NATURAL PRODUCT BIOSYNTHESIS: FRIEND OR FOE?

FROM ANTI-TUMOR AGENT TO DISEASE CAUSATION

A Dissertation

by

JENNIFER DIANNE FOULKE-ABEL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2010

Major Subject: Chemistry

NATURAL PRODUCT BIOSYNTHESIS: FRIEND OR FOE?

FROM ANTI-TUMOR AGENT TO DISEASE CAUSATION

A Dissertation

by

JENNIFER DIANNE FOULKE-ABEL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Coran M. H. Watanabe
Committee Members,	Tadhg P. Begley
	Frank M. Raushel
	Gregory D. Reinhart
Head of Department,	David H. Russell

December 2010

Major Subject: Chemistry

ABSTRACT

Natural Product Biosynthesis: Friend or Foe?
From Anti-tumor Agent to Disease Causation. (December 2010)
Jennifer Dianne Foulke-Abel, B.A., Texas A&M University
Chair of Advisory Committee: Dr. Coran M. H. Watanabe

Biosynthetic natural products are invaluable resources that have been gleaned from the environment for generations, and they play an essential role in drug development. Natural product biosynthesis also possesses the latent ability to affect biological systems adversely. This work implements recent advances in genomic, proteomic and microbiological technologies to understand further biosynthetic molecules that may influence progression of human disease.

Azinomycin A and B are antitumor metabolites isolated from the terrestrial bacterium *Streptomyces sahachiroi*. The azinomycins possess an unusual aziridine [1,2-*a*] pyrrolidine ring that reacts in concert with an epoxide moiety to produce DNA interstrand cross-links. Genomic sequencing of *S. sahachiroi* revealed a putative azinomycin resistance protein (AziR). Overexpression of AziR in heterologous hosts demonstrated the protein increases cell viability and decreases DNA damage response in the presence of azinomycin. Fluorescence titration indicated AziR functions as an azinomycin binding protein. An understanding of azinomycin resistance is important for future engineering and drug delivery strategies. Additionally, the *S. sahachiroi* draft genome obtained via 454 pyrosequencing and Illumina sequencing revealed several silent secondary metabolic pathways that may provide new natural products with biomedical application.

β-lactoglobulin (BLG), the most abundant whey protein in bovine milk, has been observed to promote the self-condensation of retinal and similar α,β-unsaturated aldehydes. BLG is a possible non-genetic instigator of cycloretinal and A2E accumulation in the macula, a condition associated with age-related macular degeneration. BLG-mediated terpenal condensation has been optimized for *in vitro* study with the retinal mimic citral. In rabbits fed retinal and BLG or skim milk, cycloretinal formation was detected in the blood by ^{1}H-NMR, and SDS-PAGE analysis indicated BLG was present in blood serum, suggesting the protein survives

ingestion and retains catalytic activity. Mass spectrometry and site-directed mutagenesis provided mechanistic insight toward this unusual moonlighting behavior.

The experiments described in this dissertation serve to further natural product biosynthesis discovery and elucidation as they relate to consequences for human health. Efforts to solve azinomycin biosynthesis via enzymatic reconstitution, characterize compounds produced by orphan gene clusters within *S. sahachiroi*, and obtain a clear mechanism for BLG-promoted cycloterpenal formation are immediate goals within the respective projects.

ACKNOWLEDGEMENTS

Gratitude is in order for my labmates and predecessors, especially the now Dr. G. Thomson Kelly, Dr. Scott E. Angell, and Dr. B. J. Bench, who not only passed on their creative methods for recycling items in the lab, but also served as my guides for biochemistry, genetics, and organic chemistry. They made me feel like I was part of the lab family. I borrowed more than a few of their good ideas and have missed them since they left for the real world. I also acknowledge Dr. Vasudha Sharma, who set the best example of postdoctoral fellow behavior: bright, responsible, engaging, and prone to infectious giggling. I hope to take at least a little of that with me as I pursue my own postdoctoral training.

Experimental advice from Dr. Benjamin (B. J.) Philmus and Professor Tadhg Begley's group was especially valuable for the resistance protein project. A thank you goes to Dr. Larry J. Dangott at the TAMU Protein Chemistry Laboratory for LC-MS analyses related to the BLG project and advice on protein purification.

To Coran, I owe a tremendous amount of gratitude and admiration. She is tough and smart, even if she still occasionally gets lost driving around Bryan. I know I wasn't the only one who found it hard to live up to her expectations, but demanding the best will bring out the best in graduate students who are willing to focus, strategize, and fight their way through the five (or six or seven) years of the indentured servitude that is grad school. We all like to make it sound bad once we've survived it, but in honesty, would any of us really have wanted it any other way? I believe the long haul was worth it. I am incredibly happy with how things turned out, and wish Coran much continued success.

Finally, to my parents and my husband, Joseph, and his parents, who all saw to it that I survived the mysterious things I do when I'm in the lab, I will always be indebted. If I play with my test tubes and bacteria long enough, eventually it does pay off. (P.S. There are no tiny sheep as a result of these efforts. I'm still working on it.)

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

Page

Page

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION: NATURAL PRODUCT BIOSYNTHESIS

INTRODUCTION

The term natural product encompasses a staggering array of compounds produced by organisms at all taxonomic levels for both basic and secondary life processes. Man has utilized natural products for medicinal purposes since pre-recorded history, relying on mystic understanding of the forces behind the physiological effects. Time, curiosity, and creativity have yielded us the technical tools to probe origin, identity and activity of ancient remedies, as well as afforded access to the previously undiscoverable gems of therapeutic benefit literally lying beneath our feet. Many medicinally-relevant natural products are borne from inessential secondary metabolic genes within organisms; they are biosynthetic pathways for which the evolutionary origin is often unknown. Recent research is becoming increasingly interdisciplinary, weaving principles of chemistry, biochemistry, genomics, proteomics, and medicine into a complex picture of submicroscale processes known as natural product biosynthesis. However, not all natural products are beneficial, and while we relentlessly search for new drug candidates, we must also take time to understand the effect of molecules being produced in our own bodies and how genes, diet and environment may play dramatic roles in catalyzing the very disease states we seek to treat with exogenous solutions. Thus, natural products are collectively both our ally and our enemy depending upon context.

THE VALUE OF NATURAL PRODUCTS

Historically, natural products have been directly used or been the inspiration for over 60% of the small molecule therapeutics available today (**Fig. 1**).[1] Beginning in the 1980s, interest in medicinal natural products began slowly shifting toward technology-driven drug development and design, employing novel techniques in combinatorial chemistry and computational molecular modeling to screen huge synthetic compound libraries for specific bioactivity in a high-throughput fashion.[2] When massive high-throughput efforts failed to reap similarly impressive rewards, the popularity of focused, structurally-related subgroup screenings took over, harkening back to old biologically-relevant structures given new life by synthetic

---

This dissertation follows the style of *Nature Chemical Biology.*

manipulation in a technique termed "diversity-oriented synthesis."[3] Pharmaceutical chemists realized the hundred thousand-membered libraries of the high-throughput era lacked the structural complexity observed in natural products, embodied in characteristics such as multiple chiral centers, heterocycles and polycyclic motifs that ultimately lead to targeted bioactivity.

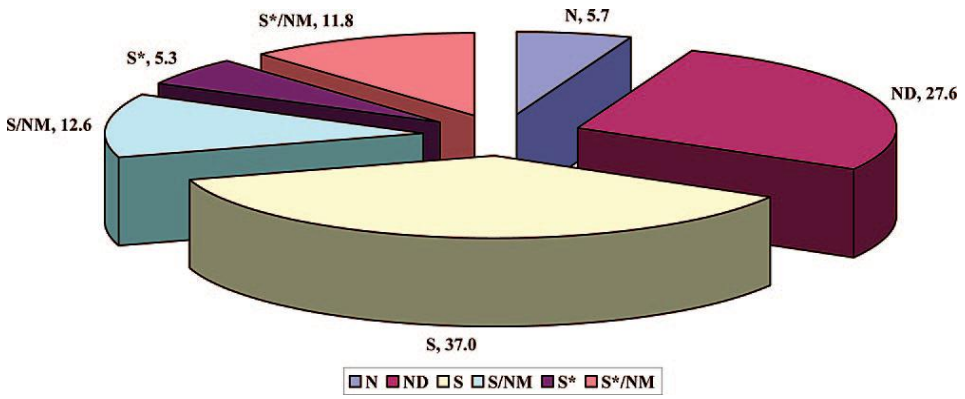


Figure 1 Distribution of small molecule drug discovery sources, 1981-2006.
n = 983, and values are expressed in percentages. N, natural product; ND, derived from a natural product and usually a semisynthetic modification; S, totally synthetic drug often found by random screening/modification of an existing agent; S*, made by total synthesis, but the pharmacophore is/was from a natural product; NM subcategory, natural product mimic.

The potential economic windfall from discovery of blockbuster natural product drugs is storied in examples of the antitumor agent taxol and the "antibiotic of last resort" vancomycin (**Fig. 2**). Isolated from bark of the Pacific yew tree[4], taxol treats a range of aggressive soft tissue cancers by over stabilizing microtubules, thus inhibiting cytoskeleton reconstruction necessary for cell division.[5] Vancomycin is a glycopeptide secondary metabolite produced by the soil-dwelling *Amycolatopsis orientalis*, and is administered primarily against methicillin-resistant *Staphylococcus aureus* and penicillin-resistant *Streptococcus pneumoniae* infections.[6] At the apex of their respective therapeutic careers, both taxol and vancomycin grossed over $1 billion in pharmaceutical sales annually.[7,8]

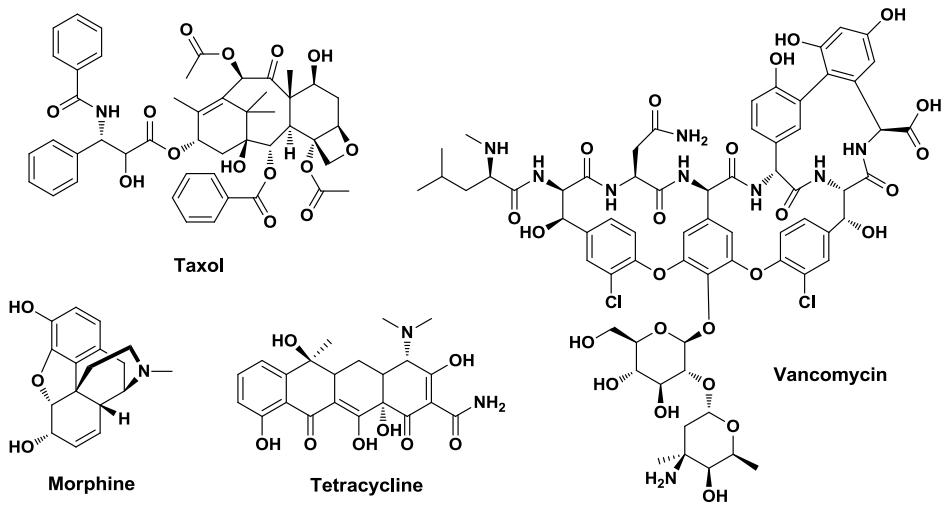


Figure 2 Examples of secondary natural product classes.
Terpene (taxol), nonribosomal peptide (vancomycin), polyketide (tetracycline), and alkaloid (morphine).

METABOLIC PATHWAYS, THE DRIVING FORCE OF LIFE

Life requires an orchestra of chemical reactions undergoing constant tempo changes. Broadly put, the collection of processes an organism uses for energy consumption and renewal comprise metabolism. But upon more detailed molecular consideration, metabolism is the building up and breaking down of every biomolecule, the response to a constantly shifting reactant/product equilibrium, and the source of distinction between inanimate object and animate being. Chemical reactions in organisms proceed primarily via catalytic macromolecules called enzymes; when two or more enzymes moderate chemical revisions of a single compound, they are said to form a biosynthetic pathway.

The essential compounds of life (nucleic acids, amino acids, vitamins, cofactors, carbohydrates, fatty acids, and hormones) are the result of primary metabolic pathways that are relatively conserved throughout all life forms. Molecule classes such as polyketides, nonribosomal peptides, alkaloids, and terpenoids (**Fig. 2**) are the products of inessential or secondary metabolic genes specific to a particular species or group. Only a fraction of the currently known secondary metabolites have a defined function, often as defense mechanisms or communication signals. Additionally, secondary pathways may be silenced or activated for any number of unknown reasons.

SYNTHETIC VERSUS BIOSYNTHETIC ROUTES TO NATURAL PRODUCTS

For every individual touting synthetic methodology as the ultimate means to gain production-scale quantities of medicinally-relevant natural products, there is the emerging concern of sustainability. While no doubt exists as to the utility and contributions of completely synthetic drugs, the exercise demands much in the way of petroleum-derived solvents and starting materials, often involves numerous labor-intensive steps, and generates much in the way of hazardous chemical waste.[9] Biosynthesis, even if only employed for select steps in a largely synthetic approach, minimizes environmental impact on both sides of the production equation. Microbes and their endogenous biosynthetic pathways can be fabulously efficient and precise under certain circumstances, operating under conditions much more harmonious with a "green" initiative. A third approach is to marry biosynthetic building blocks with conventional total synthesis to exploit natural reactivity, eliminating the need for multiple protection/deprotection steps or enzymes.[10] Realistically, all three methods have equal roles in sustainable manufacturing, as each strategy alone cannot accomplish every synthesis one may wish to undertake.

As ideas for drug delivery evolve, biosynthesis takes on a more precious role. *In situ* treatments, wherein the drug is biosynthesized selectively in cells or tissues requiring therapy, is but one application of targeted design demanding an entirely gene-encoded route for natural product production. While such an approach is still in preclinical infancy, aspirations for "healing from within" require knowledge of biosynthetic processes much more vast than our current understanding.

NATURAL PRODUCT AS FRIEND: THE ANTITUMOR AGENT AZINOMYCIN B

A push to discover new drug candidates during the mid-twentieth century resulted in a flood of substances exhibiting promise as anti-infective, antitumor, or disease therapeutics.[11] Among these compounds was carzinophilin A, an isolate from a new strain of soil-dwelling bacteria, *Streptomyces sahachiroi*, found to inhibit Yoshida sarcoma and extend lifespan in a rat model.[12] Over thirty years later the metabolites azinomycin A, azinomycin B, and epoxyamide (**Fig. 3**) were isolated from *Streptomyces griseofuscus* S42227,[13] and through advances in purification and spectroscopic assignments carzinophilin A was determined to be identical to azinomycin B.

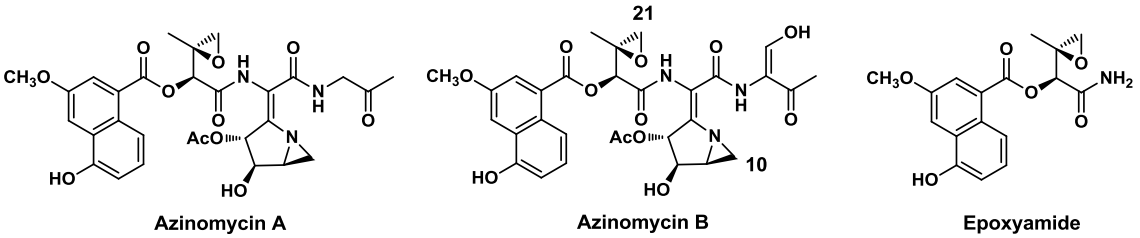


Figure 3 Structures of azinomycin A, azinomycin B, and epoxyamide.

The azinomycins and functionally related synthetic analogs have been tested against myriad disease models, including the 60 cancer cell lines maintained by the National Cancer Institute.[14] One measure of activity described by Ishizeki and coworkers showed azinomycin A and B exhibited similar *in vitro* cytotoxicity (IC_{50} of 0.07 µg/mL and 0.11 µg/mL, respectively) against the L5178Y leukemia cell line. In a P388 murine leukemic model, azinomycin B provided an increase in lifespan (ILS) of 193%, comparable to that of the current clinical drug mitomycin C (204% ILS), but at a 60-fold lower dose.[15] Azinomycin B antitumorigenicity stems from cross-linking of DNA via interaction of the electrophilic aziridine C10 and epoxide C21 with purine residues two base pairs apart on complementary DNA strands (**Fig. 4**).[16] DNA alkylation triggers depurination and strand breakage, an effect expressed in the observed upregulation of DNA synthesis/repair genes.[17] Bisalkylating activity has made the azinomycins a popular springboard for designing synthetic molecules with similar DNA-targeting ability.[14,18-22]

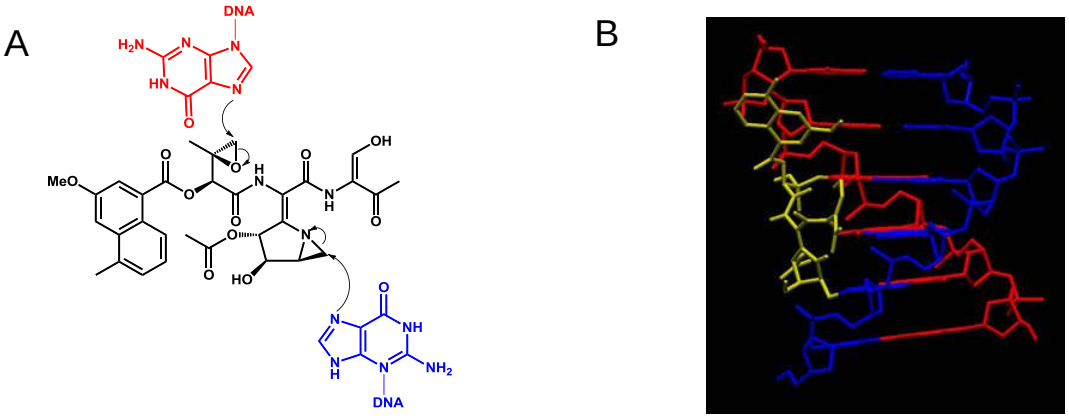


Figure 4 Azinomycin B interactions with duplex DNA.
(a) The N7 of two guanine bases attack the electrophilic epoxide and aziridine rings to form covalent cross-links. (b) Molecular modeling of azinomycin B bound in the major groove of DNA.[16] Azinomycin B is yellow and the DNA helix strands are red and blue.

AZINOMYCIN BIOSYNTHESIS

The azinomycins are interesting from a biosynthetic viewpoint due to the unprecedented azabicyclo[3.1.0]hexane ring. Of the over 100,000 natural products isolated to date, only a handful of molecules are known to contain aziridine rings (**Fig. 5**). Biosynthetic gene clusters have been proposed for mitomycin C[23], azicemicin A[24], azinomycin B[25], and maduropeptin[26], but no one has yet successfully characterized the biosynthetic reactions that transpire to form the three membered nitrogen heterocycle. The aziridine functionality could arise from a number of reasonable β-carbon activation methods such as halogenation, adenylation, sulfonylation, or phosphorylation, complicating prediction of expected intermediates and associated enzymes.

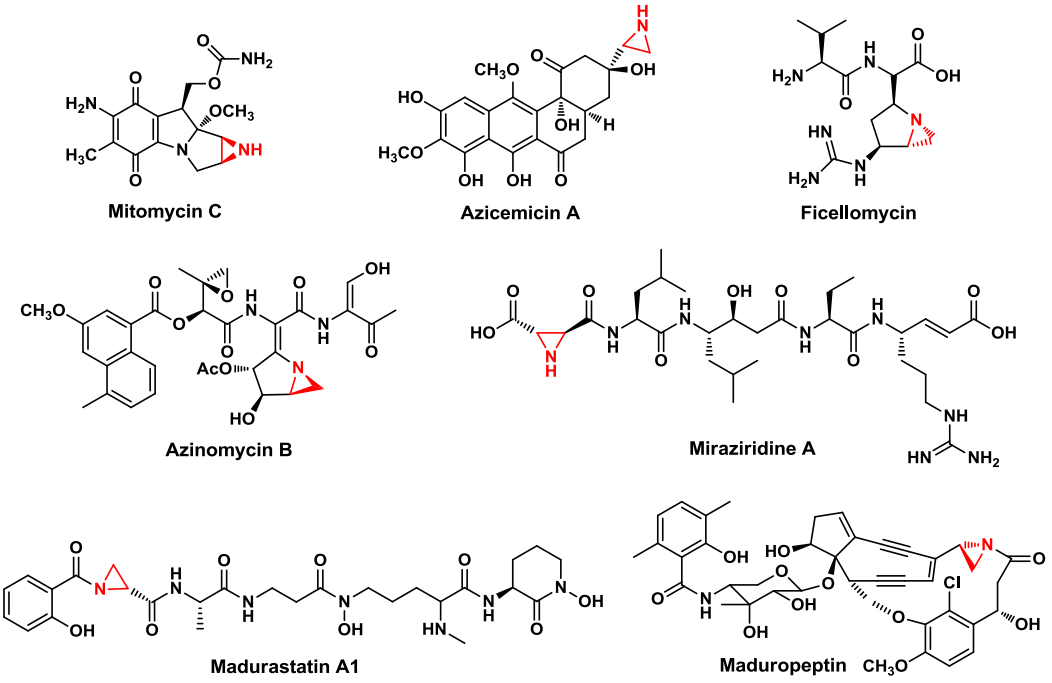


Figure 5 Aziridine-containing natural products isolated to date (2010).

The structures of azinomycin A and B suggested biosynthesis via a hybrid polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS)-containing gene cluster. An analogous naphthoate structure is found in the anti-tumor enediyene neocarzinostatin, for which the gene cluster encodes an iterative type I PKS.[27] The alternating carbonyl and amide backbone of the azinomycin right half was predicted to arise from amino acid-like substrate condensation orchestrated by modular NRPS enzymes. A variety of biosynthetic enzymes for tailoring

reactions, e.g., oxidation, reduction, methylation and acetylation were also suspected to participate.

Biosynthetic origin can be traced by providing isotopically-labeled precursor compounds to a fermentation culture, isolating the natural product, and examining isotopic signal attenuation by NMR. Corre and Lowden first used isotopically-labeled precursor compound feeding to monitor biosynthetic incorporation at specific atom positions in azinomycin B as detected by enhanced ^{13}C-NMR resonance. Providing [1-^{13}C]-, [2-^{13}C]-, or [1,2-^{13}C]-acetate to cultures of *S. sahachiroi* resulted in label incorporation consistent with polyketide synthase-mediated biosynthesis of the azinomycin naphthoate from activated acetyl and malonyl-CoA units.[28] Doubly-enriched acetate served to couple adjacent carbons and supported origin of the acetyl decoration at C13. Additional incorporation due to metabolic scrambling of acetate in the Krebs cycle showed up in C1-C4 of the keto-enol chain, which was proposed to be derived from threonine which was in turn derived from oxaloacetate. Coupled labeling also appeared between the pairs C6-C7 and C12-C13, again attributed to scrambling of acetate to form α-ketoglutarate, the precursor of glutamate, glutamine, arginine, and proline, any of which are conceivable building blocks of the azabicyclo[3.1.0]hexane ring. Synthetic deuterated naphthoic acids were fed and analyzed for incorporation by ^{2}H-NMR to determine the extent to which the ring system is tailored prior to joining the right half of azinomycin.[29] Significant uptake of all three C3' variants into the azinomycin structure was noted, evidence that the naphthoate is fully constructed prior to introduction to the rest of the molecule.

Determination of biosynthetic inhibitors, substrates, and cofactors for azinomycin B biosynthesis utilized cell-free extracts to support *in vitro* production of the compound.[30] This method involved incubation with [1-^{14}C]-malonyl-CoA as a label to track the relative amount of azinomycin B produced by the cell-free system. Reactions were subjected to TLC with standards and appropriate regions of the TLC plate were scraped and collected for analysis by scintillation counting. Cofactors NADPH and SAM were found essential for biosynthesis. Cerulenin, an inhibitor of FAS/PKS enzymes, slowed down naphthoate synthesis. Cytochrome P450 oxidases were targeted using miconazole, metyrapone, and chloramphenicol, eliciting decreased production of both naphthoate and azinomycin B. Incubations with ^{14}C-labeled amino acid building blocks demonstrated reasonable incorporation of ornithine, glycine, valine, and threonine consistent with proposed biosynthetic origins based on structure analysis.

A major obstacle preventing isotopic precursor feedings was erratic production of azinomycin B by *S. sahachiroi* using the current literature procedures. Significant effort was

invested to determine an optimal fermentation system capable of consistent azinomycin B production. Final conditions implemented dehydrated agar plates, a two stage starter culture, and a minimal fermentation media to achieve steady production.[31] This optimized method allowed study of threonine-like precursors for biosynthesis of the enol right end of azinomycin B. [U-^{13}C]-threonine feeding produced azinomycin B site-specifically labeled at the C1-C4 positions. Additional postulated enol intermediates β-ketoamino acid, β-hydroxyamino aldehyde, and β-ketoaminoaldehyde were synthesized as [U-^{13}C]-labeled compounds to evaluate the most advanced precursor accepted by biosynthetic enzymes. All three of the more advanced precursors failed to incorporate above background, indicating threonine is the recognized substrate for initial incorporation, and undergoes subsequent oxidation and reduction steps, of which the order is unknown, to provide the final keto-enol structure.

In azinomycin A, the more minor metabolite produced by *S. sahachiroi*, the enol group is absent. Thus, the origin of azinomycin A C1-C3 was suspected to arise from aminoacetone, a metabolic precursor derived from either glycine or threonine via action of a 2-amino-3-ketobutyrate coenzyme A ligase or L-threonine 3-dehydrogenase, respectively, followed by spontaneous decarboxylation. Cultures provided universally-labeled threonine, [1-^{13}C]- or [2-^{13}C]-glycine or [2-^{13}C]-aminoacetone were compared for relative incorporation of each compound into azinomycin A.[32] While minimal threonine and glycine-derived labels appeared in azinomycin A, aminoacetone incorporated at a very significant level of 26.3%. Genomic sequencing of *S. sahachiroi* also revealed homologs of the enzymes necessary for aminoacetone synthesis. Evidence of a bifurcated pathway for azinomycin A and B biosynthesis modulated by the relative availability of precursors was indicated by unlabeled substrate feeding of aminoacetone, which provided an increased amount of azinomycin A relative to azinomycin B.

A series of epoxyvaline precursors were utilized to study the biosynthetic timing of valine incorporation in azinomycin and the related metabolite epoxyamide. Surprisingly, much of the valine tailoring occurs prior to being loaded onto the NRPS machinery. Of the nine [1-^{13}C]-valine derivatives examined, incorporation levels suggest a pathway that begins with oxidation of L-valine to γ-hydroxyvaline, followed by transamination to give an α-keto acid and dehydration to form 3-methyl-2-oxobutenoic acid.[33] The exact timing of epoxidation is unclear as the instability of epoxide precursors in aqueous media may have contributed to the lack of observed incorporation in feeding experiments. An illustration of precursor incorporations directly implicated in azinomycin biosynthesis is presented in Figure **6**.

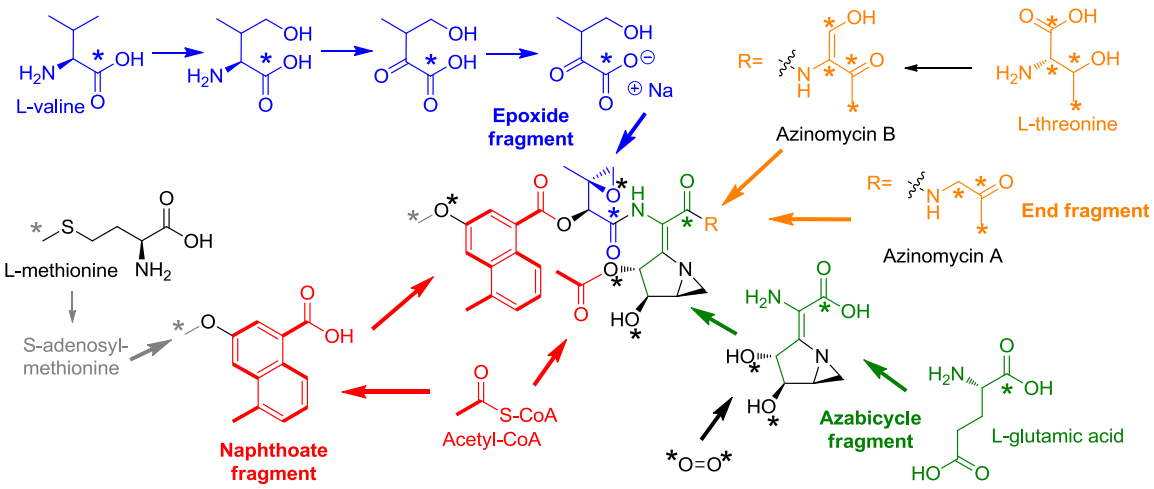


Figure 6 Biosynthesis of the azinomycins as indicated by isotopic tracing experiments. Asterisks (*) indicate isotope-labeled atom positions.

In 2008, Zhao and colleagues proposed the azinomycin B biosynthetic gene cluster based on PCR screening of genomic DNA to identify the probable type I iterative PKS responsible for naphthoic acid (NPA) formation (**Fig. 7**).[25] Screening of an *S. sahachiroi* genomic library with a probe derived from the PKS sequence fragment provided a set of overlapping clones, and further chromosome walking yielded 80 kb of discrete sequence believed to constitute the azinomycin gene cluster. Bioinformatic analysis provided annotation of genes encoding proteins associated with the naphthoate PKS, three NRPS modules, two multi-domain acyl ligase modules, and additional proteins related to possible azabicycle assembly and molecular tailoring. Verification of the pathway linkage to azinomycin biosynthesis was achieved by heterologous expression of *aziB*, the gene encoding the PKS, and detection of 5-methyl-NPA secreted in the fermentation media. Additionally, the roles of a cytochrome P450 hydroxylase (AziB1) and a SAM-dependent *O*-methyltransferase (AziB2) in NPA tailoring were demonstrated in the *Streptomyces albus* heterologous expression system, producing 3-methoxy-5-methyl-NPA as predicted.[25] Subsequent *in vitro* reconstitution of AziB1 and AziB2 overexpressed in *E. coli* presented further confirmation of hydroxylation and methyltransferase activities.[34]

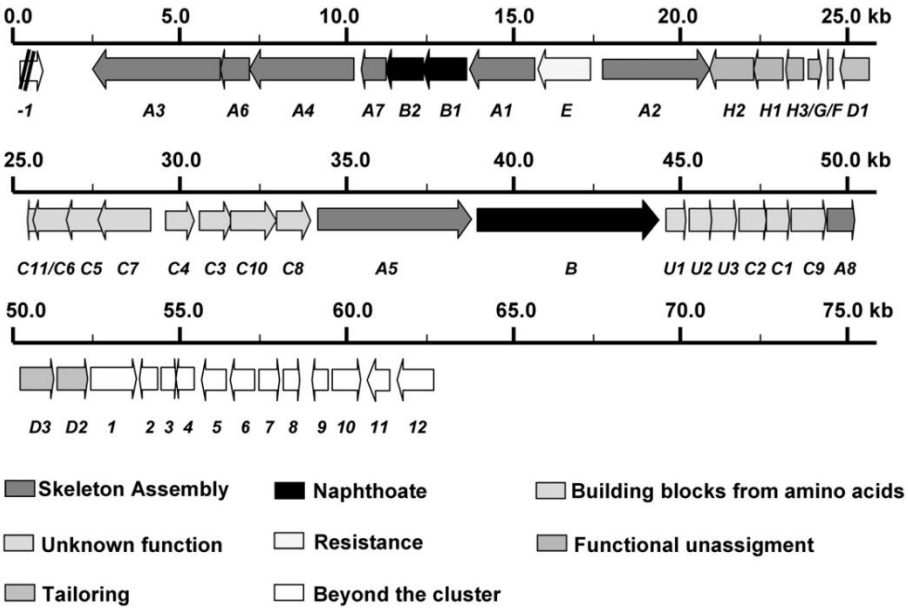


Figure 7 Genetic map of the proposed azinomycin B biosynthetic pathway from *S. sahachiroi*.

To round out the complete set of enzymes participating in naphthoic acid biosynthesis and activation, the substrate specificity of a didomain NRPS (AziA1) was investigated for loading and activation of 3-methoxy-5-methyl-NPA. In a standard ATP-[^{32}P]-PP_i exchange assay, AziA1 was found to activate 3-methoxy-5-methyl-NPA, 5-methyl-NPA, and 3-hydroxy-5-methyl-NPA.[34] A coupled PP_i release experiment demonstrated AziA1 had 3-5 fold higher preference for the most advanced substrate 3-methoxy-5-methyl-NPA. Additional NPA analogs were tolerated by AziA1 with a range of catalytic efficiencies to suggest some promiscuity, but failure to activate 6-methylsalicylic acid or orsellenic acid reveals the naphthalene ring may be a necessary structural component for substrate recognition by this type I iterative PKS.

The enzymology of epoxyvaline, azabicycle, and threonine/aminoacetone formation and incorporation remains to be explored via genetic manipulation and enzymatic reconstitution. Because bioinformatic analysis indicates the pathways may be quite complicated, as in the case of proposed mechanisms for azabicyclo[3.1.0]hexane construction, azinomycin biosynthesis represents a real challenge to reveal novel chemistry at work in bacterial systems.

STREPTOMYCES AND NATURAL PRODUCTS

The genus *Streptomyces* provides a curious wealth of molecules ranging from antibiotic and cytotoxic compounds to hormone-like differentiation signals, immunosupressants, pigments, and metal ion scavengers. Streptomycetes are known to colonize terrestrial ground, plant tissues[35], marine soil[36] and even arctic land, indicating prevalence over almost any environmental condition and hinting at immense adaptation capabilities. With only a few exceptions, Streptomycetes are not pathogenic, in contrast to the behavior of phylogenetic siblings like *Mycobacterium tuberculosis* and *Corynebacterium diphtheriae*. Sequencing of the 8.7 Mb *S. coelicolor* genome in 2002 revealed genes for primary metabolism and essential life functions centrally concentrated on the chromosome, perhaps to protect against accidental loss, and are strikingly similar among all complete *Streptomyces* genomes analyzed to date.[37-39] Localization of secondary metabolic genes near the ends or "arms" of the linear chromosome suggests trade among species by horizontal gene transfer, and these arms show little similarity between species.[40] Given the stressful conditions most Streptomycetes endure, including drought and nutrient deprivation, coupled with their immobility, leads to the conclusion that these numerous gene-encoded small molecules were acquired as the only line of defense against encroaching microorganisms. Additionally, Challis and Hopwood argue that producing several metabolites at a time may have a synergistic effect on competing microorganisms; many low-activity antibiotics can converge to be as powerful as a single superdrug.[41] Debate concerning the evolution and endogenous utility of the diverse secondary metabolic genes occupying ~5% of each *Streptomyces* genome will undoubtedly continue within the scientific literature into perpetuity.

NONRIBOSOMAL PEPTIDE SYNTHETASES, POLYKETIDE SYNTHASES, AND ENGINEERED BIOSYNTHESIS OF NATURAL PRODUCTS

Engineered biosynthesis can refer to natural products produced by enzymes in either *in vivo* or *in vitro* settings. Techniques have been proposed to manipulate pathway steps and subsequent products by domain swapping or DNA shuffling to create combinatorial libraries.[42,43] In addition, mixing and matching enzymes from various biosynthetic pathways in a cell-free system has been used to achieve *in vitro* engineered natural product biosynthesis.[44] Modular enzymes such as polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS) are particularly well-suited for swapping attempts because each synthase or megasynthase is a

collection of discrete domains that catalyze individual activation, condensation, tailoring, and cleavage steps (**Fig. 8**).

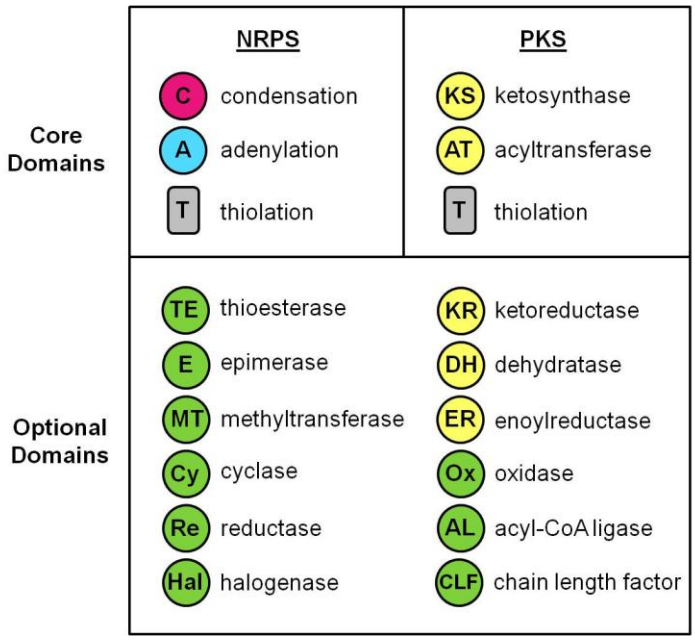


Figure 8 NRPS and PKS domains.

Both NRPS and PKS assembly lines promote thioclaisen condensation of simple activated biological monomers. NRPS modules mediate head-to-tail peptide bond formation of proteinogenic and nonproteinogenic amino acids and some carboxylic acids, requiring at minimum domains for adenylation, thiolation, and condensation. PKS modules polymerize acetate units by employing core ketosynthase, thiolation, and acyltransferase domains. NRPS and PKS products can be diversified by the presence of additional domains for oxidation, reduction, methylation, cyclization or other more specialized transformations, and are known to form hybrid systems to accomplish some syntheses.[45] Covalent cleavage of the peptide or polyketide from the enzyme is accomplished by either thioesterase or intra-molecular cyclization domains.

Thiolation domains (or acyl carrier proteins) require post-translational phosphopantetheinylation of a single serine residue by a dedicated transferase to provide the flexible tether for aminoacyl-AMP or acyl-CoA transthiolation (**Fig. 9A**). NRPS chemistry begins with the adenylation domain, which contains specific active site residues for proper

substrate recognition and attachment of adenosine 5'-monophosphate to the carboxylate end of the amino acid to give aminoacyl-AMP (**Fig. 9B**). Finally, the phosphopantetheine arm guides the monomer to the condensation domain where the free amino group commences nucleophilic attack on the downstream peptidyl thioester, causing translocation of the growing peptide chain (**Fig. 9C**).

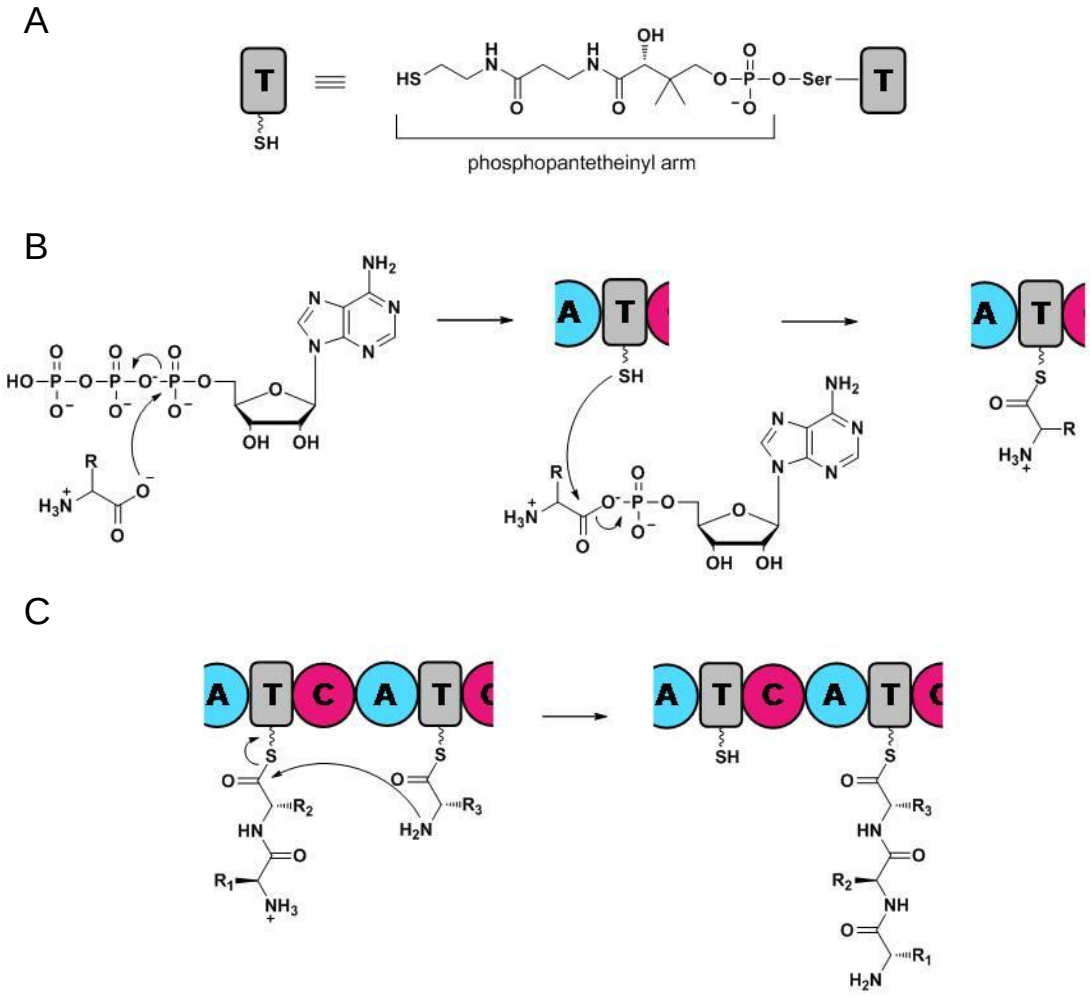


Figure 9 Chemical mechanisms catalyzed by NRPS domains.
(a) The apo-form of a thiolation domain requires phosphopantetheinylation from coenzyme A catalyzed by a phosphopantetheinyl transferase. (b) Amino acid activation occurs from reaction with ATP within A domain followed by covalent thioester linkage in T domain. (c) Condensation reaction mechanism in the C domain.

The PKS mechanism involves a thiolation step similar to that seen in NRPS systems, except the initial activated monomer is acetyl-CoA. Acyltransferase domains commonly operate

on malonyl- or methylmalonyl-CoA (**Fig. 10A**) by nucleophilic attack using a serine residue, forming a covalent acyl-*O*-AT intermediate that is subsequently transferred to the thiolation domain phosphopantetheine arm. A ketosynthase domain performs chain elongation by catalyzing decarboxylation of malonyl- or methylmalonyl-*S*-T to give the enolate for nucleophilic attack of an upstream thioester (**Fig. 10B**).

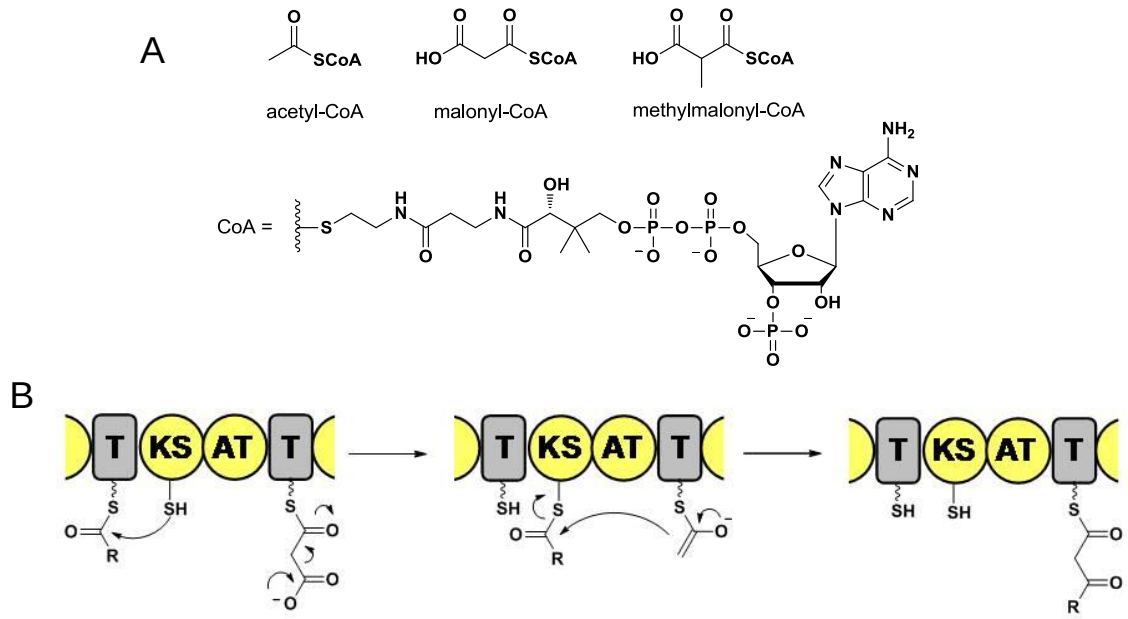


Figure 10 Structures and chemical mechanism of acetate unit incorporation by PKS domains.
(a) common acyl-CoA units operated on by PKS domains, (b) KS and AT domain catalyze decarboxylation and Claisen condensation of each monomer unit onto the growing polyketide chain.

