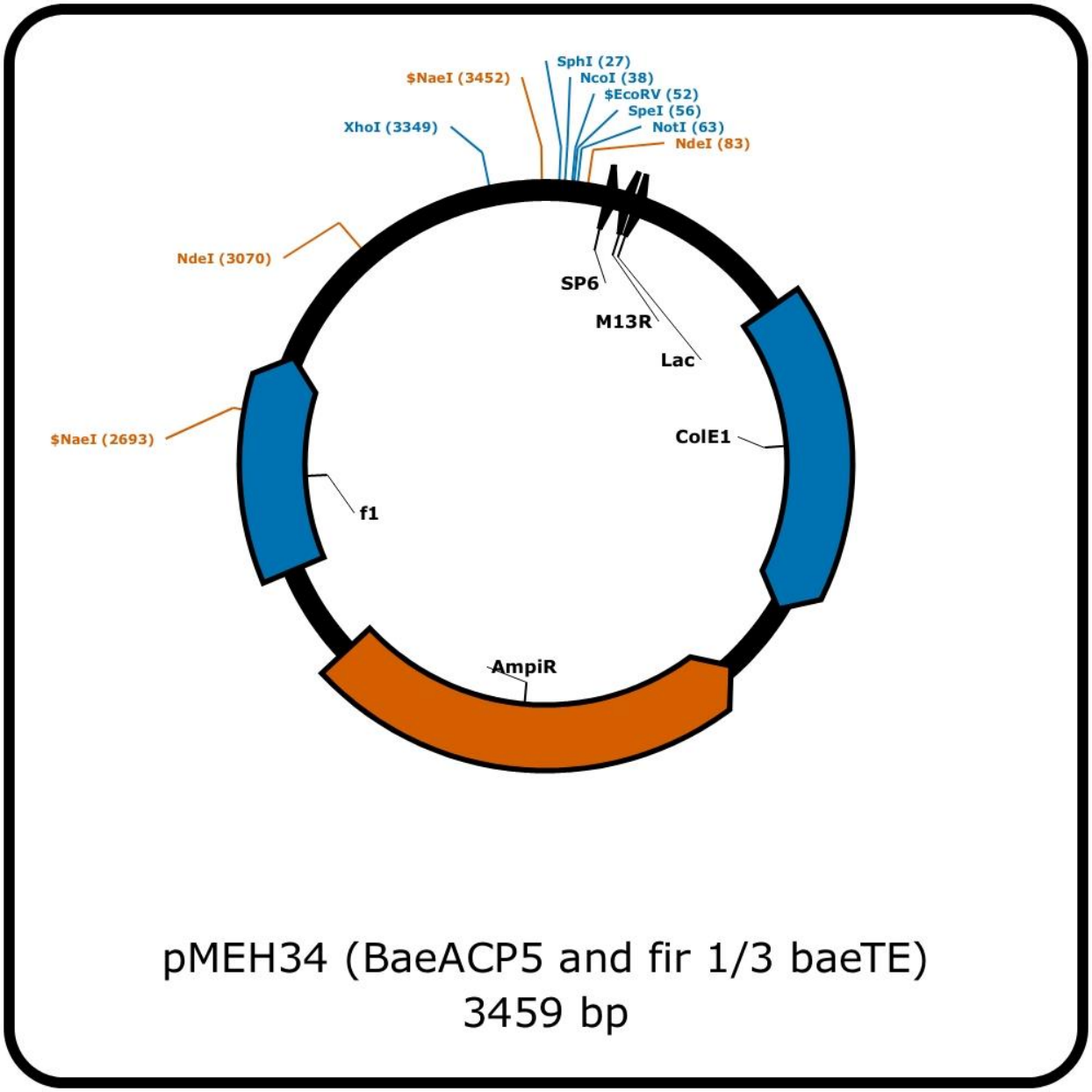


Figure II-34 pMEH34 BaeACP5 and TE fraction mid pathway from mutant strain JM33 in pGemT. PCR product from JM33 colony with XhoI and *EcoRI* sites for later TE knockouts.

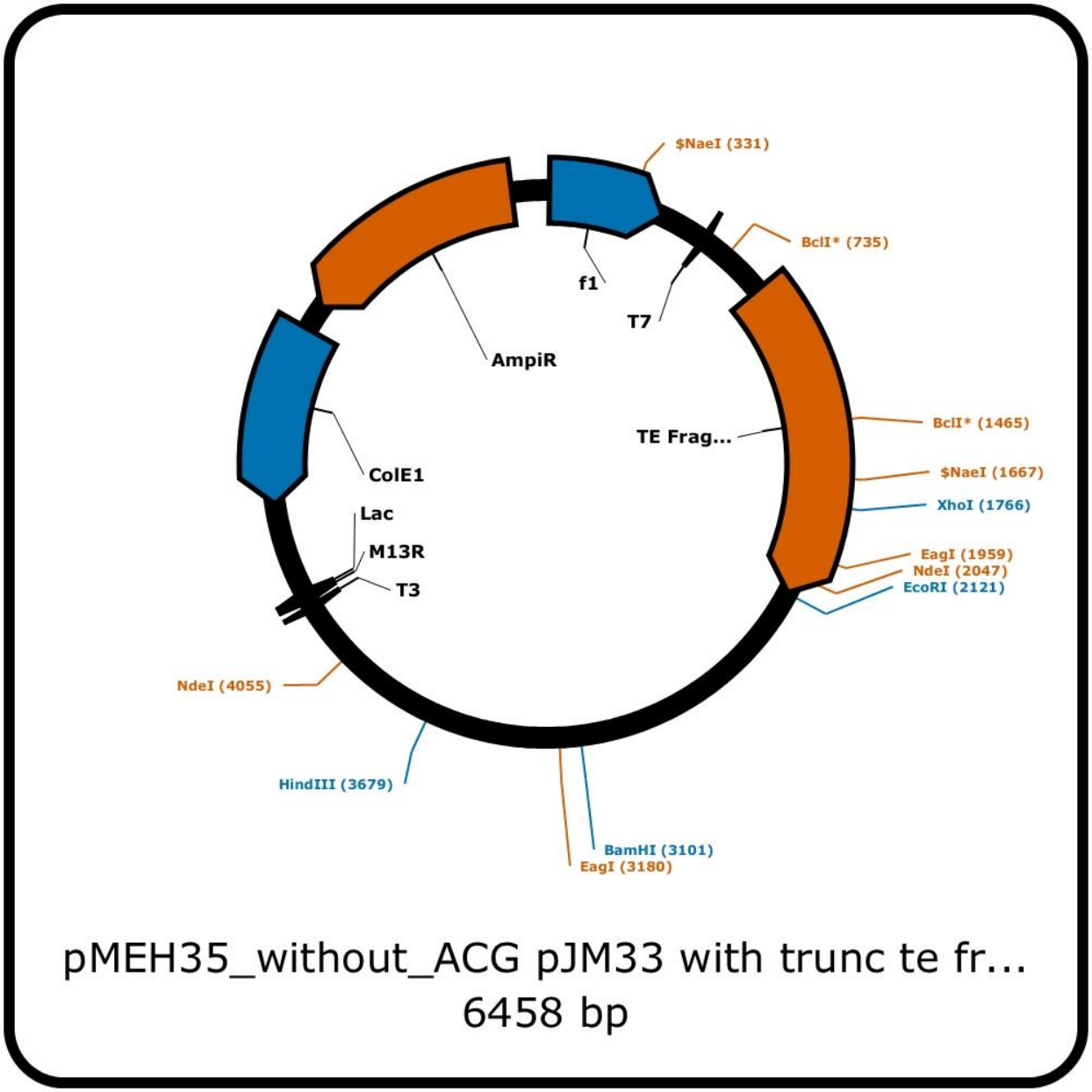


Figure II-35 pMEH35 BaeACP5 and TE fraction in pJM33 from pMEH34. Ligation of *XhoI* and *EcoRI* sites for complementation of genome copy in *Bacillus amyloliquefaciens* FZB42.

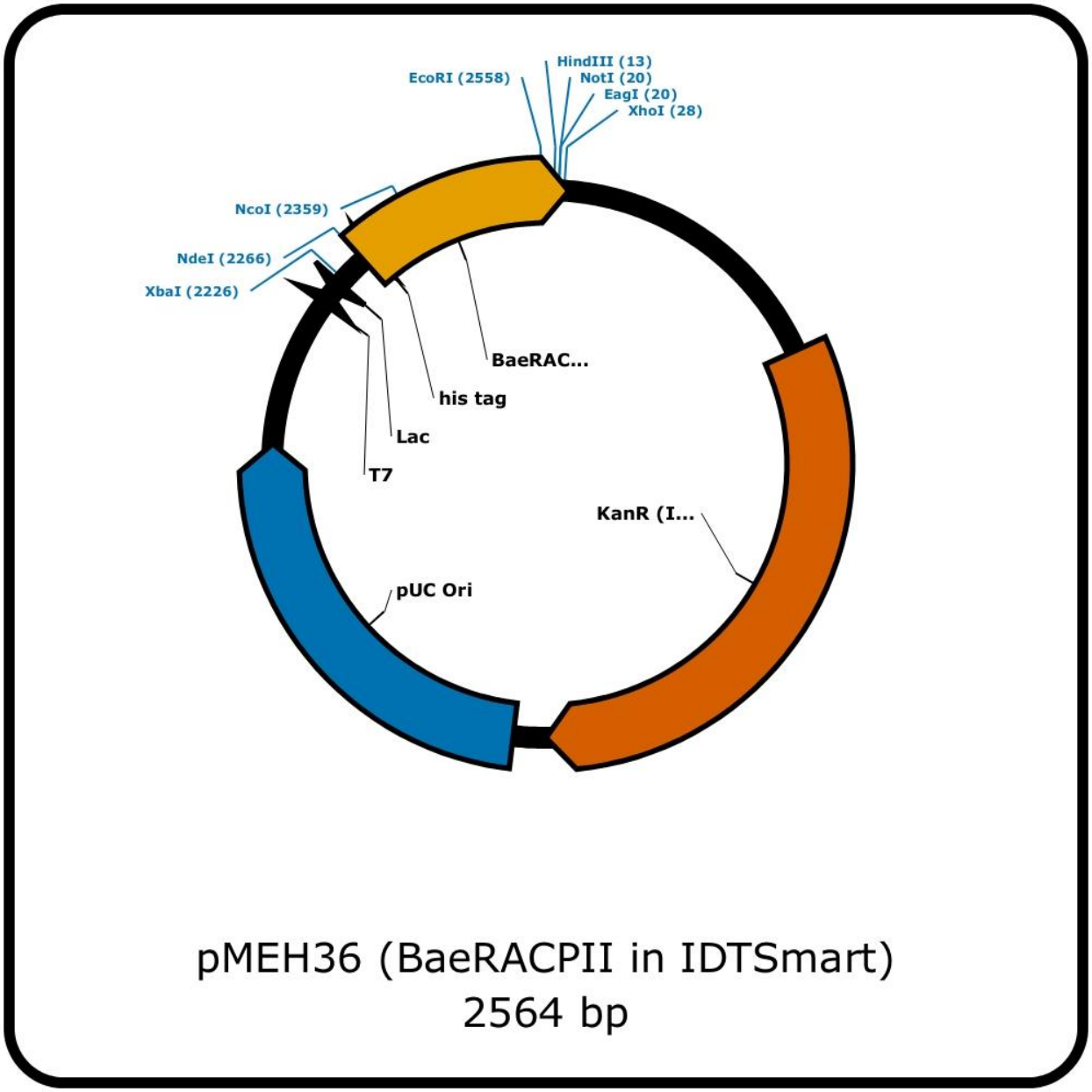


Figure II-36 pMEH36 BaeRACPII repaired from pMEH15 via PCR in IDTSmart.

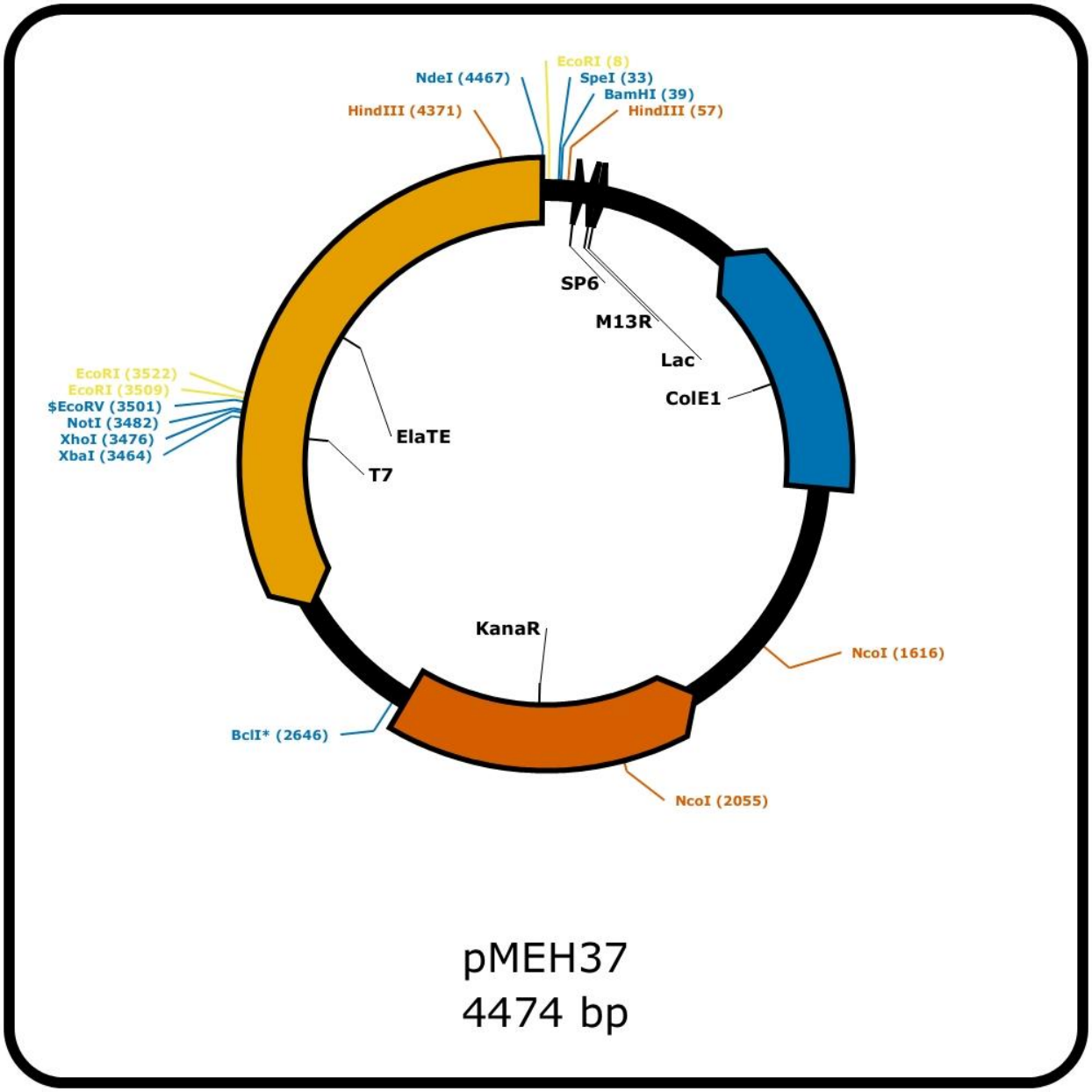


Figure II-37 pMEH37 ChiTE repaired from pMEH26 in pBluntII. Repaired via PCR.

