Ellen Johnson

Research Proposal

Optimization of a Biological Pathway in *E. coli*

**Purpose**

 The purpose of the research will be to optimize a biological pathway in *E. coli* towards production of a desired product. Pathway optimization may be defined as either the time it takes to produce the desired output or maximum production of desired output. Pathway optimization will occur through manipulation of promoters, RBS sequences, protein degradation tags, and order of steps.

**Concept**

**Pathway 1: Sucrose degradation**

Preexisting sucrose degradation pathways in *E. coli*

 Sucrose degradation is one potential biological pathway for optimization research (see figure 1). *E. coli* naturally degrades sucrose via two pathways. In one pathway enzyme 2.4.1.7 degrades sucrose to D-fructose (see figure 2) 1. In the other pathway enzyme 3.2.1.20 degrades sucrose to β-D-fructose, which is converted to β-D-fructose by enzyme 2.7.1.4(see figure 3) 1. β-D-fructose connects to longer pathways within the starch and sucrose metabolic pathway.

Figure 1. Sucrose degradation pathways as part of the starch and sucrose metabolic pathway. Larger image can be viewed at http://www.kegg.jp/kegg-bin/show\_pathway?org\_name=eco&mapno =00500&mapscale=&show\_description=hide.

Sucrose degradation pathway for optimization

 Sucrose degradation by enzyme 3.2.1.20 may be the preferred for optimization compared to the pathway with enzyme 2.4.1.7. The pathway with enzyme 3.2.1.20 is longer and therefore more suitable for expanding the number of steps being manipulated in future research.

Figure 2. One naturally occurring sucrose degradation pathway as part of the starch and sucrose metabolic pathway. Highlighted in figure 1 by purple box.

Pathway Linked to Cell Fitness

Sucrose degradation can be linked to cell fitness by the expression of the sacB gene, which encodes for levansucrase 2. Levansucrase synthesizes levans and is lethal to *E. coli* in the presence of sucrose 2. Bacteriathat produce levansucrase will survive only in the absence of sucrose. Formation of a successful sucrose degradation pathway can be detected by providing sucrose to the cell population, because only those cells that degrade sucrose will survive. Selecting for cell survival is preferred to selecting for a negative phenotypic output, such as sucrose production in *E. coli* with the sacB gene, because it is more certain that the cells survived as a result of a successful pathway and the cells can be used again.

Figure 3. One naturally occurring sucrose degradation pathway as part of the starch and sucrose metabolic pathway. Highlighted in figure 1 by blue box.

*E. coli* strain characteristics

*E. coli* with neither enzyme 3.2.1.20 nor 2.4.1.7 should be used. Without 2.4.1.7, the cells must degrade sucrose using the biological pathway being optimized. Preventing the cells from producing 3.2.1.20 is preferred to preventing production of an enzyme further down the pathway, because it will stop sucrose degradation by 3.2.1.20 and ensure that any sucrose degradation is due to the selected pathway rather than endogenous pathways. The *E. coli* will need to have the sacB gene so that optimization of a sucrose degradation pathway is connected to cell fitness. Researchers should select for the cells that are the most resistant to the presence of sucrose.

**Pathway 2: Tetracycline biosynthesis**

Tetracycline biosynthesis pathway for optimization

 *E. coli* does not naturally produce any enzymes or substrates present in the tetracycline biosynthesis pathway (see figure 4) 3. The entirely synthetic pathway may be preferred to a naturally occurring pathway, because it can be certain that production of tetracycline is entirely the result of the synthetic pathway. The pathway is linear which may make it easier to control than a non-linear pathway with many different inputs needed for one step. The pathway is also long (27 steps from acetyl-coA to tetracycline) making it suitable for manipulations of additional pathway steps in future research.

Figure 4. Tetracycline biosynthesis pathway. Larger image can be viewed at http://www.genome.jp/kegg/pathway/map/map00253.html.