Polyketide systems closely resemble fatty acid synthases (FAS), except that PKSs do not necessarily contain all catalytic domains to form the fully-reduced alkyl chains produced by FASs. Two principal types of PKS systems are found in most bacteria. Type I PKSs are multidomain proteins encoded by a single gene, while type II PKSs function as an in trans protein complex in which each domain is encoded by a separate gene. In general, the type II systems are iterative, meaning domains are reused multiple times for chain elongation, whereas type I domains are used only once per catalytic cycle. However, in the case of systems for some aromatic PKS-derived structures such as orsellenic acid, 6-methylsalicylic acid and 5-methylnaphthoic acid, type I systems have been observed to display iterative behavior.[25,46,47]

The engineering potential of distinct catalytic domain architecture was initially investigated to generate triketide lactones using recombined 6-deoxyerythronolide B synthase (DEBS) modules from *Saccharopolyspora erythraea* (**Fig. 11**).[48-50] Modules 1 and 2, along with the substrate loading module from DEBS1, were genetically fused to the TE domain from module 6 of DEBS3 to determine if a recombinant PKS would retain function. Both heterologous expression and purified enzyme incubations provided the anticipated lactones, supporting the plausibility of domain swapping to generate novel natural product structures.

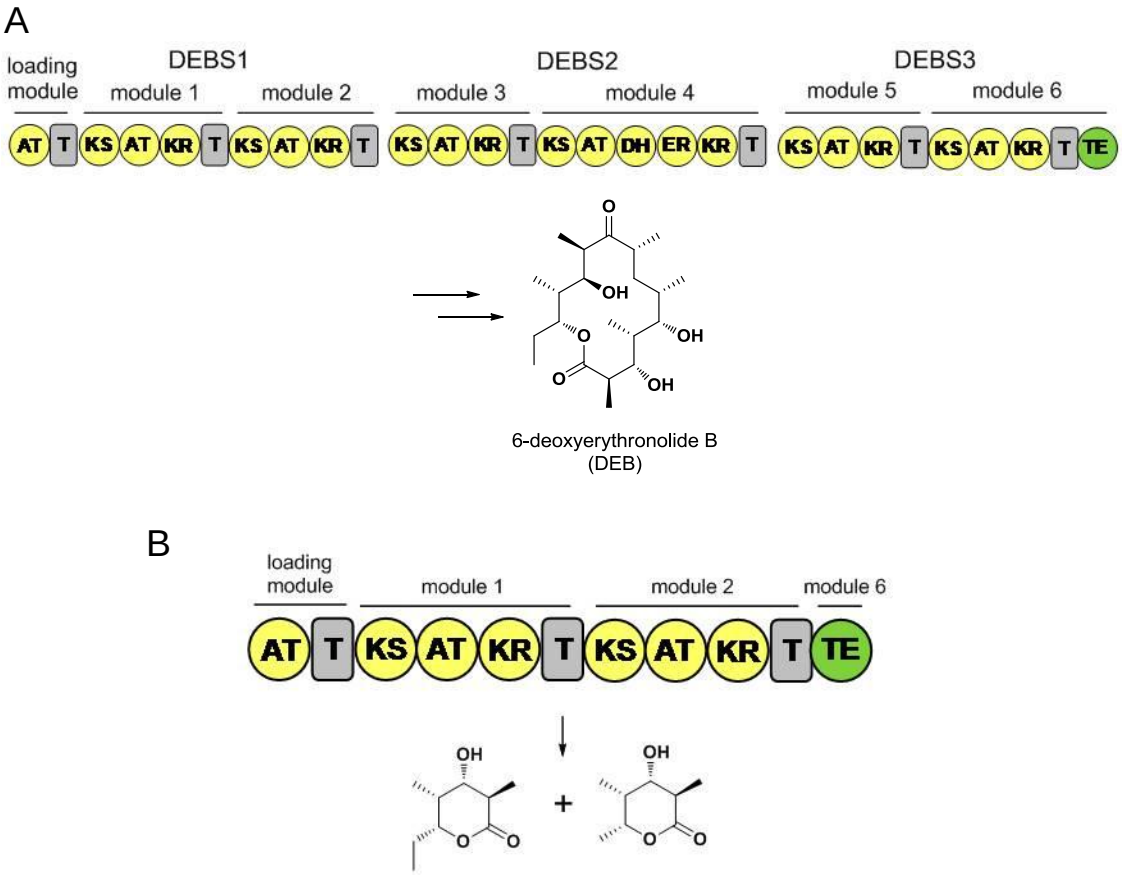


Figure 11 Genetic engineering of DEBS modules.
(a) Native DEBS1, 2, and 3 combined activity produces 6-deoxyerythronolide; (b) Recombined DEBS modules synthesize the predicted triketide lactones *in vivo*, depending upon availability of malonyl-derived substrates.

True combinatorial assembly was later completed using a domain linker strategy to guide logical assembly of domains from a mixture of multiple modules (**Fig. 12**). Each module

or linker gene fragment is flanked by different restriction endonuclease recognition sites. Once digested, the sticky ends ensure that only proper fragments will be ligated together to clone a presumably functional PKS. Again using modules from DEBS, a proof-of-principle experiment showed combinatorial module swapping is a viable option for producing functional enzymes.[51] Care was taken to install the restriction endonuclease cleavage sites in the DNA sequence with minimal peptide sequence disturbance. Sites requiring amino acid substitution used residues known not to significantly disrupt native structure and subsequent functionality. These sequence limitations on module use are somewhat inhibitory of large scale combinatorial enzyme library applications.

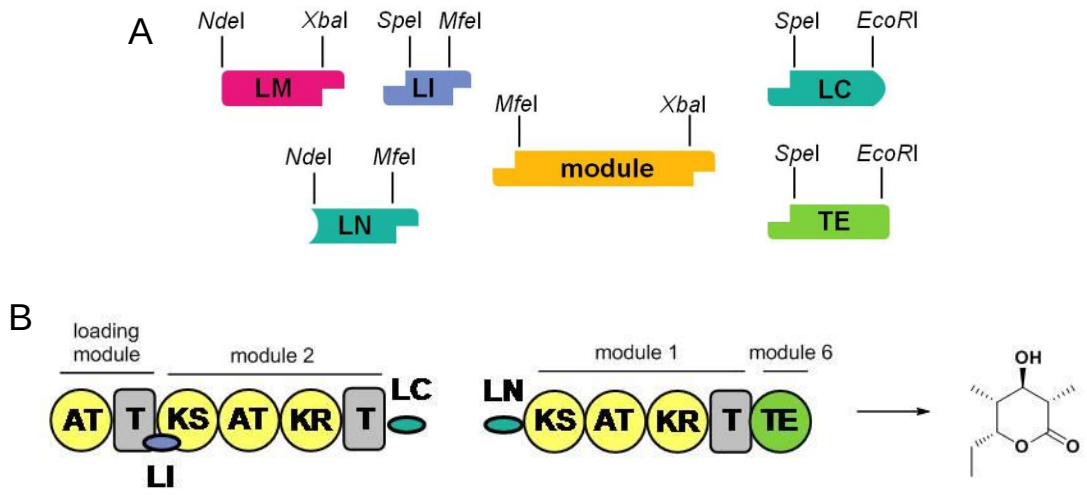


Figure 12 Combinatorial strategy for assembly of PKS modules using synthetic genes.
(a) Gene fragments are flanked with unique restriction sites to guide logical assembly of modules and linkers. LM, loading module; LI, intrapeptide linker; LC, C-terminal interaction domain; LN, N-terminal interaction domain. (b) Experimental recombination of DEBS modules yielded the predicted triketide lactone. Figure after Menzella *et al*.[51]

BACTERIAL DRUG RESISTANCE MECHANISMS

Drug resistance occurs in nature either as an intrinsic property or is acquired as a response to threatening bioactive compounds. All resistance properties fall into the study of what has been termed the 'resistome.'[52] While the evolution of resistance mechanisms is essential for survival and important in relation to engineered biosynthesis of natural products, it poses real economic and epidemiological issues when drugs begin to fail against bacterial, viral, or parasitic infection. Repetitive chemical exposure in increasing concentrations exacerbates the

acquired resistance phenomenon by exerting selective pressure to accelerate development of resistant phenotypes. Virulent drug-resistant strains such as methicillin-resistant *Staphylococcus aureus* and multi-drug resistant *Mycobacterium tuberculosis* are some of the recent publicized consequences of overuse or misuse of antibiotic treatments. Pathogens often employ more than one strategy to circumvent the effects of drug exposure, complicating efforts to develop alternative therapies. Understanding the biochemical mechanisms of drug resistance keeps us one step ahead in the seemingly endless quest to overcome microorganismal adapations to new drugs.

The first antibiotic resistance enzyme discovered was *B. coli* penicillinase, a class A β-lactamase that catalyzes penicillin β-lactam hydrolysis. Abraham and Chain found penicillin incubated with *B. coli* cell extract was deactivated when tested against *S. aureus*.[53] Degradative resistance proteins like β-lactamases can alter a functional group or cause structural rearrangement, rendering the compound unreactive toward the intended target. Drug-modifying resistance proteins can also promote covalent transfer of phosphate, acetate, adenylate, glutathione, or glycosides to inactivate a drug by blocking a reactive group or creating a structure that no longer fits within the target active site. An example is aminoglycoside phosphotransferase, which operates on aminoglycoside antibiotics such as kanamycin, streptomycin and neomycin, transferring the γ-phosphate of ATP to a carbohydrate hydroxyl to prevent drug binding to the A-site of the 30S ribosomal subunit.[54]

The resistome includes not only enzymes for chemical inactivation, but also drug target mutations, influx reduction mechanisms, and expression of transport proteins, binding proteins and efflux pumps (**Fig. 13**). Bacterial genomes code for numerous small molecule transport proteins, some of which are non-specific toxin pumps that protect against a range of environmental chemicals, such as the multi-drug binding capacity of AcrB in *E. coli*.[55] Other pumps are more specialized, as in the case of TetA for tetracycline resistance.[56] A case of drug target modification is evident in *Streptomyces erythraeus*, in which ribosomal RNA is methylated by ErmE to protect against erythromycin.[57] A class of drug binding proteins confer resistance to bleomycin, phleomycin, and tallysomycin in several actinomycetes, protecting genomic DNA from cleavage by oxygen radicals produced when the drugs chelate iron and become pseudoenzymes.[58] Binding proteins do not effect chemical change, but serve to sequester the drug from potential targets until removed from the cell.

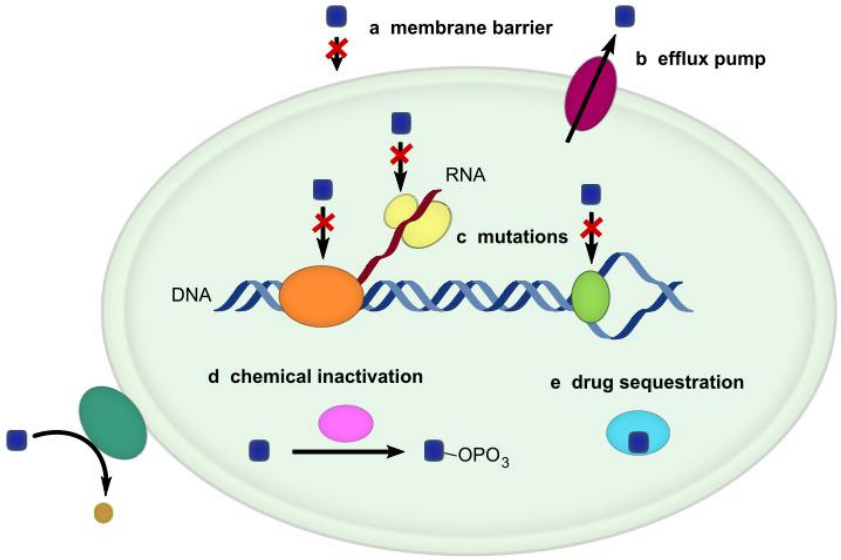


Figure 13 A summary of biochemical drug resistance mechanisms in the cell.
(a) Membrane acts as a diffusion barrier for some small molecules. (b) Efflux pumps remove toxic compounds from the cell. (c) Genetic mutations produce proteins lacking the original drug binding site. (d) Proteins add acetate, phosphate, or otherwise convert the molecule into an inactive form. (e) Drug binding proteins sequester the drug from reactive targets. Figure after Allen et al.[59]

Efflux pump and target modifying proteins can be difficult to introduce in heterologous hosts, especially for purposes of natural product biosynthetic engineering, because they pose the possibility of biological system upset. Pumps must be integrated in the cell membrane, and target modification must be tolerated by all other macromolecules dependent upon the target for normal cell function. Binding proteins or chemical modifying enzymes are theoretically more facile resistance methods to introduce in heterologous expression hosts because less punishment is brandished on cellular housekeeping proteins or physical cell structure.

CRYPTIC/ORPHAN BIOSYNTHETIC PATHWAYS

As resistance mechanisms continue to evolve, so must efforts to mine new medicinal natural products. Current estimates of isolated natural products compared to the number of possible secondary biosynthetic gene clusters available in microorganisms suggest much opportunity is being missed by conventional environmental sample collection, fractionation and screening methods. As little as 1% of the global microbial potential for biologically-relevant molecules may be represented in the archived natural product pool, in part because of our unawareness regarding the number of microorganisms inhabiting the planet.[60]

Completion of the *Streptomyces coelicolor* A3(2) genome in 2002 was one of several bacterial sequencing projects to reveal natural product-like gene clusters for which the product was unknown.[37] Prior to the disclosure of these pathways natural product drug discovery was deemed unsustainable, judging from the gradual decline in new natural product structures reported during the 1980s and 1990s and the frequent reisolation of known compounds from new microbes. The clusters, termed 'cryptic' or 'orphan' pathways to designate not having a natural product pairing, offered fresh perspective on gene-encoded small molecule sources. Cognizance of orphan pathways spurred specialization in the area of genome mining as a promising next-generation technique for drug discovery.

Orphan biosynthetic gene clusters have unassociated natural products for a number of reasons. If the compound is produced in reasonable quantity, screening methods simply may not be optimized to identify particular physical attributes. However, if the compound is expressed at undetectably low levels, the pathway may be transcriptionally silenced or downregulated under normal laboratory fermentation conditions. This case is observed for many of the secondary metabolic compounds produced by Streptomycetes, and requires special conditions (stress) or environmental cues (molecular signals from symbionts) to jumpstart clustered gene expression. In other cases, the pathway may have been acquired by horizontal gene transfer and lacks the necessary transcriptional promoters for recognition as coding sequence. A robust resistance mechanism may be lacking, so a nonproducing phenotype was selected for over subsequent generations. Mutations and frameshifts could silence a previously productive pathway. The explanations for the presence and perceived silence of orphan biosynthetic pathways in microorganisms are endless, setting the stage for studies bridging evolution, transcriptional biology, and bioinformatic prediction.

Methods to uncover orphan clusters and their associated natural products focus on DNA sequence-based tools in the form of whole genome sequencing, degenerate PCR screening, genomic scanning, or metagenomics. Whole genome sequencing and annotation, while the most time and labor intensive, gives a complete picture of an organism's metabolism and is searchable for favorite biosynthetic protein classes, which in turn reveal clustered pathways. Degenerate PCR queries genomic DNA or genomic cosmid libraries using oligonucleotides designed from conserved regions of NRPS and PKS-type proteins involved in secondary metabolism, and resulting sequence fragments aid discovery of complete clusters. Similar to degenerate PCR screening, genomic scanning uses a random gene sequence tag library to determine appropriate sequences, then uses the sequences to fish out longer pathway fragments from a BAC library and

ultimately piece together a finished pathway. For the case of unculturable or undiscovered organisms, metagenomics provides the means to identify new natural product pathways via heterologous expression of environmental DNA samples, but is the most risky and difficult because host codon bias or sequence intolerance may destroy any chance of successful metabolite production.

Following genetic identification of a putative biosynthetic cluster there are several options for product prediction and verification. Generally, bioinformatics provide a starting point for determining specific small-molecule building blocks and tailoring reactions by examining conserved domain motifs in encoded proteins and comparing them to pathway enzymes with known products. From there, induced expression and gene inactivation or isotope tracer feedings are equally valuable tools to identify pathway intermediates. Genetic knock-out requires disruption or deletion of a specific gene to prevent normal transcription, usually aided by homologous recombination with a plasmid-delivered copy of the gene interrupted by a selectable genetic marker to phenotypically indicate gene replacement. Isotopically-enriched building blocks provided to a fermentation culture incorporate the label in the resulting natural product for detection by NMR signal attenuation. If the pathway is not functional due to missing promoters or other lack of activation, or metabolite production is not robust, *in vitro* enzymatic reconstitution serves as a bypass for *in vivo* expression profiling. Individual purified enzymes predicted to catalyze pathway steps are assayed for tolerance and turnover of appropriate substrates. Entire multi-enzyme biosynthetic pathways and complete *in vitro* biosyntheses have been elucidated following reconstitution methodology.[61-63]

NEXT-GENERATION GENOME SEQUENCING

As whole genome sequencing becomes more financially accessible and requires less DNA starting material, time, and effort, it reigns as an optimal technique for sequence-driven biosynthetic natural product discovery. Complete genomic information is essential for detection of unclustered genes, fragmented pathways, and primary metabolic pathways that provide some of the unusual building blocks for secondary metabolite biosynthesis. The natural process of polymerase-mediated DNA replication was the inspiration for both early and current sequencing-by-synthesis (SBS) approaches, and future prospects promise to miniaturize the process to single-molecule imaging of base pair incorporation. While arguably the greatest impetus for evolving DNA sequence acquisition has been the International Human Genome Project, the relevance of improved techniques for sequencing of all living organism genomes is obvious.

DNA sequencing began a little over three decades ago with publication of the Sanger method for enzymatically incorporating chain-terminating 2',3'-dideoxynucleotide triphosphates (ddNTPs) in strand replication.[64] The ddNTPs are present in very small amounts relative to deoxynucleotide triphosphates (dNTPs) in a sequencing reaction, producing a stochastic distribution of synthesized single-strand DNA fragments. In the original method, [^{32}P]-end labeled fragments were separated by polyacrylamide gel electrophoresis and imaged by autoradiography to deduce the corresponding DNA sequence. Shortly after, ddNTP-conjugated fluorophores and capillary electrophoresis separation with laser detection came to constitute the standard protocol now employed for routine sequence verification of cloned DNA in plasmids.[65,66]

Whole genome sequencing was once a literally insurmountable task, even for very small (~1Mb) single chromosome species. While DNA sequencing methodologies had been in existence since the 1970s, avenues for assembly of the 25,000+ pieces of sequence data minimally generated in a theoretical genome sequencing application did not exist. Computational assembly strategies and high-throughput sequencing was until on hold until 1995 with the reported 1.8 Mb whole genome shotgun sequence of *Haemophilus influenzae* Rd.[67] By employing a randomized approach to sequence acquisition, the *H. influenzae* project circumvented the barrier imposed by direct sequencing (chromosome or primer walking). Primer walking can be hindered by DNA secondary structures, self-annealing sequence regions (hairpins) and high G+C content. Fragmentation and subsequent reassembly overcomes this obstacle, provided the shotgun library is sequenced statistically deep enough for complete genome coverage. Though the exact technique pioneered for the *H. influenzae* project is no longer a primary route to whole genome sequencing, the shotgun concept is present in all of the more advanced approaches available today.

Current genome sizes range from 160,000 bp (*Carsonella ruddi*) to over 670 Gb (*Polychaos dubia*). The human genome sequence, reported as a draft form in 2001[68] and considered roughly complete in 2003, rests between these at 2.9 Gb. Genome length is not a predictor of organism size, complexity, or encoded proteins as evidenced by a sampling of the 1,196 fully sequenced genomes publicly available as of February 2010 (**Table 1**). The European toad genome is three times the size of the human genome, but has yet to fully sequenced. Some of the smallest single-cell eukaryotes, amoebas, have genomes 100 to 300-fold larger than that of humans, and will be enormous sequencing projects if they are undertaken. Much of the mystery

surrounding disparate genome size and organism complexity begs to be explored by future whole-genome sequencing applications.

Table 1 A sampling of the distribution of genome sizes among living creatures. Values are from the Database of Genome Sizes.[69]

Organism	Common Name or Description	Genome Size	Coding Genes[a]
Carsonella ruddi	bacterial symbiont of psyllids (sap-feeding insects)	160 kb	182
Escherichia coli	rod-shaped intestinal bacterium	4.1 Mb	4,800
Streptomyces coelicolor	terrestrial bacterium	8.9 Mb	7,825
Saccharomyces cerevisiae	fungi, baker's yeast	12 Mb	6, 275
Caenorhabditis elegans	roundworm	97 Mb	19,000
Drosophila melanogaster	fruit fly	120 Mb	13,601
Boa constrictor	large, heavy-bodied snake	2.1 Gb	unknown
Homo sapiens	human	2.9 Gb	30,000
Bufo bufo	European or common toad	6.9 Gb	unknown
Amoeba proteus	single-cell eukaryote; uses pseudopodia to move/eat	290 Gb	unknown
Polychaos dubia	large freshwater amoeba	670 Gb	unknown

[a]Entries noted as unknown refer to genome sequencing either not yet undertaken or incomplete.

Because read length and throughput contribute greatly to the cost and effort invested in genome sequencing, the last decade has been witness to optimization attempts for three types of massively parallel high-throughput sequencing platforms termed 'next-generation' applications: Illumina Genome Analyzer II (GAII), Applied Biosystems Sequencing by Oligonucleotide Ligation and Detection (ABI SOLiD) system, and Roche/454 Life Sciences Genome Sequencer FLX (GS FLX) platform. All three offerings operate on the SBS premise, but implement the process in very different ways (**Table 2**). Sequencing platforms are primarily characterized as next-generation because they employ not only unique approaches to DNA base incorporation and detection, but also because the number of simultaneous sequence fragments analyzed can reach into the millions. A further advantage of these techniques is the avoidance of bacterial subcloning and enzymatic digestions, both of which introduce bias in representation of the

genome if stretches of DNA are unstable in the bacterial host or lack suitable endonuclease recognition sequences.

Table 2 Comparison of next-generation sequencing strategies.
Statistics reported by Mardis.[70]

	Roche/454 GS FLX	**Illumina GAII**	**ABI SOLiD**
Sequencing Chemistry	coupled pyrophosphate release	polymerase-based	ligase-based
Amplification approach	emulsion PCR	bridge amplification	emulsion PCR
Paired end separation	3 kb	200 bp-5 kb	600 bp-10 kb
Bases sequenced	400-600 Mb	2-2.5 Gb	20 Gb
Time (paired ends)	7 h	4 d	5 d
Read length	≈250 bp	≈25-100 bp	≈50-100 bp
Cost per single-end run	≈$8,500	≈$9,000	≈$17,500

The Illumina GAII platform consists of four stages for DNA library preparation, cluster generation, sequencing, and base calling image analysis. The library is generated from randomly sheared and size-selected genomic DNA to which oligonucleotide adapters are ligated. The adapters permit PCR-guided enrichment of the library, serve to anchor the DNA fragments to the surface of a flow cell, and prime the bridged amplification of fragments to form ~2000-molecule clonal clusters necessary for fluorescent signal detection by a charge-coupled device (CCD). Sequencing relies on reversible terminators for each step (**Fig. 14**). Once an incorporated base has been imaged, the fluorophore blocking the 3'-hydroxyl end is cleaved to permit another round of base addition. The process is repeated anywhere from 25-100+ times depending on the specific sequencing aim. The sequencing process occurs in parallel across more than six million clusters per run, generating 1.5 Gb per day with an accuracy of 99.999% at three-fold or greater coverage. Because each sequence generated is ~100 bp or less, specialized open-source sequence assembly software packages such as SSAKE, VCAKE, SHARCGS, Velvet, and Edena were developed to handle very short reads and some have the option for assembly of mate-paired end sequences.[71] Also, the number of reads required for *de novo* genome sequencing using Illumina technology can be 25-fold greater than for the same data obtained from Sanger reads, translating to a much larger data file occupying computational memory. The demand for

resources is arguably mediated by the relative ease and speed with which Illumina sequence permits access to genomes that fail sequencing by traditional methods.

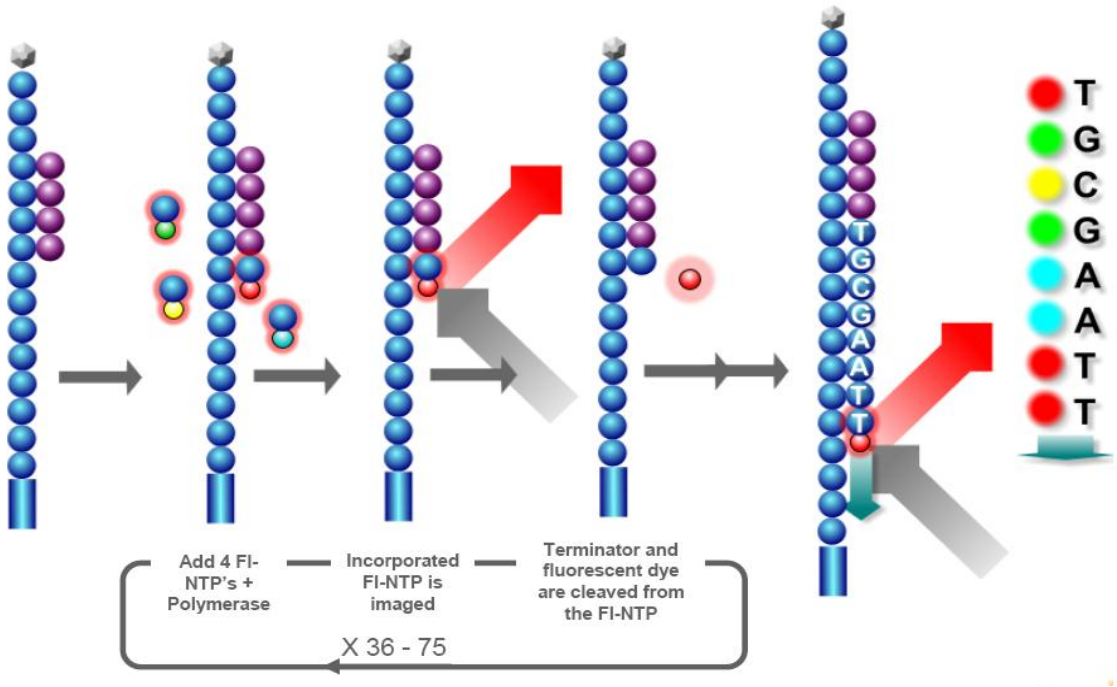


Figure 14 Illumina GAII sequencing method employing reversible terminator chemistry.
A primer is annealed to the single-stranded DNA immobilized on the flow cell, and then 3'-OH fluorescently-capped NTPs are introduced. The incorporated base is imaged and the fluorophore cap is cleaved to enable the next round of polymerization. Figure from Illumina corporate literature.

The ABI SOLiD platform carries out adapter ligation similar to Illumina GAII, but amplifies sheared genomic DNA fragments in emulsion PCR (ePCR) with magnetic beads displaying PCR primers. ePCR creates aqueous 'microreactors' suspended in an oil, each containing a single DNA template and bead. The beads are then purified and covalently attached to a glass slide for sequencing by ligation (**Fig. 15**), in which a primer hybridizes to the adapter sequence and a pool of degenerate fluorescently labeled octamers compete for ligation to the primer. Only octamers homologous to the first five bases of the DNA template will anneal for subsequent joining by a DNA ligase. Colors correspond to the combination of bases at position 4 + 5. Following imaging, bases 6-8 and the tag are chemically cleaved and the pool of

fluorescent probes is again introduced to sequence positions 9 + 10, then 14 + 15, 19 + 20, and so on. Synthesized strands are finally stripped from the template and probe annealing and ligation is repeated with primers offset by 1, 2, 3, or 4 bases to fill in sequence information for the gaps leftover from the initial round of sequencing. The number of beads deposited on the glass slide is theoretically limited only by available space, but signals must have a certain degree of separation to be distinguishable in the captured image and thus is a parameter that limits not only ABI SOLiD but also the flow cell capacity of the Illumina GAII system.

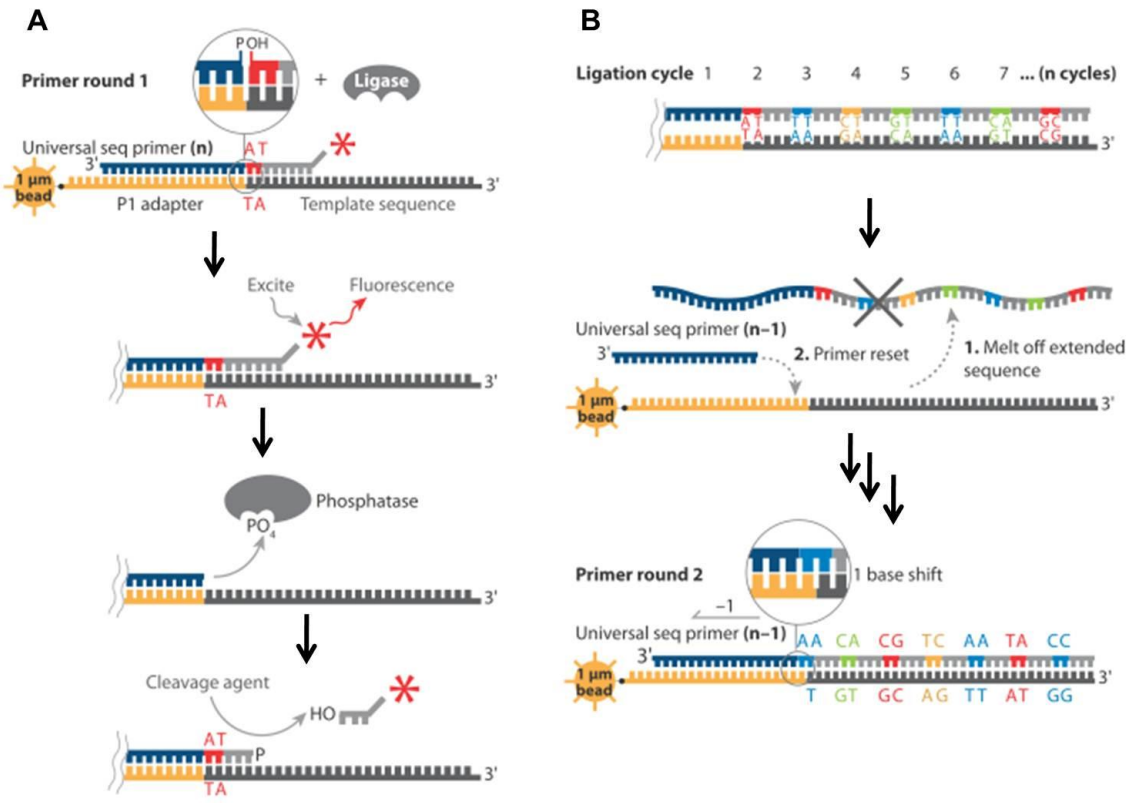


Figure 15 The ABI SOLiD process.
(a) Octamer annealing and ligation, tag imaging, capping of unprimed sequences, and tag cleavage. (b) Completed synthesized strand is removed and a new sequencing primer is annealed with a 1 base offset to fill in sequence gaps in the same manner depicted in panel **a**. Process is repeated with 2, 3, and 4 base offset primers. Figure after Mardis.[70]

The Roche/454 Life Sciences GS FLX platform also utilizes bead-bound DNA templates obtained through ePCR, but arrays individual beads in unique wells of a picotiter plate (PTP) to

provide a fixed location for sequencing, and is analogous to the flow cell or glass slide used in the previously detailed platforms. DNA polymerase and beads containing ATP sulfurylase and firefly luciferase are added to the PTP to perform pyrosequencing, which couples the release of inorganic pyrophosphate (PP_i) from standard nucleotide incorporation with a chemiluminescent reaction[72] monitored by a CCD positioned at the base of each well (**Fig. 16**). Nucleotides are sequentially admitted to the PTP for reaction, and the intensity of light emitted is directly proportional to the number of As, Ts, Cs, or Gs incorporated during each step.

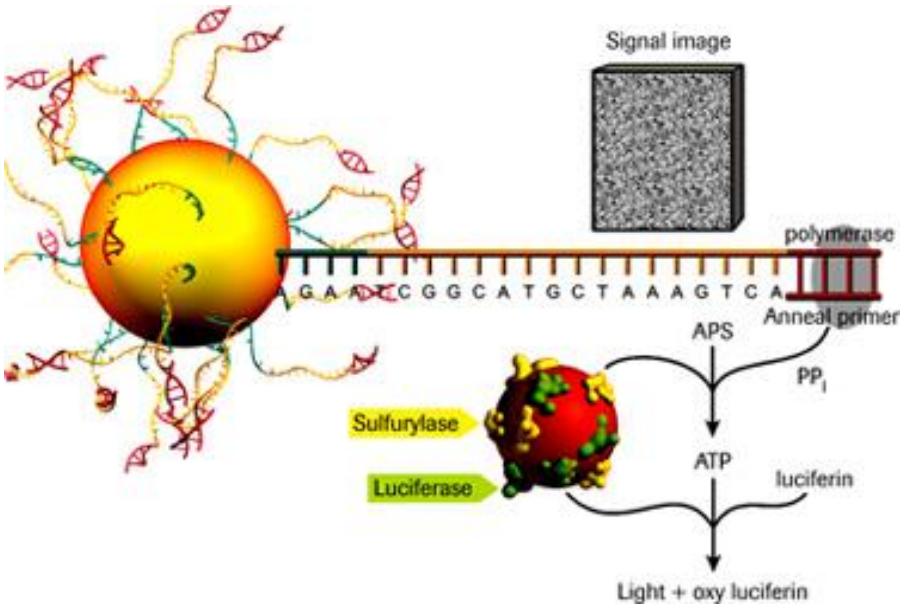


Figure 16 Roche/454 Life Sciences GS FLX bead-based pyrosequencing implementation.
Figure from Roche corporate literature.

The pyrosequencing mechanism (**Fig. 17**) relies on a standard DNA polymerase for ligation of new nucleotides to a DNA strand. This process release inorganic phosphate (PP_i). ATP sulfurylase uses PPi to convert adenosine phosphosulfate (APS) to adenosine triphosphate (ATP), which serves as a cofactor for luciferin oxidation via activation with addition of adenosine monophosphate. A side product of subsequent oxyluciferin formation is photon release, permitting detection of the coupled reaction.

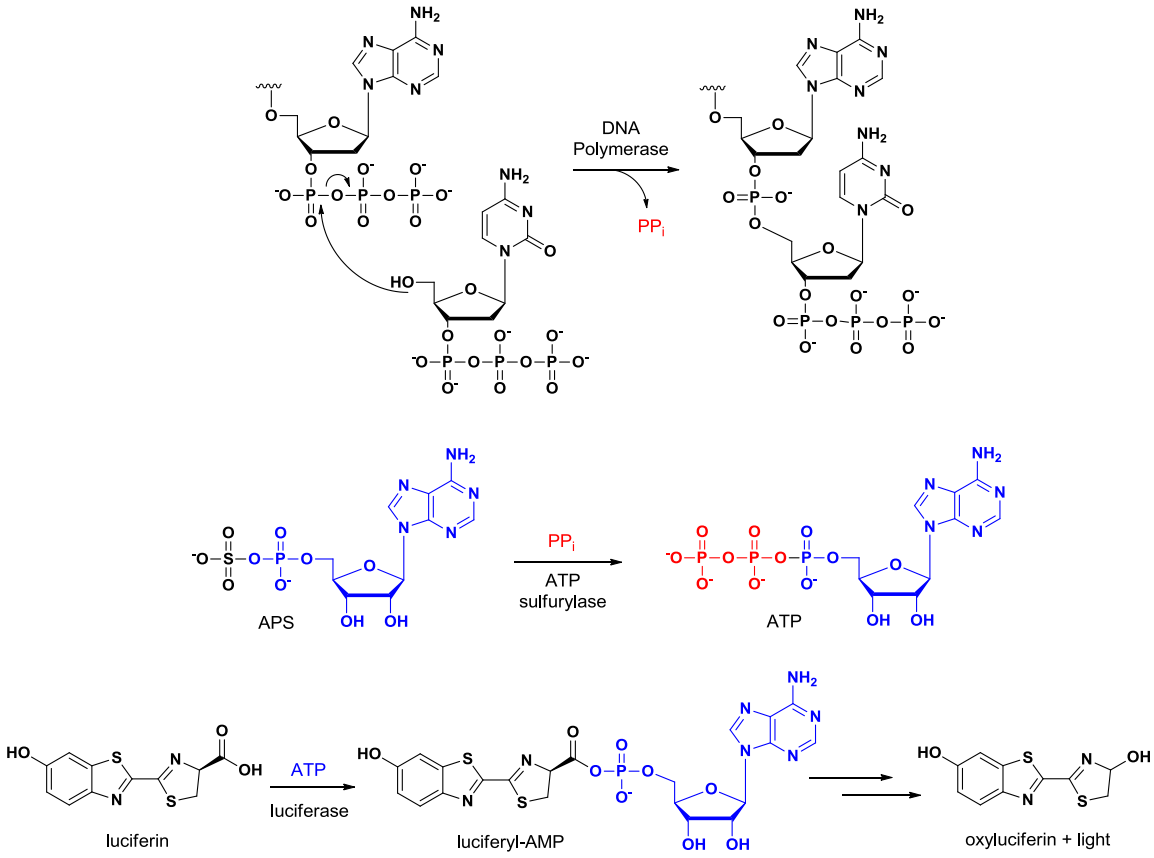


Figure 17 Roche/454 Life Sciences pyrosequencing mechanism.

Each of the next-generation sequence platforms has application in other areas of genomic science beside whole-genome sequencing.[70] For eukaryotic gene expression studies, chromatin immunoprecipitation (ChIP) first identifies specific proteins bound to DNA using an antibody, then the associated DNA is extracted for sequencing to reveal DNA-protein interactions. The high-throughput power of next-generation sequencing overcomes two obstacles encountered in low-throughput ChIP: statistical representation of putative binding sites and genome-wide evaluation from a single experiment. Quantitative approaches to understanding gene expression are also facilitated by whole-genome analysis of sequences from short-read technologies. Additional applications of Illumina, 454, or ABI SOLiD sequencing include noncoding RNA discovery, ancient genome sequencing, and metagenomic analyses. Short reads in each of these cases are optimal because the RNA/DNA fragments are inherently smaller and often available in only minute quantities.

NATURAL PRODUCT AS FOE: RETINOID BY-PRODUCTS OF THE VISION CYCLE

The vision cycle is a light-dependent process involving isomerization and regeneration of photoresponsive vitamin A-derived molecules (**Fig. 18**). The cycle is considered to begin with RPE65 isomerohydrolase-mediated conversion of all-*trans*-retinyl fatty acid ester to 11-*cis*-retinol, which then undergoes dehydrogenation and translocation from the retinal pigment epithelium (RPE) to the rod outer segment (ROS). Rhodopsin forms a Schiff base with 11-*cis*-retinal to prepare for photoisomerization, the step during which vision occurs. The all-*trans*-retinilidine-rhodopsin conjugate is hydrolyzed to give all-*trans*-retinal (ATR), which must be reduced, removed to the RPE, and acylated to continue the cycle anew. Because several steps of the vision cycle require independent enzymes with unique kinetic capabilities, pathway intermediates such as ATR can accumulate in the ROS. RPE cells consume the ROS through phagocytosis, from which incomplete digestion results in by-products known as lipofuscins, autofluorescent species implicated in a number of ophthalmologic disease states including age-related macular degeneration.

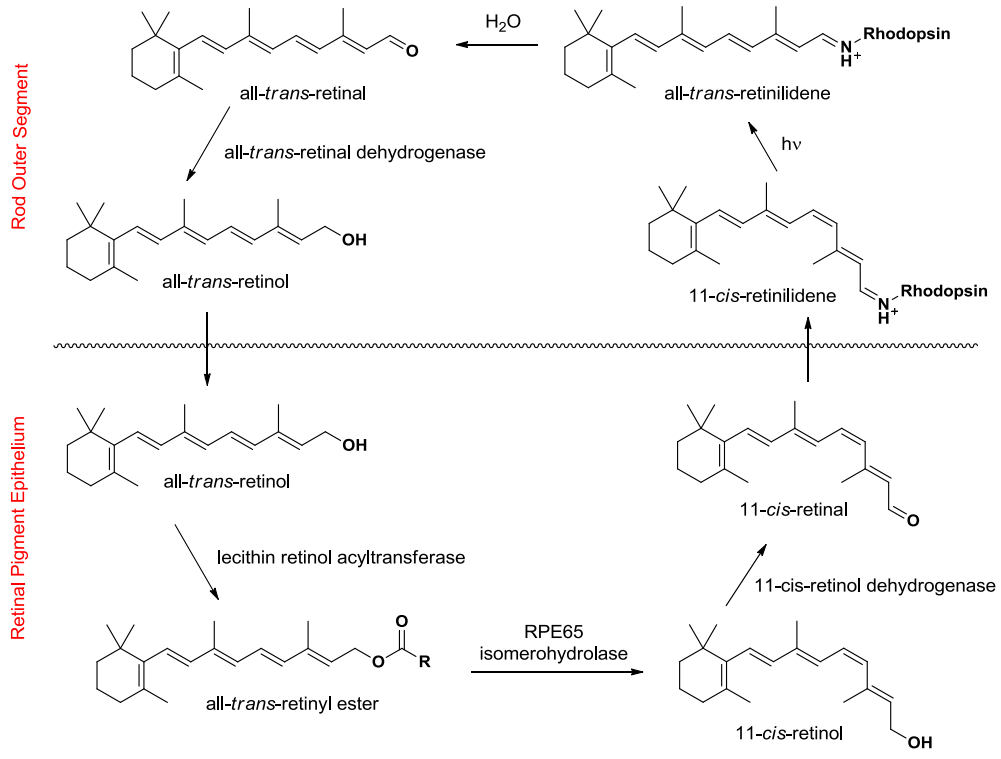


Figure 18 The human vision cycle.
Upper reactions take place in the rod outer segments and lower reactions take place in the retinal pigment epithelium.

The autofluorescent by-products of ATR include the bisretinoids A2E and ATR-dimer (cycloretinal), both of which have been isolated from RPE lipofuscins.[73] These adducts appear to be biosynthesized from two molecules of ATR in the case of cycloretinal, and an additional molecule of phosphatidylethanolamine (PE) in the case of A2E (**Fig. 19**).

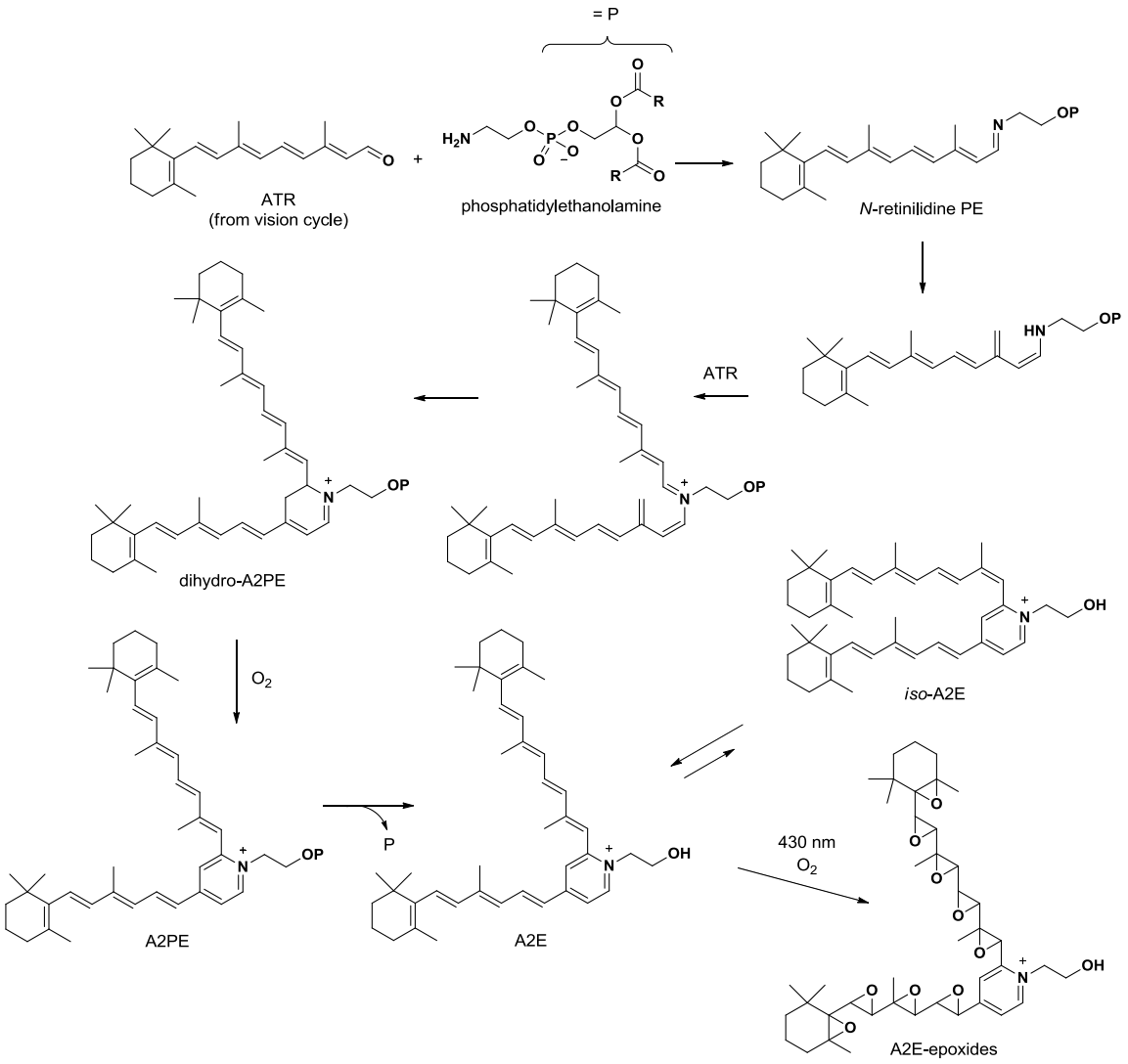


Figure 19 Proposed biosynthesis of retinoid-derived compounds found to accumulate in lipofuscins of the RPE. R groups in phosphatidylethanolamine designate the long alkyl chains.