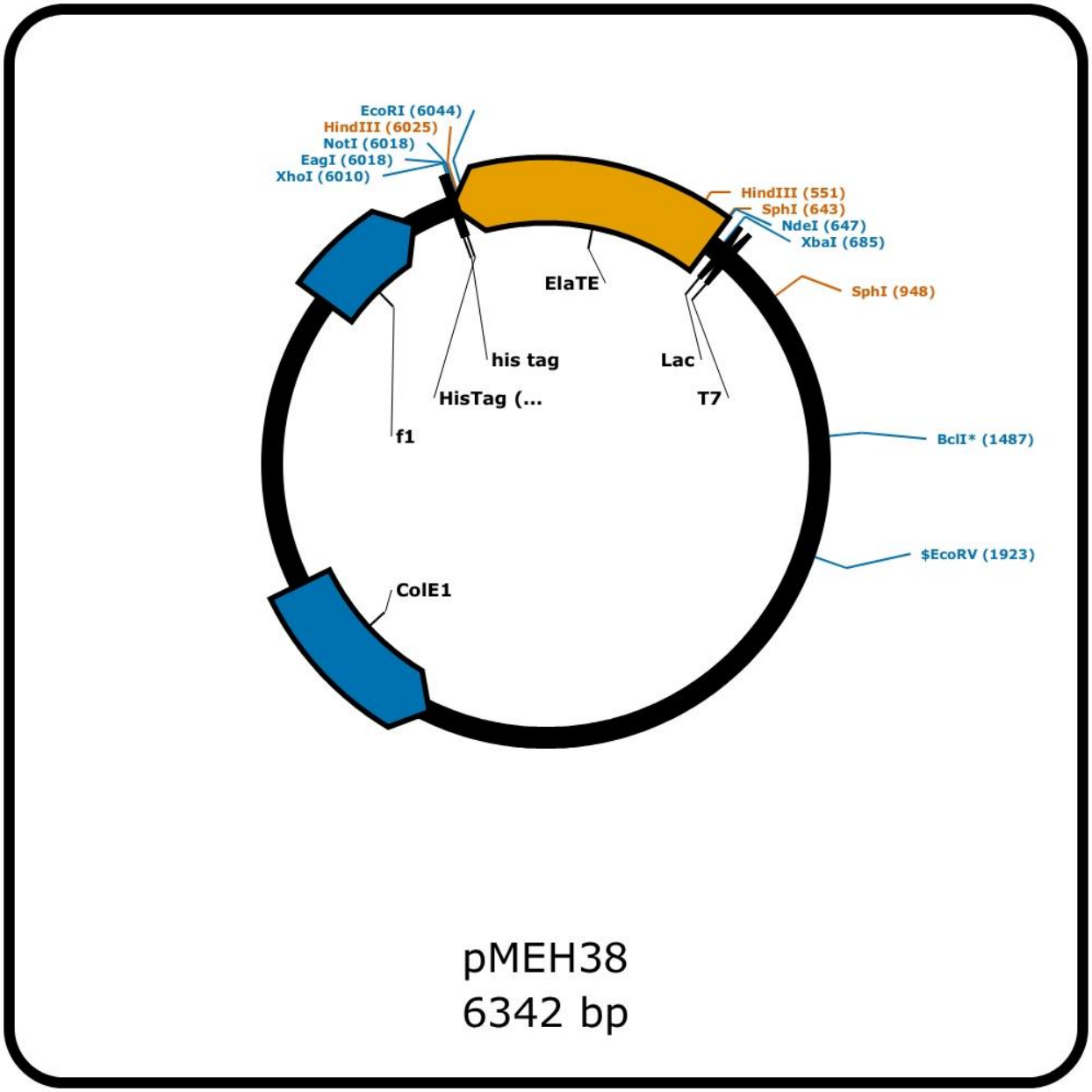


Figure II-38 pMEH38 ChiTE from pMEH37 in pET21. Ligation via *NdeI* and *EcoRI*.

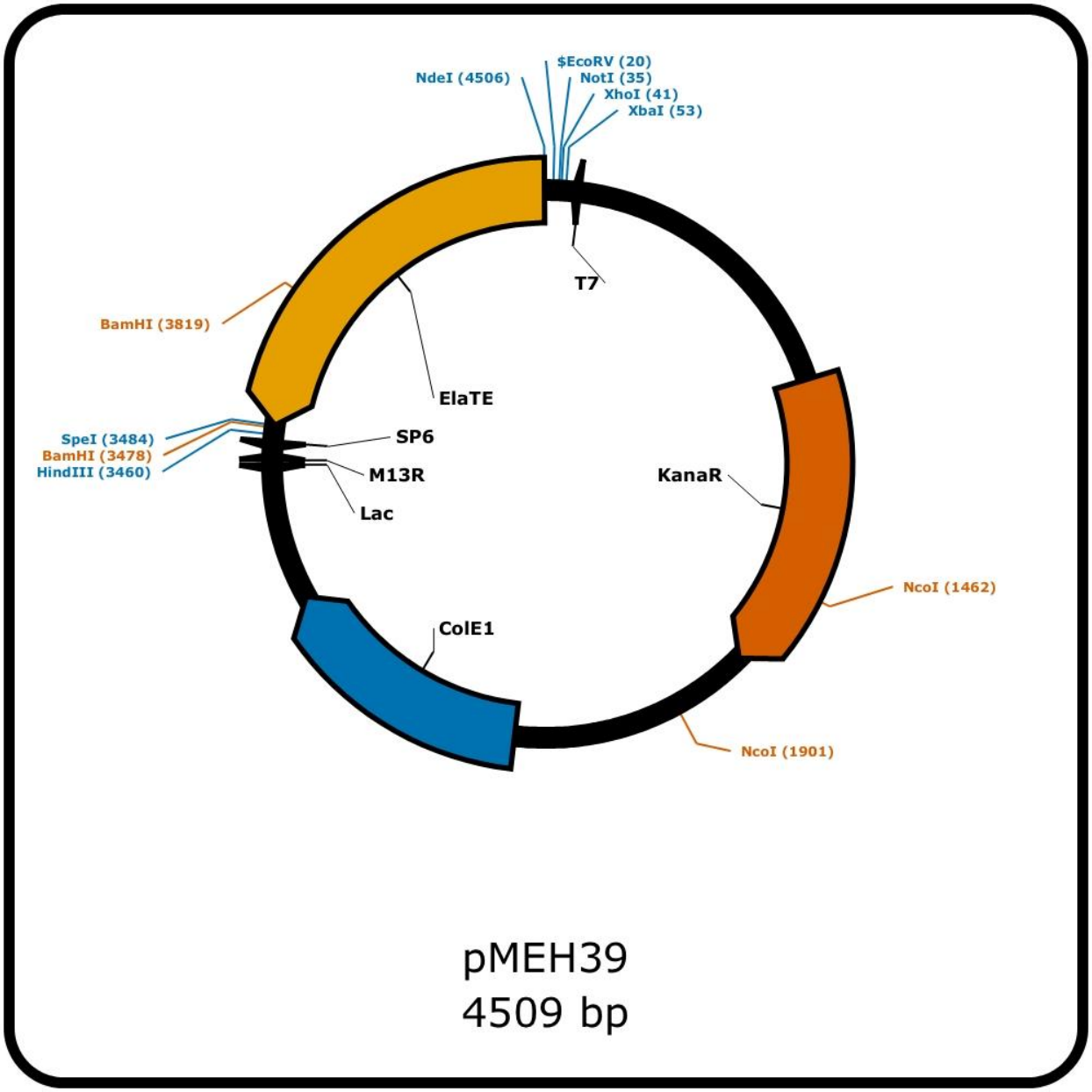


Figure II-39 pMEH39 elansolid TE synthetic gene in pBluntII. Cloning sites *NdeI* and *EcoRI*.

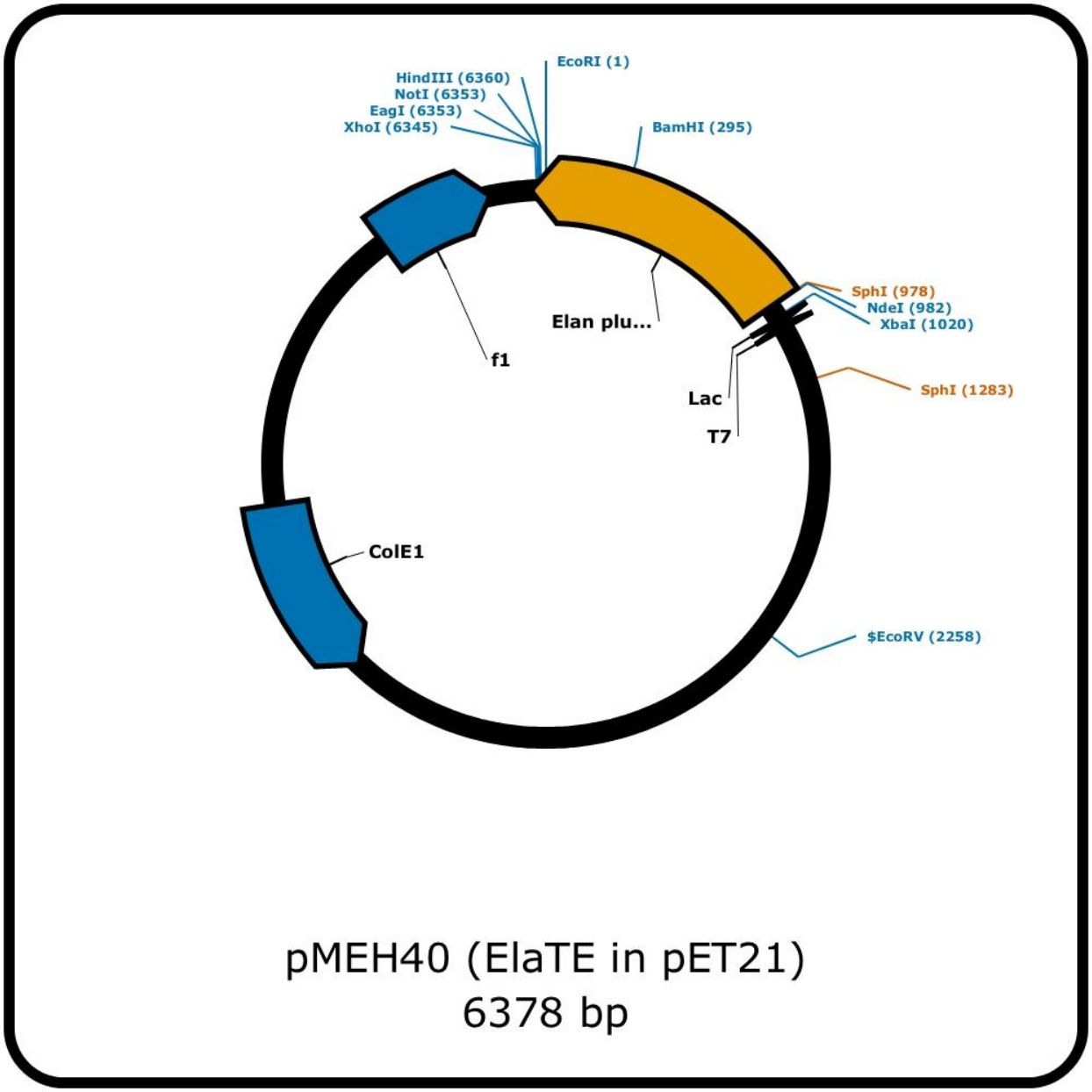


Figure II-40 pMEH40 ElaTE from pMEH39 in pET21. Ligation via *NdeI* and *EcoRI*.

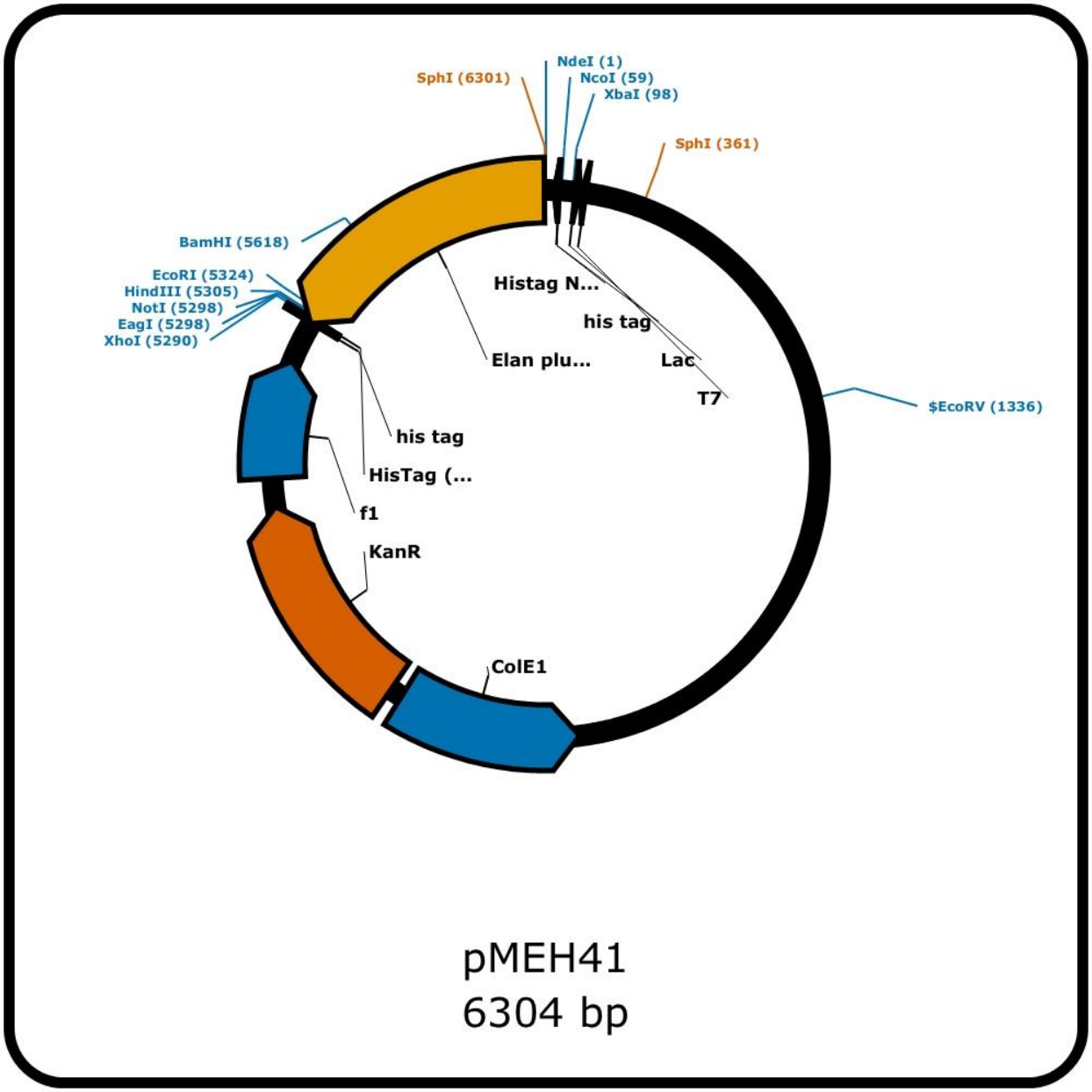


Figure II-41 pMEH41 ElaTE from pMEH39 in pET28. Ligation via *NdeI* and *EcoRI*.