Pathway Linked to Cell Fitness

 Tetracycline production can be linked to cell fitness using a mixed population of *E.coli* cells in which one strain of *E. coli* is tetracycline-resistant and the other strain of *E. coli* is tetracycline-susceptible. The susceptible strain would produce RFP so that it could be distinguished from the tetracycline resistant strain. *E. coli* cells that successfully produce tetracycline would act as predators against the susceptible strain. Tetracycline synthesis would be detected by the rate of loss of RFP, which correlates to the death of susceptible cells.

 All tetracycline-resistant cells from the mixed population system would still be present regardless of whether or not they actually produced tetracycline. Including a sensor plasmid in the tetracycline-resistant cells could prove a useful way to detect which cells actually produced tetracycline. The sensor could also be used to compare the relative amounts of tetracycline produced by the tetracycline-resistant cells. Sensor plasmid pTetLux1 could distinguish between tetracycline-resistant cells that produce tetracycline and tetracycline-resistant cells that do not produce tetracycline. The sensor plasmid emits blue light in the presence of tetracycline 4. The more blue light emitted by the cells, the more tetracycline present 4. The cells that emit blue light would be those that have successfully produced tetracycline. Researchers would select for the cells that exhibit the highest amount of bioluminescence, because those cells would have produced the most tetracycline.

 Alternatively researchers could use just one strain of *E.* coli. The strain would be tetracycline-resistant and would contain sensor plasmid pTetLux1. The cells that first emit blue light would be those that have successfully produced tetracycline the fastest. The cells that exhibit the highest amount of bioluminescence would be those that have successfully produced the most tetracycline. Such a set-up does not link tetracycline to cell fitness and thus may not be as desirable for an initial system of study. It also may be more labor intensive to determine which cells emit blue light first.

In the initial study of the sensor plasmid, it was found that high levels of tetracycline reduce the amount of bioluminescence. It was hypothesized that such a trend is due to the “toxicity of tetracycline to the sensor cells protein synthesis machinery” (Korpela et. al. 4461) 4. The reduced bioluminescence may pose a problem if the optimized pathway produces more than 80ng of tetracycline, because the cells that produced the most tetracycline would start to die. In this case, researchers would need to use a different sensor to determine which cells produced tetracycline and in the highest amounts.

**Pathway 3: Vanillin synthesis**

Vanillin synthesis pathway for optimization

 *E. coli* does no naturally produce any enzymes or substrates that are part of the vanillin synthesis pathway (see figure 5) 5. The pathway is entirely synthetic so it can be certain that all vanillin production is the result of the synthetic pathway. The pathway’s linearity may make it easier to control than a non-linear pathway that requires multiple inputs for one step. The pathway can be expanded to up to 11 before becoming part of naturally occurring *E. coli* pathways. The step prior to trans-4-hydroxy-cinnamate involves trans-cinnamate, which *E.coli* can naturally degrade with enzyme 1.14.12.19 5. In future research, the pathway may be expanded beyond 11 steps by using an *E. coli* mutant that lacks enzyme 1.14.12.19. The external factors for maximum vanillin production in the JM109 strain of *E. coli* described by Barghini et al. can be used when optimizing the vanillin production pathway to increase the amount of vanillin produced by the optimized pathways 6.

Figure 5. Vanillin synthesis pathway highlighted by orange box as part of the phenylalanine metabolism pathway. Larger image can be viewed at http://www.kegg.jp/kegg-bin/show\_pathway? org\_name=ebw&mapno=00360&mapscale=1.0&show\_description=hide.

Pathway Linked to Cell Fitness

 Vanillin synthesis can be linked to cell fitness by using a vanillin biosensor 7. When the biosensor detects vanillin in a cell’s periplasm, it goes through a series of steps that ultimately lead to transcription of the pOmpR promoter. Having a gene that makes the cells resistant to an antibiotic like tetracycline after the pOmpR promoter would result in vanillin-producing cells surviving in the presence of tetracycline and non-vanillin-producing cells dying in the presence of tetracycline. Researchers would select for the cells that exhibited the most resistance to the antibiotic. The biosensor needs to be tested in wet lab before incorporating it into pathway optimization research, because it has only been tested computationally.

