Introducing an Anti-Terminator Paralog Gene to Induce Production of Natural Products in *Clostridium* Species

By: Alexander Leung; Research Sponsor (PI): Dr. Wenjun Zhang

ABSTRACT

LoaP is a class of proteins that has the potential to induce bacterial species into creating natural products such as antibiotics. LoaP can be placed as an insert into bacterial genomes, where it can potentially drive gene expressions and subsequently activate secondary biosynthetic pathways. Here, a construct was cloned by placing LoaP as an insert into the plasmid of *Clostridium beijerinckii*. However, this LoaP overexpression strain failed to produce a compound that its wild-type creates. LoaP appeared to decrease production rather than increase production of certain metabolites, as the wild-type strain gave rise to more natural products than the transformant LoaP strain.

Major, Year, and Department: Bioengineering, 2017, Department of Chemical and Biomolecular Engineering.

Keywords: B-598, nusG, LoaP, Pbdh, paralog, anti-terminator, secondary metabolites, XCMS, mass spectrometry, transformation.

INTRODUCTION

Bacteria have been studied for their ability to form natural products that can be useful for drug discovery and other applications, such as dyes and other synthetic materials. Some natural products come in the form of secondary metabolites, which have been observed from *Clostridium*.[1] The genomes of different *Clostridium* species have been studied for secondary metabolite sequences, such as polyketide synthase and non-ribosomal peptide synthetase genes, which are widespread amongst these species.[1] These genes are only activated by particular stimuli, such as environmental conditions found in aqueous soil extract.[1] Since the necessary stimulatory environmental conditions are difficult to recreate in a laboratory setting, bacteria are often incapable of producing secondary metabolites *in vitro* due to the dormancy of usual secondary metabolite biosynthetic pathways. In a previous study, overexpression of anti-terminator genes corresponding to biosynthetic gene clusters in *Clostridium celluloyticum* yielded novel natural products such as antibiotics known as closthioamides and related thioamides.[1]

In the former study, an anti-terminator gene was used to induce a secondary metabolite pathway. This class of proteins is referred to as LoaP. LoaP is a paralog from the *nusG* family of proteins, which tend to play an important role in transcription elongation.[2] Anti-terminator proteins prevent termination of transcription at terminator sites, and allow for expression in the sequences beyond them.[3] In the absence of an anti-terminator protein, RNA polymerase, or RNAP, will cease transcription at a terminator. In this context, the anti-terminator serves to increase the overall rate of transcription.[1] In RNAP complexes, *nusG* moves transcription forward past termination sites (known as p-dependent sites), which increases RNA polymerase processivity and helps with both making transcription efficient[1] and driving the creation of polycistronic mRNA.[1] Overall, this process helps the progression of protein production. We hypothesized that this may drive a different pattern of gene expression, which may in turn drive some biosynthetic gene pathways.

Overexpression of these anti-terminator genes in biosynthetic clusters may give rise to novel natural products produced by *Clostridium beijerinckii*, which is also known as B-598. In B-598, a paralog of *nusG* called LoaP was discovered.[2] If the B-598 paralog were to be excised from B-598's bacterial genome and placed as an insert into B-598's genome under a different constitutive promoter, it should activate a biosynthetic gene cluster and, potentially, create cryptic natural products. In our study, the LoaP protein was expressed under a constitutive promoter called Pbdh, which is from a different bacterial species known as *Clostridium saccharoperbutylacetonicum*.[2] Pbdh typically regulates butanol dehydrogenase. Promoters are sequences of DNA where transcription is initiated. Promoters are also important in regulating expression of downstream genes. There is a possibility that anti-terminator proteins will not be expressed in certain promoters. The promoter Pbdh was selected due to its constitutive nature, which ensures that it is continuously active and, therefore, does not need to be induced. Once the anti-terminator proteins are expressed under this promoter, a cryptic biosynthetic pathway can be driven in B-598 to yield novel natural products such as secondary metabolites. Figure 1 shows that the sequence used was associated with Cluster 10, which is a biosynthetic cluster that was hypothesized to be responsible for secondary metabolite activity. From this, it is possible that overexpression of LoaP in B-598 using a constitutive promoter may yield novel natural products.

