UC Berkeley
UC Berkeley Previously Published Works

Title

PKS-NRPS Enzymology and Structural Biology: Considerations in Protein Production.

Permalink

https://escholarship.org/uc/item/0273m6bv

Authors

Skiba, Meredith

Maloney, Finn

Dan, Qingyun

et al.

Publication Date

2018

DOI

10.1016/bs.mie.2018.01.035

Peer reviewed

PKS-NRPS Enzymology and Structural Biology: Considerations in Protein Production

Meredith A. Skiba[*,†,‡], **Finn P. Maloney**[*,†,§], **Qingyun Dan**[*,†,‡], **Amy E. Fraley**[*,†,¶], **Courtney C. Aldrich**[£], **Janet L. Smith**[*,†,‡,1], and **W. Clay Brown**[*,†,1]

[*]University of Michigan, Ann Arbor, MI, USA

[†]Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA

[‡]Department of Biological Chemistry, University of Michigan, Ann Arbor, MI, USA

[§]Program in Chemical Biology, University of Michigan, Ann Arbor, MI, USA

[¶]Department of Medicinal Chemistry, University of Michigan, Ann Arbor, MI, USA

[£]Department of Medicinal Chemistry, University of Minnesota, Minneapolis, MN, USA

Abstract

The structural diversity and complexity of marine natural products have made them a rich and productive source of new bioactive molecules for drug development. The identification of these new compounds has led to extensive study of the protein constituents of the biosynthetic pathways from the producing microbes. Essential processes in the dissection of biosynthesis have been the elucidation of catalytic functions and the determination of 3D structures for enzymes of the polyketide synthases and nonribosomal peptide synthetases that carry out individual reactions. The size and complexity of these proteins present numerous difficulties in the process of going from gene to structure. Here, we review the problems that may be encountered at the various steps of this process and discuss some of the solutions devised in our and other labs for the cloning, production, purification and structure solution of complex proteins using *E. coli* as a heterologous host.

Keywords

1. Introduction

Natural product megasynthases, including polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS), encompass a wealth of fascinating enzymes that catalyze a vast array of chemical transformations and thus are prime candidates for development as synthetic tools (Akey, Gehret, Khare & Smith, 2012; Walsh, 2016; Winn, Fyans, Zhuo &

[1]Corresponding author JanetSmith@umich.edu, wclayb@umich.edu.

Micklefield, 2016; Yuzawa, Keasling & Katz, 2016; Keatinge-Clay, 2017; Payne, Millar, Jackson & Ochs, 2017). Purified, recombinant proteins are key reagents in the discovery of the catalytic activity, chemical mechanism, substrate scope and 3D structure of PKS and NRPS enzymes and in the manipulation of their activity. Recombinant production of individual enzymes in a heterologous host presents several challenges, including coding sequences adapted for proteins of low natural abundance, uncertain boundaries within multi-domain polypeptides, unusual G+C content, and poor solubility when excised from the full-length natural protein. Multi-domain proteins present additional challenges for folding to a native state in a heterologous host cell.

Fortunately an abundance of annotated and unannotated sequences for PKS and NRPS gene clusters provides a rich database of target proteins and, in combination with 3D structures from the growing number of homologs in the structure database (Keatinge-Clay, 2012; Miller & Gulick, 2016), can inform experiment design. Many reagents and tools are available to aid in production of active recombinant proteins, however each protein behaves in a unique manner when produced in a heterologous host cell. Thus, the success or failure of an expression and purification method for a particular protein can be poorly predictive for the applicability of that strategy to a homologous protein. A certain amount of trial-and-error is generally needed. Any one of several problems can be vexing, leading us to adapt the Anna Karenina Principle: "All happy recombinant proteins are alike, each unhappy recombinant protein is unhappy in its own way." The challenge is to identify a range of experimental possibilities by applying all knowledge about a target protein and to investigate these possibilities rapidly and economically.

Here we describe the approaches developed in our laboratory for gene expression in an *Escherichia coli* host, for purification of soluble recombinant PKS and NRPS proteins, and for addressing the most common challenges of these target proteins, with examples from our and others research. We have applied these methods widely to proteins of marine cyanobacterial origin, but they are equally applicable to many other systems. Although the host expression system is bacterial, we have used it for production of eukaryotic proteins. While smaller quantities, reduced purity and soluble aggregates are generally well tolerated in enzyme assays, our approach is directed to producing milligram quantities of highly pure, monodisperse protein because solving crystal structures is a major activity in our lab. Our method to produce recombinant proteins is based on using all available sequence and 3D-structure information to design an experiment, and then applying the fundamentals of *E. coli* physiology to survey several conditions in parallel on a small scale.

2. cDNA Selection

2.1 Amino Acid Sequence Selection

The design of a construct encoding a protein of interest generally starts with a multiple sequence alignment of the target and several related proteins, which can be found with tools such as BLAST (Johnson, Zaretskaya, Raytselis, Merezhuk, McGinnis & Madden, 2008). The sequence database is deep for PKS and NRPS enzymes, and generally a subset of hits must be selected as a workable number (10-20) for alignment. Selection criteria may favor sequences of proteins that are characterized biochemically or are from the biosynthetic

pathways of characterized natural products. Sequence identities with the target protein sequence are an important consideration; those in a range of 20% to 80% should yield an alignment with enough information to identify approximate domain boundaries, conserved hotspots, insertion/deletion sites, and intrinsically disordered regions. A search of the structure database (PDB, http://www.rcsb.org/pdb/), (Berman, Westbrook, Feng, Gilliland, Bhat, Weissig et al., 2000) with the sequence of the target protein can aid in understanding the multiple sequence alignment and inform decisions about domain boundaries. The structures of any related proteins (> 20% identity) can be powerful tools to generate hypotheses about function. We generally search the PDB with the BLAST server, using "Protein Data Bank proteins (pdb)" as the search set database.

The rapidly decreasing cost of synthetic DNA makes the entire sequence database accessible for protein production and has the added advantage of codon optimization for gene expression in a heterologous host. This can be extremely helpful when a target protein proves to be recalcitrant to purification in a stable form, or when other complications arise.

2.2 Translation Start Sites

The amino acid sequences for most PKS and NRPS proteins are based on annotation of genome or gene-cluster sequences, thus they should be regarded as working hypotheses for the translation start and stop sites. Incorrect annotation of start sites can impede biochemical and structural studies of recombinant proteins. For example, the SpnF Diels-Alderase was initially characterized in a recombinant protein that could be produced in soluble form only with chaperones (Kim, Ruszczycky, Choi, Liu & Liu, 2011). Researchers later identified a start codon upstream and in-frame with the annotated *spnF*, resulting in stable protein amenable to crystallization (Fage, Isiorho, Liu, Wagner, Liu & Keatinge-Clay, 2015). Start-codon identification can be challenging for polycistronic, overlapping genes, which are common in prokaryotes (Kozak, 1983; Osterman, Evfratov, Sergiev & Dontsova, 2013). This can be assisted by locating the Shine-Delgarno (SD) ribosome binding site; SD sites 4-12 nucleotides upstream of the start codon have optimal function. For example, a polycistronic gene encodes the CurD hydroxymethylglutaryl-synthase (HMGS) of PKS β-branching in the gene cluster for curacin A, and the *curD* start codon was originally annotated at a site 59 nucleotides upstream of the stop codon for the preceding gene *curC*. The correct start codon, four nucleotides downstream of the *curC* stop codon, was recognized by finding a consensus SD sequence immediately upstream of the *curC* stop (Maloney, Gerwick, Gerwick, Sherman & Smith, 2016). A multiple sequence alignment of homologs can also aid in detection of incorrectly annotated translation start site, as for the marine haloalkane dehalogenase DmmA (Gehret, Gu, Geders, Brown, Gerwick, Gerwick et al., 2012).

It is also important to consider the start codon identity of the producing organism. For example, a GTG codon is frequently used for initiation in G+C-rich *Streptomyces* species (Villegas & Kropinski, 2008). Consideration of several potential GTG start codons for the LipPKS1 loading module from the lipomycin biosynthetic pathway led to the identification of alternative start sites, some of which were not predicted by gene-finding tools (Yuzawa, Bailey, Fujii, Jocic, Barajas, Benites et al., 2017). In some cases, use of a different start site altered the production level and catalytic activity of recombinant protein from *E. coli* and a

heterologous *Streptomyces* host. Alternative start codons have also been identified in cyanobacterial species. (Sazuka & Ohara, 1996). Additionally, identification of translation start sites can be complicated in the case of "leaderless" genes, which exist in many bacterial genomes (Nakagawa, Niimura, Miura & Gojobori, 2010). Leaderless genes lack the SD sequence, but may encode a "TA-like signal" ~12 nucleotides upstream of the translation start codon (Zheng, Hu, She & Zhu, 2011). Cyanobacterial genomes also include a substantial number of "atypical" leaderless genes that lack both an SD sequence and a TA-like signal, although some of these may have a dipyrimidine immediately upstream of the translation start site (Nakagawa et al., 2010; Zheng et al., 2011).

2.3 Domain Boundaries

If the goal is to produce the target protein in its full natural length, then the critical decision is a straightforward choice of biological source. However, if the target protein consists of one or more domains from a megasynthase, then excision sites must be chosen within inter-domain linker peptides, typically of length 5-40 amino acids. The 3D structures of homologs can be extremely helpful in this task, as can the established boundaries of adjacent domains. The limits of the well conserved region in a multiple sequence alignment generally define the folded core of a protein domain, but this may not include all regions essential to function or stability. It is often necessary to investigate multiple constructs to identify the most stable and/or active fragment, bearing in mind that "linker" regions longer than ~100 amino acids may be folded domains, and if so, should be included or excluded in their entirety.

Domain boundaries within many PKS megasynthases were determined originally through proteolysis, and later rationalized when crystal structures revealed important conserved details that could be applied to other pathways. For example, the sequence for the first ketoreductase (KR) crystal structure from the erythromycin PKS (Figure 1a) (Keatinge-Clay & Stroud, 2006) was based on fragments from limited proteolysis (Aparicio, Caffrey, Marsden, Staunton & Leadlay, 1994). The structure revealed that the KR consists of tandem structural (KR_S) and catalytic (KR_C) domains and that an enoylreductase (ER) domain, when present, is inserted between KR_S and KR_C (Figure 1b and c). A more subtle feature was a conserved β-strand that precedes KR_S in the amino acid sequence, but serves as a critical element of the NADPH-binding core of KR_C.

Much later, this structural detail was critical to understanding how *C*-methyltransferases (CMT) fit into PKS modules (Skiba, Sikkema, Fiers, Gerwick, Sherman, Aldrich et al., 2016). In absence of a 3D structure for any close relative, it was impossible to ascertain the CMT domain boundaries beyond the easily recognized methyltransferase core motifs (Kozbial & Mushegian, 2005), so we determined the boundaries experimentally by constructing twelve plasmids for the CurJ CMT with N-termini spanning 30 residues from the end of the preceding dehydratase (DH) domain and three different C-termini. Nine constructs yielded soluble protein, and one yielded a crystal structure, which defined the true inter-domain linkers (Skiba et al., 2016). The structure led to the unexpected finding that the KR N-terminal β-strand precedes the *C*MT (Figure 1d), and the conclusion that this critical element is essential to producing soluble KR domains from homologous PKS modules.

A similar approach was used to identify soluble, active MT domains in "GNAT" loading modules, which contain a GCN5-related *N*-acetyltransferase (GNAT) domain (Skiba, Sikkema, Moss, Tran, Sturgis, Gerwick et al., 2017). In this context, the MT is located between an "adaptor region" (AR) (Gu, Geders, Wang, Gerwick, Hakansson, Smith et al., 2007) and a GNAT domain. We screened constructs encoding five different N- and C-termini for the production of a soluble MT domain from two biosynthetic pathways (Grindberg, Ishoey, Brinza, Esquenazi, Coates, Liu et al., 2011; Young, Stevens, Carmichael, Tan, Rachid, Boddy et al., 2013). Only those constructs encoding an AR-MT-GNAT tri-domain yielded stable, active protein. The crystal structure of the AprA AR-MT-GNAT tri-domain from apratoxin A revealed that the AR is a large MT lid and the GNAT is closely associated with the MT core, explaining the insolubility of shorter constructs (Skiba et al., 2017).

Crystallization behavior is not predictable and can be highly sensitive to the termini of excised domains. Thus, several termini within the putative inter-domain linker regions should be tested for crystallization behavior. Domain termini were screened (three N-termini and two C-termini) to optimize DHs from the curacin A pathway, resulting in crystal structures of all four DHs (Akey, Razelun, Tehranisa, Sherman, Gerwick & Smith, 2010). Crystal quality was improved by inclusion of five C-terminal amino acids that are not highly conserved but form a short 3_{10} helix.

3. Selection of Expression Plasmid and Bacterial Strain

3.1 Cloning Method

The production of recombinant protein generally begins with the cloning of the gene encoding the protein(s) of interest. The first choice in a cloning strategy is the method of cloning itself. Cloning was originally done using restriction digestion to generate complimentary ends on the insert and vector. A large number of commercially available vectors have multiple-cloning sequences (MCS) to accommodate many combinations of restriction sites to facilitate gene cloning for expression. The probability of a conflict arising from sequences in the gene matching restriction sites in the MCS increases with larger inserts. Over the last twenty-five years many ligation-independent cloning (LIC) systems have been developed to address this and other disadvantages of restriction cloning. LIC systems may be used universally; they function independently of gene sequence.