A number of pathway intermediates (*N*-retinilidine PE, dihydro-A2PE, and A2PE) have also been identified to support proposed biosynthetic origins. A2E and *iso*-A2E exist in a 1:4 equilibrium, and the highly conjugated system has been shown to undergo oxidation in the presence of certain wavelengths of light to generate DNA damaging A2E-epoxides.[74] ATR is the aldehyde analog of vitamin A/retinoic acid, which in instances of dietary deficiency has been linked to reduced RPE lipofuscin accumulation,[75] and provides further evidence that A2E and similar product formation is tied to availability of retinoid precursors driving the vision cycle.

While A2E and related conjugates could conceivably arise from spontaneous chemical reactions, formation of cycloretinal would require enzymatic intervention (**Fig. 20**). ATR and PE are a logical nucleophile-electrophile pair, but condensation of two molecules of ATR would require activation in a biological system, perhaps via Schiff base formation with lysine residues in a protein. The analogous biomimetic reaction using L-proline as a catalyst is already an established synthetic route to cycloretinal.[76]

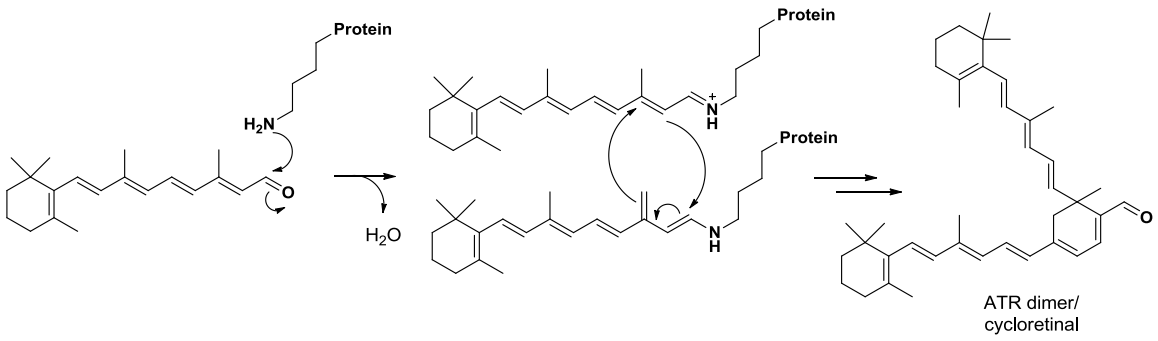


Figure 20 Proposed mechanism for cycloretinal formation mediated by proteinaceous lysine residues.

MACULAR DEGENERATION DISORDERS

Age-related macular degeneration (AMD) is one of several retinal diseases marked by angiogenesis, the proliferation of new blood vessels from existing vascular structures (**Fig. 21**). In AMD patients, angiogenesis is localized under the macula in the central region of the light-sensing retina. A disc less than 4 mm in diameter and 250 μm thick, the macula houses cells and specialized structures necessary for 20/20 vision.[77] Neovascularization beneath the macula is often fragile and hemorrhages blood or other fluids, which displaces the macula and ultimately leads to loss of central vision. While this condition adversely affects daily tasks such as reading,

driving, and face recognition, the disease is painless and often develops slowly, affecting either one or both eyes. Retinoid-like fluorophores found in the lipofuscin of RPE cells are associated with AMD progression and RPE cell death. Extracellular deposition of drusen, which are yellow granules composed of proteins, lipids, carbohydrates and zinc, is also often observed as a precursor to the development of maculopathies such as AMD. Proteomic studies have indicated the presence of β-lactoglobulin in the drusen of AMD donors,[78] a protein observed to catalyze *in vitro* dimerization of retinoid compounds.[79]

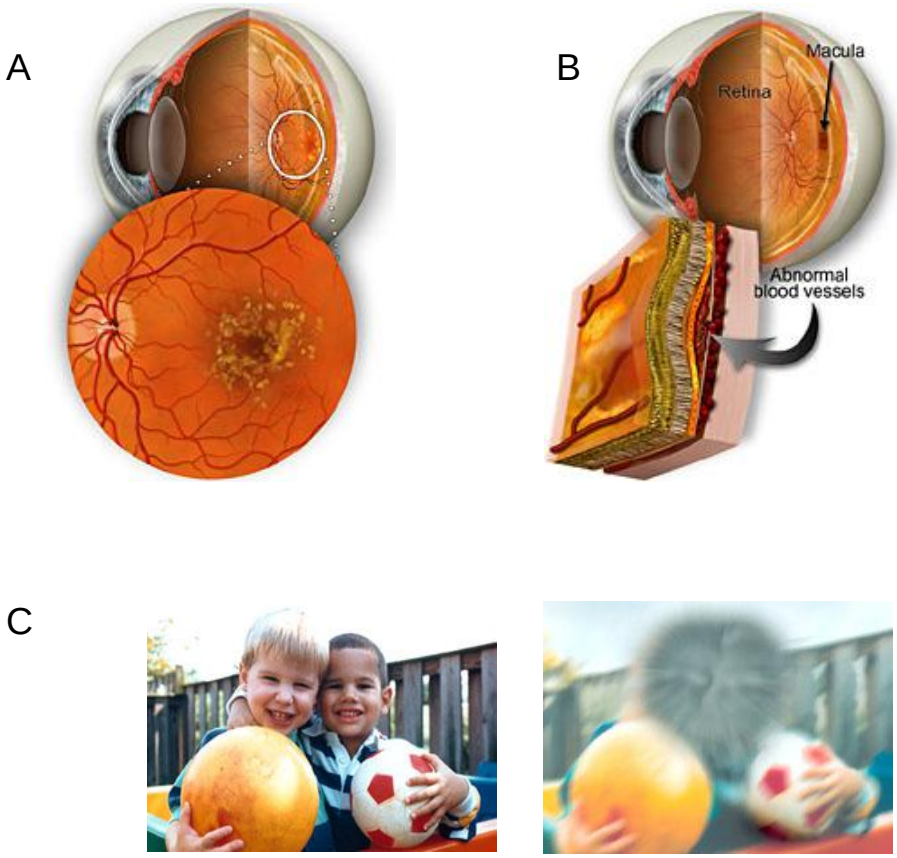


Figure 21 Physiological effects of age-related macular degeneration.
(a) Dry AMD (early stage) is characterized by yellow drusen deposits in the macula. (b) Wet AMD (advanced stage) is diagnosed when fragile, abnormal blood vessels growing beneath the macula begin to leak, raising the it from the normal position. (c) Representative images of normal eyesight compared to that of an advanced AMD patient. Images courtesy of the National Eye Institute.

Stargardt disease is an AMD-equivalent maculopathy but affects juveniles with mutations in both alleles of the ABCA4 (ABCR) gene, a membrane-bound flippase believed to transport *N*-retinylidene-phosphotidylethanolamine into the ROS for conversion to ATR for the vision cycle.[80] Individuals who are heterozygous for the Stargardt mutation do not experience symptoms at a young age but are at high risk for developing AMD later in life.[81] An Abcr null mutant mouse model was developed to study the Stargardt phenotype, in which lipofuscin accumulation occurs at an accelerated rate compared to AMD, and this model has proved useful in determining the identity of a number of lipofuscin fluorophores including cycloretinal.[73]

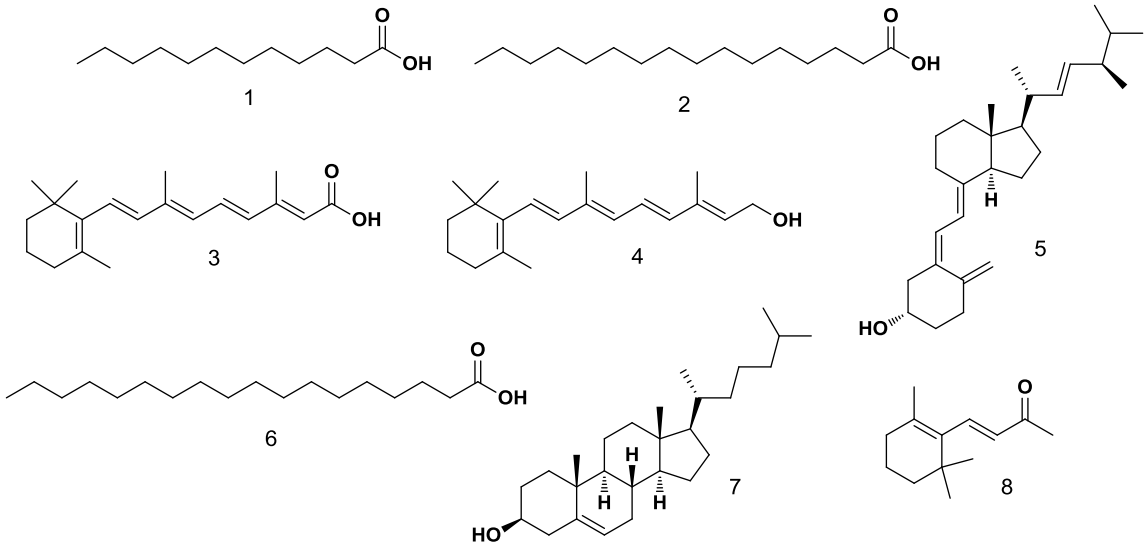
As the leading cause of vision impairment in aged populations of Western societies, AMD is poised to become a more frequent occurrence as human lifespans statistically increase. A cure for AMD is not yet available, and current treatments such as antioxidant supplements, vascular endothelial growth factor (VEGF) inhibitor injections, photodynamic therapy and laser surgery can only temporarily alleviate the symptoms of advanced stage AMD. Thus, a method to block or minimize progression of the disease in the early stage is an attractive target for limiting vision loss.

β-LACTOGLOBULIN, A PROTEIN OF UNKNOWN FUNCTION

As the principle whey protein in many mammalian milks, β-lactoglobulin (BLG) has been studied extensively over the last 70 years due to accessibility, abundance, heat and acid stability, and other desirable physical properties.[82] BLG variants are present in the milk of many mammals, including cows, cats, dogs, sheep, dolphins, water buffalo, horses, goats, and baboons, and most exist as dimers under physiological conditions. The 18.4 kDa (monomer mass) protein is a member of the lipocalin family, a diverse set of proteins known to bind hydrophilic ligands. Ligand dissociation constants measured for bovine BLG indicate appreciable binding of substrates such as retinol, palmitate, and cholesterol (**Table 3**)[83], and numerous x-ray crystal structures indicate ligand binding occurs primarily within the calyx created by the central β-barrel structure.[82] Binding data additionally indicates one molecule binds per monomer subunit. While these observations contribute to the postulated function of BLG being a lipophilic carrier protein, a definite physiological role has yet to be ascribed.[83]

Table 3 Selected ligand binding constants for bovine BLG.

Entry	Ligand	K_d (M)	Entry	Ligand	K_d (M)
1	Lauric acid	7.0×10^{-7}	5	Vitamin D_2	4.91×10^{-9}
2	Palmitate	1.0×10^{-7}	6	Stearate	1.2×10^{-7}
3	Retinoic Acid	2.0×10^{-7}	7	Cholesterol	3.49×10^{-8}
4	Retinol	1.5×10^{-7}	8	β-Ionone	6.0×10^{-7}



Studies on the BLG-directed photoisomerization of retinal and related compounds unexpectedly revealed the protein was capable of catalyzing the asymmetric self-condensation of α,β-unsaturated aldehydes to produce ring-fused homodimers (**Fig. 22**).[79] This was the first evidence showing that BLG was capable of catalytic behavior. No further studies on this mode of transformation have been reported in the literature to date.

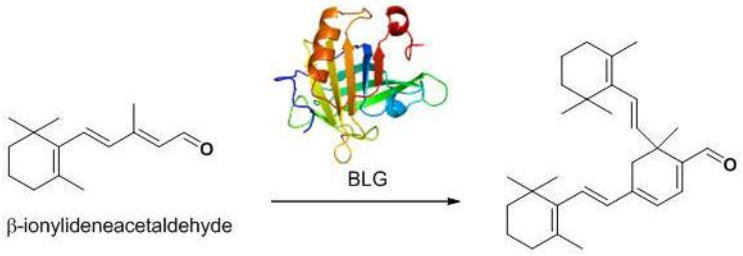


Figure 22 Cyclo-β-ional generated from β-ionylideneacetaldehyde in the presence of BLG.

Humans lack a direct homolog, but bovine BLG has been detected in blood serum at levels of 1-4 μg/L and thus must be entirely diet-derived.[84] Whey proteins are found not only in dairy products, but are also added to some breads, processed meats, and convenience snack foods to enhance texture and nutritional protein content. These foods must be avoided by some individuals because whey proteins are the major cause of human milk allergies. Highly expressed human intestinal receptors for lipocalins have been shown to bind bovine BLG for cellular uptake,[83] and as mentioned previously, BLG was among the proteins detected in drusen deposits of both healthy and AMD donor RPE cells. This evidence suggests that BLG is present in the eye and could play a role in the formation of vision cycle by-products like cycloretinal by catalyzing dimer formation and contributing to the progression of age-related macular degeneration.

STATEMENT OF PURPOSE

The introductory material of this chapter highlights not only the issues to be confronted but also illuminates the goals of our natural product biosynthetic studies. With regard to azinomycin B resistance and biosynthesis, our aim is to elucidate each step in the pathway for purposes of future engineering and development of a site-specific drug delivery system. The intended benefit of uncovering orphan biosynthetic pathways via whole genome sequencing is to broaden the catalog of bioactive natural products and provide avenues for implementing gene expression and regulation strategies. Our study of β-lactoglobulin promoted biosynthesis of α,β-unsaturated aldehyde dimers is a step toward identifying a non-genetic route to age-related macular degeneration and could contribute to methods for disease prevention and treatment. These efforts are a representation of the basic research necessary to fuel clinical approaches to therapeutics.

CHAPTER II

AZIR, A RESISTANCE PROTEIN OF THE DNA CROSS-LINKING AGENT AZINOMYCIN B

INTRODUCTION

The azinomycins (**Fig. 23**) are among a group of natural products known to act within the major groove of DNA and produce covalent interstrand cross-links. Azinomycin B was initially isolated as carzinophilin A from the fermentation broth of the soil-dwelling bacterium *Streptomyces sahachiroi,*[12] and over thirty years later was reisolated, along with azinomycin A, from *Streptomyces griseofuscus.*[13] The azinomycins were found to exhibit antitumor activity in the submicromolar range[15]. The specific activity of azinomycin towards the cell stems from interaction between DNA-reactive epoxide (C21) and aziridine (C10) functionalities with the N7 centers of purines two base pairs apart on complementary DNA strands.[16] DNA cross-linking/alkylation leads to depurination and strand breakage. As the damage escalates, cells cannot reliably replicate DNA for division and ultimately submit to apoptosis.

Biosynthesis of the azinomycins has been explored via cell-free extract systems[30] and ^{13}C-labeled precursor compound feedings.[28,29,31-33] The biosynthetic gene cluster has been proposed based on heterologous expression of an iterative type I polyketide synthase (PKS) involved in naphthoic acid production.[25] Identification of the azinomycin B gene cluster has recently been corroborated by whole genome sequencing of *S. sahachiroi*.[85] Recent studies have reconstituted all four enzymes in the pathway for 3-methoxy-5-methyl-naphthoic acid biosynthesis and activation *in vitro*, lending further support to the putative biosynthetic mechanisms stitching together the azinomycins.[34]

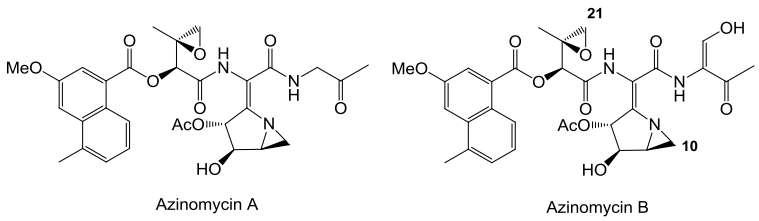


Figure 23 Azinomycin A and B structures.

From our experience manipulating *S. sahachiroi*, it is reasonable to assume a biosynthetic capacity for metabolites of unknown structure outside the azinomycin analog family. A search for NPRS or PKS biosynthetic genes involved in downregulated or silent pathways could provide information toward the discovery of novel natural products. We employed a screening strategy to search for cryptic/orphan biosynthetic pathways within a genomic library of *S. sahachiroi*. While a clear candidate for azinomycin resistance was not immediately evident from bioinformatic analysis of sequence data, a putative secondary metabolite resistance gene was chosen for further evaluation by virtue of nearby gene identities. A BLAST search indicated the resistance gene product is homologous to aminoglycoside phosphotransferases, a class of enzymes that mediate bacterial antibiotic resistance. A DNA repair protein and non-ribosomal peptide synthetase (NRPS) domains are clustered with the putative resistance protein, indicating possible association with a DNA-damaging compound produced by the biosynthetic pathway.

Bacterial resistance mechanisms are not only essential to the survival of antibiotic-producing microorganisms, but also serve purpose in bioengineering pursuits. Significant effort has been invested in studying more potent or stable synthetic azinomycin analogs for DNA alkylation.[86] Access to an analog-tolerant resistance protein, in combination with an understanding of the azinomycin biosynthetic machinery, could promote eventual engineered biosynthesis of structures previously inaccessible by synthetic methodology. This chapter explores the physical properties and protective measures the resistance protein AziR affords against DNA damage and cell death in the presence of the DNA cross-linking agent azinomycin B.

RESULTS AND DISCUSSION

Identification of a Resistance Protein

A genomic library of *S. sahachiroi* was constructed to screen for novel NRPS and PKS-type biosynthetic pathways. A fosmid-based library in an *E. coli* host was chosen to provide quick growth and regeneration time for downstream sequencing applications. Library screening was accomplished using radiolabeled degenerate hybridization probes based on either NRPS or PKS conserved regions. Positive clones were used as templates in PCR amplification of the hybridizing region to obtain fragments for DNA sequencing, confirming the presence of desired genes. Based on sequence identity, specific clones were chosen for comprehensive sequencing.

Transposon mutagenesis was employed to create a library of clones, each with randomly-inserted primer annealing sites for bidirectional sequencing.

An 801 bp ORF identified in one of the sequenced fosmids translated to a 266 amino acid sequence with homology to the aminoglycoside 3'-phosphotransferase (APH) and choline kinase (ChoK) family (cd05120, E-value 5e-9), a comparison determined using the Conserved Domain Database[87]. APH is a member of the protein kinase superfamily, whose members catalyze the transfer of γ-phosphate from ATP or CTP to the hydroxyl group of the target substrate. APH inactivates aminoglycosides, macrolides, and antibiotics upon phosphorylation[88], so a similar type of mechanism was initially suspected for the protein in this study. Protein BLAST homologs from other *Streptomyces* species showed high identity and similarity to AziR (**Table 4** and **Appendix Fig. 57**), but the homologs have not been assayed for substrate specificity or catalytic activity beyond bioinformatic annotation. A Ku-like non-homologous end-joining DNA repair protein (cd00789, E-value $6x10^{-78}$) is located 7 kb upstream and non-ribosomal peptide synthetase (NRPS) domains are 13 kb upstream, suggesting the resistance protein may be associated with a type of secondary metabolite biosynthetic pathway, the product of which may induce DNA damage.

Table 4 Top protein BLAST homologs using AziR as query.[a]

GenBank accession no.	Organism	Protein	E-value	Identity/ Similarity (%)
EDY58692	*Streptomyces sviceus* ATCC 29083	APH	4e-127	86/90
ZP_05529454	*Streptomyces viridochromogenes* DSM 40736	SvirD4_03212	8e-113	76/84
EEW71709	*Streptomyces flavogriseus* ATCC 33331	APH	4e-100	76/82
BAG20810	*Streptomyces griseus subsp. griseus* NBRC 13350	SGR_ 3981	8e-77	60/70
EFB82381	*Streptomyces sp.* ACT-1	APH	4e-73	58/69
EFE75675	*Streptomyces roseosporus* NRRL 15998	APH	8e-73	69/70

[a] Date of analysis was April 2010.

Comparison of putative active site residues in AziR with the conserved motif determined for the APH family[54] indicates AziR lacks two residues, Lys44 and Glu60, that form a salt bridge for ATP binding and coordination (**Fig. 24**). Because AziR shares similarities with resistance-

associated proteins and is proximal to NRPS domains of a putative biosynthetic gene cluster[85], the protein was cloned and overexpressed to determine if it would protect cells treated with a known DNA-damaging metabolite produced by *S. sahachiroi*, azinomycin B.

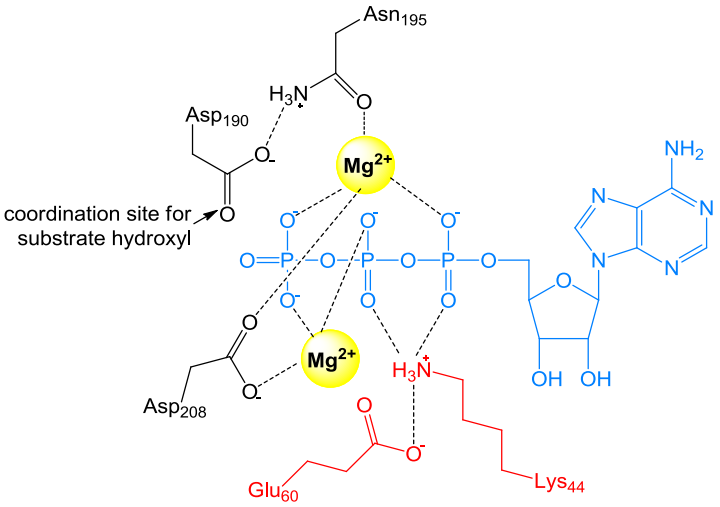


Figure 24 Active site architecture of the aminoglycoside phosphotransferase (APH) family. AziR lacks Lys44 and Glu60 (highlighted in red) The residue numbering system is derived from APH(3')-IIIa from *Enterococcus faecalis*.

In vivo Analysis of AziR Activity in the Presence of Azinomycin B

Our initial attempt to assay activity of AziR was through expression in a heterologous host, *Streptomyces lividans* TK24. The *E. coli-Streptomyces* shuttle vector pIJ86 was chosen to permit initial cloning manipulations in *E. coli* prior to introduction in *Streptomyces*. In pIJ86, expression is under control of *ermE*p*, a constitutive mutant promoter from the erythromycin biosynthetic gene cluster[89]. The vector was modified to include a ribosome binding site[90], His_{10} tag, and *att*B sites for use in Gateway system cloning.[91] The *azi*R gene was introduced to yield the final construct pIJ86G-aziR. Once transformed in *S. lividans*, expression was verified by immunoblot with His-tag monoclonal antibody. We then examined the ability of the protein to confer resistance to azinomycin B, methyl methanesulfonate (DNA alkylating agent), or streptomycin sulfate (aminoglycoside antibiotic). Relative cell density was compared to control cells harboring only pIJ86G (**Fig. 25**). Cells expressing the putative resistance protein exhibited enhanced survival only in the presence of azinomycin B, suggesting the protein serves a specific

resistance function. It was also observed that expression of the resistance protein promoted production of a red pigment indigenous to *S. lividans*.

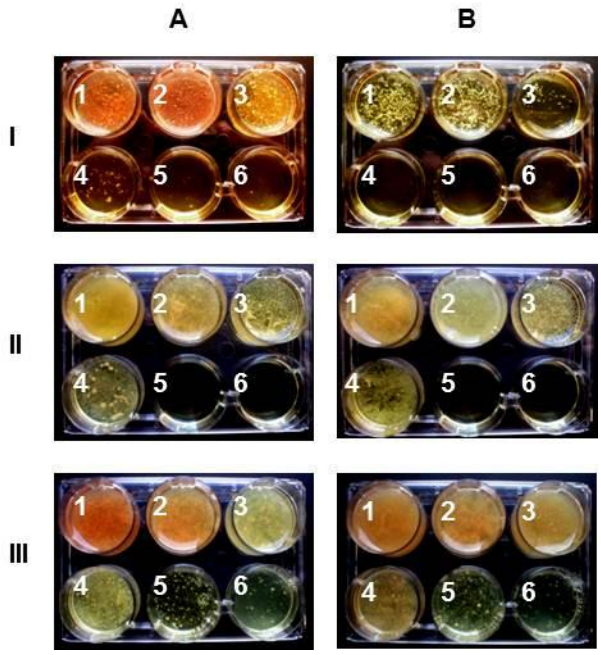


Figure 25 *In vivo* survival and resistance specificity conferred by AziR.
Column A is plasmid pIJ86G-aziR transformed in *S. lividans*. Column B is control plasmid pIJ86G transformed in *S. lividans*. Cells are treated as follows: Row I, azinomycin B at (1) 0, ethanol control; (2) 1; (3) 2.5; (4) 5; (5) 10; and (6) 15 µg/mL. Row II, methyl methanesulfonate at (1) 0; (2) 0.01; (3) 0.025; (4) 0.05; (5) 0.075 and (6) 0.1% v/v. Row III, streptomycin sulfate at (1) 0; (2) 1; (3) 2.5; (4) 5; (5) 10 and (6) 20 µg/mL.

As AziR proved difficult to purify in appreciable quantities from *S. lividans*, an *E. coli* codon-optimized version of the *aziR* gene was introduced in the expression vector pET16b. Further examination of *in vivo* AziR behavior was also carried out using this construct.

In an experiment similar to that carried out previously in our lab using yeast treated with azinomycin[17], DNA shearing effected by azinomycin B treatment in AziR-expressing *E. coli* was compared to uninduced *E. coli* (**Fig. 26**). Despite normalizing cell density prior to DNA isolation, treated cultures of the control consistently contained less total DNA, suggesting that azinomycin induced some degree of cell death and subsequent leakage of nucleic acids. Cells harboring the resistance protein contained more intact total DNA, especially evident at the 50 µg/mL treatment (**Fig. 26**, lane 6).

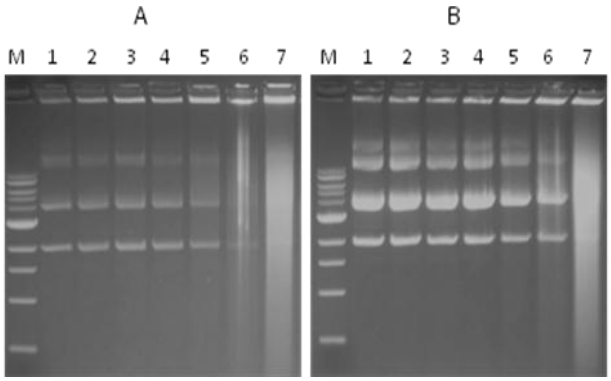


Figure 26 *In vivo* DNA shearing induced by azinomycin B.
(A) BL21(DE3)/pET16b-aziR, uninduced. (B) BL21(DE3)/pET16b-aziR, induced. Cells were treated with azinomycin B for 3 h. Lane M, 1 kb DNA ladder; lane 1, ethanol control; lane 2, 1 µg/mL azinomycin B; lane 3, 5 µg/mL azinomycin B; lane 4, 10 µg/mL azinomycin B; lane 5, 25 µg/mL azinomycin B; lane 6, 50 µg/mL azinomycin B; lane 7, 100 µg/mL azinomycin B.

To further examine the survival enhancement AziR provides during azinomycin B exposure, cells treated in a manner similar to the aforementioned experiment were washed and stained with a mixture of SYTO 9 and propidium iodide fluorescent dyes. SYTO 9 is a green cell-permeable nucleic acid binding molecule and generally stains both live and dead cells in a sample. Propidium iodide (PI) red fluorescence is maximized when it enters compromised cell membranes and intercalates within DNA. PI fluorescence will mask the SYTO 9 stain if both dyes are present within the cell, and measurement of the relative fluorescent intensities thus gives a ratio of live and dead cells within a given population. The results of a representative live/dead assay are illustrated in **Figure 27**. At all treatment levels examined, AziR provided a survival advantage.

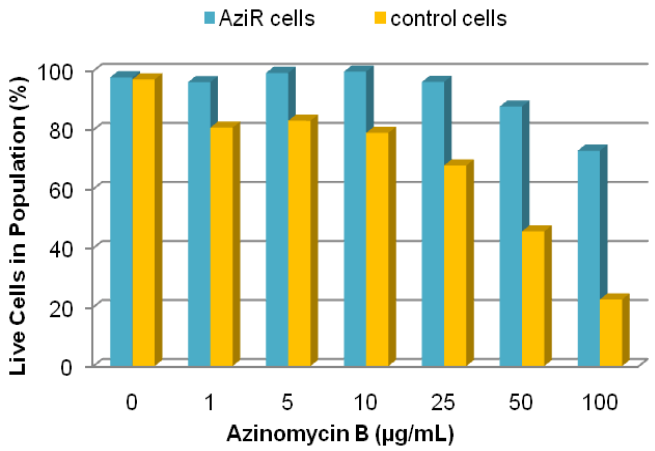


Figure 27 Comparison of percent live *E. coli* cell populations following treatment with azinomycin B.

DNA Repair Gene Transcription Changes by RT-PCR

Previous investigations indicated a strong transcriptional response related to DNA synthesis and repair proteins in the yeast genome following treatment with azinomycin B[17]. Ribonucleotide reductase (RNR1) exhibited substantial upregulation in a yeast gene chip microassay, and thus the *E. coli* ortholog of RNR1, *nrdB,* was specifically chosen with the expectation of observing a similar effect in our resistance protein *E. coli* expression system. Semi-quantitative reverse-transcriptase (RT)-PCR was employed to monitor transcription of *nrdB* relative to unaffected expression of the metabolic protein glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in treated cells. The purpose of the semi-quantitative approach is simply to identify whether or not a transcriptional effect exists in the system. Cells with or without AziR were treated with azinomycin B prior to extraction of total RNA. Gene-specific oligonucleotides were used to generate the cDNA subsequently used as a template in PCR analysis (**Fig. 28**). PCR products were gel purified, extracted, and stained with SYBR green to permit relative quantitation of products by fluorescence spectroscopy. As anticipated, *nrdB* transcription increased in response to azinomycin treatment but remained virtually unchanged when AziR was expressed by the system.

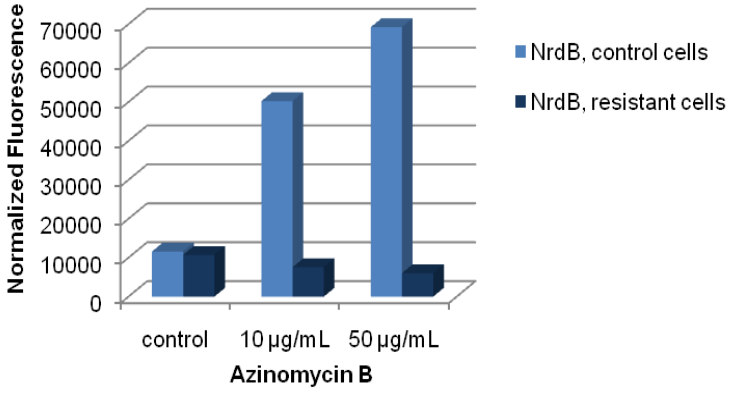


Figure 28 Semi-quantitative RT-PCR analysis of nrdB in AziR-expressing cells and control cells. Cells were treated with azinomycin B for 2.5 h prior to RNA isolation.

Phosphotransferase Activity Assays

To further assess the ability of AziR to function as a kinase despite the lack of two key active site residues, a standard assay coupling ADP release to pyruvate kinase/lactate dehydrogenase was performed with azinomycin B and also with the aminoglycoside antibiotics kanamycin, streptomycin, and neomycin.[92] AziR did not exhibit detectable phosphorylation activity towards the tested substrates. Additionally, *in vitro* incubation of AziR with azinomycin B and radiolabeled ATP or acetyl-CoA failed to indicate substrate phosphorylation or acetylation was facilitated by the protein (**Appendix Fig. 60**). Thus, it was determined that AziR does not possess the phosphotransferase activity toward azinomycin that is suggested by sequence homology.

AziR/Azinomycin B Interactions *in vitro*

AziR was expressed in an *E. coli* host, purified by immobilized metal affinity chromatography, and assayed for azinomycin B or ATP binding via equilibrium fluorescence titration (**Fig. 29A**). AziR has seven tryptophan residues, and under 295 nm excitation wavelength the protein exhibits maximum fluorescence emission at 330 nm. The dissociation constant (K_d) is 230 nM for azinomycin B, determined by nonlinear regression using a quadratic equation to express dissociation under constant enzyme concentration.[93] ATP binds to AziR with a dissociation constant of 185 µM, representing over two-fold less substrate affinity than that of aminoglycoside phosphotransferase (3')-IIIa for ATP[94] (**Fig. 29B**). Azinomycin B was

not observed to bind to a control, 2-deoxyribose 5-phosphate aldolase (DERA), a purified *E. coli* protein not involved in resistance mechanisms (**Appendix Fig**. **61**).

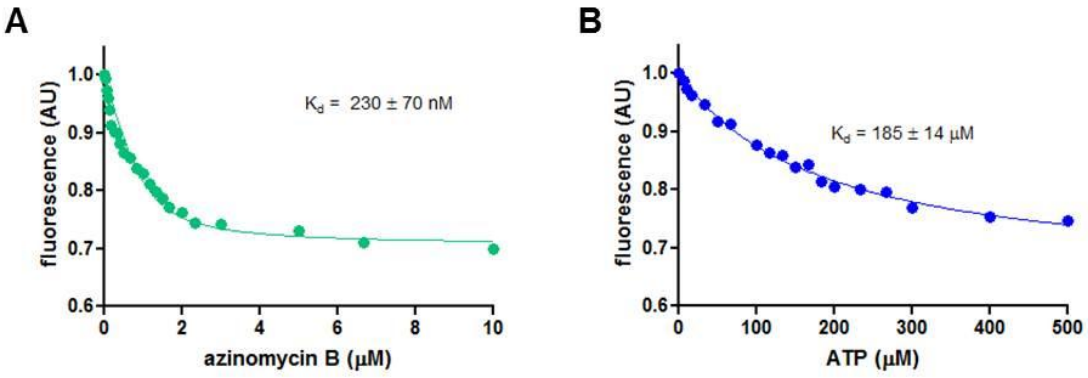


Figure 29 Equilibrium fluorescence titration curves for AziR substrate binding.
AziR titrated with (A) azinomycin B or (B) ATP. Binding constants are noted above the respective curves.

SIGNIFICANCE

AziR has been shown to exhibit a protective effect when expressed in cells exposed to azinomycin B. In cells treated with azinomycin B, AziR expression contributes to robust cell growth and suppresses DNA-targeting behavior. *In vitro*, the protein binds azinomycin B with appreciable affinity, suggesting a degree of specificity for the drug in the producing strain *S. sahachiroi*. While AziR does exhibit ATP binding, the decreased affinity relative to that displayed by a true aminoglycoside phosphotransferase may be attributed to the two missing residues deemed necessary for ATP coordination. Under all assayed circumstances, AziR failed to demonstrate phosphoryl transfer.

The proximity of AziR to the published azinomycin B biosynthetic cluster is unknown due to gaps in the draft genome of *S. sahachiroi*, and further manual sequence assembly is currently under way to address the ambiguity. Additional proteins with hypothetical resistance function have been identified through genomic sequencing and will be tested for specificity toward the azinomycins in due course. A probable protein partner in the resistance mechanism may include a membrane-bound transport protein of the major facilitator superfamily (cd06174, E-value 1.3×10^{-9}) encoded by a gene immediately upstream of *aziR*. Efflux pumps for metabolite secretion are common in bacterial cells to transport antibiotics across the cell

membrane and into the surrounding environment for protection against competing colonization.[52] A transporter is likely necessary to support translocation of azinomycin out of the producing cell.

Knowledge of a resistance mechanism directs the genetic engineering of drug analogs. An analog must be recognized by a resistance protein to protect the producing microorganism, and for the associated engineered gene cluster to be of biosynthetic use. Resistance mechanisms are also important to study because they give us a better understanding of how bacterial drug resistance can evolve. Molecular modifications such as phosphorylation,[95] acetylation,[96] degradation,[97] and oxidation[98] are among a few of the techniques employed by the "resistome."[52] Inactivation mechanisms can also lean toward more complex processes involving multiple biosynthetic enzymes such as that observed for vancomycin resistance.[99] Alternatively, resistance can also be as simple as drug binding and sequestration as evidenced by bleomycin or mitomycin resistance proteins.[100,101] Because substrate modification was not detected by available assays, an analogous binding mechanism appears to be at work in the azinomycin resistance conferred by AziR. The resistance proteins for bleomycin and mitomycin are also believed to be involved in transport/delivery of the drug to the cell membrane for removal, which is an added function for which AziR has not yet been tested.

EXPERIMENTAL PROCEDURES

Instrumentation and General Methods

DNA sequencing was performed using ABI BigDye chemistry at the Gene Technologies Lab, Texas A&M University. All commercially-available chemicals were obtained from Sigma-Aldrich. Cell density measurements were recorded in 96-well plate format on a BioTek μQuant microplate spectrophotometer, and fluorescence readings were determined using a BioTek FL800 fluorescence microplate reader. Protein fluorescence titrations were performed on a PTI QuantaMaster 4 spectrofluorometer. Protein concentrations were determined by the Bradford method[102] employing bovine serum albumin as an absorbance standard in the linear range of 0.05 to 0.5 μg/mL. Azinomycin B was produced using the fermentation and purification methods described elsewhere.[31] Strains, plasmids, and primer sequences used in this study are listed in **Table 5**. Maps of new plasmids constructed for this study are included in **Appendix Fig**. **58**.

Table 5 Strains, plasmids, and primers used in the AziR study.

Strain, plasmid, or primer	Relevant characteristic(s) or primer sequence (5'-3')	Reference or source
Strains		
S. sahachiroi	Wild type azinomycin A and B-producing strain	ATCC
S. lividans TK24	Wild type strain used for heterologous expression	ATCC
E. coli DH10B	General cloning host	Epicentre
E. coli BL21(DE3)	Host strain for recombinant protein expression under T7 promoter	Invitrogen
E. coli EPI300-T1^R	Host strain for fosmid library	Epicentre
E. coli ET12567/pUZ8002	DNA methylation-deficient strain	Paget et al[103]
Plasmids		
pCC1FOS	Vector for fosmid library construction in E. coli	Epicentre
pET-16b	T7 expression vector; encodes C-terminal His_6 tag	Novagen
pET16b-aziR	The codon-optimized *aziR* gene (801 bp) was inserted at *Nde*I and *BamH*I sites	This study
pIJ86	*Streptomyces-E. coli* shuttle vector with ermEp* constitutive promoter	Mervyn Bibb
pIJ86-RSH	*Streptomyces-E. coli* shuttle vector with ribosome binding site and C-terminal His_{10} tag	This study
pIJ86G	*Streptomyces-E. coli* shuttle vector with ribosome binding site, C-terminal His_{10} tag, and Gateway reading frame A cassette	This study
pDONR-zeo	Gateway system cloning entry vector	Invitrogen
pDONR-aziR	The *aziR* gene (801 bp) flanked by attB1 and attB2 sites was inserted using BP recombinase	This study
pIJ86G-aziR	The *aziR* gene from pDONR-aziR was inserted using LR recombinase	This study
Primers		
NRPS-A8F	TTCCGGTTCGAGCYSGGBGAGATCGA	This study
NRPS-TR	GTGVCCVCCSAGGTCGAAGAA	This study
PKS-2F	CCSCAGSAGCGCSTSCTSCTSGA	This study
PKS-2R	GTSCCSGTSCCGTGCGCCTCSA	This study
MUKAN-1 FP-1	CTGGTCCACCTACACAAAGG	This study
MUKAN-1 RP-1	AGAGATTTTGAGACAGGATCCG	This study
RSH-L	TTTTAAGCTTAGGAGGCACAGTCATGAGGC	This study
RSH-C	CACAGTCATGAGGCCTCACCACCATCACCATCACCACCAC CATC	This study
RSH-CRC	GATGGTGGTGGTGATGGTGATGGTGGTGAGGCCTCATGA CTGTG	This study
RSH-R	CACCACCACCATCACTGAAAGCTTCCCC	This study
attB1-aziR F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGACCGC GGAAGATGAAGCGC	This study
attB2-aziR R	GGGGACCACTTTGTACAAGAAAGCTGGGTCGTTCAGCAG CGCCGCGGTAAAC	This study
GAPDH-F	AAAGGCGCTAACTTCGACAA	This study
GAPDH-R	GCAGCTTTTTCCAGACGAAC	This study
NRDB-F	GCCAGACCAAAGATGTCGAT	This study
NRDB-R	CGAAGGCGAAACTCTTCAAC	This study

Construction of an *S. sahachiroi* Genomic Library

A genomic library of *S. sahachiroi* was generated using the CopyControl Fosmid Library Production Kit (Epicentre, Madison, WI) with the following modifications. *S. sahachiroi* was cultured in 250 mL YEME media including 0.5% glycine with shaking at 250 rpm, 28°C for 2 days. High molecular weight genomic DNA was isolated using the salting out procedure,[104] randomly sheared by pipet aspiration, and loaded onto a CHEF gel for size selection (2 s initial, 8 s final switch time, 12 h, 14°C, 6 V). Gel-embedded DNA fragments in the 25-75 kb size range were excised, placed in dialysis tubing with TAE buffer, and electroeluted (40 s switch time, 5 h, 14°C, 6V) in the CHEF apparatus. Buffer inside the dialysis tubing was collected and the DNA precipitated with 70% cold isopropanol. Following end repair, the DNA was ligated to the vector pCC1FOS, packaged in lambda phage, and used to transfect EPI300 *E. coli* on LB plates containing 35 μg/mL chloramphenicol. A library of 2199 clones with an average insert size of 40 kb was arrayed in 384-well plates.

***S. sahachiroi* Genomic Library Screening**

The *S. sahachiroi* fosmid library was screened for desired sequences by the standard method of hybridization with radioactive probes. ^{32}P-end-labeled degenerate oligonucleotide probes NRPS-A8F, NRPS-TR, PKS-2F, and PKS-2R were based on NRPS and PKS domain sequences taken from NCBI databases. Library colonies were cross-linked to a nitrocellulose membrane, incubated with the probe, and hybridized clones identified by autoradiography. A set of 40 clones screened NRPS positive, and 13 clones of that set also screened PKS positive. Degenerate primer PCR and sequencing of the product confirmed the hybridization as a true NRPS or PKS sequence. PCR product sequences were analyzed by the NCBI basic local alignment search tool (BLAST)[105] and grouped by result. Fosmids 2 and 36 were chosen for comprehensive sequencing based on similarity to the *S. verticillus* bleomycin and the *S. chrysomallus* actinomycin biosynthetic clusters containing a hybrid NRPS/PKS and an NRPS system, respectively.

