CHAPTER 11

Protein Engineering

SUMMARY

Enzymes, particularly those used in industrial applications, need to be more stable than their natural counterparts because they are often exposed to harsh conditions. There are multiple ways to increase enzyme stability and/or give an enzyme a novel property through protein engineering.

Disulfide bonds help maintain protein stability and 3D shape. Since disulfide bonds form between cysteine residues, introduction of more cysteine amino acids into the protein at specific locations would increase the number of disulfide bridges and thus the stability. Switching out glycine residues in the polypeptide with the more rigid proline residues also increases stability by decreasing the number of possible conformations that could exist. Also, swapping out residues that are prone to change due to pH or temperature could help stabilize the protein. However, great care must be taken not to affect the overall structure of the protein, which ultimately influences the function.

The active site of a protein can be changed by engineering in new amino acids to the region, which could increase or decrease the specificity for a particular substrate and could even change the recognized substrate altogether.

The majority of the amino acids within a protein simply provide support. Proteins could be engineered to contain only those amino acids that are important in the active site. The difficult part is engineering the protein in such a way that the active site amino acids are located in the correct positions to bind the substrate.

Enzymes with novel properties can be identified by constructing libraries for different enzymes that have been mutagenized at the genetic level, called directed evolution. Mutations are introduced randomly by error-prone PCR or by recombining different domains through DNA shuffling and combinatorial protein libraries. One interesting domain includes the zinc finger, which is a DNA-binding motif. Theoretically, any protein could be engineered with this motif to bind DNA. DNA shuffling recombines randomly cut pieces from different copies of the same gene or homologous genes from different organisms followed by mutagenesis of the DNA. The products are then screened for improved properties, such as higher activity level. Combinatorial protein libraries attempt to generate entirely new proteins that have novel or improved properties by using motifs with known properties and combining them at the genetic level to produce novel proteins. Motifs that are particularly interesting include those that bind metals or DNA.

Random polypeptide sequences are combined with known structural motifs, such as alpha helices and beta sheets, to produce *de novo* proteins. The novel proteins are produced by first generating random DNA sequences that are expressed in a host cell. Functional proteins are screened for some activity. Generation of *de novo* proteins can be used to help predict 3D structures of polypeptide sequences. Additionally, these novel proteins could be used in directed evolution experiments to select for *de novo* proteins with increased activity.

Non-natural amino acids are useful in protein engineering because they have different functional groups than those found on genetically encoded peptides. Some amino acids are categorized as natural but not genetically encoded. Rather, these types of amino acids are the result of post-translational modifications. An example includes the modification of serine to phosphoserine. Other non-natural amino acids are used to attach other reactive chemical groups, such as for cross-linking or photoactivation. And finally, others are used as probes for fluorescence, infrared, or NMR spectroscopy.

Through the use of *in vitro* protein translation systems, non-natural amino acids can be incorporated into polypeptide chains. However, introduction of non-natural amino acids *in vivo* requires a mutant aminoacyl-tRNA synthetase and a mutant tRNA. The mutant tRNA recognizes an amber stop codon but inserts the non-natural amino acids instead.

Proteins can also be engineered through the addition of non-natural amino acids containing new functional groups, such as the UV-activated *p*-benzoyl-L-phenylalanine (*p*Bpa). Changing the amino acid to a non-natural for any position on the protein, particularly for *p*Bpa, allows proteins to be covalently cross-linked upon activation by UV light. This modification *in vivo* requires a mutant aminoacyl-tRNA synthetase that can charge a tRNA with a non-natural amino acid.

Protein engineering has a wide range of applications. Manufacture of biomaterials, such as skin, vascular, or cartilage used in medicine, is often performed on protein scaffolds. These protein scaffolds could be engineered to improve support and structure, which could ultimately improve the material itself. Changing the amino acid residues could change the specific properties of the materials by enabling cross-linking. Additionally, the materials could act as an attractant to cells for regeneration of injured cells and tissue.

One last application of protein engineering is the creation of binding proteins. These proteins have the potential to bind a drug and deliver it to a specific location within a cell. This would increase the specificity of the drug and decrease the number of unwanted side effects.

