Investigation into the Activity and Specificity of the Thioesterase II, Ketoreductase and Dehydratase Domains of Multiple Polyketide Synthases

By

Heather Brianna Claxton

A dissertation submitted in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy
(Biological Chemistry)
In the University of Michigan
2011

Doctoral Committee:

Professor Janet Smith, Chair
Professor David Engelke
Professor David Sherman
Associate Professor Bruce Palfey
Assistant Professor Sylvie Garneau-Tsodikova

2011

To my ever patient and loving husband,

with whom I look forward to sharing my next grand adventure,

and to my parents,

who have been there for me since the beginning.

Acknowledgements

First of all, I would like to thank my thesis advisor, Janet Smith, for giving me the opportunity to work with her to complete my scientific endeavors, and for allowing me to pursue other opportunities that lead towards a career outside of academia. I'd also like to thank my thesis committee members: Dr. David Engelke, Dr. David Sherman, Dr. Bruce Palfey, and Dr. Sylvie Garneau-Tsodikova, for the many suggestions and insights that were shared with me throughout my candidacy.

I'd like to thank Suzanne admiral for all her help with the RifR thioesterase II project. Specifically I'd like to thank her for allowing me to pursue this project, for performing the acyl-ACP experiments, and for creating the RifM1 protein. I'd like to thank Monica Silver for creating the RifR expression plasmid as well as producing this protein for thioesterase II project. Without her help, this project would have never happened.

I'd like to thank all of my fellow smith lab members for their help, and for their many discussions about science and otherwise. Special thanks goes to David Akey for solving the structures of RifR and to Jamie Razelun for crystallizing RifR. Jamie also created the expression plasmids for the KR-ACP didomains and curacin DH proteins, as well as produced these proteins for the respective projects. .

I'd like to thank my collaborator, Robert Fecik from the University of Minnesota, and his lab for producing the product analogs used in the ketoreductase project.

I'd like to thank my faculty members, classmates and staffers of the Biological Chemistry department. Special thanks to Elizabeth Goodwin for her assistance with my many questions regarding administrational procedures. And last but certainly not least, I'd like to thank my family and friends for their unwavering support and understanding when unexpected scientific results have caused me to be less than punctual. Thank you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A	Adenylation Domain
ACP	Acyl Carrier Protein
AT	Acyl Transferase
C	Condensation Domain
Cur	Curacin
DEBS	6-Deoxyerythronolide Biosynthase
DH	Dehydratase Domain
ER	Enoyl Reductase Domain
KR	Ketoreductase Domain
Pik	Pikromycin
PKS	Polyketide Synthase
NRPS	Non-Ribosomal Peptide Synthetase
PKS/NRPS	Polyketide synthase/Non-Ribosomal Peptide Synthetase hybrid
T	Thioester Domain
Te1, Te2	Tandem Thioesterase Domains
TEI	Thioesterase I Domain
TEII	Thioesterase II Domain
Tyl	Tylosin

ABSTRACT

Investigation into the Activity and Specificity of the Thioesterase II, Ketoreductase and Dehydratase Domains of Multiple Polyketide Synthases

By

Heather Brianna Claxton

Chair: Janet L. Smith

Many pharmaceuticals are derived from natural products produced by assembly-line complexes found in bacteria, fungi, and plants. Assembly-line complexes, such as non-ribosomal peptide synthetases (NRPS) and type I polyketide synthases (PKSs), are composed of modules that work in succession to synthesize complex products. Each module extends and modifies an attached intermediate, then passes it to the next module in series. The modifications performed are determined by the domains that comprise each module. This dissertation focuses on two PKS domains, dehydratases and ketoreductases, and a class of stand-alone proteins, thioesterase IIs (TEIIs), that are associated with many NRPSs and PKSs.

The structure and substrate specificity of RifR, the TEII of the rifamycin NRPS/PKS complex was investigated by solving its crystal structure, the first of a TEII, and by testing its activity with a variety of substrate mimics. TEIIs are predicted to remove aberrant intermediates from acyl carrier proteins (ACPs) or peptide carrier proteins (PCPs) of PKSs or NRPSs, respectively. RifR was shown to prefer acyl-ACP

substrates over acyl-CoA substrate mimics, and aberrant decarboxylated acyl moieties over productive carboxylated acyl moieties. The structures of RifR in multiple crystal forms and its similarity to other thioesterases suggest RifR undergoes a conformation change during catalysis.

The substrate specificity of ketoreductase domains (KRs), which are responsible for reducing the β-carbonyl of the growing polyketide intermediate, was investigated using KR-ACP didomains from the erythromycin, pikromycin and tylosin pathways. KR-ACPs were shown to be active toward non-natural substrates. KR substrate or product mimics were loaded onto the ACP portion of the KR-ACP. Some mimics competed with non-natural substrates, providing insight into the specificity of the KR domains.

A working assay to detect dehydration and hydration by the dehydratase domains (DHs) of the curacin A PKS/NRPS pathway was developed. Little information is available regarding the substrate specificity of type I PKS dehydratases. While all the curacin DHs were active, dehydration of substrates occurred at different rates.

A better understanding of the structure and substrate specificity of these domains and proteins will help future attempts to reengineer existing assembly-line complexes to produce newer, more potent pharmaceuticals.

Chapter 1
Introduction

Assembly-line complexes

Assembly-line complexes are multifunctional proteins composed of modules that work in succession to synthesize complex chemical compounds, many of which are precursors of potent antibiotics, immunosuppressants, anti-tumor agents and other bioactive compounds (Fig 1.1) (1). Polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS), are two types of assembly-line complexes typically found in bacteria, fungi, and plants. PKS and NRPS modules extend and modify growing intermediates that are passed from module to module in precise order. The intermediate compounds are covalently attached through a thioester linkage to the phosphopantetheine arm (Ppant) of carrier domains, one associated with each module, until they are released from the synthase as the final product (2). The specific substrates and modifications used to create the final product are determined by the domains comprised by each module, with the number and type of modifying domains varying from module to module and complex to complex (1).

There is hope that by altering the domains and modules in a complex, one might customize an assembly-line complex to create designer compounds (2). Ultimately, this tool would be useful in drug development as antibiotics and immunosuppressants could easily be tweaked to produce more powerful drugs in a cheaper manner (3). Endeavors into this realm have garnered promising, though limited, success. Thus, further insights into the structure and mechanisms of these domains are needed to reach these goals.

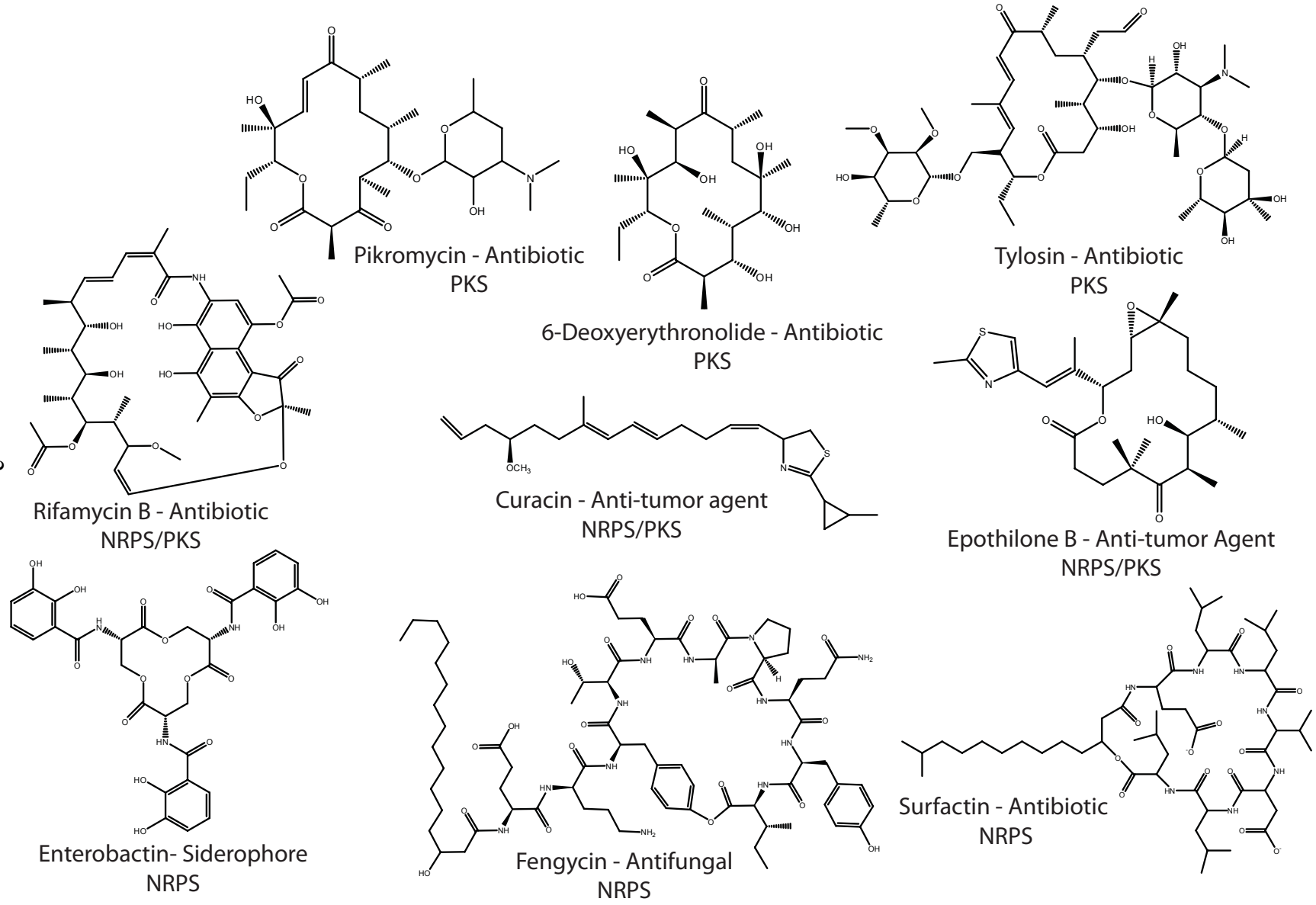


Figure 1.1 Natural products produced by assembly-line complexes.

Type I Polyketide Synthase Modules

Type I polyketide synthases efficiently build a wide range of compounds from a few basic acyl-CoA substrates through the use of consecutive, multifunctional modules. (Fig 1.2) The functions performed by each module depend on the domains that comprise the module, with the number and identity of these domains varying from module to module within a complex. A type I PKS complex module may be classified as either an elongation, an initiation or termination module.

An elongation module is responsible for 1) loading the next acyl building block onto the acyl carrier domain (ACP), 2) catalyzing a condensation reaction between the new building block and the completed intermediate from the previous module, 3) then modifying the new intermediate based on the active domains located in the module. (Fig. 1.2) Elongation modules are limited to using acyl building blocks that contain a β-carboxyl, which becomes the necessary leaving group for the condensation reaction between the building block and the upstream intermediate. This condensation reaction elongates the intermediate by a two-carbon unit. Modifying reactions are typically performed on the β-carbon of the new intermediate.

An initiation module is responsible for loading the first building block, an acyl unit, onto the complex. These modules typically consist of an acyltransferase domain (AT) and an acyl carrier domain (ACP). Active modifying or condensing domains have not been found in these modules. Because initiation modules do not catalyze condensation reactions, a β-carboxyl is not required in the acyl building block – leading to a variety of possible starting acyl building blocks.

Like the elongation modules, the termination module must also load an acyl building block onto the ACP, catalyze the condensation between the upstream intermediate and the new acyl building block, and may modify the β-carbon of the new intermediate. In addition, the termination module also removes the final product from the complex. This last role is typically performed by a thioesterase domain located in the terminal module. The final product may be released through hydrolysis, creating a linear product, or through cyclization, creating a macrolactone.

Figure 1.2 Reactions performed by a canonical PKS module.
The acyltransferase domain (AT) loads an acyl unit (malonyl- in this case) and transfers it to the Ppant arm of the ACP domain. An upstream module transfers its completed intermediate to the ketosynthase domain (KS), which then catalyzes a decarboxylative condensation reaction between the completed intermediate on the upstream ACP with the acyl unit attached to the ACP. The ketoreductase domain (KR) uses NADPH to reduce the carbonyl to an alcohol, which is then removed by the dehydrase domain (DH), leaving a double bond. The third modification domain in this module, the enoyl reductase domain (ER), uses NADPH to remove the double bond and the product is transferred to the next module in the series.

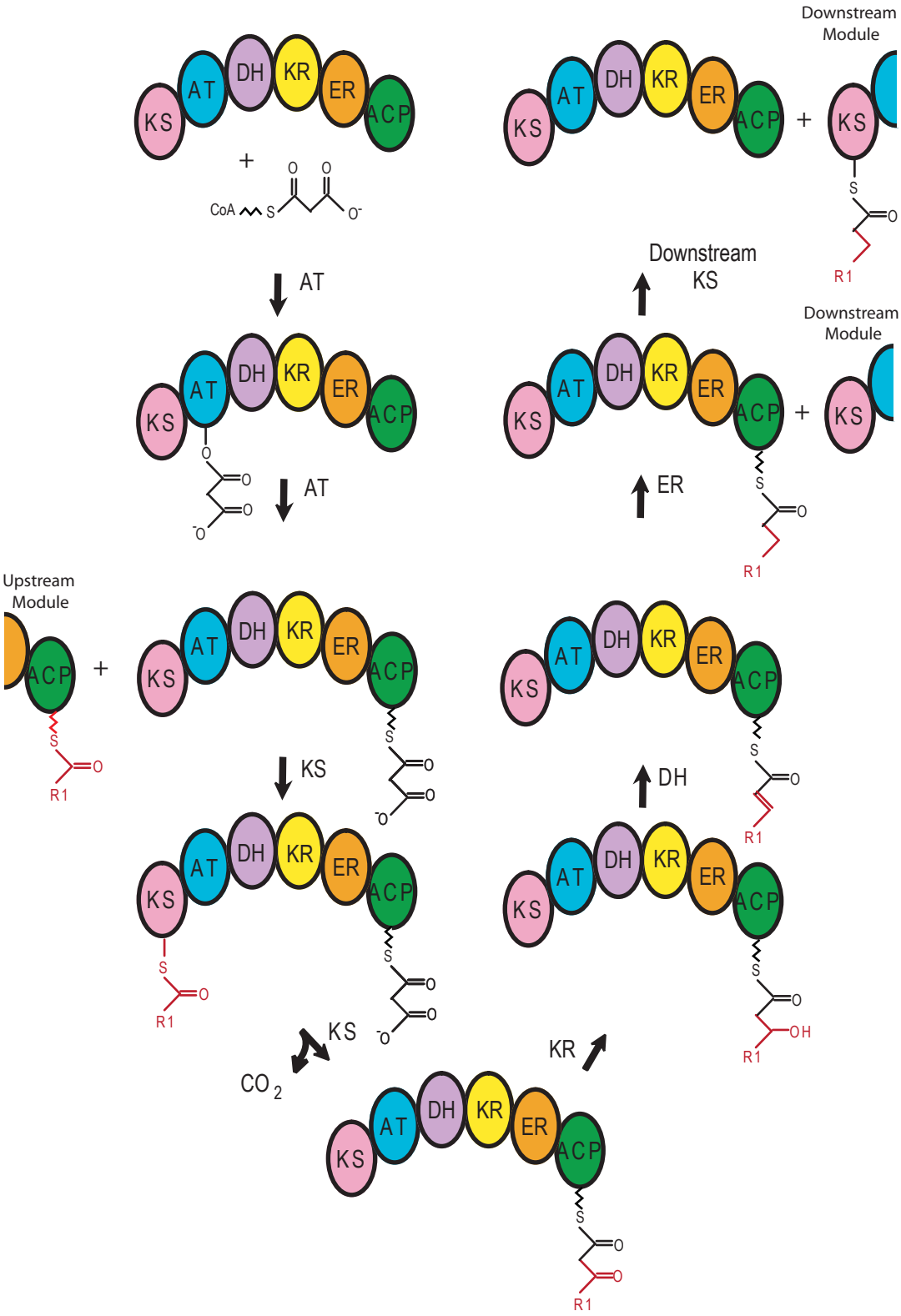


Figure 1.2 Reactions performed by a canonical PKS module.

Elongation domains

A minimal type I PKS module consists of only those domains required to elongate the growing intermediate, a ketosynthase domain (KS), acyltransferase domain (AT) and an acyl carrier domain (ACP) (4,5).

The AT domain is responsible for selecting the acyl building block and transferring it from the Ppant arm of a CoA to the Ppant arm of the ACP domain. AT domains of elongation and termination modules are highly specific for the acyl group, generally the malonyl or (2S)-methylmalonyl unit (6-8). While AT domains located in the initiation module may choose from a variety of acyl building blocks *in vitro*, one is typically preferred over all others (8-12). The preference of elongation AT domains for either malonyl- or methylmalonyl- building blocks may be determined by conserved motifs (6,7).

The ketosynthase domain (KS) is responsible for loading the completed intermediate from the upstream module onto itself, then catalyzing a Claisen condensation between the upstream intermediate and the new acyl building block attached to the Ppant of the ACP. The result of this reaction is a new intermediate that has been elongated by two carbons, and the release of CO_2. The stereochemistry of any substituent at the α position is inverted from the (2S) isomer, chosen by the AT domain, to the (2R) isomer upon condensation (13). The specificity and tolerance for unnatural upstream intermediates can vary significantly between KS domains of different modules (14).

Modification Domains

Similar to the fatty acid synthase (FAS), type I PKS modules can contain modifying domains such as the ketoreductase domain (KR), the dehydratase domain (DH), and enoyl reductase domain (ER). These domains modify the β-carbon of the growing intermediate through reduction. Unlike the FAS however, not all of these domains are present, or active, in every module. The presence or absence of these domains dictates the oxidation level of correlating regions in the final product.