 If the biosensor does not function as described, screening may be the only option to determine which cells produce vanillin. Smell could be used as an initial screening method; plates with the strongest vanillin smell would have cells that produced the most vanillin. Using smell as an initial screening method would reduce the number of colonies that need to be analyzed with HPLC. Each colony on the plates with the strongest vanillin smell could be tested for successful vanillin synthesis using HPLC 6. HPLC analysis to detect vanillin would take nine minutes 8. Such a procedure may be time-consuming and tedious because each colony would need to be tested. More would be known about the amount of time screening would take once researchers determined how many colonies would need to be tested on average.

**Execution**

 Researchers should use a short pathway for initial optimization attempts. With fewer steps the pathway will be more easily controlled and with manipulations described in the coming section, there will still be a large number of possible pathway sequences to study.

Promoters

 In initial pathway optimization attempts each step in the pathway should have a different promoter so that they can be controlled independently. Later studies could design pathways with steps that use the same promoter. The following types of promoters may be used in any desired combination.

Constitutive Promoters:

 A promoter that is always on may be a useful way to reduce pathway noise when paired with weaker transcription rates. A different promoter from the J23100 family of constitutive promoters could be used for each pathway step 9. The constitutive promoters vary in strength, and J23119 is the strongest promoter in this family 10. Each step could use a constitutive promoter and the promoter strength at each step could be manipulated for pathway optimization. Alternatively constitutive promoters could be used in combinations that include activatable promoters and/or repressible promoters.

Activatable Promoters:

 Activatable promoters would allow researchers to better control when steps in the pathways were activated, particularly if each step in the pathway uses an activatable promoter with a different activator. These promoters could also be used in conjunction with repressible promoters and/or constitutive promoters. The pBAD family, which includes a wild type, weak, and strong promoter, is one family of activatable promoters that could be used in pathway optimization research 11. These promoters are activated in the presence of arabinose. The pLux promoter is a promoter that is activated in the presence of LuxR and Al-1 12. Hybrid promoters such as the pLux/cI promoter sense two inputs 13. pLux/cI senses AHL induction and Lambda cI repression. Such promoters may be useful when two promoters in the pathway share an input. For example, the pLux/cI promoter could be used in the same pathway as a promoter that is induced by AHL because it will initiate transcription only when Lambda cI is also repressed. It should be noted that the pOmpR promoter in the vanillin biosensor is an activatable promoter.

Repressible Promoters:

 Repressible promoters would allow researchers to better control when steps in the pathways were repressed, particularly if each step in the pathway uses a promoter with a different repressor. These promoters could be used in pathways that include activatable promoters and/or constitutive promoters. The lambda cI regulated promoter, which is repressed by cI binding, is an example of a repressible promoter 14.

RBS sequences

 Manipulation of RBS sequences can be used to control translation initiation rates. In addition to acting as a variable in pathway optimization, RBS sequences can be used to reduce the amount of noise in a pathway. Low rates of translation paired with high rates of transcription will reduce noise, while the reverse will increase noise.

There are two families of constitutive RBS sequences, and each has sequences of varying strengths 15. For example in the Anderson family, RBS sequence B0064 has a translation initiation rate that is 35% as strong as RBS sequence B0034 16. RBS sequences in the Issacs RBS family regulate the translation initiation rate by blocking or opening RNA secondary structure 17. Ability to increase or decrease the rate of translation would allow researchers to study whether rate of translation rate affects the amount or rate of pathway output. Researchers could also use Salis Lab’s RBS calculator to generate RBS sequences that will yield a specified rate of translation 18.

