ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

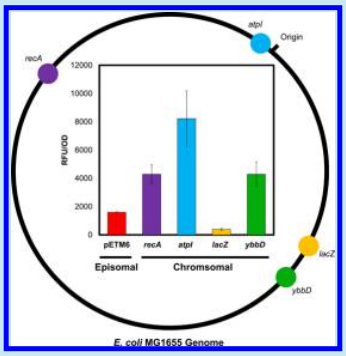
Effect of Genomic Integration Location on Heterologous Protein Expression and Metabolic Engineering in *E. coli*

Jacob A. Englaender,[†] J. Andrew Jones,[‡,▽] Brady F. Cress,[‡] Thomas E. Kuhlman,[§,∥,⊥,#] Robert J. Linhardt,[†,‡] and Mattheos A. G. Koffas*[,†,‡]

[†]Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, New York 12180, United States
[‡]Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180, United States
[§]Department of Physics, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States
[∥]Center for the Physics of Living Cells, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States
[⊥]Carl. R. Woese Institute for Genomic Biology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States
[#]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States
[▽]Department of Chemistry, Hamilton College, Clinton, New York 13323, United States

Supporting Information

ABSTRACT: Chromosomal integration offers a selection-free alternative to DNA plasmids for expression of foreign proteins and metabolic pathways. Episomal plasmid DNA is convenient but has drawbacks including increased metabolic burden and the requirement for selection in the form of antibiotics. *E. coli* has long been used for the expression of foreign proteins and for the production of valuable metabolites by expression of complete metabolic pathways. The gene encoding the fluorescent reporter protein mCherry was integrated into four genomic loci on the *E. coli* chromosome to measure protein expression at each site. Expression levels ranged from 25% to 500% compared to the gene expressed on a high-copy plasmid. Modular expression of DNA is one of the most commonly used methods for optimizing metabolite production by metabolic engineering. By combining a recently developed method for integration of large synthetic DNA constructs into the genome, we were able to integrate two foreign pathways into the same four genomic loci. We have demonstrated that only one of the genomic loci resulted in the production of violacein, and that all four loci produced *trans*-cinnamic acid from the TAL pathway.



KEYWORDS: *genomic integration, violacein, flavonoid production, metabolic burden, cinnamic acid, episomal expression*

Microbial expression of heterologous proteins and metabolic pathways has traditionally been performed in *Escherichia coli* through episomal expression, by cloning the gene of interest into a plasmid vector.[1] Plasmid-based expression systems are widely used for a multitude of reasons. Most noticeably, plasmids are easy to work with in terms of molecular biology manipulations, allowing for easy cloning of foreign DNA. Recent efforts in plasmid-based metabolic pathway optimization have resulted in the design and characterization of a large library of compatible plasmid vectors spanning several orders of magnitude of expression through tailored combinatorial design of an ever-increasing library of parts.[2−7] However, there are also drawbacks to plasmid-based expression systems. Most notably, plasmids are nonessential chromosomes. As a result, selection, usually in the form of antibiotics, is required for plasmid retention in a bacterial culture.[8,9] This effect is then compounded by the fact that cells without antibiotic resistance pressure often grow more quickly than resistant strains.[10] In industrial settings, the need to supplement large-scale bioreactors with antibiotics is cost prohibitive. Furthermore, the use of antibiotics and their subsequent introduction into the environment can lead to development of microbial antibiotic resistance worsening what has been deemed as one of the largest global health concerns of the 21st century.[11−13]

Integration of foreign DNA into the genome of the host organism allows for the stable expression of foreign DNA without the need for antibiotic selection. Site specific integration of DNA into the genome is most often achieved through homologous recombination techniques.[14−16] Homologous recombination utilizes DNA repair enzymes and homologous DNA sequences to facilitate the placement of foreign DNA into a specific location on the genome. This method generally has a low efficiency that is dependent on both the size of the foreign DNA and the amount of homology on the flanking sequence.[8] Recent methods were developed that allow for the integration of large synthetic DNA constructs into the genome of *E. coli* by introducing double stranded breaks

Received: November 20, 2016
Published: January 5, 2017

DOI: 10.1021/acssynbio.6b00350
ACS Synth. Biol. 2017, 6, 710−720

into the chromosome to combat the losses in efficiency with increasing size of the targeted integration sequence.[8,14,17]

Traditional metabolic engineering efforts employ a push, pull, block approach, where enzymes are overexpressed to push flux toward cofactors and precursors, overexpressed to pull flux through the pathway of interest, and genes are deleted to block flux through competing pathways.[18] While some up-regulation and down-regulation targets are identified empirically, a large number of these targets can be accurately predicted computationally.[19−22] Genomic integration can also be used to reduce the overall number of genetic manipulations; for example, by replacing a deleted target gene with a gene to be up-regulated, genomic integration can achieve two goals at once. Using genomic integrations in this way can result in increased yields from heterologous pathways.[23]

Multicopy plasmid DNA can also place a heavy metabolic burden on bacterial cultures at the transcriptional level.[24,25] This burden can become a problem when optimizing production of a metabolite from a heterologous host. Chromosomal integration has the ability to alleviate this burden by decreasing the copy number of the genes being overexpressed, oftentimes from 40 or more copies per cell down to a single copy, and has previously been shown to be capable of increasing production of heterologous metabolites in *E. coli*.[24,26]

Build-up of potentially toxic intermediates can also negatively impact cellular growth and metabolite production. Because of this, fine-tuning expression of each gene in a pathway is frequently required to optimize production. This is often done by laborious testing using libraries of promoters of varying strengths, ribosome binding site engineering, or by altering the copy number of each gene in a pathway until an optimum production is reached.[3,4,27−30] Studies in *E. coli* and other organisms have demonstrated that gene expression from the genome is dependent on the location of the genes.[31−33] Thus, by changing the location of a gene on the chromosome, it is possible to modulate expression equivalently and additively to the techniques previously demonstrated for plasmid-based systems.

Here, we report a combination of the pTKRED system for chromosomal integration with the ePathBrick vector system for pathway engineering.[3,8] Through slight modification of the pTKIP and pTKDP integration vectors, we were able to make the systems compatible, allowing for easy construction and integration of entire metabolic pathways into *E. coli*. Through presentation of three case studies: (1) mCherry as a fluorescent reporter protein, (2) the five-gene pathway for production of the purple pigment violacein, and (3) a single-gene pathway for the production of *trans*-cinnamic acid (Figure S1), we demonstrate that protein expression and metabolite production in *E. coli* are influenced by the location of their respective integrations on the genome.

RESULTS

Integration of the Tetracycline-Resistant "Landing-Pad" into 4 Genomic Loci. The tetracycline-resistant "landing-pad" from the plasmid pTKS/CS was integrated into the *lacZ*, *atpI-gidB*, *recA*, and *ybbD-ylbG* loci (Figure 1). Amplification of DNA across the junction of native genomic DNA and the integrated "landing-pad" was used to verify successful integration into the *lacZ* locus of the *E. coli* MG1655(DE3) genome (Figure 2). As evidenced by the agarose gel, the efficiency of integration of the "landing-pad"

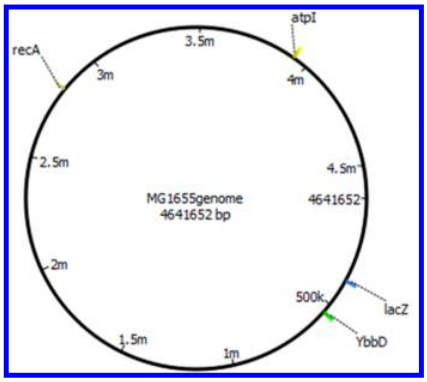


Figure 1. Locations of *recA*, *atpI-gidB*, *lacZ*, and *ybbD* loci on the genomic chromosome of *E. coli* MG1655(DE3). Genomic map generated in UGENE.

into the *lacZ* locus was 87% (n = 15). Similar results were obtained for integrating the "landing-pad" into the other genomic loci (data not shown). All colonies that showed correct amplification from colony PCR also grew on LB plates containing 25 μg/mL tetracycline, but not on plates containing 20 μg/mL chloramphenicol, indicating that they were not harboring the pTKS/CS vector, but instead had the tetracycline resistance gene integrated onto their genome.