The KR domain utilizes NADPH to reduce the β-carbonyl to an alcohol. The stereospecificity and stereotolerance of KR domains varies from module to module, with some domains producing only the (3S) isomer or (3R) isomer. The DH domain eliminates water from the C-2 position, producing an α-β double bond. The ER domain further reduces

the α-β double bond to an alkane. When annotating a PKS pathway, modifying domains are identified through sequence alignment and the oxidative levels in correlating regions of the final product. It is not unusual for a modifying domain to be present but inactive due to mutation.

Type II and Type III Polyketide Synthases.

In addition to type I polyketide synthases, there exist type II and type III polyketide synthases. Though these latter synthases are not assembly-line complexes, they share domain features with type I PKSs making their investigation relevant to the understanding of type I PKS domain function and structure.

Type II PKSs are similar to type II FASs. They are composed of stand-alone enzymes that work iteratively to produce a growing polyketide chain on a single ACP. In these pathways, elongating enzymes, such as the α-KS enzyme and the β-KS enzyme first create the backbone of the growing intermediate polyketide (15). After completion of the backbone, modifying enzymes, such as a ketoreductase or oxygenase, will modify the intermediate-ACP until it is released from the ACP as the final product (16). Unlike the type I PKS, in which each elongation and modification is performed by a different enzyme domain, the elongation enzymes and modification enzymes in the type II PKS work iteratively to produce the final product. Structures of type II KR domains has offered insights into the specificity and function of type I KR domains (17-19).

Type III polyketide synthases or, chalcone synthases, are structurally and mechanistically different from type I and type II PKSs (20). Instead of an ACP, the growing intermediates are attached to CoA. A single protein complex is responsible for elongating the intermediate, with decarboxylation, condensation, cyclization and aromatization reactions taking place in a single active site (21). Investigations into the structure, function, and specificity of these type III PKS systems may help with future attempts to reengineer PKS pathways to produce designer compounds.

Non-Ribosomal Peptide Synthetases

Non-ribosomal peptide synthetases are assembly-line complexes that utilize natural or unnatural amino acids (22) to produce bioactive polypeptides, such as the antibiotics

surfactin, and cyclosporine A (23). Like type I PKSs, an NRPS complex is composed of an initiation module, a termination module and many elongation modules. The active domains that compose each module determine the modifications performed and the appropriate amino acid building blocks used to create the final product.

The elongation domains of an NRPS module include the condensation domain (C), the adenylation domain (A) and the peptide carrier protein (PCP) (Fig. 1.3). The adenylation domain is responsible for selecting and adenylating, with ATP, the appropriate amino acid building block, which is then transferred to the PCP domain. The condensing domain catalyzes the condensation between the completed upstream intermediate and the new building block. Additionally, the C domain may catalyze cyclization to the peptide NH.

There are many types of NRPS modification domains. Examples include, epimerization domains, which epimerize the α-carbon of the new intermediate (23), and N-methyltransferase domains, which add a methyl group to the new intermediate (23). Like the type I PKS, the presence or absence of these domains determines the modifications performed by the module. By better understanding how these domains function, their structure, and mechanisms of specificity, one may alter an NRPS to create designer compounds (23).

Hybrid PKS and NRPS Systems

Hybrid PKS/NRPS systems are assembly-line complexes that utilize both NRPS modules and PKS modules to create the final product. Examples of these hybrid systems include the epothilone system, which contains one NRPS module and nine PKS modules (24,25), the curacin system, which contains at least nine PKS modules and two NRPS modules (26,27), and the rifamycin system which contains nine PKS modules and a single NRPS initiation module (28). In the case of curacin, PKS and NRPS modules are located on the same polypeptide. Combined NRPS and PKS modules in a single assembly-line complex result in an even greater diversity of natural products (24,25) and are of significant interest in attempts to create novel products from engineered pathways.

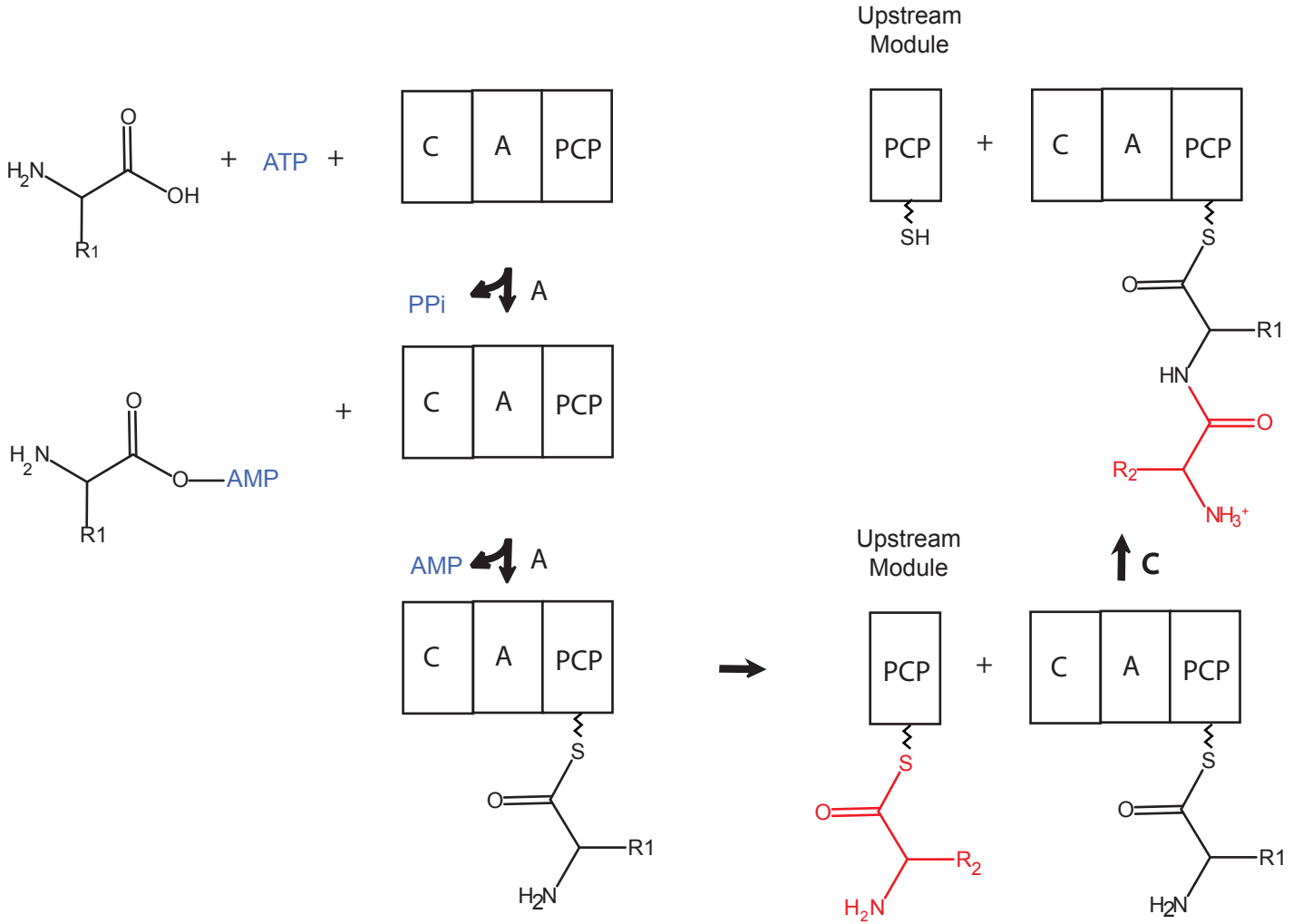


Figure 1.3 Elongation by an NRPS minimal module.
The adenylation domain (A) activates the amino acid building block from ATP and transfers it to the Ppant of the PCP domain. The condensation domain (C) catalyzes the condensation between the completed intermediate on the upstream module and the building block on the PCP domain.

Progress Toward Reengineering PKSs and NRPSs

Assembly-line complexes have been a significant resource for drug discovery. By the 1990s, over 80% of clinically available drugs were derived from natural products (29). The modularity of assembly-line complexes promises the ability to recombine domains and modules to create designer compounds (30). Nearly 200 novel compounds have been successfully created through combinatorial biosynthesis, yet many reengineered assembly-line complexes show decreased catalytic efficiency and product yields (3,31). A better understanding of the specificity and structure of assembly-line complexes is needed to address these challenges.

Novel final products may be produced by changing the building blocks used to create the final product. The specificity of an AT domain for a PKS may be changed from methylmalonyl- to malonyl- units through specific mutations in the active site (32,33). The result of these mutations has been a mix of natural products with either a malonyl or methylmalonyl starting building blocks, showing that while a novel final product could be made, complete alteration of specificity from methylmalonyl to malonyl building blocks was not observed.

Other attempts at altering the incorporation of a specific building block include swapping an AT domain for another of the desired specificity. This was performed with the geldanamycin PKS in which the AT domains of the first and fourteenth modules of the rapamycin PKS were substituted for one of the seven AT domains of the geldanamycin PKS. Of these substitutions, four of the six AT domains substitutions were active in the new module, however, decreased catalytic yields and a mix of products were observed (34).

Adenylation domain swapping (35) has shown similar challenges and moderate success. The adenylation domain controlling the specificity of the seventh amino acid incorporated in surfactin biosynthesis was swapped with adenylation domains that incorporated leucine, ornithine, pheynylalanine, cysteine and valine (35). Again, mixtures of the final natural product and the desired, novel product were obtained. Adenylation domain mutations have also resulted in altered substrate specificity, though with a similar decrease in catalytic efficiency (36). The benefits of relaxing the specificity of an adenylation domain include the ability to create peptide libraries consisting of a variety of amino acids at specified positions (37) for biological screening purposes.

Changing the state of reduction of intermediates may be performed by the inactivation or addition of modifying domains (3). The DH-KR domains of the second module of the avermectin PKS were successfully swapped with the DH-ER-KR domains from the fourth module of the pikromycin resulting in small amounts of the desired final product (3). Stereochemical modification of alcohol moieties can be performed by swapping KR domains as was done with the KR2 of the rapamycin pathway with the KR2 of the DEBS pathway (38).

Along with domain swapping, there have been attempts to swap full modules of a PKS. Interdomain linkers between modules located on the same polypeptide as well as docking domains which serve to identify and bind consecutive modules located on different polypeptides are crucial for module swapping (39,40). Attempts at mixing and matching modules from differing assembly-line complexes have proved successful (39), though further investigations will be needed before *de novo* identification of complementary modules is possible.

Overview

Though significant progress has been made towards modifying and utilizing assembly-line complexes to create novel compounds, further investigations into the structure and specificity of these complexes is still needed. Herein lies an investigation of a thioesterase II domain, ketoreductase domains and dehydratase domains of type I PKS modules.

The structure and specificity of the thioesterase II of the rifamycin NRPS/PKS pathway shows that this editing domain is able to remove a variety of acyl units from the Ppant arm of CoA and a holo-ACP domain with preference for decarboxylated substrates. Ketoreductase-ACP fusion didomains from the pikromycin, tylosin and DEBS PKSs retain reduction activity. Small chain affinity labels loaded onto the ACPs of the tylosin KR1-ACP1 and tylosin KR7-ACP7 didomains may inhibit reduction of unnatural substrates. Preliminary results show that four dehydratase domains from the curacin NRPS/PKS pathway are able to dehydrate the unnatural substrate 3-hydroxybutyryl-CoA to crotonyl-CoA. These four dehydratases are also able to catalyze the reverse reaction, hydration of the unnatural substrate crotonyl-CoA to 3-hydroxbutyryl-CoA, at differing rates.

Works Cited

1. Buchholz, T. J., Kittendorf, J. D., Sherman, D. H., and Begley, T. P. (2007) *Polyketide Biosynthesis: Modular Polyketide Synthases*, John Wiley & Sons, Inc.
2. Cane, D. E., Walsh, C. T., and Khosla, C. (1998) Harnessing the Biosynthetic Code: Combinations, Permutations, and Mutations. *Science* **282**, 63-68
3. Zhang, X., Chen, Z., Li, M., Wen, Y., Song, Y., and Li, J. (2006) Construction of Ivermectin Producer by Domain Swaps of Avermectin Polyketide Synthase in Streptomyces Avermitilis. *Applied Microbiology and Biotechnology* **72**, 986-994
4. Zhang, J., Van Lanen, S. G., Ju, J., Liu, W., Dorrestein, P. C., Li, W., Kelleher, N. L., and Shen, B. (2008) A Phosphopantetheinylating Polyketide Synthase Producing a Linear Polyene to Initiate Enediyne Antitumor Antibiotic Biosynthesis. *Proceedings of the National Academy of Sciences* **105**, 1460-1465
5. Kittendorf, J. D., and Sherman, D. H. (2009) The methymycin/pikromycin pathway: A model for metabolic diversity in natural product biosynthesis. *Bioorganic & Medicinal Chemistry* **17**, 2137-2146
6. Marsden, A. F. A., Caffrey, P., Aparicio, J. F., Loughran, M. S., Staunton, J., and Leadlay, P. F. (1994) Stereospecific Acyl Transfers on the Erythromycin-Producing Polyketide Synthase. *Science* **263**, 378-380
7. Haydock, S. F., Aparicio, J. F., Molnár, I., Schwecke, T., Khaw, L. E., König, A., Marsden, A. F. A., Galloway, I. S., Staunton, J., and Leadlay, P. F. (1995) Divergent Sequence Motifs Correlated with the Substrate Specificity of (Methyl)malonyl-CoA: Acyl Carrier Protein Transacylase Domains in Modular Polyketide Synthases. *FEBS Letters* **374**, 246-248
8. Wiesmann, K. E. H., Cortés, J., Brown, M. J. B., Cutter, A. L., Staunton, J., and Leadlay, P. F. (1995) Polyketide Synthesis *In Vitro* on a Modular Polyketide Synthase. *Chemistry & Biology* **2**, 583-589
9. Lau, J., Cane, D. E., and Khosla, C. (2000) Substrate Specificity of the Loading Didomain of the Erythromycin Polyketide Synthase. *Biochemistry* **39**, 10514-10520
10. Kao, C. M., Katz, L., and Khosla, C. (1994) Engineered Biosynthesis of a Complete Macrolactone in a Heterologous Host. *Science* **265**, 509-512
11. Kao, C. M., Luo, G., Katz, L., Cane, D. E., and Khosla, C. (1995) Manipulation of Macrolide Ring Size by Directed Mutagenesis of a Modular Polyketide Synthase. *Journal of the American Chemical Society* **117**, 9105-9106
12. Pieper, R., Luo, G., Cane, D. E., and Khosla, C. (1995) Remarkably Broad Substrate Specificity of a Modular Polyketide Synthase in a Cell-Free System. *Journal of the American Chemical Society* **117**, 11373-11374
13. Holzbaur, I. E., Ranganathan, A., Thomas, I. P., Kearney, D. J., Reather, J. A., Rudd, B. A., Staunton, J., and Leadlay, P. F. (2001) Molecular Basis of Celmer's Rules: Role of the Ketosynthase Domain in Epimerisation and Demonstration that Ketoreductase Domains can have Altered Product Specificity with Unnatural Substrates. *Chem Biol* **8**, 329-340
14. Watanabe, K. (2003) Understanding Substrate Specificity of Polyketide Synthase Modules by Generating Hybrid Multimodular Synthases. *The Journal of Biological Chemistry* **278**, 42020-42026
15. Zhan, J. (2009) Biosynthesis of Bacterial Aromatic Polyketides. *Current Topics in Medicinal Chemistry* **9**, 1958-1610

16. Hertweck, C., Luzhetskyy, A., Rebets, Y., and Bechthold, A. (2007) Type II Polyketide Synthases: Gaining a Deeper Insight into Enzymatic Teamwork. *Natural Product Reports* **24**, 162-190
17. Korman, T. P., Hill, J. A., Vu, T. N., and Tsai, S. C. (2004) Structural Analysis of Actinorhodin Polyketide Ketoreductase: Cofactor Binding and Substrate Specificity. *Biochemistry* **43**, 14529-14538
18. Korman, T. P., Tan, Y. H., Wong, J., Luo, R., and Tsai, S. C. (2008) Inhibition Kinetics and Emodin Cocrystal Structure of a Type II Polyketide Ketoreductase. *Biochemistry* **47**, 1837-1847
19. Zheng, J., Taylor, C. A., Piasecki, S. K., and Keatinge-Clay, A. T. (2010) Structural and Functional Analysis of A-Type Ketoreductases from the Amphotericin Modular Polyketide Synthase. *Structure* **18**, 913-922
20. Abe, I., and Morita, H. (2003) Structure and Function of the Chalcone Synthase Superfamily of Plant Type III Polyketide Synthases. *Natural Product Reports* **27**, 809-838
21. Staunton, J., and Weissman, K. J. (2001) Polyketide biosynthesis: a Millennium Review. *Natural Product Reports* **18**, 380-416
22. Rausch, C. (2005) Specificity Prediction of Adenylation Domains in Nonribosomal Peptide Synthetases (NRPS) Using Transductive Support Vector Machines (TSVMs). *Nucleic Acids Research* **33**, 5799-5808
23. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chemical Reviews* **97**, 2651-2673
24. O'Connor, S., Chen, H., and Walsh, C. (2002) Enzymatic Assembly of Epothilones: the EpoC Subunit and Reconstitution of the EpoA-ACP/B/C Polyketide and Nonribosomal Peptide Interfaces. *Biochemistry* **41**, 5685-5694
25. Rath, C. M., Scaglione, J. B., Kittendorf, J. D., Sherman, D. H., Lew, M., and Hung-Wen, L. (2010) NRPS/PKS Hybrid Enzymes and Their Natural Products. in *Comprehensive Natural Products II*, Elsevier, Oxford
26. Akey, D. L., Razelun, J. R., Tehranisa, J., Sherman, D. H., Gerwick, W. H., and Smith, J. L. (2010) Crystal Structures of Dehydratase Domains from the Curacin Polyketide Biosynthetic Pathway. *Structure* **18**, 94-105
27. Chang, Z., Sitachitta, N., Rossi, J. V., Roberts, M. A., Flatt, P. M., Jia, J., Sherman, D. H., and Gerwick, W. H. (2004) Biosynthetic Pathway and Gene Cluster Analysis of Curacin A, an Antitubulin Natural Product from the Tropical Marine Cyanobacterium Lyngbya majuscula *Journal of Natural Products* **67**, 1356-1367
28. Yu, T.-W., Shen, Y., Doi-Katayama, Y., Tang, L., Park, C., Moore, B. S., Richard Hutchinson, C., and Floss, H. G. (1999) Direct Evidence that the Rifamycin Polyketide Synthase Assembles Polyketide Chains Processively. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9051-9056
29. Li, J. W. H., and Vederas, J. C. (2009) Drug Discovery and Natural Products: End of an Era or an Endless Frontier? *Science* **325**, 161-165
30. Ding, Y., Sherman, D. H., Lew, M., and Hung-Wen, L. (2010) The Role of Synthesis and Biosynthetic Logic. in *Comprehensive Natural Products II*, Elsevier, Oxford