Case Study Bringing Protein Engineering and Natural Product Biosynthesis Together

Ralph A. Cacho, Yi Tang (2013). *Chemistry & Biology* 20, 3–5.

NRPSs are nonribosomal peptide synthetases that synthesize some bioactive molecules, including some antibiotics like penicillin. Each NRPS has at least three domains: adenylation (A), thiolation (T), and condensation (C) domains. The adenylation domain activates an amino acid prior to its transfer to the thiolation domain. This process depends on ATP. The condensation domain catalyzes the peptide bond formation. NRPSs are relatively simple synthetases and attractive targets for protein engineering.

In this preview, the authors discuss the protein engineering of NRPSs and a recently developed yeast cell surface display method for high-throughput screening.

Why is the A domain of particular interest for protein engineering?

The A domain is involved in the selection of the individual amino acids during the synthesis of bioactive molecules. Engineering the A domain of NRPSs will change the specificity of the synthetases and allow unnatural amino acids to be incorporated into the polypeptide chains.

How were researchers able to determine the residues within the A domain that control substrate specificity?

A high-resolution crystal structure was obtained for the phenylalanine-specific A domain (PheA) from gramicidin synthetase 1. The crystal structure indicated that the A domain is highly conserved with firefly luciferase. Additionally, the active site was sandwiched between a large N-terminal subdomain and a smaller C-terminal subdomain. Sequence alignment of A domain databases revealed a 10 amino acid (10-aa) sequence as a predictor of A domains.

How successful have researchers been with regards to the engineering of A domains?

Based on the 10-aa sequence identified from alignment studies, PheA adenylation domain's specificity was engineered to accept L-leucine instead of the natural substrate L-phenylalanine. Surfactin synthetases specificity for L-aspartate was also changed to L-asparagine. Additionally, calcium-dependent antibiotic analogs have been produced by altering the specificity of A domains from the corresponding NRPSs.

In this preview, how has protein engineering led to the expansion of the genetic code?

Successful engineering of aminoacyl-tRNA synthetases to accept unnatural amino acids has expanded the 20 natural amino acids in the genetic code.

What method was developed for high-throughput screening for A domain engineering? How successful has this method been?

Directed evolution followed by an effective screening process, such as a yeast cell surface display strategy, was employed to evolve the A domains of bacillibactin NRPS and other polypeptides. The A domain of bacillibactin NRPS (DhbE) was successfully altered to accept unnatural aromatic amino acids. In this method, DhbE was genetically fused to Aga2p for transport and adhesion on the yeast cell surface. Other moieties were added to allow the cells to be sorted by fluorescence-assisted cell sorting (FACS). Four residues from DhbE were mutagenized. Two of these residues lie within the 10-aa code, and the other two were within close proximity to the binding site. Using FACS for screening the mutants, the researchers were able to isolate DhbE mutants that selectively bound 3-HBA (3-hydroxybenzoate), an unnatural substrate, with an 11-fold higher specificity than wild-type DhbE. Additionally, this mutant had a 3-fold lower affinity toward DHB (2,3-dihydroxybenzoic acid), the natural substrate. Another mutant DhbE switched specificity from DHB to another unnatural substrate 2-ABA (anthranilate) by more than 206-fold. Based on these results, the yeast cell surface display method coupled with FACS has proven effective in screening mutant NRPS A domains.

The A domain represents the specificity domain for these synthetases. What is the next step in the production of NRPSs that can fully function to produce engineered proteins?

The yeast cell surface display protocol combined with FACS sorting is a powerful tool for screening the NRPS mutants and can be extended to other A domains and substrates. The next step would be to evolve the remaining domains in NRPSs so that unnatural amino acids could be incorporated fully into growing polypeptide chains.

The yeast cell surface display strategy combined with FACS sorting has provided a mechanism to effectively identify synthetases that recognize and can incorporate a variety of unnatural amino acids into polypeptide chains. This methodology allows researchers to efficiently screen for specific phenotypes and greatly advances the field of protein engineering.