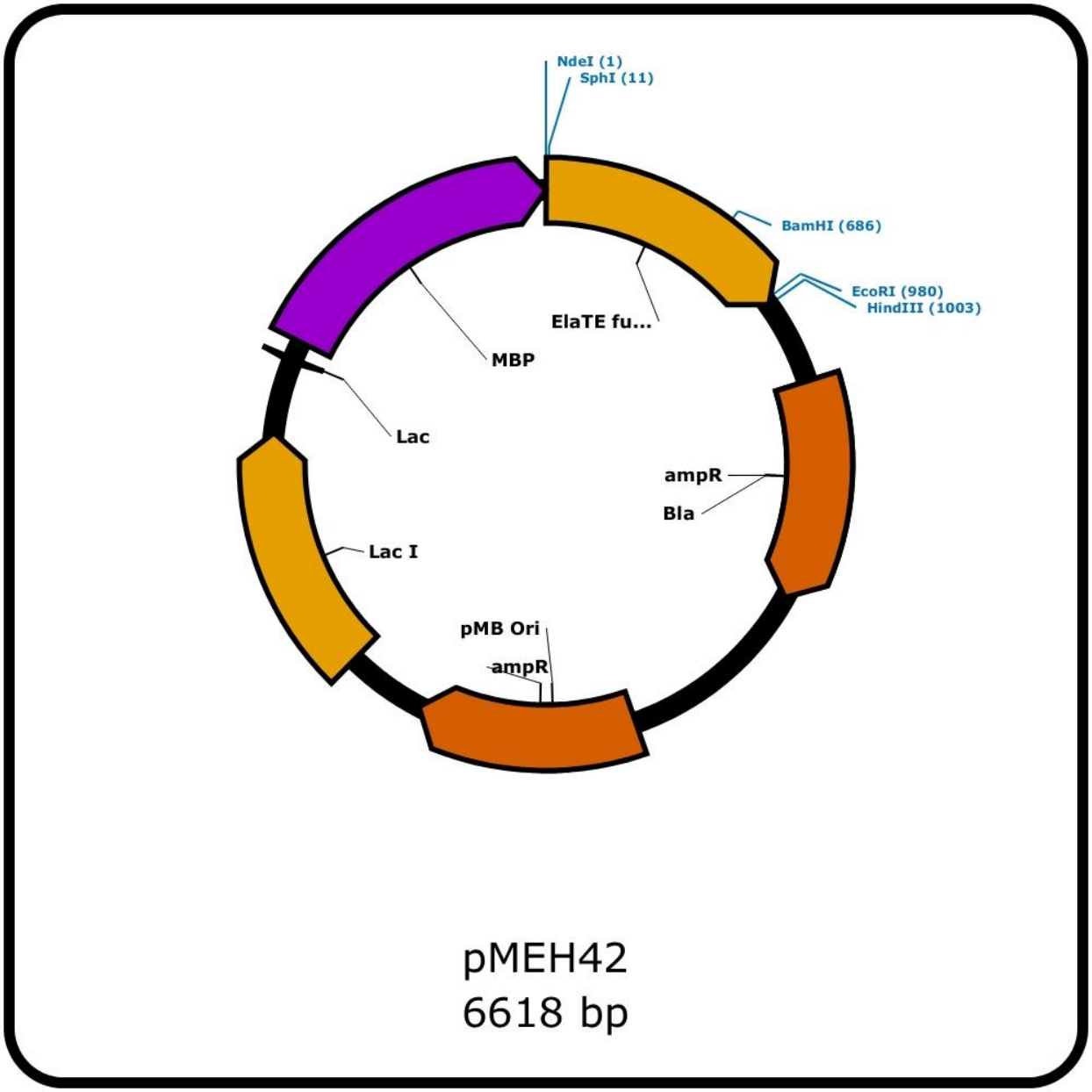


Figure II-42 pMEH42 ElaTE from pMEH39 in pMalC5x. Ligation via *NdeI* and *EcoRI*. Expression of MBP gene fusion does not require T7 polymerase.

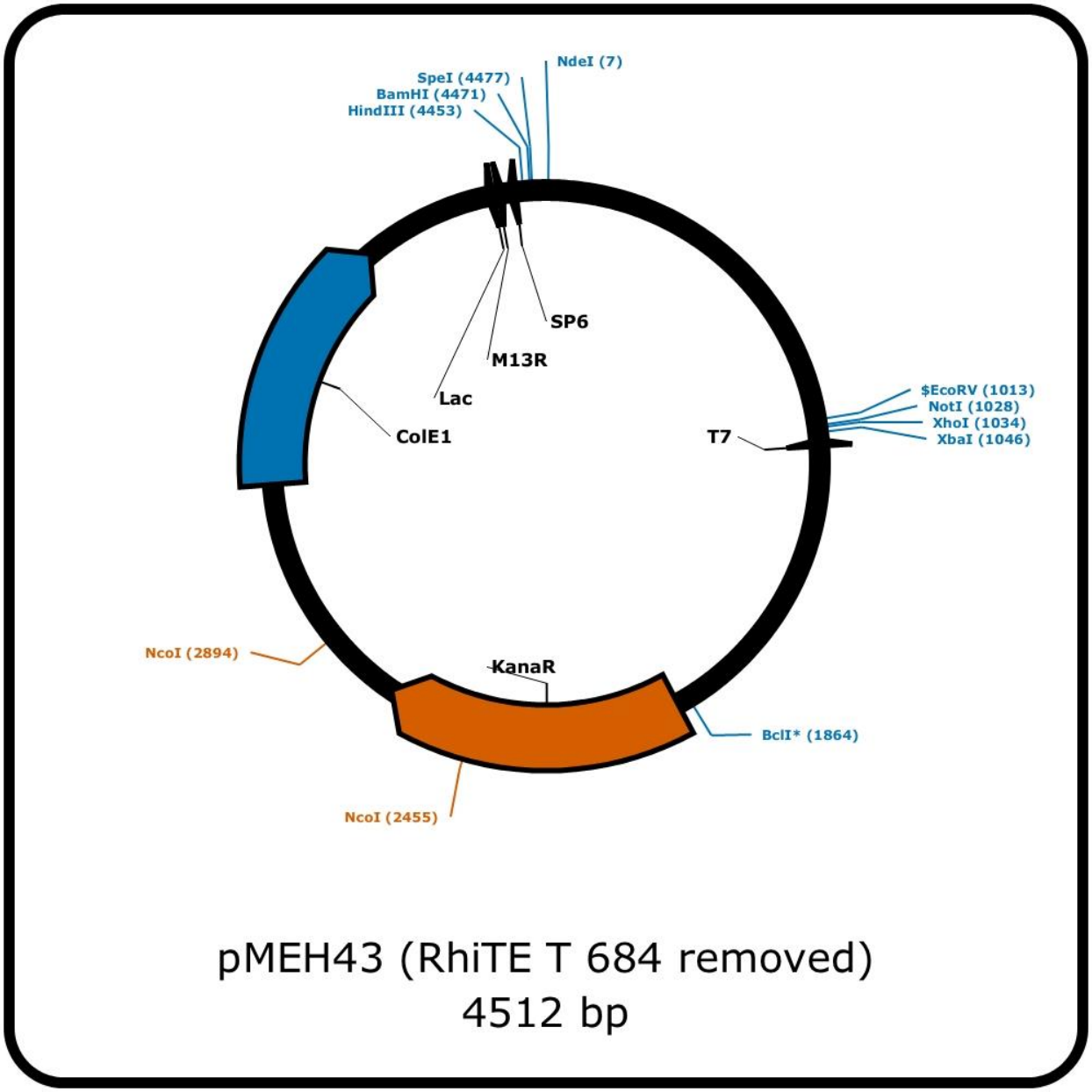


Figure II-43 pMEH43 repaired RhiTE from pMEH27 in pTopoBluntII. Repaired added T 684 via splice by overlap extension.

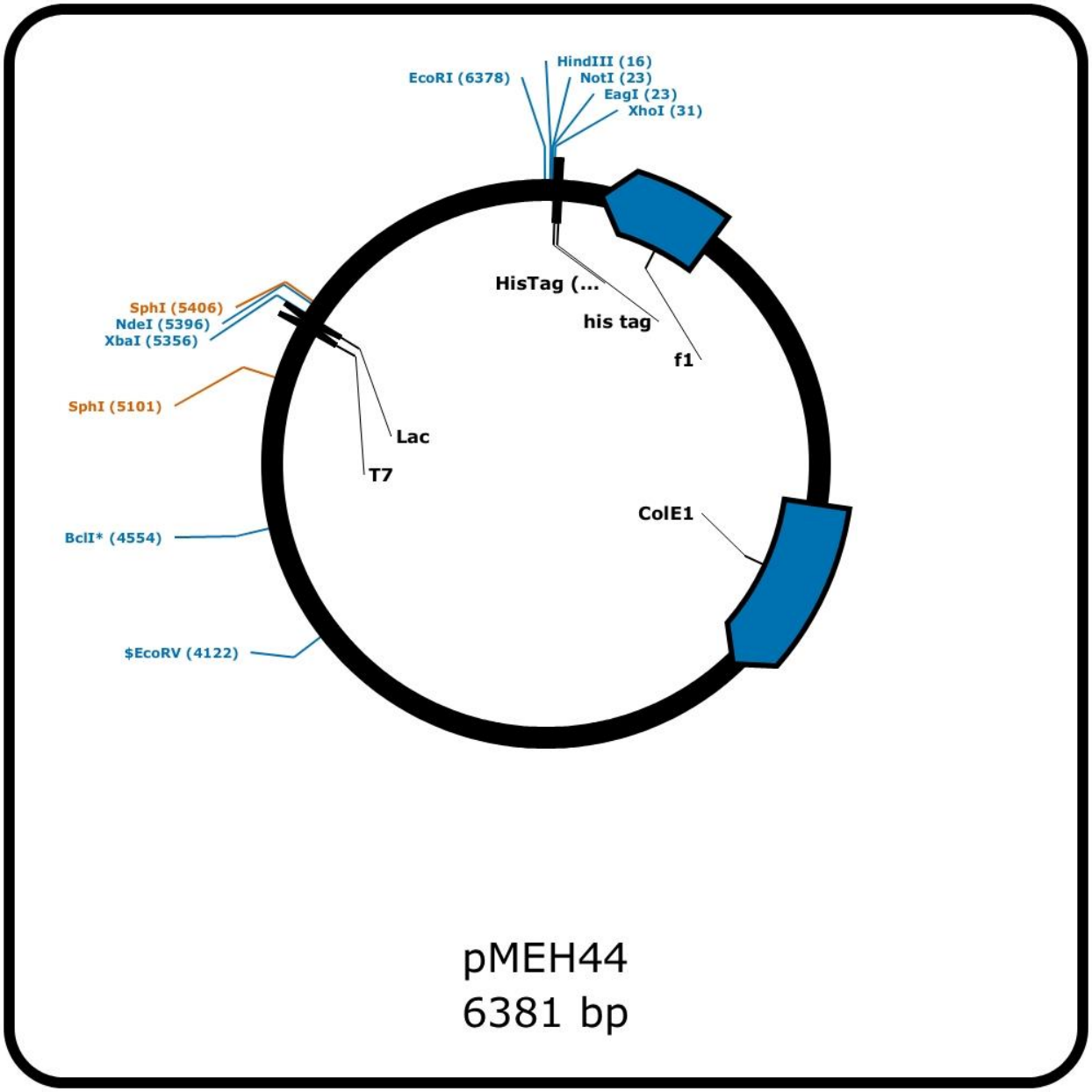


Figure II-44 pMEH44 Repaired RhiTE from pMEH43 in pET21. Ligation via *NdeI EcoRI*.

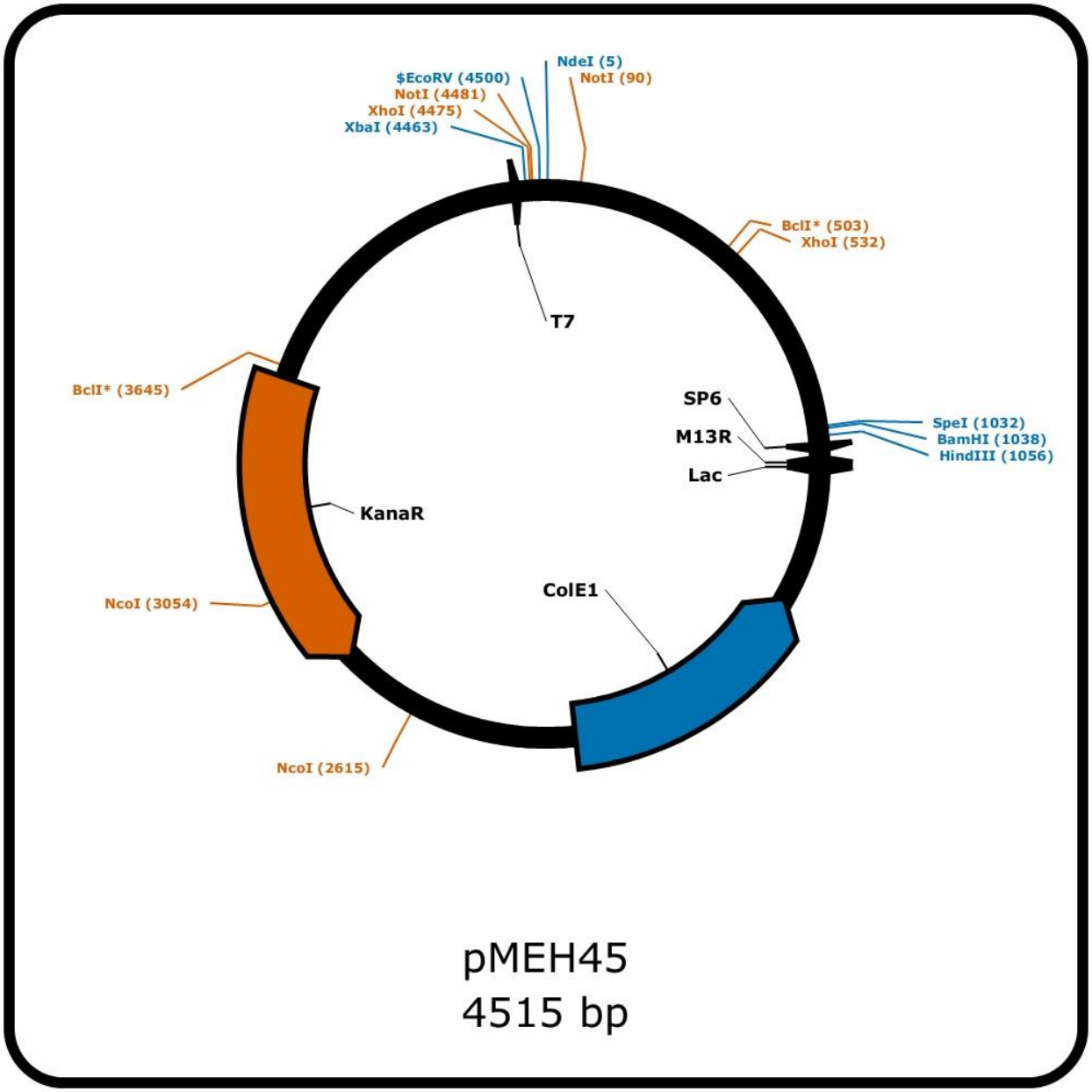


Figure II-45 pMEH45 Codon optimized BaeRTE from *B subtilis* 168 in pTopoBluntII.
Gene synthesis by IDT gBlock ligated into pTopoBluntII.

