Available on: https://jnsrr.ep-journals.org/ E-ISSN: 3051-3766 P-ISSN: 3051-3758

Journal of Natural Science Research and Review

Volume: 02 Issue: 08 August-2026

Chromatin-Optimized Crispr Activation for Enhanced Pollutant Degradation in Rhodococcus Consortia

Sebastian, Jeremia D.
Department of Laboratory Service, National Institute for Medical Research (NIMR), 4509, Tabora, Tanzania

Akpadolu, Chidinma B
Department of Applied Microbiology and Brewing, Nnamdi Azikiwe University, Awka, 420211, Anambra State, Nigeria

Benin, Sandra
Department of Biological Science, Mississippi State University, Starkville, 39762, Mississippi, United States

Okoye, Rosemary
Department of Microbiology, Federal University Gusau (FUGUSAU), Gusau, 860001, Zamfara State, Nigeria

Owusu, Nana H.
Department of Social Science, Michigan Technological University, Houghton, 49931, Michigan, United States

Published Date: August 03, 2026 **Article DOI: 10.65150/EP-jnsrr/V2E8/2026-01**

ABSTRACT: The system we are introducing allows for dynamic transcriptional targeting through an optimized CRISPR activation (CRISPRa) system that will improve the ability of Rhodococcus consortia to degrade pollutants. The majority of metabolic engineering approaches to date do not take into consideration chromatin silencing as a source of surprise that may result in unexpected transcriptional regulation of metabolic pathways within non-model organisms. The proposed approach combines chromatin accessibility profiling with a conditional variational autoencoder (cVAE) model to identify Nuclease-depleted regions (NDR's) that are chronically available for interaction with dCas9-VPR complexes to activate target downstream GbN for sustained transcriptional activation of the GbN in response to dynamic increases in pollutant concentration. Where stable NDR's proximal to key metabolic enzyme genes are identified by the cVAE as being stable chromatin, an NDR's proximity to a key metabolic enzyme locus can be used to identity likely site(s) for placement of dCas9-VPR complexes for sustained activation of the enzyme system. To prioritize the metabolically active NDR's, a provided Promoter Accessibility Score was calculated to account both local and distal regulatory controls on each NDR to identify the optimum transcriptional initiating site (TIS) for a final targeted dCas9-VPR complex placement. A plasmid-cloned, delivered CRISPRa system was developed to validate pathway function; this system provided dCas9-VPRtmp and gRNA's that were formulated for pollution-specific use, coupled with Codon-Optimized effectors, for use in metabolic pathway activation and evaluation by quantitative measures, including improvement in degradation rates of hydrocarbons as well as poly-aromatic hydrocarbons through a rated time course experiment. Ultimately, the proposed cVAE based CRISPRa system will be able to provide real-time adaptability to changes in the environment, setting the cVAE based CRISPRa apart from static dual-promoter based systems to create a greater degree of Flexibility in the metabolic response of Microbial consortia associated with biodegradation as well as bioremediation applications. This work demonstrates the relationship between Chromatin Biology and Synthetic Metabolism and provides a scalable, rapidly adaptable bioremediation system that can be ascribed with a robust response capability to fluctuations in Environmental input and ongoing target gene product variation, such as are present within non-model organisms over extended time periods, a feature that is critical to achieving desirable levels of genetic stability during bioprocessing for bioremediation and biodegradation.

KEYWORDS: Bioremediation, Crispra, Chromatin, Microbial Consortia, Pollutant Degradation, Synthetic Metabolism

INTRODUCTION

Microbial consortia represent platforms capable of advancing environmental bioremediation through the diverse metabolic capabilities of microorganisms and their resiliency to complex ecosystems (Zhang & Zhang, 2022). Rhodococcus species among all of the microbial consortia systems studied possess significant abilities to metabolically degrade recalcitrant organic pollutants such as aliphatic hydrocarbons and aromatic compounds due to their ability to utilize unique enzyme pathways (Lü et al. 2024). However, traditional methods of engineering or manipulating a gene for overexpression using plasmids or using random mutagenesis results in metabolic burden, genetic instability or unintentional cellular stress responses (Duncker et al. 2021). As a result, there is a need for more precise and stable approaches for engineering new or modifying existing metabolic pathways in microbial consortia. More recently, advances in the area of creating engineered transcriptional regulation through CRISPR

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

technologies have created new avenues for controlling gene transcription through manipulation of expression levels without making permanent genomic modifications through CRISPR activation (CRISPRa) (Cho et al., 2018) however, due to the chromatin architecture in prokaryotic organisms including Rhodococcus, the effectiveness of CRISPRa is limited by the dynamic nature of how chromatin structure alters transcriptional accessibility to response to environmental stress (Grandi et al. 2022). Unlike in eukaryotic cells where extensive studies have been conducted on profiling chromatin structure, the use of profiling to inform synthetic biology and metabolic engineering in bacterial systems has not been fully explored (Ho et al. 2020).