Fosmid Sequencing and Bioinformatic Analysis

Fosmids were sequenced using randomly introduced primer binding sites generated by the HyperMu <KAN-1> Insertion Kit (Epicentre). Clones were miniprepped and sequenced using the primers MUKAN-1 FP-1 and MUKAN-1 RP-1. Sequences were assembled using

Sequencher 4.8 (Gene Codes, Ann Arbor, MI) and open reading frames (ORFs) were defined with the aid of FramePlot 3.0beta[106] and annotated using NCBI databases.[107]

Cloning and Heterologus Expression of AziR (*S. lividans)*

The *E. coli-Streptomyces* shuttle vector pIJ86 was modified to permit use of the Gateway cloning system (Invitrogen, Carlsbad, CA) as follows. A 75 bp DNA fragment containing a ribosome binding site, start codon, *Stu*I restriction site, His_{10} tag and stop codon was constructed using a set of four oligos RSH-L, RSH-C, RSH-CRC, and RSH-R. Oligo RSH-C and the reverse complement RSH-CRC were annealed to form a double-stranded template. Oligos RSH-L and RSH-R were used as PCR primers to extend the template, resulting in the 75 bp fragment. pIJ86 was digested with *Hind*III, dephosphorylated with calf alkaline phosphatase, and ligated to the *Hind*III flanked-75 bp fragment to give plasmid pIJ86-RSH. The Gateway reading frame cassette A was introduced at the *Stu*I site of pIJ86-RSH to give plasmid pIJ86G. *aziR* was amplified by PCR with primers aziR-attB1 and aziR-attB2 to add attB1 and attB recognition sites. The aziR gene was cloned in the Gateway vector pDONR/Zeo using BP recombination, and introduced to pIJ86G via LR recombination to yield plasmid pIJ86G-aziR. The plasmid was passed through the DNA methylation-deficient *E. coli* strain ET12567 and transformed in *S. lividans* TK24 protoplasts using the rapid small-scale procedure.[104] Transformants were selected with 40 μg/mL apramycin at 28°C. Apramycin-resistant transformants were transferred to a fresh GYM agar plate for propagation. A small piece (1x1 cm) was cut from the propagated plate and used to inoculate 25 mL R2YE media containing 30μg/mL apramycin and glass beads (~3g, 3 mm diameter). The starter culture was incubated at 28°C with shaking at 250 rpm. After 48 h, the starter culture was used to inoculate 500 mL of R2YE media containing glass beads (~15 g) and incubated an additional 48 h to achieve a dense culture. The spores were collected by centrifugation at 7000 rpm, 20 min and resuspended in 40 mL phosphate buffer (20 mM NaH_2PO_4, 300 mM NaCl, 10% glycerol, 0.1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride). Spores were lysed on ice using a bead mill (BeadBeater, BioSpec Products Inc., Bartlesville, OK) containing ~25 g of 0.1 mm glass beads. The mill was operated for 10-30 sec intervals with 2 min rests for cooling. Following lysis, the lysate was centrifuged at 9800 rpm for 30 min to pellet cellular debris. AziR expression was detected by immunoblotting using a His tag monoclonal antibody (Novagen, Gibbstown, NJ).

Cloning, Overexpression, and Purification of AziR (*E. coli)*

An *E. coli* codon-optimized version of the *aziR* gene was synthesized and cloned in the vector pET-16b at *Nde*I and *BamH*I restriction sites by GenScript (Piscataway, NJ). The vector pET16b-aziR was transformed in BL21(DE3) cells for inducible expression under a T7 promoter. A 30 mL overnight culture in LB Miller media and 100 µg/mL ampicillin was used to inoculate 1 L media and grown to 0.6 OD_{600}. The culture was induced with 1 mM IPTG and transferred to 16°C for 20 h to achieve optimal soluble expression. All remaining steps were performed on ice. Cells were collected by centrifugation, resuspended in 30 mL column binding buffer (20 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 5 mM imidazole, 10% glycerol, 0.1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride), and lysed by sonication with a Branson Sonifier 450 (Branson Ultrasonics, Danbury, CT) fitted with a 5 mm microtip, output setting 6, duty cycle 50%, for 8 cycles of 30 sec each. Cellular debris was pelleted by centrifugation, and the supernatant filtered with a 0.2 µm filter before applying to a pre-equilibrated HisTrap FF 5 mL column (GE Healthcare Life Sciences, Piscataway, NJ) at 2.5 mL/min. The column was washed with 100 mL column buffer containing 20 mM imidazole, and AziR was eluted with 30 mL column buffer containing 500 mM imidazole. The eluate was dialyzed by centrifugal ultrafiltration (Amicon Ultra-15 Centrifugal Filter Unit, Millipore, Billerica, MA) to a final concentration of 1 mg/mL in 100 mM NaH_2PO_4 pH 7.5, 10% glycerol, 0.1 mM dithiothreitol, and 10 mM $MgCl_2$. Representative gels of purified AziR are included in **Appendix Fig. 62.**

To determine the multimeric form of AziR, the protein was exchanged into GFC buffer (50 mM Tris pH 7.5, 150 mM NaCl) and applied to a pre-equilibrated Superdex 200 column (GE Healthcare) at 0.5 mL/min. The column was previously calibrated with apoferritin (440 kDa), IgG (160 kDa), BSA (66 kDa), ovalbumin (43 kDa), lactoglobulin (35 kDa), myoglobin (16.7 kD), and cytochrome C (12.4 kDa). AziR eluted with an apparent molecular mass of 30 kDa (calculated 31 kDa), indicating a monomeric quaternary structure.

***In vivo* Characterization of AziR Activity (*S. lividans*)**

A starter culture of 10 mL R2YE, 30 µg/mL apramycin, and 2 g glass beads was inoculated with a 1x1 cm slice of an R2YE agar plate containing *S. lividans*(pIJ86G-aziR) spores. After growing 3 days at 28°C with shaking, 30 µL aliquots of the dense suspension were transferred to 10 mL fresh R2YE media for treatment with azinomycin B (in ethanol), methyl methanesulfonate, or streptomycin sulfate. Treated cultures were allowed to incubate in a 28°C

shaker for 96 h. The entire culture was then transferred to a 6-well plate for density comparisons.

In vivo Characterization of AziR Activity (*E. coli*)

A flask containing 100 mL LB media was inoculated with an overnight culture of BL21(DE3)/pET16b-aziR and grown to 0.6 OD_{600} at 37°C. A 50 mL aliquot was removed to a clean, sterile flask for induction with 1 mM IPTG. Both cultures were then transferred to a 16°C shaker for 16 h. Following induction, 5 mL portions were transferred to culture tubes for treatment with azinomycin B and an additional incubation period of 24 h at 16°C. The optical density of each culture was used to normalize the cell concentration relative to untreated controls, and total DNA was isolated and electrophoresed on a 1% agarose gel at 100 V for 70 min.

Cell Viability Following Azinomycin B Treatment

A 2 mL overnight starter culture of BL21(DE3) cells harboring pET16b-aziR was used to inoculate two 100 mL flasks of LB Miller media at 37°C for 2.5 h. One of the flasks was induced with 1 mM IPTG and growth continued for 1 h, at which time 3 mL aliquots of culture were divided into sterile culture tubes and treated with azinomycin B dissolved in absolute ethanol (5 mg/mL). After 3 h incubation at 37°C, optical density measurements at 600 nm were taken to normalize the cell concentration among the cultures. Cells from 1 mL of each normalized culture were pelleted and washed twice with 1 mL 0.85% NaCl, then resuspended in an additional 1 mL of solution. Cells were further diluted to 0.06 OD_{670} and assayed with the LIVE/DEAD BacLight Bacterial Viability Kit (Molecular Probes, Invitrogen) in triplicate. Optimal staining was achieved with a 1:3 ratio of SYTO 9 dye to propidium iodide. Stained cells in a 96-well plate were measured with a BioTek FL800 fluorescence microplate reader (filters: 485/30 for excitation, 528/20 and 645/40 for green and red emission, respectively).

RT-PCR Evaluation of DNA Repair Gene Transcription

Relative endpoint RT-PCR employed the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as an amplification standard to normalize mRNA levels. Ribonucleotide reductase (subunit B, nrdB) was chosen to represent the relative DNA damage/repair response. Two 50 mL cultures of pET16b-AziR(BL21(DE3)) were inoculated from an overnight starter culture and incubated at 37°C for 1.5 h. One of the cultures was

induced with 1 mM IPTG at 37°C for 1 h. Aliquots (3 mL) of culture were transferred to sterile tubes for treatment with azinomycin B at 37°C for 2.5 h. Total RNA was isolated using the RNeasy Mini Kit (Qiagen, Valencia, CA). cDNA template was produced from DNase I-digested total RNA using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA) and the forward-oriented gene-specific primer for GAPDH or nrdB. The cDNA served as a template in PCR amplification of a 359 bp fragment of *E. coli* BL21 GAPDH using primers GAPDH-F and GAPDH-R or a 333 bp fragment of *E. coli* BL21 *nrdB* using primers NRDB-F and NRDB-R. Thermal cycling was optimized so that each gene being amplified was in exponential phase and consisted of the conditions: 2 min at 96°C, followed by 20 iterative cycles of 30 sec at 96°C, 30 sec at 50°C, and 30 sec at 72°C. The PCR reaction mixtures were separated on a 1.2% agarose gel and the desired bands extracted from the agarose using a QIAquick Gel Extraction Kit (Qiagen). DNA was stained with SYBR Green (Invitrogen) in a 96-well plate and measured on a fluorescence microplate reader (filters: 485/30 for excitation and 528/20 for emission).

Aminoglycoside Phosphotransferase Activity Assay

Purified AziR was subjected to assay with kanamycin, streptomycin, neomycin or azinomycin B as described by McKay and coworkers.[92] Substrate phosphorylation was to be detected indirectly by coupling the release of ADP to the enzymatic reactions of pyruvate kinase and lactic acid dehydrogenase (PK/LDH), resulting in oxidation of NADH. Reaction progress was monitored at 340 nm on a Genesys 2 UV-Vis Spectrophotometer (ThermoFisher Scientific, Waltham, MA). Each reaction consisted of 885 μL buffer (50 mM Tris pH 7.5, 40 mM KCl, 10 mM $MgCl_2$, 0.5 g/mL NADH, 2.5 mM phosphoenolpyruvate, 1 mM ATP) to which 10 μL of a 10 mM substrate stock was added, along with 5 μL of PK/LDH solution as obtained from the supplier (Sigma P0294, PK/LDH from rabbit muscle). After incubation for 20 min at 37°C, AziR (100 μg) was quickly added to the reaction mixture just prior to spectroscopic monitoring. For all substrates tested the rate of NADH oxidation was essentially zero, indicating a phosphorylation reaction was not facilitated by AziR.

Determination of Ligand/AziR Binding Constants

The dissociation constant (K_d) for the binding of azinomycin B to AziR was determined by steady state fluorescence titration using a PTI QuantaMaster 4 Spectrofluorometer (Photon Technology International, Birmingham, NJ) in the Materials Characterization Facility, Texas

A&M University. An excitation wavelength of 295 nm was chosen such that the focus was on fluorescence changes attributed to tryptophan residues. Slit widths were set to 5 nm and spectra were autocorrected by the FeliX32 software package to account for variations in arc lamp intensity. A solution of 1 μM protein (20 mM NaH_2PO_4 pH 7.5, 50 mM NaCl, 5% glycerol, 1 mM dithiothreitol, 10 mM $MgCl_2$) was titrated with aliquots of 100 μM azinomycin B or 100 mM ATP, each prepared in the same phosphate buffer. The change in protein fluorescence at 330 nm was measured as a function of substrate concentration. Observed fluorescence was corrected (F_{corr}) for dilution using the following equation:

$$F_{corr} = F_{obs}\frac{V_0 + \mathrm{d}V}{V_0}$$

where F_{obs} is the raw observed fluorescence intensity, V_0 is the original sample volume, and dV is the change in sample volume. Data was evaluated for a potential inner filter effect due to ligand absorbance but further correction was found to be unnecessary. Each data point was taken after mixing and an equilibration time of 1 min. Fluorescence data was normalized to 1 for zero ligand binding. Nonlinear regression using the least-squares method was performed in the program GraphPad Prism 5 (GraphPad Software, La Jolla, CA) using the quadratic equation:

$$F_{obs,corr} = F_0 + \Delta F \cdot \frac{P_0 + L_0 + K_d - \sqrt{(P_0 + L_0 + K_d)^2 - 4 \cdot P_0 \cdot L_0}}{2 \cdot P_0}$$

where $F_{obs,corr}$ is the corrected observed fluorescence, F_0 is the y-intercept, ΔF is the overall change in fluorescence, P_0 is the protein concentration, and L_0 is the ligand concentration added to determine the binding constant K_d.[108] A control titration for a non-specific protein (*E. coli* 2-deoxyribose 5-phosphate aldolase, DERA) titrated with azinomycin B is included in **Appendix Figure 61**. Cloning, overexpression and purification of DERA is outlined elsewhere.[109]

CHAPTER III

ORPHAN BIOSYNTHETIC CLUSTERS REVEALED IN THE DRAFT GENOME OF *STREPTOMYCES SAHACHIROI*

INTRODUCTION

The terrestrial bacterium *Streptomyces sahachiroi* harkens from a genus renowned for its proclivity to produce structurally interesting and medicinally valuable natural products. Streptomycetes and near phylogenetic neighbors are a collective gold mine for production of compounds with antibacterial, immunosuppressive, antitumor, and herbicidal properties.[41] *S. sahachiroi* is currently best known for production of the DNA crosslinking agents azinomycin A and B and the structurally related metabolite epoxyamide (**Fig. 30**). The azinomycins have been well-studied from a biomedical and synthetic standpoint,[86] and are just beginning to be understood from a biosynthetic perspective.[25,34]

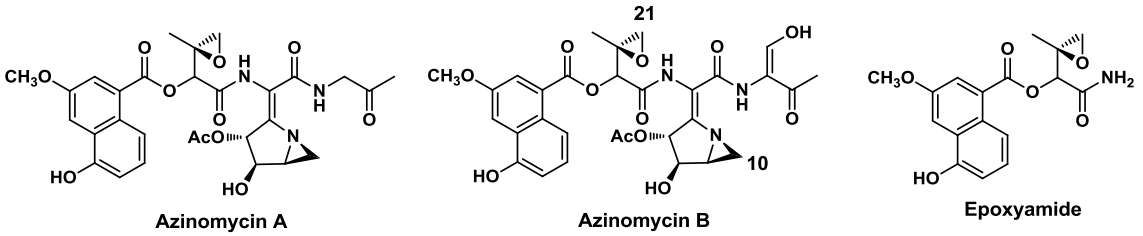


Figure 30 Azinomycin A and B and the related metabolite epoxyamide.

Our group has extensive experience with *S. sahachiroi* manipulation acquired in the course of our investigations of azinomycin A and B biosynthesis,[30-33] and those studies naturally evolved to include a curiosity toward additional latent natural products. Evidence of declining efficiency in natural product discovery plagues traditional fermentation, fractionation, and bioassay approaches. Cognizance of these difficulties was especially evident to us in our *S. sahachiroi* fermentation optimization efforts, as at times production of known metabolites like the azinomycins was difficult to induce and control, and would translationally prove impossible when coaxing biosynthesis of compounds for which immediate structural information was unknown.

The reverse approach to identification of new natural products involves genomics and biochemistry rather than end result-oriented fermentation and purification chemistry. The advent of DNA sequencing technologies has the potential to dramatically change natural product drug discovery by revealing genetically-encoded pathways for products that may be below the limit of detection or entirely absent from a microorganism's metabolic profile. Instead of being restricted to examining available compounds, we can peek at the overlooked potential within genomes to produce natural products via pathways that may be downregulated or silenced by one or more factors. Since whole-genome sequencing technologies were put into practice, proposals to explore the realm of cryptic/orphan secondary metabolic pathways, especially in microorganisms with a reputation for producing medically useful compounds, are creeping in to supplement the unending push for new antibiotics.[110,111]

This study details our efforts to sequence and annotate the draft genome of *S. sahachiroi* using a combination of traditional genomic library screening implementations and next-generation high-throughput methodologies with the aim of natural product biosynthetic pathway mining. The undertaking has revealed at least five unknown biosynthetic pathways in *S. sahachiroi* containing NRPS or PKS-type gene sequences with the potential to produce novel bioactive secondary metabolites.

RESULTS AND DISCUSSION

Genomic Library Screening for Secondary Metabolic Enzymes

Degenerate probe-based screening of a fosmid-based *S. sahachiroi* genomic library served as an initial method to identify potential genes of secondary biosynthetic pathways. Armed with structural knowledge of at least one hybrid nonribosomal peptide-polyketide compound produced by the strain, the azinomycins, the search was focused on nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) enzyme classes. Aside from structural precedence indicated by the azinomycins, modular NRPSs and PKSs are ubiquitous in microorganismal secondary metabolism and exhibit a high degree of sequence conservation among specific catalytic domains, making them a suitable target for detection with degenerate DNA hybridization oligomers. A 2,199-member fosmid library in an *E. coli* host served as a repository for 35 kb DNA fragments representing the contents of the *S. sahachiroi* genome. Separate degenerate probe sets for NRPS and PKS gene sequences were constructed containing the sequences in **Table 6**.

Table 6 Sequences of degenerate NRPS and PKS hybridization probes.

Probe name	Sequence (5'-3')	Fold-degeneracy
NRPS-A8F	TTCCGGTTCGAGCYSGGBGAGATCGA	12
NRPS-TR	GTGVCCVCCSAGGTCGAAGAA	18
PKS-2F	CCSCAGSAGCGCSTSCTSCTSGA	64
PKS-2R	GTSCCSGTSCCGTGCGCCTCSA	16

The arrayed library was immobilized on filter paper and probed with ^{32}P-end labeled degenerate oligomers. Hybridization screening with the respective probe sets revealed 40 clones containing putative NRPS-type sequences, and of that set, 13 clones also contained PKS-type sequences. The degenerate primers were then used for sequencing each clone to determine if any sequences overlapped. Based on sequence homologies determined in BLAST searches, two of the clones (fosmid2 and fosmid36) were selected for comprehensive sequencing using a random library of transposon-delivered primer binding sites.

Genomic library screening with degenerate probes has inherent limitations tied to the heterologous host. Certain sequences may be toxic and will fail to propagate in the host for library inclusion. Other sequences may not be represented, as libraries must balance statistical sequence coverage with the number of members reasonable to store and screen. With the evolution of whole-genome sequencing platforms occurring at the time of this study, we found it to be both cost-effective and more efficient to pursue genomic sequencing as an alternative to the library screening and Sanger sequencing approach.

Next-generation Sequencing Platform Comparison

We first attempted genomic sequencing of *S. sahachiroi* using the mid-range read capabilities of the Roche/454 GS FLX pyrosequencing platform. Because coverage and assembly completion were less than optimal using this approach, we also opted to try very short read-based Illumina sequencing. Each genomic sequencing run generates a basic set of statistics quantifying total base calls and reads, and following contig assembly, additional characteristics such as median contig length and contig span can be calculated. **Table 7** presents statistics to compare relative efficiencies of 454, Illumina single-read (SR) and paired-end (PE) sequencing platforms encountered in this study.

Table 7 Statistics for *S. sahachiroi* whole genome sequencing approaches.

	454	**Illumina SR**	**Illumina SR + 454**	**Illumina PE**
Total Base Calls	147 million	360 million	500 million	1.95 billion
Total Reads	590,000	9.5 million	9.8 million	27.9 million
Read Length	100 bp	36 bp	36 bp	36 + 36 bp
Contigs	5,712	4,796	4,159	8,416
N50 Contig Size	649 bp	3,358 bp	4,280 bp	9,592 bp
Largest Contigs	30, 29, 26 kb	33, 23, 19 kb	56, 23, 23 kb	63, 56, 54 kb
Contigs (>1 kb)	2,055	2,521	2,195	1,859

Illumina sequencing runs generated far larger median contig sizes in fewer total contigs than the 454 pyrosequencing experiment could provide. Combining Illumina single-read sequencing with the data obtained from 454 pyrosequencing did not serve to greatly enhance contig statistics compared to the Illumina data on its own. Even when cost was factored in, we obtained far better results using Illumina sequencing for our *Streptomyces* genome. A compilation of sequence data from Sanger, 454, Illumina SR and Illumina PE runs was used to produce the draft genome of *S. sahachiroi* referenced in subsequent sections of this chapter, a practice with precedence in studies in which *de novo* genome sequencing is practiced.[112] While Illumina sequencing can provide the bulk of sequence data with superior coverage and virtually complete sequence representation, other technologies are invaluable for evaluating sequence quality and accuracy. The shortcomings of Illumina reads can be complemented by longer 454 reads for scaffolding purposes in *de novo* genomes, and genome finishing almost always requires some extent of manual Sanger sequencing to close gaps and resolve multiple repeat regions. Unfortunately, combining these technologies is also incredibly time consuming at this point because computational algorithms cannot yet seamlessly integrate read data, quality scores, and error models from both long and short-read implementations. Coupling the lack of a reference genome to the combined assembly obstacle results in cases such as the one we have encountered. The data has produced a fragmented draft genome that will require either more efficient assemblers or additional sequencing efforts to arrive at an acceptable state of completion for comparative genomic purposes.

Automated Annotation of Sequence Data

Fortunately, genome completion is not necessary for biosynthetic pathway mining. Provided a draft genome agrees with certain parameters, automated annotation of multiple contigs is available in technologies such as batch BLAST inquiries or the SEED Rapid Annotation through Subsystems Technology (RAST) server.[113] In the latter option, searches for favorite proteins by name or enzyme class identifier in the automated output are as simple as keyword input in a spreadsheet document. However, the degree of genome fragmentation may result in identification of interesting proteins that lack full biosynthetic pathway organizational context. For these cases, areas of absent sequence data can be easily acquired from chromosome walking or genomic library screening using specific (rather than degenerate) probes to capture gene clusters for Sanger sequencing. In this study, we enlisted the aid of the SEED RAST server to provide searchable annotation of *S. sahachiroi* contigs of sizes in excess of 5 kb. The resulting data included over 5,800 genes ranging from primary metabolism to cellular regulation and propagation to the target of our study, genes of orphan secondary biosynthetic pathways (**Fig. 31**).

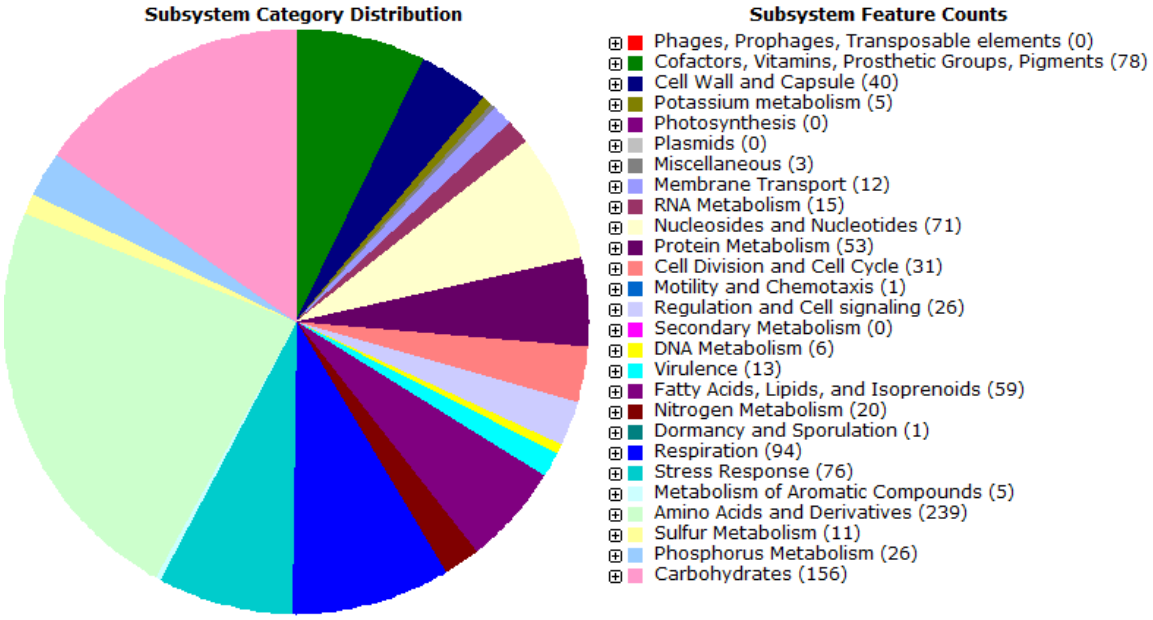


Figure 31 RAST-derived distribution of protein classes in the *S. sahachiroi* draft genome. Screenshot taken from RAST server output.

The Orphan Clusters in *Streptomyces sahachiroi*

Three presumably complete secondary metabolic pathways other than the reported azinomycin B biosynthetic gene cluster were revealed during sequence analysis, and additional partial clusters or genes characteristic of natural product pathways were also found. Each cluster was given the arbitrary designation sahA, sahB, sahC, and so on, as the cognate natural product is unknown. With the aid of specialized NRPS/PKS prediction databases and current knowledge of known biosynthetic pathways, one can propose the type of product expected from each cluster.

The sahA pathway was originally identified in the genomic library clone fosmid2 through Sanger sequencing and was extended by genomic sequencing data. Cluster sahA is an NRPS-type biosynthetic pathway with three apparent modules for incorporation of threonine, asparagine, and another unpredicted substrate (**Fig. 32** and **Table 8**) dependent upon a type II thioesterase for covalent cleavage. Tailoring enzymes include an *N*-methyltransferase domain (within *sah*A1), two oxygenases (*sah*A2 and *sah*A7), a carboxylate-amine ligase (*sah*A11), and various cofactor-dependent reductases. This gene cluster is immediately upstream of the resistance protein (AziR, *sah*A20) described as having an affinity for azinomycin B in Chapter II of this thesis. Genes upstream of *sah*A1 and downstream of *sah*A20 (corresponding to ~8 kb of known sequence not illustrated in the genetic map) do not appear to be relevant to the biosynthetic product of cluster sahA, but such assumptions cannot be proven by empirical evidence alone, which in turn lends a degree of uncertainty to all orphan pathway analyses when attempting to define the boundaries of an uncharacterized cluster.

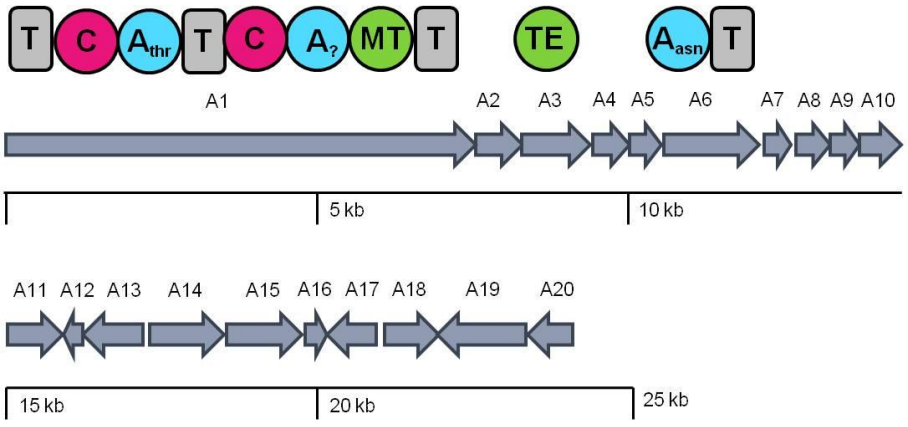


Figure 32 Orphan gene cluster sahA identified in the *S. sahachiroi* genome.

Table 8 Putative function of *orf*s in orphan gene cluster sahA of *S. sahachiroi*.

Gene	Size (aa)	Homolog, Origin, and Accession Number	Identity/ Similarity (%)	Proposed Function
*sah*A1	2824	peptide synthetase NRPS11-10, *Streptomyces verticillus* (AAG02349)	80/86	NRPS (T-C-A_{thr}-T-C-$A_{?}$-MT-T)
*sah*A2	337	clavaminic acid synthase-like protein, *Streptomyces verticillus* (AAG02347)	88/92	Oxygenase, claviminic acid synthetase-like
*sah*A3	248	thioesterase, *Streptomyces verticillus* (AAG02346)	84/89	Thioesterase
*sah*A4	427	putative transporter, *Streptomyces verticillus* (AAG02345)	88/92	Transporter
*sah*A5	333	hypothetical protein, *Streptomyces verticillus* (AAG02344)	83/90	Unknown
*sah*A6	578	peptide synthetase NRPS12, *Streptomyces verticillus* (AAG02343)	85/90	NRPS (A_{asn}-T)
*sah*A7	324	SyrP-like protein, *Streptomyces verticillus* (AAG02342)	87/92	Oxygenase, claviminic acid synthetase-like
*sah*A8	84	MbtH-like protein, *Streptomyces avermitilis* MA-4680 (NP_822026)	66/79	MbtH-like
*sah*A9	154	conserved hypothetical protein, *Streptomyces sviceus* ATCC 29083 (ZP_06914509)	63/72	Unknown
*sah*A10	237	hypothetical protein SCO7796, *Streptomyces coelicolor* A3(2) (NP_631827)	84/92	Unknown
*sah*A11	393	putative glutamate-cysteine ligase, *Streptomyces sp.* C (ZP_05504667)	62/74	Carboxylate-amine ligase
*sah*A12	150	stage II sporulation protein E (SpoIIE), *Streptomyces sp.* SPB74 (ZP_06827585)	45/55	Stage II sporulation protein
*sah*A13	355	putative secreted protein, *Streptomyces ambofaciens* ATCC 23877 (CAJ88362)	67/79	dsDNA repair protein Ku
*sah*A14	234	PAP2 superfamily domain-containing protein, *Streptomyces sp.* e14 (ZP_06712028)	69/78	Phosphatase/haloperoxidase
*sah*A15	237	hypothetical protein SBI_00322, *Streptomyces bingchenggensis* BCW-1 (ADI03443)	79/88	NAD-binding protein
*sah*A16	122	transcriptional regulatory protein, *Streptomyces sp.* Mg1 (ZP_04996370)	84/91	Transcriptional regulator HxlR
*sah*A17	252	short chain dehydrogenase, *Yersinia rohdei* ATCC 43380 (ZP_04613704)	52/73	Dehydrogenase
*sah*A18	224	putative transcriptional regulator, *Streptomyces ambofaciens* ATCC 23877 (CAJ88012)	72/80	Transcriptional regulator tetR_N
*sah*A19	408	integral membrane transport protein, *Streptomyces hygroscopicus* ATCC 53653 (ZP_05513991)	69/81	Membrane-bound transporter
*sah*A20	266	aminoglycoside phosphotransferase, *Streptomyces sviceus* ATCC 29083 (ZP_06915592)	86/90	Resistance protein
*sah*A21	91	hypothetical protein SalbJ_17933, *Streptomyces albus* J1074 (ZP_04703846)	76/84	Unknown

Table 8 continued

Gene	Size (aa)	Homolog, Origin, and Accession Number	Identity/ Similarity (%)	Proposed Function
*sah*A22	849	zinc metalloprotease (elastase-like) protein, *Streptosporangium roseum* DSM43021 (YP_003343925)	36/49	Metallopeptidase
*sah*A23	406	transcriptional regulator, TrmB, Streptomyces flavogriseus ATCC 33331(ZP_05805204)	42/54	Transcriptional regulator LuxR
*sah*A24	150	hypothetical protein SBI_09996, *Streptomyces bingchenggensis* BCW-1 (ADI13114)	96/97	Transcriptional regulator WHTH_GntR
*sah*A25	251	putative oxidoreductase, *Streptomyces bingchenggensis* BCW-1 (ADI13113)	97/98	Dehydrogenase
*sah*A26	133	hypothetical protein SAV_709, *Streptomyces avermitilis* MA-4680 (NP_821884)	78/87	Unknown
*sah*A27	237	putative hydrolase, alpha/beta fold protein, *Mycobacterium intracellulare* ATCC 13950 (ZP_05227650)	42/57	Hydrolase
*sah*A28	227	hypothetical protein Svir_08000, *Saccharomonospora viridis* DSM 43017 (YP_003132687)	40/57	Unknown
*sah*A29	212	hemerythrin HHE cation binding domain-containing protein, *Mycobacterium avium* 104 (YP_883244)	43/52	Unknown

The gene cluster dubbed sahB is a hybrid NRPS-PKS cluster with modules for loading and incorporation of threonine, proline, glycine, and an additional unknown substrate (**Fig. 33** and **Table 9**). The sahB pathway was initially elucidated during Sanger sequencing of the genomic library clone fosmid36 and was elaborated upon through genomic sequencing. The PKS may act as an iterative type I enzyme, judging from the lack of other PKS domains in the immediate vicinity. Again, the genetic map of sahB does not illustrate additional known upstream and downstream genes which are believed to operate independent of the pathway.

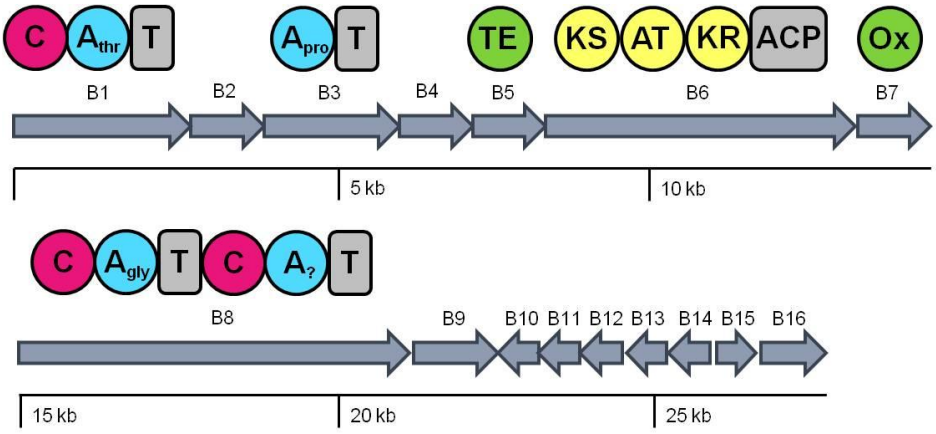


Figure 33 Orphan gene cluster sahB identified in the *S. sahachiroi* genome.

Table 9 Putative function of *orf*s in orphan gene cluster sahB of *S. sahachiroi*.

Gene	Size (aa)	Homolog, Origin, and Accession Number	Identity/ Similarity (%)	Proposed Function
*sah*B1	1119	amino acid adenylation domain-containing protein, *Nostoc punctiforme* PCC 73102 (YP_001869919)	39/55	NRPS (C-A_{thr}-T)
*sah*B2	428	MFS family major facilitator transporter, *Bacillus cereus* 172560W (ZP_04304889)	32/56	Transporter, multidrug resistance
*sah*B3	780	peptide synthetase, *Cyanothece sp*. CCY0110 (ZP_01728755)	38/59	NRPS (A_{pro}-T)
*sah*B4	422	conserved hypothetical protein, *Micromonospora aurantiaca* ATCC 27029 (ZP_06220182)	34/48	Carboxylase
*sah*B5	262	thioesterase type II, alpha proteobacterium BAL199 (ZP_02189723)	34/48	Thioesterase
*sah*B6	1477	polyketide synthase type I, *Cyanothece sp*. ATCC 51142 (YP_001804494)	37/54	PKS (KS-AT-KR-ACP)
*sah*B7	419	non-ribosomal peptide synthase, *Myxococcus xanthus* DK 1622 (YP_632700)	50/66	Flavin-dependent monooxygenase
*sah*B8	2379	amino acid adenylation domain protein, *Haliangium ochraceum* DSM 14365 (YP_003265814)	43/58	NRPS (C-A_{gly}-T-C-$A_{?}$-T)
*sah*B9	502	conserved hypothetical protein, *Pantoea sp*. At-9b (ZP_05729652)	44/61	Unknown
*sah*B10	287	NAD-dependent epimerase/dehydratase, *Kribbella flavida* DSM 17836 (YP_003382415)	66/75	Epimerase/dehydratase
*sah*B11	344	IclR family transcriptional regulator, *Micromonospora sp*. ATCC 39149 (ZP_04604445)	63/75	Transcriptional regulator
*sah*B12	322	putative vanillate demethylase reductase subunit, *Rhodococcus opacus* B4 (YP_002781284)	64/74	Reductase (FMN/FAD and NAD binding)

Table 9 continued

Gene	Size (aa)	Homolog, Origin, and Accession Number	Identity/ Similarity (%)	Proposed Function
*sah*B13	343	oxidoreductase alpha subunit, *Saccharopolyspora erythraea* NRRL 2338 (YP_001105025)	79/88	Non-heme Fe oxygenase
*sah*B14	254	phenyl acetic acid responsive transcriptional repressor, *Streptomyces sp*. SCC 2136 (CAH10131)	62/75	Transcriptional regulator
*sah*B15	300	short-chain dehydrogenase/reductase SDR, *Catenulispora acidiphila* DSM 44928 (YP_003115601)	79/86	Dehydrogenase/reductase
*sah*B16	351	amidohydrolase 2, *Catenulispora acidiphila* DSM 44928 (YP_003115602)	78/87	Metal-dependent hydrolase
*sah*B17	500	feruloyl-CoA synthetase, *Streptomyces clavuligerus* ATCC 27064 (ZP_05006456)	75/81	Acyl-CoA synthetase
*sah*B18	154	enoyl-CoA hydratase, *Streptomyces sp*. e14 (ZP_06711865)	79/88	Acyl-CoA hydratase
*sah*B19	270	AraC family transcription regulator, *Streptomyces viridochromogenes* DSM 40736 (ZP_05536279)	54/69	Transcriptional regulator, AraC family
*sah*B20	169	RacO protein, *Streptomyces ribosidificus* (CAG34704)	56/75	Unknown
*sah*B21	285	SriL03.9, Streptomyces rimosus subsp. paromomycinus (CAG44632)	85/90	Dehydrogenase/reductase
*sah*B22	244	SriL03.10, Streptomyces rimosus subsp. paromomycinus (CAG44631)	86/92	Transcriptional regulator
*sah*B23	556	oxidoreductase, *Streptomyces lividans* TK24 (ZP_06533432)	82/88	Oxidoreductase
*sah*B24	366	transcriptional repressor protein, *Streptomyces coelicolor* A3(2) (NP_624604)	90/93	Transcriptional regulator, glucokinase
*sah*B25	499	alpha-galactosidase SCF8502, *Streptomyces coelicolor* A3(2) (NP_624603)	88/92	glycosyl hydrolase
*sah*B26	449	substrate binding protein, *Streptomyces coelicolor* A3(2) (NP_733496)	73/84	ABC sugar transporter
*sah*B27	233	binding-protein-dependent transport protein, *Streptomyces coelicolor* A3(2) (NP624602)	95/99	ABC sugar transporter (transmembrane)
*sah*B28	326	binding-protein-dependent transport protein, *Streptomyces lividans* TK24 (ZP_05528614)	84/90	ABC transporter permease

Cluster sahC is anchored by a set of type II PKS domains characteristic of pathways known to produce fused six-membered multi-ring natural products (**Fig. 34A** and **Table 10**). A number of tailoring enzymes in the sahC pathway bear resemblance to proteins in the tetracenomycin gene cluster,[114] led by three polyketide cyclases (*sah*C8, 9, and 13), an acyl carrier protein (*sah*C10) and ketosynthase α and β domains (*sah*C11 and *sah*C12) to generate a proposed multicyclic natural product. Tetracenomycin is an anthracycline antibiotic produced by *Streptomyces glaucescens* and has substantial antitumor activity (**Fig. 34B**). An additional glycosyltransferase (*sah*C18) and aminoglycoside phosphotransferase (*sah*C5) suggest the cluster natural product may be glycosylated and uses the phosphotransferase for self-resistance purposes. The genetic map of sahC in the corresponding figure includes all genes currently known for the contig, and thus may not represent the content of the complete pathway.

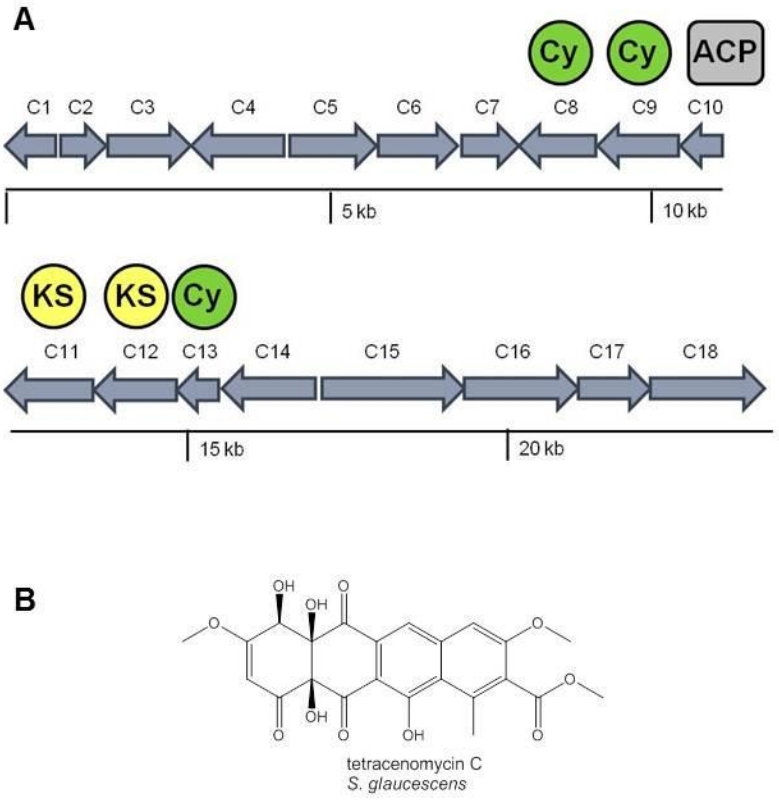


Figure 34 Orphan gene cluster sahC identified in the *S. sahachiroi* genome.
(A) Genetic map of the gene cluster. (B) The structure of tetracenomycin C, an antitumor macrolide produced by *Streptomyces glaucescens*.

Table 10 Putative function of *orf*s in orphan gene cluster sahC of *S. sahachiroi*.

Gene	Size (aa)	Homolog, Origin, and Accession Number	Identity/ Similarity (%)	Proposed Function
*sah*C1	207[a]	conserved hypothetical protein, *Streptomyces* sp. e14 (ZP_06708591)	86/91	Unknown
*sah*C2	~383[b]	hypothetical protein SCO4267, *Streptomyces coelicolor* A3(2) (NP_628439)	85/92	Unknown
*sah*C3	618	drug resistance transporter, EmrB/QacA subfamily, *Streptomyces* sp. e14 (ZP_06708592)	88/92	Membrane-bound transporter
*sah*C4	444	glycine/D-amino acid oxidase, *Streptomyces* sp. e14 (ZP_06708593)	89/93	Oxidase
*sah*C5	452[b]	hydroxyurea phosphotransferase, *Streptomyces avermitilis* MA-4680 (NP_825138)	69/79	Aminoglycoside phosphotransferase
*sah*C6	352	secreted protein, *Streptomyces* sp. e14 (ZP_06708595)	73/84	Secreted protein
*sah*C7	369	*O*-methyltransferase, *Streptomyces* sp. e14 (ZP_06708596)	92/95	SAM-dependent *O*-methyltransferase
*sah*C8	101	polyketide synthase CurG, *Streptomyces* sp. e14 (ZP_06708597)	93/96	PKS cyclase
*sah*C9	137	conserved hypothetical protein, *Streptomyces* sp. e14 (ZP_06708598)	91/95	PKS cyclase
*sah*C10	72	curamycin polyketide synthase acyl carrier protein, *Streptomyces* sp. e14 (ZP_06708599)	94/97	ACP
*sah*C11	360	polyketide beta-ketoacyl synthase 2, *Streptomyces* sp. e14 (ZP_06708600)	92/94	KS subunit β
*sah*C12	410	polyketide beta-ketoacyl synthase 1, *Streptomyces* sp. e14 (ZP_06708601)	96/97	KS subunit α
*sah*C13	124	tetracenomycin polyketide synthesis protein TcmJ, *Streptomyces* sp. e14 (ZP_06708602)	96/98	Unknown
*sah*C14	328	polyketide synthase CurD, *Streptomyces* sp. e14 (ZP_06708603)	87/93	Unknown
*sah*C15	547	tetracenomycin polyketide synthesis hydroxylase TcmG, *Streptomyces* sp. e14 (ZP_06708604)	85/90	PKS hydroxylase
*sah*C16	400	ABC transporter ATP-binding protein, *Streptomyces coelicolor* A3(2) (NP_628414)	95/97	ABC sugar transporter
*sah*C17	149	small membrane protein, *Streptomyces lividans* TK24 (ZP_05524727)	80/91	Membrane protein
*sah*C18	209[a]	guanyltransferase, *Streptomyces* sp. e14 (ZP_06708607)	93/97	Guanyltransferase/ glycotransferase

[a] incomplete gene
[b] frameshifted

Cluster sahD is an apparent partial biosynthetic gene cluster containing type II PKSs, a type III PKS, and enzymes dedicated to sugar modification and transfer (**Fig. 35** and **Table 11**). Independent enzymes in the sahD pathway do not share significant homology to known or well-characterized proteins, preventing specific structural predictions like those presented for cluster sahC/tetracenomycin. In its current state, the sequence for cluster sahD appears to have several frameshifts and incomplete gene sequences at both ends, and also lacks common cleavage domains to terminate polyketide chain elongation. Enzymes acting on sugars are located between the discrete PKS domains, suggesting they participate in the biosynthetic pathway. However, it is also possible that the Type II PKS domains (*sah*D1-D6) are part of one pathway and the chalcone synthase, dehydratase, and PKS cyclase (*sah*D12-D14) are remnants of a separate pathway. RAST output suggests the type II PKS domains (*sah*D1, D2, and D4) are β-ketoacyl synthases I, II, and III involved in fatty acid biosynthesis, a product of primary metabolism. One method for determining if an enzyme is dedicated to fatty acid or polyketide synthesis is to examine the CLF (chain length factor) domain for the presence or absence of an active site cysteine residue.[115] Without extended sequence context in both directions, it is not entirely correct to assume the sahD enzymes belong to a single cluster.