Integration of the Fluorescent Reporter Protein mCherry into Various Genomic Loci. A single copy of the gene encoding mCherry, under the control of the T7-lac promoter, was integrated into the *lacZ* locus such that its expression is driven by the native *lac* promoter. Integration was successful in 100% (n = 5) of the colonies screened by colony PCR across the 3′ integration junction (Figure 3). Similar results were observed when the gene was integrated into the *atpI-gidB*, *recA*, and *ybbD-ylbG* loci (data not shown).

Expression of Chromosomally Integrated mCherry. The mCherry expression levels of the *lacZ*, *atpI-gidB*, *recA*, and *ybbD* mCherry-integrants were compared to a high-copy plasmid pETM6-based positive control utilizing identical T7-lac promoters (Figure 4). The level of fluorescence observed was found to be dependent on the genomic locus that the gene is integrated into. When the fluorescence was normalized by the cell growth (OD_{650nm}), the highest mCherry expression was observed when the gene was integrated into the intergenic *atpI-gidB* locus, about 4-fold higher than the expression of the same gene from the high-copy plasmid. Elevated expression of mCherry was also observed when the gene was integrated into both the *recA* and *ybbD-ylbG* loci, with both showing an approximately 2-fold increase in fluorescence over the plasmid control. However, integration of mCherry into the *lacZ* locus resulted in very little expression of mCherry, although the fluorescence measured was higher than that of the noninduced control, indicating that some level of mCherry expression was induced in this strain.

The impact of the integration of mCherry on the growth rate of each strain with and without induction of mCherry expression was also analyzed to determine if induction of expression was detrimental to cell growth (Figure S2). While induction of mCherry expression had no impact on cell growth, the location of the integration did appear to have some impact on cell growth. It was observed that Δ*lacZ*::*mCherry* and Δ*ybbD*::*mCherry* integration strains were the fastest growing

DOI: 10.1021/acssynbio.6b00350
ACS Synth. Biol. 2017, 6, 710−720

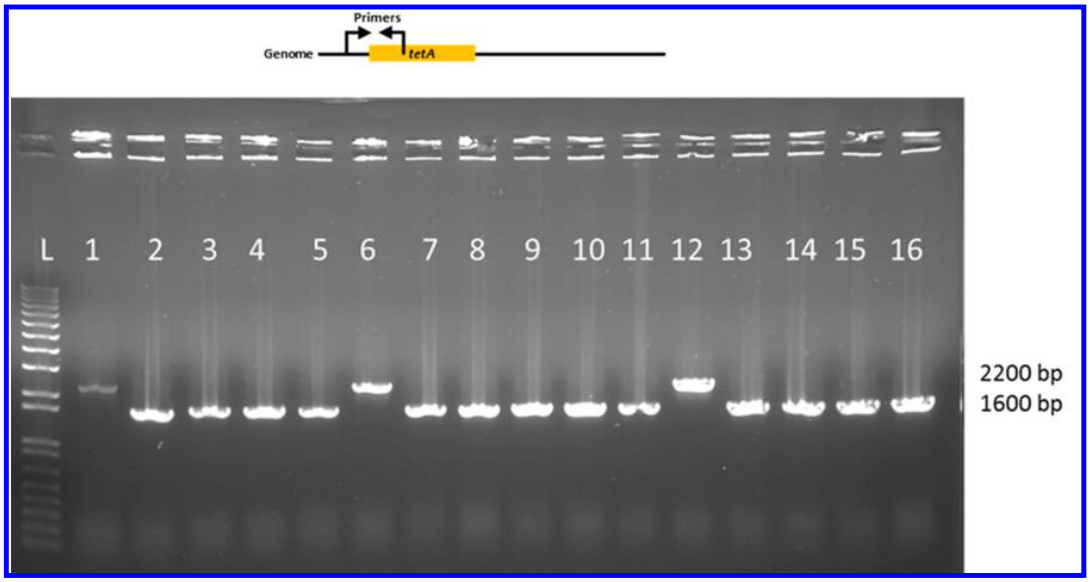


Figure 2. Colony PCR analysis to verify integration of the "landing-pad" into genomic *lacZ* locus. PCR performed with forward primer upstream of *lacZ* and reverse primers inside *lacZ* gene and "landing-pad". Lanes: L ladder, 1 wildtype negative control, 2−16 potential integrants. Positive integration results in 1600 bp amplicon, negative in 2200 bp amplicon.

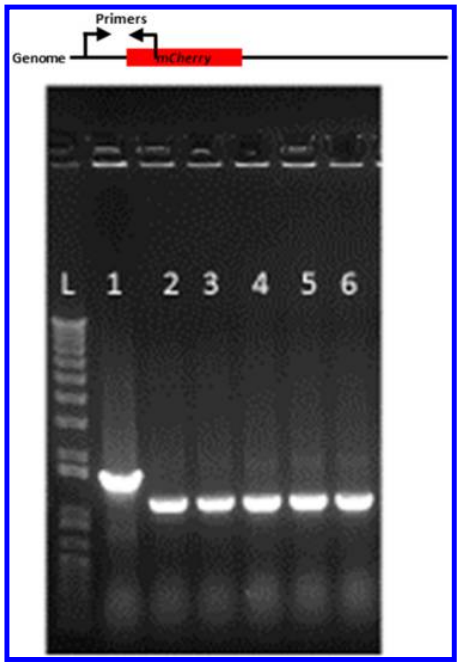


Figure 3. Colony PCR to verify integration of the gene encoding mCherry into the *lacZ* locus. PCR contained forward primer inside kanamycin resistance gene and reverse primer outside of the integration. Lanes: L ladder, 1 landing-pad negative control, 2−6 mCherry integrants.

strains, while the Δ*atpI-gidB*::*mCherry* and Δ*recA*::*mCherry* strains were slower growing.

Multicopy Expression Levels of mCherry. Fluorescence was used to measure mCherry expression of strains containing 1, 4, and 9 copies of the gene in a pseudo-operon configuration in the *atpI-gidB* locus (Figure 5). Interestingly, the amount of fluorescence negatively correlated with the copy number of the gene on the genome. Integrating four copies of the gene resulted in an approximately 8-fold decrease in mCherry expression compared to the strain containing one copy of the gene. The strain containing 9 copies of the gene resulted in an even further decrease in expression, approximately 75-fold. Similar results were also observed when the multicopy constructs were expressed from a plasmid (data not shown).

Genomic Integration of the Violacein Pathway. Integration of the five-gene pathway for the conversion of tryptophan to violacein into 4 chosen genomic loci was verified by colony PCR across the 5′ integration junction (Figure S3). The results of this analysis show that, for all four integration sites, 100% (n = 8) of colonies were positive for the integration of the violacein pathway. However, multiple amplicons were observed for the *lacZ* integration. One band present matches amplification of a proper integration event, while the other does not match the amplification of a negative integration. Instead, the smaller amplicon is a result of recombination between the genomic and plasmid copy of *lacO*. This hypothesis was verified using Sanger sequencing (IDT, Inc.). For further studies of this strain, a colony showing proper amplification was utilized. In this instance, 37.5% (n = 8) colonies showed recombination between the *lacO* sequences.

Colony PCR was used to assess the possibility that homologous recombination had occurred between the numerous regions of identical DNA sequences present in the integrated DNA, including the promoter and terminator for each gene (Figure S4). The analysis shows that the correct amplicons were observed between all of the genes in all 8 colonies, indicating that there were no recombination events occurring. The same result was observed for the other three genomic loci.