MATERIALS AND METHODS

I. Cloning and Gibson Assembly

Cloning was done by inserting the B-598 *nusG* into the backbone of B-598 under the Pbdh promoter. Lab restriction enzymes are used to cut out B-598's native *nusG* and replace it with B-598's LoaP under a different promoter. The backbone in this procedure is named PJL5. Its native *nusG* is cut out with BSP/MSSI, and generates sticky ends to facilitate insertion of the B-598's LoaP. B-598's paralog was also amplified via PCR to generate the plasmid DNA sequence of B-598. Both the PCR product and restriction digests were run in a gel electrophoresis procedure that separates out the bands so that they may be cut out of the gel and purified in a gel extraction procedure. These same restriction enzymes were also used to cut out LoaP from B-598's plasmid sequence to make it homologous to the cut-out site in PJL5. B-598's LoaP sequence was then ligated to the PJL5 plasmid, and the new plasmid was named PJL9.

However, it was discovered that restriction digest enzymes were not

Figure 1. B-598 derived LoaP Associated with Cluster 10 and Used in the Experiment.[8]

needed as the Gibson Assembly procedure helped to create sticky ends to avoid blunt-end DNA ligation. The Gibson Assembly procedure has the same principle as the restriction enzyme procedure and it was used in this experiment to generate the constructs. Without Gibson Assembly, in blunt-end DNA ligation, the LoaP insert may not have necessarily ligated to the correct site and may have ligated to any other available homologous site. Luckily, PJL9 was successfully constructed using Gibson Assembly. The theoretical construct created was seen in a model of the plasmid shown in Fig. 2. The sequence was also confirmed to be correct through a sequencing facility.

After PJL9 was generated, it was inserted into a bacterial colony via a transformation procedure. 5 μL of the plasmid's DNA was added to one 50 μL aliquot of *E. coli* XL-1 Blue competent cells. The mixture was then incubated on ice for 30 minutes, before undergoing heat shock. In heat shock, the mixture was placed in a 42°C bath for 45 seconds. During the heat-shock step of the transformation procedure, *E. coli* XL-1 Blue competent cells incorporated the plasmid. The constructed plasmid was amplified as the colony was grown on an LB plate with a Carbenicillin antibiotic. Colonies on the media surface were picked and grown in liquid LB media in sterile tubes until the media showed cloudiness or bacterial growth. The DNA was extracted from the XL-1 Blue competent cells via a Zyppy Plasmid mini-prep procedure to obtain the sequence of the constructed plasmids. In order to confirm that the sequence is correct, test digests were run on the extracted DNA with a HindIII restriction enzyme. Constructs with the correct bands were prepared for sequencing and submitted to a sequencing facility to confirm if the insert had successfully ligated to the backbone.

II. Transformation into B-598

After extracting the DNA from *E. coli* cells, running test digests, and confirming the sequences are correct through the sequencing facility, the final PJL9 plasmid was incorporated into B-598 colonies. The lab had developed a specialized transformation procedure specifically that was implemented for this experiment.[4]

III. Fermentation and Chemical Extraction

The transformants were grown in TYA media with antibiotic erythromycin at 40 μg/mL for four days.[5] In addition, a wild-type bacterial control was grown alongside the transformants (also known as LoaP overexpression strains) in order to serve as a basis for comparison. Both colonies were grown in triplicate cultures, meaning there were three samples for the transformants and three samples for the wild-type bacteria in order to ensure that there are enough data to compare and demonstrate consistency. After 4 days, the transformants and wild-type bacteria were chemically extracted with EtOAc. EtOAc was added to the liquid cultures in a 1:1 ratio, before being vortexed. Afterwards, the mixture was spun down, the top layer was removed, then dried with nitrogen lines. The resulting extract was then suspended in HPLC-grade methanol and run in a quadrupole time-of-flight (QTOF) mass spectrometer for analysis.