Here we propose a chromatin-guided approach for the design of synthetic promoters using chromatin accessibility information combined with computational analysis through machine learning to develop NDR sequences as stable genetic elements for subsequent implementation in Rhodococcus. Our methodology differs from traditional promoter design methodologies that use only motif-based information to develop potential sequences to direct expression of a gene. Rather, we propose using a conditional variational autoencoder (cVAE) trained on ATAC-seq data to model context-dependent changes in chromatin structure (Xie et al. 2020). Using the cVAE-derived latent variables from the cVAE training set allows us to identify genomic regions that retain a transcriptionally permissive state when subjected to pollutant-induced environmental stress, allowing us to use those sites as targets for activation with CRISPRa-based approaches for the up-regulation of metabolic pathways. This methodology is particularly advantageous over traditional metabolic engineering based on static promoter sequences since the lack of chromatin accessibility may prevent heterologous gene expressions due to compaction of the bacterial genome (Wang et al. 2020). Our system combines two major components to develop a framework that incorporates a dual-component strategy for maintaining robust metabolism in a degrading pathway through sustained metabolic flow from the degradation activity of an engineered microbial population while preventing the potential for creating cellular stress responses.

The first component consists of a computational algorithm that predicts NDRs that can serve as transcriptional targets through the use of the latent cVAE attributes, and the second component consists of a CRISPRa-based maintain-productivity system. The second component is composed of: (1) a set of CRISPRa (CRISPR activation) constructs with specific gRNA sequences that respond to environmental pollutants and (2) codon-optimized enhancers that can be used for maintaining and sustaining gene expression levels. We provide an example of how this type of framework can be applied for targeting the enhancement of both alkB and catA pathways to degrade alkanes and catechols, respectively, in Rhodococcus opacus. Our framework is modular and readily applicable for incorporation into existing engineering strategies for establishing coordinated control over multiple metabolic activities from engineered microorganisms depending on environmental context. Our study has taken a significant step in advancing the understanding of both chromatin biology and synthetic metabolism. Additionally, by incorporating chromatin accessibility as an engineering design factor for the development of engineered metabolic pathways, we have addressed a major challenge faced in synthetic biology by creating a reliable mechanism for preventing the silencing of engineered pathways present in non-model bacterial species. Therefore, our study offers a unique perspective on how to improve the reliability of engineered microbial systems for the degradation of pollutant compounds through an improved understanding of how metabolic pathways function relative to chromatin organization; therefore presenting opportunities for improving precision metabolic engineering practices utilizing naturally occurring bacteria with potential commercial applications.

The remainder of this manuscript will be organized as follows; Section 2 will discuss previous published work in the area of microbial consortia engineering and chromatin-based regulatory systems. Section 3 will provide background information on chromatin architecture in Rhodococcus and its implications on gene regulation by comparing examples of studies that utilize chromatin profiling, whereas Section 4 describes the details of our chromatin-guided approach for developing cVAE-based models in conjunction with CRISPRa to produce engineered genes with predictable transcriptional expression characteristics in Rhodococcus. In Section 5, we will depict the experimental validation of the effectiveness of our approach on promoting enhanced rates of gene expression and pathway stability when applied to engineered aldose reductase and glutenin pathways using Rhodococcus opacus. Section 6 will identify future opportunities for expanding the use of this approach beyond what is currently known in synthetic biology and metabolic engineering studies. Section 7 will summarize key findings that emerged from this investigation.

Related Work

Advancements made recently in synthetic biology are creating new options to develop engineered microbes as tools for cleaning up environmental systems. The nature of how this is done has previously been broken down into three general categories: 1. Conventional Metabolic Engineering, 2. Chromatin-Based Genetic Regulation, 3. Machine Learning-based Synthetic Biology.

Metabolic Engineering in Non-Model Bacteria

Heterologous expression of genes or promoter engineering for the enhancement of pollutant degradation is an approach commonly utilized by traditional biotechnology to increase pollutant degradation via the addition of specific degradative enzyme pathways to microbes. Pseudomonas and Rhodococcus both have been successfully altered using the popular (and standard) Ptac promoter, and this type of work has succeeded in using constitutive expression of metabolic pathways within these organisms (Kim et al., 2020). However, due to the potential for regulatory imbalances, the use of unguided expressions and rapid fluctuations in environmental conditions (Wu et al., 2016) make these approaches problematic for optimal degradation. There have been recent developments attempting to produce an inducible response mechanism that can respond to a pollutant, such as through the use of a benzene/alkane-binding receptor; yet, these types of systems still suffer from limitations due to the need for specific chemicals for activation (Cho et al., 2026).