31. Kittendorf, J. D., and Sherman, D. H. (2006) Developing Tools for Engineering Hybrid Polyketide Synthetic Pathways. *Current Opinion in Biotechnology* **17**, 597-605
32. Del Vecchio, F. (2003) Active-Site Residue, Domain and Module Swaps in Modular Polyketide Synthases. *Journal of Industrial Microbiology & Biotechnology* **30**, 489-494
33. Reeves, C. D., Murli, S., Ashley, G. W., Piagentini, M., Hutchinson, C. R., and McDaniel, R. (2001) Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations *Biochemistry* **40**, 15464-15470
34. Patel, K., Piagentini, M., Rascher, A., Tian, Z.-Q., Buchanan, G. O., Regentin, R., Hu, Z., Hutchinson, C. R., and McDaniel, R. (2004) Engineered Biosynthesis of Geldanamycin Analogs for Hsp90 Inhibition. *Chemistry & Biology* **11**, 1625-1633
35. Stachelhaus, T., Schneider, A., and Marahiel, M. (1995) Rational Design of Peptide Antibiotics by Targeted Replacement of Bacterial and Fungal Domains. *Science* **269**, 69-72
36. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999) The Specificity-Conferring Code of Adenylation Domains in Nonribosomal Peptide Synthetases. *Chemistry & Biology* **6**, 493-505
37. Walsh, C. T. (2002) Combinatorial Biosynthesis of Antibiotics: Challenges and Opportunities. *ChemBioChem* **3**, 124-134
38. Kao, C. M., McPherson, M., McDaniel, R. N., Fu, H., Cane, D. E., and Khosla, C. (1998) Alcohol Stereochemistry in Polyketide Backbones Is Controlled by the β-Ketoreductase Domains of Modular Polyketide Synthases. *Journal of the American Chemical Society* **120**, 2478-2479
39. Menzella, H., and Reeves, C. (2007) Combinatorial Biosynthesis for Drug Development. *Current Opinion in Microbiology* **10**, 238-245
40. Chandran, S. S., Menzella, H. G., Carney, J. R., and Santi, D. V. (2006) Activating Hybrid Modular Interfaces in Synthetic Polyketide Synthases by Cassette Replacement of Ketosynthase Domains. *Chemistry & Biology* **13**, 469-474

Chapter 2

RifR TEII

Three types of α/β hydrolase thioesterases are associated with type I PKS and NRPS assembly-line complexes, thioesterase I (TEI), tandem thioesterases (Te1, Te2), and thioesterase II (TEII). The final product of the assembly-line complex may be removed by either a TEI, tandem Tes, or a less common method such as by an amide synthase or hydrolysis. The putative function of a TEII, however, is not to remove the final product from the assembly line, but to remove intermediates from the synthase. A better understanding of the structure, function, and specificity of these thioesterases is needed to expand the assembly line reprogramming toolkit.

Though some exceptions occur (1), TEIs are usually integrated into the C-terminal end of the final module of the PKS, NRPS, or the C-terminus of the mFAS, and are responsible for removing the final product from the synthases through hydrolysis or cyclization. Like PKS modules, PKS TEIs such as PikTE and DEBSTE from the pikromycin and 6-deoxyerthronolide biosynthesis pathways, respectively are dimers with the dimerization domain located N-terminal to the α/β hydrolase fold comprising the core of the thioesterase. Both NRPS modules and NRPS TEIs are monomers and lack the dimerization domain found in PKS TEIs (Fig. 2.1). The TEI of the human mFAS is a monomer, while the mFAS complex is dimeric (2).

In some cases, two thioesterases are integrated into the final module of the NRPS (3,4). These Tes are more similar to NRPS TEIs than PKS TEIs (3). Removal of either Te results in significant decrease of product yield (100-95%). Both thioesterases are required to remove the final product from the assembly complex through cyclization (3).

TEIIs were first discovered in the mFAS cluster of lactating rat mammary glands (5), and later found to be common in both PKS and NRPS biosynthetic clusters. Similar to NRPS TEIs, TEIIs are monomers. Unlike TEIs and tandem Tes, TEIIs are not integrated into any module, allowing the TEII to act on multiple modules in the synthase. TEIIs remove substrates from the synthase through hydrolysis, not cyclization. The specificities of TEIIs vary greatly from pathway to pathway.

Structures have been reported for seven PKS/NRPS/FAS thioesterases: crystal structures for the TEIs from the pikromycin PKS (PDB 2HFJ, PikTE) (6), 6-deoxyerythronolide B PKS (PDB 1MO2, DEBSTE) (7), surfactin NRPS (PDB 1JMK, SrfTE) (8), fengycin NRPS (PDB 2CB9, FenTE) (9), and human fatty acid synthase (PDB 1XKT, hFasTE) (2) systems, and NMR structures for enterobactin TEI (PDB 2ROQ, EntF) (10), and surfactin TEII (PDB 2RON, SrfTEII) (11). There are no structures currently available for NRPS tandem thioesterases.

All TEs have an α-helical insertion after strand β5, which forms a lid over the active site. Additionally, in the PKS TEIs, the N-terminal dimer-forming helices contribute to the lid structure, forming a fixed channel which runs the length of the TE and contains the active site.

In contrast, the active site pocket of monomeric NRPS TEIs and TEIIs is flexible: two conformations of the lid and active site pocket were observed in the surfactin TEI (SrfTEI) (12) structure and chemical shift observations suggested greater flexibility for residues of the lid region in the surfactin TEII (SrfTEII) (11) solution structure. These movements seem to be of functional importance, as a movement of a linker peptide in SrfTEI determines the shape of the active site pocket, and a movement of the first lid helix appears to modulate access to the active site (8).

TEIs and TEIIs contain the cannonical Ser-His-Asp catalytic triad found in thioesterases, however, the placement of the active site Asp residue is located before the lid region in TEIs and after the lid region in TEIIs (Fig. 2.1). Only PKS TEIs contain extra residues forming two helices that compose a dimerization domain N-terminal to the α/β hydrolase core. These extra helices are not found in the monomeric mFAS TEI, NRPS TEI, nor the TEIIs (Fig. 2.1). TEIIs have greater sequence similarity to other TEIIs from either NRPS or PKS systems than to the TEI of the same pathway (13), but

TEIIs from both PKS and NRPS pathways have greater sequence similarity to PKS TEIs than to NRPS TEIs (13).

Figure 2.1 Structure-based sequence alignment of TEIs, TEIIs and Tes.
PKS/NRPS/FAS TEIIs, TEIs and Tes were first aligned using the EXPRESSO(3DCoffee) server (14), then manually edited based on 3-D structures. Conserved residues were highlighted with Jalview 2.4 (15). Asterisks mark the active site residues of both TEIs, TEIIs and Tes. TEII sequences are highlighted in pink, FAS/NRPS TEIs in blue, PKS TEIs in green and Tes in beige. The lid regions of RifR, SrfTEI, and PikTE are highlighted in yellow in the secondary structure map, the α/β hydrolase cores are highlighted in grey. Thioesterases used in this alignment are further described in Table 2.1.

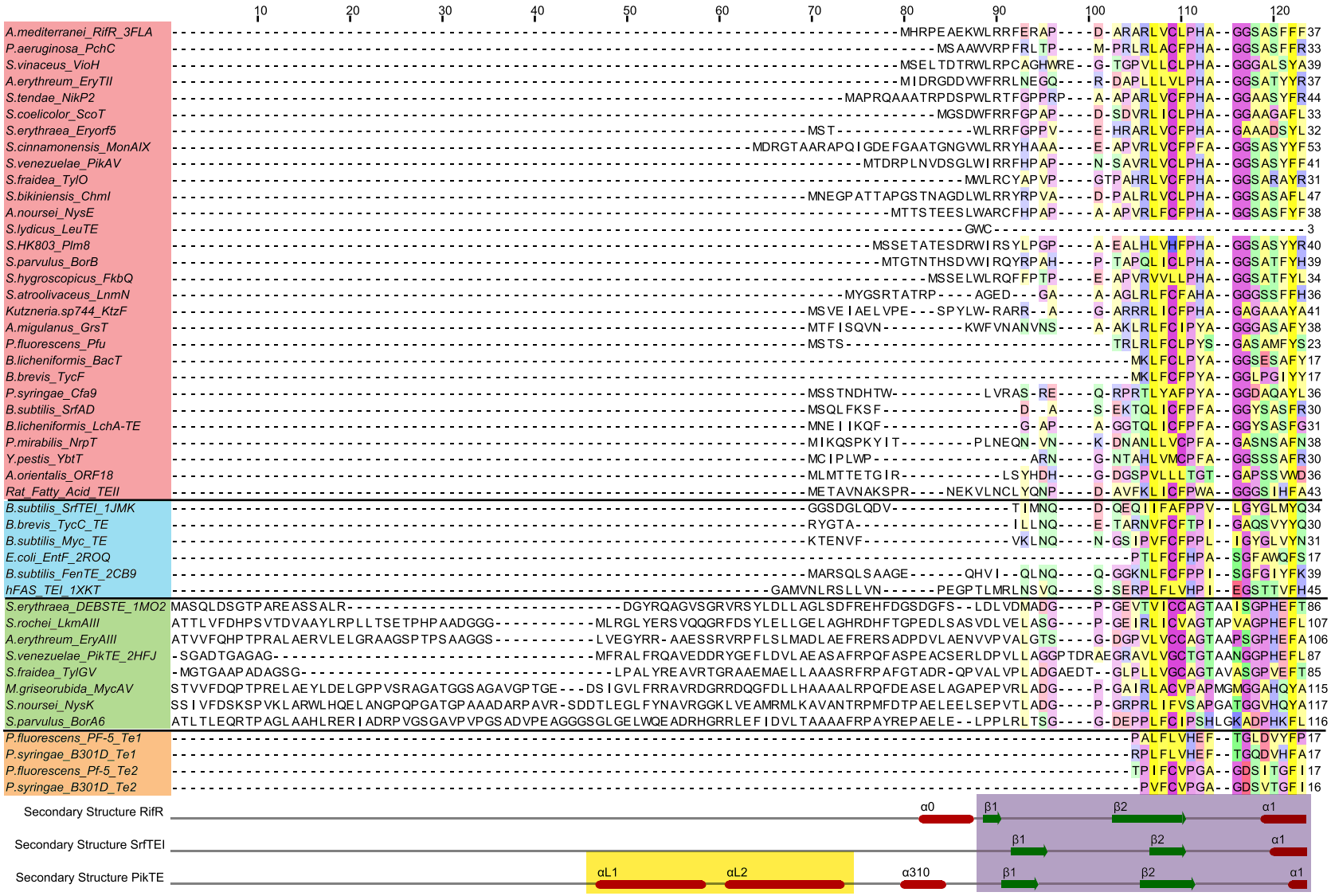
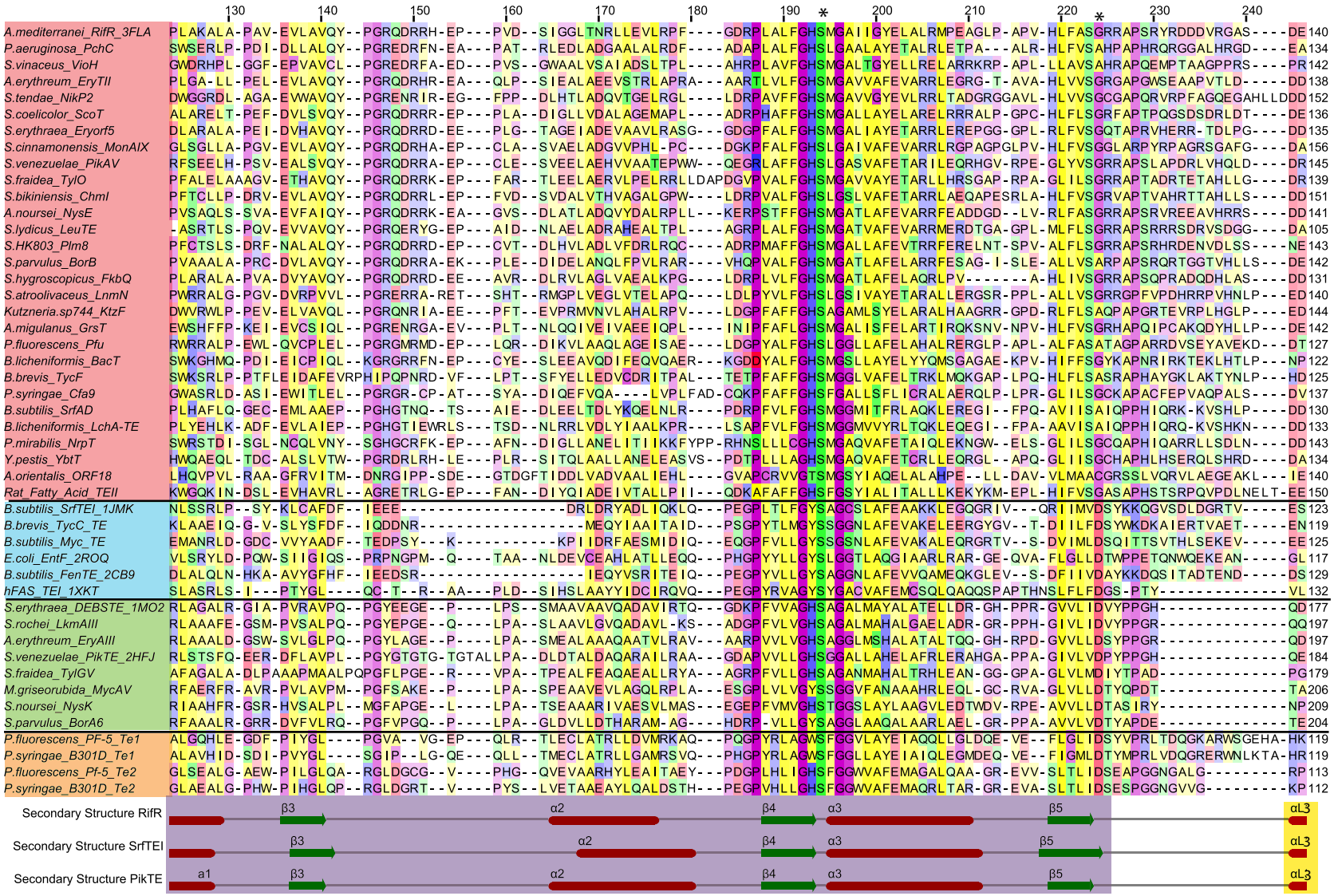
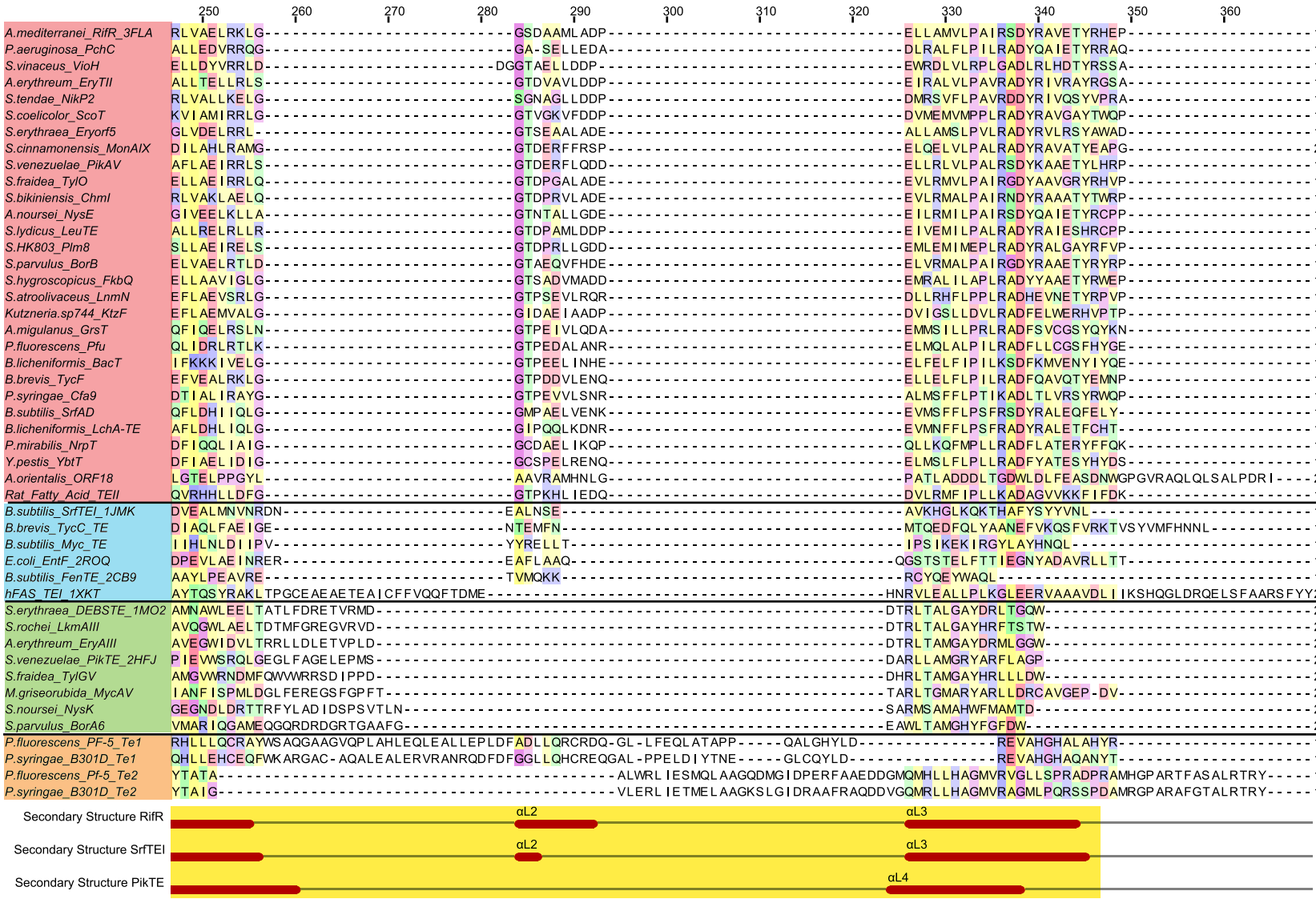


Figure 2.1 Structure-based sequence alignment of TEIs, TEIIs and Tes.