Bringing Protein Engineering and Natural Product Biosynthesis Together

Ralph A. Cacho[1] and Yi Tang[1,2,*]
[1]Department of Chemical and Biomolecular Engineering
[2]Department of Chemistry and Biochemistry
University of California, Los Angeles, Los Angeles, CA 90095, USA
*Correspondence: yitang@ucla.edu
http://dx.doi.org/10.1016/j.chembiol.2013.01.005

In this issue of *Chemistry & Biology*, Zhang and colleagues developed a yeast cell surface display strategy to effectively evolve the substrate specificity of DhbE, one of the adenylation domains of the bacillibactin synthetase complex. The method yields DbhE variants that have dramatically altered substrate specifities toward unnatural aryl substrates.

Nonribosomal peptide synthetases (NRPSs) are large multimodular enzymes that are responsible for the biosynthesis of important bioactive molecules such as penicillin, cyclosporine, and daptomycin. NRPSs catalyze the successive formation of peptide bonds using aminoacyl adenylates. Each module of the NRPS has minimally three domains: the adenylation domain (A), which specifically activates an amino acid and transfers it to the thiolation domain (T) in an ATP-dependent fashion; and a condensation domain (C) that catalyzes the peptide bond formation between the T domain-tethered aminoacyl group and the growing peptide chain. The linear juxtaposition of several modules constitutes an assembly-line-like machinery in which the growing peptide increases in size by one residue after each module.

This elegantly simple biochemical logic of NRPSs has made these megasynthetases attractive targets for protein engineering toward production of hybrid products. The A domain, in particular, has been focused on because of its role in selection of the individual building blocks. Classical domain swapping approaches have led to fruitful combinatorial biosynthesis of analogs as demonstrated in engineering of the daptomycin synthetase (Baltz et al., 2005). Directed evolution, which is an extremely powerful approach in the engineering of desired traits of target enzymes, has not been widely used on natural product biosynthetic enzymes, largely due to the lack of effective screening strategies. In this issue of *Chemistry & Biology*, Zhang et al. (2013) developed a yeast cell surface display strategy to effectively evolve the substrate specificity of one of the A domains in the bacillibactin NRPS. They were able to dramatically alter the A domain specificity toward unnatural aromatic building blocks. This approach is generally applicable to different A domain/substrate combinations, and may significantly increase the repertoire of A domains that can be used in the construction of synthetic NRPSs.

An enabling breakthrough for understanding and engineering the specificity of A domains is the first high-resolution crystal structure of a phenylalanine-specific A domain (PheA) from the gramicidin synthetase 1 (Conti et al., 1997). The crystal structure revealed that PheA shares a highly conserved domain arrangement with firefly luciferase, which is also an adenylating enzyme. PheA contains a large N-terminal subdomain and a smaller C-terminal subdomain, with the active site nestled at the interfaces of these two subdomains. Based on the PheA structural information, several groups identified the residues within the A domain binding pocket that control substrate specificity (Challis et al., 2000; Stachelhaus et al., 1999). Structure-guided sequence alignment of a large number of A domain sequences in the database led to the assignment of a ten amino acid (10-aa) code that can be used to predict A domain substrate specificity. Moreover, the 10-aa code has served as a guide in engineering substrate specificity of adenylation domains through rational mutagenesis, as first demonstrated in altering the amino acid specificity of PheA from L-phenylalanine to L-leucine (Stachelhaus et al., 1999). Subsequent examples included switching the specificity of an A domain in the surfactin synthetase from L-aspartate to L-asparagine (Eppelmann et al., 2002) and recently the production of calcium-dependent antibiotic (CDA) analogs through mutagenesis of an A domain in the CDA NRPS (Thirlway et al., 2012).