Figure 2.* nusG *B-598 insert (LoaP) placed into the backbone DNA sequence of B-598 to yield the PJL9 sequence.

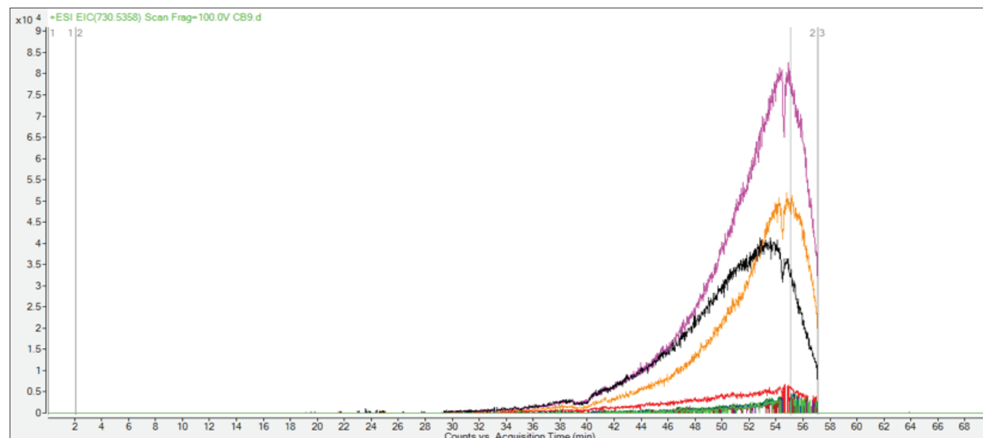

Figure 3. Chromatogram comparing B-598 LoaP overexpression strains versus wild-type. *Orange, black, and pink curves represent the wild-type bacteria, while red, blue, and green curves represented the transformants with the LoaP. The graph indicates that the 730.5358 m/z mass eluted from the column at fifty-five minutes.*

IV. QTOF/XCMS

Compounds were separated out through MS-Reverse-Phase HPLC. The mass-to-charge ratio of both the transformants and wild-type colonies were analyzed to see what compounds are being produced. However, in order to narrow down whether the transformants have a new compound or are missing compounds, an XCMS procedure was implemented. XCMS is a data-analysis approach for mass spectrometry data that examines spectrometry variables such as peak matching, peak detection, retention time alignment, etc.[7] In this study, XCMS generated a table with 1 row for every single mass peak detected. It then calculated the average size of that peak for each replicate and took the average of these values. The most significant variables in this study are median retention times, fold changes, and p-values. Median retention times signify how long the peak was observed in the sample. Since the machine has a 55-minute gradient, only a range of 2 to 57 minutes was considered.[7] Fold change reflects the difference between the average LoaP and control value and a 10-fold change is usually significant.[7] P-value indicates how statistically significant are the results, 0.01 or 0.05 being the cutoff. Using these three variables, the rows were narrowed down to obtain only significant results. Each row corresponds to a new peak, which then means that either a new mass was found or a typical mass is missing. After the rows were obtained, we returned to the raw data to analyze significant results.

RESULTS

The mass-to-charge ratio that was significant as identified from the XCMS method was 730.5358 m/z. This value satisfies the XCMS procedure by having a p-value larger than our cutoff of 0.01 and 0.05, while also having nearly a 10-fold change as indicated by Table 1. The chromatogram indicated a possible major significant difference between the LoaP overexpression and control samples as seen in Fig. 3. There are six curves in total to signify three wild-type samples and three LoaP samples. The LoaP curves were close to the baseline, indicating that the mass was not present in all three LoaP overexpression samples. Meanwhile, all wild-type curves showed high counts, indicating that the mass was present in a pretty high quantity as opposed to the LoaP overexpression samples. Another QTOF data file comparing counts to mass-to-charge ratio in Fig. 4 serves to confirm that the mass is present in the wild-type. Overall, the results indicate that a nonpolar compound was produced in the wild-type bacteria, but not the B-598 derived LoaP overexpression samples. This serves to show that the mutant strain was eliminating a product that would usually be created in the wild-type colonies.