The Gateway system (Invitrogen) relies on enzymatic recombination to create the final expression clone. The large number of Gateway-adapted vectors makes the system flexible, and the high efficiency makes it easy for first-time users. However, the requirement for proprietary enzymes may be cost prohibitive when making more than a few clones. The Gateway system also adds up to 8 amino acids to the target protein through translation of the recombination sites that are added to the gene fragment for cloning. The In-Fusion system (Clontech) relies on the use of complimentary sequences in the vector and insert, and employs a proprietary enzyme to render these sequences single-stranded and then join them. Advantages include no addition of amino acids and applicability in any vector (Berrow, Alderton, Sainsbury, Nettleship, Assenberg, Rahman et al., 2007). Our primary cloning method relies on vectors containing the LIC region developed at the Midwest Center for

Structural Genomics (MCSG) (Stols, Gu, Dieckman, Raffen, Collart & Donnelly, 2002). Complimentary sequences in the vector and insert are rendered single-stranded through the exonuclease activity of T4 DNA polymerase, a less expensive method than the In-Fusion or Gateway reagents. The MCSG-LIC and In-Fusion systems have 15-nucleotide single-stranded tails and thus comparable cloning efficiencies. However, the MCSG method adds an alanine codon at the beginning of the gene sequence to provide a guanine required for stopping the T4 exonuclease activity.

Numerous variations of LIC integrate a PCR step into the process, including polymerase incomplete primer extension (PIPE) (Klock, Koesema, Knuth & Lesley, 2008) and overlap extension cloning (OEC) (Unger, Jacobovitch, Dantes, Bernheim & Peleg, 2010). The popular Gibson assembly system (Gibson, Young, Chuang, Venter, Hutchison & Smith, 2009) uses the exonuclease step to generate single-stranded overlaps in combination with PCR to amplify the gene and vector fragments. The reaction mix is transformed directly into bacteria where positive clones are identified by the power of antibiotic selection. Commercially available enzyme kits can be used at lower volumes than suggested to reduce the cost. We use Gibson assembly with the MCSG vectors to increase cloning efficiency while keeping costs low. Although the Gibson method was initially developed for the construction of very large pathway assemblies, it has the disadvantage of increased mutation rates with larger fragments due to the reliance on a second round of PCR for native templates. Use of a synthetic template reduces this problem if the cloning-complimentary sequences are synthesized into the fragment, which can then go into the Gibson reaction directly without prior PCR amplification. Gibson assembly is the method of choice for generating chimeric proteins and multi-enzyme polypeptides (Hagen, Poust, Rond, Fortman, Katz, Petzold et al., 2016).

3.2 Base Content

Actinobacteria, such as the pikromyin producer *Streptomyces venezuelae* (He, Sundararajan, Devitt, Schilkey, Ramaraj & Melancon, 2016), are rich sources of PKS and NRPS gene clusters, but have upwards of 70% G+C content and present challenges for cloning of large multi-domain constructs especially from cosmid or genomic DNA. The use of hot-start high-fidelity *Thermococcus kodakaraensis* (KOD) DNA polymerases, such as KOD Xtreme (Novagen), can greatly improve PCR efficiency and product purity, expediting the cloning process. In contrast, cyanobacteria, such as the marine *Moorea* species with 40-50% G+C content (Leao, Castelao, Korobeynikov, Monroe, Podell, Glukhov et al., 2017), often contain long runs of A or T nucleotides, which are challenging for primer design due to the likelihood of false priming or primer dimerization. In such cases, the introduction of silent mutations into primers can assist in successful amplification.

3.3 Promoter Choice

The choice of promoter to drive transcription is a major consideration in vector selection. The bacteriophage T7 RNA polymerase-promoter system, the most widely used in bacterial production, is one of the strongest available for *E. coli* due to its polymerase binding and transcription initiation and is an ideal first-choice system. T7 polymerase travels over DNA at eightfold the rate of *E. coli* RNA polymerase (Iost, Guillerez & Dreyfus, 1992), thus

uncoupling transcription and translation. This uncoupling is not detrimental for soluble production of many proteins, but for others may lead to insolubility due to misfolding and aggregation. Slowing the rate of production can increase the solubility of these proteins. This is commonly done by lowering the growth temperature to slow the rates of transcription and translation (section 6.2). However, if the protein of interest is primarily insoluble at lower temperature, a change to a weaker promoter may be needed. A number of commercially available vectors with native *E. coli* promoters can be used for tight coupling of transcription and translation. Coupling allows for the coordinated sensitivity of RNA polymerase and ribosomes to naturally programed pausing, which may be required for some difficult folds. Improved folding often increases solubility. The *trp-lac* fusion promoter *trc* (Amann, Brosius & Ptashne, 1983) is widely used as a native-promoter alternative to the T7 system.

The choice of promoter influences the choice of host strain. If a T7 promoter plasmid is chosen, the host strain must contain the gene for T7 RNA polymerase unless it is delivered exogenously. The most common host strain is BL21 (DE3) (Studier, Rosenberg, Dunn & Dubendorff, 1990), a lambda lysogen of *E. coli* strain BL21 carrying a copy of the gene for T7 RNA polymerase under the control of the inducible *lac*UV5 promoter. Several features have been incorporated to reduce the problems caused by leakage transcription from the *lac*UV5 promoter, for example, insertion of lac repressor sites downstream of the T7 promoter in the expression vector to block progression of T7 RNA polymerase prior to induction. When used in conjunction with a (DE3) host, the use of IPTG or lactose as an inducer both stimulates transcription of the T7 RNA polymerase gene and relieves transcriptional repression of the target gene. Another level of repression can be achieved with a secondary plasmid (pLysS or E) (Studier et al., 1990) carrying a constitutively expressed copy of the T7 gene for lysozyme, which binds T7 RNA polymerase and inhibits transcription. The lysozyme can assist in cell lysis prior to protein purification, but may also cause unexpected cell lysis. The presence of pLysS also limits the ability to employ other secondary plasmids to deliver accessory functions to enhance soluble protein production.

An alternative host strain for T7 vectors is BL21 AI (Invitrogen), which carries the gene for T7 RNA polymerase under the control of the promoter *araBAD*, derived from the arabinose metabolic pathway (Guzman, Belin, Carson & Beckwith, 1995). The *araBAD* promoter is more tightly controlled than *lac*UV5 with a significantly lower level of leakage transcription but still capable of high-level expression from a T7 target plasmid when induced by the addition of arabinose (Chao, Chiang & Hung, 2002). Since the target vectors rely on the *lac* repressor to reduce leakage from the T7 promoter, the AI host requires arabinose in addition to IPTG or lactose for full induction. We have found the BL21 AI strain to be preferable to pLys-containing strains when working with genes encoding proteins toxic to *E. coli*.

When shifting from a T7 promoter plasmid to a weaker promoter such as *trc* it is prudent to change the host strain as well. If a T7-compatible strain is used with a lactose-inducible promoter plasmid, the T7 RNA polymerase will be induced in competition with the target gene. The T7 RNA polymerase message is stable and well translated to a stable protein that may accumulate to fairly high levels. It is best to remove this as a competitor, especially if the target protein is not made well and/or folds with difficulty. When using *trc* promoter plasmids we have used BL21 as the production host. Since this is the parent strain of the

(DE3) and AI variants used with the T7 system, expression results with these strains may be directly compared. Other hosts commonly used with native-promoter containing plasmids include DH5α and JM109.

4. Affinity Handles and Fusion Partners

The most widely used affinity handles are some variation of the poly-histidine tag (His-tag), usually differing by the number and placement of histidine residues. The advantages of His-tags are the relatively small size of the tag and the low cost of metal-affinity chromatography. All the vectors we use incorporate a His_6-tag so they may be used in parallel with standardized conditions for expression and purification. Small-scale expression trials including an immobilized-metal-affinity chromatography (IMAC) step conveniently determine the target solubility and ease of purification. Although most vectors have an N-terminal His-tag, this placement is not always compatible with the encoded protein. In some cases, a C-terminal tag may avoid tag interference with enzymatic activities or protein-protein interactions. For protein targets prone to truncation due to proteolysis or premature termination of translation, a C-terminal tag may allow for purification of full-length protein away from truncated fragments. A series of MCSG vectors is available for LIC construction with C-terminal His-tags (Eschenfeldt, Maltseva, Stols, Donnelly, Gu, Nocek et al., 2010).

One of the most common issues in heterologous protein production in *E. coli* is insolubility of the product. The His-tag is an effective affinity handle, but it does not generally function as a solubility fusion partner. At the cloning step, insolubility problems can be addressed through the use of solubility fusion partners that are naturally soluble and produced to high level. Numerous proteins have been developed as fusion partners with varying levels of success (Costa, Almeida, Castro & Domingues, 2014). It is unclear what properties of a protein improve the solubility of a passenger protein although high charge and an acidic pI appear to correlate with effectiveness (Fox, Routzahn, Bucher & Waugh, 2003; Su, Zou, Feng, Zhou & Cao, 2007). For example, the commonly used maltose-binding protein (MBP) is a stable, acidic protein that also has tight specific binding to amylose, making it an effective affinity handle. When combined with a His-tag, MBP allows for multi-step purification to a high level of purity (Pryor & Leiting, 1997; Donnelly, Zhou, Millard, Clancy, Stols, Eschenfeldt et al., 2006).

Problems associated with the use of MBP as a solubility partner arise from its ability to solubilize protein aggregates and from difficulties in purifying the passenger protein away from MBP after the fusion is cleaved. These problems led us to develop an additional fusion partner to complement MBP, the Mocr fusion (DelProposto, Majmudar, Smith & Brown, 2009). Mocr is a highly charged, highly acidic (pI=3.8) protein one-third the molecular weight of MBP. This compactness makes Mocr less likely to interfere with passenger protein activity while they are still fused. We and others have shown that the Mocr fusion has effective solubilizing activity (Buchholz, Geders, Bartley, Reynolds, Smith & Sherman, 2009; Gehret, Gu, Gerwick, Wipf, Sherman & Smith, 2011; Banda, Tiwari, Darici & Tse-Dinh, 2016; Key, Dydio, Clark & Hartwig, 2016), and have rarely observed difficulty in removing Mocr from the passenger protein.

To maximize the chance of producing soluble protein, we routinely test His-tag, His-MBP and His-Mocr fusions in parallel. All N-terminal tags encoded in the MCSG vectors in our lab are cleavable with tobacco etch virus (TEV) protease (Raran-Kurussi, Cherry, Zhang & Waugh, 2017), which has high specificity and only rarely detected cleavage at other sites in the protein. TEV protease leaves three amino acids at the N-terminus (Ser-Asn-Ala), which have no impact on activity or crystallization properties of most targets. Cleavage of the fusion is not always necessary, as it may not interfere with biochemical activity, particularly the smaller Mocr-tag and His-tag. For crystallography it is usually desirable to remove the solubility fusions (MBP and Mocr), but the His-tag alone often does not interfere with crystallization (Akey et al., 2010; Maloney et al., 2016; Skiba et al., 2017).

Many other fusion proteins are available with varying levels of effectiveness or utility. Although glutathione-*S*-transferase (GST) is viewed as a solubility partner (Smith & Johnson, 1988; Waugh, 2005), we have found it to be ineffective presumably because it forms homodimers. If the passenger protein is prone to form oligomers or aggregate, then the GST dimer may nucleate extensive aggregation. We have found effective solubilizing activity in the SUMO fusion partner, which is available in LIC compatible vectors (Weeks, Drinker & Loll, 2007). More importantly, SUMO-specific protease cleaves the link between the SUMO C-terminus and the target protein N-terminus to yield a target protein with its native N-terminus (Malakhov, Mattern, Malakhova, Drinker, Weeks & Butt, 2004). The SUMO fusion is an excellent option for proteins requiring a solubilizing fusion partner and a native N-terminus for biochemical activity or partner interactions (Whicher, Smaga, Hansen, Brown, Gerwick, Sherman et al., 2013).

5. Helper Plasmids

Production of protein is an integral function of the cell. It is a very complex series of closely coordinated processes requiring assorted cellular machinery and pools of resources. Occasionally the normal limitations of these resources are reached when using a cell to produce recombinant protein, requiring the provision of additional resources within the cell.

5.1 Codon Usage

Discussions of codon bias during heterologous expression usually revolve around the differences in codon usage between the host species, often *E. coli*, and the species of the gene being used for expression. For eukaryotic genes there is a clear codon bias in *E. coli* that can reduce both the quantity and quality of the protein product. When such a bias exists, it is widely accepted that supplementing the cells with additional copies of the genes for the tRNAs of low-use codons can have a significant impact on heterologous protein yield. This supplementation approach can also aid in expressing genes of bacterial origin. The tables of codon usage for a given species are constructed from averages across the genome, but this presentation obscures the heterogeneity in codon usage among genes within a genome (Sharp & Li, 1986). Although the patterns of usage in related organisms are similar, differences exist and are generally correlated with the evolutionary distance of the organisms (Sharp, Cowe, Higgins, Shields, Wolfe & Wright, 1988). When expressing genes from bacteria not closely related to *E. coli*, low-use codons can be a complication. Additionally,

highly expressed genes tend to have high-use codons whereas genes expressed at low levels display little or no bias for codon usage; they have a mix of high- and low-use codons (Sharp et al., 1988). The gene clusters for many interesting natural products are unlikely to be highly expressed in the native organism, and are likely to contain low-use codons in greater number. When working in *E. coli* with a gene for a low-expressing protein from a distantly related bacterium, supplementing the cell with additional copies of the genes for the tRNAs of low-use codons can have a significant impact on production of the heterologous protein. This is accomplished by the introduction of a plasmid carrying anywhere from two to twelve tRNA genes for low-use codons.