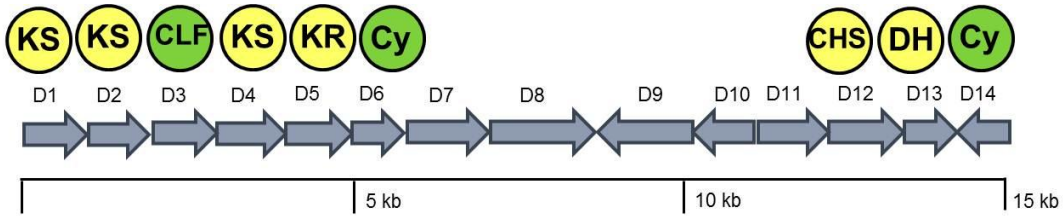


Figure 35 Orphan gene cluster sahD identified in the *S. sahachiroi* genome.

Table 11 Putative function of *orf*s in orphan gene cluster sahD of *S. sahachiroi*.

Gene	Size (aa)	Homolog, Origin, and Accession Number	Identity/ Similarity (%)	Proposed Function
*sah*D1	148	3-oxoacyl-[acyl-carrier-protein] synthase, *Streptomyces sp*. C (ZP_05508643)	67/79	Ketosynthase
*sah*D2	619	3-oxoacyl-[acyl-carrier-protein] synthase, *Streptomyces sp*. C (ZP_05508643)	74/82	Ketosynthase
*sah*D3	277	Beta-ketoacyl synthase, *Streptomyces bingchenggensis* BCW-1 (ADI05720)	51/60	Chain-length factor
*sah*D4	353	3-oxoacyl-(acyl-carrier-protein) synthase III, *Streptomyces sp*. C (ZP_05508645)	76/88	Ketosynthase
*sah*D5	310	aldo/keto reductase, *Streptomyces sp*. C (ZP_07288856)	72/84	Aldo/keto reductase
*sah*D6	344	FAD-dependent oxidoreductase, *Streptomyces sp*. C (ZP_05508648)	70/79	Hydroxylase/ monooxygenase
*sah*D7	293	hypothetical protein StreC_23256, *Streptomyces sp.* C (ZP_05508649)	66/75	PKS cyclase
*sah*D8	179	glycosyltransferase, *Streptomyces griseus subsp. griseus* (CAE17547)	30/42	Glycosyltransferase
*sah*D9	262	SARP family pathway specific regulatory protein, *Streptomyces bingchenggensis* BCW-1] (ADI09962)	67/79	Regulatory protein
*sah*D10	208	NDP-hexose 3,5-epimerase, *Streptomyces eurythermus* (ABW91159)	47/60	Sugar epimerase
*sah*D11	490	putative dNDP-4-keto-6-deoxy-glucose-2,3-dehydratase SimB3, *Streptomyces antibioticus* (AAK06810)	56/70	Sugar dehydratase
*sah*D12	387	putative type III polyketide synthase, *Saccharopolyspora erythraea* NRRL 2338 (ZP_06563505)	62/75	Type III (chalcone) PKS
*sah*D13	152	MaoC domain-containing protein dehydratase, *Micromonospora aurantiaca* ATCC 27029 (YP_003833831)	58/76	Dehydratase
*sah*D14	139	hypothetical protein StreC_23241, *Streptomyces sp*. C (ZP_05508646)	53/67	PKS cyclase

Putative cluster sahE currently consists of only a gene fragment corresponding to the C-terminal end of a potentially large type I PKS based on sequence homology to the pladienolide PKS encoded by *pldA*I (**Fig. 36A** and **Table 12**). Sequence data provides only seven domains in sahE1 resulting from the most recent assembly efforts. PldAI is an 18-domain type I PKS within the genome of *Streptomyces platensis* Mer-11107 known to participate in the biosynthesis of at least seven pladienolide structural variants.[116] Pladienolide B (**Fig. 36B**) is the variant produced at the highest level during fermentation and exhibits both VEGF inhibition and antitumor properties.[117,118] Amino acid sequence similarity (49%) suggests sahE1 may work in tandem

with additional PKSs to produce a highly reduced macrolide. While *S. sahachiroi* genomic sequence annotation has only afforded a partial gene sequence, such evidence provides sequence clues for design of gene-specific oligonucleotide probes that can be used to screen genomic libraries for clones containing additional sequence information.

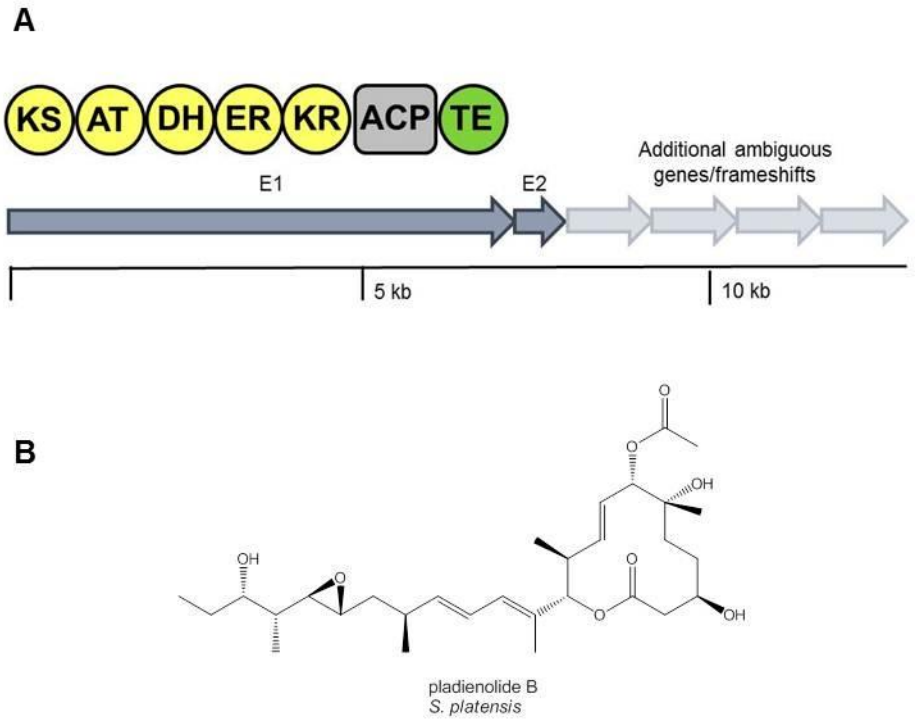


Figure 36 Orphan gene cluster fragment sahE identified in the *S. sahachiroi* genome.
(A) Genetic map of the gene fragments. (B) Structure of pladienolide B, one of seven metabolites in the pladienolide family constructed with the assistance of a large type I PKS, PldAI. *sah*E1 has 49% similarity and 61% identity in comparison with *pld*AI.

Table 12 Putative function of *orf*s in orphan gene cluster sahE of *S. sahachiroi*.

Gene	Size (aa)	Homolog, Origin, and Accession Number	Identity/ Similarity (%)	Proposed Function
*sah*E1	2427	polyketide synthase, *Streptomyces platensis* (BAH02268)	49/61	PKS (KS-AT-DH-ER-KR-ACP-TE)
*sah*E2	150	possible transposase, *Streptomyces fradiae* (AAZ23101)	91/94	transposase

The State of Genome Completion

At the conclusion of the present *S. sahachiroi* genome assembly effort by the author, the *S. sahachiroi* sequence information obtained from a combination of fosmid Sanger sequencing, 454 GS FLX, and Illumina GAII exists as a set of 9,759 contigs (898 contigs are >5 kb) with the largest contig at 93 kb and a median contig size of 839 bp. The project retains significant room for manual assembly using contig building software and human interaction to correct the often oversized arbitrary gaps introduced by the Velvet short-read assembler when it analyzes paired-end sequences. At this point it is uncertain whether further assembly would yield additional new orphan clusters pertinent to our group's interests.

SIGNIFICANCE

Identification of orphan pathways provides the necessary genetic information to pursue cluster capture and perform studies related to pathway activation or upregulation. Experiments to understand promoter regulation, promiscuity, and induction signaling may ultimately serve to support the original fermentation and fractionation practices that lead to new biosynthetic product isolation. If pathway induction could be generalized as a strategy, it would further augment traditional methods for natural product discovery or increase production of currently known metabolites for human use.

Once the pathway is successfully activated, the product of the orphan cluster will be the focus of structure determination and bioassay for suitability as an antibacterial or other medicinally-relevant species. Given the similarity observed in bioinformatic prediction of gene products to those of established bioactive compounds, *S. sahachiroi* has the potential to offer either several useful variants of compounds already included in the natural product pool, or to even supply entirely new structures for potential biomedical application. Because the genome is still in a draft stage, there is also the possibility of as-yet assembled additional secondary pathways. As gap closure and assembly closes in on a single contiguous sequence, further utility in comparative genomics to study aspects such as gene transfer, primary metabolic pathway variations, and species propagation become feasible undertakings. In short, genomic sequence data opens the door to seemingly unlimited investigation possibilities.

We are additionally interested in extending sequence information pertinent to the reported azinomycin B gene cluster within *S. sahachiroi*.[25] Despite bioinformatic analysis, the exact steps and timing related to biosynthesis of the aziridinopyrrollidine ring remain unclear. Our independent efforts to assign and enzymatically reconstitute elements of the pathway reveal

several proteins deemed necessary to complement the existing biosynthetic model are lacking, either because they are beyond the published sequence data or because they act in *trans* from a distant location within the genome. Therefore, having a draft genome facilitates the search for probable gene candidates to carry out undefined processes in azinomycin biosynthesis, transport, and resistance for future investigation.

EXPERIMENTAL PROCEDURES

Construction and Screening of the *S. sahachiroi* Genomic Library

S. sahachiroi genomic library construction and the NRPS/PKS screening method employing ^{32}P-end labeled degenerate hybridization probes are outlined in the Experimental Procedures section of Chapter II in this thesis.

454 DNA Pyrosequencing of *S. sahachiroi*

S. sahachiroi was cultured in YEME media (0.3% yeast extract, 0.5% peptone, 0.3% malt extract, 1% glucose, 34% sucrose, 5 mM $MgCl_2$, 20% glycine). Genomic DNA for all sequencing applications was isolated using the salting out procedure.[104] Pyrosequencing with a read length of 100 bp was carried out by Macrogen (Seoul, Korea) using a single microwell plate on the Roche/454 Life Sciences 2008 version of the GS FLX platform. Macrogen also provided sequence assembly services for the pyrosequencing data and delivered the final results as 5,712 contigs of varying size.

Illumina DNA Sequencing of *S. sahachiroi*

An adapter-ligated library of *S. sahachiroi* genomic DNA with average fragment size of 200 bp was generated using the Illumina DNA Sample Kit (Illumina, San Diego, CA) and the Illumina-provided protocol. Single-read 36 bp sequencing was carried out in a single flow cell lane using a Genome Analyzer II (Laboratory for Genome Technology, Norman E. Borlaug Center, Texas A&M University). Data was processed within the Illumina Genome Analyzer pipeline by the facility housing the sequencer.

A paired-end adapter-ligated library of *S. sahachiroi* genomic DNA with average fragment size of 350 bp was produced using the Illumina Paired-End DNA Sample Prep Kit with provided procedure. Paired-end 36 + 36 bp sequencing was carried out in a single flow cell lane and pipeline data processing was completed in a manner analogous to the single-read run.

Illumina Sequence Assembly

Bioinformatics specialist Serge Batalov (Andrew I. Su group, Genomics Institute of the Novartis Research Foundation, La Jolla, CA) performed all short-read assemblies. Optimal short-read sequence assembly for the single-read Illumina sequence data was accomplished using Edena with a seed length of 27.[119] The single read data was also combined with sequence from 454/pyrosequencing using Edena with an optimal seed length of 21. The 454 data was arbitrarily fragmented to resemble short-read input for assembly purposes. Paired-end sequence data was processed using Velvet with optimal seed length of 25.[120] The author completed additional manual assembly aided by the software program Sequencher 4.8 (Gene Codes, Ann Arbor, MI) to enable internal gap filling of paired-end scaffolds using contigs generated from the single-read assembly.

Sequence Annotation

Automated annotation of contigs larger than 5 kb was accomplished using the SEED Rapid Annotation Through Subsystems Technology (RAST) server developed and maintained by the National Microbial Pathogen Data Resource Center.[113] Manual curation efforts combined FramePlot 3.0beta,[106] NCBI BLAST[105] and NCBI Conserved Domain Databases[87] for general annotation. Specialized databases such as the NRPS-PKS web-based utility[121] and PKS/NRPS Prediction Website[122] were employed to predict domain specificity of modular enzymes.

CHAPTER IV

β-LACTOGLOBULIN, CYCLORETINAL, AND THE LINK TO AGE-RELATED MACULAR DEGENERATION

INTRODUCTION

In 1992, Asato and coworkers reported an unusual and unintended reaction occurring between the milk whey protein β-lactoglobulin (BLG) and the retinoid-like compound β-ionylideneacetaldehyde.[79] In the course of studying photoisomerization of retinoids immobilized by BLG (a putative retinol-binding protein, acting as a mimic for rhodopsin),[123] it was noted that β-ionylideneacetaldehyde, a C-15 α,β-unsaturated aldehyde, underwent conversion to a C-30 ring-fused homodimer to a small (~11%) but measurable extent (**Fig. 37**). Authentic standards for the C-30 homodimer, as well as a C-40 homodimer (all-*trans*-retinal dimer or cycloretinal) and other retinoid compounds were found to produce identical UV-visible absorbance traces and NMR spectra as products obtained from BLG incubation. However, no further work has been reported regarding these transformations from either a protein chemistry viewpoint nor from the molecular level. Because BLG is a protein without definite ascribed function and is prevalent in the diets of populations consuming dairy products, evidence of catalytic activity piqued our interest in this well-studied yet mysterious globular protein.

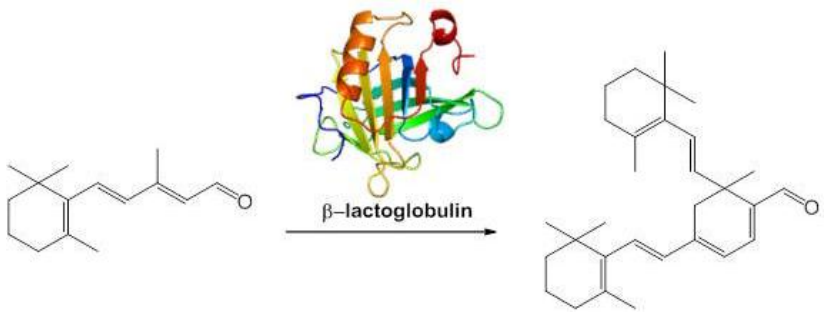


Figure 37 BLG-promoted formation of a C-30 ring-fused dimer from β-ionylideneacetaldehyde.

The results of this study are derived from the efforts of Dr. Bennie J. Bench and the author. Specific experimental roles are denoted by the initials B.J.B, J.F.A, or C.M.H.W (Prof. Coran M. H. Watanabe) following each experimental description in the Experimental Procedures section of this chapter.

The ring-fused dimers produced by assistance from BLG are members of a class of natural product aldehydes we have termed cycloterpenals to reflect both their terpenoid biosynthetic origin and cyclohexadienal structural core.[76,124] Representative cycloterpenals with immediate relevance to this particular study include cyclocitral, an antibiotic secondary metabolite produced by the North Sea bryozoan *Flustra foliacea*,[125-127] and cycloretinal (**Fig. 38**), a vision cycle by-product isolated from the human eye that is potentially associated with age-related macular degeneration (AMD),[73,128,129] the primary cause of vision loss in Americans over the age of 60.[77,130] AMD is characterized by loss of sharp central vision and stems from abnormal neovascular structures beneath the macula, the disc at the center of the retina. These fragile, proliferating blood vessels often leak fluids that displace the macula, preventing focus of fine images necessary for routine daily tasks. Early indicators of potential AMD include deposition of lipids, proteins, and other cellular elements in drusen, yellow granules outside the retinal pigment epithelium (RPE) that are ostensibly harmless and produce no direct symptoms.[78,131] Also evident near RPE cells are lipofuscins, autofluorescent aggregates of cellular debris and the remnants of photoreceptor outer segment phagocytosis carried out to re-isomerize retinoids for the visual cycle. Though not definitive disease markers, both drusen and lipofuscins are very frequently detected in AMD patient tissue.[77]

Proteomic studies indicate BLG is among the proteins crystallized in drusen.[78] Coupled with this discovery is the accumulation of fluorophores such as cycloretinal, cycloretinal-phosphatidylethanolamine conjugate, and A2E in nearby lipofuscins.[73] The highly conjugated system of A2E has been implicated in oxidative damage mechanisms in RPE cells that ultimately lead to apoptosis and is itself of concern independent of AMD.[74] The presence of BLG and retinoid-derived compounds near tissues susceptible to maculopathy, combined with knowledge of BLG's ability to promote cycloterpenal formation, suggest a possible correlation between these molecular species and disease progression. From an AMD prevention standpoint, the effect of milk consumption bears scrutiny and thus we are taking a closer look at BLG chemistry and the biological relevance of the reaction *in vivo* as the first steps toward defining a protein-based impetus for lipofuscin formation.

This study extends the initial observation of BLG-promoted α,β-unsaturated aldehyde condensation and attempts to define the limits of BLG activity in both purified and naturally-occurring milk-borne form. The series of experiments detailed here range from *in vitro* to *in vivo* observations of BLG chemistry.

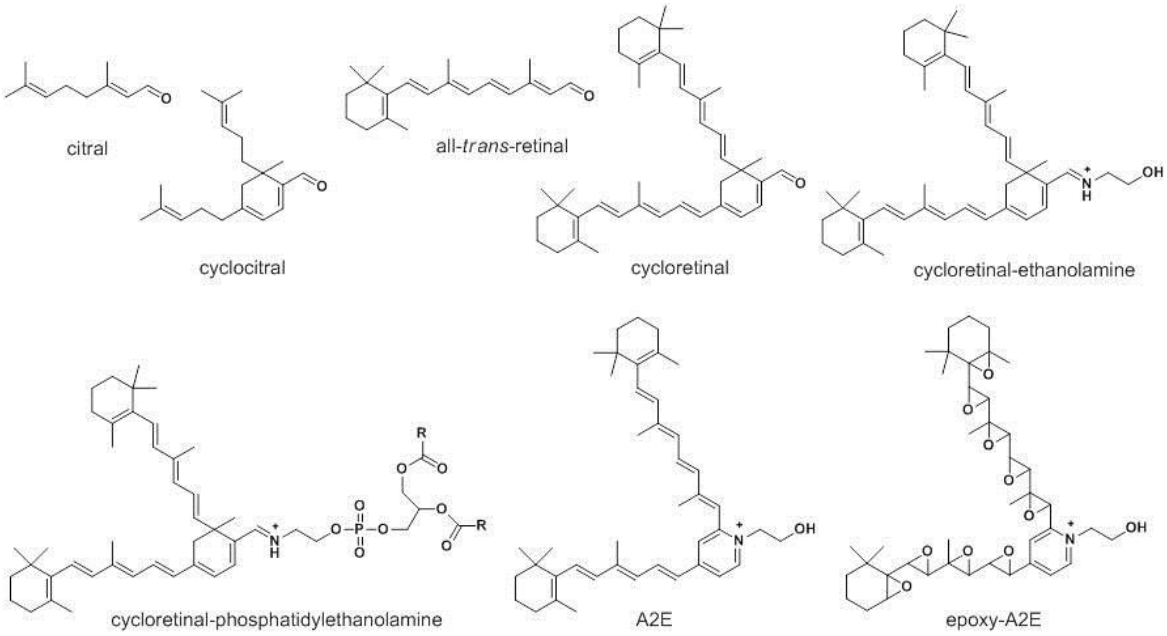


Figure 38 Molecular species relevant to the study of BLG chemistry and lipofuscin formation. R groups designate long alkyl chains of phosphatidylethanolamine.

RESULTS AND DISCUSSION

BLG-Promoted Cycloterpenal Formation

Purified bovine BLG was incubated with the model α,β-unsaturated aldehydes citral or all-*trans*-retinal to demonstrate the ability to catalyze self-condensation reactions. Scouting for optimal conditions included varying reaction duration, buffer pH, and protein to substrate ratio to achieve maximum yield. Using citral as substrate, 58% conversion to cyclocitral was observed under a 1:3 protein to substrate ratio at pH 7.0 and 37°C over 4 days (**Fig. 39** and **Appendix Fig. 64**). The reaction proceeds quite slowly by enzymatic standards, suggesting this moonlighting activity is not the primary biological role of BLG, but it is a remarkable behavior nonetheless.

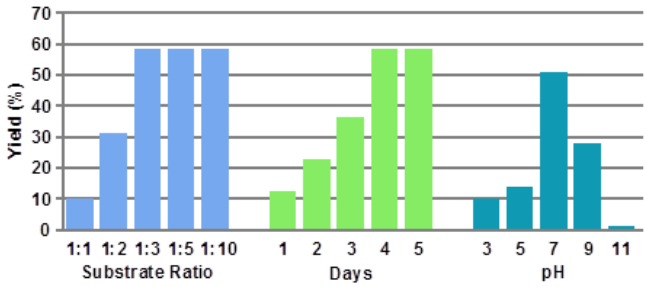


Figure 39 Optimization of BLG reaction conditions.
Yield reported as a function of protein-to-substrate ratio, reaction duration, and buffer pH.

Reaction progress can be monitored by examination of the aldehydic region of ^{1}H NMR (**Appendix Fig. 63**). The commercially available citral used as substrate is a mixture of *cis* and *trans* isomers whose aldehyde protons are centered at δ 10 ppm, and the signal corresponding to the cyclocitral aldehyde proton is shielded, shifting upfield to δ 9.45 ppm. Thus, emergence of the cyclocitral aldehyde signal indicates expected product formation in the presence of BLG. Biomimetic dimerizations catalyzed by proline implicate aldehyde activation through formation of a Schiff base[76] in a mechanism analogous to the capability of a lysine residue in a protein. This approach is precedented, as an example of biological aldehyde activation via imine formation includes a step in the vision cycle in which 11-*cis*-retinal is covalently linked to a lysine residue in opsin to give the visual pigment rhodopsin.[132] To determine if any proteinaceous lysine could promote cycloterpenal formation, bovine serum albumin (BSA), a protein containing 60 lysine residues, was evaluated for this specific catalytic activity (**Fig. 40**).

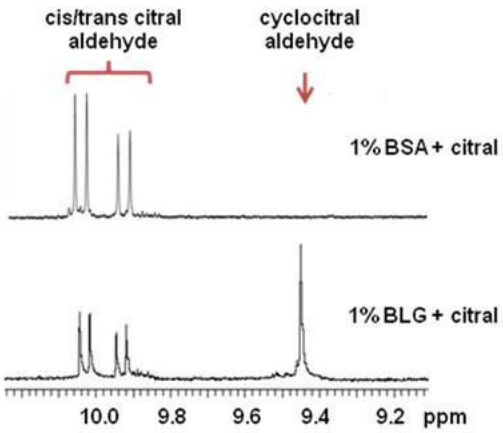


Figure 40 Aldehydic ^{1}H NMR region of reaction extract from citral incubated with BLG or BSA.

Additional protein extracts from *E. coli* and retinal pigment epithelial (RPE) cells were also assessed for assisting conversion of citral to cyclocitral. Product formation in all cases was monitored by ^{1}H NMR. The *E. coli* proteome contains over 1,000 proteins, and RPE cells are the obvious location for retinoid-derived dimer formation in the eye. If a particular class of proteins represented in a bacterial proteome or specific proteins present in the RPE could also drive cycloterpenal formation, then the behavior exhibited by BLG would remain an accidental but inconsequential discovery. However, because BSA, *E. coli* extract, and RPE extract all failed to produce the ring-fused dimer upon incubation with citral (**Fig. 41**), it was determined that the reaction was unique to BLG and not a general reaction promoted by proteins or peptides.

The reaction products of a BLG solution incubated with either citral or all-*trans*-retinal were also followed by UV-visible spectroscopy (UV-vis) to demonstrate another analytical method for product detection. Reaction aliquots extracted every 24 h were dried and resuspended in ethanol for wavelength absorbance scanning from 200-600 nm. Scans from each 24 h aliquot were overlaid to produce the spectra in **Figure 42**. Citral and all-*trans*-retinal have respective maximum absorbance at 240 nm and 385 nm, and the products cyclocitral and cycloretinal absorb most strongly at 320 nm and 290 nm/432 nm, respectively. An isobestic point is evident at 267 nm for citral/cyclocitral and 342 nm for retinal/cycloretinal. While cycloterpenal product degradation is evident in the traces corresponding to protein-mediated synthesis compared to a proline-catalyzed authentic sample, absorbance wavelength scanning does provide a glimpse of reaction progress over the incubation period. The reaction with all-*trans*-retinal appears to proceed slightly faster than an identical trial carried out with citral when characterized in this manner.

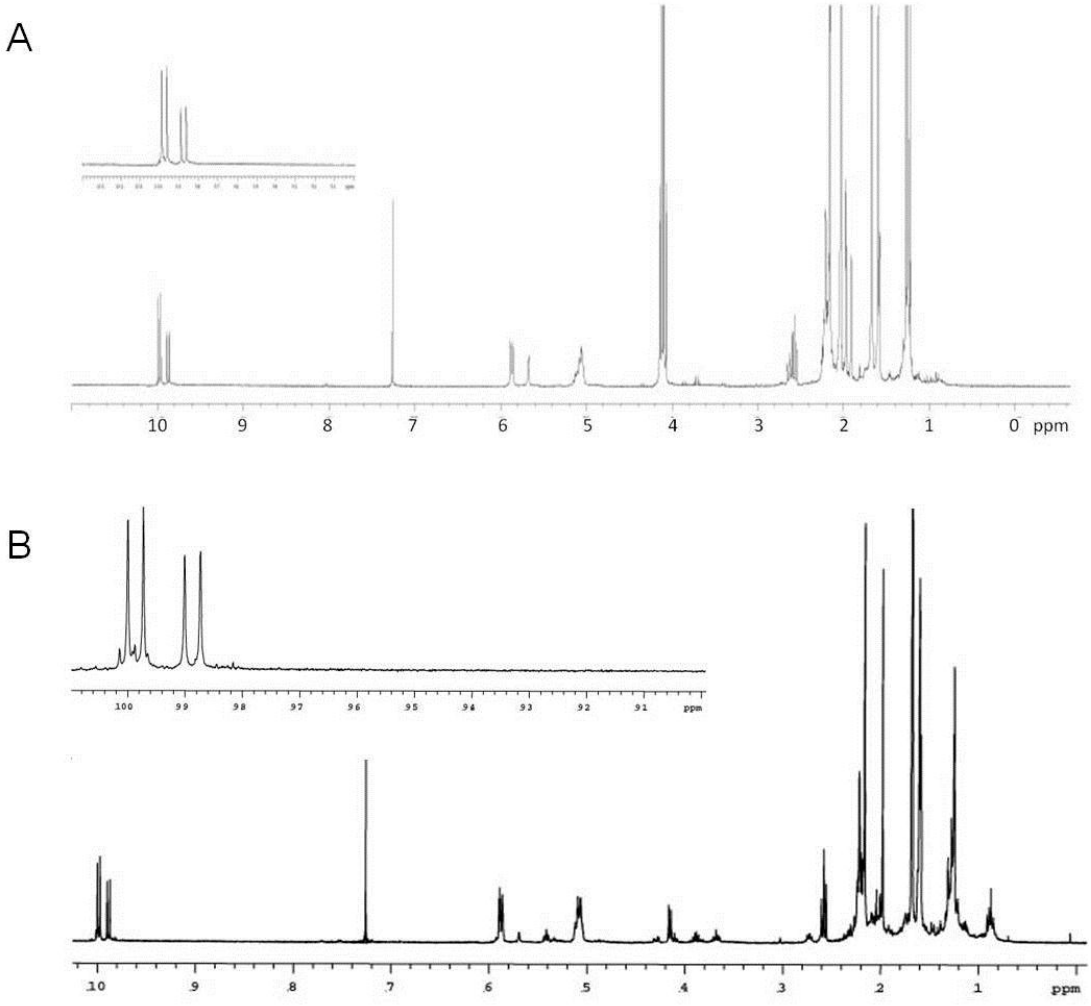


Figure 41 Evaluation of *E. coli* proteome or RPE cell extract assistance in cycloterpenal formation.

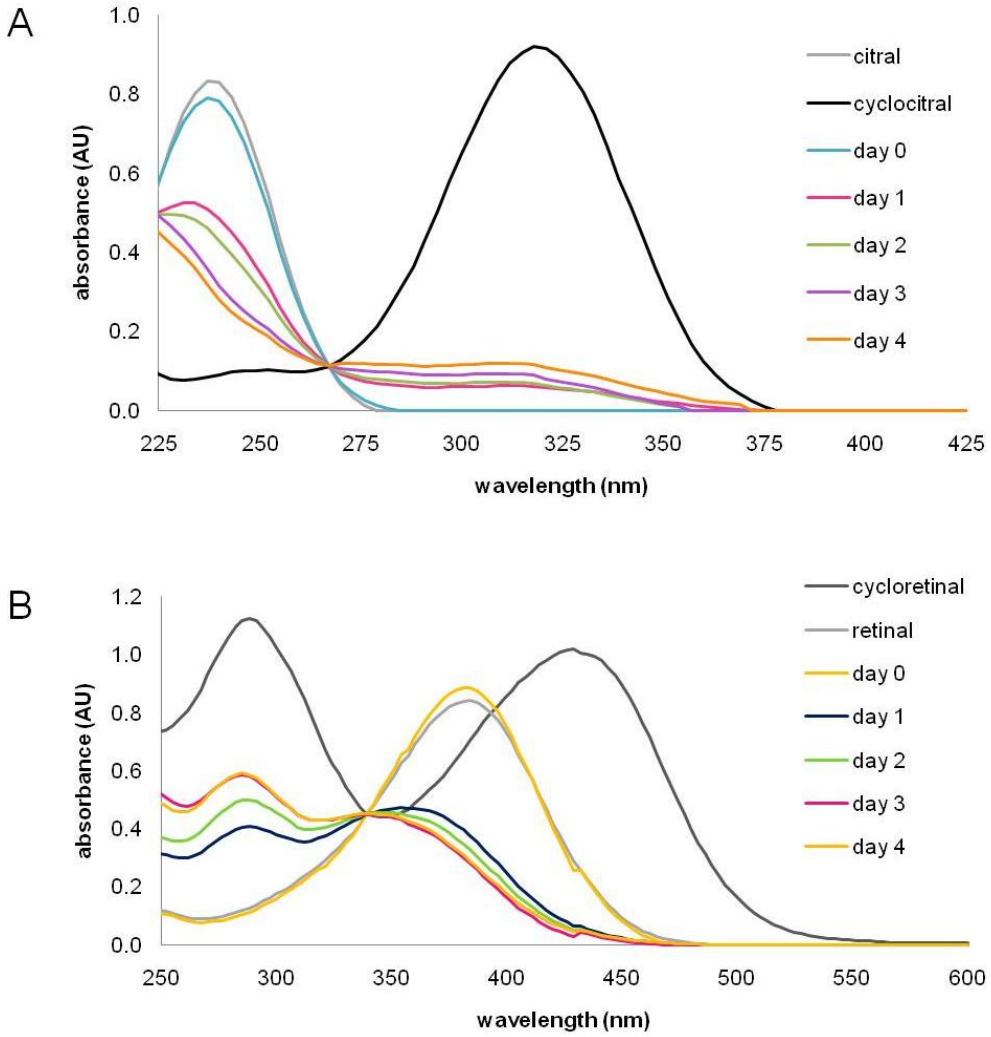


Figure 42 UV-visible absorption wavelength scans of BLG-citral or –retinal extracts.
It should be noted that the products from days 1-4 provide a spectrum equivalent to that of an authentic sample maintained at room temperature for 24 h, and thus product instability/degradation limits the utility of total sample analytical methods such as UV-vis. Fortunately, such limitations do not impede the use of NMR analysis.

Additional analysis of reaction products was carried out by HPLC to verify cycloterpenal formation employing co-injection of an authentic standard sample (**Fig. 43**). BLG incubated with all-*trans*-retinal in a standard four day treatment was subjected to HPLC separation following extraction, and the product mixture was compared to authentic cycloretinal and a BLG blank extract. Co-injection of the BLG-assisted product mixture and synthetic cycloretinal verified formation of the predicted product.

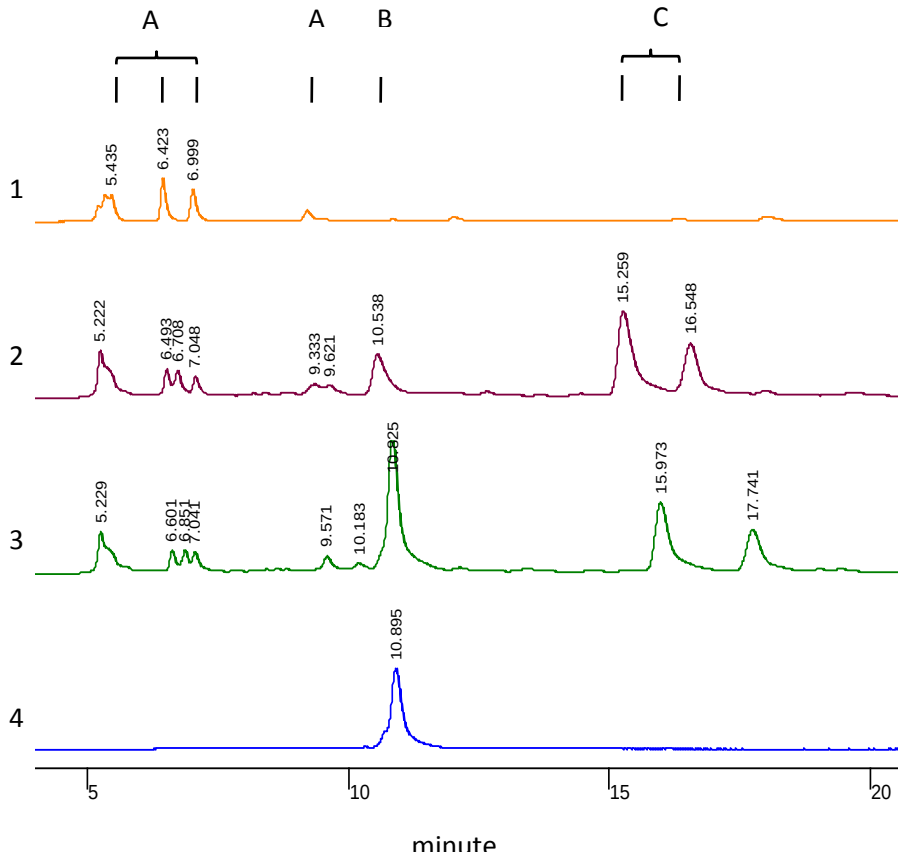


Figure 43 HPLC profile of BLG-promoted cycloretinal formation.
[1] BLG extract [2] BLG + retinal extract [3] BLG + retinal extract spiked with synthetic cycloretinal (co-injection) [4] synthetic cycloretinal. Peak identification is as follows: (A) BLG-bound metabolites (B) cycloretinal (C) retinal-derived side products.

Attempts to abolish the catalytic behavior of BLG were carried out employing standard approaches to protein denaturation, i.e., treatment with detergent, heat, or chaotropic reagent, followed by incubation with citral, extraction, and ^{1}H NMR analysis. BLG-promoted cycloterpenal formation is inhibited by pretreating BLG with detergent (10% SDS) but not with other denaturants (7 M urea or 100°C, 2 h) (**Fig. 44** and **Appendix Fig. 65**), providing further evidence that not only is BLG a robust and stable protein, but also that catalytic activity does depend upon tertiary structure preservation.

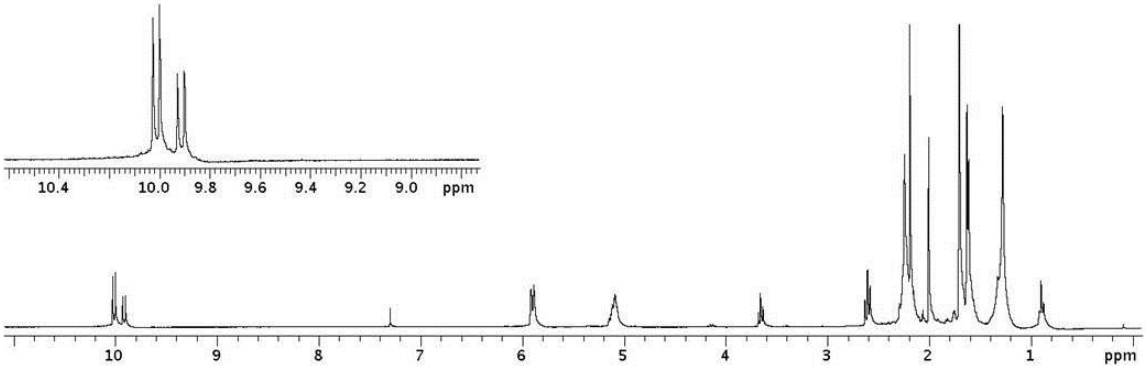


Figure 44 ^{1}H NMR of BLG catalytic inhibition in the presence of sodium dodecyl sulfate.
Only treatment with detergent (10% w/v SDS), and not heat nor urea, caused BLG to lose activity in the standard assessment using citral as substrate. The inset is an expansion of the aldehydic region to highlight lack of the cyclocitral aldehyde signal.

Milk-Promoted Formation of Cycloterpenals

With knowledge of purified BLG catalytic activity and heat stability, we next sought to evaluate the ability of pasteurized cow's milk available from a local grocer to assist in cycloterpenal biosynthesis. Reactions carried out with either whole milk (4% milk fat), 2% milk (2% milk fat), skim milk (fat-free), or buttermilk (4% milk fat) were assayed with excess citral as substrate and examined by NMR (**Fig. 45**). The highest conversion was obtained using skim milk (14%), followed by 2% milk (7.3%), whole milk (5%), and buttermilk (1%). The yield variance can be rationalized by noting BLG has a propensity to bind a number of lipophilic ligands with submicromolar affinity,[82,133] and thus greater fat content in milk translates to a relative form of inhibition. The poorest result observed with buttermilk may be further attributed to the lower pH (in the range of pH 4.4-4.8, due to lactic acid) in contrast to pH 6.7 for ordinary milk.

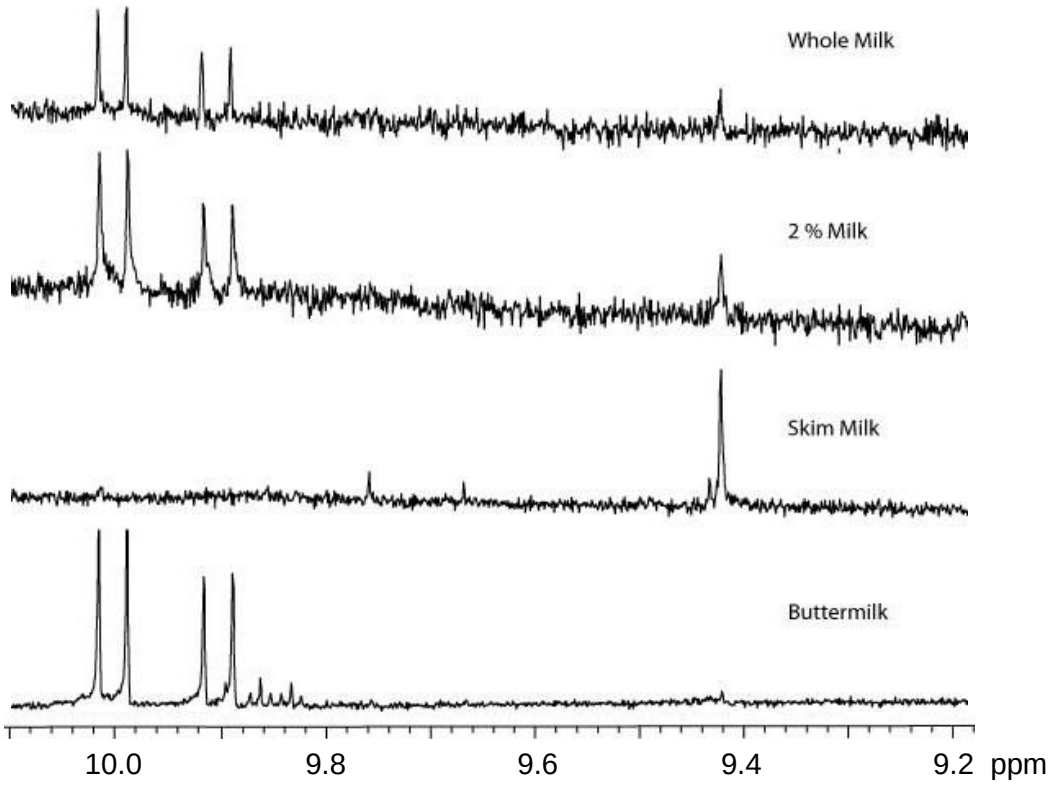


Figure 45 ^{1}H NMR evaluation of milk-assisted cyclocitral formation.

To ensure the milk-assisted reaction was due to protein and not small molecule involvement, skim milk was desalted using Sephadex G-25 medium and re-subjected to assay (**Appendix Fig. 66**). Reaction yield was unaffected, supporting the concept that heat-treated BLG retains the function to promote cycloterpenal formation.

Study of Cycloterpenal Formation in a Rabbit Model

An average person has between 1.9 and 2.5 g of BLG present in blood sera.[84] As humans do not possess a genetically-encoded BLG homolog, the BLG detected in blood is derived exclusively from consumption of foods containing whey protein. Because BLG is acid stable and receptors for lipocalins are present within the small intestine, the protein has unevaluated potential to survive ingestion and exhibit catalytic behavior *in vivo*. To further probe the physiological relevance of BLG in promoting cycloterpenal formation *in vivo*, specifically for that of cycloretinal, we examined the ability of BLG (in purified form and in milk) to support cycloretinal formation in a rabbit. Eight New Zealand white rabbits were

allocated to the control (basic diet) (n=1), BLG/water control (n=1), skim milk control (n=1), BLG/water + retinal (n=2), and skim milk + retinal (n=2). All of the rabbits had free access to a standard rabbit diet and water. They were each administered their respective solutions by gavage twice daily for 7 days. Following the feeding regimen, the rabbits were anesthetized and exsanguinated for subsequent analysis. The blood was extracted with ethyl acetate, the organics were dried over anhydrous magnesium sulfate, and concentrated *in vacuo*. The extracts were analyzed by ^{1}H-NMR spectroscopy (**Fig. 46A**). Cycloretinal was only detected in rabbits that were supplied a solution of BLG and retinal or skim milk and retinal. To verify the presence of BLG in blood, we analyzed rabbit blood serum by protein electrophoresis. The serum was diluted (1:10) with buffer and analyzed by SDS-PAGE. The gel (**Fig. 46B**) revealed that BLG was readily absorbed into the blood stream of rabbits with the exception of the control rabbit (which only received a basic diet and showed no trace of the protein.) These results are consistent with studies which have shown that absorption of intact BLG into the blood stream proceeds by intracellular transfer through the intestinal transepithelial cells.[83,134] Blood isolated from rabbits provided a BLG-supplemented diet promoted the formation of cycloretinal from retinal during subsequent *in vitro* incubation, while the control (basic diet) rabbit blood showed no signs of product formation when incubated with retinal (**Appendix Fig. 68**). This indicates dietary BLG is uptaken and is active within blood for at least the period of time necessary to assist cycloterpenal biosynthesis, and the reaction previously studied only *in vitro* does have relevance within a living system.