Vector Construction and Genomic Integration of the *TAL* Pathway. Colony PCR was used to verify the integration of *TAL* into the *lacZ*, *atpI-gidB*, *recA*, and *ybbD-ylbG* loci of *E. coli* MG1655(DE3). Colony PCR was used to amplify the junction at the 5′ end of the integrated sequence in 8 colonies for all four genomic sites (Figure S5). The gene encoding TAL

DOI: 10.1021/acssynbio.6b00350
ACS Synth. Biol. 2017, 6, 710−720

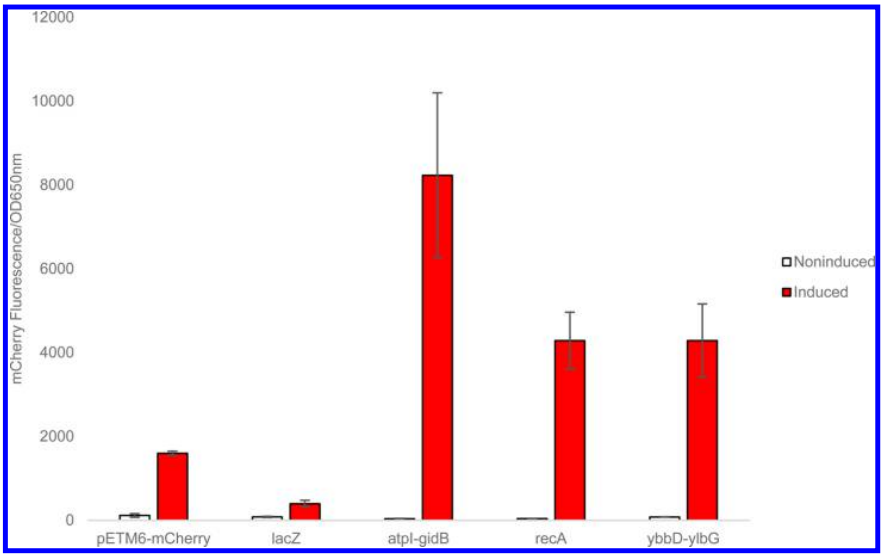


Figure 4. Expression of mCherry when integrated into four genomic loci. Positive control is mCherry expressed from the high-copy plasmid pETM6. Error bars are standard deviation of three biological replicates.

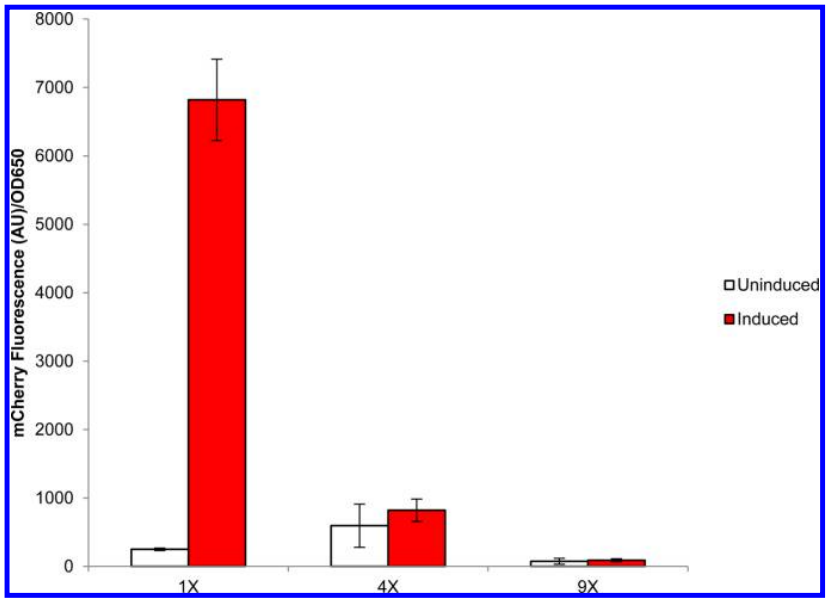


Figure 5. mCherry expression of strains containing 1, 4, and 9 copies of the gene encoding mCherry. End-point fluorescence was measured with and without addition of IPTG to induce mCherry expression. Error bars indicate standard deviation of biological triplicate.

was integrated at a high efficiency in all four genomic loci. Integration into the *lacZ* locus resulted in 25% (n = 8) of colonies in which the genomic and plasmid-based copies of *lacO* recombined.

Violacein Production. A preliminary screen of violacein production was conducted on LB plates containing IPTG to induce the expression of the genes required for violacein production. To this end, 8 colonies were streaked for each strain carrying the pathway in the different genomic loci. This screen showed that both the Δ*atpI-gidB*::*vioABECD* and Δ*ybbD*::*vioABECD* strains produced no colored colonies, indicating that these strains were not producing violacein or any of the other colored side products from the violacein pathway. Interestingly, the Δ*lacZ*::*vioABECD* and Δ*recA*::*vioABECD* strains showed mixed results. The Δ*recA*::*vioABECD* strain was the most successful strain according to this plate-based assay, as 87.5% (n = 8) of colonies screened produced a purple pigment, indicating that these strains were capable of producing violacein. However, one of these strains (#8) produced a mixture of purple and green, indicating that the strain was producing both violacein and one of the other side products of the pathway. Integration into the other three genomic loci resulted in no violacein production.

These strains were then tested for their ability to produce violacein in liquid culture. Initial attempts to produce violacein in the rich defined media AMM[34] were unsuccessful for all of the strains. The strains were tested for their ability to produce violacein in LB broth. The only one that produced any colored

DOI: 10.1021/acssynbio.6b00350
ACS Synth. Biol. 2017, 6, 710−720

compound was the Δ*recA*::*vioABECD* strain. According to HPLC analysis, 87.5% (n = 8) of the Δ*recA*::*vioABECD* colonies produced some quantifiable amount of violacein, while colony #1 produced no violacein (Figure 6). Of the colonies

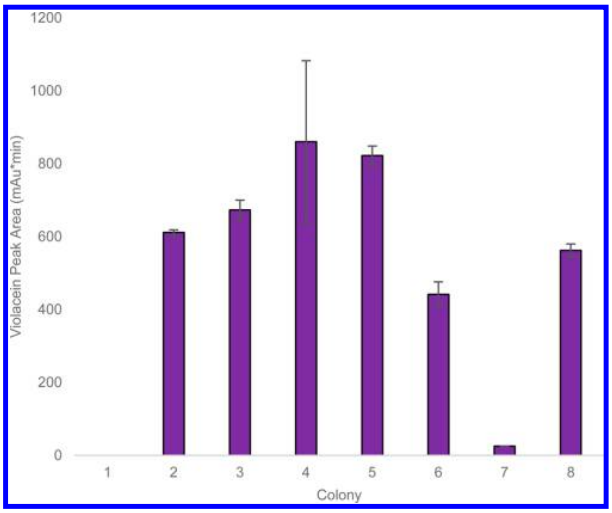


Figure 6. Violacein production from MG1655(DE3) Δ*recA*::*vioABECD*. Eight colonies that passed both colony PCR and antibiotic screening were tested for ability to produce violacein. Violacein was measured by HPLC peak area.

producing violacein, 5/7 produced similar, but statistically different (one way ANOVA, p = 0.03) amounts of violacein, while colony #7 produced very little violacein; colony #8 actually produced a smaller amount of violacein and also produced a larger amount of a green compound (Figure S6). DNA sequencing analysis did not reveal any differences between these strains. Interestingly, the peak that was initially thought to be violacein in strain #8 shows a retention time shift that correlates with the product of a strain containing the incomplete *vioABE* pathway, indicating that this peak is not violacein. As can be seen on the HPLC chromatogram (Figure S6), all of the strains that are making violacein are also making other side products from the violacein pathway in small amounts.