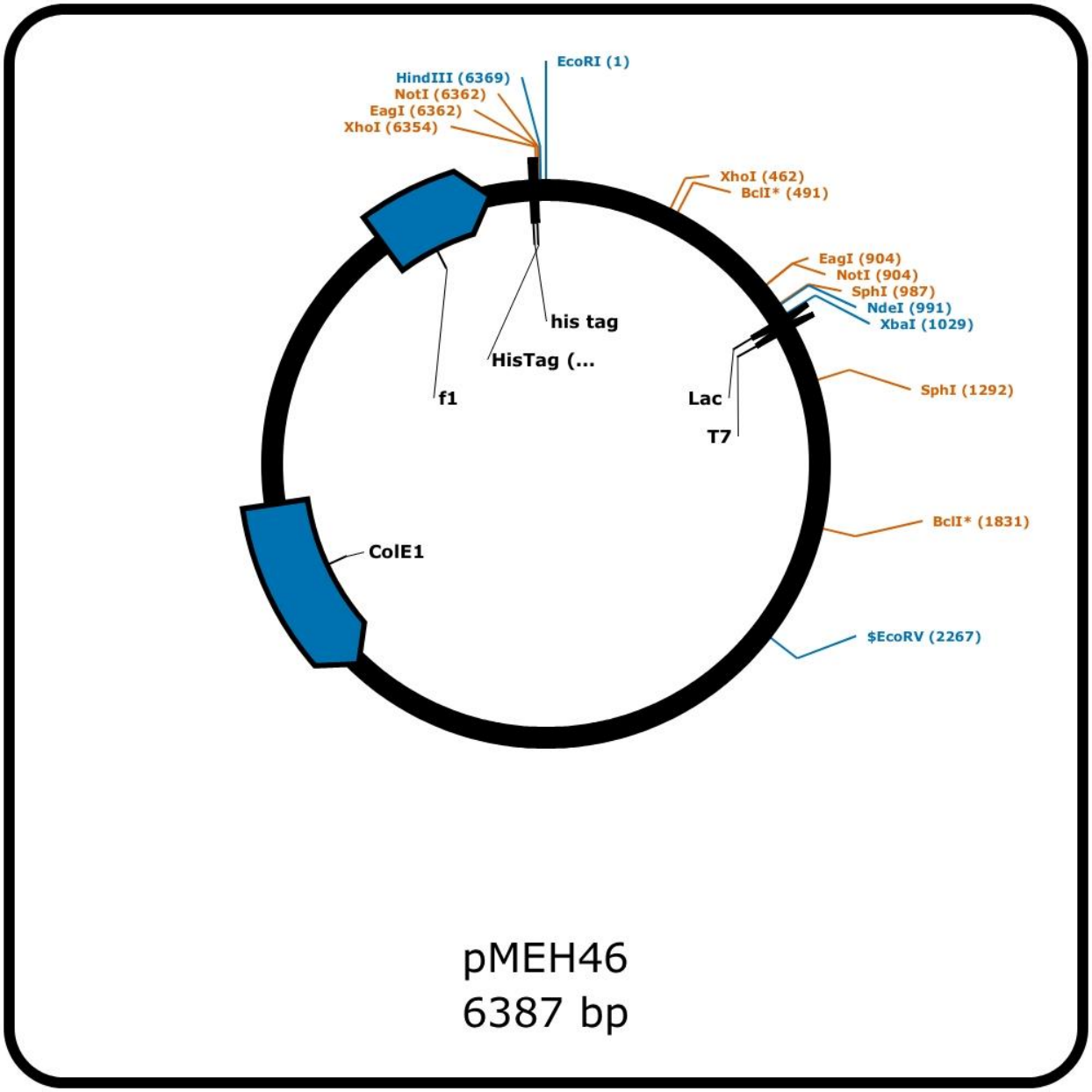


Figure II-46 pMEH46 codon optimized BaeRTE from pMEH46 in pET21. Ligation via *NdeI* and *EcoRI*.

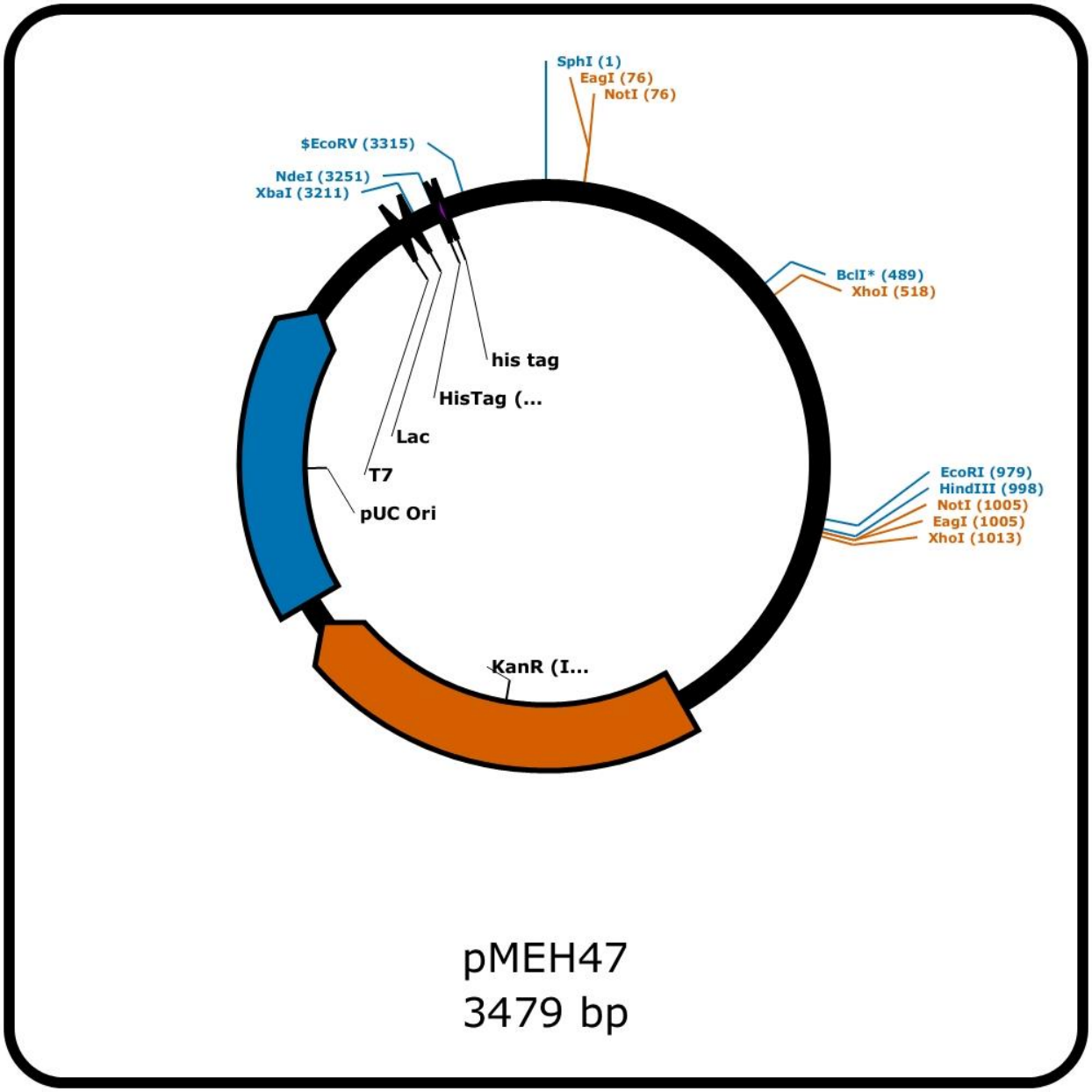


Figure II-47 pMEH47 Codon optimized BaeRTE from pMEH46 in pMEH16. Ligation via *SphI* and *EcoRI*. Forms an ACP-TE fusion between BaeRACPIII from FZB42 and the BaeRTE from *B. subtilis* 168.

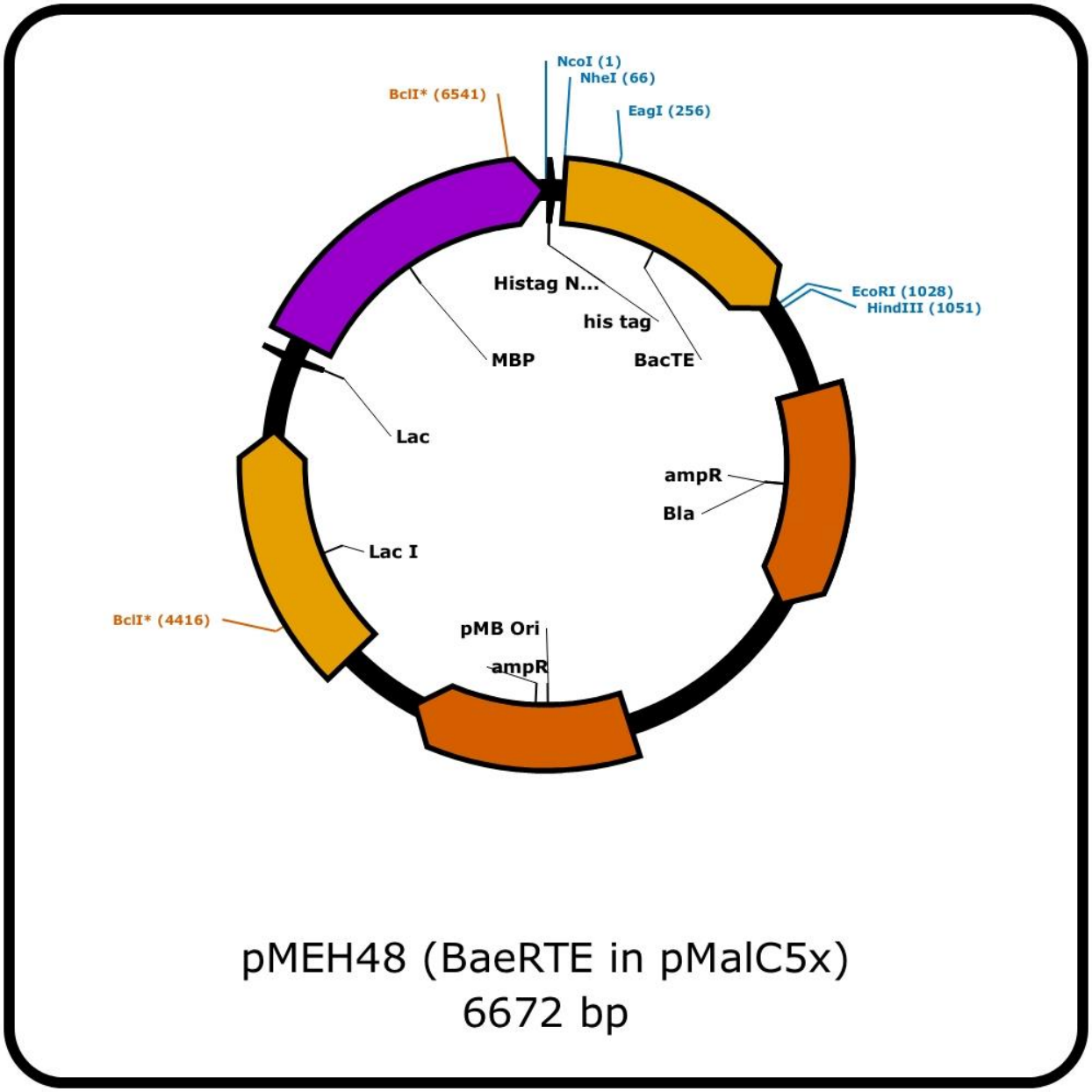


Figure II-48 BaeRTE FZB42 in PMalC5x from pMEH46. Ligation via *NdeI* and *EcoRI*. BaeRTE FZB42 in PMalC5x

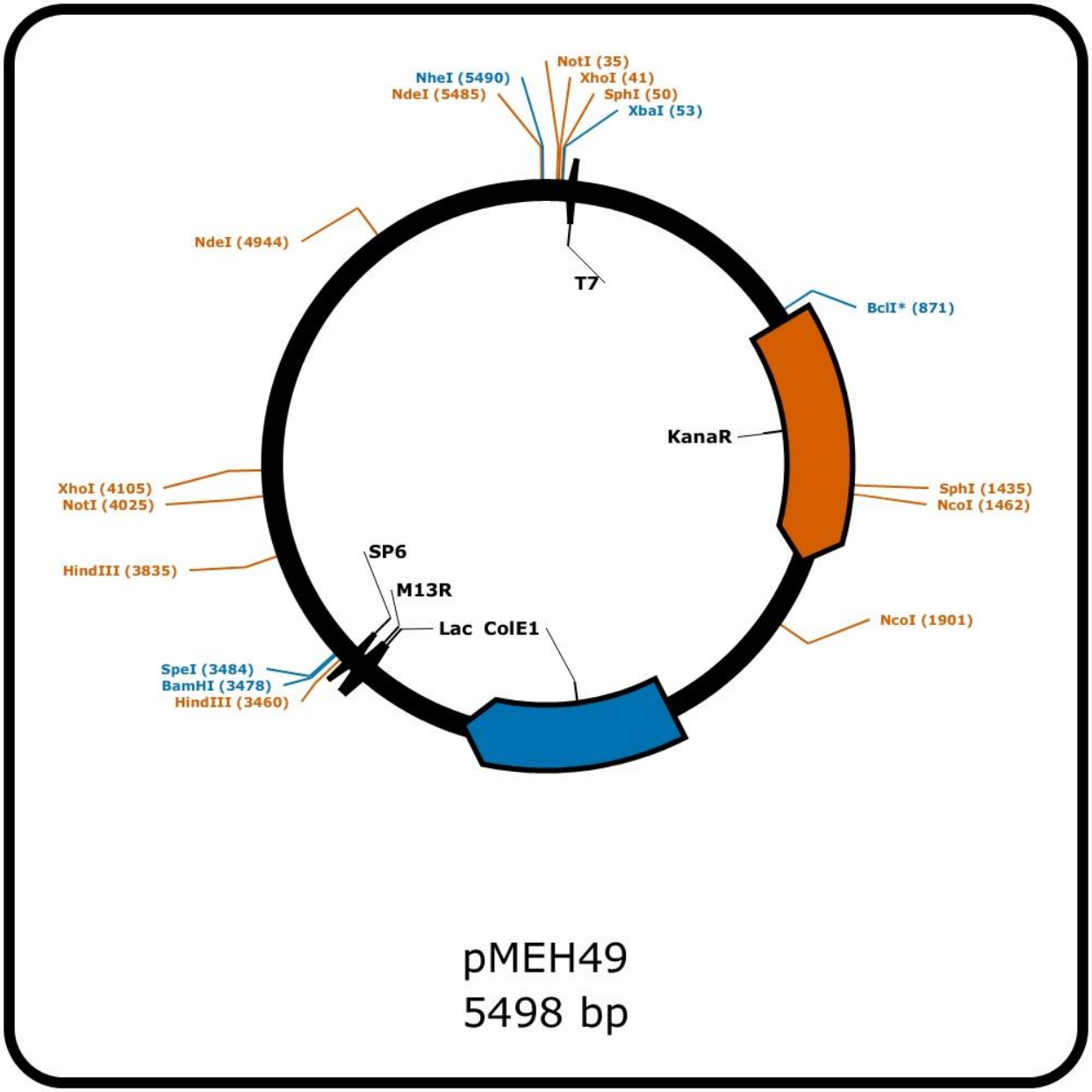


Figure II-49 pMEH49 codon optimized BaeR KS from pTopoBluntII.