Codon usage issues may also be addressed with synthetic DNA. The cost of DNA synthesis has dropped dramatically over the last few years, making synthetic gene fragments with sequences optimized for common heterologous hosts an attractive option. If fragments are made synthetically using predominantly high-use codons, then a production strain carrying a tRNA supplementation plasmid may not be needed. Fragments longer than several thousand nucleotides may be cost prohibitive at present, but the economics are changing rapidly. Constructs from synthetic fragments with optimized sequences are becoming the norm for ever larger recombinant proteins.

5.2 Chaperone Assistance

For many years it was thought that folding was a purely thermodynamic process and newly synthesized proteins could fold to their native state without assistance. As the functions of heat-shock proteins and chaperones were elucidated, it became clear that not all proteins spontaneously fold to their native state, but require chaperone assistance. Chaperones mediate folding by binding unstable conformers of client proteins, followed by a series of binding and release steps associated with the hydrolysis of ATP. Possible outcomes include proper folding, assembly of homo- and hetero-oligomers, trafficking to and from sub-cellular compartments, or proteolytic degradation (Hartl, 1996).

This understanding of chaperone function led researchers studying heterologous protein production to postulate that insufficient chaperone function in the cell can lead to target protein misfolding and aggregation. As with low-use tRNAs, supplementation could solve this problem. Two major sets of chaperones play different but complimentary roles in protein folding in *E. coli*. The DnaK-DnaJ-GrpE set binds and stabilizes nascent polypeptides very early in translation, often when still bound to the ribosome. The GroEL-GroES set acts at a later stage when a protein has been released from the ribosome with extensive secondary structure but has not reached its final tertiary structure (Hartl, 1996). A series of plasmids was constructed carrying various combinations of the genes for these chaperone sets under several different promoters (Nishihara, Kanemori, Kitagawa, Yanagi & Yura, 1998). In small-scale co-expression tests, problematic recombinant proteins can be tested with the various combinations of chaperones to find a combination that maximizes production of soluble protein. Takara has made a subset of these plasmids commercially available.

Chaperone assistance was essential to our recent structural characterization of the PKS HMGS enzyme (Maloney et al., 2016). Several HMGS proteins were incorrigibly insoluble when produced in *E. coli*, even with use of culture additives, fusion partners, and native

promoters. Co-expression of *curD*, the curacin A HMGS gene, with the Takara plasmids encoding either the GroEL/ES (Takara pGro7) or DnaK/DnaJ (Takara pKJE7) chaperones (Nishihara et al., 1998) resulted in production of soluble protein. Furthermore, this strategy could be applied to solubilize CurD homologs from several other biosynthetic pathways. Thus, when chaperone coexpression is successful for stabilizing a protein of interest, one may consider applying the strategy to homologs from other biosynthetic pathways during structural studies. Similarly, co-expression of *malA* and a Takara plasmid encoding GroEL/ GroES was essential to production of the flavin-dependent halogenase from the malbrancheamide pathway (Fraley, Garcia-Borràs, Tripathi, Khare, Mercado-Marin, Tran et al., 2017).

5.3 Simultaneous tRNA Augmentation and Chaperone Assistance

Most of the plasmids encoding low-use-codon tRNAs have a pACYC backbone and rely on chloramphenicol selection (Rosetta from Novagen, Codon Plus from Agilent). The Takara chaperone sets are also cloned in pACYC backbones and some of these also rely on chloramphenicol selection. Plasmids with a common origin of replication are thought to be incompatible (Novick, 1987), such that a second plasmid introduced into a cell causes instability of a resident plasmid. This incompatibility may not be as stringent as once thought; two plasmids from the same compatibility group may be retained over short periods through antibiotic selection (Velappan, Sblattero, Chasteen, Pavlik & Bradbury, 2007). For cell lines maintained for longer than several overnight growths, however, it is advisable to use plasmids from different compatibility groups.

Over the years we encountered several protein production projects having problems with both codon usage and proper folding, thus we created a supplemental tRNA plasmid (pRARE2-CDF) (Lopanik, Shields, Buchholz, Rath, Hothersall, Haygood et al., 2008) compatible with the Takara chaperone plasmids. To ensure long-term stability of both a tRNA plasmid and a chaperone plasmid in the cell, we needed to move one of the components to a different plasmid backbone. The five plasmids in the Takara chaperone set are pACYC-based, so we moved the tRNA genes into a plasmid from a different compatibility group (Novagen pCDF-1). The secondary plasmid pRare2 (Novagen), which carries tRNA genes corresponding to low-use codons in *E. coli* and has a copy number of 10, and the plasmid pCDF-1b (Novagen), which has a copy number of 20, were each digested with restriction endonucleases DrdI and XbaI. The larger fragment from pRare2 carrying the tRNA genes and the smaller fragment from pCDF-1b, which carries the origin of replication and the spectinomycin resistance marker gene, were each purified from a low-melting agarose gel. The two fragments were ligated to create pRARE2-CDF. This plasmid allows us to use chaperone supplementation at the same time as tRNA supplementation. Another advantage of the system is the copy-number increase from 10 to 20 for the plasmid bearing tRNA genes. Increasing the level of the tRNAs has been beneficial and does not compete at all with target protein production, as might occur if the level of chaperones were increased.

5.4 Co-expression of NRPS A Domains with MtbH-like Proteins

A strategy analogous to the co-expression of chaperones has been used to improve production of soluble NRPSs. MbtH-like proteins (MLPs) comprise a family of small proteins whose coding sequences are found in many NRPS gene clusters. Deletions of MLP-encoding genes provided genetic evidence that MLPs are essential for the production of nonribosomal-peptide natural products (Wolpert, Gust, Kammerer & Heide, 2007) while coexpression and purification identified them as integral components of NRPSs (Felnagle, Barkei, Park, Podevels, McMahon, Drott et al., 2010; Imker, Krahn, Clerc, Kaiser & Walsh, 2010; Miller, Drake, Shi, Aldrich & Gulick, 2016). It has since been shown that NRPSs from heterologous production found to be inactive or unobtainable due to insolubility could be rescued through coexpression of the cognate MLP-encoding gene (McMahon, Rush & Thomas, 2012). Cross-talk between MLPs and NRPSs from different pathways (Lautru, Oves-Costales, Pernodet & Challis, 2007) suggests that a similar effect may be obtained with non-cognate MLPs. Coexpression may be achieved by cloning a bicistronic insert of the NRPS and MLP genes into a standard expression vector such as pET (Imker et al., 2010) or through addition of a second compatible vector, such as pACYC Duet (McMahon et al., 2012) carrying the MLP gene. This coexpression strategy has become a common means to obtain soluble NRPS proteins where the use of fusions, low temperature and variations in induction conditions have failed (Zolova & Garneau-Tsodikova, 2012).

6. Growth Conditions

6.1 Media and pH

E. coli employed for heterologous protein production has been likened to a factory. As with factories, the level of production is dependent on the efficiency of operation. The best results are usually obtained by tailoring expression conditions to maintain bacterial cell growth within physiologically optimal conditions and by avoiding triggers of stress responses that divert energy and resources away from production of the target protein(s). The longstanding use of *E. coli* as a model organism provides a large body of literature to shape protocols and numerous well-characterized vectors from multiple complementation groups to use for expression and co-expression experiments.

Although most heterologous production of PKS and NRPS protein constituents is carried out in *E. coli*, alternative systems have been developed. A recently developed synthetic biology platform allows for the modular assembly of plasmid elements into cyanobacterial host-vector systems (Taton, Unglaub, Wright, Zeng, Paz-Yepes, Brahamsha et al., 2014). A shuttle vector was derived from the broad-host-range plasmid RSF1010, making it compatible with a broad range of cyanobacterial hosts. This system was employed to express a cyanobacterial P450 protein that could not be obtained from *E. coli* (Agarwal, Blanton, Podell, Taton, Schorn, Busch et al., 2017). Alternatively, the pSyn_6 vector (ThermoFisher), can be used to insert a gene of interest into a neutral site of the *Synechococcus elongatus* PCC 7942 genome via homologous recombination (Clerico, Ditty & Golden, 2007). *Rhizobium leguminosarum* has also been used as a heterologous host for expression of cyanobacterial proteins (Freeman, Help, Bhushan, Morinaka & Piel, 2017).

E. coli grow over a pH range from 5.5 to 9.0 (Padan, Zilberstein & Schuldiner, 1981). The metabolites from *E. coli* in culture can cause rapid and significant changes in the pH of an unbuffered growth medium. The carbon source affects the metabolites produced and their impact on pH (Losen, Frolich, Pohl & Buchs, 2004). Incorporating a buffer into the medium will prevent alkaline or acid stress responses from growth-induced shifts above pH 9.0 or below pH 5.5. We compared the growth of *E. coli* BL21(DE3) cultures in 96-well deep-well blocks in three media: unbuffered lysogeny broth (LB) (Bertani, 1951), buffered ZYM 5052 (Studier, 2005), and buffered Terrific Broth (TB) with glycerol (Tartoff, 1987). The 1.0-mL cultures grew to strikingly different final cell densities and pHs (Table 1). Cells in unbuffered LB went into alkaline stress and stopped growing at a fairly early time-point with the growth curve flattening at ~10 hours. For ZYM 5052 medium, the growth curve flattened at ~22 hours and for TB the growth rate had begun to slow when the experiment was stopped at 25 hours. Following this test and further optimization experiments, we settled on TB as our default medium for expression testing and larger-scale production. TB is a fermentation-type medium that avoids or reduces cellular stress from triggers for the general stress response (Storz & Hengee-Aronis, 2000) including carbon starvation, fluctuations in pH, low osmolarity and slow growth in late log phase approaching stationary phase. TB is a rich, complex medium containing relatively high amounts of yeast extract and tryptone, which ensures against carbon starvation and micronutrient deprivation at high density. TB is also a buffered medium, which reduces pH fluctuations. In our overnight growth studies we found that cultures in TB stay in log phase for over 20 hours. We usually harvest induction cultures at about 18 hours.

The standard recipe for TB contains phosphate buffer, but if a protein requires other additives that are not compatible with phosphate, which may cause precipitation, then HEPES, MOPS or another buffer with a pK_a near pH 7.4 can be used. TB recipes all include glycerol, generally at 0.4% but up to 4.0% for large-scale production or fermentation. We also use the higher 4.0% glycerol in our TB formulation (Table 2) in small-scale expression tests so the results will parallel large-scale production up to 20 L fermentation. Glycerol acts as an alternative carbon source, reduces pH fluctuations by altering the metabolite profile during growth (Losen et al., 2004), and may be a chemical chaperone promoting proper protein folding (Brown, Hong-Brown, Biwersi, Verkman & Welch, 1996; Rariy & Klibanov, 1997). We compared several target protein constructs for soluble production in three media and found that insoluble protein was produced in LB and ZYM 5052, whereas soluble protein was produced in TB (Lopanik et al., 2008; Buchholz et al., 2009).

6.2 Temperature

The 4-10 fold higher rates of transcription and translation in *E. coli* compared to eukaryotic cells may cause folding problems for eukaryotic proteins in *E. coli* (Widmann & Christen, 2000), leading to aggregation and/or proteolysis. Similarly, rapid production in an *E. coli* over-expression system can also cause folding problems for more complex bacterial proteins, particularly those that are not normally expressed to very high levels. This is a common problem for the multi-domain proteins of PKS and NRPS systems. Slowing the rate of production through reduced temperature is a widely used technique to limit aggregation and promote solubility during heterologous protein production (Schein, 1988).

E. coli grow over a temperature range from ~8°C (Broeze, Solomon & Pope, 1978) to ~49°C (Herendeen, VanBogelen & Neidhardt, 1979). Maintenance cultures and starter cultures are usually grown at 37°C. Production cultures are generally started at 37°C and grown to a specific density (OD_{600}), and then the temperature may be downshifted. Literature reports have a very wide variety of temperatures, ranging from 12°C to 37°C, for induction of protein production. *E. coli* has a linear growth range from 21°C to 37°C (Herendeen et al., 1979) within which the steady-state levels of protein do not vary. Outside this range growth rates decrease sharply, and the relationship of temperature and growth rate is nonlinear (Herendeen et al., 1979). We generally induce cultures after they have reached 20°C when downshifted from 37°C.

For small-scale (0.5 mL) cultures we allow an hour to equilibrate at 20°C before induction. The 17°C downshift in temperature is large enough to trigger cold-shock in these small cultures, which reach the target temperature very quickly. They emerge from cold-shock after an hour (Jones, Cashel, Glaser & Neidhardt, 1992; Thieringer, Jones & Inouye, 1998). Cold-shock is not a significant problem with larger cultures (500 mL to 1 L) since the culture itself does not change temperature as quickly as the incubator. Often, a 1 L culture may not have fully reached 20°C after one hour from the time of the downshift. It is essential that the culture itself has reached 20°C before induction, so we monitor the temperature of larger cultures, beginning at 1 hour post-downshift. We have observed that proteins prone to insolubility above the reduced growth temperature tend to go into the insoluble fraction if cultures are induced before reaching the reduced temperature. This suggests that the first protein made sets the pattern for protein made subsequently. Once the protein starts to misfold and aggregate at the higher temperature, subsequent protein made after the culture has reached the target temperature continues to misfold and aggregate even though it would fold and be soluble if it hadn't been induced until the culture was at the lower temperature.

Growth temperatures below 20°C may yield soluble protein, but as these temperatures are outside the linear range for *E. coli* growth, the results may not be reproducible. If the protein of interest is insoluble when produced at 20°C, we suggest shifting to an alternative weaker promoter (Section 6.4) to obtain a more extensive decrease in the rates of transcription and translation rather than shifting to even lower temperature. A weaker promoter combined with 20°C temperature will result in dramatic decreases in the speed and level of transcription as well as the rate of translation.