130
140
150
160
170
180
190
200
210
220
230
240
A.mediterranei_RifR_3FLA
P.aeruginosa_PchC
S.vinaceus_VioH
A.erythreum_EryTII
S.tendae_NikP2
S.coelicolor_ScoT
S.erythraea_Eryorf5
S.cinnamonensis_MonAIX
S.venezuelae_PikAV
S.fraidea_TylO
S.bikiniensis_ChmI
A.noursei_NysE
S.lydicus_LeuTE
S.HK803_Plm8
S.parvulus_BorB
S.hygroscopicus_FkbQ
S.atroolivaceus_LnmN
Kutzneria.sp744_KtzF
A.migulanus_GrsT
P.fluorescens_Pfu
B.licheniformis_BacT
B.brevis_TycF
P.syringae_Cfa9
B.subtilis_SrfAD
B.licheniformis_LchA-TE
P.mirabilis_NrpT
Y.pestis_YbtT
A.orientalis_ORF18
Rat_Fatty_Acid_TEII
B.subtilis_SrfTEI_1JMK
B.brevis_TycC_TE
B.subtilis_Myc_TE
E.coli_EntF_2ROQ
B.subtilis_FenTE_2CB9
hFAS_TEI_1XKT
S.erythraea_DEBSTE_1MO2
S.rochei_LkmAIII
A.erythreum_EryAIII
S.venezuelae_PikTE_2HFJ
S.fraidea_TylGV
M.griseorubida_MycAV
S.noursei_NysK
S.parvulus_BorA6
P.fluorescens_PF-5_Te1
P.syringae_B301D_Te1
P.fluorescens_Pf-5_Te2
P.syringae_B301D_Te2
Secondary Structure RifR
Secondary Structure SrfTEI
Secondary Structure PikTE
a1
β3
α2
β4
α3
β5
αL3


250
260
270
280
290
300
310
320
330
340
350
360
A.mediterranei_RifR_3FLA
P.aeruginosa_PchC
S.vinaceus_VioH
A.erythreum_EryTII
S.tendae_NikP2
S.coelicolor_ScoT
S.erythraea_Eryorf5
S.cinnamonensis_MonAIX
S.venezuelae_PikAV
S.fraidea_TylO
S.bikiniensis_ChmI
A.noursei_NysE
S.lydicus_LeuTE
S.HK803_Plm8
S.parvulus_BorB
S.hygroscopicus_FkbQ
S.atroolivaceus_LnmN
Kutzneria.sp744_KtzF
A.migulanus_GrsT
P.fluorescens_Pfu
B.licheniformis_BacT
B.brevis_TycF
P.syringae_Cfa9
B.subtilis_SrfAD
B.licheniformis_LchA-TE
P.mirabilis_NrpT
Y.pestis_YbtT
A.orientalis_ORF18
Rat_Fatty_Acid_TEII
B.subtilis_SrfTEI_1JMK
B.brevis_TycC_TE
B.subtilis_Myc_TE
E.coli_EntF_2ROQ
B.subtilis_FenTE_2CB9
hFAS_TEI_1XKT
S.erythraea_DEBSTE_1MO2
S.rochei_LkmAIII
A.erythreum_EryAIII
S.venezuelae_PikTE_2HFJ
S.fraidea_TylGV
M.griseorubida_MycAV
S.noursei_NysK
S.parvulus_BorA6
P.fluorescens_PF-5_Te1
P.syringae_B301D_Te1
P.fluorescens_Pf-5_Te2
P.syringae_B301D_Te2
Secondary Structure RifR
Secondary Structure SrfTEI
Secondary Structure PikTE
αL2
αL3
αL4

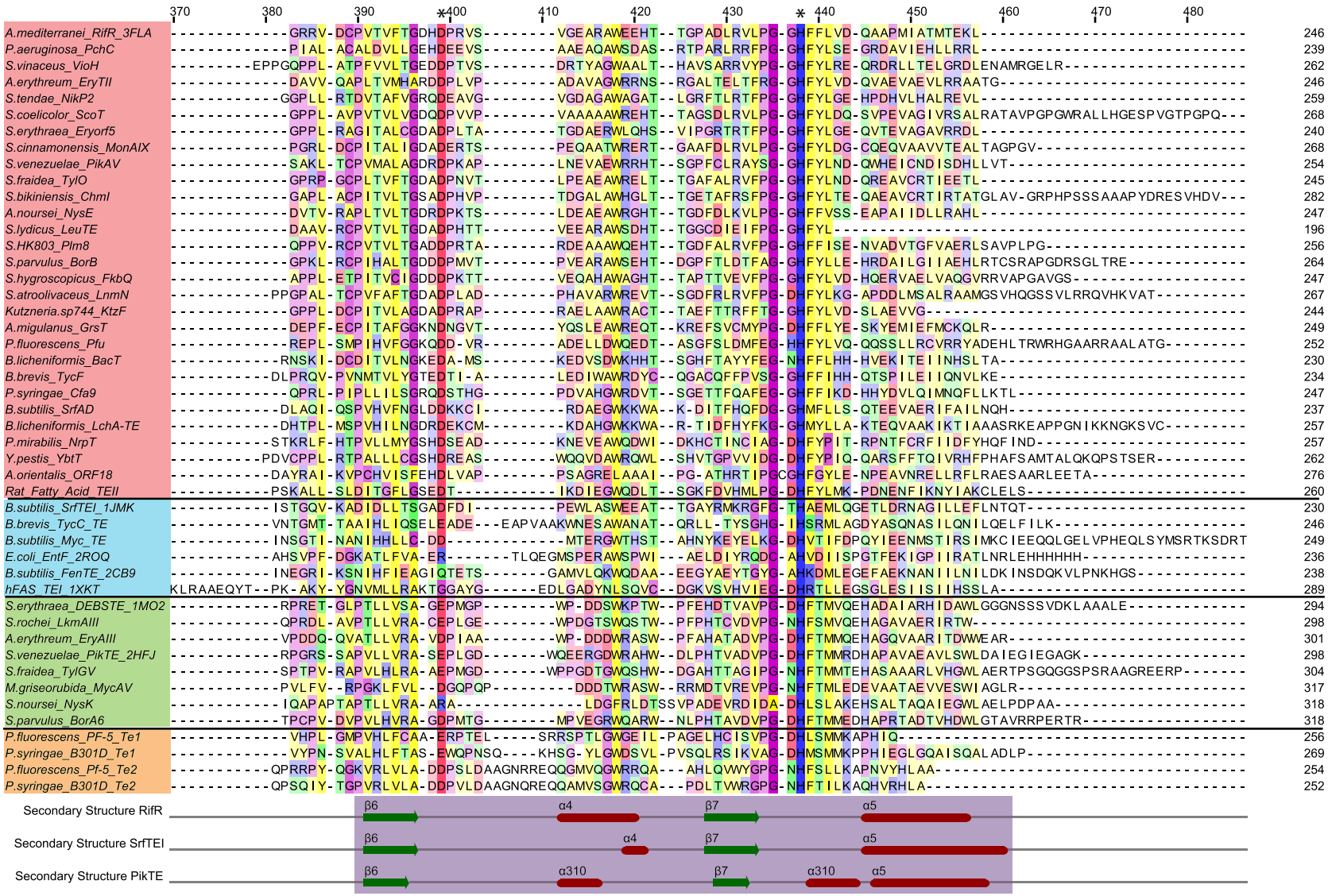

370
380
390
400
410
420
430
440
450
460
470
480
A.mediterranei_RifR_3FLA
P.aeruginosa_PchC
S.vinaceus_VioH
A.erythreum_EryTII
S.tendae_NikP2
S.coelicolor_ScoT
S.erythraea_Eryorf5
S.cinnamonensis_MonAIX
S.venezuelae_PikAV
S.fraidea_TylO
S.bikiniensis_ChmI
A.noursei_NysE
S.lydicus_LeuTE
S.HK803_Plm8
S.parvulus_BorB
S.hygroscopicus_FkbQ
S.atroolivaceus_LnmN
Kutzneria.sp744_KtzF
A.migulanus_GrsT
P.fluorescens_Pfu
B.licheniformis_BacT
B.brevis_TycF
P.syringae_Cfa9
B.subtilis_SrfAD
B.licheniformis_LchA-TE
P.mirabilis_NrpT
Y.pestis_YbtT
A.orientalis_ORF18
Rat_Fatty_Acid_TEII
B.subtilis_SrfTEI_1JMK
B.brevis_TycC_TE
B.subtilis_Myc_TE
E.coli_EntF_2ROQ
B.subtilis_FenTE_2CB9
hFAS_TEI_1XKT
S.erythraea_DEBSTE_1MO2
S.rochei_LkmAIII
A.erythreum_EryAIII
S.venezuelae_PikTE_2HFJ
S.fraidea_TylGV
M.griseorubida_MycAV
S.noursei_NysK
S.parvulus_BorA6
P.fluorescens_PF-5_Te1
P.syringae_B301D_Te1
P.fluorescens_Pf-5_Te2
P.syringae_B301D_Te2
Secondary Structure RifR
Secondary Structure SrfTEI
Secondary Structure PikTE
β6
α4
β7
α5
α310

Table 2.1 TEIIs, TEIs, and Tes in alignment

Accession Number	*Name*	*Biosynthetic Molecule*	*NRPS/PKS/ FAS*
TEIIS			
AAG52991, 2FLA	A.mediterranei_RifR_2FLA	rifamycin	NRPS/PKS
CAA57967	P.aeruginosa_PchC	salicylate, pyochelin, dha	NRPS
AAP92498.1	S.vinaceus_VioH	viomycin (tuberactinomycin)	PKS
AAU93793	A.erythreum	erythromycin	PKS
CAC11138	S.tendae_NikP2	nikkomycin	NRPS
AAF43096.2	S.coelicolor3_ScoT	actinorhodin	NRPS
AAA26497	S.erythraea_eryOrf5	erythronolide (DEBS)	PKS
AA065795	S.cinnamonensis_MonAIX	monensin	PKS
AAC69333	S.venezuelae_PikAV	pikromycin	PKS
AAA21345	S.fraidea_TylO	tylocine	PKS
AAS79448	S.bikiniensis_ChmI	chalcomycin	PKS
AAF71777	A.noursei_NysE	nystatin	PKS
AAZ20309	S.lydicus_LeuTE	streptolydigin	PKS/NRPS
AAQ84143	S.HK803_Plm8	phoslactomycin B	PKS
CAE45660.1	S.parvulus_BorB	borrelidin	PKS
AAF86400	S.hygroscopicus_ var_ascomyceticus_FkbQ	FK520 (ascomycin)	PKS
AAN85527.1	S.atroolivaceus_LnmN	leinamycin	PKS/NRPS
ABV56586	Kutzneria_KtzF	kutznerides	NRPS
P14686	A.migulanus_GrsT	gramicidin	NRPS
AAL59667	P.fluorescens_Pfu	pyoverdine	NRPS
BAA36683	B.licheniformis_bacT	bacitracin	NRPS
AAC45933	B.brevis_TycF	tyrocidin	NRPS
AAC38657	P.syringae_pv_Cfa9	coronafacic acid	PKS
Q08788	B.subtilis_SrfAD	surfactin	NRPS
CAA06326	B.licheniformis_LchA-TE	llichenysin	NRPS
AAD10392	P.mirabilis_NrpT	putative NRPS	NRPS
AAC69590	Y.pestis_YbtT	yersiniabactin	NRPS
CAA11784	A.orientalis_ORF18	chloroeremomycin	NRPS
CAA68411	Rat_FAS_TEII		FAS

Accession number	*Name*	*Biosynthetic Molecule*	*NRPS/PKS/ FAS*
TEIs			
1JMK	B.subtilis_SrfTEI_1JMK	surfactin	NRPS
AAC45930	B.brevis_TycC_TE		
AAF08797	B.subtilis_Myc	mycosubtilin	NRPS
2ROQ	E.coli_EntF_2ROQ	enterobactin	NRPS
2CB9	B.subtilis_FenTE_2CB9	fengycin	NRPS
1XKT	hFAS_TEI_1XKT	Human FAS	FAS
CAA39583.1, 1MO2	S.erythraea_DEBSTE_ 1MO2	DEBS	PKS
BAC76491	S.rochei_LkmaIII	lankamycin	PKS
AAU93805.2	A.erythreum_EryAIII	erythromycin	PKS
2HFJ	S.venezuelae_PikTE_2HFJ	Pikromycin	PKS
AAB66508.1	S.fradei_TylGV	tylosin	PKS
BAC57032.1	M.griseorubida_MycAV		PKS
AAF71768	S.noursei_NysK	nystatin	PKS
CAE45672.1	S.parvulus_BorA6	borrelidin	PKS
Te1 & Te2s			
AAY91421	P.fluorescencs_Pf-5_Te1	orfamide	NRPS
AAY91421	P.fluorescencs_Pf-5_Te2	orfamide	NRPS
AA072425	P.syringae_B301D_Te1	syringopeptin	NRPS
AA072425	P.syringae_B301D_Te2	syringopeptin	NRPS

Disruption of the TEI function results in a complete loss of product, while disruption of the TEII function results in a significant decrease in product yield (0-95%) as shown in Table 2.2. (1,3,16-23) Neither TEIs nor TEIIs may rescue the disrupted function of the other, (24,25) but a TEII may rescue the function of another disrupted TEII from a different pathway (26). The presence of a TEII can reduce the deleterious effect of misprimed carrier proteins. An increase of misprimed carrier proteins of the dihyroaeruginoate/pyochelin pathway containing PchC, the TEII of the pathway, resulted in a 37% decrease of pyochelin yield. (21) An increase of misprimed carrier proteins of the ΔPchC mutant dihyroaeruginoate/pyochelin pathway resulted in a 90% reduction of pyochelin yield. (21).

TEIIs from different pathways have different specificities, but general trends include a preference for decarboxylated acyl units over carboxylated acyl units (24,27,28), substrates linked to a carrier domain over substrates linked to CoA or the phosphopantetheine mimic SNAC (12,29) and single amino acids over di- or tri-peptides (12,24) as seen in Tables 2.3-2.5. TEIIs are able to hydrolyze substrates attached to carrier domains from their native pathway as well as other pathways (22,24,29).

Table 2.2 Knockout studies of TEIIs

Product Yield Decrease	TEII Name	Synthase Product	Host species	Reference
100%	YbT	Yersiniabactin	*Y. pestis*	Geoffroy, 2000
>95%	PikAV	Pikromycin	*S. venezuelae ATCC 15439*	Xue. 1998
95%	ArfC_Te2	Arthrofactin	*Psuedomonas sp. MIS38*	Roongsawang, 2007
85%	TylO	Tylactone	*S. fradiae*	Butler, 1999
84%	SrfA-TE (SrfTEII)	Surfactin	*B. subtilis,*	Schneider, 1998
65%	MonAIX & MonAX	Monensin A	*S. cinnamonensis*	Harvey, 2006
60%	RifR	Rifamycin	*A. mediterranei*	Doi-Katayama, 2000
60%	PchC	Dihydroaeruginoate, pyochelin	*P. aeruginosa*	Reimmann, 2004
<20%	Ery-Orf5	6-Deoxyerthronolide B	*Sac. erythraea*	Hu, 2003
0%	PikAV	Pikromycin	*S. venezuelae ATCC 15432*	Chen, S. 2001

Table 2.3 TEII activity with acyl-CoAs

	k_{cat}/K_M (M^{-1} s^{-1})	
Substrate	RifR (PKS)(31) (Claxton, 2009)	Rat (FAS)(30) (Witkowski, 1992)
Acetyl-CoA	11	
Propionyl-CoA	25	
Butyryl-CoA	13	
Isobutyryl-CoA	9.6	
Malonyl-CoA	1.5	
Methylmalonyl-CoA	1.8	
Hexanoyl-CoA	5.9	
Octanoyl-CoA	31	
Decanoyl-CoA	160	7.67E+03

Table 2.4 TEII activity with substrate mimics

	k_{cat}/K_M (M^{-1} s^{-1})		
Substrate	Tyl TEII (PKS)(27) (Heathcote, 2001)	TycF (NRPS)(12) (Yeh, 2004)	Rat (FAS)(30) (Witkowski 1992)
Acetyl-SNAC	2.5	2.2	
Propionyl-SNAC	13		
Butyryl-SNAC	6.5		
Pentanoyl-SNAC	1.7		
Diketide-SNAC	0.9		
Leu-SNAC		4	
AcetylLeu-SNAC		23	
Decanoyl-NPO			6.90E+04
Acetyl-NPO	83		
Propionyl-NPO	439		
Butyryl-NPO	306		
Pentanoyl-NPO	284		
Triketide-NPO	39		

Table 2.5 TEII activity with acyl-carrier proteins

	k_{cat}/K_M (M^{-1} s^{-1})					
Substrate	Pik TEII (PKS)(28) (Kim, 2002)	DEBS TEII (PKS)(32) (Hu, 2003)	RifR (PKS)(31) (Claxton 2009)	ZhuC (PKS)(29) (Tang 2004)	SrfTEII (NRPS)(24,33) (Linne, 2004 & Schwarzer, 2002)	Rat (FAS)(30) (Witkowski 1992)
Acetyl-AT-ACP	4.8	1.50E+04				
Acetyl-PCP		est from Methods			1.83E+06	
Propionyl-AT-ACP	15.8					
Butyryl-AT-ACP	18.3					
Malonyl-AT-ACP	3.8			0.0011		
Methylmalonyl-AT-ACP	3.3					
Methylmalonyl-PikAIII	2.8					
Decanoyl-ACP						2.10E+05
Acetyl-ACP (ZhuG)				0.44		
Acetyl-ACP (FrenN)				0.24		
Acetyl-ACP			150			
Methylmalonyl-ACP			54			
Propionyl-ACP			210	0.42		
Butyryl-ACP				0.05		
Hexanoyl-ACP				0.0021		
Octanoyl-ACP				0.0033		

Whereas TEIs are covalently attached to the terminal module and generally process only the final product of an assembly-line complex (Fig 2.2A), TEIIs are discrete proteins that can remove intermediates from any module in the complex. A variety of functions have been attributed to TEIIs, the most prevalent of which is a 'housekeeping function' – the removal of aberrant acyl units from carrier domains. These aberrant acyl units may be due to premature decarboxylation by the ketosynthase domain on PKSs (27) (Fig 2.2B) or by mispriming of the carrier domain by a promiscuous phosphopantetheinyl transferase (12,24,33) (Fig 2.2C). Other functions proposed for TEIIs include the removal of specific intermediates from the synthase as in the case of the mammary gland rat fatty acid synthase (mFAS) TEII in lactating rats, which removes medium chain C_8-C_{12} fatty acids from the ACP domain (5) (Fig 2.2D), and the removal of amino acid derivatives from an NRPS carrier domain (23,34-36) allowing these derivatives to be incorporated into the natural product by a later module in the assembly-line complex (Fig 2.2E).