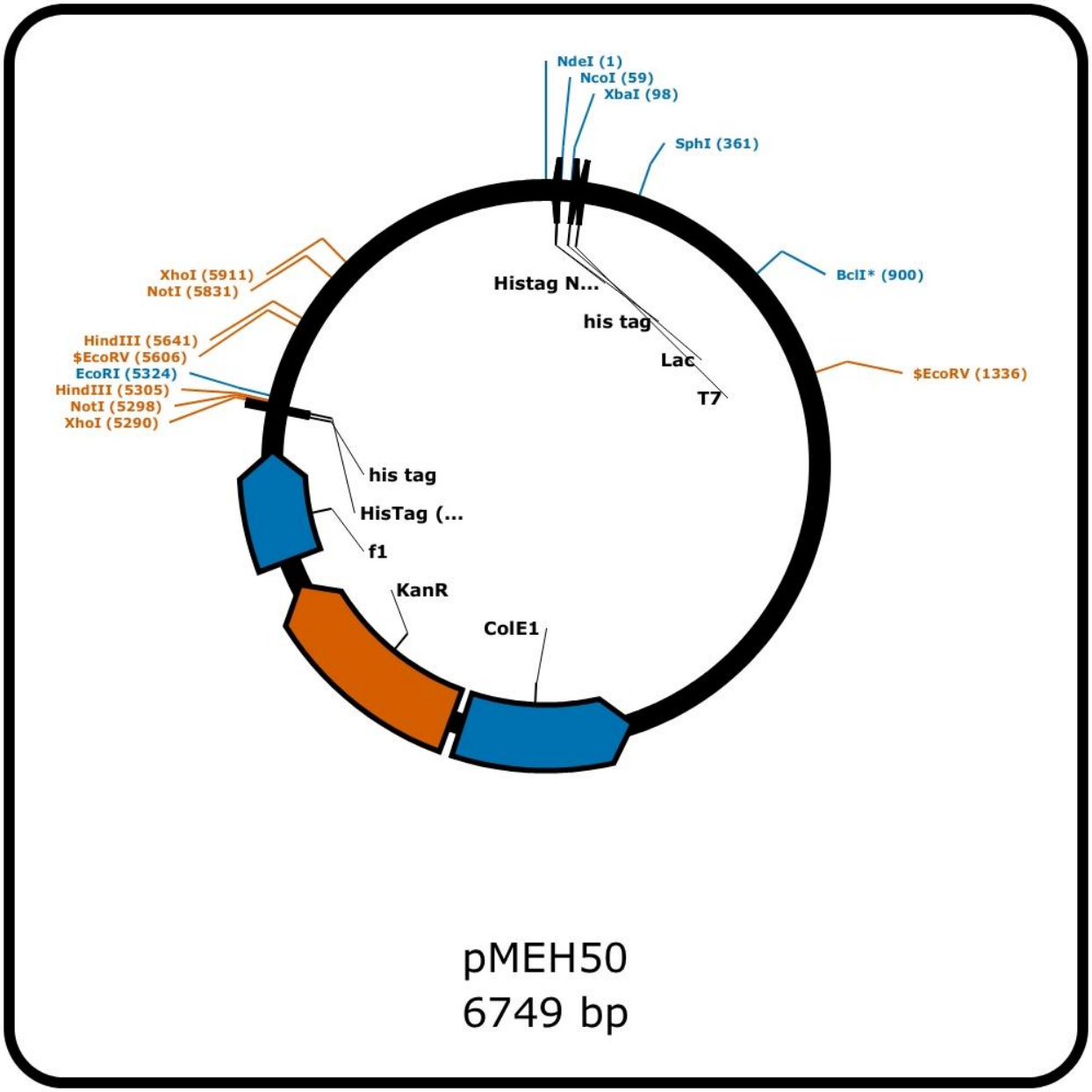


Figure II-50 pMEH50 BaeR KS from pMEH49 in pET28. Ligation via *NheI* and *EcoRI*.

References

1. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science (New York, N.Y.)*, 2001, 291, 1790-1792. (pMEH21)

Appendix III

Figure III-1 Alignment of bacillaene TE domain constructs

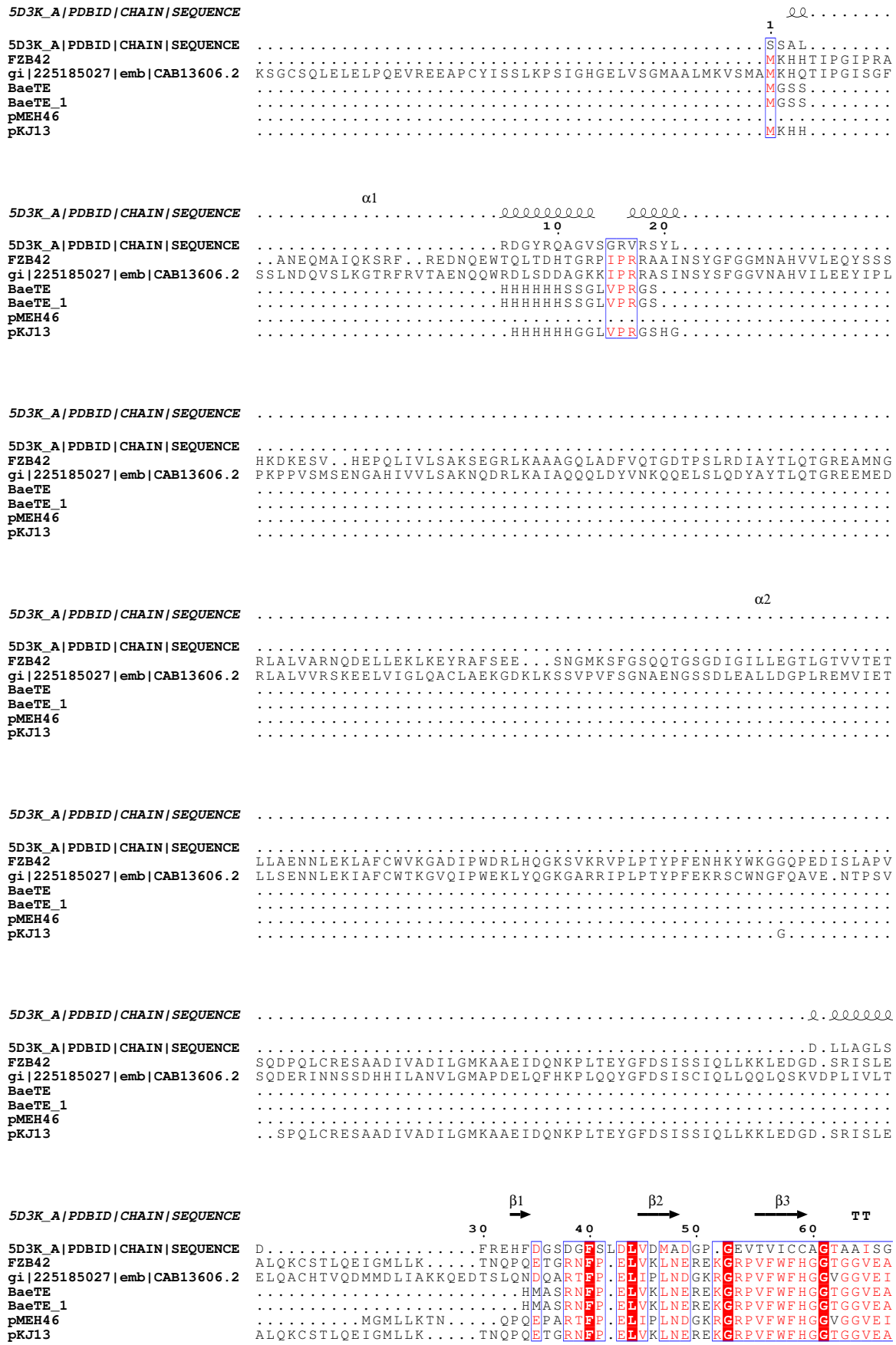
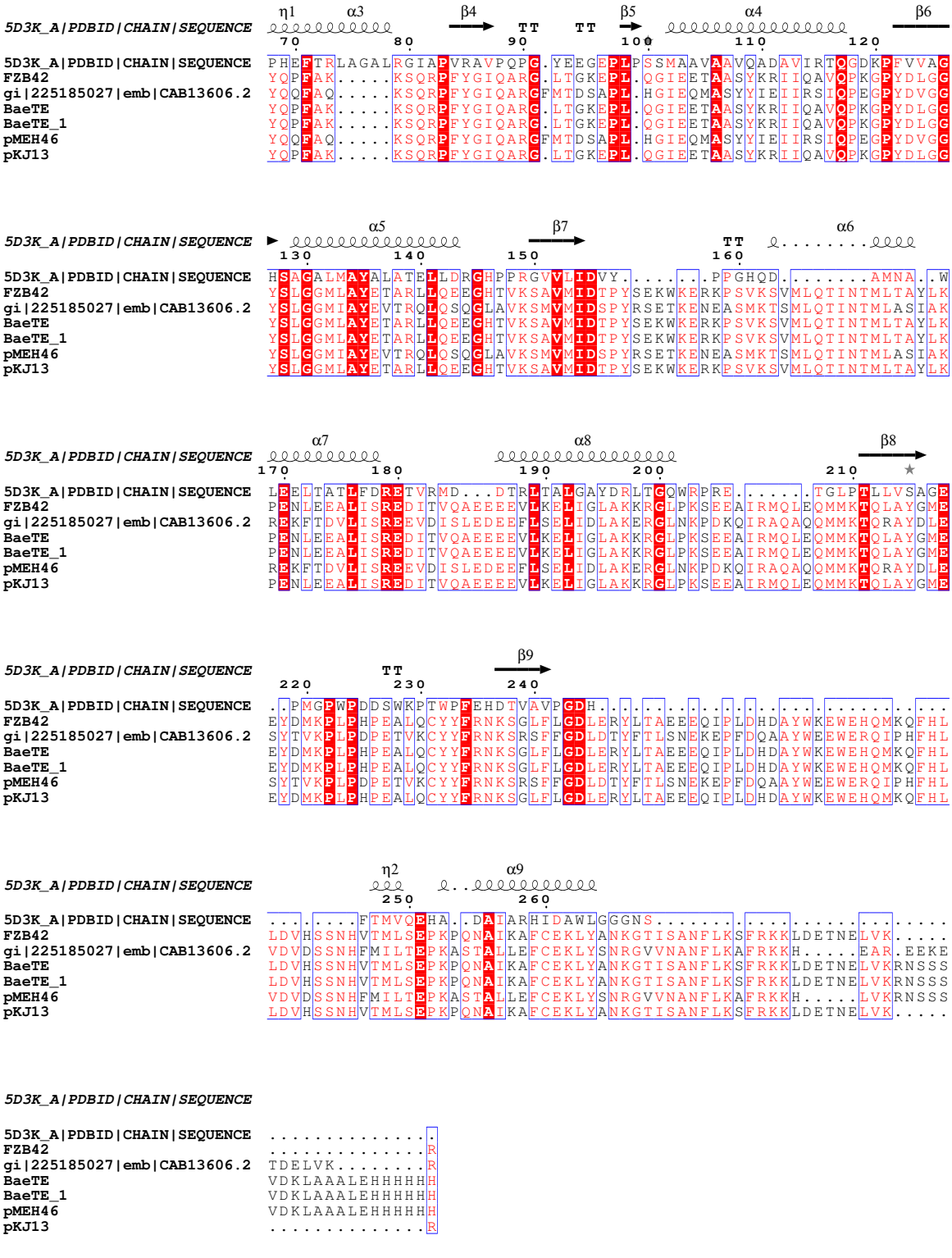


Figure III-1 (continued)



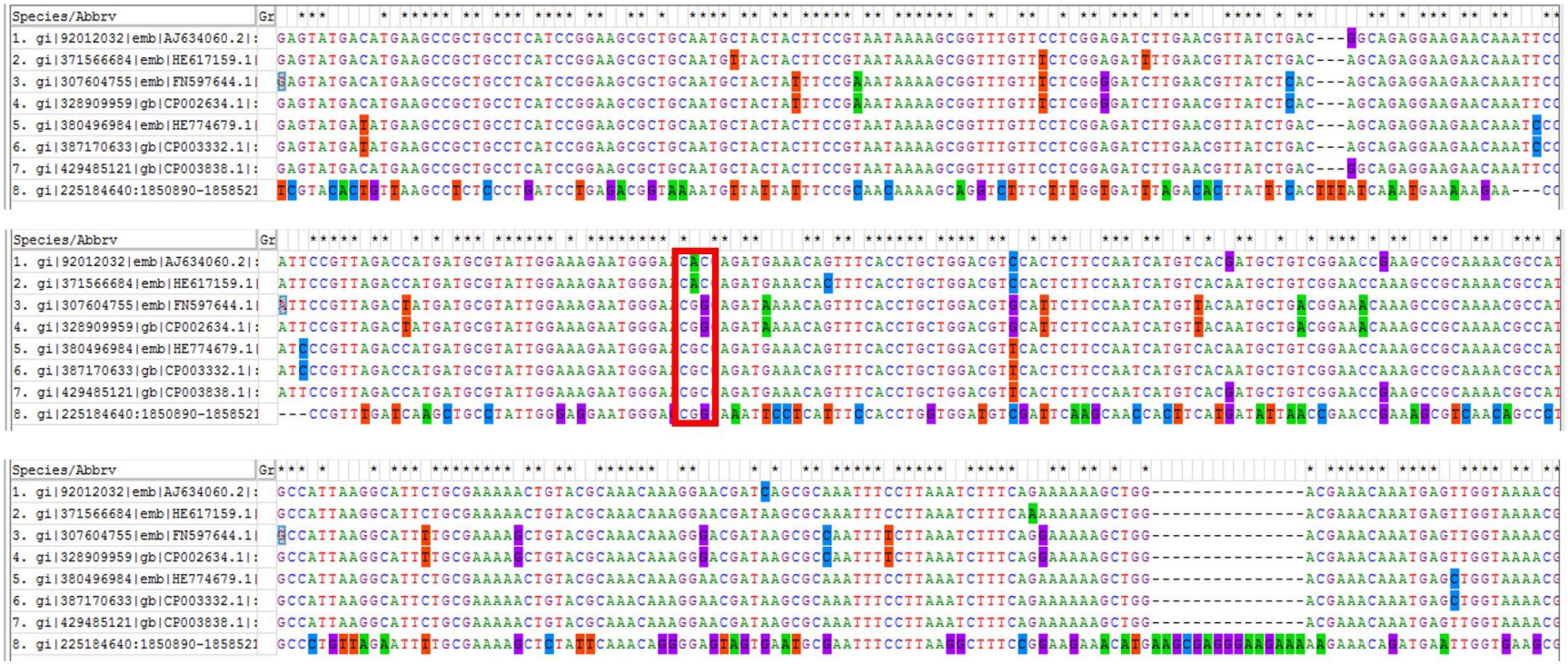


Figure III-2 A 7400-7763 alignment region of BaeR modules from 8 Bacillus strains. The first row contains the FZB42 sequence and any substitution occurring in less than 50% of the sequences is highlighted. Figure generated from MEGA6[1] in 120 base segments. The Arg to His substitution in FZB42 is highlighted in the red box