Temperature shifts are sometimes achieved by removing flasks from a shaker and placing in a cold-room or on ice for rapid cooling, with the great disadvantage that the cultures will become hypoxic, a stress trigger detrimental to producing soluble protein. Rapid temperature reduction also may place the cultures at temperatures below the growth range for *E. coli* and present risk of some cell death. Too rapid cooling may trigger cold shock, and the cells will be unresponsive to inducer until they acclimate to the final induction temperature, which for larger cultures may be longer than the time to cool naturally. In absence of refrigerated incubators, cultures can be shifted from a 37 °C incubator to one at room temperature to maintain optimal aeration. The cells may require more than 1 hour to reach the target temperature, but they should grow well and remain in log phase without significant stress.

6.3 Aeration

Aeration is another factor in obtaining optimal *E. coli* culture growth. To maximize the surface-to-volume ratio, the vessel should be wide and the volume of medium relatively shallow. For small-scale production to parallel large-scale production, shaking 50 mL conical tubes with 5 mL of medium produces well aerated cultures. When using 96-well blocks, a culture volume of 0.5 mL and a shake speed of at least 400 RPM will maximize aeration. For large-scale cultures in shake-flasks, the surface of the medium should not go above the point of maximum width of the flask (one-quarter volume of an Erlenmeyer flask, *i.e.* 1 L of medium in a 4 L flask; 800-1000 mL for a 2.8 L Fernbach flask), and the shake speed should be at least 300 RPM. When using TB, optimal aeration is important to take full advantage of the achievable high cell densities.

6.4 Induction

Considerations for inducing heterologous gene expression include choice of inducer, inducer concentration and time of induction. Lactose, the native inducer for the most widely used expression systems, is much less expensive than isopropyl β-ᴅ-1-thiogalactopyranoside (IPTG). Lactose can be present in the medium at the time the culture is inoculated, but induction is delayed until other carbon sources have been exhausted (Studier, 2005). Auto-induction requires less handling, an advantage for high-throughput expression trials. The timing of induction in auto-induction medium is dependent on a number of factors including plasmid composition, arrangement of promoter and repressor elements, culture aeration during growth, and auto-induction medium composition (Blommel, Becker, Duvnjak & Fox, 2007). Proper coordination of the timing of induction with the change in growth temperature is critical to produce soluble protein and requires significant effort to perfect. For this reason, we induce with added IPTG, which enters the cell through passive diffusion, resulting in nearly immediate induction after addition.

We measured protein production as a function of inducer concentration for various promoters at various temperatures (Table 3). At levels of inducer commonly reported in the literature (400 μM up to 1 mM), protein production was lower than with the optimal concentrations. Titrating the inducer concentration is another way to reduce the rate of transcription, especially with a native *E. coli* promoter where limiting the inducer has a continuing impact on the native promoter. With a T7 system, limiting the inducer will slow the production of only the T7 RNA polymerase; once T7 RNA polymerase saturates the available plasmid, the system will be insensitive to the inducer limitation.

We studied the effect of culture optical density (OD_{600}) at the time of induction on protein yield. Literature reports suggest OD_{600} values for induction ranging from 0.6 to 1.0. We grew parallel cultures in TB, induced production beginning at an OD_{600} of 0.6 then at OD_{600} of 1.0 and at intervals of 1.0 OD_{600} unit up to an OD_{600} of 7.0, grew the cultures overnight, purified the target protein by metal-affinity, and compared yields by SDS-PAGE analysis. We observed no significant difference in yield between the cultures induced at the lowest ODs and those induced at the highest. These experiments were run with MBP as the target protein. As MBP is made well, these results may not extrapolate completely to a more problematic protein, but they suggest that the culture density at induction is not a critical

factor in the success of the production. We generally induce at an OD_{600} of roughly 1.0 to 2.0.

Based on the considerations summarized in sections 6.1-6.4, we developed an optimized set of conditions for initial production of PKS and NRPS proteins (Table 4). Variations of these conditions are used as needed.

6.5 Additives

When produced alone in an over-expression system, some proteins do not fold quickly or completely enough for purification from the soluble fraction of a cell lysate. For such proteins, proper folding may be stimulated by co-expression with a natural partner protein. The production of soluble protein for the dehalogenase Bmp8 from *M. mediterranea* MMB1 required co-expression of the thioesterase Bmp1 from the same pathway (El Gamal, Agarwal, Rahman & Moore, 2016). It was reasoned that, as Bmp1 is the substrate for Bmp8, the two proteins form a functional complex at some point and interaction with Bmp1 could help Bmp8 fold to its native stable form in an over-expression system. Similarly, production of soluble halogenase CylC from the cylindrocyclophane biosynthetic gene cluster required co-expression with the partner fatty acid-ACP ligase (CylA) and ACP (CylB) (Nakamura, Schultz & Balskus, 2017). Untagged CylC was co-purified with tagged CylB, indicating formation of a halogenase-ACP complex, which required prior formation of fatty acyl-CylB by CylA.

Other proteins may be stabilized *in vivo* by the addition of a ligand, either a co-factor or an agonist or antagonist of activity. A group of small organic compounds have been identified as "chemical chaperones" based on their reversal of protein mis-localization or aggregation (Brown et al., 1996; Perlmutter, 2002). They have since been adapted to improve recombinant protein folding (Prasad, Khadatare & Roy, 2011). The group includes polyols such as glycerol (discussed above), triethylamines such as triethylamine-*N*-oxide (TMAO), some amino acid derivatives, and sugars such as trehalose and sorbitol. Chemical chaperones may be employed to enhance folding *in vivo* through addition to the growth medium (100 mM–1 M) or *in vitro* by addition to the lysis and purification buffers. When used at lysis, the chemical chaperones or co-solvents, such as kosmotropic salts or detergents, may rescue partially folded proteins, prevent aggregation and partitioning into the insoluble fraction, and assist in achieving a native fold (Lindwall, Chau, Gardner & Kohlstaedt, 2000; Bondos & Bicknell, 2003). The presence of the osmolytes is thought to stabilize the properly folded regions of the protein and alter solvent interactions in unfolded regions to favor a more compact and folded structure. Kosmotropic salts are thought to work in a similar fashion by modulating the interactions of the protein surface with solvent to favor more compact and folded structures.

Proteins requiring cofactors or ligands for stability may benefit from the addition of the cofactor, ligand or a precursor at the time of induction. For example, cultures producing flavin-dependent enzymes can be stabilized by addition of 50 μM riboflavin, a precursor to the cofactors FAD and FMN. Similarly, cultures co-expressing *cylC* (halogenase), *cylA* (fatty acyl-ACP ligase) and *cylB* (ACP) required supplementation with the CylA substrate

decanoic acid in order to produce stable CylC as a complex with CylB, which was produced as the decanoyl-ACP (Nakamura et al., 2017).

Occasionally components of PKS or NRPS pathways are membrane associated, such as fatty acid desaturases (Zhu, Liu & Zhang, 2015; Zhu, Su, Manickam & Zhang, 2015) and ε-poly-L-lysine synthetase (Yamanaka, Maruyama, Takagi & Hamano, 2008). Such proteins are assayed in the membrane fraction or extracted with detergents.

7. Production of ACPs in Defined State

For functional and structural studies, it is usually necessary to produce acyl carrier proteins (ACPs) or peptidyl carrier proteins (PCPs) in a defined state such as the inactive apo-ACP or the native activated ACP with either a 4'-phosphopantetheine (Ppant) arm (holo-ACP) or modified 4'-phosphopantetheine arm (crypto-ACP). The latter crypto-ACPs containing unnatural acyl substrates on the Ppant arm of the ACP have been widely used for functional studies to elucidate the substrate specificity of PKS and NRPS domains (Belshaw, Walsh & Stachelhaus, 1999; Sieber, Walsh & Marahiel, 2003; Vitali, Zerbe & Robinson, 2003; Fridman, Balibar, Lupoli, Kahne, Walsh & Garneau-Tsodikova, 2007; El Gamal, Agarwal, Diethelm, Rahman, Schorn, Sneed et al., 2016). Modification of carrier proteins with reactive acyl groups incorporating epoxides, chloroacetates, imidazoles, vinylsulfonamides, and 3-alkynyl sulfones have facilitated cross-linking of ACPs with other domains for structural studies (Worthington, Rivera, Torpey, Alexander & Burkart, 2006; Liu & Bruner, 2007; Haushalter, Worthington, Hur & Burkart, 2008; Kapur, Worthington, Tang, Cane, Burkart & Khosla, 2008; Hur, Meier, Baskin, Codelli, Bertozzi, Marahiel et al., 2009; Worthington, Porter & Burkart, 2010; Liu, Zheng & Bruner, 2011; Mitchell, Shi, Aldrich & Gulick, 2012; Sundlov, Shi, Wilson, Aldrich & Gulick, 2012; Ishikawa, Haushalter, Lee, Finzel & Burkart, 2013; Haslinger, Brieke, Uhlmann, Sieverling, Süssmuth & Cryle, 2014; Drake, Miller, Shi, Tarrasch, Sundlov, Allen et al., 2016). While modification of ACPs with bioorthogonal handles, fluorophores, and biotin have been explored for a variety of chemical biology applications (La Clair, Foley, Schegg, Regan & Burkart, 2004; Clarke, Mercer, La Clair & Burkart, 2005; Meier, Mercer, Rivera & Burkart, 2006).

ACPs are typically isolated as a mixture of apo-ACP and holo-ACP during expression in *E. coli*. To ensure complete conversion to the holo-form, we co-express with a phosphopantetheinyl transferase from a low-copy plasmid following the general protocol described in Methods in Enzymology (Sunbul, Zhang & Yin, 2009) or in the Bap1 cell line, in which a phosphopantetheinyl transferase is integrated into the *E. coli* genome (Pfeifer & Khosla, 2001). The promiscuous Sfp from *Bacillus subtilis* has been the most widely employed because of its high catalytic activity and relaxed substrate specificity (Lambalot, Gehring, Flugel, Zuber, LaCelle, Marahiel et al., 1996), but *Streptomyces verticillus* Svp has proven to be equally versatile (Sánchez, Du, Edwards, Toney & Shen, 2001). Additionally, an Sfp mutant (R4-4 Sfp) will load the simpler and more synthetically accessible 3-dephospho-CoA analogues (Zou & Yin, 2009; Kittilä & Cryle, 2017).

Preparation of apo-ACPs required for synthesis of crypto-ACPs on the other hand can be more challenging due to the presence of the endogenous phosphopantetheinyl transferase

EntD, which we have found in some cases results in partial phosphopantetheinylation of recombinantly expressed ACPs. The *E. coli* gene *entD* is part of an operon that encodes for enterobactin, an NRPS-derived siderophore (small molecule iron-chelator), whose biosynthetic genes are induced under iron restricted conditions (Lambalot et al., 1996). The LB and TB media commonly used for recombinant protein production do not contain added iron and may become iron depleted as the medium is exhausted during growth, causing induction of *entD* and other genes required for enterobactin biosynthesis and transport. Indeed, the production of enterobactin in LB medium has been observed, indicating expression of *entD* (Goetz, Holmes, Borregaard, Bluhm, Raymond & Strong, 2002).

We have found that apo-ACPs can be reproducibly obtained by addition of iron to the medium along with other trace metals to repress induction of *entD*. Our general protocol involves supplementation of 1 L of medium containing the appropriate antibiotic with 1 mL of a 1000× trace metals mix to provide final concentrations of the following trace metals: iron (50 μM), calcium (20 μM), manganese (10 μM), zinc (10 μM), cobalt (2 μM), copper (2 μM), nickel (2 μM), molybdenum (2 μM), selenium (2 μM), and boron (2 μM) (see Table 5). This method is noted for its experimental simplicity and effectiveness with a wide variety of ACPs from PKS and NRPS systems (Maloney et al., 2016; Skiba et al., 2016; Skiba et al., 2017).

A complimentary approach to generate apo-ACP employs an ACP hydrolase (AcpH) to cleave the Ppant arm of holo-ACPs (Kosa, Haushalter, Smith & Burkart, 2012). AcpH orthologs from *E. coli, Pseudomonas aeruginosa, Shewanella oneidensis*, and *P. fluorescens* have been investigated and shown to catalyze phosphopantetheine hydrolysis of holo-ACPs from PKS, NRPS, and fatty-acid synthase (FAS) proteins. However, these AcpHs are marred by variable activity toward PKS and NRPS proteins (Kosa, Pham & Burkart, 2014). Despite these clear limitations, AcpH provides an interesting alternative method for phosphopantetheine removal.

8. Production of Multi-Domain Proteins

Multi-domain proteins pose special problems for over-expression systems because domains generally fold independently and may do so at very different rates, posing a risk of aggregation before all domains are folded. The risk of aggregation is increased if any of the domains are discontinuous in the amino acid sequence (Figure 1), as more polypeptide must emerge from the ribosome before folding is complete.

A deficiency in production of recombinant cyanobacterial PKS and NRPS multi-domain proteins in *E. coli* compared to those from other bacterial species (Roberts, Staunton & Leadlay, 1993; Yuzawa, Eng, Katz & Keasling, 2013; Dutta, Whicher, Hansen, Hale, Chemler, Congdon et al., 2014; Edwards, Matsui, Weiss & Khosla, 2014; Whicher, Dutta, Hansen, Hale, Chemler, Dosey et al., 2014; Miller et al., 2016; Lowell, DeMars, Slocum, Yu, Anand, Chemler et al., 2017; Tarry, Haque, Bui & Schmeing, 2017) indicates that the cyanobacterial targets may require additional strategies. We encountered problems with several cyanobacterial multi-domain proteins when using T7-based expression systems that had been used successfully for actinobacterial multi-domain proteins. Slowing the rate of

protein production by use of LIC-compatible expression plasmids containing the trc promoter (Smith, Brown, Harms & Smith, 2015), and induction with 200 μM IPTG at 20°C as described in Section 3.3, led to robust expression levels for several previously problematic proteins. Coupled rates of transcription and translation provided by the native *E. coli* RNA polymerase rather than the more rapid T7 RNA polymerase, may have enhanced production of the large multi-domain proteins in soluble form. Previously, the *trc* promoter was used successfully for the expression of the NRPS SrfA-C (Tanovic, Samel, Essen & Marahiel, 2008).