Notwithstanding these successes in modification of A domain specificity, it is evident that this domain represents an ideal target for engineering by directed evolution approaches, such as large scale saturation mutagenesis at multiple positions within the 10-aa code. Access to significantly larger sequence space, coupled with a high-throughput screening strategy, should lead to variants that have dramatically altered substrate specificity. This is evidenced in the highly successful engineering of the aminoacyl-tRNA synthetases that control fidelity of the translation system, which has led to the expansion of the 20 natural amino acid genetic code (Wang et al., 2001). The first demonstration of directed evolution of A domain was that of the A-T didomain AdmK from the andrimid pathway. Through saturation mutagenesis of three residues within the 10-aa code, followed by a product-based screening of 14,330 clones using liquid chromatography-tandem mass spectrometry, the substrate tolerance of the A domain was expanded from L-valine to L-isoleucine, L-leucine, L-alanine, or L-phenylalanine (Evans et al., 2011). Whereas this is a significant breakthrough in NRPS engineering, the use of MS-based screening is clearly the rate-limiting

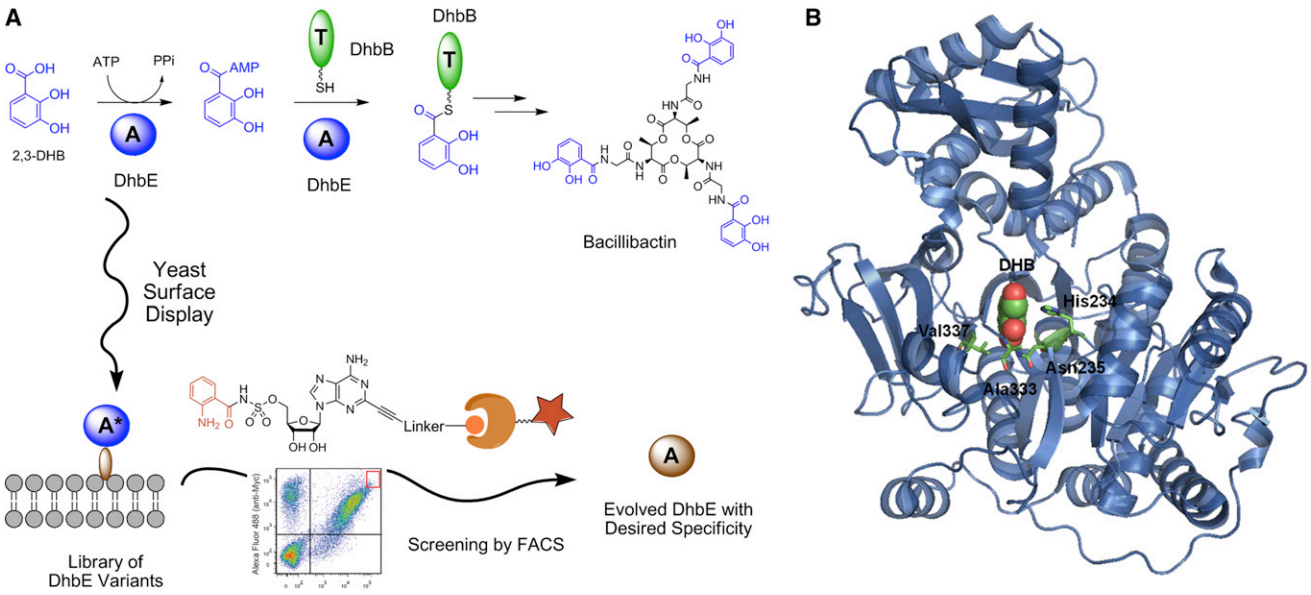

Figure 1. Scheme of the Directed Evolution Study of DhbE

(A) DhbE is a stand-alone adenylation domain that activates 2,3-dihydroxybenzoic acid in the biosynthesis of the siderophore bacillibactin. Zhang et al. (2013) utilized a chemical probe that mimics the acyl-adenylate intermediate of DhbE and screened using fluorescence assisted cell sorting (FACS) to isolate desired mutants.

(B) Crystal structure of DhbE (accession number 1MDB) showing the sites targeted for saturation mutagenesis by Zhang et al. (2013) for the directed evolution study.

factor in exploring additional sequence space.