[1] Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30:2725-2729

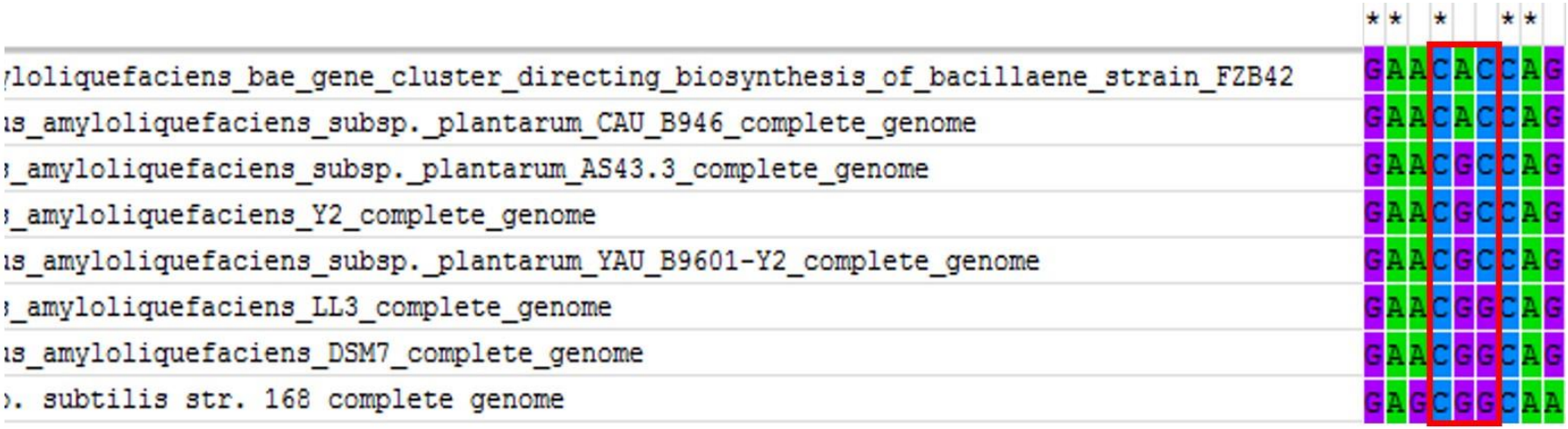


Figure III-2 B Highlighted codon coding for surface Arg (CGC or CGG) or His (CAC).

Appendix IV

Figure IV- 1 Initial SNAC hydrolysis time-courses with other BaeRTE constructs.