Some recombinant multi-domain proteins were insoluble even with use of the *trc* promoter and an MBP solubility tag. In some cases, much improved levels of soluble protein were achieved by supplementing the growth medium with 250-500 mM of the osmolyte TMAO, which promotes protein folding (Bandyopadhyay, Saxena, Kasturia, Dalal, Bhatt, Rajkumar et al., 2012) and is found in marine organisms (Yancey, Clark, Hand, Bowlus & Somero, 1982). Ni-affinity chromatography, followed by cleavage of the fused solubility tag and size exclusion chromatography yielded multiple milligrams of >90% pure protein per liter of culture.

Additional strategies have been applied to successfully express constructs encoding multi-domain PKS and NRPS proteins. In the case of HMWP1 and HMWP2 from the yersiniabactin synthetase, reducing the rate of protein production by lowering growth temperatures to 16°C -18°C and reducing inducer concentration (50-75 μM IPTG) yielded soluble protein (Keating, Miller & Walsh, 2000; Suo, Tseng & Walsh, 2001). Alternatively, coexpression with *Streptomyces* chaperones has also facilitated expression (Herbst, Jakob, Zahringer & Maier, 2016) and improved activity (Betancor, Fernandez, Weissman & Leadlay, 2008).

9. Cell Lysis

To achieve high yields of active recombinant protein, it is critical to achieve full cell lysis under conditions appropriate for stabilizing the protein of interest. Incomplete lysis can be a problem with heterologous expression in *E. coli*, due to presence of a rigid peptidoglycan layer. (For heterologous expression in insect cells, incomplete lysis is rarely a problem.) Resuspension of cells in 4 mL of lysis buffer containing 0.05-0.2 mg/mL lysozyme per gram of cell paste assists in degrading the *E. coli* peptidoglycan layer. To achieve complete lysis, physical methods such as homogenization or sonication are often used. However, temperature should be monitored during lysis to prevent over-heating of the lysate, which may damage the target protein. In the cell lysis process, we add Benzonase nuclease (MilliporeSigma) or DNase to degrade nucleic acid, reduce viscosity, and facilitate later purification steps. Cofactor-dependent proteins can be provided with the cofactor at the cell lysis step if this helps stabilize the protein (Fraley et al., 2017).

Cell lysis is a drastic change from the cellular environment to a rather simple buffer, which can stress a delicate target protein causing denaturation, proteolysis, and/or aggregation. Protease inhibitors added during cell lysis and purification can inhibit proteolysis. For

problems with protein stability and aggregation, we have had success with a systematic screening process to identify conditions that promote protein stabilization.

The perception that protein aggregation during over-production in *E. coli* is driven by formation of a "glob" of unfolded polypeptides has given way to the understanding that aggregates may be highly ordered and relatively close to their native state but have become associated through non-native β-sheet structures (Ventura & Villaverde, 2006). The current view suggests that aggregated proteins could be dissociated and encouraged to complete folding to their native state. An indicator that a protein may have rescue potential is the partitioning of the majority of the protein to the insoluble fraction of the cell lysate with a small portion in the soluble fraction. Proteins with this profile may be candidates for a lysis buffer screen in which a variety of chemical co-solvents are added to the buffer to disaggregate the protein and promote greater solubility.

Based on literature reports (Lindwall et al., 2000; Bondos et al., 2003), we designed a 96-condition screen and tested it on proteins from several natural product biosynthetic pathways. While the more extreme conditions displayed negative effects, a number of conditions appeared to have positive effects in a variety of different combinations. We identified a small set of components for use in focused combinatorial screening rather than a large set of standard conditions (Table 6). The elements that have shown the most success in increasing soluble yield include: pH, ionic salts, kosmotropic salts, organic co-solvents, and detergents.

Components of the lysis buffer can have a large impact on protein stability by affecting physical and chemical properties of the buffer, by interfering with interactions between solvent molecules and the target protein, or by direct interactions with the protein. As pH affects protein charge state, finding a suitable pH is of vital importance for many proteins. We use pH 7.0 and pH 8.5 in initial trials. Ionic salts interact with protein surface charges and reduce electrostatic interactions, while kosmotropic salts impact protein-solvent interactions in favor of the native state (Bondos et al., 2003). Organic co-solvents weaken hydrophobic interactions and may disaggregate proteins or cause localized unfolding. Detergents disrupt hydrophobic interactions and provide interactions with the solvent. Arginine has been shown to have a stabilizing effect on several proteins, and is proposed to reduce hydrophobicity by interacting with aromatic amino acids on the protein surface (Ito, Shiraki, Matsuura, Okumura, Hasegawa, Baba et al., 2011).

In order to screen lysis conditions for improved protein solubility, we developed a 14-condition small-scale lysis screen (Table 7, Figure 2). To carry out the buffer screen, a His-tagged target protein is over-expressed in *E. coli* cell culture. The culture is split into 14 equal-volume aliquots (1 mL), and cells in each aliquot are resuspended in one of the test buffers and lysed by sonication. After separation, the soluble and insoluble fractions are assayed by SDS-PAGE. An optimal buffer is selected for use in large-scale purification. The lysis buffer screen not only benefits protein purification, but also provides useful information on buffer choice for biochemical assays (Fiers, Dodge, Li, Smith, Fecik & Aldrich, 2015; Fraley et al., 2017).

10. Protein Purification

Once appropriate conditions for cell lysis have been identified, protein purification can proceed. In general, the first purification step of a His-tagged protein involves affinity chromatography, such as metal affinity chromatography for a His-tagged protein. The inclusion of a low level of imidazole (20 mM) in buffers used during loading and wash steps can inhibit non-specific binding of *E. coli* proteins to the resin, and improve purity of the target protein during the first purification step. Use of an imidazole gradient or stepwise elution with increasing imidazole concentrations can improve protein purity over the simpler wash-and-elute protocol. If the target protein is sensitive to imidazole, the use of cobalt resin rather than nickel resin allows elution of the protein at lower imidazole concentrations. Despite the near universal application of metal-affinity resins to His-tagged proteins, complications can ensue. Care must be taken to avoid a protein buffer containing strong chelators that could strip metal from the resin. Similarly, weakly bound native metals can be stripped from metalloproteins by the high imidazole concentrations used to elute proteins from the resin. Disulfide reducing agents may be essential to the stability of some proteins, but can be deleterious to metal affinity resins; if a reducing agent is needed, the resin manufacturer's instructions should be consulted for compatibility.

In cases where the gene of interest is co-expressed with a chaperone-bearing plasmid, the target protein may retain significant amounts of chaperone during affinity purification. Addition of Mg-ATP can cause the chaperone to release the target protein (Maloney et al., 2016; Fraley et al., 2017). This can be done while the target protein is bound to metal-affinity resin, which should not bind the chaperone. Size exclusion chromatography can also separate the target protein from chaperones. The addition of ATP alone is not always effective in releasing the chaperone from the target protein. Another release method exploits the affinity of GroEL for denatured protein, which acts as a competitor of the target protein (Rohman & Harrison-Lavoie, 2000). The addition of urea- or heat-denatured protein with ATP was found to be necessary for the successful purification of the A-domain of MtbA away from bound chaperones (Somu, Wilson, Bennett, Boshoff, Celia, Beck et al., 2006).

A solubility fusion partner or N-terminal His-tag can be removed from the target by incubation with TEV protease in dialysis to remove imidazole. The tag-free target protein can be separated from uncleaved protein and the fusion partner by a second pass over metal affinity resin. As TEV protease is a cysteine protease, a low level of reducing agent is added to the cleavage reaction to maintain protease activity, but not interfere with any additional metal affinity steps. If the protein of interest is sensitive to the presence of strong reducing agents, a glutathione redox buffer can be substituted during the cleavage step to maintain protease activity (Saez, Cristofori-Armstrong, Anangi & King, 2017).

A single affinity-chromatography step can yield protein of sufficient purity for many applications. However, greater purity is usually required for crystallization. We generally include a second orthogonal step of size-exclusion chromatography, which provides information about protein oligomer state and monodispersity and also eliminates soluble aggregates. Some biochemical assays are susceptible to interference from contaminating proteins such as ATPases (Drake, Duckworth, Neres, Aldrich & Gulick, 2010) or

acyltransferases (Florova, Kazanina & Reynolds, 2002), particularly when carrier proteins are used as substrates or in single-turnover experiments. For such applications, most remaining trace contaminants can be eliminated by a third purification step, for example ion-exchange chromatography.

11. Refinements for Crystallization

Even after conditions to produce stable recombinant protein have been established, an ensuing crystallization campaign may require additional refinements. For crystallization of enzymes from natural product biosynthetic pathways, we have found successful strategies include screening multiple homologs of a target, altering the N- or C-terminus of the protein, co-crystallization with substrates/cofactors, and targeted mutagenesis.

The addition of cofactor or removal of a buffer component, such as glycerol, may make a protein more amenable to crystallization in sparse matrix screening. In cases where poorly diffracting crystals have been obtained, additional components identified through additive screening can greatly improve crystal quality (Bergfors, 2007). For example, initial CurJ CMT crystals were multi-crystal spherulites. The addition of acetone, identified in a round of additive screening, produced poorly diffracting crystals (8 Å). Further addition of glutathione, identified in a second round of additive screening, yielded single crystals that diffracted to 2 Å (Skiba et al., 2016).

Mutagenesis may be considered in cases where crystallization trials fail, when poor crystal quality precludes structure solution, or where crystallization is too unreliable for a study requiring structures in several states of catalysis or substrate binding. For several projects in our lab, changes in the boundaries of an expression construct were crucial to producing protein amenable to crystallization. Exploration of polypeptide chain termini is strongly recommended for proteins excised from a PKS or NRPS megasynthase and intended for crystallization. As described in section 2.3, exploration of domain boundaries was necessary to produce soluble CurJ CMT, however, among several soluble CMT variants, only one crystallized (Skiba et al., 2016). Similarly, the GNAT-MT2 didomain from the apratoxin loading module was produced in soluble form from a construct that, upon comparison to the CurJ CMT structure, included a large C-terminal linker region. Deletion of the linker led to successful crystallization. Finally, a change of even a single amino acid at the terminus may affect crystallization. Initial crystallization trials of the CurF enoylreductase failed, but removal of the C-terminal amino acid yielded crystals diffracting to 0.96 Å (Khare, Hale, Tripathi, Gu, Sherman, Gerwick et al., 2015).

A strategy known as surface entropy reduction (SER) has been effective in cases where crystals are not obtained. SER, developed for proteins that cannot be induced to leave solution, aims to replace flexible amino acid side chains on the protein surface with alanine to reduce the entropic penalty of forming a crystal contact at that surface (Goldschmidt, Cooper, Derewenda & Eisenberg, 2007). Predicted SER substitutions may be compared to a homolog of known structure (if available) to eliminate candidate substitutions that are near the active site, internal, or at an oligomer interface. SER is typically applied in cases where a target is exceedingly soluble, as exemplified by the CurM sulfotransferase (McCarthy,

Eisman, Kulkarni, Gerwick, Gerwick, Wipf et al., 2012). The CurD HMGS is a more unorthodox example of SER-assisted crystallization (Maloney et al., 2016). CurD is marginally stable and utterly insoluble unless it is produced with chaperones. Extensive screening of HMGS proteins from four biosynthetic pathways yielded no crystals. Despite the apparent instability of these proteins, a SER variant of CurD was stable, yielded diffraction-quality crystals in hours, and had activity comparable to the wild type. The SER substitutions are near a crystal contact but do not directly participate, contrary to the usual logic for SER. Occasionally, mutagenesis can be serendipitous. Initial crystallization conditions for AprA MT1-ΨGNAT were obtained from a protein containing two PCR-induced substitutions (S274I, Q528P) (Skiba et al., 2017). The substitutions are near crystal contacts in two different crystal forms, and the wild type protein did not crystallize. Similarly, MalA from the terrestrial fungus *Malbranchea aurantiaca* RRC1813 was recalcitrant to crystallization. However, the 99% identical homolog MalA' from the marine fungus *Malbranchea graminicola* (086937A), which differs at two amino acid positions (L276P and R428P), yielded crystals (Fraley et al., 2017). The Pro428 occurs in a crystal contact that would be incompatible with an Arg side chain.

In some cases, mutagenesis to repair poor crystal lattice contacts can solve a problem of poor crystal reproducibility. This may be considered when the scope of a study extends beyond a single structure, requiring the capture of multiple states or substrate/cofactor bound complex. This was the case for the MycF *O*-methyltransferase (Bernard, Akey, Tripathi, Park, Konwerski, Anzai et al., 2015), for which crystals grew in multiple space groups and typically had poor-quality diffraction. In the initial MycF structure, we discovered that repulsion of two glutamate side chains in a crystal contact was likely the culprit for this behavior. A double glutamine/alanine substitution of these residues reproducibly yielded crystals that diffracted beyond 2 Å, allowing three new substrate complexes to be captured. Targeted mutagenesis may also be used to introduce methionine substitutions in proteins where they are lacking and are needed for crystallographic phasing (Newmister, Li, Garcia-Borràs, Sanders, Yang, Lowell et al., 2018).