Production of *trans*-Cinnamic Acid from Cells Containing an Integrated Copy of *TAL*. The impact on *trans*-cinnamic acid production from the insertion of the tetracycline-resistant "landing-pad" into all four of the chosen genomic loci was analyzed by HPLC (Figure S7). Surprisingly, integration of the "landing-pad" into all of these loci resulted in a decrease in *trans*-cinnamic acid production from plasmid. Interestingly, a 14.4% decrease in *trans*-cinnamic acid production was observed between the highest and lowest producing strains in this analysis. Integration into the *atpI-gidB* intergenic locus had no impact on *trans*-cinnamic acid production, followed by *lacZ* (4.6% decrease), *recA* (11.1% decrease), and *ybbD* (14.4% decrease) compared to the control when TAL was induced after 4 h of growth.

After seeing multiple phenotypes present when expressing the violacein pathway after integrating the genes onto the genome, it was important to test the TAL pathway integrants for varying phenotypes. Four colonies screened for *trans*-cinnamic acid production when the gene was integrated into the *ybbD* locus showed nearly identical levels of production when grown in media supplemented with phenylalanine (Figure S8). The issue of *lacO* recombination at the *lacZ* locus described earlier was again observed (Figure S9). Of the 7 colonies that passed the antibiotic screening for proper integration of the TAL gene into the *lacZ* locus on the genome, five produced equal amounts of *trans*-cinnamic acid. The other two colonies produced equal, but lower amounts of *trans*-cinnamic acid. When compared to colony PCR analysis of these strains (Figure S5), it was determined that in these two strains, recombination had occurred between the *lacO* sequences present on the genome and integration vector, resulting in lower *trans*-cinnamic acid production.

Next, the impact of both the IPTG inducer concentration and the timing of the induction on *trans*-cinnamic acid production were evaluated. The timing of induction of TAL expression has a major impact on *trans*-cinnamic acid production in *E. coli* MG1655(DE3) Δ*atpI-gidB*::TAL (Figure S10). Similar results were obtained for TAL expressed in the other 3 integration strains. Inducing expression of TAL both too early and too late in the culture's growth negatively impacted the culture's ability to produce *trans*-cinnamic acid from phenylalanine. In contrast, IPTG inducer concentration was shown to have little-to-no impact on the strains' ability to produce *trans*-cinnamic acid. At each induction time point, a similar amount of *trans*-cinnamic acid was produced regardless of whether TAL expression was induced with 0.1 mM or 2 mM IPTG.

trans-Cinnamic acid production of strains harboring the *TAL* gene integrated into the *lacZ*, *atpI-gidB*, *recA*, and *ybbD* loci was also compared to a strain containing TAL on a high-copy plasmid (Figure 7). Overall, the strains harboring genomic copies of TAL produced approximately 50% less *trans*-cinnamic acid than the plasmid-based strain. There was little difference in *trans*-cinnamic acid production between the integration strains (one-way ANOVA, p = 0.02). The difference between the

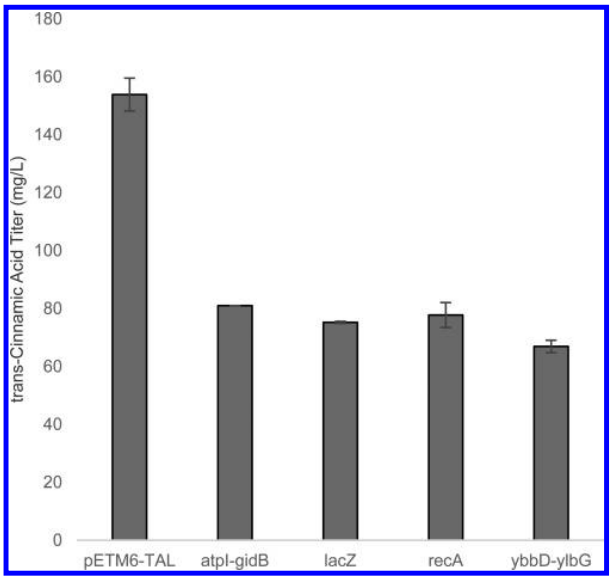


Figure 7. Comparison of *trans*-cinnamic acid production between strains containing the *TAL* gene integrated into different genomic loci and plasmid-based expression. Cells were induced with 2 mM IPTG after 2 h of growth and allowed to grow for 24 h. Error bars are standard deviation of biological duplicates.

DOI: 10.1021/acssynbio.6b00350
ACS Synth. Biol. 2017, 6, 710−720

highest producing strain (*atpI-gidB*) and lowest producing strain (*ybbD-ylbG*) is approximately 17%.

DISCUSSION

Here we have shown that, when integrated into the genome of *E. coli*, the expression of the fluorescent protein mCherry, under the control of the T7 promoter and *lacO* operator, is dependent upon the location in the genome. When integrated into four genomic loci, three distinct levels of expression were obtained. Integration into the intergenic *atpI-gidB* locus resulted in the highest level of expression, which was measured to be over 4-fold higher than expression of the gene from a high-copy plasmid. The same gene integrated into either the *recA* or *ybbD-ylbG* loci resulted in a "medium" level of expression, which showed just over 2-fold higher mCherry expression than the plasmid-based control. Finally, we observed that there was approximately 8-fold less mCherry expression compared to the plasmid-based control when the gene was integrated into the genomic *lacZ* locus.

We observed unwanted recombination between our integration construct and the genome under certain circumstances. That is, the *lacO* DNA sequence found on the integration vector constructs would recombine with the same sequence found on the genome upstream of the *lacZ* gene. This situation only occurred when we were trying to integrate into the *lacZ* locus, and occurred in approximately 25−33% of the integrants. The genomic *lacO* sequence is 740 bp upstream of the I-SceI restriction site that is introduced when the tetracycline-resistant "landing pad" is integrated into the *lacZ* locus. Normally, the double-strand break produced here is required for the integration of constructs larger than approximately 2000 bp, as λ-Red is capable of repairing double stranded breaks with homologous recombination.[8] This recombination indicates that homologous DNA as close as 740 bp away from the double-stranded break can be used to repair the damaged DNA. When this specific recombination occurs, the T7 promoter, which is supposed to control the overexpression of the gene, is not integrated. As a result, reduced protein expression would be expected from cells in which this recombination occurred.

However, this recombination event does not explain the lower mCherry expression when the gene is integrated into the *lacZ* locus. Colony PCR and sequencing analysis were used to confirm that the *lacZ* strains used for protein expression were proper integrants with intact T7 promoter sequences. Thus, there must be another mechanism causing the stunted level of protein expression when mCherry is integrated into the *lacZ* locus. Interestingly, a recent study showed that, under certain conditions, protein expression under *lacO* control is more tightly regulated when the gene is located in close proximity to *lacI*, which encodes the lac repressor.[17] This could offer an explanation for the lower expression observed, though it would obviously require further investigation. However, this explanation does not adequately address why mCherry expression was higher when the gene was integrated into the *atp-gidB* locus than the *recA* or *ybbD* loci. None of these loci are close enough to the *lacI* gene to be influenced as the *lacZ* locus was (Figure 1). Importantly, we have shown the capability of genomic integration to express higher quantities of protein than high-copy plasmid-based expression systems.

We hypothesized that in order to obtain a similar level of protein expression as seen in plasmid-based expression systems, an increased copy number present on the genome would be required. This was thought to be the case because when we integrate the mCherry gene onto the genome, there is a single copy per cell, as opposed to the high-copy pETM6-mCherry strain where there is upward of 40 copies of the gene per cell.[3] Interestingly, we have shown that mCherry expression level negatively correlated with copy number in the pseudo-operon configuration. There is no known mechanism to explain this observation, though increased metabolic burden is a likely cause. Our results indicate that increased copy number was not useful for increased expression level of genes integrated onto the chromosome.

Next, we have shown the ability to combine the ePathBrick system for synthetic pathway construction with the pTKRED system for genomic integration.[3,8] Together, these two methods offer nearly endless possibilities for further studies on the metabolic engineering of novel pathways on the genome of *E. coli*.