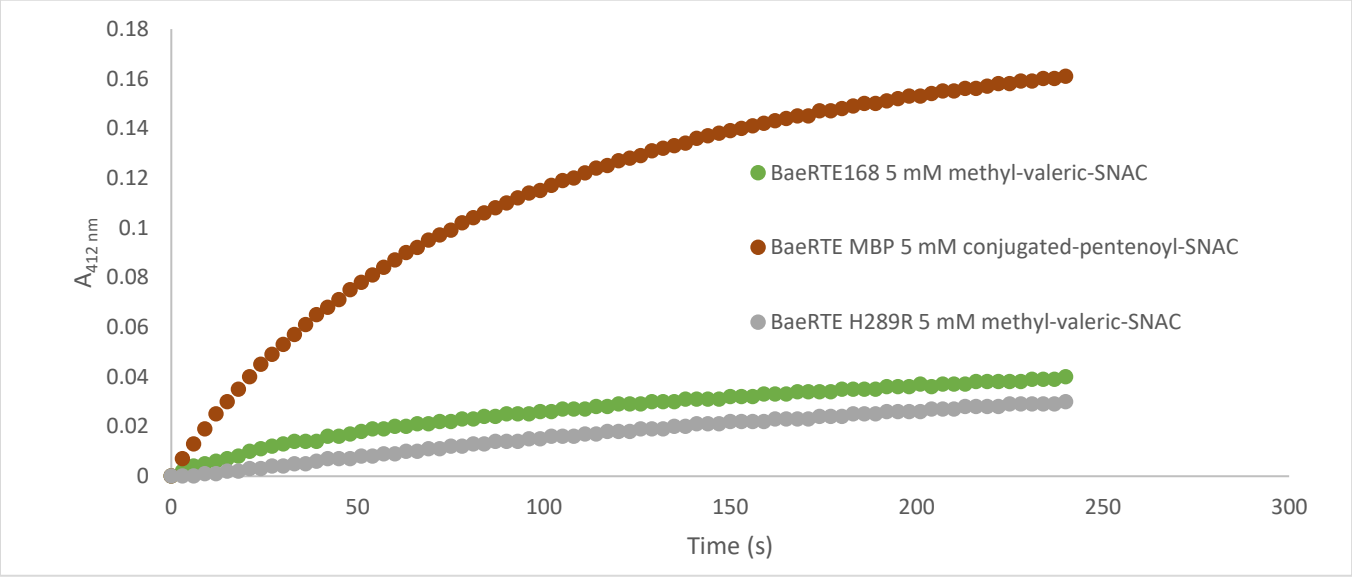


Figure IV- 2 400 MHz $CDCl_3$ ^1H NMR of *N*-Acetylcysteamine

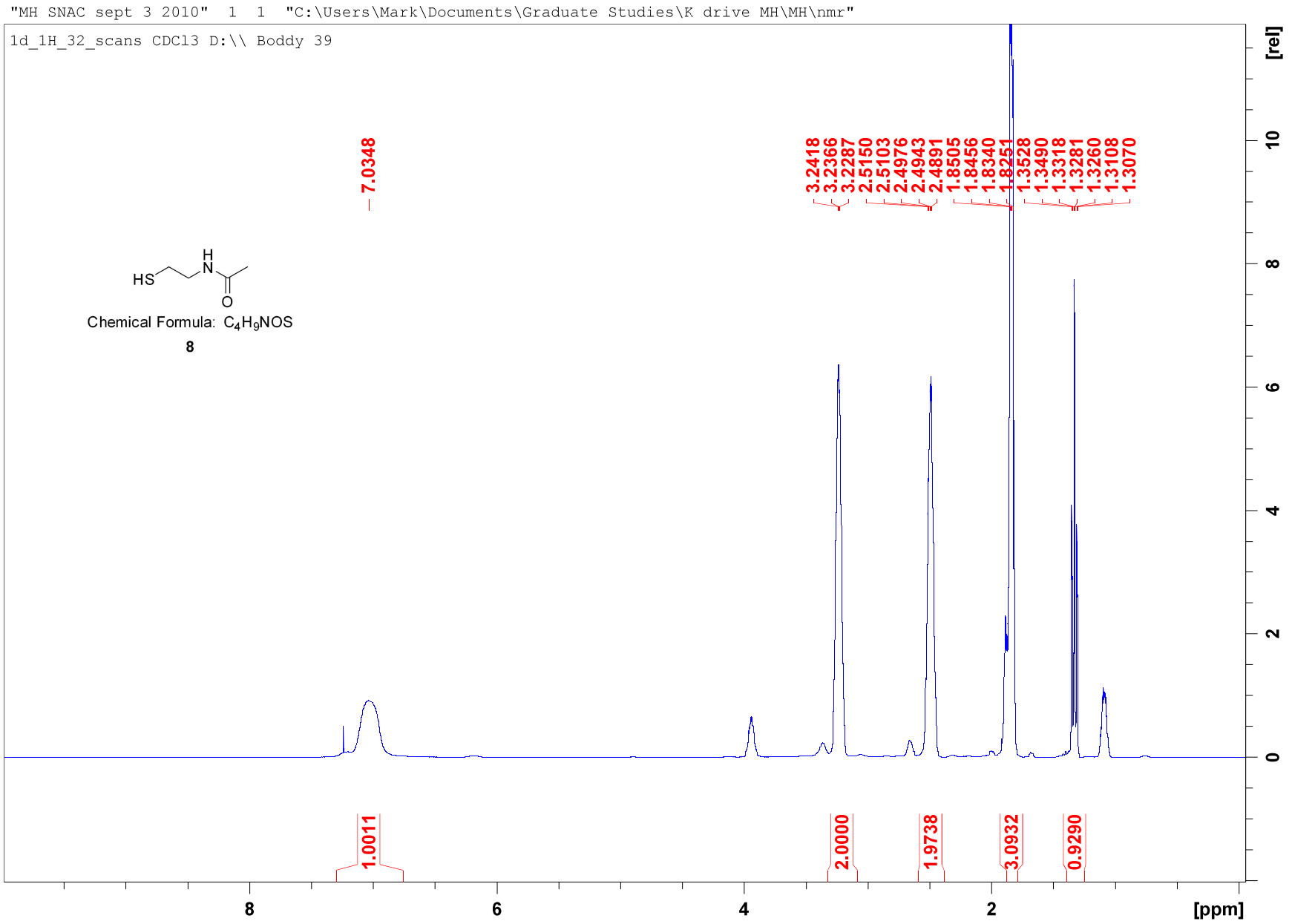


Figure IV- 3 400 MHz $CDCl_3$ ^1H NMR of racemic 2-methyl-valeric-SNAC

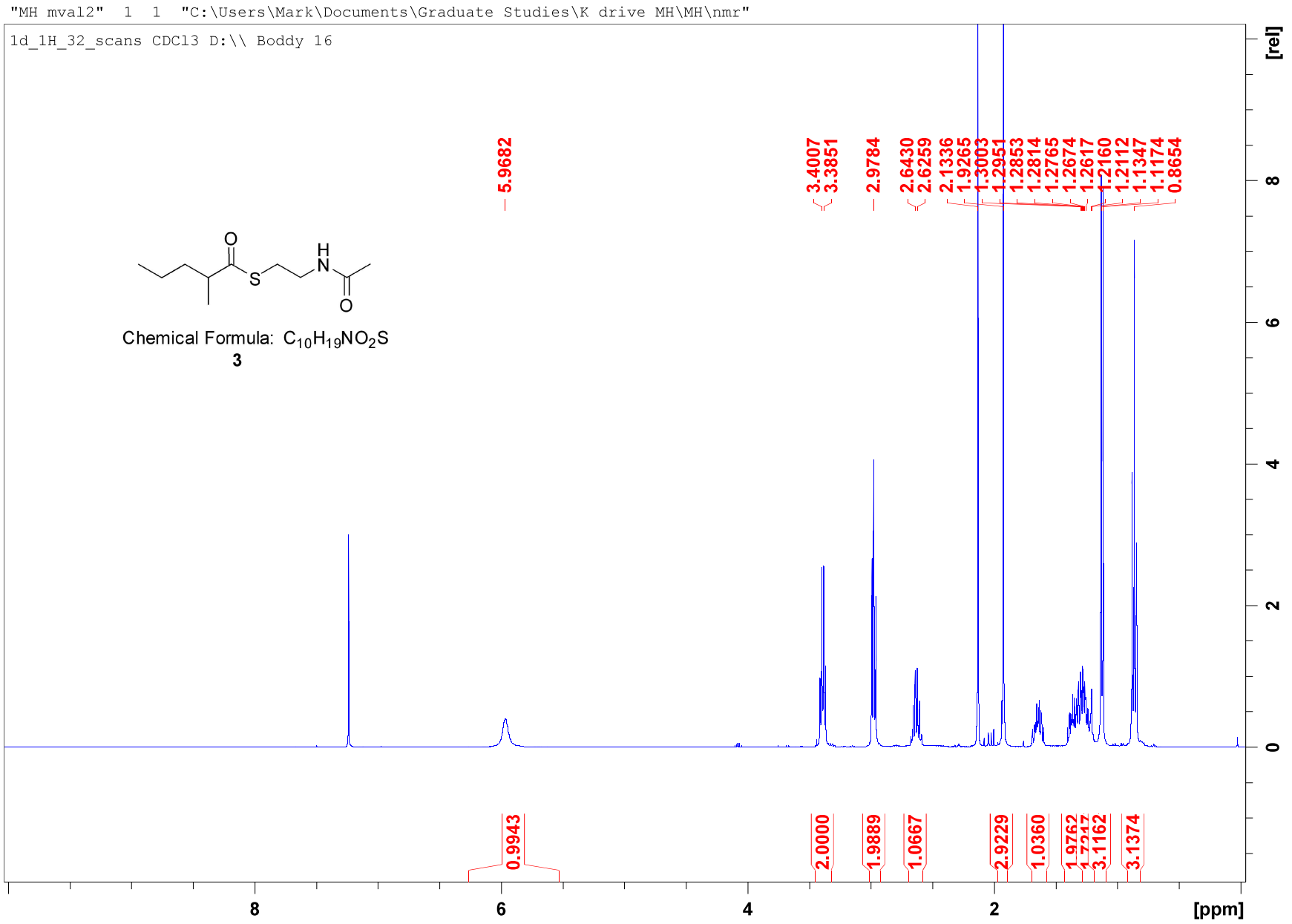


Figure IV- 4 ^{13}C NMR of racemic 2-methyl-valeric-SNAC

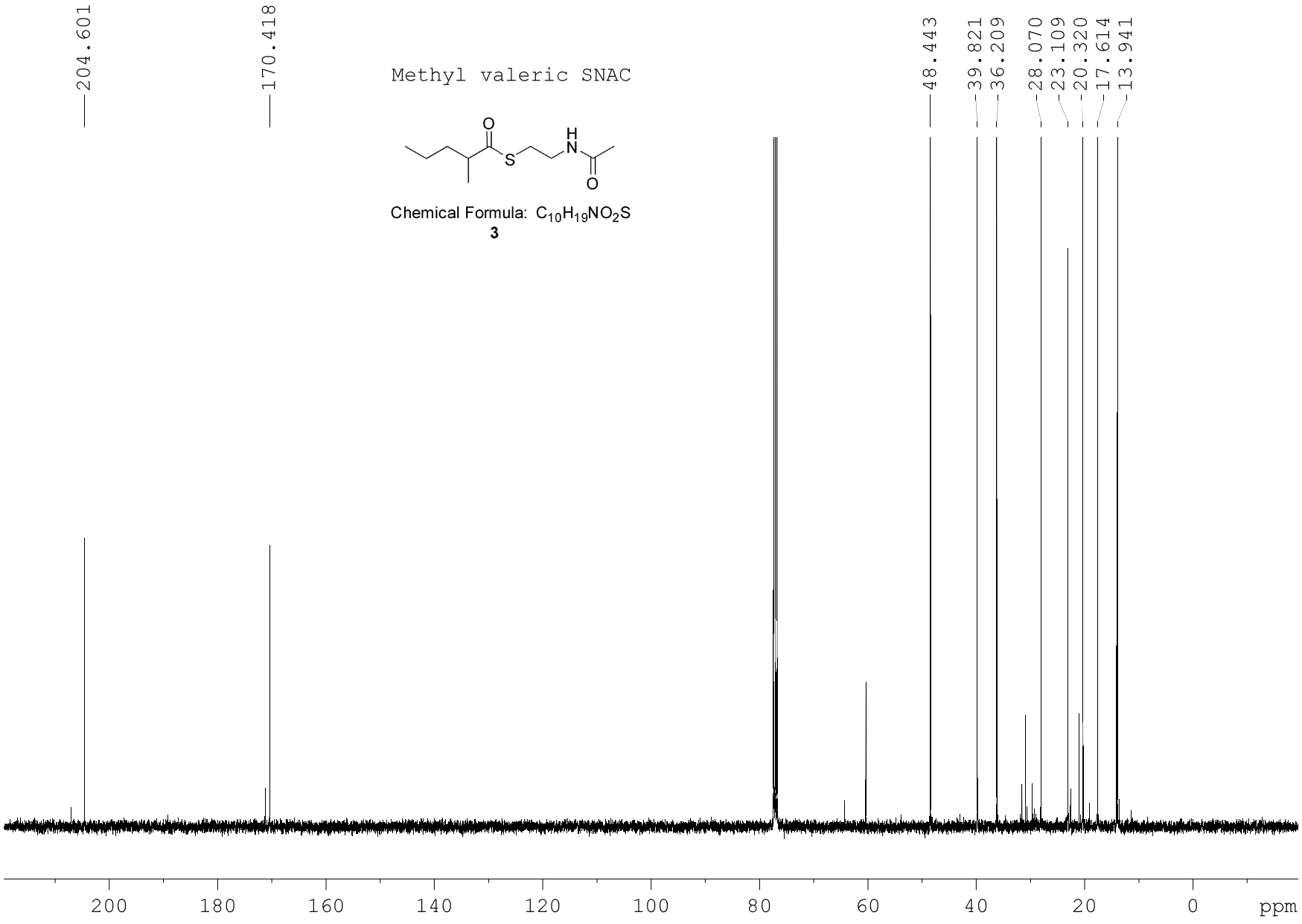


Figure IV- 5 $_{13}$C NMR of *trans*-2-methyl-2-pentenoyl-SNAC

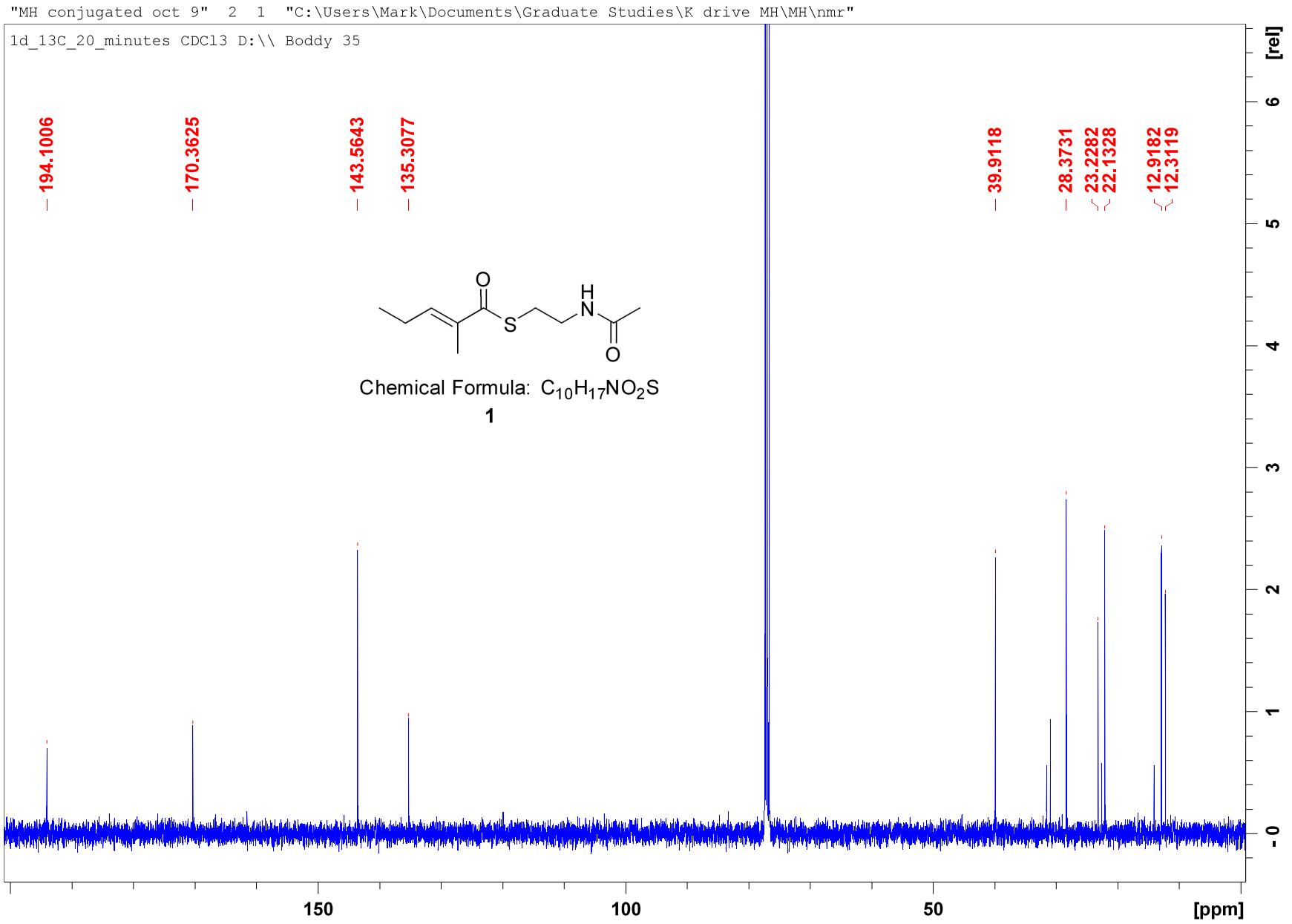


Figure IV- 6 400 MHz $CDCl_3$ ^1H NMR of *cis*-3-methyl-2-pentenoic acid

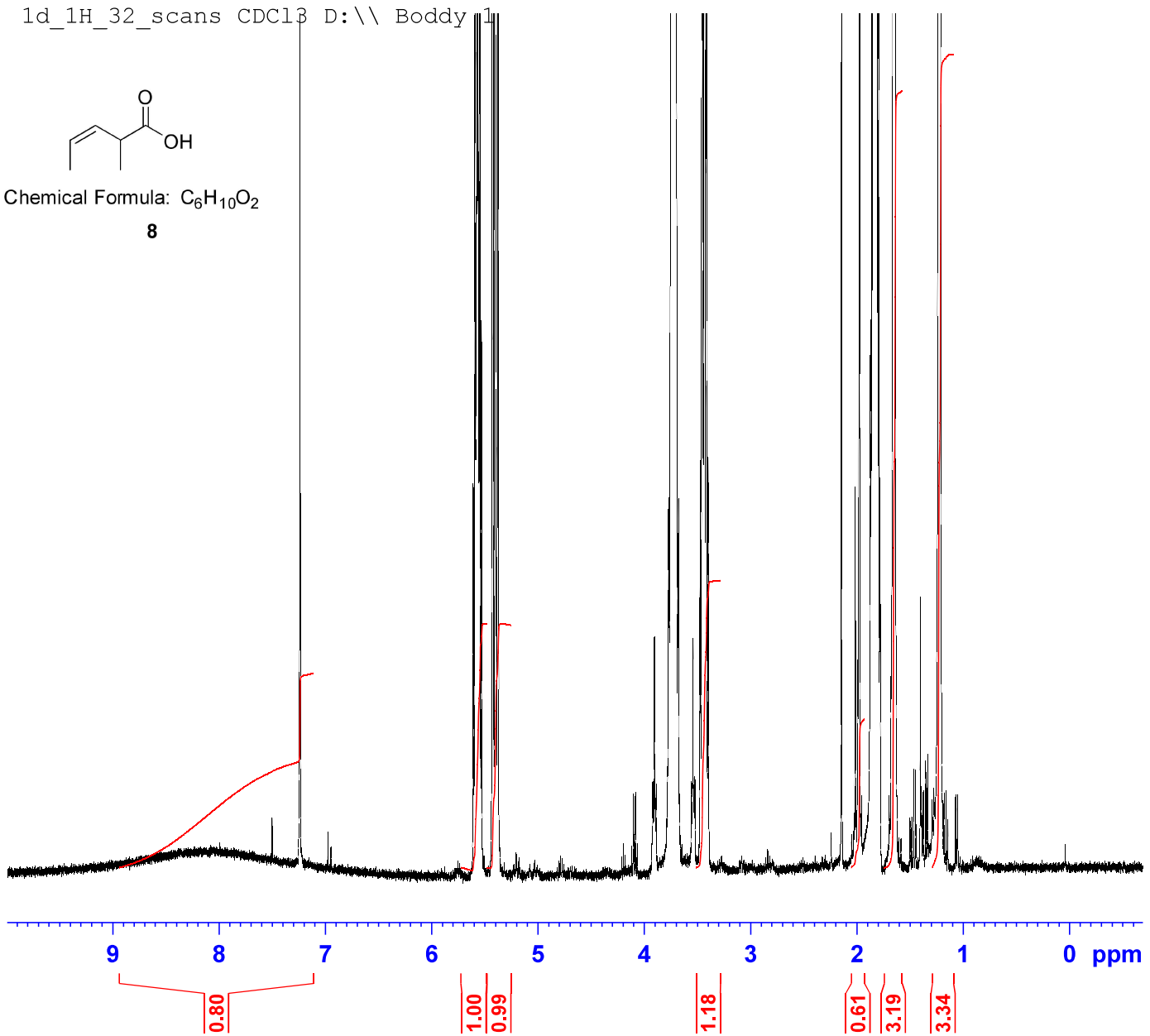