Each step in the pathway could have C. dog, which involves two ribosomal binding sites to control the rate of translation 19. The first site is fixed and the second site is variable. It is unclear how including an RBS sequence that yields a specified rate of translation will function within the C. dog system. It is also unclear how different RBS sequences at the second site will affect C. dog. Manipulations of RBS sequences within C. dog may be a topic researchers would like to explore. Initial pathway optimizations may be best left without C. dog to reduce the level of complexity in this particular part of the pathway steps.

Protein degradation tags

 Protein degradation tags can be used to control how long proteins are present in the cell and available to react with other steps in the pathway 20. Incorporation of degradation tags should be taken on a case-by-case basis depending on the enzymes and substrates produced by each pathway step being optimized. In general slow enzymes should not be degraded, as they take longer to find and bind with the target substrate. Degradation may prevent an enzyme-substrate interaction necessary for the desired output whereas enzymes that act quickly may be degraded to prevent overstraining the cell. Degradation tags may not be needed in each step of the pathway or in any step if the pathway involves only enzymes that function slowly. Researchers can use degradation tags in all of the steps, in none of the steps, or in any other combination desired.

Degradation tags of different strengths degrade proteins at different rates. For example, AANDENYALAA is a very fast ssrA degradation tag that degrades proteins via SspB-mediated binding to both ClpX and ClpA proteases 21. AANDENYADAS degrades proteins at a moderate rate, and AANDENYNYADAS degrades proteins at a rate between that of AANDENYALAA and AANDENYADAS 22. Degradation tags of varying strengths can be used in any desired combination for pathway optimization.

Order of steps

 Given the large number of variables described above that can be manipulated to optimize the suggested pathways, only manipulation of the order of pathway steps will be discussed. For initial optimization attempts, researchers should focus on classifying the effects of a manageable number of variables on the pathway. It should be kept in mind that future research could also manipulate the orientation (forward or reverse) of pathway steps to study how such changes affect the pathway. Orientation changes could be accomplished by incorporating hix sites between steps of the pathway 23.

 Changing the order of pathway steps may affect how well the pathway functions. For example, if a gene that leads to substrate formation is transcribed after a gene that leads to enzyme formation, the enzyme-substrate reaction may take a longer time because the enzyme is essentially waiting for the substrate to be produced. If a protein degradation tag degrades an enzyme very quickly, forming the substrate first may be preferred so that the enzyme has a higher probability of encountering substrate immediately upon its formation. These examples are two scenarios that represent potential ways the order of the pathway steps affect its optimization.

 Barcode primers could be used to detect the order of the pathway steps in the cell 23. Each step in the pathway should be associated with a different barcode so that they can be distinguished from one another once incorporated into *E. coli*. A barcode primer should be placed in front of each step in the pathway, and PCR can be used to determine the order of the steps.

Incorporation in Cells

 The aforementioned pathway parts may be incorporated into the bacteria cells with CRIM plasmids. CRIM plasmids insert a segment of the pathway through binding their attP sites to attB sites in *E. coli* 20. A helper plasmid that makes Int protein aids in incorporating the CRIM plasmids into the genome. CRIM plasmids can be removed without alteration of the plasmid with helper plasmids that make Int protein and Xis protein 20. The ability to remove the plasmids may be used to verify that the phenotype observed is due to the gene on the plasmid rather than other cellular processes. Such verification would be most beneficial with semi-synthetic or natural pathways where researchers are unsure whether output production is due to the pathway being optimized or due to another pathway already present in the cell. Alternatively 8-cutters and standard plasmid integration could be used. Each segment would need an 8-cutter recognition site so that it can be excised from the bacterial genome independent of the other segments 23. Some examples of potential 8-cutters include FseI, PacI, and AscI 24.

Figure 6. Type IIs restriction enzymes can be used to swap out particular parts of a pathway step.

Once the pathway segments are incorporated into the cells, type IIs restriction enzymes can reinsert and swap out parts of each pathway step as described by Missouri Western State University 25 (see figure 6).

