Duke DeLoache

Independent Study

Drs. Campbell and Heyer

5/12/2012

Davidson College Synthetic Biology

Summer 2012 Research Proposal

**Introduction**

Violacein is a purple bacteria-produced pigment produced from L-tryptophan via a pathway involving five enzymes. This pigment has anti-tumor, antiviral, and antifungal properties, which make it a valuable object of study1. Additionally, research has shown that this production pathway has interesting recombinant expression properties. The five enzymes in the pathway are vioABCDE, and the biosynthetic pathway is VioA 🡪 VioB 🡪 VioE 🡪 VioD 🡪 VioC (see Figure 1). I will denote the operon as a string of characters, like ABCDE

The genes encoding the pathway have been expressed in *E. coli* by several research teams, including two iGEM teams: Cambridge (2009) and Slovenia (2010). These teams have experimented with various recombinant forms and discovered some interesting expression patterns. In 2009, the Cambridge iGEM team used biobrick assembly to remove various parts of the pathway and, in doing so, showed varying pigmentation expression. The operon ABCDE produced a purple pigment as expected, but ABDE produced a dark green pigment (deoxyviolacein), and ABCE produced a light green pigment (deoxychromoviridans)2.

The Cambridge team used the biobrick approach to remove individual genes, but did not vary the order in which they assembled the genes along the plasmid: ABCDE. Jiang *et al.* tested nine different recombinant vectors with the vio gene cluster, showing that various strains yield significant differences in crude violacein production. Additionally, they tested the various recombinant vectors across different hosts, showing that differing vectors showed different expression across various strains of *E. coli* (as well as some other bacteria). In some strains, deoxyviolacein was produced while crude violacein accumulated in others1.

While the few research efforts I have mentioned are well documented, it seems that there is a great deal that is unknown about the pathway characteristics of violacein, specifically with repect to the effect of various recombinant vectors. I propose that the Davidson College Synthetic Biology Team look into experimentally manipulating the vioABCDE gene cluster by producing a library of genetic circuits using the power of Golden Gate Assembly, the C-dog construct, CRIM plasmids, and barcode-like primer sequences in order to vary the strength, presence, and order of the violacein-producing genes and their constituent components in such a way that many strains can be produced, tested, and kept track of in an efficient manner (see Figure 2). One goal would be to optimize control of violacein expression, and another would be to discover more about the pathway characteristics, such as instances in which deoxyviolacein or deoxychromoviridans tend to accumulate.

**Procedures**

While the Cambridge iGEM team took their genes from *Chromobacterium voilaceum*, I think that the vioABCDE genes should be cloned from *Duganella* sp. B2, as Jiang *et al.* describe this source as more reliable in terms of the final product: expression of violacein in *E. coli* 1. I recommend that the team use the C. dog construct to both control and vary gene expression levels (see Figure 3). One advantage of using C. dog would be that for the various organizations of cassettes along the genetic circuit, the expression levels of the five genes of interest (GOIs) would be more consistent5. Additionally, by varying the RBS2, these expression levels could be manipulated in a more controlled and predictable manner than could a circuit with traditional ribosomal binding sites. Upstream of C. dog on each cassette will be a promoter, and downstream of C. dog on each cassette will be one of the GOIs followed by a degradation tag.

There will need to be a way to insert and replace these parts as well as various promoters and degradation tags. To solve this problem, I propose the team use a double cloning approach with type II restriction enzymes. Each component (promoter, C. dog, etc.) will be assigned one pair of type II restriction enzymes. A library of components will be produced, wherein two copies of each component will be built. One copy of each component will be flanked by binding sites for one of the type II restriction enzymes, and the other copy with be flanked by binding sites for the other type II restriction enzyme. Each enzyme will cut the plasmid in two locations, slicing out the designated component and leaving the same length overhang as its paired enzyme partner.

 If we want to switch out a promoter, we simply have to add the appropriate enzyme to cut out the promoter that is already present before ligating the new promoter with the other restriction enzyme binding site via Golden Gate Assembly (see Figure 4). Similarly, the other enzyme could be added remove this promoter. This process could theoretically be repeated back and forth, swapping components in and out as many times as one desired, as type II restriction enzymes leave no scar sites. It is important that each pair of enzymes not only have the same overhang length but also operate reliably at the same temperature and with a similar buffer solution. The following pairs of restriction enzymes were found using New England BioLabs’ enzyme finder and obey the previously mentioned constraints:

BbsI, BsaI

FokI, SfaNI

BsmAI, BsmBI

BsmFI, BtgZI

BseYI, BssSI

EarI, SapI

While these are not the only pairs of enzymes in the database that have the same length overhang, buffer type and temperature, they all leave at least a three base pair overhang post-cut, which may allow for more reliable rebinding6.