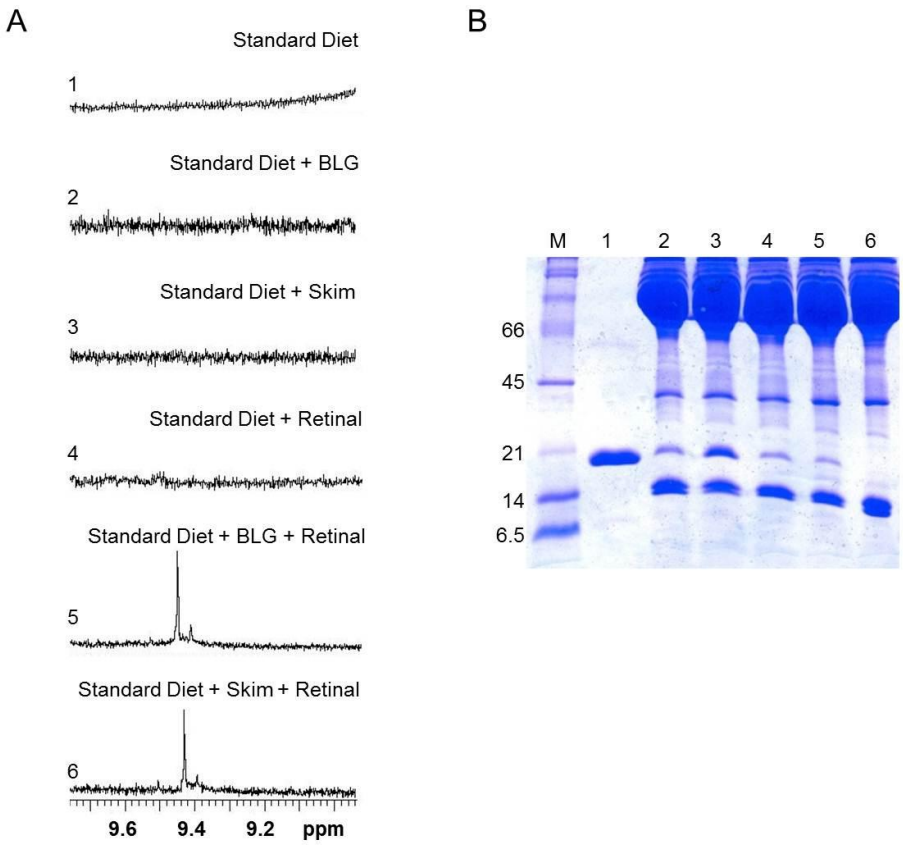


Figure 46 Analysis of cycloretinal formation in rabbit blood.
[A] Shows alignments of the aldehydic region of the ^{1}H-NMR spectrum of organics extracted from rabbit blood samples: (1) Control (basic diet) (2) BLG control (3) skim milk (4) retinal control (5) BLG + retinal (6) skim + retinal; Detection of BLG in rabbit blood. [B] SDS-PAGE analysis of blood serum: (M) molecular weight marker, kD (1) BLG standard (2) BLG control (3) BLG + retinal (4) skim milk (5) skim milk + retinal (6) control (basic diet).

SIGNIFICANCE

The studies presented herein describe terpenal ring-fused dimer formation aided by the milk protein β-lactoglobulin in an unprecedented biosynthetic fashion, the products of which may bear physiological implication in disease states such as age-related macular degeneration. Because AMD is marked by retinoid dimer accumulation within lipofuscins, and we have demonstrated BLG is capable of milk-borne and *in vivo* catalysis of similar chemical species, there is compelling interest to further our understanding of BLG chemistry. Moreover, proteomic studies performed on eye tissue of macular degeneration patients have revealed the presence of BLG within drusen (retinal pigment epithelium extracellular deposits containing assorted lipids, polysaccharides, glycosaminoglycans and proteins),[78] setting up speculation of

BLG-assisted lipofuscin biosynthesis. While synthetic reactions employing proline and triethylamine as catalysts yield cycloretinal,[135] the reaction fails to proceed under standard physiological conditions in the presence of proline (**Appendix Fig. 67**). BLG-promoted cycloterpenal formation is probably not the sole physiological route to molecules like cycloretinal given the relatively slow rate observed for this protein-catalyzed process, but because the reaction does not appear to be spontaneous nor does it arise from reaction with endogenous proline, activation of retinoids *in vivo* logically requires assistance from another biomolecule. BLG may provide a plausible missing link in understanding the formation of lipofuscins from retinoids, but for the time being serves as a willing model to probe the mechanistic requirements for biosynthetic terpenal self-condensation reactions.

Focused epidemiological studies are necessary to draw more definitive links between dairy product consumption and the promotion of disorders characterized by accumulation of retinoid-derived lipofuscin fluorophores like cycloretinal and A2E. Unfortunately, such studies may be increasingly difficult to organize and manage as world populations historically devoid of dietary dairy intake begin to adopt milk products as a nutritious protein source. At the same time, increased world dairy consumption intensifies the need for investigations into the possible correlation between BLG and macular disease progression. AMD is poised to become increasingly prevalent in societies as average life expectancies exceed 80 years or more. Individuals with milk protein allergies or lactose intolerance may provide a population suitable for cross-analysis, but ultimately human studies toward this investigation will require decades of pursuit to obtain the most far-reaching and meaningful results.

EXPERIMENTAL PROCEDURES

Reagents and Instrumentation

All solvents and reagents were used without further purification unless noted. Experiments were carried out using BLG from Sigma (L3908, St. Louis, MO, 90% BLG by PAGE) or Davisco Foods International, Inc. (JE-003-6-922, La Sueur, MN, 93.6% BLG). NMR for ^{1}H (300 MHz) and ^{13}C (75 MHz) spectra were obtained on Varian Inova or Mercury300 NMR instruments. Mass data was collected on an API QSTAR PULSAR (ES) instrument. HPLC separations were performed using a Varian ProStar chromatography system.

General Procedure for BLG-Promoted Biosynthesis of Cycloterpenal Homodimers

In a 1000 mL Erlenmeyer flask, 500 mL of a 1% (w/v) BLG solution in phosphate buffered saline (PBS, 10 mM phosphate, 2.7 mM KCl, 137 mM NaCl, pH 7.0) and 3 equivalents of substrate (0.5 M in absolute ethanol, 1.630 mL) were combined. This mixture was placed in a shaker (37 °C, 250 rpm) for 4 days. The reaction was quenched by the addition of 500 mL of ethyl acetate and stirred for 30 min at room temperature to extract the product. To avoid gathering precipitated protein, the mixture was centrifuged at 9,000 rpm at 4 °C for 30 min. The organic layer was collected, dried over anhydrous magnesium sulfate, concentrated *in vacuo*, then purified and characterized according to Bench *et. al*.[76,124] (B.J.B)

Milk-Promoted Biosynthesis of Cycloterpenals

To demonstrate dimerization with processed milk, skim milk, 2% milk, whole milk and buttermilk were purchased from a local grocery market. The milk contained approximately 8-9 grams of protein according to their labels. Initial experiments were performed with 200 mL of milk, which was diluted with 300 mL of PBS and incubated with aldehyde substrate (2.20 mL of 0.5 M solution in absolute ethanol) at 37°C with shaking (250 rpm) for 4 days. Following incubation, the solutions were extracted with 300 mL of ethyl acetate, concentrated *in vacuo* and the metabolites analyzed using the methods described above.

To ensure that there were no small metabolites responsible for the dimerization process, each milk was passed through a Sephadex G-25 column (GE Healthcare, Uppsala, Sweden, Product # 17-0033-01). The G-25 resin (60 g) was swollen to 300 mL with PBS (pH 7.0) and packed into a glass column (5 x 30.5 cm). Milk (90 mL) was passed through the column followed by the addition of PBS until a total volume of 250 mL was obtained. The filtered milk was incubated with aldehyde substrate (1.10 mL of 0.5 M solution in absolute ethanol) and retested utilizing the same conditions as detailed above. (B.J.B. and J.F.A.)

***In Vivo* Study Utilizing New Zealand White Rabbits**

The following protocol (AUP#2008-70) was approved by the Texas A&M University Institutional Animal Care and Use Committee (IACUC). Eight adult New Zealand white rabbits (2.5 kg, Myrtle's Rabbitry, Thompsons Station, TN) were randomly allocated to the control (normal diet, n=1), BLG/water control (n=1), skim milk control (n=1), retinal control (n=1), BLG/water + retinal (n=2), and skim milk + retinal (n=2). All rabbits had free access to a standard rabbit diet and water. The rabbits were housed separately in standard cages in the

Laboratory Animal Resources and Research (LARR) facility at Texas A&M University and maintained under standard conditions. The rabbits were provided the solutions by gavage twice daily for a week. Each solution was administered in 50 mL volumes with the exception of the retinal control, which was supplied as a 1 mL solution (40 mg/kg; retinal was solubilized in 150 μL of ethanol). The BLG/water solution was supplied at 1.3 g/kg in 50 mL of water.

After seven days, the rabbits were anesthetized with a solution containing 10 mg/kg of ketamine with 3 mg/kg xylezine by intravenous injection allowing 5 min for the cocktail to take effect. Depth of anesthesia was monitored prior to blood removal by squeezing the foot of the rabbit. Cardiac blood was removed by opening the thorax to visualize the heart before carrying out cardiac puncture, which also results in exsanguination. This process allowed us to obtain approximately 40-50 mL of blood from each rabbit. The blood (40 mL) was processed by extraction. Ethyl acetate (300 mL) was added and stirred for 1 h. The suspension was centrifuged for 30 min. at 9,000 rpm, 4°C. The organic layer was removed, dried with anhydrous magnesium sulfate, and concentrated *in vacuo*.

To assess BLG content in the samples, blood from the control (normal diet), BLG/water control, BLG/water + retinal, skim milk control, and skim milk + retinal were centrifuged at 4,000 rpm for 30 min at 4°C to separate red blood cells from serum. After centrifugation, 1 μL of serum was diluted with 9 μL of PBS pH 8.0 and 10 μL of SDS-PAGE Buffer (125 mM Tris pH 6.8, 4% SDS, 20% glycerol, 0.2 mg/ml bromophenol blue, 0.2 mM DTT). Samples were heated to 90°C for 10 min and analyzed by SDS-PAGE (15%, 200V for 35 min) stained with Coomassie blue. (B.J.B.)

Determination of Optimal Protein to Substrate Ratio

To determine the optimal protein and substrate concentrations for dimer formation, a range of protein to substrate ratios from 1:1 to 1:10 were tested. A 1% (w/v) BLG solution (5 g) was prepared in PBS (500 mL, pH 7.0). An aliquot of 0.5 M citral in ethanol was added to the BLG solution to achieve the desired ratio, and the reaction was incubated at 37°C, 250 rpm for 4 days. The mixture was extracted, concentrated, and analyzed as outlined in the general procedure for BLG-promoted biosynthesis. (B.J.B.)

pH Profile Analysis

To examine the behavior of the BLG-promoted reaction as a function of pH, we evaluated the reaction at pH values of 1, 3, 5, 7, 9, and 11. PBS was either buffered with 1 M

HCl or 1 M NaOH to reach the desired pH. BLG (5 g) was dissolved in each buffered PBS solution to give a 1% (w/v) solution and incubated with 3 equivalents of citral (0.5 M in ethanol, 1.630 mL) at 37°C, 250 rpm for 4 days. The BLG mixture was extracted and analyzed by ^{1}H-NMR. (B.J.B.)

HPLC Profile of BLG-Promoted Biosynthesis of Cycloterpenals

The products of a 500 mL BLG-retinal reaction were extracted as previously described. An aliquot of the extract (20%) was dissolved in 40 µL dichloromethane and analyzed by HPLC (Phenomenex Luna silica column, 5 µm particle size, 250 x 10.00 mm) with a dichloromethane mobile phase, 3 mL/min flow rate, and UV absorption monitoring at 254 nm. Additionally, an extract of 1% (w/v) BLG in 500 mL PBS, synthetic (proline-catalyzed) cycloretinal, and an aliquot of the BLG-retinal reaction extract spiked with 0.25 mg synthetic cycloretinal were analyzed in the same manner. (J.F.A.)

Absorbance Spectroscopy of BLG-Promoted Cycloterpenal Biosynthesis

Solutions of 1% (w/v) BLG in PBS pH 7.4 (5 mL) were incubated with 3 molar equivalents of either citral or all-*trans*-retinal in glass culture tubes (16 x 125 mm) with metal caps at 37°C with agitation at 250 rpm. Reactions containing all-*trans*-retinal were protected from ambient light at all times to prevent photo-induced degradation. At each time point, 750 µL of reaction was transferred to a 1.5 mL polypropylene microcentrifuge tube and extracted twice with an equal volume of ethyl acetate by vortexing, centrifuging for 3 min at 14000 rpm, and removal of the organic layer. A stream of nitrogen was used to evaporate the solvent, and the residue was subsequently resuspended in absolute ethanol for UV-visible absorbance scanning using a quartz cuvette in a Genesys 2 UV-Vis spectrophotometer (ThermoFisher Scientific, Waltham, MA) with ethanol serving as the absorbance blank. (J.F.A.)

Microbial Protein Crude Extract Controls

E. coli DH10B was cultured on agar plates using standard methods. A single colony was picked and placed in 3 mL of LB medium and allowed to grow overnight before being transferred to 500 mL of fresh medium. After 24 h, the culture was centrifuged and the cell pellet flash frozen in liquid nitrogen before being stored at -80°C. Cells were lysed with a bead mill (Bead Beater, Biospec Products, Bartlesville, OK) equipped with 0.1 mm glass beads. Frozen cell pellets, PBS, and glass beads were added to the mill chamber on ice. Cells were

lysed with ten (30 sec) pulses with 1 min cooling intervals. The lysate was centrifuged and the resulting supernatant served as the crude extract. Protein concentration was measured by Bradford assay (Bio-Rad, Hercules, CA) and diluted to a 1% (w/v) solution with PBS.

Prior to carrying out the dimerization assay, a metabolite extract blank was generated. The protein extract (40 mL) was extracted with 200 mL of ethyl acetate, dried with anhydrous magnesium sulfate, and concentrated *in vacuo*. A ^{1}H-NMR spectrum was obtained to ensure that there were no metabolite peaks in the aldehydic region of interest. Results were also confirmed by mass spectrometry. To establish whether crude protein extracts could support cycloterpenal formation, protein lysates were incubated with citral at 37°C for 4 days in an orbital shaker (250 rpm). As detailed previously, following incubation, the suspension was extracted and analyzed. (B.J.B.)

Evaluation of RPE Cell Extract for Cycloterpenal Formation

RPE cells (ARPE-19, ATCC) were subcultured to give a total of 10 plates at 95% confluency. Cells were propagated in ATCC-formulated Dulbecco's modified Eagles medium (DMEM:F12) supplemented with 10% fetal bovine serum (FBS) in a 5% carbon dioxide atmosphere at 37°C. Cells were dislodged from the plate with 3 mL of a 0.05% (w/v) Trypsin-EDTA solution and transferred to a 50 mL conical tube. Cells were pelleted by centrifugation for 20 min. at 1,000 rpm. To remove traces of medium, the cells were washed twice with DPBS, (1X without calcium or magnesium), and subsequently resuspended in 10 mL of DPBS. The cell suspension was lysed with a 50 mL dounce homogenizer and the resulting lysate centrifuged at 5,000 rpm (4 °C) for 30 min to pellet cellular debris. Protein concentration was measured by Bradford assay and diluted to a 1% solution with PBS buffer.

A metabolite control was generated by extracting the protein lysate (5 mL) with ethyl acetate (50 mL), which was dried over anhydrous $MgSO_4$, concentrated *in vacuo*, and evaluated by ^{1}H NMR spectroscopy. To establish whether the crude protein extract (10 mL) could support cycloterpenal formation, the protein lysate was incubated with citral at 37 °C with shaking (250 rpm) for 4 days, extracted and analyzed in a manner identical to the microbial cell extract. (B.J.B.)

SDS Denaturation of BLG

In a 1000 mL Erlenmeyer flask, 500 mL of a 1% BLG solution (PBS, pH 7.0) was stirred with 10% SDS (w/v, 50 g) for 2 h. Three equivalents of citral (0.5 M in absolute ethanol,

1.630 mL) was added to the SDS-protein solution. The mixture was placed in a 37°C shaker and incubated at 250 rpm for 4 days. The BLG mixture was extracted and analyzed as detailed previously. (J.F.A.)

Evaluation of Cycloretinal Formation Under Physiological Conditions

L-proline (54.4 mg) was solubilized in 10 mL of buffer (10 mM phosphate, 2.7 mM KCl, 137 mM NaCl; pH 7.4) to which was added dropwise retinal (10 mg/1 mL ethanol). The reaction was incubated at room temperature (in the dark) for 11 h and subsequently extracted with 50% ethyl ether in hexane. The organics were dried over magnesium sulfate and concentrated *in vacuo* prior to NMR analysis. (C.M.H.W.)

CHAPTER V

THE MECHANISM OF β-LACTOGLOBULIN-MEDIATED CYCLOTERPENAL BIOSYNTHESIS

INTRODUCTION

Our previous studies with α,β-unsaturated aldehydes and β-lactoglobulin (BLG) indicate catalyst involvement is necessary to mediate self-condensation to yield ring-fused dimers. In the synthetic sense, L-proline, with the assistance of an organic base in some instances, provides activation of aldehydes via formation of a Schiff base. An analogous mechanism to support self-condensation by a protein scaffold would likely involve lysine residues, an enzymatic proposal exemplified by rhodopsin conjugated to retinilidine through an imine bond to ready for the photoisomerization step that invokes sight in the vision cycle (**Fig. 47**). While rhodopsin conjugation does not directly assist in further chemical modifications to retinilidene, the reactivity of an amine and aldehyde in a biological context is proven.

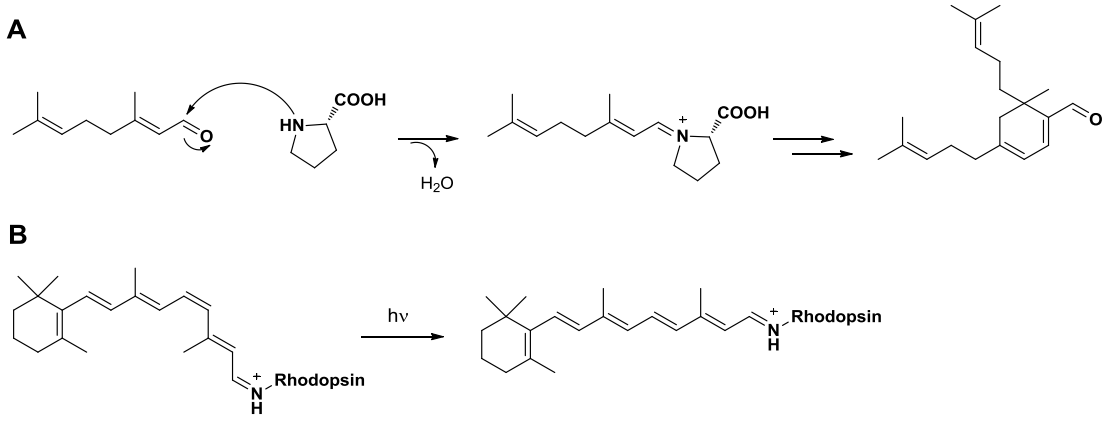


Figure 47 Aldehydes converted to imine conjugates in biologically relevant reactions.
(A) L-proline synthetic activation of citral. (B) Rhodopsin-conjugated 11-*cis*- and all-*trans*-retinilidine in the vision cycle.

Activated aldehydes, whether by proline or protein-based catalysts, may undergo condensation through one of two possible pathways, either via stepwise Michael-like imine addition or a concerted Diels-Alder-type mechanism (**Fig. 48**). Current kinetic isotope

experiments being carried out suggest a concerted route to ring formation, but additional evidence is required to verify this preliminary observation.

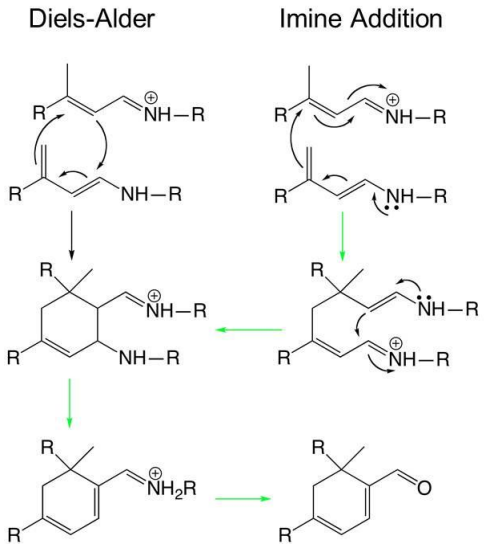


Figure 48 Proposed reaction mechanisms for imine-activated α,β-unsaturated aldehyde condensation.

Given these precedents for aldehyde reactivity, we sought to determine if a pair of lysine residues in BLG is directly involved in promoting cycloterpenal formation. Conceivable permutations to effect catalysis include participation of two different lysine residues within a monomer subunit or a single lysine residue acting in tandem with another residue on a separate subunit of a dimer configuration. BLG contains 15 lysine residues, and while the crystal structure (**Fig. 49**) suggests certain residues may be better situated for aldehyde conjugation than other residues (e.g., residues that are solvent exposed versus buried), we conducted an evaluation of the potential role each lysine could play in assisting self-condensation using site-directed mutagenesis and trypsin-digest proteomic mass spectrometry.

This chapter details cloning and expression of soluble BLG in an *E. coli* host, site-directed mutagenesis rationale, and preliminary peptide mass spectrometry data to support involvement of lysine residues in the formation of Schiff bases to facilitate dimer condensation reactions. It also documents efforts to screen lysine mutants for loss of activity using HPLC separation of reaction extracts and mass spectrometry or UV-visible spectroscopy.

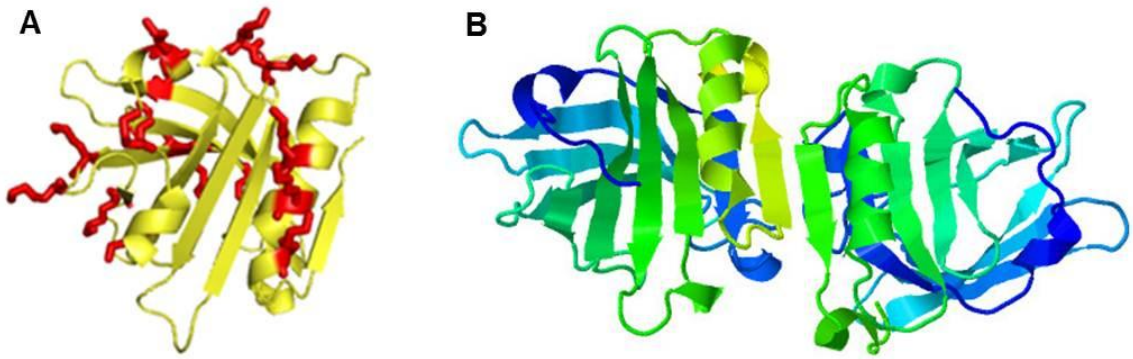


Figure 49 X-ray crystal structure cartoon diagrams of bovine BLG.
(A) PDB structure 1B8E, monomer subunit with lysine residue side chains are highlighted in red. (B) PDB structure 1B8E, native dimer.

RESULTS AND DISCUSSION

Proteomic Mass Spectrometry to Probe Catalytic Residues

The initial approach to detecting specific lysine residue involvement in ring-fused dimer formation catalyzed by BLG involved trapping the Schiff base formed between lysine and the aldehyde substrate citral by reducing the imine bond with sodium cyanoborohydride. Although retinal is a more biologically-relevant substrate, it is also light and temperature-sensitive, so we employed citral as a more stable and inexpensive model substrate to study BLG chemistry. Once the citral-lysine Schiff base is covalently stabilized by reduction, the protein-substrate complex was fragmented by trypsin digest to obtain short peptides for analysis by electrospray ionization mass spectrometry (ESI-MS). An additional round of fragmentation yielded spectral signatures for specific peptides predicted to be obtained by trypsin digest. MASCOT-assisted prediction of peptide sequences with expected mass to charge ratios, adjusted to reflect possible added mass of citral or cyclocitral covalently bound to a lysine residue within a peptide, were used as search parameters in the tandem MS/MS data set. Mining of fragment signatures revealed peptides containing residues K77 and K91 with citral substrate bound, and residue K77 with the product cyclocitral bound (**Fig. 50**).

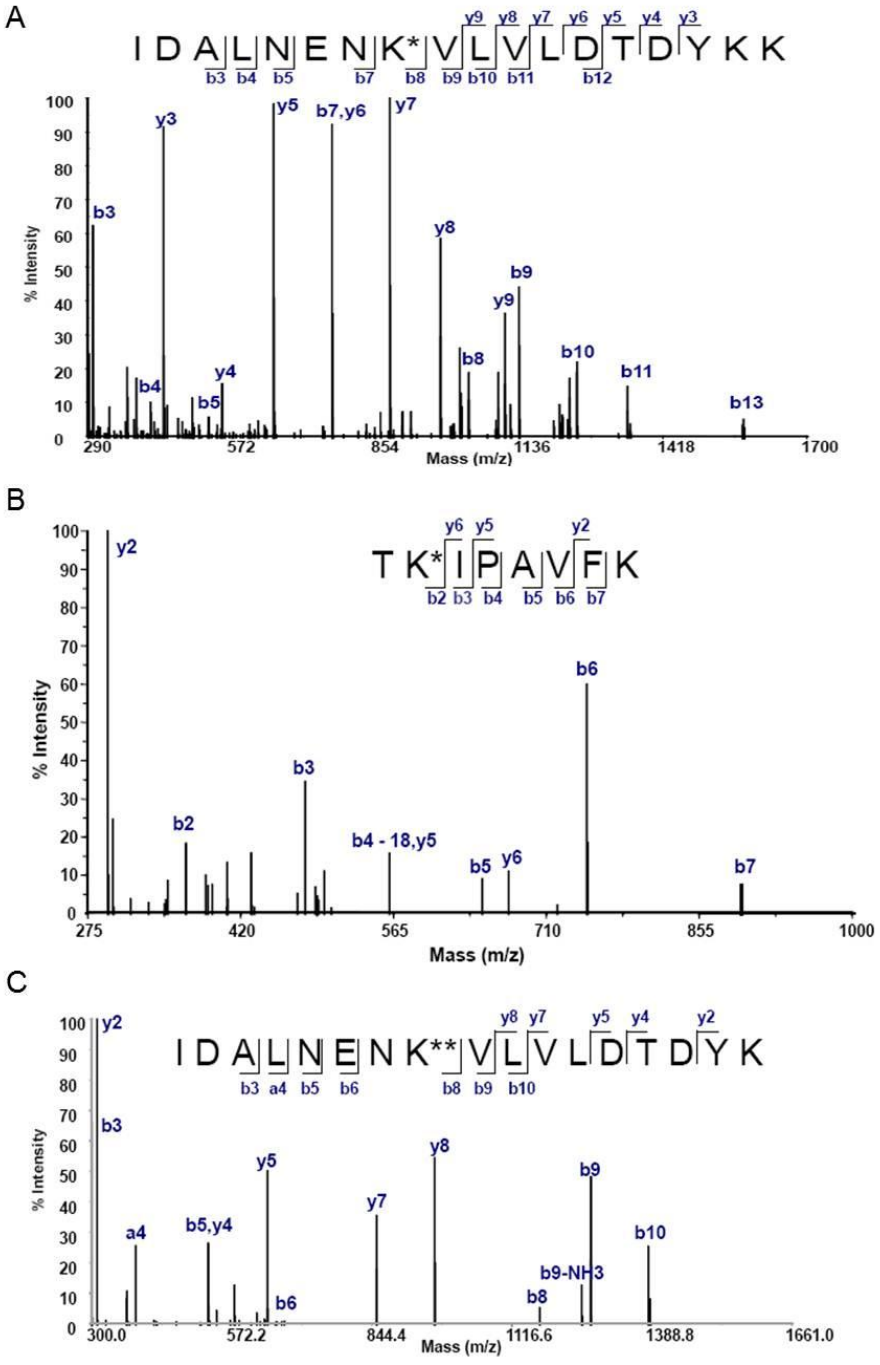


Figure 50 Mass spectra of peptides containing lysine residues with citral or cyclocitral bound.
(A) K77 with citral bound, denoted by K* in the peptide sequence. (B) K91 with citral bound, denoted by K* in the peptide sequence. (C) K77 with cyclocitral bound, denoted by K** in the peptide sequence. Figures were produced by Dr. B. J. Bench and Dr. William Russell.

Residues K77 and K91 are situated within plausible proximity (~12 Å, based on crystal structure 1B8E) to guide condensation (**Fig. 51A**). K91 would be considered fairly immobile on the outer wall of the β-barrel, but K77 resides on a flexible loop structure. Thus, it is reasonable to suggest K77 could move near K91 to facilitate the intended reaction. However, the participation of these residues cannot be viewed as definitive because the possibility of random substrate binding exists, especially considering the excess citral used to achieve target residue saturation. In a standard extraction of BLG incubated with citral, the aqueous phase retains a deep orange color (indicative of a citral-protein Schiff base) even after exposure to organic solvent. Additional fragments identified in the mass spectrometry experiment corroborate random covalent binding of citral to lysine residues throughout the protein. Residues K60 and K69 are perhaps more optimally situated to facilitate substrate condensation, as they reside on anti-parallel sheets within the β-barrel interior wall within 4.8 Å of one another. Additionally, K60 and K69 are members of the central calyx known to bind lipophilic species (**Fig. 51B**) and the residue pair exists in a solvent-protected environment theoretically more amenable to retinoid condensation. Despite the ambiguity presented by binding studies and proteomic mass spectrometry, the results provide a starting point for evaluation of BLG lysine mutants to verify the role of specific residues in catalysis. Mass data can be re-mined to produce evidence of binding should mutant studies indicate participation of other lysine residues.

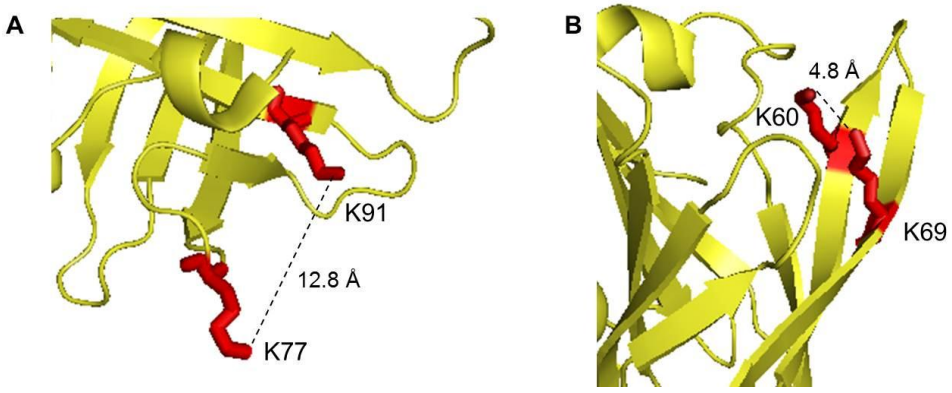


Figure 51 BLG lysine residue pairs postulated to be involved in cycloterpenal catalysis.
(A) Residues K77 on the flexible loop and K91 on the β-barrel exterior. (B) Residues K60 and K69 on the β-barrel wall interior. Lysine side chains are highlighted in red.

Expressing Soluble and Active BLG in *E. coli*

A survey of recombinant BLG literature reveals significant difficulty expressing soluble and properly folded protein in the most common bacterial host, *E. coli*.[136-139] For our purposes, we had to overcome an additional barrier of low-yielding cultures, as ring-fused dimer formation is slow and requires gram quantities of protein to generate sufficient product for analysis. While soluble BLG expression is successful in yeast systems like *Pichia pastoris*,[140] cloning, transformation, and expression is much more time consuming than in *E. coli*, rendering *P. pastoris* suboptimal for generating numerous lysine mutants. In our laboratory, attempts to clone *E. coli* codon-optimized BLG in *E. coli* as a His_6 variant in popular pET or pQE vectors employing T7 or lac promoters failed to produce soluble protein even at reduced temperature. After contemplating the variety of chaperones and fusion proteins/peptides available to mitigate protein solubility problems inherent in many *E. coli* expression systems, the maltose-binding protein (MBP) was chosen as a fusion candidate for BLG. In our hands the MBP-BLG fusion construct dubbed pMAL-BLG proved highly rewarding, generating soluble fusion protein in excess of 100 mg/L under a *tac* promoter when induced with 0.3 mM isopropyl-thiogalactopyranoside at 16°C for 20 h.

The purified fusion protein was assessed for activity comparable to wild type BLG by incubation with citral and subsequent HPLC analysis of the reaction extract. Because the initial aim was to simply determine qualitative catalytic behavior, we opted to assay the protein as a fusion. Additionally, Factor Xa cleavage of MBP-BLG was observed to be an inefficient process and was impractical to carry out at gram scale. Comparison of chromatogram peaks confirmed the MBP-BLG fusion did not prevent formation of cyclocitral, although it was somewhat less efficient than wild type BLG as judged by peak area (**Fig. 52**).

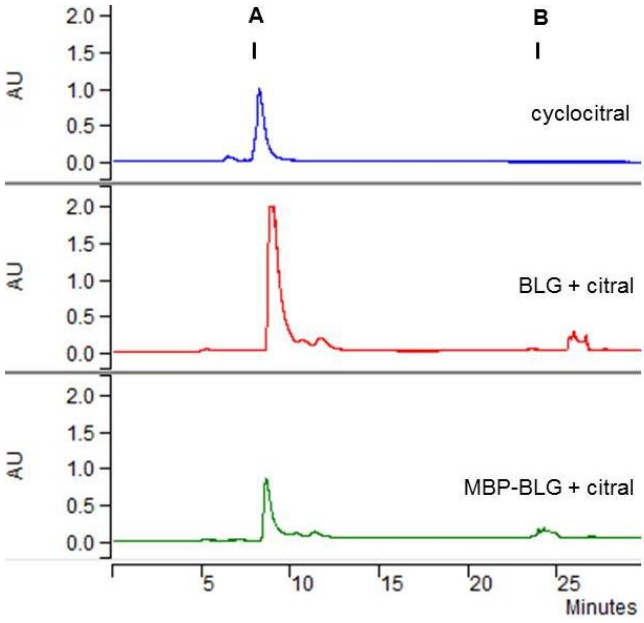


Figure 52 MBP-BLG and wild type BLG activity comparisons as judged by HPLC product analysis.
Peak (A) is cyclocitral and peak (B) is excess citral. Due to the detector wavelength setting at 320 nm, absorbance attributed to cyclocitral is maximized while citral absorbance is minimal, and thus the traces do not reflect relative actual concentrations present in the extract mixture.

For additional product confirmation, extracts could be derivatized with dansyl hydrazine (dns) (**Fig. 53A**) to provide a better candidate for mass spectrometry (MS). The single aldehyde functionality of cyclocitral is tricky to ionize and was thus difficult to reliably detect by routine methods. Addition of dns via simple acid-catalyzed aldehyde-hydrazine coupling provides additional functionality and the resulting dns-cyclocitral adduct can be detected by ESI-MS (**Fig. 53B** and **C**) in the nanogram range.

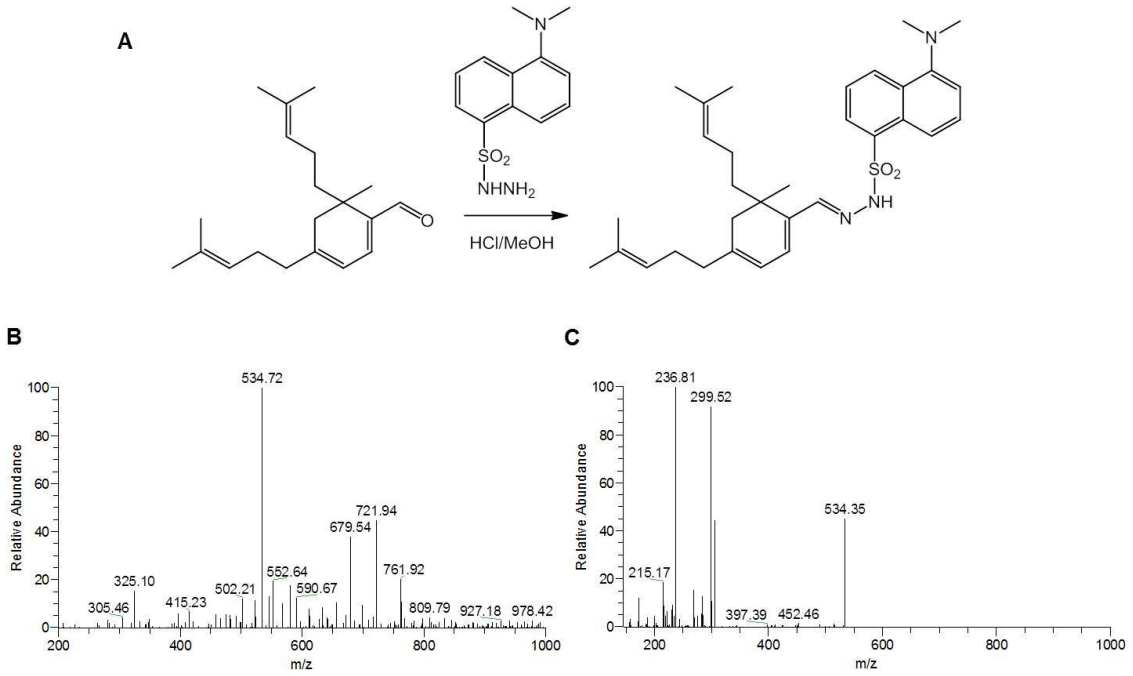


Figure 53 Cyclocitral derivatization for MS detection.
(A) Reaction between cyclocitral and dansyl hydrazine. (B) Principal dns-cyclocitral ion, m/z 534. (C) Subsequent ionization fragments of the m/z 534 species.

Monomeric Versus Dimeric BLG

As bovine BLG exists in equilibrium as both monomer and dimer in solution, we also sought to determine if activity is contingent to multimeric status. The BLG homolog from horse's milk (*Equus caballus* BLG I, accession NP_001075962, eBLG) is known to exist primarily as a monomer in solution,[141] contrary to most other BLG orthologs studied to date. An *E. coli* codon-optimized eBLG gene was cloned as a maltose-binding protein fusion to generate a highly-expressed and easily purified protein analogous to the bovine BLG-MBP fusions mentioned previously. Native PAGE analysis confirmed MBP-eBLG exists primarily as a monomer (with two apparent isoforms), while the MBP-BLG wild type and lysine mutants adopt a monomer-dimer equilibrium with some degree of oligomerization (**Fig. 54**). Equine BLG is not a perfect monomeric equivalent to bovine BLG due to lack of lysine residues at 9 of the 15 positions (residues 8, 14, 47, 60, 77, 91, 101, 138, and 141). However, eBLG provides one of the only known options for analysis of monomer behavior. Mutants of bovine BLG at residues believed to support electrostatic interactions of the dimer interface have been reported to be monomers based on ultracentrifugation association measurements,[142] but the mutants R40D (AB

loop mutant), H146P (I strand mutant), and R40D/H146P (AB loop/I strand double mutant) constructed (**Appendix Table 14**) and tested in our MBP fusion expression system all appeared to maintain a monomer-dimer equilibrium similar to wild type (**Appendix Fig. 71**) and thus were not viable options for strict monomer catalytic comparison.

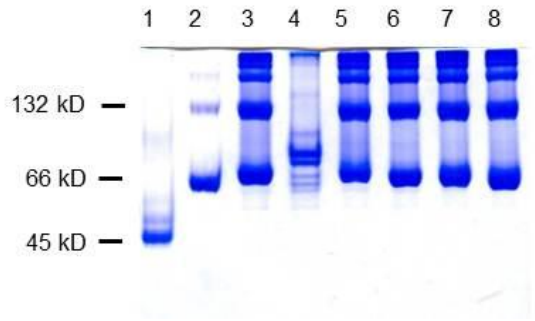


Figure 54 Native PAGE analysis of purified MBP-eBLG, MBP-BLG wild type and lysine mutants.
Lane 1, ovalbumin standard (45 kDa); lane 2, BSA standard (66 kDa monomer, 132 kDa dimer); lane 3, MBP-BLG; lane 4, MBP-eBLG; lane 5, MBP-BLG K69A; lane 6, MBP-BLG K77A; lane 7, MBP-BLG K91A; lane 8, MBP-BLG K77A/K91A double mutant.

MBP-eBLG is not expressed in *E. coli* with an efficiency equal to MBP-BLG. Identical preparations yield eBLG fusion in amounts that are only 10% of that acquired from BLG fusion purification. Suprisingly, incubation of citral and MBP-eBLG supported cyclocitral formation in quantities equivalent to that derived from incubation with MBP-BLG fusion (**Fig. 55**), but at a protein concentration of 0.1% (w/v), which is 10-fold less protein than the standard 1% (w/v) bovine BLG used in a routine incubation.

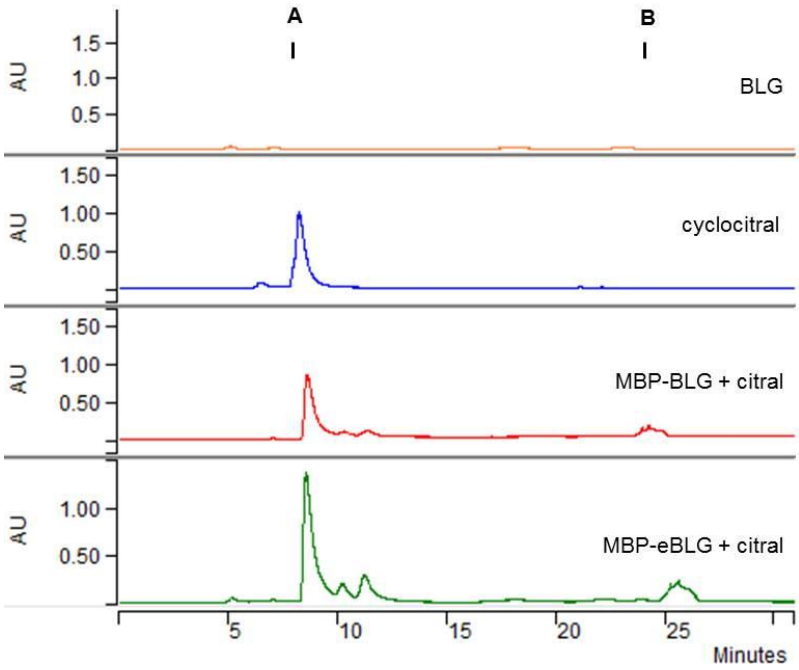


Figure 55 HPLC trace comparison of reaction extracts from bovine and equine BLG.
Peak (A) is cyclocitral and peak (B) is excess citral. Due to the detector wavelength setting at 320 nm, absorbance attributed to cyclocitral is maximized while citral absorbance is minimal, and thus the traces do not reflect relative actual concentrations present in the extract mixture.

BLG Lysine Mutants

To identify lysine residues suspected to contribute to the catalytic activity of BLG, standard site-directed mutagenesis was employed.[143] Replacement of individual lysine residues with alanine within the pMAL-BLG construct conceivably eliminates the potential for Schiff base formation with an aldehyde substrate at a specific position. A total of 15 individual lysine to alanine point mutants at positions 8, 14, 47, 60, 69, 70, 75, 77, 83, 91, 100, 101, 135, 138, and 141 were prepared for evaluation of cyclocitral formation. Additional double mutants at residue pairs K60/K69 and K77/K91 were also assayed after considering both the proteomic mass spectrometry results and the residues that were most likely to be involved due to retinoid binding site proximity.

HPLC Analysis of BLG Mutant Activity

Reaction scale was designed to balance the need for maximum product formation with reasonable culture volumes and protein purification suited to laboratory scale. Thus, a 30 mL

reaction volume of phosphate-buffered saline (PBS) pH 7.4 containing 1 g of MBP-BLG fusion protein (the molar equivalent to a 1% w/v wild type BLG solution) was deemed sensible, which demanded purification of a 10 L culture for each mutant to be assayed. Following a standard 4 day incubation with citral, the extract was analyzed by HPLC to detect the presence or absence of the expected product cyclocitral (**Fig. 56**).