Noise

As mentioned in earlier sections, manipulating different pathway parts may affect the amount of noise present. Reducing the amount of noise is the most likely goal of the research manipulations described above. Noise may make the pathway harder to control and may make it harder to determine which version of the pathway is the optimized pathway. Strong promoters followed by weak translation can be used to minimize noise 26. A constitutive promoter paired with weaker transcription rates can reduce noise. Selecting a shorter pathway, manipulating fewer steps, or manipulating steps closer to the desired output can reduce noise from the output, because noise tends to increase with longer pathways. If noisy inputs are the problem, choosing a longer pathway may be a better option, because longer pathways are not as easily activated by sub-threshold levels of activation and are thus more tolerant of noisy inputs 26.

**References**

1. “Starch and Sucrose Metabolism: Reference Pathway.” *Krypto Encyclopedia of Genes and Genomes*. Feb. 1, 2012. http://www.kegg.jp/kegg-bin/show\_pathway?org\_name=eco&mapno =00500&mapscale=&show\_description=hide (accessed Apr. 27, 2012).

2. French, Chris and Maria Kowel. “B. subtilis levansucrase. Lethal to *E. coli* in presence of sucrose.” *Registry of Standard Biological Parts.* Sept. 24, 2009. http://partsregistry.org/ Part:BBa\_K322921 (accessed Apr. 27, 2012).

3. “Tetracycline Biosynthesis: Reference Pathway.” *Krypto Encyclopedia of Genes and Genomes*. Feb. 1, 2012. http://www.genome.jp/kegg/pathway/map/map00253.html (accessed Apr. 27, 2012).

4. Korpela, Matti T. et al. “A Recombinant *Escherichia coli* sensor strain for the Detection of Tetracyclines.” *Analytical Chemistry, Vol. 70, No. 21*. Nov. 1, 1998. http://pubs.acs.org/doi/pdf /10.1021/ac980740e (accessed Apr. 27, 2012).

5. “Phenylalanine Metabolism: Reference Pathway.” *Krypto Encyclopedia of Genes and Genomes*. Feb. 1, 2012. http://www.kegg.jp/kegg-bin/show\_pathway?org\_name=map&mapno= 00360&mapscale=1.0&show\_description=hide (accessed Apr. 27, 2012).

6. Barghini, Paolo et al. “Vanillin production using metabolically engineered *Escherichia coli* under non-growing conditions.” *Microbial Cells Factories, 6:13.* Apr. 16, 2007. http://www.ncbi. nlm.nih.gov/pmc/articles/PMC1857700/ (accessed Apr. 27, 2012).

7. Aparici, A. et. al. “Synthetic periplasmic binding protein that docks a vanillin molecule.” *Registry of Standard Biological Parts*. Oct. 26, 2006. http://partsregistry.org/wiki/index.php? title=Part:BBa\_J58105 (accessed Apr. 27, 2012).

8. “Fast Determination of Vanillin and its Synthesis Precursor by HPLC.” *Dionex Corporation*. 2009. http://www.dionex.com/en-us/webdocs/81981-AU170-HPLC-Vanillin-SynthesisPrecursor-06Oct2009-LPN2314.pdf (accessed Apr. 27, 2012).

9. “*E. coli* Constitutive Promoters.” *Registry of Standard Biological Parts*. http://partsregistry. org/Promoters/Catalog/Ecoli/Constitutive (accessed Apr. 27, 2012).

10. Anderson, John. “Constitutive Promoter Family Member JM23119.” *Registry of Standard Biological Parts*. Aug. 24, 2006. http://partsregistry.org/wiki/index.php?title=Part:BBa\_J23119 (accessed Apr. 27, 2012).

11. “pBAD Promoter Family.” *Registry of Standard Biological Parts*. http://partsregistry.org/ PBAD\_Promoter\_Family (accessed Apr. 27, 2012).

12. Feeney, Erin. “pLux.” *Registry of Standard Biological Parts*. July 25, 2008. http://partsregistry.org/wiki/index.php?title=Part:BBa\_K091156 (accessed Apr. 27, 2012).