We have successfully integrated the five-gene, 8 kb pathway for the conversion of tryptophan to violacein into four genomic loci. Again, violacein production was dependent on the location of the integration. Interestingly, the ability to produce violacein did not correlate with mCherry expression levels observed earlier. Instead, integration of the pathway into the *atpI-gidB* locus, which saw the highest mCherry expression, resulted in no violacein production. Integration into either the *lacZ* or *ybbD-ylbG* loci did not result in any violacein production. Surprisingly, integration into the *lacZ* locus resulted in the production of a green compound, most likely prodeoxyviolacein.[35] Prodeoxyviolacein is a side-product of the violacein pathway, where VioA, VioB, and VioE convert tryptophan to a reduced intermediate, which is then nonenzymatically converted to prodeoxyviolacein. Normally, VioD would convert the intermediate to protoviolaceinic acid, which can either be utilized to produce proviolacein, or converted to violacein by VioC. Production of prodeoxyviolacein from the *lacZ* integrants would, thus, indicate a lack of *vioD* expression.

We hypothesize that this lack of expression could be the result of unwanted homologous recombination occurring inside of the pathway. This might be due to the presence of many regions of identical DNA in the pathway. In the case of the violacein pathway, all 5 genes are controlled by identical copies of the T7 promoter, *lacO*, ribosome-binding site, and terminator. In total, there is approximately 1600 bp of identical DNA between each of the genes. One possible explanation could be that when the λ-Red recombinase is overexpressed, recombination events could occur between any of these identical regions. However, we have shown with PCR and sequencing analysis that these events do not occur. No other mechanism for production of this green compound in this strain is readily apparent.

Surprisingly, violacein was only produced when the pathway was integrated into one of the four genomic loci, the *recA* site. There is no apparent reason that this locus would be better for violacein production than the other three. In terms of protein expression, the *recA* integration site showed equal mCherry expression to the *ybbD* locus, which was about half as high as the *atpI-gidB* integrant and roughly 20-fold higher than the *lacZ* integrant. One hypothesis is that instead of the *recA* locus itself being responsible for the increased production level, it could be that a lack of RecA expression is the reason the strain is able to produce violacein. RecA is a protein that is essential for the repair of damaged DNA with homologues found in all known organisms. Even in this study, RecA is essential for utilizing

DOI: 10.1021/acssynbio.6b00350
ACS Synth. Biol. 2017, 6, 710−720

homologous recombination to integrate foreign DNA into the genome, and a copy of the gene is found on the integration vector pTKRED.[8] In homologous recombination, the protein works by binding single-stranded DNA and promoting strand invasion. In molecular biology, *E. coli* strains used for cloning and plasmid propagation are generally deficient in *recA* in order to promote plasmid stability, as RecA has been shown to promote recombination between regions of homology within plasmids.[36] It is unlikely that RecA is responsible for postintegration recombination events in our system, as we have shown with PCR and sequencing analysis that all five genes are still present in the strains. However, RecA could play some other role in disturbing the expression of all of the pathway genes.

Next, we integrated the gene encoding tyrosine ammonia lyase (TAL) onto the genome in four separate genomic loci of *E. coli*. Here, the TAL protein functions to convert phenylalanine into *trans*-cinnamic acid, a metabolite that is easily quantifiable with analytical HPLC. We have shown here that cells harboring the TAL gene on the genome produced significantly less *trans*-cinnamic acid than a strain expressing the gene from a plasmid. The most obvious explanation for this result is that the plasmid-based strains are expressing TAL at a more optimal level than the genome-based strains. In our optimization studies, we determined that growth phase at the time of induction has the largest effect on the cells' ability to produce *trans*-cinnamic acid. We also found that inducing with different IPTG concentrations, at least between 0.1 mM and 2 mM, had little-to-no impact on the *trans*-cinnamic acid production by the strains. We also observed that replacing the *lacZ*, *atpI-gidB*, *recA*, and *ybbD* loci on the genome with the tetracycline "landing-pad" had an impact on trans-cinnamic acid production when TAL was expressed on a high-copy plasmid. These results indicate that deleting genes, even if they are not related to cellular metabolism, alters the overall metabolic state of the cell enough to impact production of metabolites. We showed here approximately a 20% difference between these strains, indicating that deleting some genes plays a larger role than others.

Importantly, we have shown that integrating large DNA sequences with many regions of identical DNA results in an issue of reproducibility between strains. Here, we saw that when integrating the violacein pathway into the *recA* locus, most of the strains that were tested produced violacein. However, two of the strains made little-to-no violacein, and another strain produced a green side-product instead of violacein. We hypothesized that this was a result of recombination events occurring between any of the identical regions of DNA between the genes in the pathway, but PCR and sequencing analysis determined that the sequences were identical.

CONCLUSIONS

Overall, our results demonstrate that chromosomal integration may be a promising avenue for protein expression and metabolic engineering. We have shown that, at least for some proteins, integration onto the genome may result in higher levels of protein expression. Further, we have shown the ability to integrate and express five-gene, and one-gene heterologous pathways on the genome of *E. coli*. Importantly, we have shown that the location of these integrations plays a role in the protein expression and metabolite production of the strains. Interestingly, the ability to produce high amounts of protein does not correlate with the ability to produce products of heterologous pathways. While with the violacein and TAL pathways our chromosomal integrants were unable to match the production levels of the same pathways expressed on high-copy plasmids, other studies have demonstrated the ability of chromosomal integration to surpass production from plasmids with other pathways.[26,38] Our results indicate that when it comes to genomic integration of heterologous genes, not all genomic loci are equivalent.

MATERIALS AND METHODS

Bacterial Strains, DNA Vectors, and Media. All plasmid cloning was performed in the stable strain *E. coli* DH5α. *E. coli* MG1655(DE3) was used for homologous recombination. The ePathBrick plasmid pETM6[3] was utilized to clone integration constructs, including multicopy and multigene vectors. For later experiments, pTKS/CS[8] was used for the amplification of the tetracycline resistance "landing-pad". The integration vectors pTKIP-neo and pTKDP-neo[14] were used for homologous recombination, and pTKRED was used for expression of λ-Red recombinase and the restriction enzyme I-SceI. All cultures were grown in LB broth (Sigma) for general cloning and expression studies. For integration experiments, cultures were grown in MOPS EZ Rich Defined Medium (Teknova) supplemented with 0.5% (v/v) glycerol, referred from here on as RDM.

Construction of ePathBrick Vectors Containing Multiple Copies of Genes. The ePathBrick plasmid pETM6 was utilized to build vectors containing between one and 10 copies the gene encoding mCherry in a monocistronic configuration. The gene encoding mCherry was previously cloned into pETM6[3]. This vector containing one copy of mCherry (1X), was digested in two separate reactions to build the 2X constructs. In one reaction, to generate a backbone fragment containing one copy of the gene, the vectors were digested with *SpeI* and *Sal*I. In the other reaction, to generate the inset fragment containing another copy of the gene, the vector was digested with *AvrII* and *Sal*I. These digestions were then run on 0.8% agarose gel and the relevant DNA fragment was extracted from the gel (MicroElute Gel Extraction Kit, Omega) and then ligated together. This 2X construct was then used in conjunction with the 1X construct to build both the 3X and 4X constructs, which were subsequently used to build the 5X, 6X, 7X, and 8X constructs. Finally, the 8X construct was used to build the 9X and 10X constructs. In situations where multiple combinations could be used to build a new construct, the constructs were built using the highest previously built construct as the inserted DNA.