Figure IV- 7 400 MHz d_6-DMSO ^{1}H NMR of *cis*-3-methyl-2-pentenoyl-SNAC

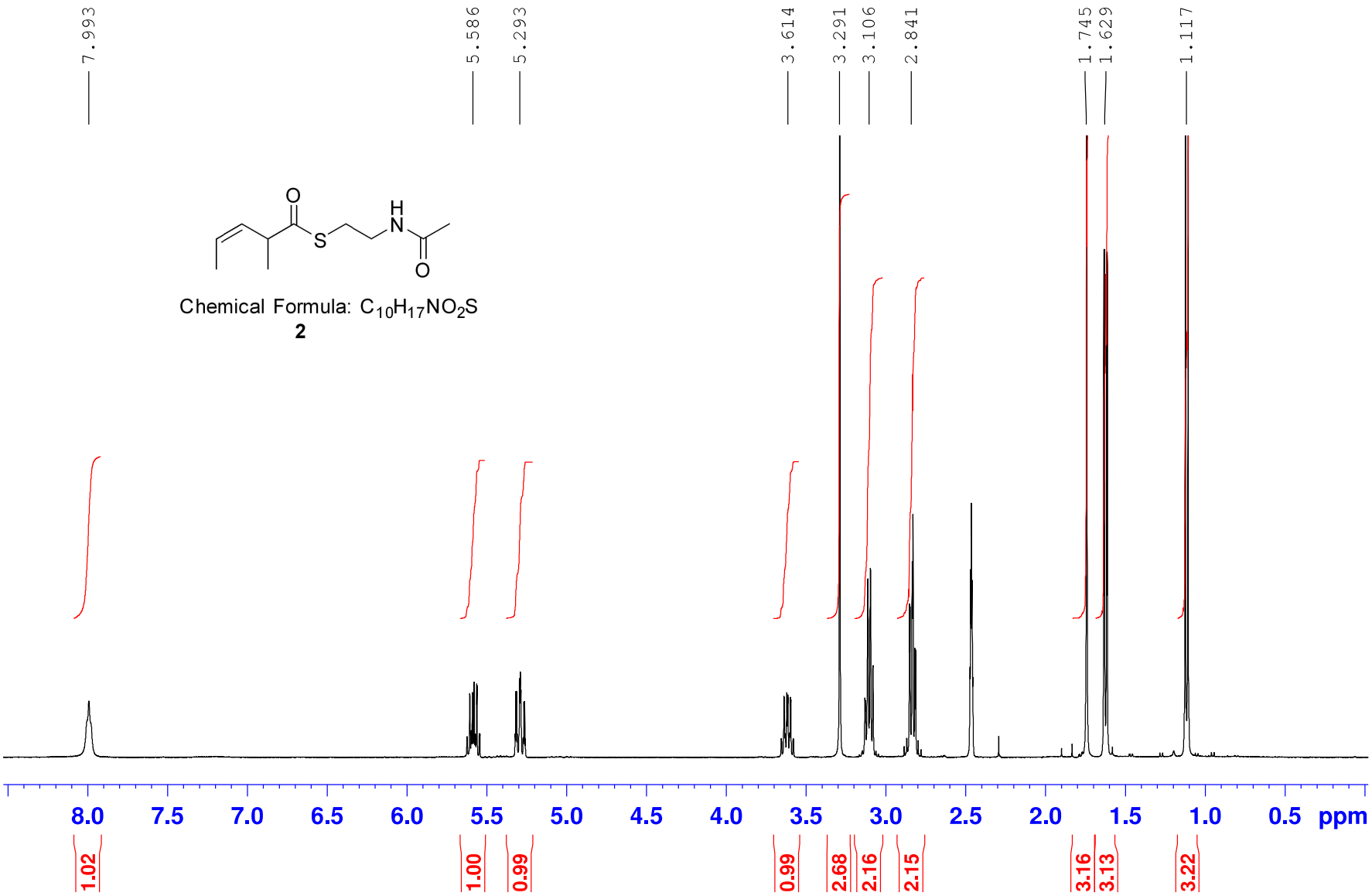


Figure IV-8 400 MHz $CDCl_3$ ^1H NMR of *(2E,4E)*-Hexadienoyl-SNAC

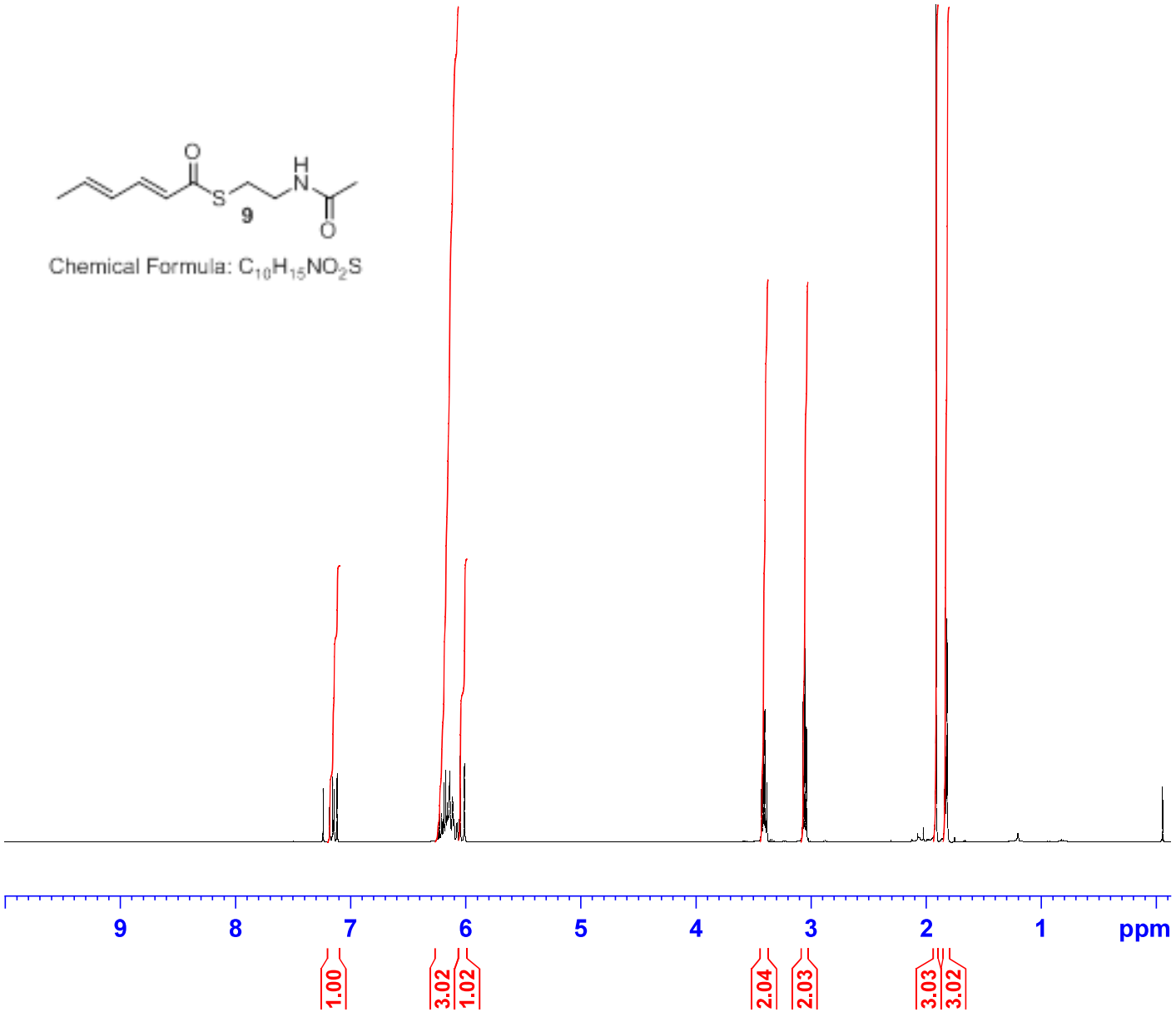


Figure IV-9 ^{13}C NMR of *(2E,4E)*-Hexadienoyl-SNAC

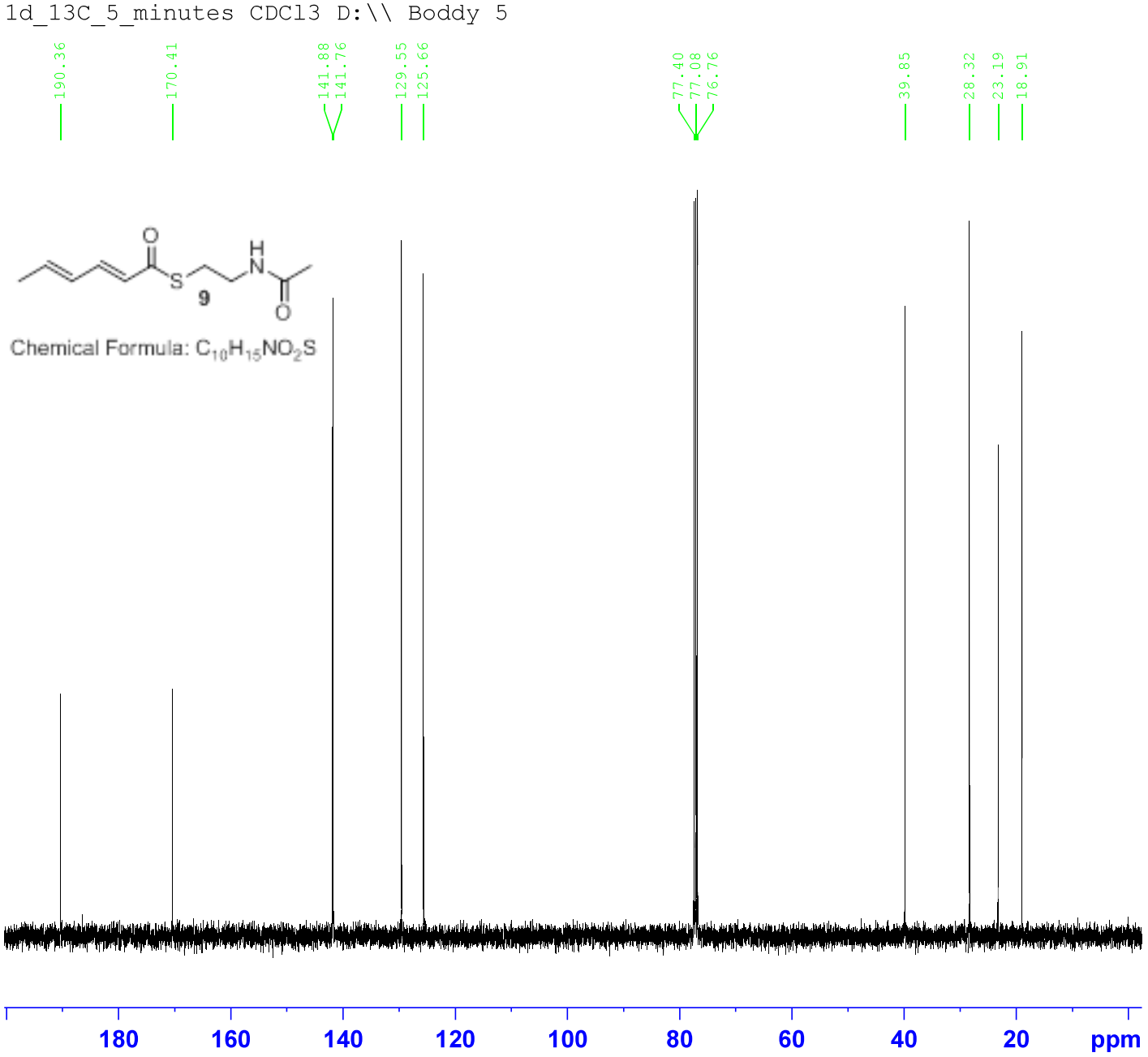


Figure IV-10 HPLC purification of racemic 2-methyl-valeric-CoA

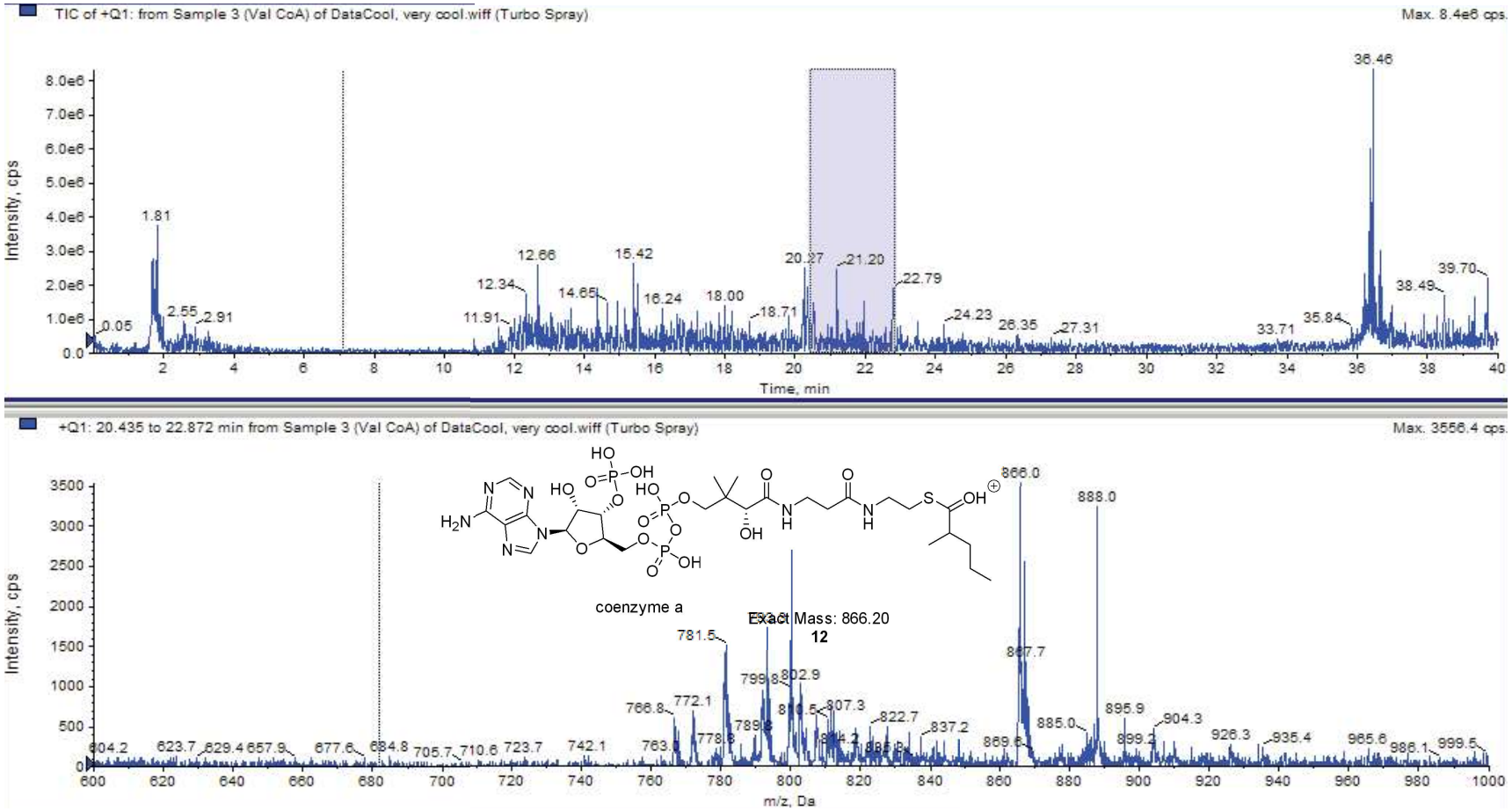


Figure IV-11 400 MHz $CDCl_3$ ^1H NMR of 2-methyl-valeric-CoA

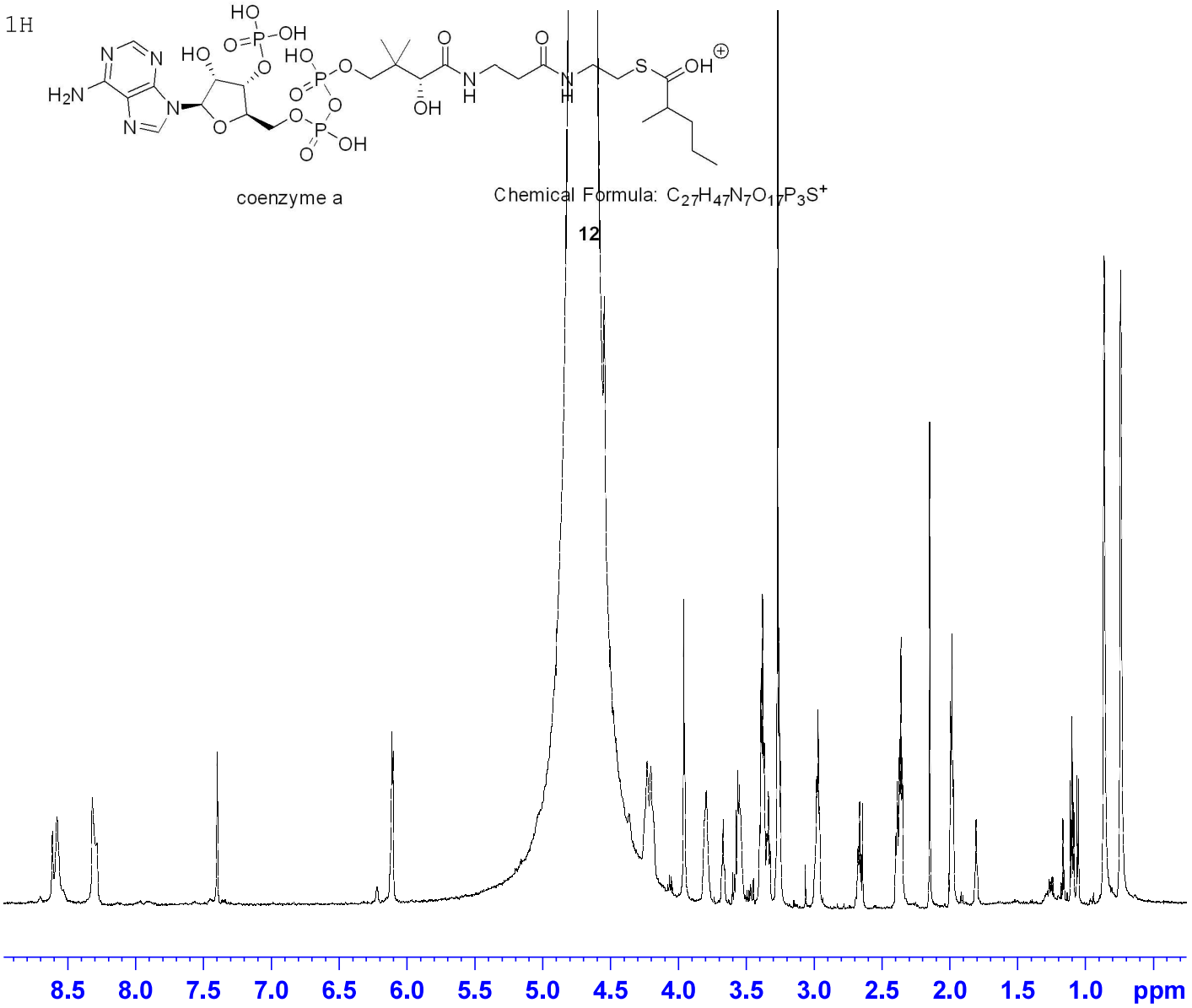


Figure IV-12 LCMS detection of hydrolyzed *trans*-2-methyl-2-pentenoyl-CoA