Chromatin-Based Regulation in Prokaryotes

There is a body of research investigating the relationship between chromatin regulation and gene expression in eukaryotes; however, its use for understanding bacterial gene regulation is becoming more recognized as a valuable tool of analysis. ATAC-seq has been developed to investigate open chromatin regions in E. coli and Bacillus subtilis to exhibit geospatial accessibility variation based upon environmental conditions (Cantin et al., 2024). Very limited research has taken advantage of this knowledge base within synthetic biology. Keung et al. (2015) reported that nucleosome positioning impacts the functionality performance of synthetic promoters within yeast; however no similar studies have been conducted using bacteria. Fontana et al. (2020) also noted that the local chromatin state, in addition to the dCas9 activation mechanism, affects

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

CRISPR activation in Streptomyces. Therefore, researchers should focus on generating synthetic designs that account for chromatin accessibility for improved efficiency/utility of CRISPRa tools.

Machine Learning for Synthetic Biology

The development of powerful deep generative models for designing genetic parts has enabled the design of synthetic promoters in E. coli using VAE's (Wang et al. 2020), and the prediction of regulatory elements in a variety of bacteria using transformer-based architectures (Xia et al. 2026). However, most prior work has focused only on the use of sequence data rather than accounting for the chromatin environment. There are a few recent examples where chromatin context was integrated into the design of genetic components; namely, a study by (DaSilva et al. 2024) that used diffusion models and chromatin accessibility data to create functional regulatory elements in mammalian cells.

The proposed methodology links chromatin profiling with a conditional generative model to predict stable NDRs for CRISPRa targeting. Our method differs from previously describe methods by taking into consideration the impact that toxin stressors have on dynamic changes to chromatin structure, resulting in a reliable mechanism to stimulate pathways that lead to activation of non-model bacterial systems. This work advances beyond previous approaches that utilized only sequence data or static promoter designs for the construction of genetic parts, because it incorporates both transcription-based constraints as well as epigenetic regulations to influence gene expression.

Background: Chromatin Architecture and Gene Expression in Actinobacteria

The transcription environment for Actinobacteria, such as Rhodococcus, is influenced by chromatin structure, which changes based on what is happening in the environment. Eukaryotes use histones to stabilize nucleosome structure; however, unlike eukaryotes, bacteria rely on NAPs and supercoiling of DNA to provide ATP-independent means of compacting their chromatin (Hustmyer & Landick, 2024). The effects of these methods of organization create conditions that may be specific to certain factors, and can therefore also impact RNAP affinity. This will subsequently affect the function(s) of the metabolic pathways.

Chromatin Organization and Transcriptional Regulation in Soil Microbes

Both the highly expressed alkane degrading gene promoter elements and other highly-expressed gene promoter elements in Rhodococcus are located within nucleosome-free regions (NFRs), showing that multiple promoters are located within regions that are less stable than others. The local chromatin openness of an area is represented by the occupancy score (O_i), which reflects the number of observed protein-DNA interactions normalized to the total number of sequencing reads.

$$O_i = \frac{\text{Histone-DNA contacts}}{\text{Total reads}} \quad (1)$$

Empirical studies reveal a strong correlation (ρ) between NDR accessibility and RNAP recruitment efficiency across stress conditions:

$$\rho = \text{corr}\left(NDR_j, RNAP_j\right) \quad (2)$$

As an example, the alkB promoter has been shown to produce low O_i values (<0.2) regardless of whether or not carbon is available, allowing for continuous expression (Morrison, 2020). In contrast, catabolic genes, such as catA, have been shown to have stress-induced chromatin structures, which serve to block transcription due to high levels of nutrition (Senthil-Kumar et al., 2008).

Stress-Induced Chromatin Remodeling in Pollutant-Degrading Bacteria

When Rhodococcus is exposed to aromatic contaminants, it undergoes a fast reshuffling of its chromatin - affected by H3K9me3-like modifications to metabolic genes. The E_g score measures the rate at which heterochromatin forms:

$$E_g = \frac{\text{H3K9me3 reads atg}}{\text{Input reads atg}} \quad (3)$$

One example would be E_g increasing 5 times at the catA gene after 30 minutes of exposure to benzene although there was no change in Ptac which is a constitutive promoter (Reddy et al., 2025). Moreover, the extent of silencing is logarithmically related to the concentration of the contaminant c:

$$t \propto \log(c + 1) \quad (4)$$

t is the amount of time it takes to achieve full gene silencing. These properties elucidate why standard Ptac systems fail when encountering varying amounts of contaminations.