Once this library of components is built (promoters, C. dog constructs, vioABCDE, and degradation tags all flanked by the appropriate restriction enzymes) a genetic circuit can be built and put on a plasmid that can be transformed into *E.* *coli* cells. However, so that the team may vary the order of the cassettes along the cascade, it should be built in such a way that 8-cutter restriction enzyme binding sites exist between each cassette and at each end of the entire circuit (see Figure 5). This will allow entire cassettes to be reshuffled and also for cassettes to be removed for manipulation, say, for the swapping in/out of a promoter. It would also simply allow for a cassette to be removed from the plasmid entirely, as done via biobrick assembly by the Cambridge iGEM team.

One important issue to address is how the team will keep track of the order in which the cassettes are shuffled. To manage this issue, I propose the team assign each specific arrangement of components (each of the potential combinations of promoter/C. dog/GOI/degradation tag to be used) a specific primer sequence, referred to as a “barcode.” Additionally, each genetic circuit will be flanked by general primer sequences that will remain unchanged throughout the research process. These two barcodes will be oriented opposite of one another, the idea being that PCR could be used to screen for length in both the forward and reverse orientation (and for each of the barcodes present) in order to determine the order in which the cassettes are arranged for a particular population of cells. The general barcode further upstream will be oriented in such a way that if the team were to use the hin/hix system to flip any of the cassettes in order to test their efficacy in the reverse orientation, these could be primed for and screened for length in the PCR process. Using the barcodes to keep track of order could allow the team to shuffle the cassettes in such a way that many different combinations could be produced simultaneous. Then cells could be selected for how much they express various attributes, like the intensity of purple or green pigmentation.

Determining the sequences for the barcodes themselves, however, is a challenge. Each individual cassette’s barcode, ideally, would be as different as possible from all of the other barcodes, subject to the constraint that they all have the same (or very similar) melting temperatures. I see here an opportunity for the math/computer science portion of the synthetic biology team to step in and make a useful webtool. The optimization problem would likely be solved with a linear program and would be formulated as follows: given length of primers, an ideal melting temperature, and a total number of primers to generate as inputs, the objective would be to generate a set of primer sequences, each of the user-defined length, subject to the constraint that each have the same G/C content. This could evolve into an online database specific to individual users wherein cassettes are recorded and associated with specific barcodes such that none would be reused.

Another important consideration is the manner in which the plasmids containing the genetic circuits will be transformed into *E. coli*. I recommend CRIM plasmids be used, as these will maximize the modularity of the system. Their integration allows for direct transformation of plasmids with Integrase. Furthermore, removal is site-specific and keeps the integration site (attB) intact such that integration may be repeated, similar to the double cloning process with type II restriction enzymes. Also, removed plasmids are identical to those inserted, so that the exact organization of the cassettes selected for can be verified and reproduced or varied as desired7.



**Figure 1** Biosynthetic pathway of L-tryptophan into violacein, deoxychromoviridans, and deoxyviolacein3.



**Figure 2** Example genetic circuit composed of five cassettes, each containing a promoter, RBS, vioABCDE gene, and degradation tag.



 **Figure 3** C. dog construct4.



**Figure 4** Double cloning with type II restriction enzymes. Example swap-out with Golden Gate Assembly.



**Figure 5** Genetic circuit with 8-cutter restriction enzyme binding sites (purple circles).



**Figure 6** Final example genetic circuit with barcodes (thin, straight arrows).

**Sources**

1. Jiang P, Wang H, Chong Z, Lou K, Xing X (2009) Reconstruction of the violacein

biosenthetic pathway from *Duganella* sp. B2 in different heterologous hosts. Appl

Microbiol Biotechnol 86:1077-1088.

1. Cambridge iGEM 2009. [http://2009.igem.org/Team:Cambridge/Project/VI01](http://2009.igem.org/Team%3ACambridge/Project/VI01).

Accessed May 11, 2012.

1. Registry of Standard Biological Parts.

[http://partsregistry.org/Image:SloViolacein-pot.jpg](http://partsregistry.org/Image%3ASloViolacein-pot.jpg). Accessed May 11, 2012.

1. Mutalik, Vivek *et al.* BIOFAB, Emryville.

[http://gcat.davidson.edu/mediawik i1.15.0/images/5/5b/C\_dog\_concept.pdf](http://gcat.davidson.edu/mediawik%20%20%20i1.15.0/images/5/5b/C_dog_concept.pdf). Accessed May 11, 2012.

1. Endy, D. *C. dog*. Bicistronic Translational Junction. Stanford.
2. New England BioLabs. <http://www.neb.com/nebecomm/default.asp>. Accessed May

11, 2012.

1. Haldimann A, Wanner BL (2001) Conditional-Replication, Integration, Excision, and

Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria.

Journal of Bacteriology 182(21):6284-6393.