12. Summary

The production of recombinant proteins is essential for elucidating the biochemical steps of PKS and NRPS pathways, and for solving structures of the component enzymes. However, poor gene expression, protein solubility or protein stability is often an impediment. We and others have identified several strategies to overcome such obstacles during the cloning, expression and purification processes. After initial protein production, additional steps can improve efforts towards structural characterization. Here we have described many strategies that led to the successful production of protein targets from PKS and NRPS pathways, but this is not an exhaustive list. Challenging protein targets require careful screening of multiple constructs and conditions for expression and purification, along with an ounce of creativity.

Acknowledgments

References

Agarwal V, Blanton JM, Podell S, Taton A, Schorn MA, Busch J, et al. Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. Nat Chem Biol. 2017; 13(5):537–543. [PubMed: 28319100]

Akey DL, Gehret JJ, Khare D, Smith JL. Insights from the sea: structural biology of marine polyketide synthases. Nat Prod Rep. 2012; 29(10):1038–1049. [PubMed: 22498975]

Akey DL, Razelun JR, Tehranisa J, Sherman DH, Gerwick WH, Smith JL. Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. Structure. 2010; 18(1):94–105. [PubMed: 20152156]

Amann E, Brosius J, Ptashne M. Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in. Escherichia coli Gene. 1983; 25(2-3):167–178. [PubMed: 6363212]

Aparicio JF, Caffrey P, Marsden AF, Staunton J, Leadlay PF. Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. J Biol Chem. 1994; 269(11):8524–8528. [PubMed: 8132579]

Banda S, Tiwari PB, Darici Y, Tse-Dinh YC. Investigating direct interaction between Escherichia coli topoisomerase I and RecA. Gene. 2016; 585(1):65–70. [PubMed: 27001450]

Bandyopadhyay A, Saxena K, Kasturia N, Dalal V, Bhatt N, Rajkumar A, et al. Chemical chaperones assist intracellular folding to buffer mutational variations. Nat Chem Biol. 2012; 8(3):238–245. [PubMed: 22246401]

Belshaw PJ, Walsh CT, Stachelhaus T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. Science. 1999; 284(5413):486–489. [PubMed: 10205056]

Bergfors T. Screening and optimization methods for nonautomated crystallization laboratories. Methods Mol Biol. 2007; 363:131–151. [PubMed: 17272840]

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28(1):235–242. [PubMed: 10592235]

Bernard SM, Akey DL, Tripathi A, Park SR, Konwerski JR, Anzai Y, et al. Structural basis of substrate specificity and regiochemistry in the MycF/TylF family of sugar O-methyltransferases. ACS Chem Biol. 2015; 10(5):1340–1351. [PubMed: 25692963]

Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res. 2007; 35(6):e45. [PubMed: 17317681]

Bertani G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic. Escherichia coli J Bacteriol. 1951; 62(3):293–300. [PubMed: 14888646]

Betancor L, Fernandez MJ, Weissman KJ, Leadlay PF. Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after coexpression with *Streptomyces* chaperonins in *Escherichia coli*. Chembiochem. 2008; 9(18):2962–2966. [PubMed: 19021139]

Blommel PG, Becker KJ, Duvnjak P, Fox BG. Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. Biotechnol Prog. 2007; 23(3):585–598. [PubMed: 17506520]

Bondos SE, Bicknell A. Detection and prevention of protein aggregation before, during, and after purification. Anal Biochem. 2003; 316(2):223–231. [PubMed: 12711344]

Broeze RJ, Solomon CJ, Pope DH. Effects of low temperature on *in vivo* and *in vitro* protein synthesis in *Escherichia coli* and *Pseudomonas fluorescens*. J Bacteriol. 1978; 134(3):861–874. [PubMed: 96102]

Brown CR, Hong-Brown LQ, Biwersi J, Verkman AS, Welch WJ. Chemical chaperones correct the mutant phenotype of the delta F508 cystic fibrosis transmembrane conductance regulator protein. Cell Stress Chaperones. 1996; 1(2):117–125. [PubMed: 9222597]

Buchholz TJ, Geders TW, Bartley FE 3rd, Reynolds KA, Smith JL, Sherman DH. Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. ACS Chem Biol. 2009; 4(1):41–52. [PubMed: 19146481]

Chao YP, Chiang CJ, Hung WB. Stringent regulation and high-level expression of heterologous genes in *Escherichia coli* using T7 system controllable by the araBAD promoter. Biotechnol Prog. 2002; 18(2):394–400. [PubMed: 11934312]

Clarke KM, Mercer AC, La Clair JJ, Burkart MD. In vivo reporter labeling of proteins via metabolic delivery of coenzyme A analogues. J Am Chem Soc. 2005; 127(32):11234–11235. [PubMed: 16089439]

Clerico EM, Ditty JL, Golden SS. Specialized techniques for site-directed mutagenesis in cyanobacteria. Methods Mol Biol. 2007; 362:155–171. [PubMed: 17417008]

Costa S, Almeida A, Castro A, Domingues L. Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. Front Microbiol. 2014; 5:63. [PubMed: 24600443]

DelProposto J, Majmudar CY, Smith JL, Brown WC. Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. Protein Expr Purif. 2009; 63(1):40–49. [PubMed: 18824232]

Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, et al. An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. Protein Expr Purif. 2006; 47(2):446–454. [PubMed: 16497515]

Drake EJ, Duckworth BP, Neres J, Aldrich CC, Gulick AM. Biochemical and structural characterization of bisubstrate inhibitors of BasE, the self-standing nonribosomal peptide synthetase adenylate-forming enzyme of acinetobactin synthesis. Biochemistry. 2010; 49(43): 9292–9305. [PubMed: 20853905]

Drake EJ, Miller BR, Shi C, Tarrasch JT, Sundlov JA, Allen CL, et al. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. Nature. 2016; 529(7585):235–238. [PubMed: 26762461]

Dutta S, Whicher JR, Hansen DA, Hale WA, Chemler JA, Congdon GR, et al. Structure of a modular polyketide synthase. Nature. 2014; 510(7506):512–517. [PubMed: 24965652]

Edwards AL, Matsui T, Weiss TM, Khosla C. Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. J Mol Biol. 2014; 426(11):2229–2245. [PubMed: 24704088]

El Gamal A, Agarwal V, Diethelm S, Rahman I, Schorn MA, Sneed JM, et al. Biosynthesis of coral settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase-thioesterase enzyme pair. Proc Natl Acad Sci U S A. 2016; 113(14):3797–3802. [PubMed: 27001835]

El Gamal A, Agarwal V, Rahman I, Moore BS. Enzymatic reductive dehalogenation controls the biosynthesis of marine bacterial pyrroles. J Am Chem Soc. 2016; 138(40):13167–13170. [PubMed: 27676265]

Eschenfeldt WH, Maltseva N, Stols L, Donnelly MI, Gu M, Nocek B, et al. Cleavable C-terminal His-tag vectors for structure determination. J Struct Funct Genomics. 2010; 11(1):31–39. [PubMed: 20213425]

Fage CD, Isiorho EA, Liu Y, Wagner DT, Liu HW, Keatinge-Clay AT. The structure of SpnF, a standalone enzyme that catalyzes [4 + 2] cycloaddition. Nat Chem Biol. 2015; 11(4):256–258. [PubMed: 25730549]

Felnagle EA, Barkei JJ, Park H, Podevels AM, McMahon MD, Drott DW, et al. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. Biochemistry. 2010; 49(41): 8815–8817. [PubMed: 20845982]

Fiers WD, Dodge GJ, Li Y, Smith JL, Fecik RA, Aldrich CC. Tylosin polyketide synthase module 3: stereospecificity, stereoselectivity and steady-state kinetic analysis of beta-processing domains via diffusible, synthetic substrates. Chem Sci. 2015; 6(8):5027–5033. [PubMed: 26366283]

Florova G, Kazanina G, Reynolds KA. Enzymes involved in fatty acid and polyketide biosynthesis in *Streptomyces* glaucescens: role of FabH and FabD and their acyl carrier protein specificity. Biochemistry. 2002; 41(33):10462–10471. [PubMed: 12173933]

Fox JD, Routzahn KM, Bucher MH, Waugh DS. Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. FEBS Lett. 2003; 537(1-3):53–57. [PubMed: 12606030]

Fraley AE, Garcia-Borràs M, Tripathi A, Khare D, Mercado-Marin EV, Tran H, et al. Function and structure of MalA/MalA', iterative halogenases for late-stage C-H functionalization of indole alkaloids. J Am Chem Soc. 2017; 139(34):12060–12068. [PubMed: 28777910]

Freeman MF, Helf MJ, Bhushan A, Morinaka BI, Piel J. Seven enzymes create extraordinary molecular complexity in an uncultivated bacterium. Nat Chem. 2017; 9(4):387–395. [PubMed: 28338684]

Fridman M, Balibar CJ, Lupoli T, Kahne D, Walsh CT, Garneau-Tsodikova S. Chemoenzymatic formation of novel aminocoumarin antibiotics by the enzymes CouN1 and CouN7. Biochemistry. 2007; 46(28):8462–8471. [PubMed: 17580964]

Gehret JJ, Gu L, Geders TW, Brown WC, Gerwick L, Gerwick WH, et al. Structure and activity of DmmA, a marine haloalkane dehalogenase. Protein Sci. 2012; 21(2):239–248. [PubMed: 22124946]

Gehret JJ, Gu L, Gerwick WH, Wipf P, Sherman DH, Smith JL. Terminal alkene formation by the thioesterase of curacin A biosynthesis: structure of a decarboxylating thioesterase. J Biol Chem. 2011; 286(16):14445–14454. [PubMed: 21357626]

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009; 6(5):343–345. [PubMed: 19363495]

Goetz DH, Holmes MA, Borregaard N, Bluhm ME, Raymond KN, Strong RK. The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. Mol Cell. 2002; 10(5):1033–1043. [PubMed: 12453412]

Goldschmidt L, Cooper DR, Derewenda ZS, Eisenberg D. Toward rational protein crystallization: A web server for the design of crystallizable protein variants. Protein Sci. 2007; 16(8):1569–1576. [PubMed: 17656576]

Grindberg RV, Ishoey T, Brinza D, Esquenazi E, Coates RC, Liu WT, et al. Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. PLoS One. 2011; 6(4):e18565. [PubMed: 21533272]

Gu L, Geders TW, Wang B, Gerwick WH, Hakansson K, Smith JL, et al. GNAT-like strategy for polyketide chain initiation. Science. 2007; 318(5852):970–974. [PubMed: 17991863]

Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol. 1995; 177(14):4121–4130. [PubMed: 7608087]

Hagen A, Poust S, Rond T, Fortman JL, Katz L, Petzold CJ, et al. Engineering a polyketide synthase for in vitro production of adipic acid. ACS Synth Biol. 2016; 5(1):21–27. [PubMed: 26501439]

Hartl FU. Molecular chaperones in cellular protein folding. Nature. 1996; 381(6583):571–579. [PubMed: 8637592]

Haslinger K, Brieke C, Uhlmann S, Sieverling L, Süssmuth RD, Cryle MJ. The structure of a transient complex of a nonribosomal peptide synthetase and a cytochrome P450 monooxygenase. Angew Chem Int Ed Engl. 2014; 53(32):8518–8522. [PubMed: 25044735]

Haushalter RW, Worthington AS, Hur GH, Burkart MD. An orthogonal purification strategy for isolating crosslinked domains of modular synthases. Bioorg Med Chem Lett. 2008; 18(10):3039–3042. [PubMed: 18249538]

He J, Sundararajan A, Devitt NP, Schilkey FD, Ramaraj T, Melancon CE 3rd. Complete genome sequence of *Streptomyces* venezuelae ATCC 15439, producer of the methymycin/pikromycin family of macrolide antibiotics, using PacBio technology. Genome Announc. 2016; 4(3)

Herbst DA, Jakob RP, Zahringer F, Maier T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. Nature. 2016; 531(7595):533–537. [PubMed: 26976449]

Herendeen SL, VanBogelen RA, Neidhardt FC. Levels of major proteins of Escherichia coli during growth at different temperatures. J Bacteriol. 1979; 139(1):185–194. [PubMed: 156716]

Hur GH, Meier JL, Baskin J, Codelli JA, Bertozzi CR, Marahiel MA, et al. Crosslinking studies of protein-protein interactions in nonribosomal peptide biosynthesis. Chem Biol. 2009; 16(4):372–381. [PubMed: 19345117]

Imker HJ, Krahn D, Clerc J, Kaiser M, Walsh CT. N-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. Chem Biol. 2010; 17(10):1077–1083. [PubMed: 21035730]

Iost I, Guillerez J, Dreyfus M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. J Bacteriol. 1992; 174(2):619–622. [PubMed: 1729251]

Ishikawa F, Haushalter RW, Lee DJ, Finzel K, Burkart MD. Sulfonyl 3-alkynyl pantetheinamides as mechanism-based cross-linkers of acyl carrier protein dehydratase. J Am Chem Soc. 2013; 135(24):8846–8849. [PubMed: 23718183]

Ito L, Shiraki K, Matsuura T, Okumura M, Hasegawa K, Baba S, et al. High-resolution X-ray analysis reveals binding of arginine to aromatic residues of lysozyme surface: implication of suppression of protein aggregation by arginine. Protein Eng Des Sel. 2011; 24(3):269–274. [PubMed: 21084280]

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008; 36(Web Server issue):W5–9. [PubMed: 18440982]

Jones PG, Cashel M, Glaser G, Neidhardt FC. Function of a relaxed-like state following temperature downshifts in *Escherichia coli*. J Bacteriol. 1992; 174(12):3903–3914. [PubMed: 1597413]

Kapur S, Worthington A, Tang Y, Cane DE, Burkart MD, Khosla C. Mechanism based protein crosslinking of domains from the 6-deoxyerythronolide B synthase. Bioorg Med Chem Lett. 2008; 18(10):3034–3038. [PubMed: 18243693]