Construction of Integration Vectors. All DNA constructs to be integrated had to be cloned into the integration vectors pTKIP-neo or pTKDP-neo to achieve genomic integration.[8,14] Initially, there was no commonality between the integration vectors and the pETM6 constructs that would allow for convenient subcloning. The gene encoding mCherry was amplified from the pETM6-mCherry vector using primers 1 and 2 (Table S1) and cloned into pTKIP-neo and pTKDP-neo between the *ApaI* and *Sal*I restriction sites, generating pTKIP-mCherry and pTKDP-mCherry. This introduces the *AvrII* and *Sal*I restriction sites into the integration vectors, allowing for subcloning from any ePathBrick construct. Next, all ePathBrick constructs were digested with *AvrII* and *Sal*I and subcloned into the multiple cloning sites of the integration vectors pTKIP or PTKDP.

DOI: 10.1021/acssynbio.6b00350
ACS Synth. Biol. 2017, 6, 710−720

Integration of Large Constructs into the *lacZ*, *atpI-gidB*, *recA*, and *ybbD* Genomic Loci on the MG1655(DE3) Genome. Homologous recombination of large constructs (over 2.5 kb) was achieved a modified version of the earlier described protocol for homologous recombination.[8,14] Here, the plasmid pTKRED encodes both λ-Red recombinase and a yeast restriction enzyme I-SceI. The genomic locus of interest is first replaced with a small "landing-pad" that contains a tetracycline resistance gene flanked by a novel 25 bp sequence of DNA termed a landing-pad as well as the recognition site for I-SceI to facilitate the integration of large constructs. A third plasmid in this system, pTKS/CS, contains this tetracycline "landing-pad". For this study, four loci spread approximately equally throughout the genome were chosen as target locations for integration of exogenous DNA: *lacZ*, *atpI-gidB*, *recA*, and *ybbD*.

The strain of choice is first transformed with pTKRED and plated onto LB agar containing 100 μg/mL spectinomycin at 30 °C to integrate this small "landing-pad". An overnight seed culture of these cells was then diluted 1:50 to inoculate a new culture in 20 mL of LB supplemented with 100 μg/mL spectinomycin and 2 mM IPTG at 30 °C to produce electro-competent cells expressing λ-Red recombinase. The "landing-pad" was amplified using PCR from pTKS/CS utilizing primers 3−10 (Table S1) that contains at least 40 bp on both ends that is identical to one of four genomic locations chosen for this study to be replaced. After restriction digestion with *DpnI* for 2 h, 10 μL of this linear fragment is transformed into 100 μL of electro-competent cells in a 0.2 cm cuvette (VWR). The mixture was shocked (GenePulser Xcell, BioRad) at 2.5 kV, 25 μF, 200 Ω, immediately resuspended in 1 mL of ice-cold SOC, incubated in a shaking incubator at 30 °C for 3 h, and then 300 μL was plated onto LB plates containing 100 μg/mL spectinomycin and 25 μg/mL tetracycline and incubated at 30 °C. Colonies that grew on the plate were then transferred onto a new LB plate containing spectinomycin and tetracycline, and also an LB plate containing 25 μg/mL chloramphenicol to screen for chloramphenicol resistance from residual pTKS/CS. Colonies that grew on the plate containing spectinomycin and tetracycline, but not chloramphenicol, were then screened using colony PCR for proper integration of the "landing-pad".

Next, electro-competent cells were created from the strain containing the tetracycline "landing-pad" on the genome. Cells (100 μL) were electroporated with 10 μL of the desired integration construct, which were purified from an overnight culture (Plasmid DNA Mini Kit, Omega), and 300 μL of the cells were plated onto LB plates containing 100 μg spectinomycin, 25 μg tetracycline, and 50 μg kanamycin and incubated at 30 °C overnight. A small patch of colonies from these transformation plates were then used to inoculate 5 mL of RDM supplemented with 0.5% glycerol, 100 μg/mL spectinomycin, 50 μg/mL kanamycin, 2 mM IPTG, and 0.2% (w/v) arabinose. These cultures were incubated at 30 °C for approximately 24 h, or until the culture was saturated with bacterial growth, determined by turbidity. From here, the cultures were diluted, based on the measured optical density, such that roughly 100 CFU were plated onto LB plates containing 50 μg/mL kanamycin.

Usually, counter selection was required to obtain colonies with successful integration.[14] Here, 100 μL of the saturated culture was used to inoculate 5 mL of RDM supplemented with 0.5% glycerol, 50 μg/mL kanamycin, 6 mM $NiCl_2$, and 5% (w/v) sucrose at 37 °C. Growth of cells expressing the tetracycline resistance marker TetA is inhibited by $NiCl_2$ and cells expressing the *sacB* gene present on pTKDP are inhibited by sucrose. As a result, only cells containing the kanamycin resistance marker from the integration event are able to grow in the counter-selection media. After the cultures were saturated with bacterial growth, they were diluted with sterile water and 250 μL was spread on LB plates containing 50 μg/mL kanamycin and incubated at 37 °C.

Screening Integration Colonies. Colonies were patched onto three LB plates containing either 50 μg/mL kanamycin, 25 μg/mL tetracycline, or 100 μg/mL ampicillin to ensure that any expression observed by modified strains was a result of genomic integration, and not residual plasmid copies of the genes. Colonies that grew on only the plate containing kanamycin were then subsequently screened using colony PCR.

Colony PCR was performed to verify integration of the desired DNA into the target genomic locus by amplifying DNA across the junction of the host's genomic DNA and the newly integrated exogenous DNA (GoTaq Hot Start Master Mix, Promega). Briefly, individual colonies were suspended in 10 μL of sterile water, and then this suspension was used as the template DNA in a PCR reaction with an extended initial denaturing step at 95 °C to facilitate cell lysis to free genomic DNA. In each reaction, three primers were present: one from the genomic DNA, one from the "landing-pad," and one from the newly integrated DNA, such that a positive and negative integration event results in amplicons of different sizes. Primers 11−36 (Table S1) were used for PCR analysis, depending on what was being screened. Generally, colony-PCR was used at both junctions of the integrated DNA to ensure the exact integration was performed.

mCherry Expression Measurements. Strains expressing the gene encoding mCherry, under control of the T7-lac promoter were incubated overnight at 37 °C in 2 mL of LB broth (Sigma) supplemented with appropriate antibiotics. These seed cultures were then used to inoculate fresh 2 mL LB cultures in a 1:50 ratio, which were grown at 37 °C. After 2 h of growth, expression of mCherry was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The expression of mCherry was measured on a plate reader (Synergy 4, Biotek) using fluorescence. Cell culture (200 μL) was loaded into a black-walled, clear bottom 96-well plate (Greiner Bio One, Polystyrene) and fluorescence was measured with an excitation wavelength of 588 nm and emissions were measured at 618 nm. Optical densities were calculated based on OD_{650nm}[37] after cultures were diluted down into the linear range of OD_{650nm} for the instrument, and then fluorescence and OD_{650nm} were measured from the same well. Fluorescence and cell density measurements were taken approximately hourly.

Construction of Violacein and *TAL* Pathway Integration Vectors. The genes for the five-gene pathway for the conversion of tryptophan into the purple pigment violacein were previously cloned into the ePathBrick vector pETM6.[4] Site-directed mutagenesis had to be performed on *vioB* and *vioC* to remove natural *Sal*I restriction sites to clone the entire pathway into the integration vectors. To this end, primers 37−40 (Table S1) were designed to create silent, single nucleotide mutations in the *Sal*I cleavage site in both *vioB* and *vioC*, and site directed mutagenesis was achieved by following standard protocols, followed by verification using Sanger sequencing (Genewiz, Inc.). The resulting mutants were termed *vioB** and *vioC** to distinguish them from the wildtype sequences. Construction of the 5-gene pathway, including *vioB** and

DOI: 10.1021/acssynbio.6b00350
ACS Synth. Biol. 2017, 6, 710−720

*vioC**, was first done in pETM6 and then subcloned into pTKIP-neo and pTKDP-neo using *AvrII* and *Sal*I restriction sites. Each of these pETM6 constructs was restriction digested with *AvrII* and *Sal*I and the band corresponding to the gene was extracted and purified from agarose gel. Next, pETM6-vioA and pETM6-vioC* were digested with *SpeI* and *Sal*I, and ligated with the previously digested *vioB** and *vioD*, respectively, creating both pETM6-m-vioA-m-vioB and pETM6-m-vioC-m-vioD. In the same fashion, *vioE* was cloned downstream for *vioB* to create pETM6-m-vioA-m-vioB-m-vioE. Finally, pETM6-m-vioA-m-vioB-m-vioE that was restriction digested with *SpeI* and *Sal*I, and pETM6-m-vioC-m-vioD that had been digested with *AvrII* and *Sal*I were ligated together to generate pETM6-m-vioA-m-vioB*-m-vioE-m-vioC*-m-vioD. The final construct was verified by restriction digestion. This vector, containing the entire violacein pathway, was then digested with *AvrII* and *Sal*I and subcloned into pTKDP-mCherry that was digested with the same enzymes, resulting in the integration vector pTKDP-m-vioA-m-vioB*-m-vioE-m-vioC*-m-vioD.