Two models have been proposed for the TEII housekeeping function (27). In the high-specificity model, the TEII scans the complex and efficiently removes only aberrant acyl units. In the low-specificity model, the TEII removes both correct and incorrect acyl units from the phosphopantetheine arm (ppant) at an inefficient rate. As correct acyl units are quickly incorporated into the growing intermediate compound, incorrect acyl units stall on the carrier domain providing a longer window of opportunity for removal by a TEII. Thus a slow low-specificity enzyme can be effective.

Figure 2.2. Proposed functions of thioesterase proteins.
A. Cyclization and removal of the synthase product from the final module by a thioesterase I (TEI). **B**. Production of a decarboxylated acyl unit by the ketosynthase (KS) domain and the subsequent hydrolysis by a TEII. **C**. Mispriming of a PKS by transfer of an acyl-phosphopantetheine arm by a promiscuous phosphopantetheinyl transferase (Pptase) and the subsequent hydrolysis by a TEII. **D**. Removal of a medium chain fatty acid from the ACP of the FAS module by a TEII **E.** Hydrolysis of an amino acid derivative by a TEII from an NRPS module comprising an adenylation domain (A) and a peptide carrier protein (PCP) domain.

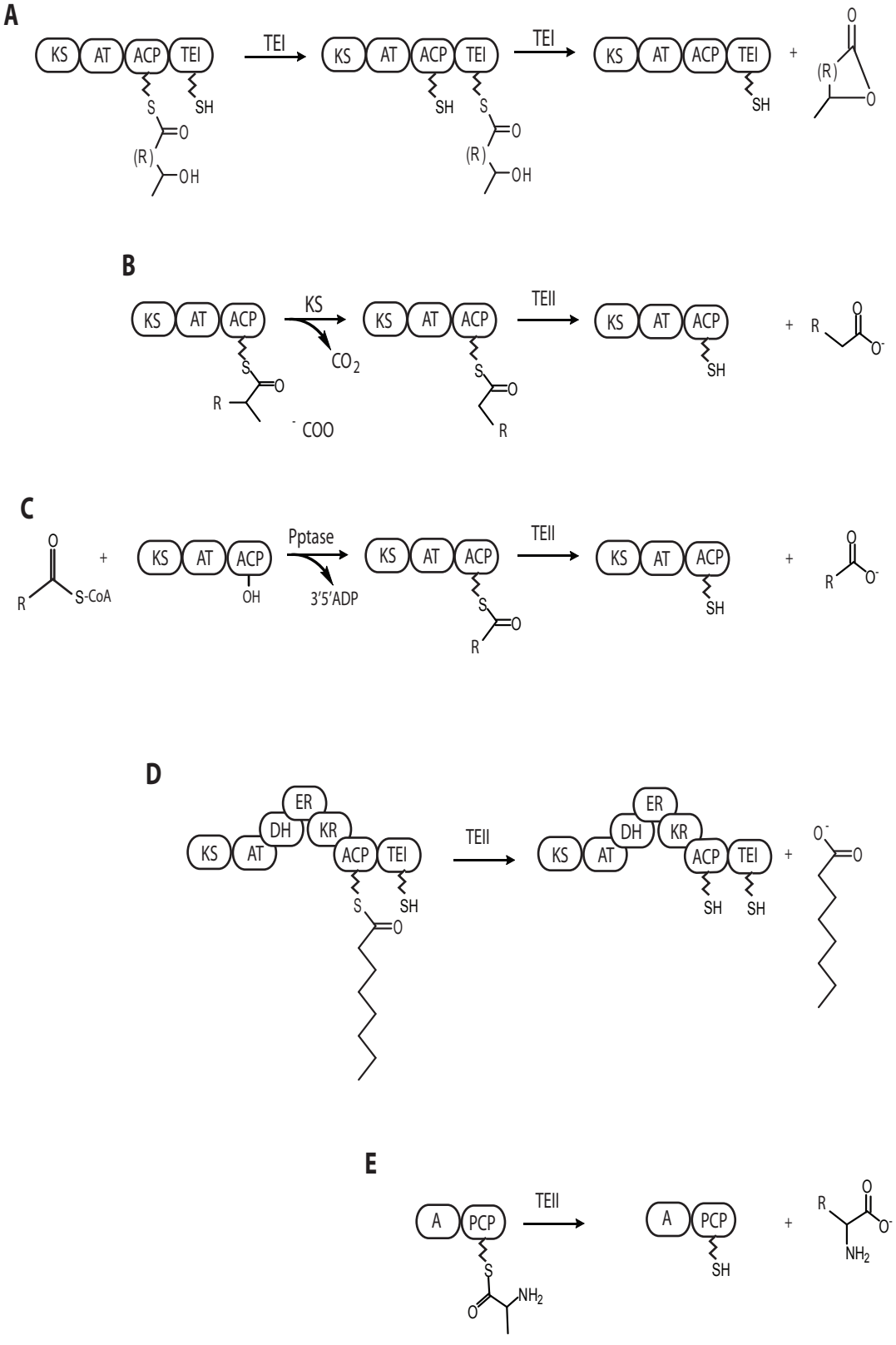


Figure 2.2 Proposed functions of thioesterase proteins.

Reported in this chapter is the structure and activity of recombinant RifR, the TEII of the rifamycin biosynthetic cluster (Figure 2.3). Proansamcyin X is produced by the rifamycin assembly-line complex, which is an NRPS/PKS hybrid system composed of six proteins comprising one NRPS-like loading module and ten PKS modules (20). The synthase does not contain a TEI. Instead, it utilizes an amide synthase, RifF, to produce the macrolactamide product.

Steady-state kinetic analysis of the hydrolytic activity of RifR on a wide range of acyl-CoA and acyl-ACP substrates demonstrates that acyl-ACP substrates are preferred over the acyl-CoAs. Aberrant, decarboxylated acyl units are processed more efficiently than the natural rifamycin building blocks, carboxylated acyl units. The crystal structure of RifR, the first for any hybrid PKS/NRPS TEII, shows an active site pocket similar to an NRPS TEI. The size and shape of the active site pocket are variable, as one of the elements forming the pocket, an extended linker segment, is highly flexible, and different crystal forms reveal different shapes for the active site pocket. Access to the active site is severely restricted, and structural comparisons to other thioesterases suggest that a conformational change in the lid and the flexible linker region is required for access to the substrate pocket.

Figure 2.3 Rifamycin NRPS/PKS hybrid biosynthase.
Proansamcyin X is the precursor of rifamycin and ultimately the anti-tuberculosis drug rifampicin. The synthase consists of 6 proteins, RifA through RifF, and an NRPS loading domain. Inactive DH domains are designated by an 'X'.

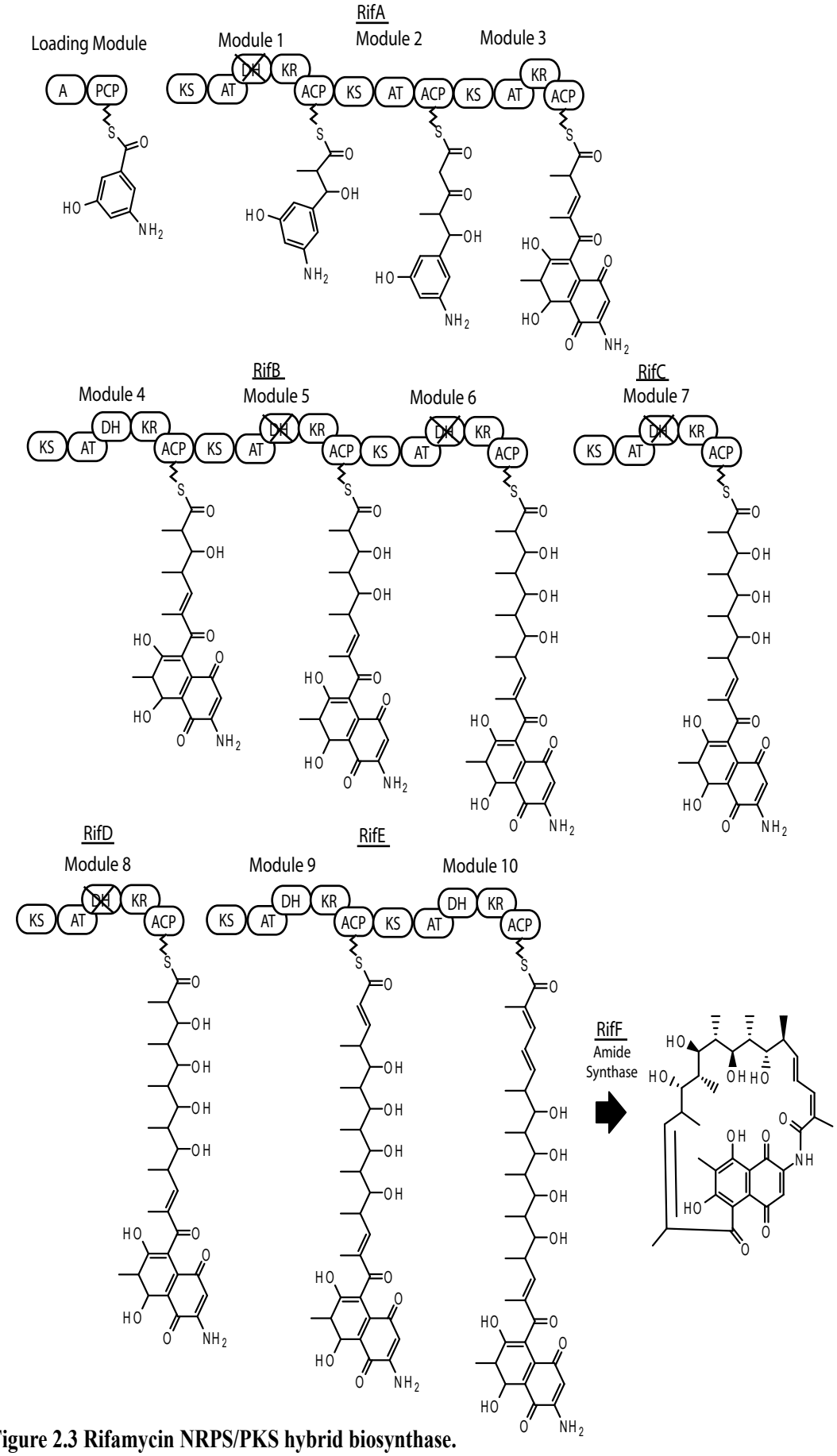


Figure 2.3 Rifamycin NRPS/PKS hybrid biosynthase.

Materials and Methods

Materials. Non-radioactive acyl-CoAs were obtained from Sigma at the highest purity available. *DL*-2-[methyl-^{14}C]-Methylmalonyl-CoA (54 mCi/mmol) and [malonyl-2-^{14}C]-malonyl-CoA (52 mCi/mmol) were obtained from PerkinElmer, and [acetyl-1-^{14}C]-acetyl-CoA (54 mCi/mmol) and [propionyl-1-^{14}C]-propionyl-CoA (53 mCi/mmol) from Moravek Biochemicals. Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and restriction enzymes were obtained from Invitrogen.

Manipulation of DNA and Strains. DNA manipulations were performed in *E. coli* Novablue (Novagen) or DH5α using standard culture conditions (37). Polymerase chain reactions were carried out using Platinum Pfx polymerase (Invitrogen) as recommended by the manufacturer.

Construction of Expression Vectors for Wild-type and S94A RifR. A DNA template for RifR, was unavailable. Instead, PCR-based gene synthesis was used to assemble the *rifR* gene (GenBank AF040570, nt 96034-96813) from a set of 34 overlapping oligonucleotides (38). The terminal 5'- and 3'-oligonucleotides were designed to flank the synthetic gene with *Nde*I and *Xho*I restriction sites, respectively. After assembly, the gene was PCR-amplified, digested with *Nde*I and *Xho*I, and ligated to pET21 (Novagen) that had also been digested with the same enzymes. This produced pMS8, an expression vector for RifR with a natural N-terminus and a hexahistidine sequence appended to its C-terminus. The identity of the *rifR* synthetic gene was confirmed by DNA sequencing. Creation of the pMS8 plasmid was done by Monica Silver. The Quikchange method (Stratagene) was used to generate the S94A mutant of RifR: the serine nucleophile of the catalytic triad was converted to alanine by mutating AGT to GCT at the appropriate location in pMS8 to give expression vector pHC2. The mutation was confirmed by sequencing.

Construction of an Expression Vector for S639A Rif M1. The natural sequence 5'-CGCGCC-3' at nt 24260-24265 (GenBank AF040570), corresponding to the C-terminal end of Rif Module1 (M1), was chosen on the basis of an alignment of DEBS and Rif

thiolation (T) domain sequences (39) for replacement with the *Spe*I recognition sequence 5'-ACTAGT-3'. The *Bsa*BI-*Spe*I fragment encoding Rif M1 was then fused to the SpeI-*Eco*RI fragment encoding the DEBS TE via replacement of the *Bsa*BI-*Spe*I fragment encoding DEBS M3 in pST132 (40) to give pSA10. The presence of the DEBS TE domain was undesirable for this study, so its coding sequence was eliminated by ligating the *Nde*I-*Spe*I fragment of pSA10 encoding RifM1 to the *Nde*I-*Nhe*I fragment of pET25b (Novagen). This yielded pMS24, an expression vector for RifM1 with hexahistidine appended to the C-terminus. The Quikchange method (Stratagene) was used to generate Rif M1 with an inactive acyltransferase (AT) domain: the active site serine of the AT domain was converted to alanine by mutating TCG at nt 21434-21436 of the original sequence to GCG to give expression vector pMS25, which was fully sequenced to confirm its identity. Creation of pMS24 and pMS25 was done by Monica Silver. Creation of pSA10 was done by Suzanne Admiraal.

Expression and Purification of Proteins. Expression plasmids were transformed into *E. coli* strain BL21 Star (DE3) (Invitrogen). One liter cultures were grown at 37°C in 2-L flasks containing LB medium supplemented with 0.1 mg/mL ampicillin. Protein expression was induced with 100 μM IPTG at an optical density at 600 nm of 0.8. After induction, incubation was continued for 20 h at 15°C. The cells were then harvested by centrifugation at 2500*g* and resuspended in 50 mM sodium phosphate (pH 8.0), 300 mM NaCl, 10 mM imidazole, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mg/mL DNase I, 10% v/v glycerol.

All purification procedures were performed at 4°C. The resuspended cells were disrupted by two passages through a French press at 16,000 psi, and the lysate was collected by centrifugation at 47,800*g* and loaded onto a Histrap HP column (1 mL, GE Healthcare) previously equilibrated with 10 mM imidazole in 50 mM sodium phosphate (pH 8), 300 mM NaCl, 10% v/v glycerol. Proteins were eluted with an imidazole gradient (10-100 mM) in the same solution.

For RifM1, pooled fractions containing S639A RifM1 were diluted with 20 mM Tris (pH 7.5), 50 mM NaCl, 1 mM EDTA, 10% v/v glycerol and loaded onto a previously equilibrated HiTrapQ HP anion exchange column (1 mL, GE Biosciences).

The column was washed with 50 mM NaCl in 20 mM Tris (pH 7.5), 1 mM EDTA, 10% v/v glycerol, and S639A Rif M1 was eluted with a NaCl gradient (50-500 mM) in the same solution. Pooled fractions containing S639A RifM1 were buffer exchanged into 50 mM HEPES (pH 7.5), 50 mM NaCl, 1 mM EDTA, 1 mM TCEP, 10% v/v glycerol and concentrated with an Amicon Ultra-15 centrifugal filter unit (Millipore).