DISCUSSION

The results obtained were unexpected, and contradicted the original hypothesis. Based on previous studies[2] we expected that inserting anti-terminator genes into bacterial genomes could change gene expression and activate biosynthetic pathways that can create new secondary metabolites. However, the results indicated that an unknown compound that was produced in the wild-type was not produced by the B-598 derived LoaP samples. In fact, it seemed as if the anti-terminator protein may have decreased production of one of the compounds usually produced. However, re-analysis of the B-598 genome indicated there are actually two LoaP sequences within the genome. The sequence used in the experiment was hypothesized to be associated with secondary metabolite biosynthetic gene clusters. Perhaps the loss of one product could be due to the fact that the sequence had some minor secondary effects that inhibited the creation of that one product. What exactly are the secondary effects is unknown at this time, although it can be theorized that LoaP served as a negative regulator for that product. Namely,

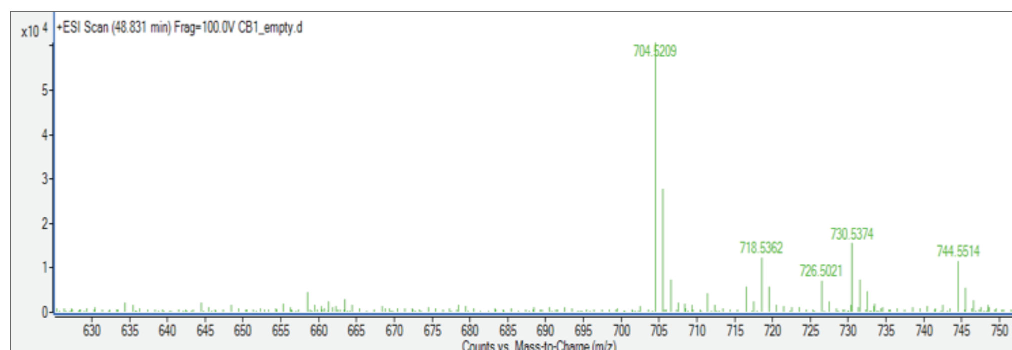

Figure 4. Counts versus mass charge confirming the presence of the 730.5358 m/z in the wild-type bacteria.

name	fold	log2fold	tstat	pvalue	qvalue	mzmed
M731T50	905.8176	-9.8230	-2.9431	0.0986	0.3131	730.5358

Table 1. Statistical variables from XCMS Procedure Corresponding to the 730.5358 m/z.

it would prevent a certain gene from being expressed, thereby inhibiting the pathway that would otherwise create the product. In the future, the copy of LoaP found (associated with Cluster 18, Fig. 6), will be used instead to see if that induces the B-598 strain to create new natural products. Analysis of the cluster indicates that Cluster 18 might be a biosynthetic pathway for saccharides, which are primary metabolites. Regardless, overexpressing this copy might yield more fruitful results than the B-598 sequence in Cluster 10. In future work, the cloning procedure could be repeated for this homolog and inserted into B-598 to be overexpressed under the same promoter to test its ability to induce secondary metabolites. We hypothesize that a LoaP copy that is associated with a saccharide cluster could induce cryptic biosynthetic pathways of new metabolites when inserted into B-598. Another alternative is to search for other proteins often associated with heavy biosynthetic gene clusters (preferably anti-terminators) and undergo the same procedure to study secondary metabolite effects. Preferably, these proteins need to be positive regulators so that they simply can be expressed under the same promoters using the same cloning procedure mentioned above.

CONCLUSION

Theoretically, overexpression of the *nusG* vector in bacteria would help to induce bacteria into creating natural products such as antibiotics. However, the LoaP overexpression strain demonstrated the opposite of what was expected. In the chromatogram results, the wild-type bacteria were discovered to contain a peak with a mass that the transformant strain lacked. This runs contrary to the hypothesis that the insertion of the B-598 derived LoaP into B-598 would induce biosynthetic pathways in B-598, which could in turn create useful products like closthioamide.[2] The sequence inserted into B-598's decreased production in the bacteria. This could be due to the fact that the sequence used had some secondary effects that inhibited creation of the expected product. In the future, another LoaP sequence associated with a saccharide gene cluster can be inserted into the B-598 genome and studied for secondary metabolite activity. Other proteins similar to LoaP can also be used in a similar fashion to induce secondary metabolite activity.