Keating TA, Miller DA, Walsh CT. Expression, purification, and characterization of HMWP2, a 229 kDa, six domain protein subunit of Yersiniabactin synthetase. Biochemistry. 2000; 39(16):4729–4739. [PubMed: 10769129]

Keatinge-Clay AT. The structures of type I polyketide synthases. Nat Prod Rep. 2012; 29(10):1050–1073. [PubMed: 22858605]

Keatinge-Clay AT. The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines. Chem Rev. 2017; 117(8):5334–5366. [PubMed: 28394118]

Keatinge-Clay AT, Stroud RM. The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. Structure. 2006; 14(4):737–748. [PubMed: 16564177]

Key HM, Dydio P, Clark DS, Hartwig JF. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. Nature. 2016; 534(7608):534–537. [PubMed: 27296224]

Khare D, Hale WA, Tripathi A, Gu L, Sherman DH, Gerwick WH, et al. Structural basis for cyclopropanation by a unique enoyl-acyl carrier protein reductase. Structure. 2015; 23(12):2213–2223. [PubMed: 26526850]

Kim HJ, Ruszczycky MW, Choi SH, Liu YN, Liu HW. Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. Nature. 2011; 473(7345):109–112. [PubMed: 21544146]

Kittilä T, Cryle MJ. An enhanced chemo-enzymatic method for loading substrates onto carrier protein domains. Biochem Cell Biol. 2017

Klock HE, Koesema EJ, Knuth MW, Lesley SA. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins. 2008; 71(2):982–994. [PubMed: 18004753]

Kosa NM, Haushalter RW, Smith AR, Burkart MD. Reversible labeling of native and fusion-protein motifs. Nat Methods. 2012; 9(10):981–984. [PubMed: 22983458]

Kosa NM, Pham KM, Burkart MD. Chemoenzymatic exchange of phosphopantetheine on protein and peptide. Chem Sci. 2014; 5(3):1179–1186. [PubMed: 26998215]

Kozak M. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol Rev. 1983; 47(1):1–45. [PubMed: 6343825]

Kozbial PZ, Mushegian AR. Natural history of S-adenosylmethionine-binding proteins. BMC Structural Biology. 2005; 5(1):19. [PubMed: 16225687]

La Clair JJ, Foley TL, Schegg TR, Regan CM, Burkart MD. Manipulation of carrier proteins in antibiotic biosynthesis. Chem Biol. 2004; 11(2):195–201. [PubMed: 15123281]

Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, et al. A new enzyme superfamily - the phosphopantetheinyl transferases. Chem Biol. 1996; 3(11):923–936. [PubMed: 8939709]

Lautru S, Oves-Costales D, Pernodet JL, Challis GL. MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces* coelicolor M145. Microbiology. 2007; 153(Pt 5):1405–1412. [PubMed: 17464054]

Leao T, Castelao G, Korobeynikov A, Monroe EA, Podell S, Glukhov E, et al. Comparative genomics uncovers the prolific and distinctive metabolic potential of the cyanobacterial genus Moorea. Proc Natl Acad Sci U S A. 2017; 114(12):3198–3203. [PubMed: 28265051]

Lindwall G, Chau M, Gardner SR, Kohlstaedt LA. A sparse matrix approach to the solubilization of overexpressed proteins. Protein Eng. 2000; 13(1):67–71. [PubMed: 10679532]

Liu Y, Bruner SD. Rational manipulation of carrier-domain geometry in nonribosomal peptide synthetases. Chembiochem. 2007; 8(6):617–621. [PubMed: 17335097]

Liu Y, Zheng T, Bruner SD. Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. Chem Biol. 2011; 18(11):1482–1488. [PubMed: 22118682]

Lopanik NB, Shields JA, Buchholz TJ, Rath CM, Hothersall J, Haygood MG, et al. *In vivo* and *in vitro* trans-acylation by BryP, the putative bryostatin pathway acyltransferase derived from an uncultured marine symbiont. Chem Biol. 2008; 15(11):1175–1186. [PubMed: 19022178]

Losen M, Frolich B, Pohl M, Buchs J. Effect of oxygen limitation and medium composition on Escherichia coli fermentation in shake-flask cultures. Biotechnol Prog. 2004; 20(4):1062–1068. [PubMed: 15296430]

Lowell AN, DeMars MD 2nd, Slocum ST, Yu F, Anand K, Chemler JA, et al. Chemoenzymatic total synthesis and structural diversification of tylactone-based macrolide antibiotics through late-stage polyketide assembly, tailoring, and C-H functionalization. J Am Chem Soc. 2017; 139(23):7913–7920. [PubMed: 28525276]

Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, Butt TR. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J Struct Funct Genomics. 2004; 5(1-2):75–86. [PubMed: 15263846]

Maloney FP, Gerwick L, Gerwick WH, Sherman DH, Smith JL. Anatomy of the β-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate. Proc Natl Acad Sci U S A. 2016; 113(37):10316–10321. [PubMed: 27573844]

McCarthy JG, Eisman EB, Kulkarni S, Gerwick L, Gerwick WH, Wipf P, et al. Structural basis of functional group activation by sulfotransferases in complex metabolic pathways. ACS Chem Biol. 2012; 7(12):1994–2003. [PubMed: 22991895]

McMahon MD, Rush JS, Thomas MG. Analyses of MbtB, MbtE, and MbtF suggest revisions to the mycobactin biosynthesis pathway in *Mycobacterium tuberculosis*. J Bacteriol. 2012; 194(11): 2809–2818. [PubMed: 22447909]

Meier JL, Mercer AC, Rivera H Jr, Burkart MD. Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. J Am Chem Soc. 2006; 128(37):12174–12184. [PubMed: 16967968]

Miller BR, Drake EJ, Shi C, Aldrich CC, Gulick AM. Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. J Biol Chem. 2016; 291(43):22559–22571. [PubMed: 27597544]

Miller BR, Gulick AM. Structural Biology of Nonribosomal Peptide Synthetases. Methods Mol Biol. 2016; 1401:3–29. [PubMed: 26831698]

Mitchell CA, Shi C, Aldrich CC, Gulick AM. Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. Biochemistry. 2012; 51(15):3252–3263. [PubMed: 22452656]

Nakagawa S, Niimura Y, Miura K, Gojobori T. Dynamic evolution of translation initiation mechanisms in prokaryotes. Proc Natl Acad Sci U S A. 2010; 107(14):6382–6387. [PubMed: 20308567]

Nakamura H, Schultz EE, Balskus EP. A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. Nat Chem Biol. 2017; 13(8):916–921. [PubMed: 28671684]

Newmister SA, Li S, Garcia-Borràs M, Sanders JN, Yang S, Lowell AN, et al. Structural basis of the Cope rearrangement and cyclization in hapalindole biogenesis. Nat Chem Biol. 2018

Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an

allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. Appl Environ Microbiol. 1998; 64(5): 1694–1699. [PubMed: 9572938]

Novick RP. Plasmid incompatibility. Microbiol Rev. 1987; 51(4):381–395. [PubMed: 3325793]

Osterman IA, Evfratov SA, Sergiev PV, Dontsova OA. Comparison of mRNA features affecting translation initiation and reinitiation. Nucleic Acids Res. 2013; 41(1):474–486. [PubMed: 23093605]

Padan E, Zilberstein D, Schuldiner S. pH homeostasis in bacteria. Biochim Biophys Acta. 1981; 650(2-3):151–166. [PubMed: 6277371]

Payne JT, Millar JJ, Jackson CR, Ochs CA. Patterns of variation in diversity of the Mississippi river microbiome over 1,300 kilometers. PLoS One. 2017; 12(3):e0174890. [PubMed: 28350888]

Perlmutter DH. Chemical chaperones: a pharmacological strategy for disorders of protein folding and trafficking. Pediatr Res. 2002; 52(6):832–836. [PubMed: 12438657]

Pfeifer BA, Khosla C. Biosynthesis of polyketides in heterologous hosts. Microbiol Mol Biol Rev. 2001; 65(1):106–118. [PubMed: 11238987]

Prasad S, Khadatare PB, Roy I. Effect of chemical chaperones in improving the solubility of recombinant proteins in *Escherichia coli*. Appl Environ Microbiol. 2011; 77(13):4603–4609. [PubMed: 21551288]

Pryor KD, Leiting B. High-level expression of soluble protein in Escherichia coli using a His6-tag and maltose-binding-protein double-affinity fusion system. Protein Expr Purif. 1997; 10(3):309–319. [PubMed: 9268677]

Raran-Kurussi S, Cherry S, Zhang D, Waugh DS. Removal of affinity tags with TEV protease. Methods Mol Biol. 2017; 1586:221–230. [PubMed: 28470608]

Rariy RV, Klibanov AM. Correct protein folding in glycerol. Proc Natl Acad Sci U S A. 1997; 94(25): 13520–13523. [PubMed: 9391058]

Roberts GA, Staunton J, Leadlay PF. Heterologous expression in *Escherichia coli* of an intact multienzyme component of the erythromycin-producing polyketide synthase. Eur J Biochem. 1993; 214(1):305–311. [PubMed: 8508800]

Rohman M, Harrison-Lavoie KJ. Separation of copurifying GroEL from glutathione-S-transferase fusion proteins. Protein Expr Purif. 2000; 20(1):45–47. [PubMed: 11035949]

Saez NJ, Cristofori-Armstrong B, Anangi R, King GF. A Strategy for Production of Correctly Folded Disulfide-Rich Peptides in the Periplasm of *E. coli*. Methods Mol Biol. 2017; 1586:155–180. [PubMed: 28470604]

Sánchez C, Du L, Edwards DJ, Toney MD, Shen B. Cloning and characterization of a phosphopantetheinyl transferase from *Streptomyces* verticillus ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. Chem Biol. 2001; 8(7):725–738. [PubMed: 11451672]

Sazuka T, Ohara O. Sequence features surrounding the translation initiation sites assigned on the genome sequence of Synechocystis sp. strain PCC6803 by amino-terminal protein sequencing. DNA Res. 1996; 3(4):225–232. [PubMed: 8946162]

Schein CH, N MHM. Formation of soluble recombinant proteins in *Escheichia coli* is favored by lower growth temperatures. Nature Biotechnology. 1988; 6:291–294.

Sharp PM, Cowe E, Higgins DG, Shields DC, Wolfe KH, Wright F. Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. Nucleic Acids Res. 1988; 16(17):8207–8211. [PubMed: 3138659]

Sharp PM, Li WH. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic Acids Res. 1986; 14(19):7737–7749. [PubMed: 3534792]

Sieber SA, Walsh CT, Marahiel MA. Loading peptidyl-coenzyme A onto peptidyl carrier proteins: a novel approach in characterizing macrocyclization by thioesterase domains. J Am Chem Soc. 2003; 125(36):10862–10866. [PubMed: 12952465]

Skiba MA, Sikkema AP, Fiers WD, Gerwick WH, Sherman DH, Aldrich CC, et al. Domain organization and active site architecture of a polyketide synthase C-methyltransferase. ACS Chem Biol. 2016; 11(12):3319–3327. [PubMed: 27723289]

Skiba MA, Sikkema AP, Moss NA, Tran CL, Sturgis RM, Gerwick L, et al. A mononuclear iron-dependent methyltransferase catalyzes initial steps in assembly of the apratoxin A polyketide starter unit. ACS Chem Biol. 2017; 12(12):3039–3048. [PubMed: 29096064]

Smith AM, Brown WC, Harms E, Smith JL. Crystal structures capture three states in the catalytic cycle of a pyridoxal phosphate (PLP) synthase. J Biol Chem. 2015; 290(9):5226–5239. [PubMed: 25568319]

Smith DB, Johnson KS. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene. 1988; 67(1):31–40. [PubMed: 3047011]

Somu RV, Wilson DJ, Bennett EM, Boshoff HI, Celia L, Beck BJ, et al. Antitubercular nucleosides that inhibit siderophore biosynthesis: SAR of the glycosyl domain. J Med Chem. 2006; 49(26): 7623–7635. [PubMed: 17181146]

Stols L, Gu M, Dieckman L, Raffen R, Collart FR, Donnelly MI. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr Purif. 2002; 25(1):8–15. [PubMed: 12071693]

Storz, G., Hengee-Aronis, R. Bacterial Stress Responses. Washington D.C: ASM Press; 2000.

Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 2005; 41(1):207–234. [PubMed: 15915565]

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. 1990; 185:60–89. [PubMed: 2199796]

Su Y, Zou Z, Feng S, Zhou P, Cao L. The acidity of protein fusion partners predominantly determines the efficacy to improve the solubility of the target proteins expressed in *Escherichia coli*. J Biotechnol. 2007; 129(3):373–382. [PubMed: 17374413]
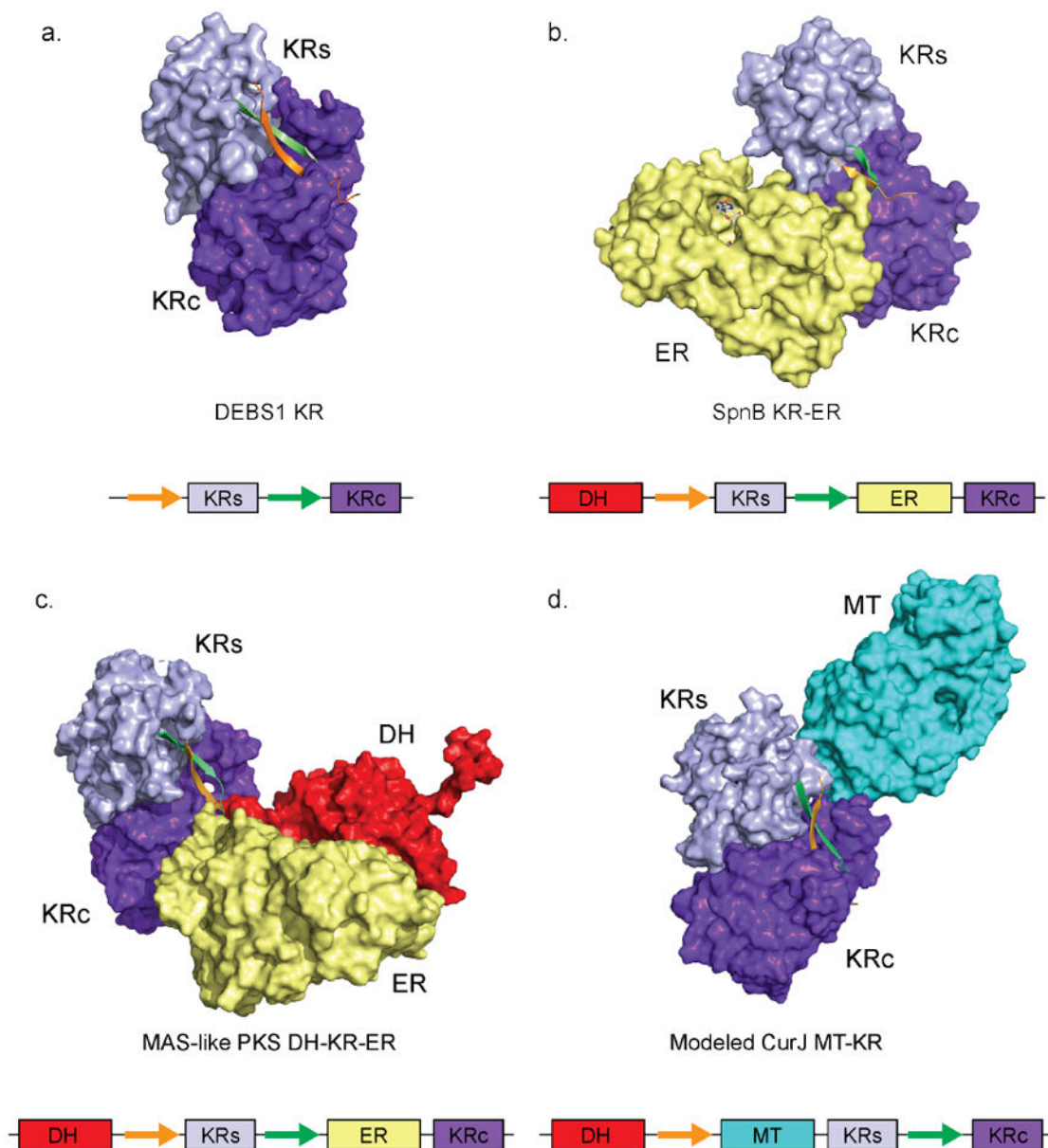
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

a.
KRs
KRc
DEBS1 KR
KRs
KRc
b.
KRs
ER
KRc
SpnB KR-ER
DH
KRs
ER
KRc
c.
KRs
DH
KRc
ER
MAS-like PKS DH-KR-ER
DH
KRs
ER
KRc
d.
MT
KRs
KRc
Modeled CurJ MT-KR
DH
MT
KRs
KRc

Figure 1.

Architectures of PKS modification domains: DH (red), MT (cyan), KR (light and dark purple KRs and KRc sub-domains), ER (yellow). A two-stranded β-ribbon (orange and green β-strands) is an integral part of the KR regardless of the domain context. The presence of other modifying domains such as ER or MT can complicate identification of the full KR (β-orange – KRs – β-green – KRc). a) DEBS1 KR (PDB 2FR0): A β-strand (orange) at the KRs N-terminus pairs with a β-strand (green) between the KRs and KRc sub-domains (Keatinge-Clay et al., 2006). b) SpnB KR-ER (PDB 3SLK): The ER domain is located after the green β-strand following KRs but prior to the KRc (Zheng, Gay, Demeler, White & Keatinge-Clay, 2012). c) MAS-like PKS DH-KR-ER (PDB 5BP4): The orange β-strand preceding KRs follows the DH domain (Herbst et al., 2016). d) Modeled CurJ MT-KR (MT

PDB 5THY, KR homology model): CMT domains are inserted between the orange β-strand and KRs in cis-AT PKS (Skiba et al., 2016).
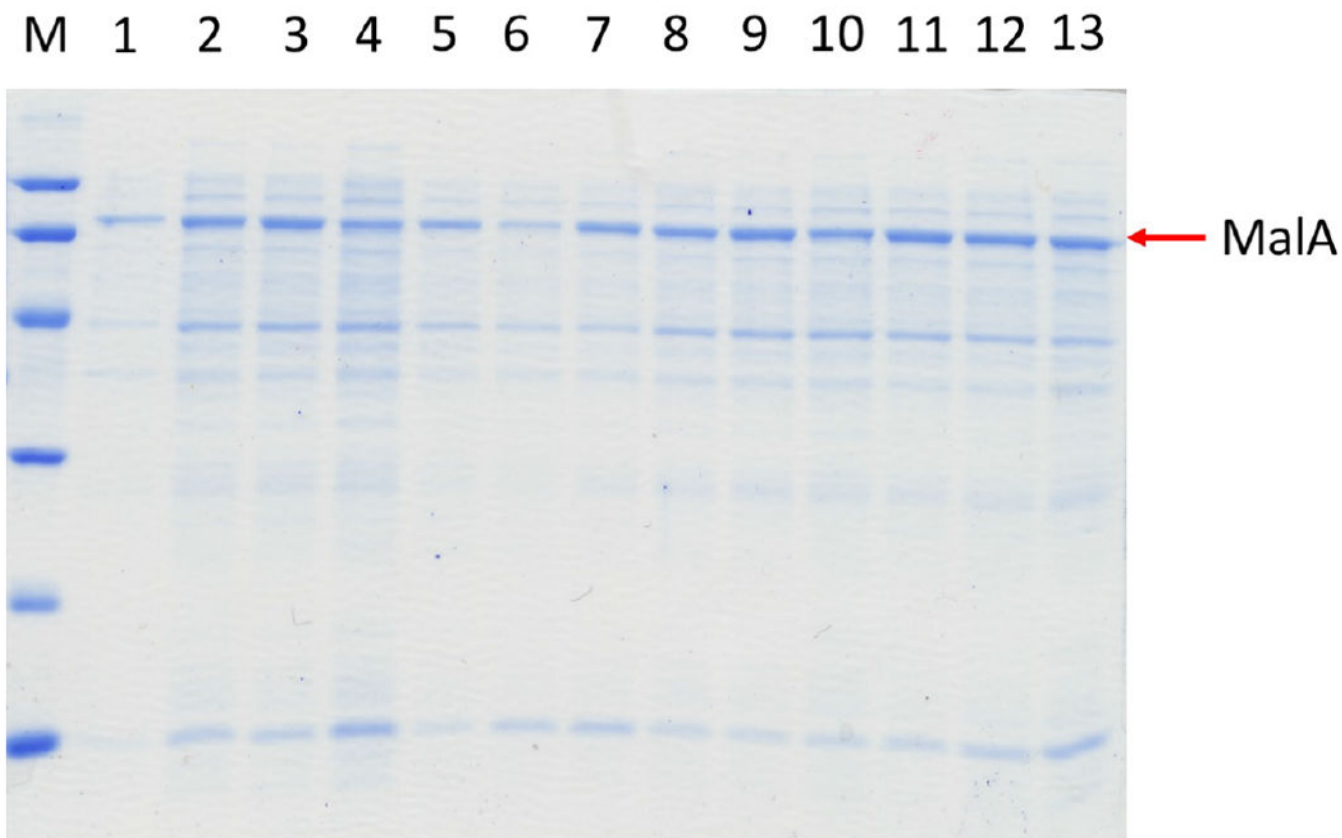
Figure 2.
Lysis buffer screen used in developing a purification protocol for the MalA halogenase (Fraley et al., 2017). Cells from equal-volume aliquots of an *E. coli* culture were lysed by sonication into 13 different lysis solutions (Table 6), and quickly purified by Ni affinity chromatography, and analyzed by SDS-PAGE. Lanes in the gel are molecular weight markers (M: 97.4, 66.2, 45.0, 31.0, 21.5, 14.4 kDa) and lysis solutions 1-13. MalA molecular weight is 73 kDa.

Table 1

Comparisons of Growth Parameters for Different Media

Medium	Time to exit from log phase (hr)	Culture density (OD_{600})	pH
LB	10	4.0-4.7	9.5
ZYM5052	22	12-16	6.7-7.3
TB	>25	18-20	7.3

Culture density and pH were measured when the cells began to emerge from log-phase growth. Data points represent ranges or averages of 96 cultures for each medium, grown in 96-well blocks.

Table 2

Recipe and Protocol for Phosphate-Buffered TB Medium

	Nutrient solution	Buffer solution
Components	12 g tryptone	23.1 g KH_2PO_4
	24 g yeast extract	125.4 g K_2HPO_4
	40 mL of glycerol	
Volume	900 mL in deionized H_2O	1 L in deionized H_2O
Treatment	Autoclave separately	Autoclave separately
TB (mix immediately before use)	900 mL	100 mL

Table 3

Optimal Inducer Concentrations for Different Promoters and Temperatures

Promoter	Temperature	Optimal inducer concentration
T7	20 °C	150 μM
	37 °C	50 μM
trc	20 °C	300 μM
	37 °C	125 μM

Table 4

Default Bacterial Production Conditions

Parameter	Condition
Cloning	pMCSG vectors (LIC variants of pET) T7 promoter with fusions (His-tag, MBP, Mocr)
Expression cell line	*E. coli* BL21AI/pRare2CDF
Medium	Terrific Broth w/ 4% glycerol
Aeration	Optimal surface-to-volume ratio and shaker speed for optimal aeration
Induction	200 μM IPTG, 0.2% arabinose (300 μM IPTG with trc promoter)
Temperature	Grow at 37°C, shift to 20°C gradually (& confirm), induce, continue growth at 20°C for ~18 hr, spin and freeze cell pellet at -80°C
Extras	Protein and/or chemical chaperones as needed

Table 5

1000x Trace Metals Mix (100 mL)[1]

Chemical[2]	Volume	MW	[Final]
0.1 M FeCl$_3$ ·6H$_2$O[3]	50 mL	270.3	50 μM iron
1 M CaCl$_2$	2 mL	111.0	20 μM calcium
1 M MnCl$_2$·4H$_2$O	1 mL	197.9	10 μM manganese
1 M ZnSO$_4$·7H$_2$O	1 mL	287.6	10 μM zinc
0.2 M CoCl$_2$·6H$_2$O	1 mL	237.9	2 μM cobalt
0.1 M CuCl$_2$·2H$_2$O	2 mL	170.5	2 μM copper
0.2 M NiCl$_2$·6H$_2$O	1 mL	237.7	2 μM nickel
0.1 M Na$_2$MoO$_4$·2H$_2$O	2 mL	242.0	2 μM molybdenum
0.1 M Na$_2$SeO$_3$·5H$_2$O[4]	2 mL	263.0	2 μM selenium
0.1 M H$_3$BO$_3$	2 mL	61.8	2 μM boron
H$_2$O	36 mL	18.0	-

[1] Add all solutions consecutively to 36 mL water, then filter sterilize; when preparing phosphate buffered growth media, add trace metals prior to buffer to minimize precipitation

[2] All stock solution of metals except FeCl$_3$·6H$_2$O are autoclaved, then stored at room temperature.

[3] Dissolved in 0.1 M HCl.

[4] A brief precipitation appears upon addition of Na$_2$SeO$_3$, which rapidly redissolves.

Table 6

Parameters of a Lysis Buffer Screen

Parameter	Condition
pH	7.4, 8.5
Ionic salt	none; 50 mM or 100 mM NaCl
Kosmotropic salt	50 mM $(NH_4)_2SO_4$; 50 mM Na_2SO_4
Organic co-solvent	none; 2.5% or 5% DMSO; 5% or 10% glycerol; 2.5% or 5% ethanol
Detergent	none; 0.5% or 1.0% CHAPS; 0.5% or 1.0% Triton; 1.0% octylglucoside
Other	none; 100 mM urea; 50 mM arginine; 50 mM glutamate

Table 7

Lysis Buffer Screen Formulation

Condition	Buffer	Additives
1	50 mM Tricine pH 8.5	500 mM NaCl, 10% glycerol
2	50 mM Tricine pH 8.5	50 mM $(NH_4)_2SO_4$, 10% glycerol
3	50 mM Tricine pH 8.5	50 mM NaCl, 5% DMSO
4	50 mM Tricine pH 8.5	50 mM NaCl, 5% ethanol
5	50 mM Tricine pH 8.5	50 mM NaCl, 1% Triton X-100
6	50 mM Tricine pH 8.5	50 mM NaCl, 100 mM urea
7	50 mM Tricine pH 8.5	50 mM NaCl, 50 mM Arg, 50 mM Glu
8	50 mM HEPES pH 7.0	50 mM $(NH_4)_2SO_4$, 10% glycerol
9	50 mM HEPES pH 7.0	50 mM NaCl, 5% DMSO
10	50 mM HEPES pH 7.0	50 mM NaCl, 5% ethanol
11	50 mM HEPES pH 7.0	50 mM NaCl, 1% Triton X-100
12	50 mM HEPES pH 7.0	50 mM NaCl, 100 mM urea
13	50 mM HEPES pH 7.0	50 mM NaCl, 50 mM Arg, 50 mM Glu
14	50 mM HEPES pH 7.0	500 mM NaCl, 10% glycerol