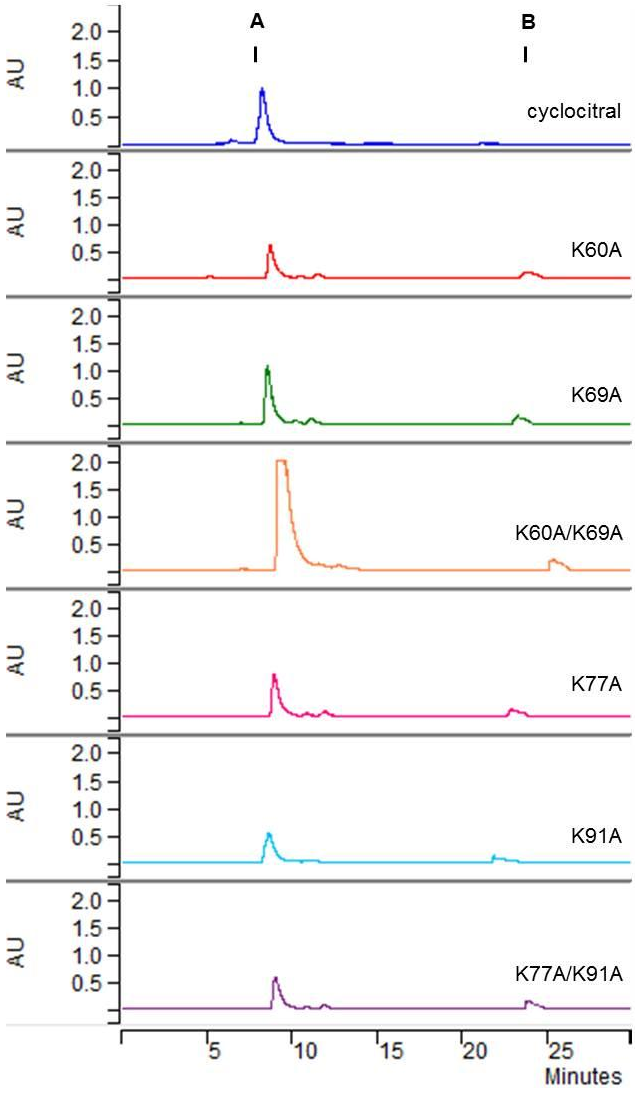


Figure 56 HPLC analysis of BLG lysine mutant activity.
Peak (A) is cyclocitral and peak (B) is excess citral. Due to the detector wavelength setting at 320 nm, absorbance attributed to cyclocitral is maximized while citral absorbance is minimal, and thus the traces do not reflect relative actual concentrations present in the extract mixture.

Each of the assayed mutants (K60A, K69A, K77A, K91A, and their respective paired double mutants) promoted cyclocitral formation similar to that of wild type MBP-BLG. Judging from relative peak intensities, the K60A/K69A double mutant performed exceptionally well in catalysis. It is plausible to suggest the removal of K60 and K69 frees up space inside the protein calyx for easier substrate entry and exit. Unfortunately, these results disprove the proposed catalytic mechanism involving appropriately situated lysine residues. The remaining MBP-BLG lysine mutants are currently being tested for activity. We have also begun analysis using available crystal structures to explore the possibility of other residues in or near the calyx (such as residues N63, N109, D28, D33, E62, or E89) that may experience a local charge environment appropriate for facilitating aldehyde activation and nucleophilic attack.

SIGNIFICANCE

The identification of specific residues involved in BLG catalysis of α,β-unsaturated aldehydes to form ring-fused dimers is useful in determining the greater biological implications of BLG present in the body from ingestion of dairy products. Because BLG has been detected in the human eye proximal to tissues containing retinal-derived fluorophores linked to maculopathies, the importance of understanding the mechanism by which BLG promotes formation of cycloterpenals may lead to methods for circumventing these reactions, thus lowering disease risk. The option for development of active site inhibitors requires investigations such as those undertaken in this study. While a definitive correlation between milk consumption, BLG-promoted cycloterpenal formation, and the progression of diseases such as age-related macular degeneration will likely remain unproven for some time, understanding this unusual activity serves as a springboard toward studies in substrate promiscuity, inhibitor design, and preparation for managing therapeutics to target this potential disease mechanism.

As the principle whey protein in milk, BLG is expected to play a distinct role in nutrition although that function remains slightly ambiguous. The catalytic actions carried out by a protein for which a definite role remains unassigned may provide clues toward explaining the evolution and presence of BLG in many mammalian milks, and spur more comparative ortholog studies beyond the most studied β-lactoglobulin, bovine BLG. Furthermore, because there have not been protein candidates of human origin identified in promoting catalysis of retinal-derived dimers, the BLG model may prove useful in identifying endogenous proteins that aid formation of lipofuscins. While a number of high quality crystal structures containing lipophilic species

within the protein calyx exist,[144-147] the images do not convey covalent linkage nor do they prove that catalytic activity must be confined to that site. Determination of active site architecture, specifically from crystal structures with covalently-bound retinoid substrates, could be used as a guide for such investigations. And in combination with planned top-down proteomic mass spectrometry (rather than the bottom-up method described in this study), support for an alternative active site residue model should be within reach.

EXPERIMENTAL PROCEDURES

BLG-Citral Trapping and Protein Mass Spectrometry

The BLG-citral trapping described here was performed by Dr. B. J. Bench. Trypsin digest and liquid chromatography-mass spectrometry (LC-MS) and the associated data mining procedures were carried out by Dr. William K. Russell of the Laboratory for Biological Mass Spectrometry at Texas A&M University. One set of BLG solutions (1 mL, 1% w/v) were incubated with 4 molar equivalents of citral and 10 molar equivalents of $NaCNBH_3$ at 37°C for periods of 6, 24, 48, 72, and 96 h. A second set of BLG solutions were identical except that reduction with $NaCNBH_3$ was instated after the initial incubation period indicated above. Two 75 μL aliquots of each time point were desalted using a Micro Bio-Spin P30 column (BioRad, Hercules, CA). Protein concentration was adjusted to 0.1 mg/mL with $(NH_4)_2HCO_3$ buffer and reduced with 5 mM dithiothreitol at 60°C for 1 h. Subsequent protein alkylation was achieved using 20 mM iodoacetamide at room temperature for 10 min. Protein samples were then digested with trypsin overnight (protein:enzyme ratio of 1:50) at 37°C. Separation and mass spectrometry were carried out on a NanoFrontier LC-MS (Hitachi High Technologies, Dallas, TX) equipped with a nanospray ESI source. A 200 ng peptide sample was separated on a Vydac C18 capillary column (Grace Davison Discovery Sciences, 150 x 0.075 mm) at a flow rate of 200 nL/min under the following gradient routine employing water/acetonitrile and 0.1% formic acid in all conditions: 2% water/acetonitrile for 5 min, 2-10% over 0.1 min, 10-40% over 29.9 min, 40-60% over 10 min, 60-98% over 5 min, 98% for 6 min, 98-2% over 1 min, 2% for 13 min. MASCOT-assisted predictions for modified lysine and carboamidomethyl groups on peptide fragments were used in manual examination of tandem MS/MS data. Final spectra were produced by deconvolution to show the *m/z* +1 peaks and labeled to indicate the ion fragments resulting from b and y-type peptide cleavage.

Maltose Binding Protein-BLG Fusion Cloning, Overexpression, and Purification

An *E. coli* codon-optimized version of *Bos taurus* BLG variant B (accession CAA88303) was synthesized and ligated to the TA cloning vector pQE-30UA (Qiagen, Valencia, CA) by GenScript (Piscataway, NJ) to generate the plasmid pQE30-BLG. The maltose binding protein-BLG fusion was produced by PCR with the primers BLG-EcoRI and BLG-HindIII using pQE30-BLG template with Phusion DNA Polymerase. After agarose gel purification, the 504 bp PCR product was digested with *EcoR*I and *Hind*III, dephosphorylated with calf alkaline phosphatase, and ligated to the vector pMAL-c4x (New England Biolabs) using T4 DNA ligase at room temperature to give the vector pMAL-BLG (**Appendix Fig. 69**). Following transformation in *E. coli* DH10B, colonies were screened by PCR to identify positive clones and sequenced using pMALseqF and pMALseqR primers to verify no mutations had occurred.

A single colony of pMAL-BLG(DH10B) was grown 16 h at 37°C in 25 mL Luria-Bertani (LB) Miller broth and used to inoculate 1 L LB Miller broth containing 0.2% glucose to suppress endogenous amylases. The culture was shaken at 37°C to an optical density at 600 nm of 0.5, then cooled to 16°C and induced with 0.3 mM isopropyl β-thiogalactopyranoside for 20 h. The cells were pelleted (7000 rpm, 10 min), resuspended in 35 mL column buffer (20 mM NaH_2PO_4 pH 7.5, 200 mM NaCl, 1 mM EDTA, 10% glycerol, 1 mM phenylmethylsulfonyl fluoride), and stored frozen at -80°C.

The cell suspension was thawed and sonicated on ice using a Branson Sonifier 450 fitted with a 5 mm microtip (10-30 s pulses at 50% duty cycle, output setting 6, with 2 min cooling intervals). Debris was pelleted at 9800 rpm for 30 min and the supernatant filtered to 0.2 μm and diluted with buffer to 180 mL before applying to 25 mL amylose resin slurry (New England Biolabs) packed in a gravity-flow column (Kontes Flex-Column, 2.5 x 10 cm) equilibrated with column buffer. The run was conducted at ambient temperature using buffers on ice. Following a wash with 12 column volumes of buffer (180 mL), MBP-BLG was eluted with column buffer containing 10 mM maltose (60 mL). The eluate was ultrafiltered to a final concentration of 100 mg/mL (Amicon Ultra-15 Centrifugal Filter Unit, Millipore, Billerica, MA) and maintained at 4°C for short-term use or was flash-frozen in liquid nitrogen for long-term storage at -80°C. MBP-BLG fusions were reliably produced and isolated in excess of 100 mg/L culture under the described conditions.

Purified protein was quantitated by the Bradford method[102] using wild-type bovine BLG (Sigma-Aldrich, >90% by PAGE) as a standard for the linear range of 500 μg/mL to 5 mg/mL in

a microassay. It was noted that when the traditional Bradford assay standard bovine serum albumin was used the assay indicated fusion protein concentrations were ~10-fold less than that estimated from A_{280} measurements and SDS-PAGE analysis.

Site-Directed Mutagenesis

Lysine to alanine point mutations were introduced in pMAL-BLG using the QuikChange II Site-Directed Mutagenesis Kit (Stratagene, Cedar Creek, TX) with slight modification. Each 50 µL mutagenesis PCR reaction contained 50 ng pMAL-BLG DNA, 125 ng of each mutagenic primer (**Table 13**), 1x reaction buffer, 1 µL dNTP mix, and 1 unit of Phusion High-Fidelity DNA Polymerase (Finnzymes Oy, Espoo, Finland). PCR conditions consisted of 16 cycles of 95°C for 30 s, 55°C for 1 min, and 68°C for 7 min. *Dpn*I digest of methylated template DNA was carried out overnight at 37°C. The digestion reaction was purified using the MinElute PCR Purification Kit (Qiagen, Valencia, CA) and transformed in *E. coli* DH10B for mutant fusion expression. Gene mutations were confirmed by DNA sequencing using primers pMALseqF and pMALseqR, which are complementary to the plasmid sequence immediately preceding or following the BLG gene. Several of the site-directed mutagenesis PCR reactions and sequencing confirmations were carried out by Hillary C. Agbo.

Double point mutants of pMAL-BLG at K77A/K91A and K60A/K69A were constructed in an identical manner except that the template DNA consisted of one of the respective mutants and the primers were such that they introduced the second respective lysine-to-alanine point mutation.

Table 13 MBP-BLG cloning and site-directed mutagenesis primer sequences.

Primer	**Sequences** (5'-3')[a]
BLG-EcoRI	TATAGAATTCCTGATTGTGACCCAGACCATGA
BLG-HindIII	TACCAAGCTTTTAAATGTGACACTGTTCTTCC
pMALseqF	TCGATGAAGCCCTGAAAGAC
pMALseqR	GTGCTGCAAGGCGATTAAGT
K8A sense	GATTGTGACCCAGACCATGGCAGGCCTGGATATTCAGAAA
K8A antisense	TTTCTGAATATCCAGGCCTGCCATGGTCTGGGTCACAATC
K14A sense	ATGAAAGGCCTGGATATTCAGGCAGTGGCGGGTACCTG
K14A antisense	CAGGTACCCGCCACTGCCTGAATATCCAGGCCTTTCAT
K47A sense	GTGTGTATGTTGAAGAACTGGCACCGACCCCGGAAGG
K47A antisense	CCTTCCGGGGTCGGTGCCAGTTCTTCAACATACACAC
K60A sense	TCTGGAAATTCTGCTGCAGGCATGGGAAAACGGCGAATGC
K60A antisense	GCATTCGCCGTTTTCCCATGCCTGCAGCAGAATTTCCAGA
K69A sense	GAAAACGGCGAATGCGCGCAGGCAAAAATTATCGCGGAAAAAACC
K69A antisense	GGTTTTTTCCGCGATAATTTTTGCCTGCGCGCATTCGCCGTTTTC
K70A sense	CGGCGAATGCGCGCAGAAAGCAATTATCGCGGAAAAAACC
K70A antisense	GGTTTTTTCCGCGATAATTGCTTTCTGCGCGCATTCGCCG
K75A sense	CGCAGAAAAAAATTATCGCGGAAGCAACCAAAATTCCGGCGGTG
K75A antisense	CACCGCCGGAATTTTGGTTGCTTCCGCGATAATTTTTTTCTGCG
K77A sense	AGAAAAAAATTATCGCGGAAAAAACCGCAATTCCGGCGGTGTTTAAAATTG
K77A antisense	CAATTTTAAACACCGCCGGAATTGCGGTTTTTTCCGCGATAATTTTTTTCT
K83A sense	AACCAAAATTCCGGCGGTGTTTGCAATTGATGCGCTGAATGAAAAC
K83A antisense	GTTTTCATTCAGCGCATCAATTGCAAACACCGCCGGAATTTTGGTT
K91A sense	GTTTAAAATTGATGCGCTGAATGAAAACGCAGTGCTGGTGCTGGATAC
K91A antisense	GTATCCAGCACCAGCACTGCGTTTTCATTCAGCGCATCAATTTTAAAC
K100A sense	GCTGGTGCTGGATACCGATTATGCAAAATATCTGCTGTTTTGCATG
K100A antisense	CATGCAAAACAGCAGATATTTTGCATAATCGGTATCCAGCACCAGC
K101 sense	GTGCTGGTGCTGGATACCGATTATAAAGCATATCTGCTGTTTTGCA
K101 antisense	TGCAAAACAGCAGATATGCTTTATAATCGGTATCCAGCACCAGCAC
K135 sense	GTGGATGATGAAGCCCTGGAAGCATTTGATAAAGCGCTGAAAGC
K135 antisense	GCTTTCAGCGCTTTATCAAATGCTTCCAGGGCTTCATCATCCAC
K138A sense	GATGATGAAGCCCTGGAAAAATTTGATGCAGCGCTGAAAGCGC
K138A antisense	GCGCTTTCAGCGCTGCATCAAATTTTTCCAGGGCTTCATCATC
K141A sense	GAAAAATTTGATAAAGCGCTGGCAGCGCTGCCGATGCATATTCG
K141A antisense	CGAATATGCATCGGCAGCGCTGCCAGCGCTTTATCAAATTTTTC

[a] restriction sites or mutations are underlined.

Cloning and Overexpression of Equine BLG

The reported monomeric protein *Equus caballus* (horse) BLG variant I was synthesized as an *E. coli* codon-optimized gene and ligated to pMAL-c4x by GenScript using the *Eco*RI and *Hind*III sites to produce vector pMAL-eBLG. Overexpression and purification conditions were identical to those used for MBP-BLG to obtain milligram quantities of MBP-eBLG fusion protein. The purified protein was analyzed for monomeric or dimeric states by tris-glycine nondenaturing discontinuous PAGE (10% gel, 0.15 mA, 55 min). BSA was chosen as an appropriate standard due to similar molecular weight and pI as that of the MBP-eBLG fusion.

MBP-BLG Incubations, Mass Spectrometry, and HPLC Analysis

To assess the catalytic capability of MBP-BLG fusions, a 30 mL reaction of 1 g protein in PBS pH 7.4 (10 mM phosphate, 2.7 mM KCl, 137 mM NaCl) and 3 equivalents of citral (100 μL of 0.5 M ethanol solution) were incubated at 250 rpm and 37°C for 4 days. This reaction composition is the molar equivalent of a 1% (w/v) wild type BLG solution (i.e., the amount of BLG is the same). The reaction was diluted to 100 mL with PBS and stirred with 100 mL ethyl acetate for 30 min to extract organics. The mixture was centrifuged for 30 min at 6000 rpm to settle the denatured protein, and organics were collected, dried over anhydrous magnesium sulfate, and concentrated *in vacuo*. Crude extract was derivatized at room temperature by stirring with dansyl hydrazine (1 mg, Invitrogen, Carlsbad, CA) in 1% HCl in methanol (2 mL) for 30 min. Following dilution to 30 mL, the mixture was extracted with 3-30 mL portions of dichloromethane, and organics were dried over anhydrous magnesium sulfate and concentrated to give a viscous orange oil. Derivatized crude extract was stored at -80°C to prevent degradation of the unreduced imine bond.

Mass spectrometry (MS) was performed at the Protein Chemistry Laboratory within the Biochemistry/Biophysics Department at Texas A&M University. Samples were directly infused into a DecaXP ion trap mass spectrometer (ThermoFisher, Waltham, MA) for positive mode electrospray ionization analysis and tandem MS/MS.

Normal phase HPLC was carried out on a Varian ProStar Liquid Chromatography system with a Luna 5 silica column (10 x 250 mm, 5 μm, 100 Å, Phenomenex) and a mobile phase gradient of 0-5% methanol in dichloromethane over 50 min with a 3 mL/min flow rate. Samples were directly injected and species elution detected by UV absorbance at 320 nm via an on-board photo diode array.

CHAPTER VI

CONCLUSION

Natural products are inextricably linked to both the survival and detriment of humanity. The molecules produced by biosynthetic pathways may serve to protect an organism's interests by warding off predators or competition, while at the same time may promote unintended consequences in other living systems. Adopting compounds manufactured by microorganisms and vegetation is an age-old practice which has inspired a wealth of medicinally-valuable chemicals, and the search for untold therapeutic opportunities will undoubtedly continue far into the foreseeable future. A biochemical understanding of the processes that bring about natural products is one of the most valuable tools we have for control and manipulation of the world surrounding us, and the aim of the studies undertaken here was to contribute to that goal.

If biochemical transformations can truly be compartmentalized and segregated into modular entities, the tinkering opportunities are limitless. Admittedly, conceptually vast options for drug development are currently known to be reined in by many factors, some understood and others not yet clear. Within the biochemical landscape one may prefer to think small instead of big, and thus much of the potential for wide utility is lost in the wonder of delineating just one tiny detail. If studies such as these eventually lead to greater and grander things, so much the better, but we should also take a moment to cull immense satisfaction with the baby steps made toward experiencing nature at its most fundamental, awe-inspiring and inexplicably exquisite.

From a therapeutic engineering standpoint, studies on azinomycin biosynthesis and resistance provide not only a rationalization for complex synthesis to select specific biological targets, but also in a general sense serve the ambitions of *de novo* biosynthetic drug design. As exact enzymatic transformations and timing of azinomycin epoxyvaline, aziridinopyrrollidine, and keto-enol biosynthesis are fleshed out, they will become the essential pieces upon which stability and specificity engineering can proceed. When a complete captured cluster is in hand, efforts to develop a gene therapy approach for *in situ* drug delivery mediated by plasmids and cancer cell-specific promoters is but one idea relying on the outcome of this and other planned studies within our research group.

Genomic sequencing of the azinomycin-producing *Streptomyces* strain has presented additional secondary metabolic pathways to explore, and perhaps the resistance protein and related information characterized herein will prove relevant to elucidation of prospective novel

compound structure and associated biosynthetic studies. The *S. sahachiroi* draft genome sequence represents both a big step forward, in terms of long-awaited genetic information, as well as a hefty sequence finishing task. Significant scaffolding for contig ordering using a close genetic relative has not yet been begun. Because the genome only displays a roughly estimated 60% or less overall organizational homology to completed (single contig) *Streptomyces* strains, scaffolding attempts alone may be insufficient, especially within the chromosome end regions where unconserved nonessential metabolic pathways typically reside. The recently released genome sequences of 13 *Streptomyces* species by the Broad Institute may be useful if a nearer genetic neighbor than *S. coelicolor* or *S. avermitilis* can be identified. We are, of course, unable to directly compete with the resources and manpower available to institutional groups, and our finishing efforts may require significantly more time than that invested elsewhere.

Alongside the *S. sahachiroi* genome, we performed Illumina paired-end sequencing of six additional microbial strains of significant interest to our laboratory. These strains were collected during various field trips within marine environments in Hawaii, Texas, and Florida. Assembly and analysis of the sequence data is also within the future plans of the group, as these strains are suspected to harbor secondary biosynthetic pathways of potentially novel natural products, and likely include orphan pathways like those discovered in the *S. sahachiroi* genome. Genomics is very much an accelerant for biosynthetic pathway discovery, and has the power to unleash a torrential overload of information as science moves ever closer to the $1,000 genome.

The converse of promising biosynthetic natural product leads is the latent damage invoked by formation of detrimental biosynthetic compounds. β-Lactoglobulin-mediated cycloterpenal formation is a potential example of the nefarious behavior natural product biosynthetic enzymes can assume. While not catalyzed by a genetically-encoded human enzyme, the transformation attributed to the milk protein BLG is no less significant to medical conditions associated with lipofuscin formation. Because the mechanism of retinoid-derived fluorophore accumulation is as unclear as the biological role of BLG, they are both at this time well-matched explanations for their observed deposition within macular tissue. Conclusions based on human studies are not available at this time, and thus the link between whey protein consumption, retinoid dimer accumulation, and macular degeneration remain a plausible but untested hypothesis.

The structure of a protein environment suited to α,β-unsaturated aldehyde activation via Schiff base and subsequent facilitation of ring-fused dimer formation is in itself an interesting query independent of the disease implications associated with BLG. Experiments described in

this thesis mark the beginning of a long series of planned investigations to clarify the perceived moonlighting activity of this ubiquitous whey protein, with the aim of shedding light on possible strategies for blocking catalytic activity such that it would prevent disease progression. BLG active site architecture would pave the way for comparative analysis of human proteins that may carry out identical or similar transformations, thus explaining the progression of age-related macular degeneration in the absence of a dietary link. Finally, confirmation of catalytic activity observed by a protein previously only suspected of a carrier function suggests proteins may be more versatile than they are often imagined to be. Under specific conditions, rare activities could be exploited for biosynthesis of previously unattainable targets. The unfortunate requisite is that moonlighting behavior is often an accidental discovery, as in the case of BLG, and is not to be expected from all proteins and all substrates. Even high-throughput screening for moonlighting activity would appear to be a blind undertaking, although *de novo* peptide and protein engineering groups are actively seeking such discoveries as biosynthetic solutions to synthetic problems.

REFERENCES

1. Newman, D.J. Natural products as leads to potential drugs: An old process or the new hope for drug discovery? *J. Med. Chem.* **51**, 2589-2599 (2008).

2. Cragg, G.M., Newman, D.J. & Snader, K.M. Natural products in drug discovery and development. *J. Nat. Prod.* **60**, 52-60 (1997).

3. Newman, D.J., Cragg, G.M. & Snader, K.M. Natural products as sources of new drugs over the period 1981−2002. *J. Nat. Prod.* **66**, 1022-1037 (2003).

4. Wani, M.C., Taylor, H.L., Wall, M.E., Coggon, P. & McPhail, A.T. Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J. Am. Chem. Soc.* **93**, 2325-2327 (1971).

5. Fuchs, D.A. & Johnson, R.K. Cytologic evidence that taxol, an antineoplastic agent from *Taxus brevifolia*, acts as a mitotic spindle poison. *Cancer Treat. Rep.* **62**, 1219-1222 (1978).

6. Levine, D.P. Vancomycin: A history. *Clin. Infect. Dis.* **42**, S5-S12 (2006).

7. Kingston, D.G.I. Taxol, a molecule for all seasons. *Chem. Commun.*, 867-880 (2001).

8. Abad, F., Calbo, F., Zapater, P., Rodríguez-Vilanova, F., García-Pérez, L.-E. *et al.* Comparative pharmacoeconomic study of vancomycin and teicoplanin in intensive care patients. *Int. J. Antimicrob. Agents* **15**, 65-71 (2000).

9. Takahashi, H., Liu, Y.-n. & Liu, H.-w. A two-stage one-pot enzymatic synthesis of TDP-L-mycarose from thymidine and glucose-1-phosphate. *J. Am. Chem. Soc.* **128**, 1432-1433 (2006).

10. Baran, P.S., Maimone, T.J. & Richter, J.M. Total synthesis of marine natural products without using protecting groups. *Nature* **446**, 404-408 (2007).

11. Cooper, E.L. Drug discovery, CAM and natural products. *eCAM* **1**, 215-217 (2004).

12. Hata, T., Koga, F., Sano, Y., Kanamori, K., Matsumae, A. *et al.* Carzinophilin, a new tumor inhibitory substance produced by *Streptomyces*, I. *J. Antibiot., Ser. A* **7**, 107-112 (1954).

13. Nagaoka, K., Matsumoto, M., Oono, J., Yokoi, K., Ishizeki, S. *et al.* Azinomycins A and B, new antitumor antibiotics I. Producing organism, fermentation, isolation, and characterization. *J. Antibiot.* **39**, 1527-1532 (1986).

14. Casely-Hayford, M.A., Pors, K., James, C.H., Patterson, L.H., Hartley, J.A. *et al.* Design and synthesis of a DNA-crosslinking azinomycin analogue. *Org. Biomolec. Chem.* **3**, 3585-3589 (2005).

15. Ishizeki, S., Ohtsuka, M., Kazuhiko, I., Kukita, K.-i., Nagaoka, K. *et al.* Azinomycins A and B, new antitumor antibiotics III. Antitumor activity. *J. Antibiot.* **40**, 60-65 (1987).

16. Coleman, R.S., Perez, R.J., Burk, C.H. & Navarro, A. Studies on the mechanism of action of azinomycin B: Definition of regioselectivity and sequence selectivity of DNA cross-link formation and clarification of the role of the naphthoate. *J. Am. Chem. Soc.* **124**, 13008-13017 (2002).

17. Kelly, G.T., Liu, C., Smith III, R., Coleman, R.S. & Watanabe, C.M.H. Cellular effects induced by the antitumor agent azinomycin B. *Chem. Biol.* **13**, 485-492 (2006).

18. LePla, R.C., Landreau, C.A.S., Shipman, M., Hartley, J.A. & Jones, G.D.D. Azinomycin inspired bisepoxides: influence of linker structure on in vitro cytotoxicity and DNA interstrand cross-linking. *Bioorg. Med. Chem. Lett.* **15**, 2861-2864 (2005).

19. Hashimoto, M., Matsumoto, M., Yamada, K. & Terashima, S. Synthetic studies of carzinophilin. Part 4: Chemical and biological properties of carzinophilin analogues. *Tetrahedron* **59**, 3089-3097 (2003).

20. Hashimoto, M., Matsumoto, M., Yamada, K. & Terashima, S. Synthesis, chemical property, and cytotoxicity of the carzinophilin congeners carrying a 2-(1-acylamino-1-alkoxycarbonyl)methylidene-1-azabicyclo[3.1.0]hexane system. *Tetrahedron Lett.* **35**, 2207-2210 (1994).

21. Armstrong, R.W., Combs, A.P., Tempest, P.A., Brown, S.D. & Keating, T.A. Multiple-component condensation strategies for combinatorial library synthesis. *Acc. Chem. Res.* **29**, 123-131 (1996).

22. Shishido, K., Haruna, S., Iitsuka, H. & Shibuya, M. Synthesis and evaluation of enantiomerically pure azinomycin-lexitropsin hybrid molecules with DNA-cleaving activity. *Heterocycles* **49**, 109-112 (1998).

23. Mao, Y., Varoglu, M. & Sherman, D.H. Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from *Streptomyces lavendulae* NRRL 2564. *Chem. Biol.* **6**, 251-263 (1999).

24. Ogasawara, Y. & Liu, H.-w. Biosynthetic studies of aziridine formation in azicemicins. *J. Am. Chem. Soc.* **131**, 18066-18068 (2009).

25. Zhao, Q., He, Q., Ding, W., Tang, M., Kang, Q. *et al.* Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis. *Chem. Biol.* **15**, 693-705 (2008).

26. Van Lanen, S.G., Oh, T.-j., Liu, W., Wendt-Pienkowski, E. & Shen, B. Characterization of the maduropeptin biosynthetic gene cluster from *Actinomadura madurae* ATCC 39144 supporting a unifying paradigm for enediyne biosynthesis. *J. Am. Chem. Soc.* **129**, 13082-13094 (2007).

27. Basundhara, S., Tae-Jin, O., Rajan, L., Kwangkyoung, L., Hei Chan, L. *et al.* Neocarzinostatin naphthoate synthase: an unique iterative type I PKS from neocarzinostatin producer *Streptomyces carzinostaticus*. *FEBS Lett.* **566**, 201-206 (2004).

28. Corre, C. & Lowden, P.A.S. The first biosynthetic studies of the azinomycins: Acetate incorporation into azinomycin B. *Chem. Commun.*, 990-991 (2004).

29. Corre, C., Landreau, C.A.S., Shipman, M. & Lowden, P.A.S. Biosynthetic studies on the azinomycins: The pathway to the naphthoate fragment. *Chem. Commun.*, 2600-2601 (2004).

30. Liu, C., Kelly, G.T. & Watanabe, C.M.H. *In vitro* biosynthesis of the antitumor agent azinomycin B. *Org. Lett.* **8**, 1065-1068 (2006).

31. Kelly, G.T., Sharma, V. & Watanabe, C.M.H. An improved method for culturing *Streptomyces sahachiroi*: Biosynthetic origin of the enol fragment of azinomycin B. *Bioorg. Chem.* **36**, 4-15 (2008).

32. Sharma, V., Kelly, G.T., Foulke-Abel, J. & Watanabe, C.M.H. Aminoacetone as the penultimate precursor to the antitumor agent azinomycin A. *Org. Lett.* **11**, 4006-4009 (2009).

33. Sharma, V., Kelly, G.T. & Watanabe, C.M.H. Exploration of the molecular origin of the azinomycin epoxide: Timing of the biosynthesis revealed. *Org. Lett.* **10**, 4815-4818 (2008).

34. Ding, W., Deng, W., Tang, M., Zhang, Q., Tang, G. *et al.* Biosynthesis of 3-methoxy-5-methyl naphthoic acid and its incorporation into the antitumor antibiotic azinomycin B. *Mol. Biosyst.* **6**, 1071-1081 (2010).

35. Tokala, R.K., Strap, J.L., Jung, C.M., Crawford, D.L., Salove, M.H. *et al.* Novel plant-microbe rhizosphere interaction involving *Streptomyces lydicus* WYEC108 and the pea plant (*Pisum sativum*). *Appl. Environ. Microbiol.* **68**, 2161-2171 (2002).

36. Mincer, T.J., Jensen, P.R., Kauffman, C.A. & Fenical, W. Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. *Appl. Environ. Microbiol.* **68**, 5005-5011 (2002).

37. Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147 (2002).

38. Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H. *et al.* Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotech* **21**, 526-531 (2003).

39. Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M. *et al.* Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J. Bacteriol.* **190**, 4050-4060 (2008).

40. Hopwood, D.A. The *Streptomyces* genome - be prepared! *Nat. Biotechnol.* **21**, 505-506 (2003).

41. Challis, G.L. & Hopwood, D.A. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14555-14561 (2003).

42. Oliynyk, M., Brown, M.J.B., Cortés, J., Staunton, J. & Leadlay, P.F. A hybrid modular polyketide synthase obtained by domain swapping. *Chem. Biol.* **3**, 833-839 (1996).

43. Stemmer, W.P.C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389-391 (1994).

44. Roessner, C.A. & Scott, A.I. Genetically engineered synthesis of natural products: From alkaloids to corrins. *Annu. Rev. Microbiol.* **50**, 467-490 (1996).

45. Du, L. & Shen, B. Biosynthesis of hybrid peptide-polyketide natural products. *Curr. Opin. Drug Discov. Dev.* **4**, 215 - 228 (2001).

46. Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S.D. *et al.* The neocarzinostatin biosynthetic gene cluster from *Streptomyces carzinostaticus* ATCC 15944 involving two iterative type I polyketide synthases. *Chem. Biol.* **12**, 293-302 (2005).

47. Shao, L., Qu, X.-D., Jia, X.-Y., Zhao, Q.-F., Tian, Z.-H. *et al.* Cloning and characterization of a bacterial iterative type I polyketide synthase gene encoding the 6-methylsalicyclic acid synthase. *Biochem. Biophys. Res. Commun.* **345**, 133-139 (2006).

48. Bedford, D., Jacobsen, J.R., Luo, G., Cane, D.E. & Khosla, C. A functional chimeric modular polyketide synthase generated via domain replacement. *Chem. Biol.* **3**, 827-831 (1996).

49. Pieper, R., Gokhale, R.S., Luo, G., Cane, D.E. & Khosla, C. Purification and characterization of bimodular and trimodular derivatives of the erythromycin polyketide synthase. *Biochemistry* **36**, 1846-1851 (1997).

50. McDaniel, R., Kao, C.M., Hwang, S.J. & Khosla, C. Engineered intermodular and intramodular polyketide synthase fusions. *Chem. Biol.* **4**, 667-674 (1997).

51. Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J. *et al.* Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23**, 1171-1176 (2005).

52. Wright, G.D. The antibiotic resistome: The nexus of chemical and genetic diversity. *Nat Rev Micro* **5**, 175-186 (2007).

53. Abraham, E.P. & Chain, E. An enzyme from bacteria able to destroy penicillin. *Nature* **146**, 837 (1940).

54. Boehr, D.D., Thompson, P.R. & Wright, G.D. Molecular mechanism of aminoglycoside antibiotic kinase APH(3′)-IIIa: Roles of conserved active site residues. *J. Biol. Chem.* **276**, 23929-23936 (2001).

55. Yu, E.W., McDermott, G., Zgurskaya, H.I., Nikaido, H. & Koshland, D.E., Jr. Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science* **300**, 976-980 (2003).

56. Yamaguchi, A., Ono, N., Akasaka, T., Noumi, T. & Sawai, T. Metal-tetracycline/H+ antiporter of *Escherichia coli* encoded by a transposon, Tn10. The role of the conserved dipeptide, Ser65-Asp66, in tetracycline transport. *J. Biol. Chem.* **265**, 15525-15530 (1990).

57. Katz, L., Brown, D., Boris, K. & Tuan, J. Expression of the macrolide-lincosamide-streptogramin-B-resistance methylase gene, *erm*E, from *Streptomyces erythraeus* in *Escherichia coli* results in N6-monomethylation and N6,N6-dimethylation of ribosomal RNA. *Gene* **55**, 319-325 (1987).

58. Dumas, P., Bergdoll, M., Cagnon, C. & Masson, J.-M. Crystal structure and site-directed mutagenesis of a bleomycin resistance protein and their significance for drug sequestering. *EMBO J.* **13**, 2483-2492 (1994).

59. Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J. *et al.* Call of the wild: Antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* **8**, 251-259 (2010).

60. Watve, M., Tickoo, R., Jog, M. & Bhole, B. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **176**, 386-390 (2001).

61. Shen, B. & Hutchinson, C. Enzymatic synthesis of a bacterial polyketide from acetyl and malonyl coenzyme A. *Science* **262**, 1535-1540 (1993).

62. Qian, C., Longkuan, X., Izumikawa, M., Meluzzi, D. & Moore, B.S. Enzymatic total synthesis of enterocin polyketides. *Nat. Chem. Biol.* **3**, 557-558 (2007).

63. Kharel, M.K., Pahari, P., Lian, H. & Rohr, J.r. Enzymatic total synthesis of rabelomycin, an angucycline group antibiotic. *Org. Lett.* **12**, 2814-2817 (2010).

64. Sanger, F., Nicklen, S. & Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467 (1977).

65. Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C. *et al.* Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674-679 (1986).

66. Swerdlow, H. & Gesteland, R. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucl. Acids Res.* **18**, 1415-1419 (1990).

67. Fleischmann, R., Adams, M., White, O., Clayton, R., Kirkness, E. *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512 (1995).

68. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).

69. Blom, N. & Rapacki, K. Database of Genome Sizes. March 5, 2010. (Center for Biological Sequence Analysis, Technical University of Denmark)

70. Mardis, E.R. Next-generation DNA sequencing methods. *Annu. Rev. Genom. Hum. Genet.* **9**, 387-402 (2008).

71. Pop, M. & Salzberg, S.L. Bioinformatics challenges of new sequencing technology. *Trends Genet.* **24**, 142-149 (2008).

72. Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M. & Nyrén, P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**, 84-89 (1996).

73. Kim, S.R., Jang, Y.P., Jockusch, S., Fishkin, N.E., Turro, N.J. *et al.* The all-*trans*-retinal dimer series of lipofuscin pigments in retinal pigment epithelial cells in a recessive Stargardt disease model. *Proc. Natl. Acad. Sci.* **104**, 19273-19278 (2007).

74. Sparrow, J.R., Vollmer-Snarr, H.R., Zhou, J., Jang, Y.P., Jockusch, S. *et al.* A2E-epoxides damage DNA in retinal pigment epithelial cells: Vitamin E and other antioxidants inhibit A2E formation. *J. Biol. Chem.* **278**, 18207-18213 (2003).

75. Eldred, G.E. in *The Retinal Pigment Epithelium: Function and Disease* (eds Marmor, M.F. & Wolfensberger, T.J.) 651-668 (Oxford Univ Press, New York, 1998).

76. Bench, B.J., Liu, C., Evett, C.R. & Watanabe, C.M.H. Proline promoted synthesis of ring-fused homodimers: Self-condensation of alpha,beta-unsaturated aldehydes. *J. Org. Chem.* **71**, 9458-9463 (2006).

77. Stone, E.M. Macular degeneration. *Annu. Rev. Med.* **58**, 477-490 (2007).

78. Crabb, J.W., Miyagi, M., Gu, X., Shadrach, K., West, K.A. *et al.* Drusen proteome analysis: An approach to the etiology of age-related macular degeneration. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14682-14687 (2002).

79. Asato, A.E., Watanabe, C., Li, X.Y. & Liu, R.S.H. The proline and beta-lactoglobulin mediated asymmetric self-condensation of beta-ionylideneacetaldehyde, retinal and related-compounds. *Tetrahedron Lett.* **33**, 3105-3108 (1992).

80. Molday, R., Beharry, S., Ahn, J. & Zhong, M. Binding of *N*-retinylidene-PE to BACA4 and a model for its transport across membranes. in *Retinal Degenerative Diseases*, Vol. 572 (eds. Hollyfield, J.G., Anderson, R.E. & LaVail, M.M.) 465-470 (Springer US, 2006).

81. Allikmets, R. Simple and complex ABCR: Genetic predisposition to retinal disease. *Am. J. Hum. Genet.* **67**, 793-799 (2000).

82. Sawyer, L. & Kontopidis, G. The core lipocalin, bovine β-lactoglobulin. *Biochim. Biophys. Acta* **1482**, 136-148 (2000).

83. Caillard, I. & Tome, D. Modulation of beta-lactoglobulin transport in rabbit ileum. *Am. J. Physiol.* **266**, G1053-G1059 (1994).

84. Lovegrove, J.A., Osman, D.L., Morgan, J.B. & Hampton, S.M. Transfer of cow's milk beta-lactoglobulin to human serum after a milk load: A pilot study. *Gut* **34**, 203-207 (1993).

85. Foulke-Abel, J., Batalov, S., Su, A.I. & Watanabe, C.M.H. *unpublished work*.

86. Hodgkinson, T.J. & Shipman, M. Chemical synthesis and mode of action of the azinomycins. *Tetrahedron* **57**, 4467-4488 (2001).

87. Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C. *et al.* CDD: specific functional annotation with the Conserved Domain Database. *Nucl. Acids Res.* **37**, D205-210 (2009).

88. Umezawa, Y., Yagisawa, M., Sawa, T., Takeuchi, T., Umezawa, H. *et al.* Aminoglycoside 3'-phosphotransferase III, a new phosphotransferase resistance mechanism. *J. Antibiot. (Tokyo)* **28**, 845-853 (1975).

89. Bibb, M.J., White, J., Ward, J.M. & Janssen, G.R. The mRNA for the 23S rRNA methylase encoded by the ermE gene of *Saccharopolyspora erythrea* is translated in the absence of a conventional ribosome-binding site. *Mol. Microbiol.* **14**, 533-545 (1994).

90. Page, N., Kluepfel, D., Shareck, F. & Morosoli, R. Increased xylanase yield in *Streptomyces lividans*: Dependence on number of ribosome-binding sites. *Nat. Biotechnol.* **14**, 756-759 (1996).

91. Hartley, J.L., Temple, G.F. & Brasch, M.A. DNA cloning using *in vitro* site-specific recombination. *Genome Res.* **10**, 1788-1795 (2000).

92. McKay, G.A., Thompson, P.R. & Wright, G.D. Broad spectrum aminoglycoside phosphotransferase type III from *Enterococcus*: overexpression, purification, and substrate specificity. *Biochemistry* **33**, 6936-6944 (1994).

93. Anderson, K.S., Sikorski, J.A. & Johnson, K.A. Evaluation of 5-enolpyruvoylshikimate-3-phosphate synthase substrate and inhibitor binding by stopped-flow and equilibrium fluorescence measurements. *Biochemistry* **27**, 1604-1610 (1988).

94. Wu, L. & Serpersu, E.H. Deciphering interactions of the aminoglycoside phosphotransferase(3')-IIIa with its ligands. *Biopolymers* **91**, 801-809 (2010).

95. Wright, G.D. & Thompson, P.R. Aminoglycoside phosphotransferases: Proteins, structure, and mechanism. *Front. Biosci.* **4**, d9-21 (1999).

96. Davies, J. & O'Connor, S. Enzymatic modification of aminoglycoside antibiotics: 3-*N*-Acetyltransferase with broad specificity that determines resistance to the novel aminoglycoside apramycin. *Antimicrob. Agents Chemother.* **14**, 69-72 (1978).

97. Livermore, D. Beta-lactamases in laboratory and clinical resistance. *Clin. Microbiol. Rev.* **8**, 557-584 (1995).

98. Johnson, D.A., August, P.R., Shackleton, C., Liu, H.-w. & Sherman, D.H. Microbial resistance to mitomycins involves a redox relay mechanism. *J. Am. Chem. Soc.* **119**, 2576-2577 (1997).

99. Walsh, C.T., Fisher, S.L., Park, I.S., Prahalad, M. & Wu, Z. Bacterial resistance to vancomycin: Five genes and one missing hydrogen bond tell the story. *Chem. Biol.* **3**, 21-28 (1996).

100. Gatignol, A., Durand, H. & Tiraby, G. Bleomycin resistance conferred by a drug-binding protein. *FEBS Lett.* **230**, 171-175 (1988).

101. Martin, T.W., Dauter, Z., Devedjiev, Y., Sheffield, P., Jelen, F. *et al.* Molecular basis of mitomycin C resistance in *Streptomyces*: Structure and function of the MRD protein. **10**, 933-942 (2002).

102. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254 (1976).

103. Paget, M.S.B., Chamberlin, L., Atrih, A., Foster, S.J. & Buttner, M.J. Evidence that the extracytoplasmic function sigma factor sigma E Is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **181**, 204-211 (1999).

104. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. & Hopwood, D.A. *Practical Streptomyces Genetics*, 613 (John Innes Foundation, Norwich, UK, 2000).

105. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).

106. Ishikawa, J. & Hotta, K. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. *FEMS Microbiol. Lett.* **174**, 251-253 (1999).

107. Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **36**, D13-21 (2008).

108. Jenkins, A.H., Schyns, G., Potot, S., Sun, G. & Begley, T.P. A new thiamin salvage pathway. *Nat. Chem. Biol.* **3**, 492-497 (2007).

109. DeSantis, G., Liu, J., Clark, D.P., Heine, A., Wilson, I.A. *et al.* Structure-based mutagenesis approaches toward expanding the substrate specificity of 2-deoxyribose-5-phosphate aldolase. *Biorg. Med. Chem.* **11**, 43-52 (2003).

110. Challis, G.L. Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* **154**, 1555-1569 (2008).

111. Gross, H. Strategies to unravel the function of orphan biosynthesis pathways: Recent examples and future prospects. *Appl. Microbiol. Biotechnol.* **75**, 267-277 (2007).

112. DiGuistini, S., Liao, N.Y., Platt, D., Robertson, G., Seidel, M. *et al. De novo* genome sequence assembly of a filamentous fungus using Sanger, 454 and Illumina sequence data. *Genome Biol.* **10**, R94.91-R94.12 (2009).

113. Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T. *et al.* The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 975 (2008).

114. Summers, R.G., Wendt-Pienkowski, E., Motamedi, H. & Hutchinson, C.R. The tcmVI region of the tetracenomycin C biosynthetic gene cluster of *Streptomyces glaucescens* encodes the tetracenomycin F1 monooxygenase, tetracenomycin F2 cyclase, and, most likely, a second cyclase. *J. Bacteriol.* **175**, 7571-7580 (1993).

115. Tang, Y., Tsai, S.-C. & Khosla, C. Polyketide chain length control by chain length factor. *J. Am. Chem. Soc.* **125**, 12708-12709 (2003).