For wild-type and S94A RifR, metal-affinity column fractions containing RifR were pooled, diluted with 20 mM Tris (pH 7.5), 50 mM NaCl, 1 mM EDTA, 10% v/v glycerol and loaded onto a previously equilibrated Mono Q 5/50 GL anion exchange column (GE Biosciences). RifR was present in the column flow-through and was buffer exchanged into 50 mM HEPES (pH 7.5), 50 mM NaCl, 1 mM EDTA, 1 mM TCEP, 10% v/v glycerol and concentrated with an Amicon Ultra-15 centrifugal filter unit (Millipore).

Purified proteins were flash-frozen in liquid nitrogen and stored at -80°C. Protein concentrations were determined using the calculated extinction coefficients (41) at 280 nm: 18,450 $M^{-1}cm^{-1}$ for RifR, and 166,840 $M^{-1}cm^{-1}$ for S639A Rif M1. Typical 1-L cultures yielded 10 mg purified RifR or 4 mg purified S639A RifM1.

Selenomethionyl (SeMet) RifR was produced with a protocol as for RifR, modified according to Guerrero *et al*. (42), in which a 50-mL overnight culture was pelleted and added to minimal media supplemented with SeMet prior to induction. Production of RifM1 was performed by Monica Silver.

Measurement of RifR Activity Toward Acyl-CoA Substrates. Starting acyl-CoA stocks contained a small amount of CoA. Acyl-CoAs (25-1000 μM) were incubated with RifR or S94A RifR (2.5-25 μM) or no enzyme in the presence of 50 mM HEPES (pH 7.5), 25 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP, 5% v/v glycerol at 25°C. To ensure accurately measurable hydrolysis for all acyl-CoAs over the same time-frame, slower hydrolyzing acyl-CoAs, (acetyl-CoA, isobutyryl-CoA, hexanoyl-CoA, malonyl-CoA and methylmalonyl-CoA (250-1000 μM)) were incubated with 25 μM RifR, and faster hydrolyzing acyl-CoAs, (butyryl-CoA, octanoyl-CoA and propionyl-CoA (25-1000 μM)) were incubated with 2.5 μM RifR. Due to its limited solubility, decanoyl-CoA was incubated at a lower concentration (25-250 μM) with RifR (2.5 μM) than were the other faster hydrolyzing substrates. At each time point, aliquots were quenched to a final

concentration of 5% trichloroacetic acid and the precipitated protein was removed by centrifugation at 20,800 x g for 5 min. The ratio of acyl-CoA to CoA in the supernatant was quantified by HPLC using a C18 reverse phase column (Altima, 5 μM, 250 x 4.6 mm) monitored by absorbance at 259 nm. Separation was performed using a modification of a published protocol (43) Briefly, a linear gradient of buffer A (75 mM potassium phosphate pH 4.5) and buffer B (0.1% trifluoroacetic acid in acetonitrile) was used at a constant flow rate of 1.0 mL/min. Initial conditions were 96% buffer A, and 4% buffer B. At 5 min, buffer B was increased to 7% over 5 min, then increased to 9% over 4 min. At 14 min, buffer B was increased to 50% over 5 min and maintained for 8 min. At 27 min, buffer B was decreased to 4% over 1 min, and the column was equilibrated at 4% buffer B for 8 min between injections. Retention times were as follows: acetyl-CoA, 18.6 min; butyryl-CoA, 20.5 min; CoA, 14.8 min; decanoyl-CoA, 22.6 min; hexanoyl-CoA, 21.4 min; isobutyryl-CoA, 20.5 min; malonyl-CoA, 13.5 min; methylmalonyl-CoA, 16.8 min; octanoyl-CoA, 22.0 min; and propionyl-CoA, 20.0 min. With the exception of isobutyryl-CoA, which was shown to saturate wild-type RifR, hydrolysis of acyl-CoAs was linearly dependent on enzyme concentration in the wild-type RifR reactions. No hydrolysis was detected in the control reactions without RifR, nor was hydrolysis observed in the S94A reactions except with isobutyryl-CoA and propionyl-CoA. Data analysis was performed using Kaleidagraph (Synergy Software). The hydrolysis progress plot was fit to the equation: CoA-fraction = $1 - (1 - \text{CoA-fraction}_0)e^{-t(kobs)}$, where CoA-fraction = [CoA]/([CoA]+[Acyl-CoA]), t = time, and kobs = V_{max}/K_M. At initial velocities, the rate of substrate consumption is equivalent to $(V_{max}/K_M)[S_0]$ where S_0 is initial substrate concentration. Initial velocities were calculated and plotted against initial substrate concentrations to create Michaelis-Menten plots.

Measurement of RifR Activity Toward Acyl-S639A RifM1 Substrates. To generate [^{14}C]-acyl-S639A Rif M1 substrates, [^{14}C]-acyl groups were installed on the *apo* T domain of S639A Rif M1 by pre-incubating the *apo* protein with [^{14}C]-acyl-CoA and the promiscuous phosphopantetheinyl transferase Sfp (44,45). Pre-incubation reactions were performed at 25°C for 60-90 min and contained 25 μM *apo* S639A RifM1, 25 μM Sfp, and 25 μM [^{14}C]-acyl-CoA in 50 mM HEPES (pH 7.5), 25 mM NaCl, 5 mM $MgCl_2$, 1

mM TCEP, 5% v/v glycerol. Aliquots of the pre-incubation reactions were then distributed into reaction tubes containing wild-type RifR, S94A RifR, or no TEII, for final reactions that consisted of varying concentrations of [^{14}C]-acyl-S639A Rif M1 (2-12 μM) and wild-type or S94A RifR (0-4 μM) in 50 mM HEPES (pH 7.5), 25 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP, 5% v/v glycerol (plus residual Sfp and the 3',5'-ADP product of pre-incubation reactions). Final reactions were incubated at 25°C, and at desired time points 10-μL aliquots were quenched in an equal volume of 10% TCA. The protein precipitate was pelleted by centrifugation, washed with 150 μL 5% TCA, and solubilized in 20 μL 2% SDS, 50 mM Tris (pH 8). This solution was combined with 5 mL of liquid scintillation fluid (Ultima Gold, Perkin Elmer), and the amount of [^{14}C]-acyl-S639A Rif M1 remaining at each time point was quantified by liquid scintillation counting. Disappearance of [^{14}C]-acyl-S639A RifM1 substrates was linearly dependent on enzyme concentration in the wild-type RifR reactions, but little or no breakdown of [^{14}C]-acyl-S639A Rif M1 substrates was observed in the no-TEII and S94A RifR control reactions. Data analysis was performed using Kaleidagraph (Synergy Software), and exponential fits to the data typically gave $R \geq 0.95$.

To determine the identity of the acyl products of the RifR reactions, the TCA supernatants of late reaction time points were analyzed by radio-HPLC. Samples were injected onto a System Gold HPLC (Beckman) equipped with an Aminex HPX-87H ion exclusion column (Bio-Rad) and a Radiomatic 150TR flow scintillation analyzer (Perkin Elmer) to separate and detect [^{14}C]-labeled species. Separations were performed isocratically in 0.008 M sulfuric acid over 30 min with a flow rate of 0.6 mL/min, and flow scintillation analysis was performed on the column eluant after it was mixed with Ultima Flo liquid scintillation fluid (Perkin Elmer) in a 1 to 2 ratio. As expected, [^{14}C]-acetate, [^{14}C]-propionate, and [^{14}C]-methylmalonate predominated in the TCA supernatants from reactions that contained [^{14}C]-acetyl-S639A Rif M1, [^{14}C]-propionyl-S639A RifM1, and [^{14}C]-methylmalonyl-S639A RifM1, respectively. However, significant amounts of both [^{14}C]-malonate and [^{14}C]-acetate were detected in TCA supernatants from [^{14}C]-malonyl-S639A RifM1 reactions; [^{14}C]-acetate presumably results from decarboxylation of [^{14}C]-malonyl-S639A RifM1 to [^{14}C]-acetyl-S639A RifM1 followed by RifR-catalyzed hydrolysis during the reaction period. The concurrent

decarboxylation of [^{14}C]-malonyl-S639A RifM1 prevented us from obtaining a reliable k_{cat}/K_M value for its hydrolysis by RifR, but the accumulation of [^{14}C]-malonate over time indicates that malonyl-S639A Rif M1 is indeed a substrate. Measurement of RifR activity toward acyl-S63A Rif M1 substrates was performed by Suzanne Admiraal.

Crystallization. RifR was crystallized by hanging-drop vapor diffusion at 4°C. Crystallization drops were set by addition of protein stock (5-13.5 mg/mL RifR, 10 mM HEPES pH 7.0, 2 mM DTT) to reservoir solution, (8%-23% PEG 8000, 100 mM HEPES pH 7.0-7.6, 35-50 mM $CaCl_2$, 2 mM DTT) in a ratio of 1:2 to 3:2. Crystallization of SeMet RifR required microseeding from native RifR crystals. Before flash freezing in liquid nitrogen, crystals were cryoprotected by soaking 5-10 seconds in a solution equivalent to the reservoir solution with addition of 10% PEG 400. SeMet crystallization was performed by Jamie Razelun.

Crystallization with acyl-CoA substrates. Attempts were made to bind acetyl-CoA, propionyl-CoA, isobutyryl-CoA, and decanoyl-CoA in the active site of S94A RifR by adding acyl-CoAs to crystallization droplets, by transferring crystals to a new drop containing an equivalent reservoir solution with acyl-CoA and by co-crystallization. These attempts were unsuccessful as no substrate was observed in the active site during data processing. For all crystals, protein solution (5 - 13.5 mg/mL) in 10 mM HEPES pH 7.0, and 2 mM DTT was added to a reservoir solution, 8%-23% PEG 8000, 100 mM HEPES pH 7.0-7.6, 35-50 mM $CaCl_2$, and 2 mM DTT, in a 1:2 to 3:2 ratio utilizing the vapor diffusion method in a hanging-drop tray at 4°C.

For crystal soaking, crystals were transferred to a drop of reservoir solution containing either 7.6 mM acetyl-CoA, 2.5 mM propionyl-CoA or 7.0 mM malonyl-CoA. Crystals were soaked for 8-24 hrs and were transferred to a cryoprotectant solution (15% PEG 400, 85% reservoir solution with acyl-CoA) for 20 seconds. Crystals were frozen in liquid nitrogen.

For addition of acyl-CoA to crystallization droplet, a solution containing 92-95% reservoir solution and 5 mM acetyl-CoA, propionyl-CoA, or decanoyl-CoA was added to the crystallization droplet. After 60-80 min, crystals were transferred to a cryoprotectant

solution (70% PEG 400, 30% reservoir solution) for 2 seconds, and were frozen in liquid nitrogen.

For co-crystallization, crystals were grown in the presence of substrate by adding protein solution to the reservoir solution (20% PEG 8000, 40 mM $CaCl_2$, 100 mM HEPES pH 7.6, and 2 mM DTT) in a 1.5:1 ratio. 25 mM propionyl-CoA was added to the crystallization droplet in a 1:10 ratio. The crystal was transferred to a cryoprotectant solution (15% PEG 400, 85% reservoir solution with appropriate acyl-CoA) for 20 seconds and then frozen in liquid nitrogen.

Crystallography. X-ray diffraction data were collected at the GM/CA beamline (23ID-D) at the Advanced Photon Source (APS, Argonne National Laboratory). A three-wavelength MAD data set was recorded from a SeMet RifR crystal for structure determination. Data were processed using the HKL2000 package (46) (Table 2.6). Determination of Se atomic positions, experimental phasing, density-modification phase refinement and initial model building were performed using the programs SOLVE and RESOLVE (47,48). Twelve of fourteen expected selenium sites were identified. Model building was carried out with Coot (49), and the model was refined using REFMAC5 in the CCP4 suite (50), collaborative computational project). Rigid-body motion was modeled as six translation/libration/screw (TLS) groups per monomer, assigned with the aid of the TLSMD server (51). The structure was solved from monoclinic crystals with two RifR polypeptides in the asymmetric unit ($P2_1$: a = 39.5 Å, b = 94.6 Å, c = 63.2 Å, b = 90.55°). Non-crystallographic symmetry restraints were employed in refinement. Subsequent crystal forms, which were orthorhombic with a single molecule in the asymmetric unit, varied in the dimension of the long unit cell axis (82 Å – 108 Å) and were solved with molecular replacements using AMORE (52). Of the subsequent crystal forms, only one contained a fully ordered protein chain (see below) and is reported here in addition to the original crystal form. Gel filtration analysis indicates that RifR is a monomer in solution (data not shown). The final model contains residues 2-247 in both chains. The structures were validated using MolProbity (53) and secondary structure assignment used the Stride server (54,55). The crystal structures are deposited in the

Protein Data Bank (PDB 3FLA for form 1 and 3FLB for form 2). Crystallography was performed by David Akey.

Table 2.6 Crystallographic data

	Form 1 - SeMet			**Form 2**
Diffraction data				
	Peak	Inflection	Remote	Native
Spacegroup	$P2_1$			$P2_12_12_1$
Cell dimensions				
a, b, c (Å)	39.51, 94.65, 63.17			38.94, 62.50, 82.47
α, β, γ (°)	90, 90.55, 90			90, 90, 90
Wavelength (Å)	0.97942	0.97959	0.95446	
Resolution (Å)	50.00—1.80 (1.86—1.80)	50.00—1.90 (1.97—1.90)	50.00—1.86 (1.94—1.86)	50.00—1.80 (1.86—1.80)
Avg I / σ_I	10.5 (1.5)	19.2 (2.2)	13.5 (2.6)	12.5 (2.2)
R_{symm}	0.115 (0.540)	0.112 (0.488)	0.131 (0.665)	0.069 (0.376)
Completeness (%)	99.8 (87.5)	99.0 (97.6)	98.4 (97.1)	96.1 (88.3)
Avg redundancy	3.6 (2.5)	3.6 (3.0)	3.5 (2.9)	2.9 (2.5)
Refinement				
Resolution (Å)	50.0—1.80			50.0—1.80
No. reflections	39950			17702
R_{work} / R_{free}	0.171 / 0.200			0.195 / 0.237
No. atoms				
Protein	3888			1926
Ligand/ion	2			14
Water	482			168
B-factors				
Protein	17.2			22.9
Ligand/ion	22.7			48.2
Water	29.9			33.6
R.m.s deviations				
Bond lengths (Å)	0.009			0.008
Bond angles (°)	1.204			1.181

Results

The ability of RifR to hydrolyze a variety of substrates from a Ppant delivered by both CoA (Fig. 2.4) and ACP carriers was tested. Of particular interest was the ability to remove carboxylated acyl units vs. decarboxylated acyl units, short chain acyl units vs. medium chain acyl units, and acyl units attached to a carrier domain (ACP) vs. those attached to CoA (Table 2.7). RifR hydrolyzed all substrates tested with catalytic efficiencies over a range of 1 $M^{-1}s^{-1}$ to 200 $M^{-1}s^{-1}$. Background hydrolysis was undetectable. With the exception of isobutyryl-CoA, saturation kinetics were not observed.

Hydrolysis of Carboxylated and Decarboxylated Acyl-CoAs. The catalytic efficiency of RifR was compared directly for two natural Rif building blocks (malonyl and methylmalonyl thioesters) and their corresponding decarboxylated variants (acetyl and propionyl thioesters). RifR hydrolyzed the decarboxylated substrates, acetyl-CoA and propionyl-CoA, 7-fold to 14-fold more efficiently, respectively, than the corresponding carboxylated substrates, malonyl-CoA and methylmalonyl-CoA (Table 2.7). In fact, the carboxylated substrates were the poorest of all substrates tested with catalytic efficiencies of 1 $M^{-1}s^{-1}$. While the increased activity against decarboxylated over carboxylated substrates is suggestive of the high-specificity editing model, the discrimination is modest, and the relatively slow rate of reaction is consistent with the low-specificity model, in which hydrolysis of natural carboxylated building blocks occurs inefficiently and does not compete with chain elongation.

Hydrolysis of Medium Chain and Short Chain Acyl-CoAs. The ability of RifR to hydrolyze acyl groups that resemble neither the natural Rif building blocks nor their decarboxylated variants was tested. Unlike previously tested TEIIs from PKS and NRPS pathways, which had little or no activity towards acyl units of medium length (C4 – C10) (5,24,29), RifR hydrolyzed several medium-chain acyl units (Table 2.7). Catalytic efficiency was uncorrelated with chain length: C10 > C8 > C3 > C4 ≈ C2 > C6. It was not possible to determine kinetic constants for these reactions, so it is unknown whether the difference in efficiency is due to differences in k_{cat} and/or K_M values.

Figure 2.4 Michaelis-Menten plots for deacylation of acyl-CoAs.
Plots of WT RifR with **A.** decanoyl-CoA, **B.** octanoyl-CoA, **C.** propionyl-CoA, **D.** butyryl-CoA, **E.** hexanoyl-CoA, and **F.** isobutyryl-CoA, were fit with the Michaelis-Menten equation. Plots of **G.** acetyl-CoA, **H.** methylmalonyl-CoA, and **I**. malonyl-CoA were fit with a linear fit. The k_{cat}/K_M values, determined by these graphs, were later divided by 60 to convert from $M^{-1}min^{-1}$ to $M^{-1}s^{-1}$, as listed in Table 2.7.