Toward developing a high throughput screening method for engineering the A domain, Zhang et al. (2013) employed a yeast-surface displayed method that has been successfully used in the evolution of a wide range of polypeptides, such as antibody fragments (Boder and Wittrup, 1997). Their target was the stand-alone DhbE A domain, which is part of the NRPS machinery for biosynthesis of the siderophore bacillibactin in *Bacillus subtilis*. The goal is to expand the substrate specificity of DhbE from the natural 2,3-dihydroxybenzoic acid (DHB). To display DhbE and generated mutants, wild-type DhbE was first genetically fused to Aga2p, which allows the transport and adhesion of the fusion protein to the yeast cell surface. Furthermore, DhbE was fused to an N terminus HA tag and a C terminus myc tag, which can be recognized by anti-HA and anti-myc antibodies, respectively. This enabled the use of fluorescence assisted cell sorting (FACS) with a flow cytometer to capture the yeast cells displaying full-length DhbE.

In order to capture DhbE fusion proteins that can bind to the desired substrate or acyl-adenylate, probes consisting of an acyl-adenosyl-monosulfamate (acyl-AMS) covalently linked to biotin were synthesized. The acyl-AMS moiety is a nonhydrolyzable mimic of the acyl-AMP intermediate that acts as a "bait" to bind DhbE, while the biotin moiety allows conjugation to a fluorophore tethered to streptavidin. Therefore, in this design, cells displaying DhbE, as shown by binding of fluorescent anti-myc antibody, and those that bind to the acyl-AMS probe, as shown by conjugation of the fluorescent streptavidin probe, can be captured through FACS (Figure 1A). To prove that the acyl-AMS based probe can bind to DhbE, a salicylyl-AMS probe, in lieu of the redox sensitive DHB, was synthesized and was shown to select for the cells that display DhbE. Two other chemical probes that contain 3-hydroxybenzoate (3-HBA) and anthranilate (2-ABA) were also synthesized and were subsequently used for the FACS-based selection of mutants. The chemical probe design also allows one to screen a variety of unnatural substrates by simply changing the acyl-AMS portion of the probe to the targeted acyl group.

Based on the X-ray crystal structure of DhbE (May et al., 2002) and the 10-aa code, four residues (His234, Asn235, Ala333, and Val337) were chosen for saturation mutagenesis (Figure 1B). Asn235 and Val337 were part of the 10-aa code, while His234 and Ala333 were shown to be in proximity of the binding pocket. The synthesized acyl-AMS of 3-HBA and 2-ABA were then incubated with DhbE mutants displayed on yeast surface. Thereafter, streptavidin bound to phycoerythrin was incubated with the yeast cells to capture the DhbE mutants that bind the acyl-AMS probes. Using FACS to screen through 5×10^6 colonies, the authors were able to isolate a DhbE mutant (KZ4) that has an 11-fold higher k_{cat}/K_M compared to wild-type DhbE toward 3-HBA and 3-fold lower k_{cat}/K_M compared to wild-type DhbE toward DHB, thus switching its selectivity 30-fold. On the other hand, the KZ12 mutant has a 6.2-fold higher k_{cat}/K_M compared to wild-type DhbE toward 2-ABA, and an 11-fold lower k_{cat}/K_M compared to wild-type DhbE toward DHB, which accounts for a dramatic 206-fold switch in selectivity toward 2-ABA. Thus, Zhang et al. (2013) were able to prove that yeast cell surface display coupled with FACS is an effective high-throughput screening method for the directed evolution of NRPS A domains. This method can be extended to any

pairing of A domain and substrate through the preparation of the corresponding acyl-AMS probe. The next challenge would be to further evolve the rest of the NRPS assembly line to process the unnatural building block efficiently.

The marriage of directed evolution and natural product biosynthesis is a tantalizing union toward production of "unnatural" natural products. The difficulties associated with high throughput screening of desired phenotypes have severely limited progresses in this regard. Zhang et al. (2013) demonstrated that powerful techniques, such as yeast surface display, which are widely used in the protein engineering field, can be effectively designed to allow rapid screening of enzymes in secondary metabolism. This important work has only scratched the surface. It is anticipated that additional ingenious approaches can significantly improve our abilities to tailor the properties of these enzymes toward customized synthesis of target small molecules.