13. Iwasaki, Kenichiro. “pLux-cI hybrid promoter.” *Registry of Standard Biological Parts*. Oct. 25, 2007. http://partsregistry.org/wiki/index.php?title=Part:BBa\_I751501 (accessed Apr. 27, 2012).

14. Mahajan, Vinay S. et. al. “promoter (lambda cI regulated).” *Registry of Standard Biological Parts*. Jan, 31, 2003. http://partsregistry.org/wiki/index.php?title=Part:BBa\_R0051 (accessed Apr. 27, 2012).

15. “Ribosome Binding Sites: Constitutive Community Collection.” *Registry of Standard Biological Parts*. http://partsregistry.org/Ribosome\_Binding\_Sites/Prokaryotic/Constitutive/ Community\_Collection (accessed Apr. 27, 2012).

16. Kelly, Jason. “Tuned RBS for Q04401.” *Registry of Standard Biological Parts*. Feb. 23, 2008. http://partsregistry.org/wiki/index.php/Part:BBa\_B0064 (accessed Apr. 27, 2012).

17. “Ribosome Binding Sites: Regulated Isaacs.” *Registry of Standard Biological Parts*. http://partsregistry.org/Ribosome\_Binding\_Sites/Prokaryotic/Regulated/Isaacs (accessed Apr. 27, 2012).

18. “RBS Calculator.” *Salis Lab of Pennsylvania State University*. https://salis.psu.edu/software/ (accessed Apr. 27, 2012).

19. Gittin, Alex and Dancho Penev. “Reports on Circuits: C. dog.” *Synthetic Biology Network Research*. Apr. 10, 2012. http://gcat.davidson.edu/mediawiki1.15.0/index.php/ WEEK\_TWELVE\_(April\_2\_-\_6) (accessed Apr. 27, 2012).

20. Evans, Becca and Kirsten Huffer. “Reports of Circuits: CRIM Plasmids, Degradation Tags, and Transposons.” *Synthetic Biology Network Research*. Apr. 10, 2012. http://gcat.davidson.edu/ mediawiki1.15.0/index.php/ WEEK\_TWELVE\_(April\_2\_-\_6) (accessed Apr. 27, 2012).

21. Moser, Felix. “AANDENYALAA. (Very fast) SsrA degradation tag.” *Registry of Standard Biological Parts*. Dec. 5, 2007. http://partsregistry.org/wiki/index.php?title=Part:BBa\_M0050 (accessed Apr. 27, 2012).

22. “Protein Domains: Degradation.” *Registry of Standard Biological Parts*.http://partsregistry.org/Protein\_domains/Degradation (accessed Apr. 27, 2012).

23. Clarkson, Ben and Duke DeLoache. “Reports of Circuits: Double Cloning with Type IIs Restriction Enzymes.” *Synthetic Biology Network Research*. Apr. 10, 2012. http://gcat.davidson.edu/ mediawiki1.15.0/index.php/ WEEK\_TWELVE\_(April\_2\_-\_6) (accessed Apr. 27, 2012).

24. “Recleavable Filled in 5' Overhangs.” *New England BioLabs Inc.* http://www.neb.com/ nebecomm/tech\_reference/restriction\_enzymes/recleavable\_filled\_5\_overhangs.asp#.T5X41-24Ldk (accessed Apr. 27, 2012).

25. “Programming Bacteria for Optimization of Genetic Circuits.” *Missouri Western State University*. Mar. 7, 2012. http://gcat.davidson.edu/mediawiki-1.15.0/index.php/Synthetic\_ Biology\_Network\_Research (accessed Apr. 27, 2012).

26. Baker, Erich et al. “Reports of Circuits: Noise and Optimal Pathway Characteristics.” *Synthetic Biology Network Research*. Apr. 10, 2012. http://gcat.davidson.edu/ mediawiki1.15.0/index.php/ WEEK\_TWELVE\_(April\_2\_-\_6) (accessed Apr. 27, 2012).