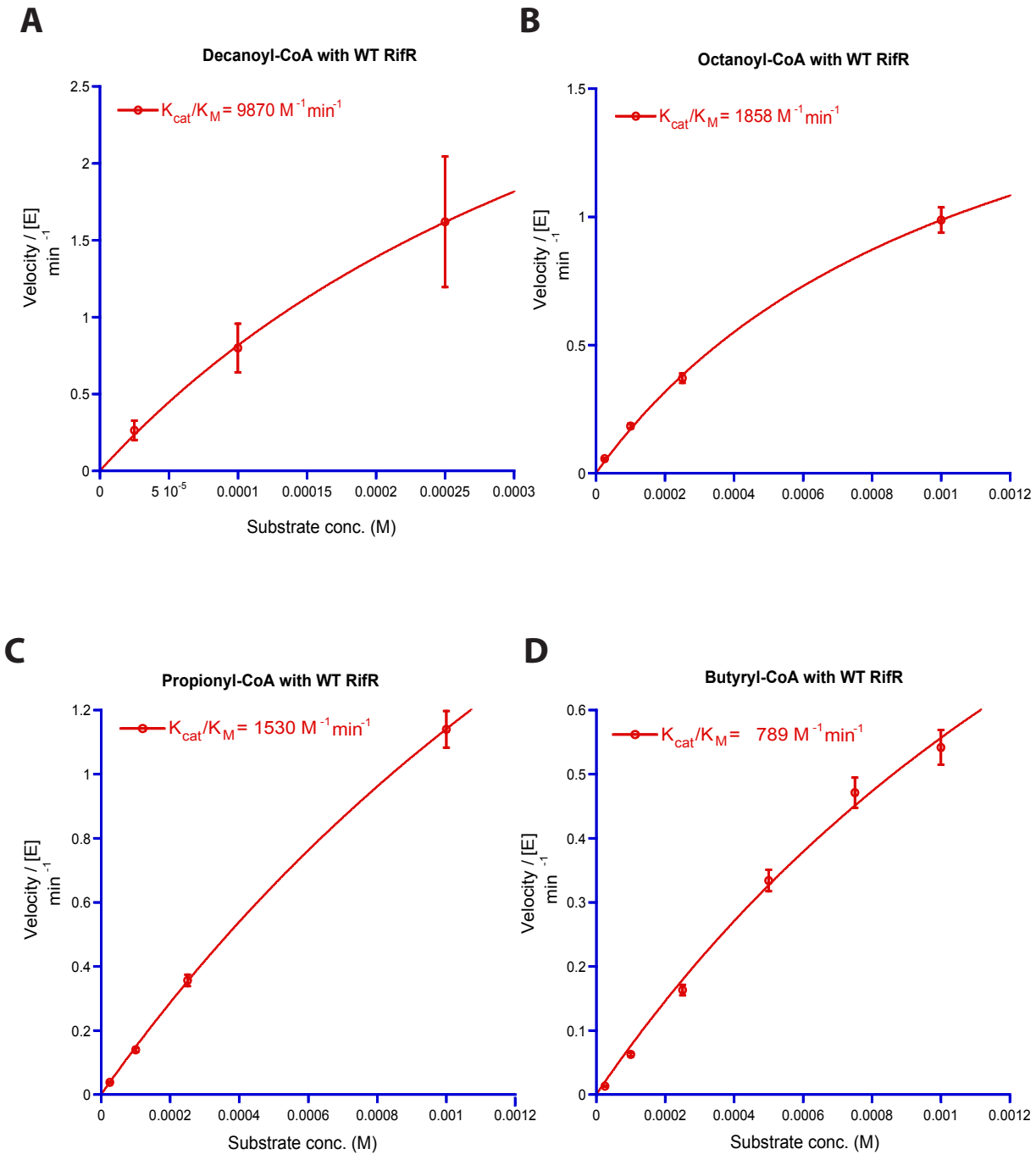
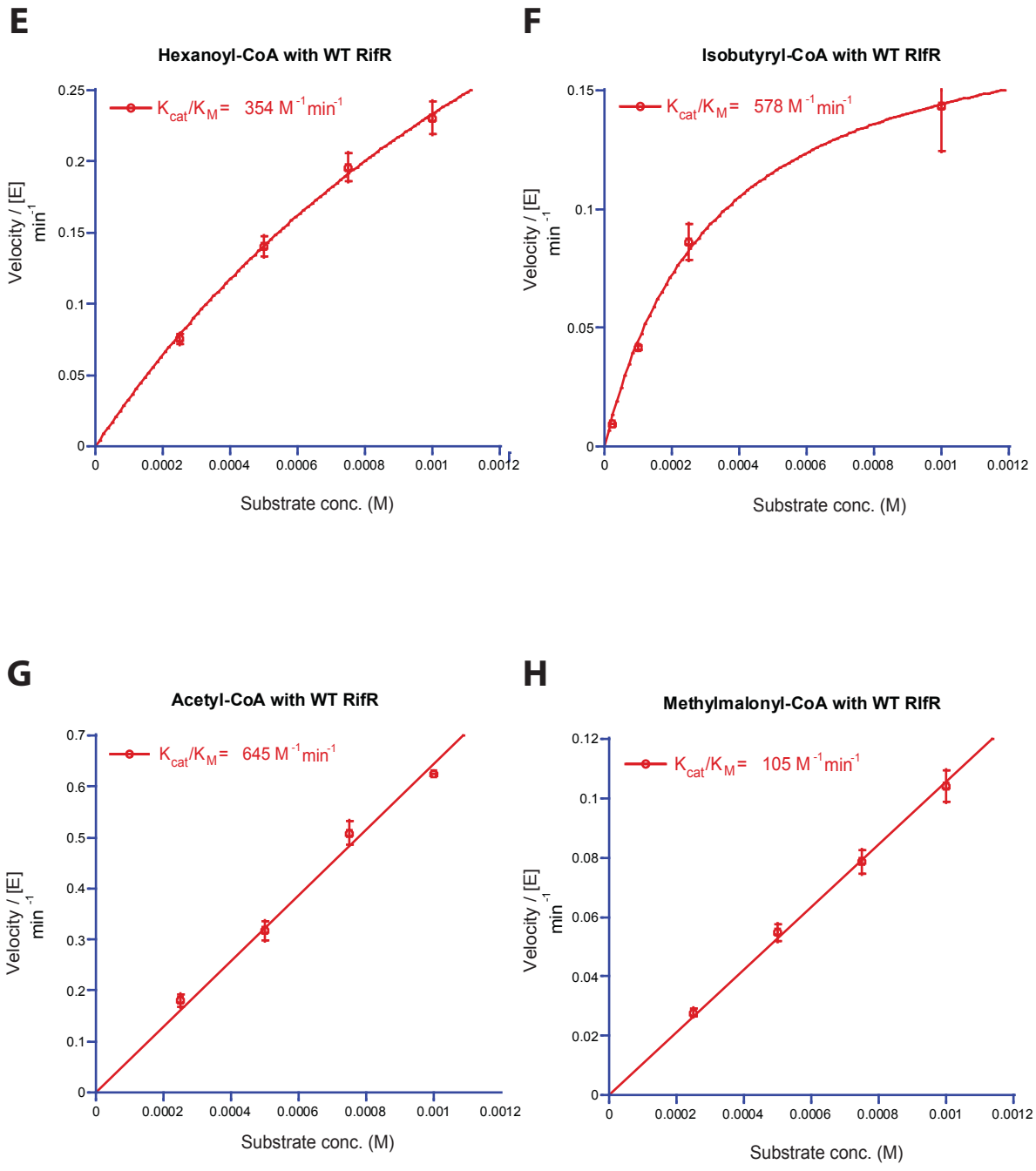
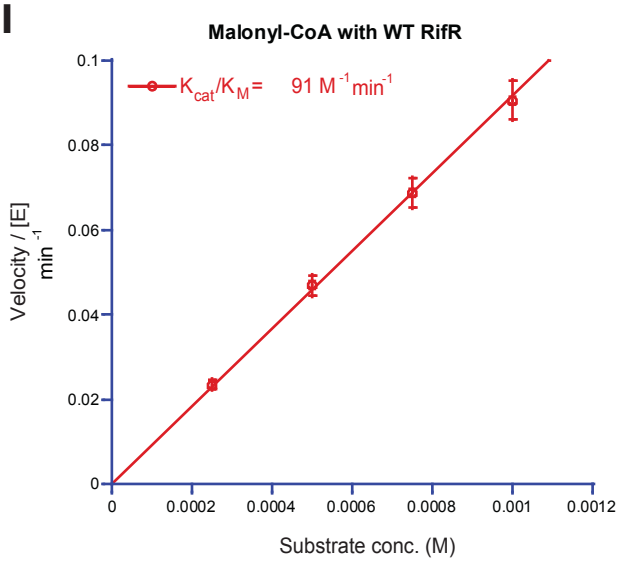


Figure 2.4 Michaelis-Menten plots for deacylation of acyl-CoAs.


E
Hexanoyl-CoA with WT RifR
K_{cat}/K_M = 354 $M^{-1}min^{-1}$
Velocity / [E] min^{-1}
Substrate conc. (M)
F
Isobutyryl-CoA with WT RIfR
K_{cat}/K_M = 578 $M^{-1}min^{-1}$
Velocity / [E] min^{-1}
Substrate conc. (M)
G
Acetyl-CoA with WT RifR
K_{cat}/K_M = 645 $M^{-1}min^{-1}$
Velocity / [E] min^{-1}
Substrate conc. (M)
H
Methylmalonyl-CoA with WT RIfR
K_{cat}/K_M = 105 $M^{-1}min^{-1}$
Velocity / [E] min^{-1}
Substrate conc. (M)

I


Malonyl-CoA with WT RifR
K_{cat}/K_M= 91 $M^{-1}min^{-1}$
Velocity / [E] min^{-1}
0.1
0.08
0.06
0.04
0.02
0
0
0.0002
0.0004
0.0006
0.0008
0.001
0.0012
Substrate conc. (M)

Hydrolysis of Acyl-ACPs. The catalytic efficiency of RifR was compared directly for acyl-ACP and acyl-CoA substrates using the natural Rif module 1 building block (methylmalonyl thioester), its decarboxylated variant (propionyl thioester), and a potential mis-acylated substrate (acetyl thioester). For these experiments, the ACP from Rif module 1 was used in the context of the full module (RifM1), a 115-kDa multifunctional protein. PKS acyltransferase domains can possess deacylation activity toward cognate acyl-ACP domains. To avoid the possibility of deacylation by the acyltransferase in RifM1, an acyltransferase-inactive variant (S639A) was used. RifR hydrolyzed all three ACP substrates (Table 2.7). It was not possible to obtain saturating concentrations of the acyl-ACP substrates. As for acyl-CoA substrates, RifR showed a slight (4-fold) preference for the decarboxylated substrate (propionyl-ACP) over the carboxylated (methylmalonyl-ACP) substrate. In contrast to the slight discrimination among acyl substrates, under matched reaction conditions, RifR displayed a stronger preference for acyl-ACP substrates over the corresponding acyl-CoA substrates: 8-fold for the propionyl unit, 14-fold for the acetyl unit, and 30-fold for the methylmalonyl unit (Table 2.7).

Hydrolysis of Acyl-CoAs by S94A RifR. Catalytic activity of wild-type RifR was compared to an active-site RifR mutant, in which the catalytic serine was substituted by alanine (S94A) (Table 2.7). Thioesterase activity of S94A RifR was effectively eliminated for all substrates excepting isobutyryl-CoA and propionyl-CoA. These substrates may be capable of binding in the active site such that hydroxide ion derived from water acts as the nucleophile in place of the active site serine hydroxylate, allowing hydrolysis of the acyl unit, albeit at a decreased rate.

Table 2.7 Kinetic parameters for RifR hydrolysis of acyl substrates

Substrate		k_{cat}/K_M ($M^{-1}s^{-1}$)		Ratio	
		WT RifR	S94A RifR	WT/ S94A	ACP/ CoA
CoA Substrates					
Decanoyl-CoA	CH_3-$(CH_2)_8$-CO-S-CoA	160 ± 18	<0.16	>1000	
Octanoyl-CoA	CH_3-$(CH_2)_6$-CO-S-CoA	31 ± 2.5	<0.16	> 190	
Propionyl-CoA	CH_3-CH_2-CO-S-CoA	25 ± 0.5	0.96 ± 0.37	26	
Butyryl-CoA	CH_3-$(CH_2)_2$-CO-S-CoA	13 ± 3.2	<0.04	> 340	
Acetyl-CoA	CH_3-CO-S-CoA	11 ± 0.2	<0.03	> 320	
Isobutyryl-CoA	$(CH_3)_2$-CH-CO-S-CoA	9.6 ± 0.08	4.5 ± 0.08	2.1	
Hexanoyl-CoA	CH_3-$(CH_2)_4$-CO-S-CoA	5.9 ± 0.33	< 0.04	> 140	
Methylmalonyl-CoA	CO_2-$(CH_3)CH_2$-CO-S-CoA	1.8 ± 0.17	< 0.03	> 72	
Malonyl-CoA	CO_2-CH_2-CO-S-CoA	1.5 ± 0.08	< 0.07	> 21	
ACP Substrates					
Propionyl-RifM1	CH_3-CH_2-CO-S-ACP	210 ± 20			8.4
Acetyl-RifM1	CH_3-CO-S-ACP	150 ± 38			14
Methylmalonyl-RifM1	CO_2-$(CH_3)CH_2$-CO-S-ACP	54 ± 6.3			30

Overall Structure of RifR Type II Thioesterase. RifR is a monomeric protein (Fig. 2.5A) and a member of the α/β-hydrolase family, with a fold similar to the folds of an NRPS TEII (SrfTEII) (11), three NRPS TEIs (SrfTEI) (8), FenTE (9) and EntTE (10), two PKS TEIs (DEBS TEI (7) and Pik TEI (6)), and the human fatty acid synthase thioesterase (hFAS TE) (2). The α/β-hydrolase core fold is predominantly a parallel β-sheet surrounded by α-helices. The hydrolytic active site is a triad of amino acids located on loops at the C-terminal edge of the core β-sheet. Members of the diverse family differ in the location of some triad residues and in the number and location of helices that decorate the core fold. RifR contains a small sub-domain (residues 130-180), which forms a three α-helix 'lid' inserted between strands β5 and β6 of the α/β-hydrolase fold (Fig. 2.5B). The first two helices of the lid (αL1 and αL2) form a short hairpin structure comprising the top of the lid, and the third helix (αL3) forms the back of the lid.

Catalytic Triad. The active site of RifR is a classic catalytic triad comprising residues Ser94, Asp200 and His228 (Fig. 2.5C). The triad residues of RifR, like those of other TEIIs (Fig. 2.1), follow strands β4 (Ser), β6 (Asp) and β7 (His) of the α/β-hydrolase fold (Fig. 2.5). The active site serine, found within the signature sequence Gly92-His93-Ser94-Xaa-Gly96, is between strand β4 and helix α3 and has the constrained geometry typical of a nucleophilic elbow, a hallmark of the α/β-hydrolase family. A number of hydrogen bonds position residues in the catalytic center. Notably, His93 of the signature sequence forms a hydrogen bond with the backbone carbonyl of the active site His228, stabilizing its alignment within the triad (Fig. 2.5C). The oxyanion hole, which stabilizes the tetrahedral intermediate, is formed by the backbone amides of Met95 and Ala29 and contains a single chloride ion in the crystal structures (Fig. 2.5C). The aspartate of the catalytic triad follows strand β6, in contrast to the PKS, NRPS and FAS TEIs, where it follows strand β5 (Fig 2.1).

Figure 2.5 Structure of RifR.
A. Stereodiagram of RifR TEII. In this ribbon diagram, the polypeptide is colored as a rainbow from blue at the N-terminus to red at the C-terminus. The lid domain is colored yellow. Active site triad residues (Ser94, Asp200 and His228) are shown as sticks. Two conformations (from different crystal forms) are shown for the flexible linker region between strand β5 and the lid domain. **B**. Topology of RifR. **C**. Stereodiagram of the active site. The catalytic triad (Ser94, Asp200 and His228) and surrounding residues are shown as sticks. A chloride ion (green) occupies the oxyanion hole formed by backbone amides of residues Ala29 and Met95 (side chain not shown). Atomic colors are used in **A** and **C** for stick figures with yellow C, red O, blue N and green Cl. Figure was created by David Akey.

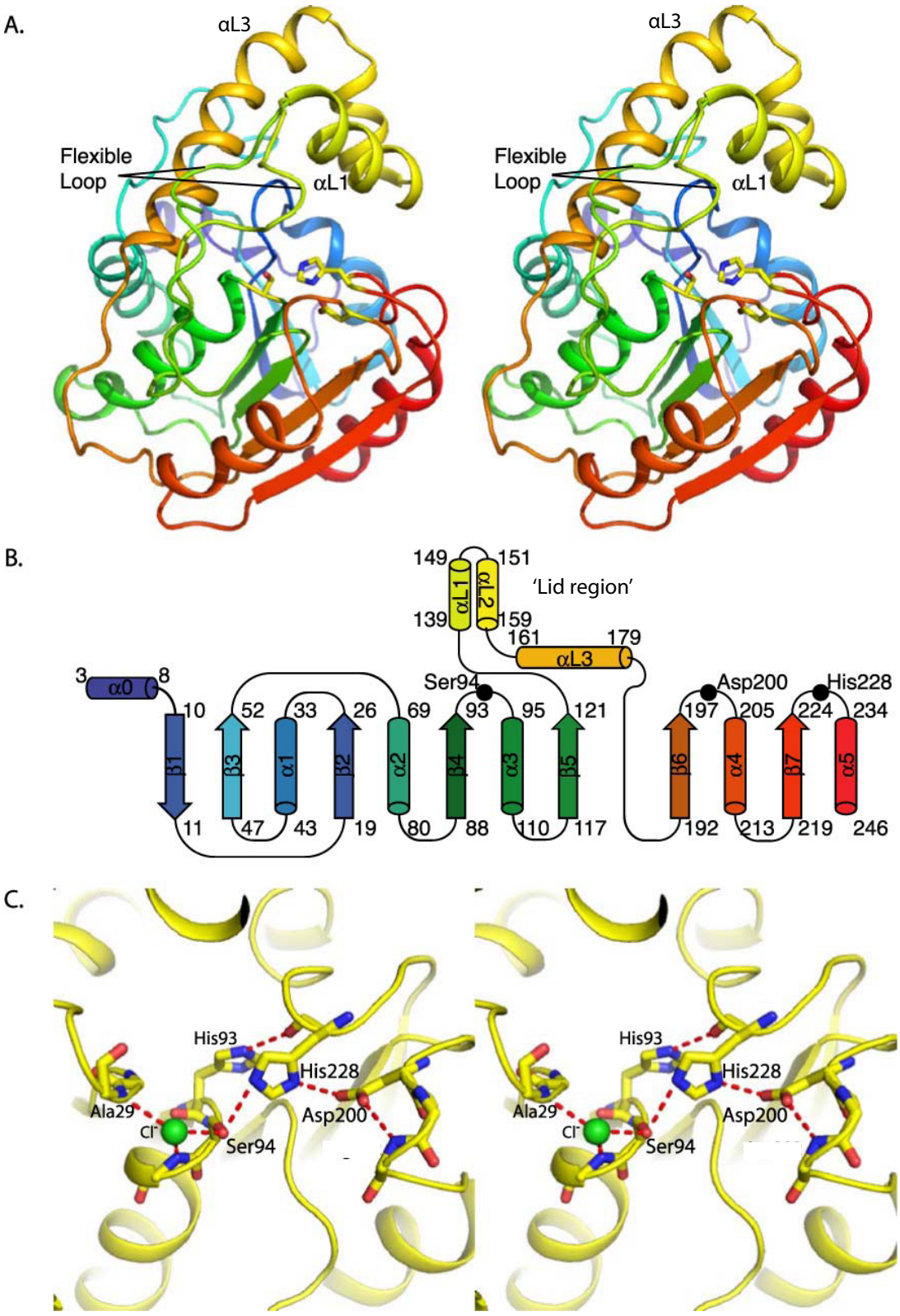


Figure 2.5 Structure of RifR.