116. Machida, K., Arisawa, A., Takeda, S., Tsuchida, T., Aritoku, Y. *et al.* Organization of the biosynthetic gene cluster for the polyketide antitumor macrolide, pladienolide, in *Streptomyces platensis* Mer-11107. *Biosci. Biotechnol. Biochem.* **72**, 2946-2952 (2008).

117. Sakai, T., Sameshima, T., Matsufuji, M., Kawamura, N., Dobashi, K. *et al.* Pladienolides, new substances from culture of *Streptomyces platensis* Mer-11107 I. Taxonomy, fermentation, isolation and screening. *J. Antibiot.* **57**, 173-179 (2004).

118. Mizui, Y., Sakai, T., Iwata, M., Uenaka, T., Okamoto, K. *et al.* Pladienolides, new substances from culture of *Streptomyces platensis* Mer-11107 III. *In vitro* and *in vivo* antitumor activities. *J. Antibiot.* **57**, 188-196 (2004).

119. Hernandez, D., François, P., Farinelli, L., Østerås, M. & Schrenzel, J. *De novo* bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Res.* **18**, 802-809 (2008).

120. Zerbino, D.R. & Birney, E. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821-829 (2008).

121. Ansari, M.Z., Yadav, G., Gokhale, R.S. & Mohanty, D. NRPS-PKS: A knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* **32**, W405-413 (2004).

122. Bachmann, B.O. & Ravel, J. Chapter 8 Methods for *in silico* prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. in *Methods Enzymol.*, Vol. 458 (ed. Hopwood, D.A.) 181-217 (Academic Press, San Diego, 2009).

123. Li, X.-y., Asato, A.E. & Liu, R.S.H. β-Lactoglobulin directed photoisomerization of retinal and related compounds. *Tetrahedron Lett.* **31**, 4841-4844 (1990).

124. Bench, B.J., Tichy, S.E., Perez, L.M., Benson, J. & Watanabe, C.M.H. Synthesis and cellular effects of cycloterpenals: Cyclohexadienal-based activators of neurite outgrowth. *Bioorg. Med. Chem.* **16**, 7573-7581 (2008).

125. Peters, L., Konig, G.M., Wright, A.D., Pukall, R., Stackebrandt, E. *et al.* Secondary metabolites of *Flustra foliacea* and their influence on bacteria. *Appl. Environ. Microbiol.* **69**, 3469-3475 (2003).

126. Peters, L., Wright, A.D., Kehraus, S., Gundisch, D., Tilotta, M.C. *et al.* Prenylated indole alkaloids from *Flustra foliacea* with subtype specific binding on NAChRs. *Planta Med.* **70**, 883-886 (2004).

127. Peters, L., Wright, A.D., Krick, A. & Konig, G.M. Variation of brominated indoles and terpenoids within single and different colonies of the marine bryozoan *Flustra foliacea*. *J. Chem. Ecol.* **30**, 1165-1181 (2004).

128. Fishkin, E.N., Sparrow, J. R., Allikmets, RA., Nakanishi, K. Isolation and characterization of a retinal pigment epithelial cell fluorophore: An all-*trans*-retinal dimer conjugate. *Proc. Natl. Acad. Sci.* **102**, 7091-7096 (2005).

129. Fishkin, N., Pescitelli, G., Sparrow, J.R., Nakanishi, K. & Berova, N. Absolute configurational determination of an all-*trans*-retinal dimer isolated from photoreceptor outer segments. *Chirality* **16**, 637-641 (2004).

130. Evans, J.R. Risk factors for age-related macular degeneration. *Prog. Retin. Eye Res.* **20**, 227-253 (2001).

131. Wang, L., Clark, M.E., Crossman, D.K., Kojima, K., Messinger, J.D. *et al.* Abundant lipid and protein components of drusen. *PLoS One* **5**, e10329 (2010).

132. Sakmar, T.P., Franke, R.R. & Khorana, H.G. The role of the retinylidene Schiff base counterion in rhodopsin in determining wavelength absorbance and Schiff base pKa. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 3079-3083 (1991).

133. Flower, D.R. Beyond the superfamily: The lipocalin receptors. *Biochim. Biophys. Acta* **1482**, 327-336 (2000).

134. Marcon-Genty, D., Tome, D., Kheroua, O., Dumontier, A.M., Heyman, M. *et al.* Transport of beta-lactoglobulin across rabbit ileum *in vitro*. *Amer J Physiol* **256**, G943-948 (1989).

135. Asato, A.E., Peng, A., Hossain, M. Z., Mirzadegan, T. and Bertram, J. S. Azulenic retinoids: Novel nonbenzenoid aromatic retinoids with anticancer activity *J. Med. Chem.* **36**, 3137-3147 (1993).

136. Batt, C.A., Rabson, L.D., Wong, D.W.S. & Kinsella, J.E. Expression of recombinant bovine β-lactoglobulin in *Escherichia coli*. *Agric. Biol. Chem.* **54**, 949-955 (1990).

137. Ariyaratne, K.A.N.S., Brown, R., Dasgupta, A., de Jonge, J., Jameson, G.B. *et al.* Expression of bovine β-lactoglobulin as a fusion protein in *Escherichia coli*: A tool for investigating how structure affects function. *Int. Dairy J.* **12**, 311-318 (2002).

138. Chatel, J.-M., Adel-Patient, K., Créminon, C. & Wal, J.-M. Expression of a lipocalin in prokaryote and eukaryote cells: Quantification and structural characterization of recombinant bovine β-lactoglobulin. *Protein Expression Purif.* **16**, 70-75 (1999).

139. Cho, Y., Batt, C.A. & Sawyer, L. Probing the retinol-binding site of bovine beta-lactoglobulin. *J. Biol. Chem.* **269**, 11102-11107 (1994).

140. Invernizzi, G., Ragona, L., Brocca, S., Pedrazzoli, E., Molinari, H. *et al.* Heterologous expression of bovine and porcine beta-lactoglobulins in *Pichia pastoris*: Towards a comparative functional characterisation. *J. Biotechnol.* **109**, 169-178 (2004).

141. Ikeguchi, M., Kato, S.-i., Shimizu, A. & Sugai, S. Molten globule state of equine β-lactoglobulin. *Protein Struct Funct Genet* **27**, 567-575 (1997).

142. Sakurai, K. & Goto, Y. Manipulating monomer-dimer equilibrium of bovine β-Lactoglobulin by amino acid substitution. *J. Biol. Chem.* **277**, 25735-25740 (2002).

143. Hutchison, C.A., Phillips, S., Edgell, M.H., Gillam, S., Jahnke, P. *et al.* Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551-6560 (1978).

144. Yang, M.-C., Guan, H.-H., Liu, M.-Y., Lin, Y.-H., Yang, J.-M. *et al.* Crystal structure of a secondary vitamin D_3 binding site of milk beta-lactoglobulin. *Proteins: Struct. Funct. Bioinform.* **71**, 1197-1210 (2008).

145. Kontopidis, G., Holt, C. & Sawyer, L. The ligand-binding site of bovine β-Lactoglobulin: Evidence for a function? *J. Mol. Biol.* **318**, 1043-1055 (2002).

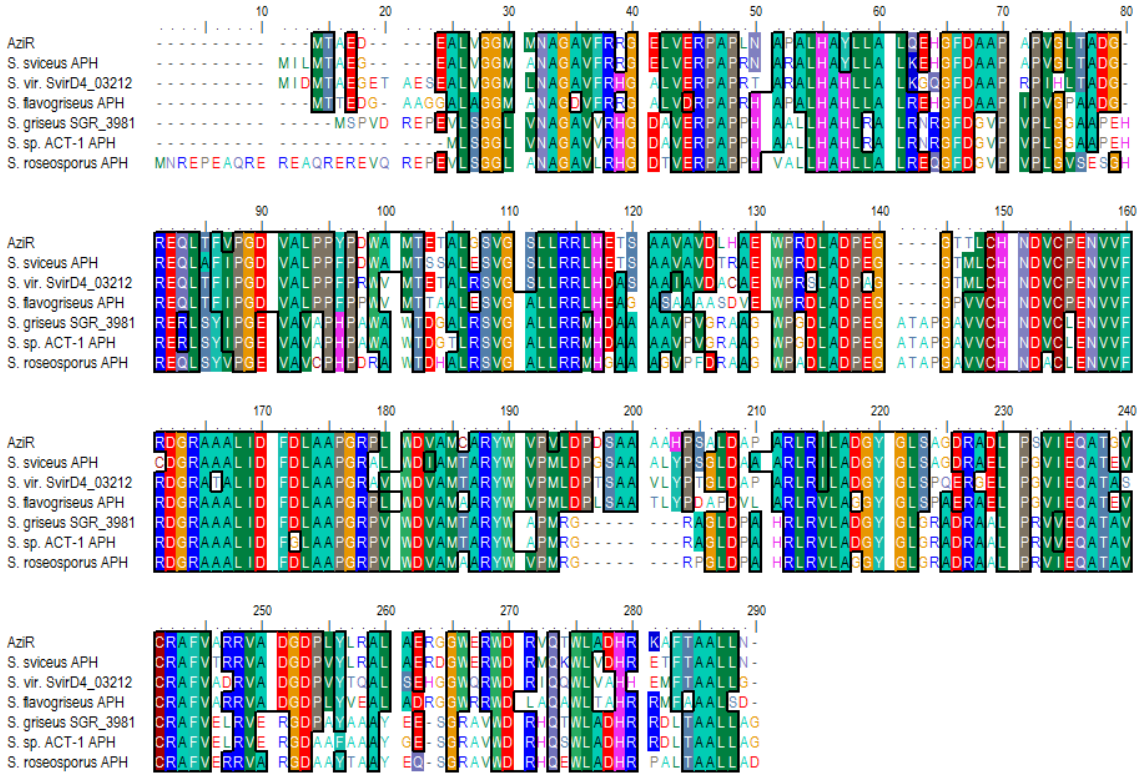
146. Qin, B.Y., Creamer, L.K., Baker, E.N. & Jameson, G.B. 12-Bromododecanoic acid binds inside the calyx of bovine β-lactoglobulin. *FEBS Lett.* **438**, 272-278 (1998).

147. Wu, S.-Y., Pérez, M.D., Puyol, P. & Sawyer, L. β-Lactoglobulin binds palmitate within its central cavity. *J. Biol. Chem.* **274**, 170-174 (1999).

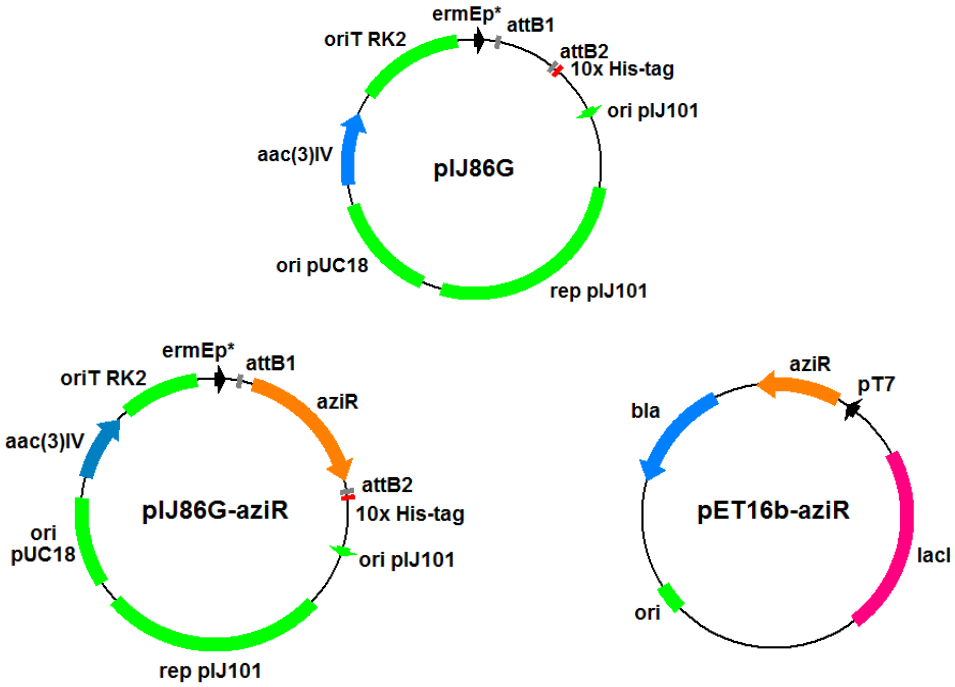
APPENDIX

ADDITIONAL FIGURES

SUPPLEMENTAL FIGURES FOR CHAPTER II



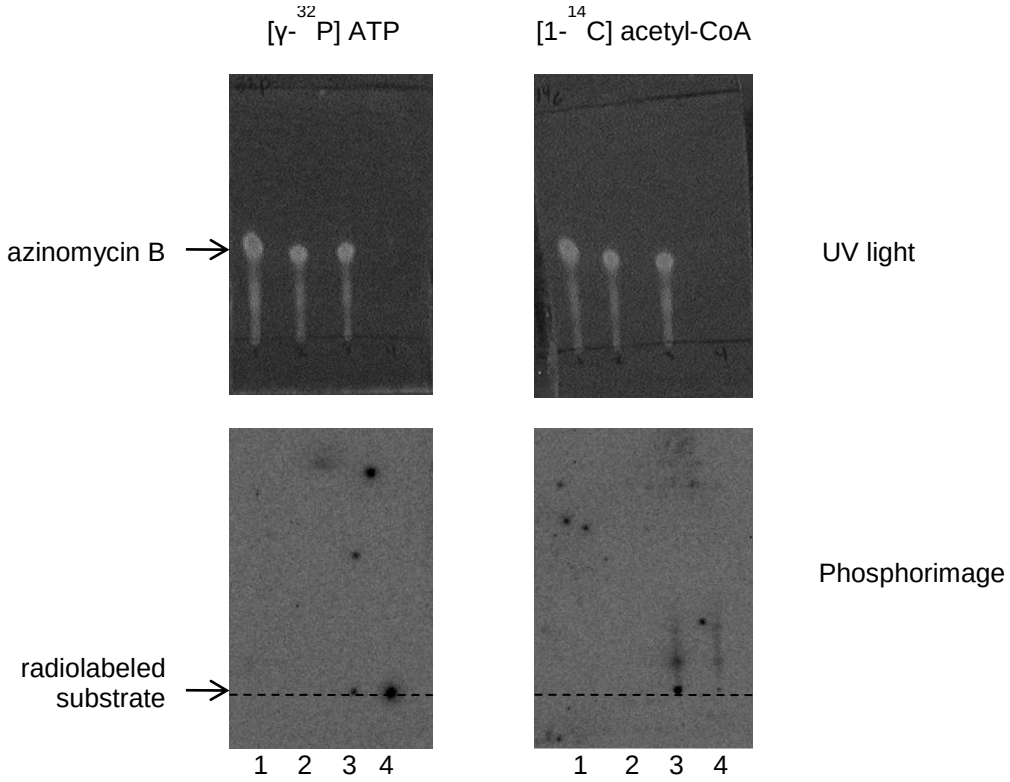
Appendix Figure 57 Multiple alignment of AziR and *Streptomyces* sp. homologs.



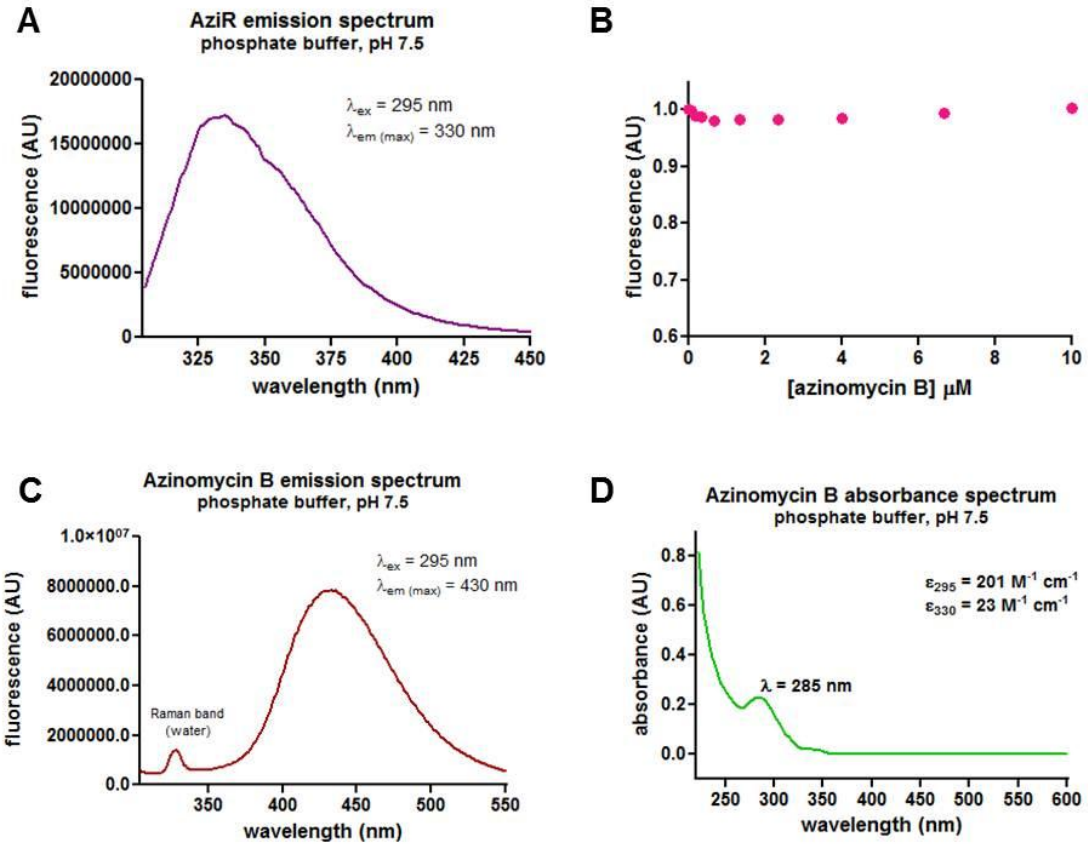
Appendix Figure 58 Maps of plasmids constructed in the course of AziR studies.

>aziR (*S. sahachiroi*), codon-optimized, 801 bp
ATGACCGCGGAAGATGAAGCGCTGGTGGGCGGCATGATGAACGCGGGCGCGGTGTTTCGTCGTGGCGAACTGGTGGAAC
GTCCGGCGCCGCTGAACGCGCCGGCGCTGCATGCGTATCTGCTGGCGCTGCAGGAACATGGCTTTGATGCGGCGCCGGC
GCCGGTGGGCCTGACCGCGGATGGCCGTGAACAGCTGACCTTTGTGCCGGGCGATGTGGCGCTGCCGCCGTATCCGGAT
TGGGCGATGACCGAAACCGCGCTGGGCAGCGTGGGCAGCCTGCTGCGTCGTCTGCATGAAACCAGCGCGGCGGTGGCGG
TGGATCTGCATGCGGAATGGCCGCGTGATCTGGCCGATCCGGAAGGCGGCACCACCCTGTGCCATAACGATGTGTGCCC
GGAAAACGTGGTGTTTCGTGATGGCCGTGCGGCGGCGCTGATTGATTTTGATCTGGCGGCGCCGGGCCGTCCGCTGTGG
GATGTGGCGATGTGCGCGCGTTATTGGGTGCCGGTGCTGGACCCGGATAGCGCGGCGGCGGCGCATCCGAGCGCGCTGG
ATGCGCCGGCGCGTCTGCGTATTCTGGCGGATGGCTATGGCCTGAGCGCGGGCGATCGTGCGGATCTGCCGAGCGTGAT
TGAACAGGCGACCGGCGTGTGCCGTGCGTTTGTGGCGCGTCGTGTGGCGGATGGCGATCCGCTGTATCTGCGTGCGCTG
GCGGAACGTGGCGGCTGGGAACGTTGGGATCGTGTGCAGACCTGGCTGGCGGATCATCGTAAAGCGTTTACCGCGGCGC
TGCTGAACTAA

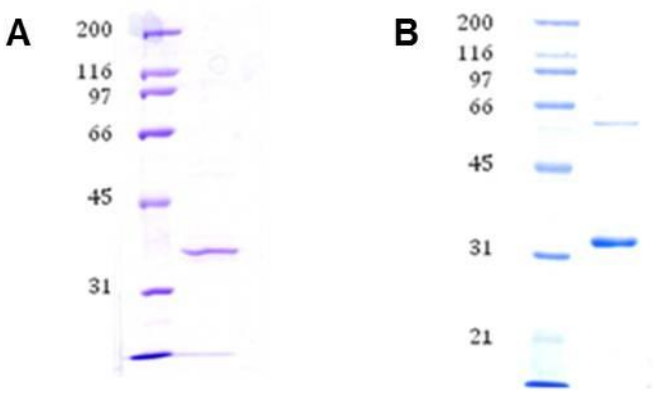
Appendix Figure 59 Gene sequence of *E. coli* codon-optimized AziR.



Appendix Figure 60 Assessing AziR modification of azinomycin B using radiolabeled substrates. TLC separation and UV or phosphorimage detection of azinomycin B incubated with AziR and [γ-^{32}P] ATP or [1-^{14}C] acetyl-CoA. Lane 1, azinomycin B; lane 2, azinomycin B + AziR; lane 3, azinomycin B + AziR + radiolabeled substrate; lane 4, radiolabeled substrate. Reactions consisted of 500 μM azinomycin B and 500 μM radiolabeled substrate, to which 100 μM AziR in 50 mM Tris pH 7.5 and 10 mM NaCl was added. After 1 h incubation at room temperature, reactions were extracted with dichloromethane and analyzed by TLC with 5% methanol in dichloromethane as mobile phase.

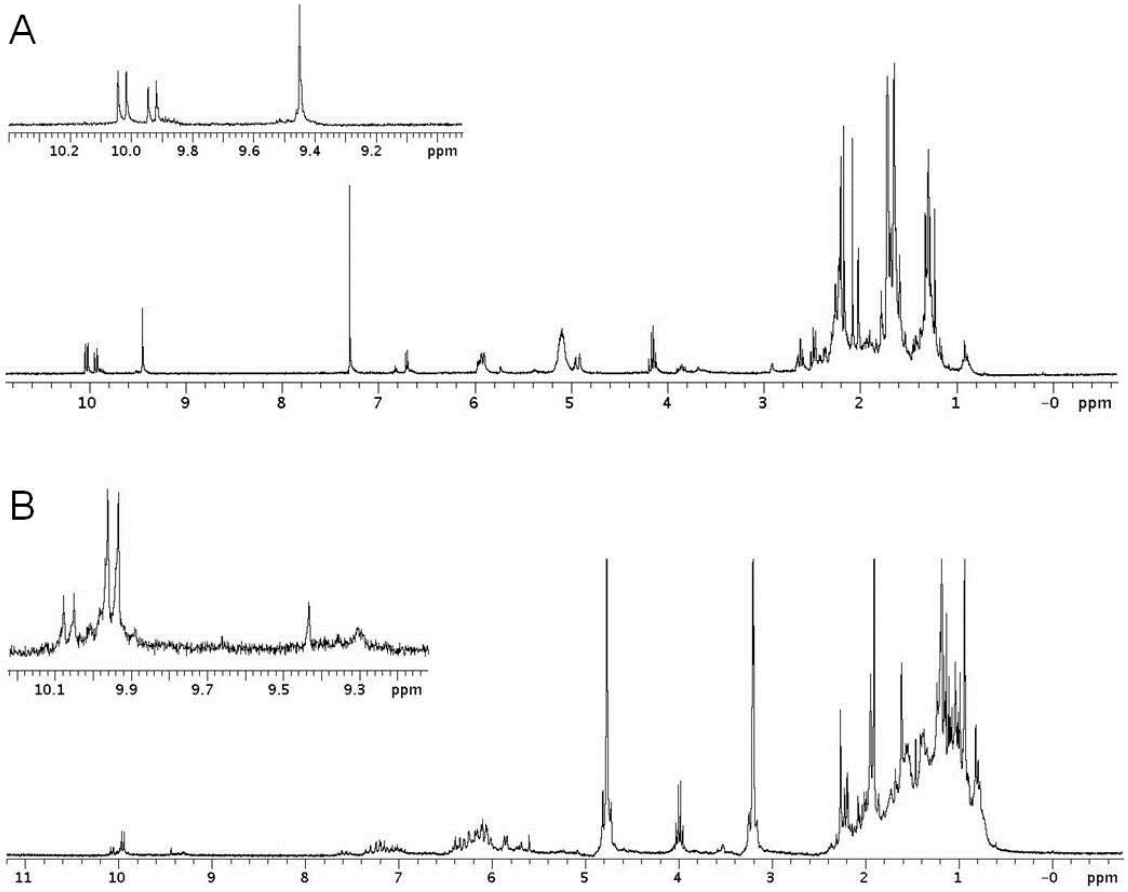


Appendix Figure 61 AziR control fluorescence titration curves and emission spectra.
(A) Fluorescence emission spectrum of AziR, λ_{ex} = 295 nm. (B) DERA titrated with azinomycin B. (C) Azinomycin B (1 mM) fluorescence emission spectrum, λ_{ex} = 295 nm, included to show that ligand does not emit within a range that may interfere with the protein signal. (D) Azinomycin B (1 mM) absorbance spectrum with a local maximum at 285 nm. The ligand does not have a significant absorbance at the excitation wavelength at the concentrations used for titration to necessitate correction for an inner filter effect; corrected data is virtually equivalent to uncorrected data.



Appendix Figure 62 SDS-PAGE of purified AziR from heterologous hosts.
(A) *S. lividans* TK24 and (B) *E. coli* BL21(DE3). Ladder molecular weights are in kDa.

SUPPLEMENTAL FIGURES FOR CHAPTER IV



Appendix Figure 63 ^{1}H NMR analysis of BLG-promoted cycloterpenal biosynthesis.
(A) Extract from 1% BLG incubated with 3 equivalents of citral. (B) Extract from 1% BLG incubated with 3 equivalents of retinal.

A

Day	Yield (%)
1	12.3
2	22.4
3	36.2
4	58
5	58

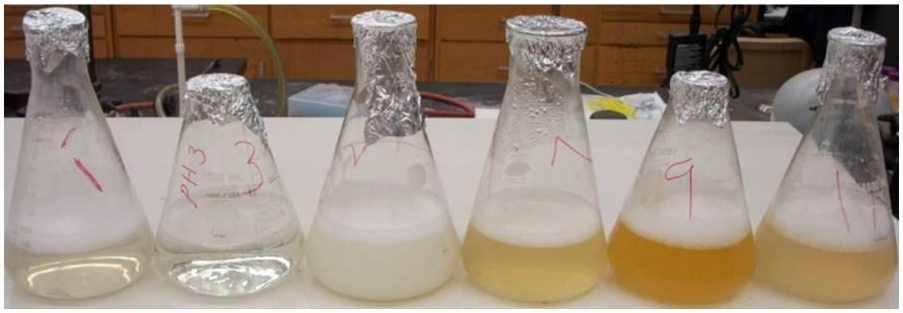
B

pH	Yield (%)
1.0	0
3.0	9.8
5.0	13.6
7.0	50.7
9.0	27.9
11.0	0

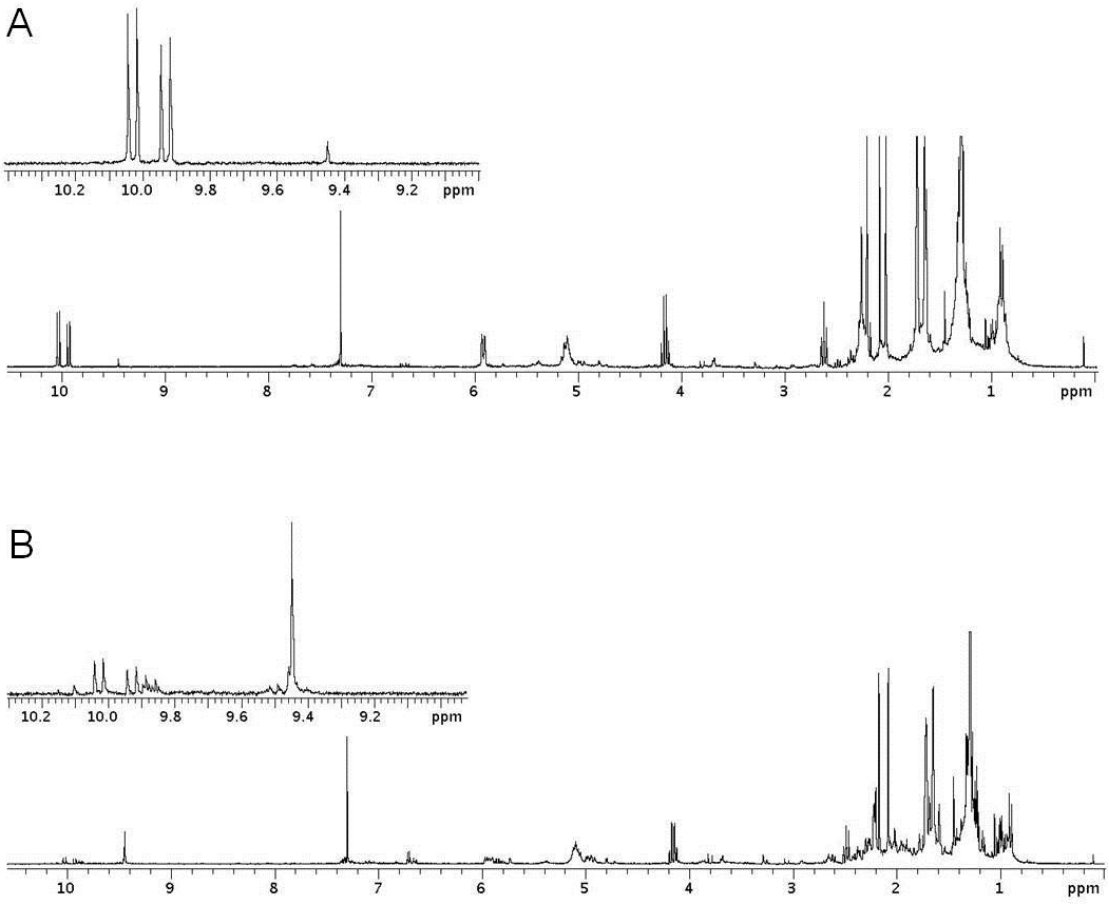
C

BLG:Substrate Equivalents	Volume 0.5 M citral (mL)	Yield (%)
1:1	0.54	10
1:2	1.09	31
1:3	1.63	58
1:5	2.72	58
1:10	5.43	58

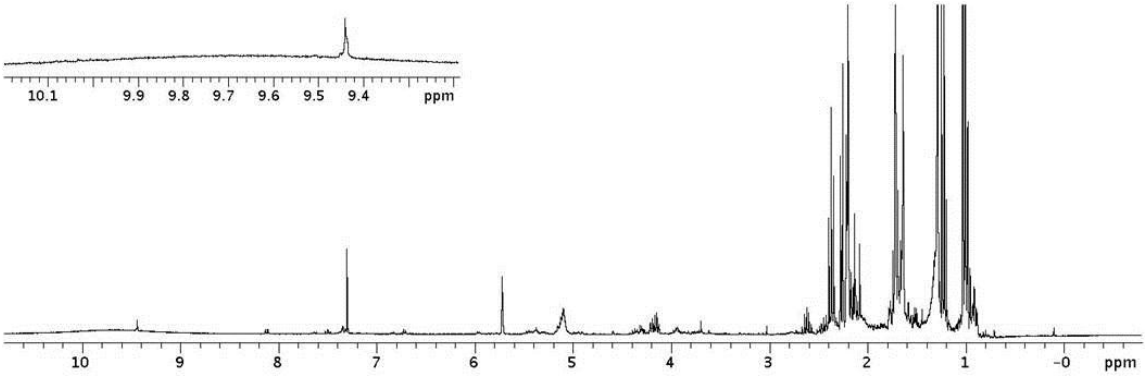
D



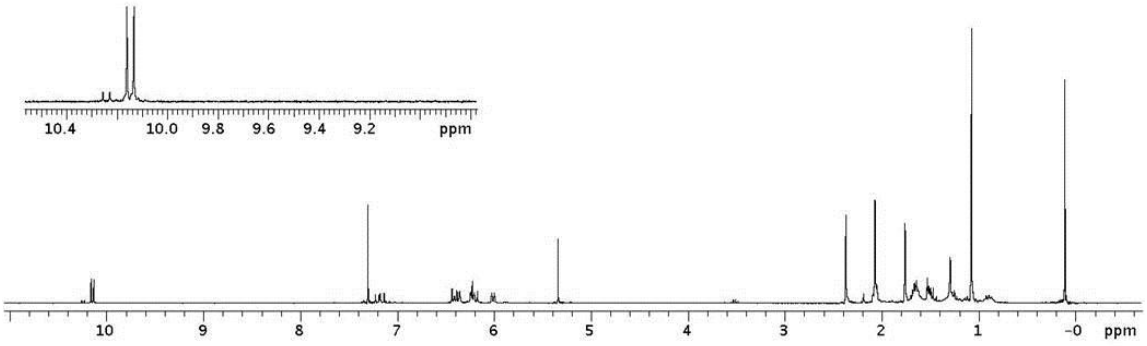
Appendix Figure 64 Raw data for BLG-promoted reaction optimization.
Percent yield for ranges of (A) protein to substrate ratio, (B) pH, and (C) reaction duration. (D) Color changes along with pH. From left to right: pH 1.0, 3.0, 5.0, 7.0, 9.0, 11.0.



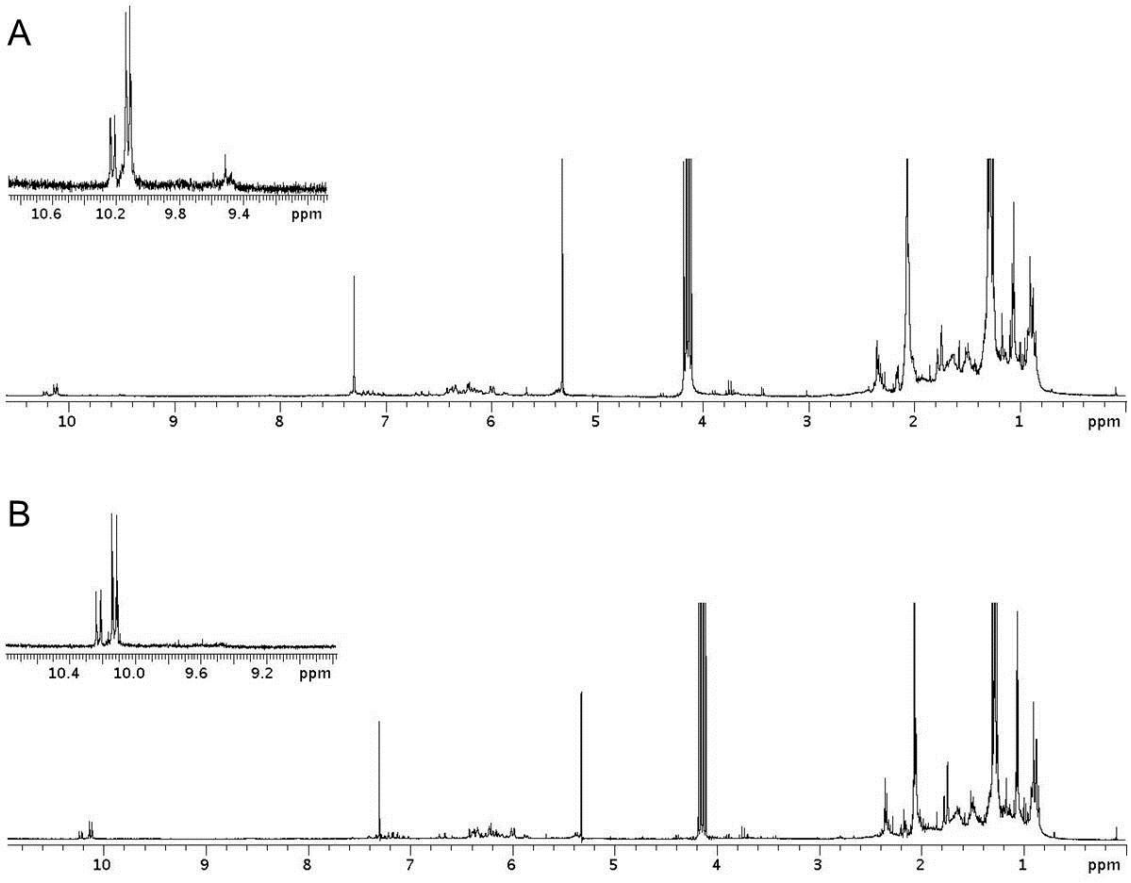
Appendix Figure 65 ^{1}H NMR of reaction extract from BLG denaturation attempts using urea or heat. (A) BLG stirred with 7 M urea for 2 h prior to substrate addition and standard incubation. (B) BLG heated at 100°C for 2 h with stirring prior to substrate addition and standard incubation. Despite significant aggregation and precipitation at high temperatures, the protein retained catalytic activity (cyclocitral aldehyde signal at δ 9.45 ppm) as shown in the NMR spectra insets.



Appendix Figure 66 ^{1}H NMR spectrum of extract from Sephadex G-25 filtered skim milk incubated with citral.

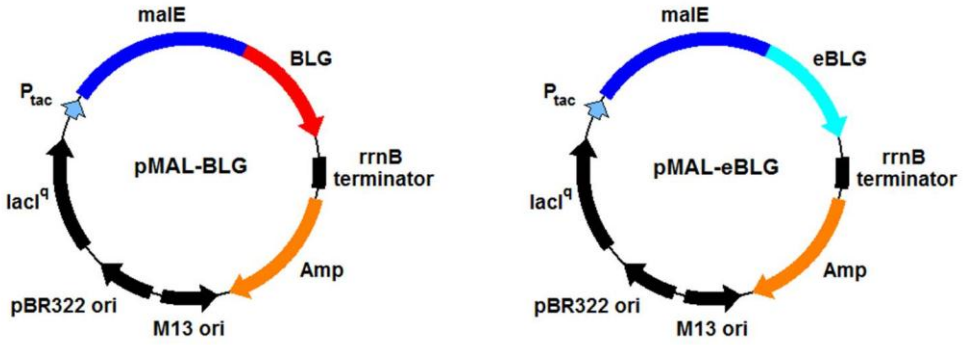


Appendix Figure 67 ^{1}H NMR spectrum of extract from L-proline in PBS pH 7.4 incubated with retinal.



Appendix Figure 68 ^{1}H NMR spectra of rabbit blood incubated with retinal *in vitro*.
(A) Rabbit fed BLG-supplemented diet. (B) Rabbit fed control (basic) diet.

SUPPLEMENTAL FIGURES FOR CHAPTER V



Appendix Figure 69 Plasmid maps of pMAL-BLG and pMAL-eBLG for expression in *E. coli*.

>BLG variant B (*Bos taurus*), codon-optimized, 489 bp
CTGATTGTGACCCAGACCATGAAAGGCCTGGATATTCAGAAAGTGGCGGGTACCTGGTATAGCCTGGCCATGGCGGCGA
GCGATATTAGCCTGCTGGATGCGCAGAGCGCGCCGCTGCGTGTGTATGTTGAAGAACTGAAACCGACCCCGGAAGGCGA
TCTGGAAATTCTGCTGCAGAAATGGGAAAACGGCGAATGCGCGCAGAAAAAAATTATCGCGGAAAAAACCAAAATTCCG
GCGGTGTTTAAAATTGATGCGCTGAATGAAAACAAAGTGCTGGTGCTGGATACCGATTATAAAAAATATCTGCTGTTTT
GCATGGAAAATAGCGCGGAACCGGAACAGAGCCTGGCGTGCCAGTGTCTGGTGCGCACCCCGGAAGTGGATGATGAAGC
CCTGGAAAAATTTGATAAAGCGCTGAAAGCGCTGCCGATGCATATTCGCCTGAGCTTTAATCCGACCCAGCTGGAAGAA
CAGTGTCACATTTAA

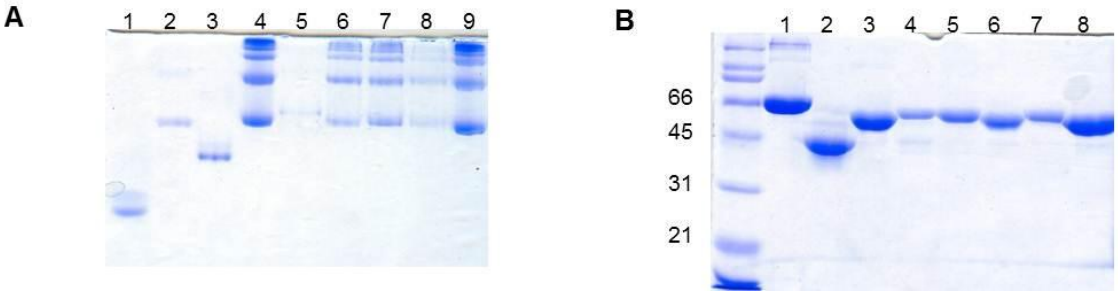
>BLG I (*Equus caballus*), codon-optimized, 489 bp
ACGAACATCCCGCAAACGATGCAAGACCTGGACCTGCAAGAAGTGGCTGGCAAATGGCACAGCGTGGCAATGGCGGCGA
GCGATATTAGCCTGCTGGATAGCGAATCTGCACCGCTGCGTGTCTATATTGAAAAACTGCGCCCGACCCCGGAAGACAA
CCTGGAAATTATCCTGCGTGAAGGCGAAAACAAAGGTTGCGCAGAAAAGAAAATTTTCGCTGAAAAAACGGAAAGTCCG
GCGGAATTTAAAATCAACTACCTGGATGAAGACACCGTTTTTGCCCTGGATACGGACTATAAAAACTACCTGTTCCTGT
GCATGAAAAATGCAGCAACCCCGGGCCAGTCCCTGGTGTGTCAATACCTGGCACGCACGCAGATGGTTGATGAAGAAAT
CATGGAAAAATTCCGTCGCGCTCTGCAACCGCTGCCGGGTCGTGTCCAAATCGTGCCGGATCTGACGCGCATGGCCGAA
CGCTGTCGTATCTAA

Appendix Figure 70 Gene sequence of *E. coli* codon-optimized bovine BLG and equine BLG.

Appendix Table 14 Primers for monomeric bovine BLG site-directed mutagenesis.

Primer	**Sequences** (5'-3')[a]
R40D sense	GCAGAGCGCGCCGCTGGATGTGTATGTTGAAGAAC
R40D antisense	GTTCTTCAACATACACATCCAGCGGCGCGCTCTGC
H146P sense	CTGAAAGCGCTGCCGATGCCGATTCGCCTGAGCTTTAATC
H146P antisense	GATTAAAGCTCAGGCGAATCGGCATCGGCAGCGCTTTCAG

[a] mutations are underlined.



Appendix Figure 71 PAGE analysis of proposed monomeric BLGs expressed as MBP fusions. (A) Native discontinuous PAGE. Lane 1, BLG; lane 2, BSA; lane 3, ovalbumin; lane 4, MBP-BLG; lane 5, MBP-eBLG; lane 6, MBP-BLG R40D; lane 7, MBP-BLG H146P; lane 8, MBP-BLG R40D/H146P; lane 9, MBP-BLG K77A. (B) SDS-PAGE to demonstrate purity of samples used in native PAGE. Lane 1, BSA; lane 2, ovalbumin; lane 3, MBP-BLG; lane 4, MBP-eBLG; lane 5, MBP-BLG R40D; lane 6, MBP-BLG H146P; lane 7, MBP-BLG R40D/H146P; lane 8, MBP-BLG K77A. Ladder molecular weights are in kDa.

VITA

Name: Jennifer Dianne Foulke-Abel

Current Address: The Johns Hopkins University
Department of Chemistry
3400 N. Charles St.
Baltimore, MD 21218

Email Address: jfoulkeabel@gmail.com

Education: B.A. Chemistry, Texas A&M University, 2004
Ph.D. Chemistry, Texas A&M University, 2010

Publications:

Sharma, V., Kelly, G.T., Foulke-Abel, J. & Watanabe, C.M.H. Aminoacetone as the penultimate precursor to the anti-tumor agent azinomycin A. *Org. Lett.* **11**, 4006-4009 (2009).

Bench, B.J., Foulke-Abel, J. & Watanabe, C.M.H. Milk, revealed "silent" chemistry: A new mode of cycloretinal synthesis. *Mol. BioSyst.* (2010). In press.

Foulke-Abel, J., Agbo, H., Zhang, H., Mori, S. & Watanabe, C.M.H. Biosynthesis and mode of action of the azabicycle containing natural products azinomycin and ficellomycin. *Nat. Prod. Rep.* (2010). Manuscript submitted.

Foulke-Abel, J., Kelly, G.T., Zhang, H. & Watanabe, C.M.H. Characterization of AziR, a resistance protein of the DNA cross-linking agent azinomycin B. (2010). Manuscript in preparation.