The gene encoding tyrosine ammonia lyase (TAL) had previously been cloned into the ePathBrick vector pETM6. pETM6-TAL^{syn} and the integration vector pTKDP-mCherry were both digested with *AvrII* and *Sal*I and the bands corresponding to the gene and the plasmid backbone were extracted and purified from agarose gel. These two fragments were ligated together and transformed into *E. coli* to produce pTKDP-TAL^{syn}, which was verified by restriction digestion.

Sequence Verification of the Integrated Violacein Pathway. Multiple PCRs were run to amplify most of the pathway to determine if recombination events were taking place between the genes in the violacein pathway after integration into the genome. Primers 25−34 and 41−43 were used to amplify the pathway (Table S1). These same primers were then used to sequence the pathway by Sanger sequencing (Genewiz, Inc.).

Violacein Production. Colonies with genomic copies of the violacein pathway, which had passed both the antibiotic and colony-PCR screening, were patched onto LB plate containing 50 μg kanamycin and 1 mM IPTG to induce the expression of the pathway to screen each colonies ability to produce violacein. Colonies that were able to produce violacein on the plate were purple, and those that could not were white.

Experiments were then performed to measure the optimal liquid growth medium, and IPTG concentration for violacein production in these strains. Violacein production studies were performed as 2 mL cultures in 48 well plates (5 mL, VWR). Briefly, individual colonies were inoculated into 2 mL of a defined rich media (AMM)[4] to create overnight seed cultures in a 48-well plate, and incubated at 30 °C at 225 rpm for 14 h. These cultures were then diluted 50-fold into 2 mL fresh media, and variables were changed as described. Cultures were grown at 37 °C until an hour before induction, when they were transferred to 20 °C, where they stayed after induction. Cultures were allowed to grow for 3 h before they were induced 1 mM IPTG. Cultures were allowed to grow for 18 h postinduction before violacein production was measured.

***trans*-Cinnamic Acid Production.** Production of *trans*-cinnamic acid from phenylalanine was optimized in regards to induction time and IPTG concentration. Individual colonies were inoculated into 2 mL AMM, supplemented with 100 μg/L ampicillin for plasmid-based expression, in a 48-well plate and incubated overnight at 37 °C at 225 rpm. This seed culture was then diluted 50-fold into 2 mL fresh AMM supplemented with 100 mg/L phenylalanine and ampicillin for plasmid-based expression and allowed to grow for varying amounts of time before TAL expression was induced with 0.1 mM or 2 mM IPTG. The cultures were then grown for 24 h at 37 °C before the cultures were processed to analyze *trans*-cinnamic acid production.

HPLC Analysis of Violacein and *trans*-Cinnamic Acid Production. Violacein was measured as previously described.[4] Briefly, Cells were pelleted (20 000*g*, 10 min) and the supernatant was removed. The violacein was then extracted from the pellet by adding twice the original culture volume of pure methanol and boiling in a 95 °C water bath for 5 min or until the pellet appeared completely white. In samples with elevated violacein levels, subsequent extractions were required. The extract was then centrifuged (20 000*g*, 10 min) to pellet cell debris and 10 μL of extract was directly injected into the HPLC.

Violacein analysis was carried out using Agilent 1200 series HPLC with diode array detector (DAD) and ZORBAX SB-C18 StableBond analytical column (150 mm × 5 mm, 5 μm) maintained at 30 °C. The mobile phases were acetonitrile (A) and water (B), both containing 0.1% formic acid. The following gradient was used at a flow rate of 1 mL/min: 0 min, 5% A; 1 min, 5% A; 5 min, 45% A; 7 min, 55% A; 9 min, 95% A; 10 min, 5% A; 12 min, 5% A. Violacein (7.95 min) and deoxyviolacein (9.11 min) were analyzed by peak area integration at 565 nm.

trans-Cinnamic acid production was measured using HPLC as previously described.[38] Briefly, the culture was mixed with equal volume of absolute ethanol, briefly vortexed, and centrifuged (20 000*g*, 10 min) to remove cell debris. The supernatant (10 μL) was then used for analysis. Cinnamic acid analysis was carried out using Agilent 1200 series HPLC equipped with a ZORBAX SB-18 column (150 mm × 5 mm, 5 μm) and a diode array detector. The mobile phase was acetonitrile (solvent A) and water (solvent B) (both contain 0.1% formic acid) at a flow rate of 1 mL/min. HPLC program was as follows: 10−40% A (0−10 min) and 40−60% A (10−15 min). Absorbance at 280 nm was monitored. The titer of cinnamic acid (12.0 min) was determined using authentic standard purchased from Sigma-Aldrich (St. Louis, MO).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.6b00350.

Table S1; Figures S1−S10 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (518) 276-2220. Fax: (518) 276-3405. E-mail: koffam@rpi.edu.

ORCID

Brady F. Cress: 0000-0002-2948-2846

Author Contributions

J.A.E., R.J.L., and M.A.G.K. designed the project and wrote the manuscript. J.A.E., J.A.J., and B.F.C. performed experiments. T.E.K. provided technical assistance.

Notes

The authors declare no competing financial interest.

DOI: 10.1021/acssynbio.6b00350
ACS Synth. Biol. 2017, 6, 710−720

■ ACKNOWLEDGMENTS

Partial support for this work was provided by NSF MCB-1448657 to M.K. and R.L. T.E.K. was supported by the NSF Center for the Physics of Living Cells (PHY 1430124) and the Alfred P. Sloan Foundation Research Fellowship in Physics (FG-2015-6553).