Flexible Lid Sub-domain. The lid sub-domain of RifR covers the active site like similar lids in other PKS, NRPS and FAS TEs. In each TE, the lid is centered over the catalytic triad and defines a 'Ppant entrance' on one side of the triad and a 'substrate chamber' on the other side (Fig. 2.6). Several lines of evidence are consistent with these functional assignments. The substrate of each TE is delivered to the active site on the Ppant arm of a carrier domain. The Ppant entrance is inferred from the position of the TE N-terminus, where the carrier domains are fused to PKS TEIs (7), and from a solution structure of EntTE in complex with its cognate ACP domain (10). The substrate chamber is inferred from the structures of substrate-analog affinity-labeled PKS TEIs (32) and of an inhibitor complex of hFAS TE (2). The size, shape and character of the substrate chamber determine which substrates can be accommodated and whether the TE hydrolyzes the thioester to a linear product or, like many PKS TEIs, forms a macrolactone.

All the TE lids are helical, however they differ in the number and disposition of helices and in their flexibility. The RifR lid is similar to the lids of the monomeric NRPS TEIs and TEII, which are continuous in sequence and flexible. In contrast to the RifR lid, the lids of the dimeric PKS TEIs lack flexibility and contain four non-consecutive α-helices, two of which are an N-terminal extension of the sequence and form the dimer interface (Fig. 2.1). Variation among RifR crystal structures provides evidence for lid flexibility. RifR crystallized in a range of related forms with similar crystal packing along two shorter unit cell axes of ~39 Å and ~64 Å. The longer unit-cell axis displayed remarkable variation from 82 Å to 109 Å. The lid sub-domain participated in a crystal lattice contact along the direction of this long unit-cell edge.

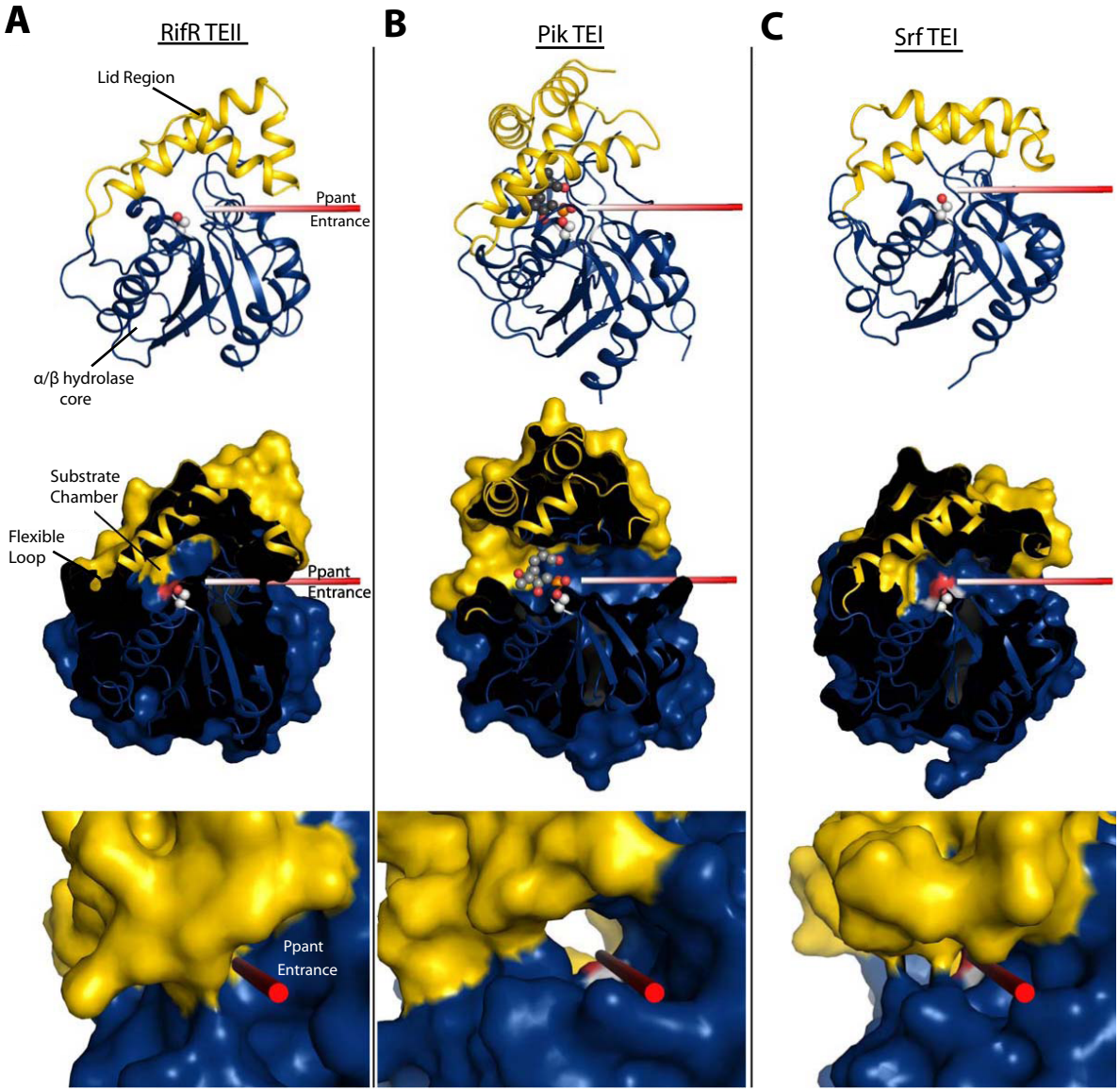


Figure 2.6 Cartoon and surface diagrams of TEIs and TEII.
Cartoon and surface diagrams in equivalent orientations for **A**, RifR TEII, **B**, Pik TEI (PDB 2HFJ) with affinity label and **C**, Srf TEI (PDB 1JMK). The rods show the Ppant entrance path to the catalytic serine (shown in spheres). The middle panels show a cut-away surface diagram in the same orientation as the top panels. Access to the active site for RifR TEII is blocked by the lid helices. The bottom panels are close up views along the Ppant entrances, showing the closed entrance in RifR, the tunnel-like entrance characteristic of dimeric PKS TEIs (Pik TEI) and trough-like entrance of the monomeric NRPS TEI (Srf TEI in the open form). Figure was created by David Akey.

The various crystal forms captured different conformations of the lid. The structures fall into three distinct classes, which differ in the conformation of a flexible 'lid loop' (residues 122 – 138) that is an integral part of the substrate chamber and links strand β5 of the α/β-hydrolase core to the first helix of the lid domain (αL1). In 'form 1' crystals (long axis 94 – 99Å), the lid loop is positioned towards the active site; in 'form 2' structures (long axis ~82 Å), the lid loop lies along αL3; and in other crystal forms (long axis 88 – 92 Å and 108 – 109 Å), the lid loop is disordered. The atomic mobility (B) factors are higher and the electron density for the lid loop is poorer in form 2 than in form 1 (Fig. 2.7A, 2.7B). Additionally, αL1 is shifted towards the α/β-hydrolase core and rotated inward in form 2 with respect to form 1 (Fig. 2.7C). Movement of the lid helices and the flexible lid loop has dramatic effects on the size and shape of the substrate chamber (Fig. 2.7D, 2.7E). This flexibility in the substrate chamber is consistent with the modest substrate preferences and wide substrate range exhibited by RifR (Table 2.7).

The inside of the RifR Ppant entrance port contains residues that are conserved across TEIIs, but, consistent with the kinetic results, neither the entrance port, nor the substrate chamber contain any obvious structural features that would confer exclusive preference for decarboxylated substrates over carboxylated ones, (Table 2.7). The lid movements also affect access to the catalytic triad from the presumed Ppant entrance. The substrate entrance of RifR is bounded by helix αL1 of the lid and helix α1 of the α/β hydrolase core. In all crystal forms, the Ppant entrance is blocked by contact of these α-helices (Fig. 2.6A). In part for this reason we think that movement of helix αL1 is required to open the binding site for the Ppant arm.

Figure 2.7 Two crystal forms of RifR.
Electron density and model for flexible linker region for form 1 (**A,** – experimental density contoured at 1σ) and for form 2 (**B,** – 2Fo-Fc refined density contoured at 1σ). **C**. Stereodiagram of αL1 helices from form 1 (blue) and form 2 (gold). Form 1 **D**, and form 2 **E,** substrate chambers have unique shapes, and different access to the exterior of RifR. Figure was created by David Akey.

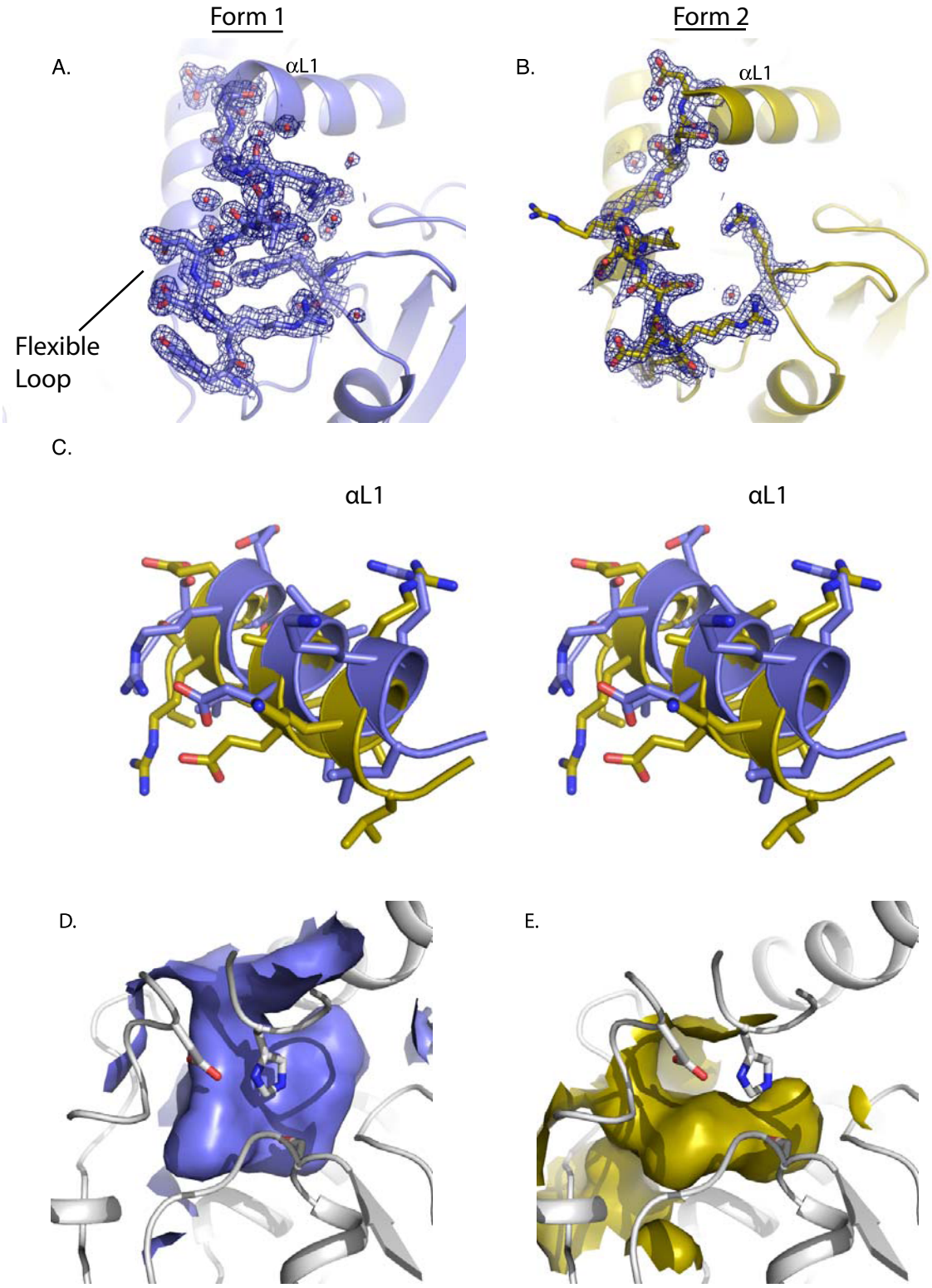


Figure 2.7 Two crystal forms of RifR.

Discussion

RifR displayed broad substrate specificity, hydrolyzing carboxylated and decarboxylated acyl thioesters, as well as short-, medium- and branched-chain substrates. Despite the broad substrate range, RifR preferentially hydrolyzed aberrant decarboxylated acyl thioesters over natural Rif building blocks. However, the preference for decarboxylated over carboxylated substrates (4-fold to 14-fold) was modest (Table 2.7). Methylmalonate is the building block for most modules in the Rif pathway, so the decarboxylated variant of methylmalonyl-ACP (propionyl-ACP) should be a primary target of any editing enzyme. RifR had a modest preference for propionyl-ACP over methylmalonyl-ACP (4-fold, Table 2.7) consistent with the low-specificity model for TEII editing in which both aberrant and natural acyl thioesters are hydrolyzed from carrier domains more slowly than the assembly line pathway processes the natural building blocks. The rate of Rif pathway throughput is unknown, as is the catalytic efficiency of individual ketosynthase condensing domains, so it is not possible to compare throughput and editing rates. Nevertheless RifR is a rather slow enzyme with efficiencies between 1 $M^{-1}s^{-1}$ and 200 $M^{-1}s^{-1}$.

The structural variability of the RifR substrate chamber matches the observed broad specificity of the enzyme. The chamber is malleable due to the flexibility of the lid loop (residues 122-138) and loop helix αL1. The plasticity of the substrate chamber likely allows it to accommodate a variety of acyl groups, accounting in part for the broad substrate specificity. The crystal structures captured two variations of the substrate chamber, as well as a highly open chamber in which the lid loop is disordered. These variants likely represent a small subset of substrate chamber shapes that are accessible to the protein in solution. In addition to plasticity, the interior surface of the substrate chamber appears able to accommodate a variety of substrates. The surface of the substrate chamber is hydrophilic in both crystal forms, and appears unable to distinguish between charged and uncharged substrates, or short, medium and branched acyl thioesters. The chamber is accessible to bulk solvent in all crystal forms, also consistent with the broad substrate specificity. A closed Ppant entrance was observed in all crystal forms of RifR, but differences among these crystal structures provided evidence of lid motion. A substantially populated closed-lid form of RifR in solution could account for

the observed slow turnover of the enzyme. Lid flexibility is a hallmark of monomeric PKS and NRPS TEs. The surfactin (Srf) TEI crystallized with two independent molecules, one with an open Ppant entrance, the other closed (8). Solution (NMR) structures of Ent TEI (10) and Srf TEII(11) also suggest movement in the lid region. In fact, the flexible lid of Srf TEII was reported in an extremely open conformation with no contacts to the α/β hydrolase core. In contrast, no flexibility has been observed for lid α-helices or lid loops in the dimeric PKS TEIs. The extra N-terminal helices, which comprise the dimerization domain in the Pik TEI and DEBS TEI lids, likely stabilize the lid loop region.

The 8-fold to 30-fold preference of RifR for substrates carried by ACP over those carried by CoA is consistent with an editing function for RifR. If RifR is a scavenger of aberrant acyl units that stall the Rif pathway, then it should have poor or no activity with CoA substrates. The observed carrier preference could be due to either favorable interactions of RifR with Rif ACP or unfavorable interactions with CoA. The Ppant arm, common to ACP and CoA carriers, is long enough to reach the catalytic triad from the enzyme surface at the Ppant entry. The RifR surface surrounding the Ppant entrance is neither strongly hydrophobic nor strongly electronegative, and thus lacks features that could lead to unfavorable electrostatic or van der Waals interactions with CoA. It seems more likely that favorable protein-protein interactions with Rif ACP account for the carrier preference.

Our working model for RifR editing invokes the dynamic property of the lid. The lid must be open for acylated Ppant to reach the active site, and any RifR molecules in a closed-lid form are temporarily unavailable for catalysis. Evidence of lid motion comes from differences in RifR crystal forms (Fig 2.6) and from larger-scale motions observed or implied in structures of SrfTEI and SrfTEII. We propose that helix αL1 moves to allow proper substrate binding. In this manner, lid dynamics could be a strategy to prevent wasteful hydrolysis of CoA substrates. If Rif ACPs interact preferentially with an open-lid form of RifR and if CoA and non-Rif ACPs have no such preference, then Rif ACPs would be the preferred RifR substrate carriers. Thus, the “correct” carrier increases editing efficiency by facilitating lid opening. Most characterized editing TEs have weak or no acyl-group specificity. It may be a general feature of editing thioesterases that