ACKNOWLEDGEMENTS

I want to thank Professor Zhang and Jeffrey Li for allowing me to work under their facility and to conduct experiments. The Zhang Lab in University of California, Berkeley has offered me abundant resources and guidance in helping me to conduct scientific research of my own.

REFERENCES

1. Behnken, S., Lincke, T., Kloss, F., Ishida, K., & Hertweck, C. (2012). Antiterminator-Mediated Unveiling of Cryptic Polythioamides in an Anaerobic Bacterium. *Angewandte Chemie International Edition, 51*(10), 2425-2428. doi: 10.1002/anie.201108214
2. Goodson, J. R., Klupt, S., Zhang, C., Straight, P., & Winkler, W. C. (2017). LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in Bacillus amyloliquefaciens. *Nature microbiology, 2*(5), 17003. doi: 10.1038/nmicrobiol.2017.3
3. Clark, D., & Pazdernik, N. (n.d.). Antitermination. *ScienceDirect*. Retrieved from https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/antitermination
4. Herman, N. A., Li, J., Bedi, R., Turchi, B., Liu, X., Miller, M. J., & Zhang, W. (2017). Development of a high-efficiency transformation method and implementation of rational metabolic engineering for the industrial butanol hyperproducer Clostridium saccharoperbutylacetonicum strain N1-4. *Appl. Environ. Microbiol., 83*(2), e02942-16. doi: 10.1128/AEM.02942-16
5. Chang, Y. Y., & Cronan, J. E. (1982). Mapping nonselectable genes of Escherichia coli by using transposon Tn10: location of a gene affecting pyruvate oxidase. *Journal of bacteriology, 151*(3), 1279-1289. doi: 10.1007/BF00328074
6. Division of Endocrinology, Metabolism and Lipid Research. (2013, Nov. 19). Time-of-Flight Fundamentals. *Biomedical Mass Spectrometry Resource*. Retrieved from https://msr.dom.wustl.edu/time-of-flight-fundamentals/
7. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical chemistry, 78*(3), 779-787. doi: 10.1021/ac051437y
8. Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., ... & Breitling, R. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic acids research, 43*(W1), W237-W243. doi: 10.1093/nar/gkv437

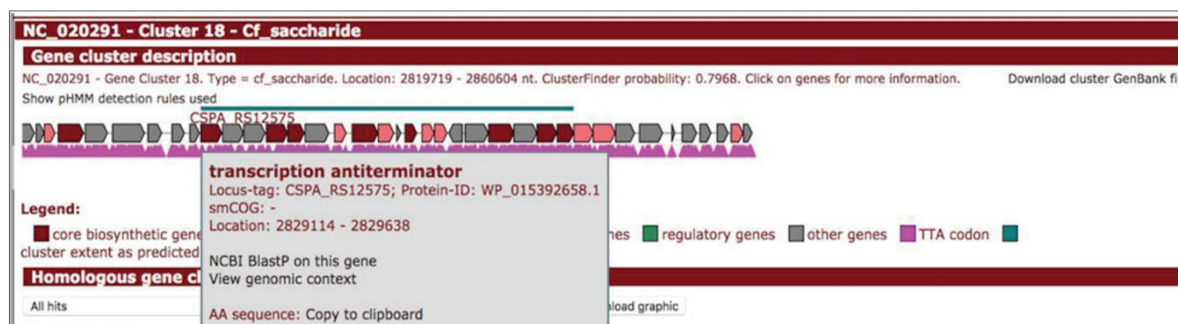

Figure 6. B-598 derived LoaP copy associated with Cluster 18, a saccharide biosynthetic cluster for future approaches.[8]