REFERENCES

Baltz, R.H., Miao, V., and Wrigley, S.K. (2005). Nat. Prod. Rep. *22*, 717–741.

Boder, E.T., and Wittrup, K.D. (1997). Nat. Biotechnol. *15*, 553–557.

Challis, G.L., Ravel, J., and Townsend, C.A. (2000). Chem. Biol. *7*, 211–224.

Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997). EMBO J. *16*, 4174–4183.

Eppelmann, K., Stachelhaus, T., and Marahiel, M.A. (2002). Biochemistry *41*, 9718–9726.

Evans, B.S., Chen, Y., Metcalf, W.W., Zhao, H., and Kelleher, N.L. (2011). Chem. Biol. *18*, 601–607.

May, J.J., Kessler, N., Marahiel, M.A., and Stubbs, M.T. (2002). Proc. Natl. Acad. Sci. USA *99*, 12120–12125.

Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). Chem. Biol. *6*, 493–505.

Thirlway, J., Lewis, R., Nunns, L., Al Nakeeb, M., Styles, M., Struck, A.-W., Smith, C.P., and Micklefield, J. (2012). Angew. Chem. Int. Ed. Engl. *51*, 7181–7184.

Wang, L., Brock, A., Herberich, B., and Schultz, P.G. (2001). Science *292*, 498–500.

Zhang, K., Nelson, K.M., Bhuripanyo, K., Grimes, K.D., Zhao, B., Aldrich, C.C., and Yin, J. (2013). Chem. Biol. *20*, this issue, 92–101.

Small Molecule Disruption of *B. subtilis* Biofilms by Targeting the Amyloid Matrix

Emma K. Andersson[1] and Matthew Chapman[2,*]
[1]Department of Chemistry, Umeå University, 90187 Umeå, Sweden
[2]Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI 48109-1048, USA
*Correspondence: chapmanm@umich.edu
http://dx.doi.org/10.1016/j.chembiol.2013.01.004

Small molecule inhibitors of amyloid aggregation have potential as treatment for a variety of conditions. In this issue of *Chemistry & Biology*, Romero and colleagues use amyloid-dependent *B. subtilis* biofilm formation to screen for amyloid inhibitors, identifying compounds that not only inhibit *B. subtilis* biofilm formation but also ones that disrupt preformed biofilms.

Amyloid fibers are historically associated with diseases such as Alzheimer's, Huntington, and the prion diseases. Many proteins can adopt the amyloid conformation—essentially a β-rich repeating structure where the β strands orient perpendicular to the fiber axis. Ordered amyloid protein polymers upset cellular proteostasis because they can inappropriately interact with membranes or sequester chaperone machinery. Therefore, the search is on for ways to ameliorate amyloid-related diseases by targeting amyloid formation. A major hurdle in this endeavor is that disease-associated amyloid formation is sporadic and difficult to faithfully reproduce in model organisms.

However, a growing number of "functional" amyloids have been identified that are assembled by dedicated biogenesis systems. Functional amyloids and their assembly systems have been found in nearly all walks of cellular life, including mammals, fungi, and bacteria (Hammer et al., 2008, Fowler et al., 2007). Functional amyloids contribute to cellular biology in various ways, including regulation of melanin synthesis, information transfer, or as a structural component. Furthermore, some of these functional amyloid systems provide a unique platform for understanding how amyloid formation can be directed and controlled so that cellular toxicity is minimized.

Amyloids are commonly found as the major protein component of the extracellular matrix in bacterial biofilms. Bacteria within the biofilm are protected from environmental insults, including disinfectants and antibiotics, making biofilms a major concern in hospital and industrial settings. Therefore, factors that can disrupt bacterial amyloid formation would be potential lead compounds for targeting bacterial biofilms or amyloid formation in general. The study presented in this issue of *Chemistry & Biology* by Romero et al. (2013) describes *Bacillus subtilis* pellicle biofilm as a model system to screen for amyloid inhibitors. The extracellular