■ REFERENCES

(1) Xu, P., Bhan, N., and Koffas, M. A. G. (2013) Engineering plant metabolism into microbes: From systems biology to synthetic biology. *Curr. Opin. Biotechnol. 24*, 291−299.
(2) Shetty, R. P., Endy, D., and Knight, T. F. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng. 2*, 5.
(3) Xu, P., Vansiri, A., Bhan, N., and Koffas, M. a G. (2012) ePathBrick: A Synthetic Biology Platform for Engineering Metabolic Pathways in E. coli. *ACS Synth. Biol. 1*, 256−66.
(4) Jones, J. A., Vernacchio, V. R., Lachance, D. M., Lebovich, M., Fu, L., Shirke, A. N., Schultz, V. L., Cress, B., Linhardt, R. J., and Koffas, M. A. G. (2015) ePathOptimize: A Combinatorial Approach for Transcriptional Balancing of Metabolic Pathways. *Sci. Rep. 5*, 11301.
(5) Smanski, M. J., Bhatia, S., Zhao, D., Park, Y., B A Woodruff, L., Giannoukos, G., Ciulla, D., Busby, M., Calderon, J., Nicol, R., Gordon, D. B., Densmore, D., and Voigt, C. A. (2014) Functional optimization of gene clusters by combinatorial design and assembly. *Nat. Biotechnol. 32*, 1241−1249.
(6) Na, D., Kim, T. Y., and Lee, S. Y. (2010) Construction and optimization of synthetic pathways in metabolic engineering. *Curr. Opin. Microbiol. 13*, 363−370.
(7) Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res. 40*, 8773−8781.
(8) Kuhlman, T. E., and Cox, E. C. (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res. 38*, e92.
(9) Ganusov, V. V., and Brilkov, A. V. (2002) Estimating the instability parameters of plasmid-bearing cells. I. Chemostat culture. *J. Theor. Biol. 219*, 193−205.
(10) Godwin, D., and Slater, J. H. (1979) The influence of the growth environment on the stability of a drug resistance plasmid in Escherichia coli K12. *J. Gen. Microbiol. 111*, 201−210.
(11) French, G. L. (2010) The continuing crisis in antibiotic resistance. *Int. J. Antimicrob. Agents 36*, S3−S7.
(12) Gould, I. M. (2010) Coping with antibiotic resistance: The impending crisis. *Int. J. Antimicrob. Agents 36*, S1−S2.
(13) Bush, K., Courvalin, P., Dantas, G., Davies, J., Eisenstein, B., Huovinen, P., Jacoby, G. A., Kishony, R., Kreiswirth, B. N., Kutter, E., Lerner, S. A., Levy, S., Lewis, K., Lomovskaya, O., Miller, J. H., Mobashery, S., Piddock, L. J., Projan, S., Thomas, C. M., Tomasz, A., Tulkens, P. M., Walsh, T. R., Watson, J. D., Witkowski, J., Witte, W., Wright, G., Yeh, P., and Zgurskaya, H. I. (2011) Tackling antibiotic resistance. *Nat. Rev. Microbiol. 9*, 894−896.
(14) Tas, H., Nguyen, C. T., Patel, R., Kim, N. H., and Kuhlman, T. E. (2015) An integrated system for precise genome modification in Escherichia coli. *PLoS One 10*, 1−19.
(15) Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A. 97*, 6640−6645.
(16) Sabri, S., Steen, J. A., Bongers, M., Nielsen, L. K., and Vickers, C. E. (2013) Knock-in/Knock-out (KIKO) vectors for rapid integration of large DNA sequences, including whole metabolic pathways, onto the Escherichia coli chromosome at well-characterised loci. *Microb. Cell Fact. 12*, 60.
(17) Kuhlman, T. E., and Cox, E. C. (2012) Gene location and DNA density determine transcription factor distributions in Escherichia coli. *Mol. Syst. Biol. 8*, 610.
(18) Cress, B. F., Trantas, E. A., Ververidis, F., Linhardt, R. J., and Koffas, M. A. G. (2015) Sensitive cells: Enabling tools for static and dynamic control of microbial metabolic pathways. *Curr. Opin. Biotechnol. 36*, 205−214.
(19) Burgard, A. P., Pharkya, P., and Maranas, C. D. (2003) OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. *Biotechnol. Bioeng. 84*, 647−657.
(20) Cress, B. F., Toparlak, O. D., Guleria, S., Lebovich, M., Stieglitz, J. T., Englaender, J. A., Jones, J. A., Linhardt, R. J., and Koffas, M. A. G. (2015) CRISPathBrick: Modular Combinatorial Assembly of Type II-A CRISPR Arrays for dCas9-Mediated Multiplex Transcriptional Repression in E. coli. *ACS Synth. Biol. 4*, 987−1000.
(21) Khodayari, A., Chowdhury, A., and Maranas, C. D. (2014) Succinate Overproduction: A Case Study of Computational Strain Design Using a Comprehensive Escherichia coli Kinetic Model. *Front. Bioeng. Biotechnol. 2*, 76.
(22) Xu, P., Rizzoni, E. A., Sul, S.-Y., and Stephanopoulos, G. (2016) Improving Metabolic Pathway Efficiency by Statistical Model-Based Multivariate Regulatory Metabolic Engineering. *ACS Synth. Biol.*, DOI: 10.1021/acssynbio.6b00187.
(23) Perlova, O., Fu, J., Kuhlmann, S., Krug, D., Stewart, A. F., Zhang, Y., and Muller, R. (2006) Reconstitution of the myxothiazol biosynthetic gene cluster by red/ET recombination and heterologous expression in Myxococcus xanthus. *Appl. Environ. Microbiol. 72*, 7485−7494.
(24) Mairhofer, J., Scharl, T., Marisch, K., Cserjan-Puschmann, M., and Striedner, G. (2013) Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free Escherichia coli expression systems under production conditions. *Appl. Environ. Microbiol. 79*, 3802−3812.
(25) Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., and Koffas, M. A. G. (2016) Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol. 34*, 652−664.
(26) Wang, J., Niyompanich, S., Tai, Y.-S., Wang, J., Bai, W., Mahida, P., Gao, T., and Zhang, K. (2016) Engineering Highly Efficient *E. coli* strain for Mevalonate Fermentation by Chromosomal Integration. *Appl. Environ. Microbiol. 82*, 7176.
(27) Zhao, S., Jones, J. A., Lachance, D. M., Bhan, N., Khalidi, O., Venkataraman, S., Wang, Z., and Koffas, M. A. G. (2015) Improvement of catechin production in Escherichia coli through combinatorial metabolic engineering. *Metab. Eng. 28*, 43−53.
(28) Xu, P., Gu, Q., Wang, W., Wong, L., Bower, A. G. W., Collins, C. H., and Koffas, M. A. G. (2013) Modular optimization of multi-gene pathways for fatty acids production in E. coli. *Nat. Commun. 4*, 1409.
(29) Jones, J. A., Toparlak, O. D., and Koffas, M. A. G. (2015) Metabolic pathway balancing and its role in the production of biofuels and chemicals. *Curr. Opin. Biotechnol. 33*, 52−59.
(30) Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol. 27*, 946−50.
(31) Fu, J., Wenzel, S. C., Perlova, O., Wang, J., Gross, F., Tang, Z., Yin, Y., Stewart, A. F., Muller, R., and Zhang, Y. (2008) Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res. 36*, e113.
(32) Sauer, C., Syvertsson, S., Bohorquez, L. C., Cruz, R., Harwood, C. R., van Rij, T., and Hamoen, L. W. (2016) Effect of Genome Position on Heterologous Gene Expression in Bacillus subtilis: An Unbiased Analysis. *ACS Synth. Biol. 5*, 942.
(33) Bassalo, M. C., Garst, A. D., Halweg-Edwards, A. L., Grau, W. C., Domaille, D. W., Mutalik, V. K., Arkin, A. P., and Gill, R. T. (2016) Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. *ACS Synth. Biol. 5*, 561.
(34) He, W., Fu, L., Li, G., Jones, J. A., Linhardt, R. J., and Koffas, M. (2015) Production of chondroitin in metabolically engineered E. coli. *Metab. Eng. 27*, 92−100.
(35) Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J., and Dueber, J. E. (2013) Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Res. 41*, 10668−10678.
(36) Lovett, S. T., Hurley, R. L., Sutera, V. A., Aubuchon, R. H., and Lebedeva, M. A. (2002) Crossing over between regions of limited

homology in Escherichia coli: RecA-dependent and RecA-independent pathways. *Genetics 160*, 851−859.
(37) Hecht, A. H., Endy, D., Salit, M., and Munson, M. S. (2016) When Wavelengths Collide: Bias in Cell Abundance Measurements due to Expressed Fluorescent Proteins. *ACS Synth. Biol. 5*, 37010.
(38) Jones, J. A., Collins, S. M., Vernacchio, V. R., Lachance, D. M., and Koffas, M. A. G. (2016) Optimization of naringenin and p-coumaric acid hydroxylation using the native E. coli hydroxylase complex, HpaBC. *Biotechnol. Prog. 32*, 21−25.

DOI: 10.1021/acssynbio.6b00350
ACS Synth. Biol. 2017, 6, 710−720