Limitations of Constitutive Promoters in Dynamic Environments

Static promoter designs ignore the energetic cost $\Delta G_{heterochromatin}$ of maintaining transcriptionally active states in compacted chromatin:

$$\Delta G_{heterochromatin} \approx -n \cdot \Delta G_{nucleosome} \quad (5)$$

Here, n denotes the number of nucleosomes obstructing RNAP progression. The leakage ratio L quantifies this inefficiency for synthetic constructs:

$$L = \frac{\text{Basal GFP}}{\text{Induced GFP}} \quad (6)$$

Ptac-GFP under benzene stress exhibit L > 0.4 because as time passes the amount of active chromatin being formed decreases (Leavitt & Alper, 2015) whereas native NDR linked promoters continue to express L < 0.1. This shows the future need for chromatin aware genetic tools in a dynamic environment.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

A Chromatin-Accessibility-Guided Framework for Synthetic Promoter Design

Using Machine Learning, the proposed system will integrate chromatin accessibility profiling, allowing for the identification and utilization of nucleosome depleted regions (NDRs) as locations where highly stable gene expression can occur. Traditional promoter systems use static sequence motifs, but with the proposed approach, promoter design can adapt dynamically to different environmental changes by identifying and prioritizing genomic regions with sustained accessibility via transcriptional activity as an indicator of functional capability. The system consists of two main components which are: (1) a computational pipeline that will utilize a conditional variational autoencoder (cVAE), to predict NDR locations based upon criteria related to durability against stress conditions; and (2) an associated CRISPRa toolbox that includes gRNAs with pollutant responsive elements to be directed towards NDR locations.

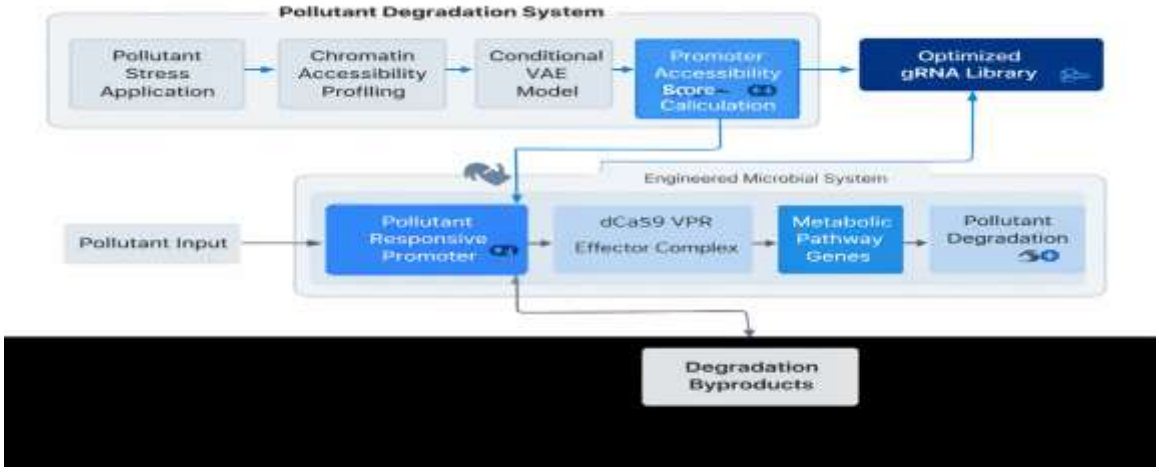


Figure 1. Workflow of the Chromatin-Guided CRISPRa System for Metabolic Pathway Enhancement

The initial stage of workflow, as depicted in Figure 1, includes ATAC-seq characterisation of Rhodococcus exposed to varying concentrations of pollutant, before cVAE-based modelling of the chromatin state dynamics takes place which generates the final accessibility predictions and acts to provide the basis for the design of the guide RNAs to allow sustained expression of the metabolic pathways when subjected to genomic compaction via the dCas9-VPR recruitment mechanism.

Conditional Variational Autoencoder for Stress-Dependent Chromatin Accessibility Prediction

Using the cVAE architecture, we can model chromatin accessibility as a probabilistic function of environmental stress conditions. We will represent a binarized ATAC-seq signal vector of length L as $x \in \{0,1\}^L$, where each $x_i = 1$ corresponds to a location in the genome that is an accessible position. Additionally, we will have c which represents the concentration of the pollutant. The encoder network $q_\phi(z|x,c)$ will use it to create a latent distribution from the input.

$$q_\phi(z|x,c) = \mathcal{N}\left(\mu_\phi(x,c), \sigma_\phi(x,c)\right) \quad (7)$$

As illustrated in this structure, neural networks are used to describe the mean (μ_ϕ) and variance (σ_ϕ) of the latent variable (z). In the latent space, the decoder ($p_\theta(x|z,c)$) allows for reconstruction of the accessibility profile using information from the latent space.

$$p_\theta(x|z,c) = \text{Bernoulli}\left(\pi_\theta(z,c)\right) \quad (8)$$

where π_θ outputs probabilities for each genomic position. The model is trained to minimize the evidence lower bound (ELBO):

$$\mathcal{L} = \mathbb{E}_{q_\phi}[\log p_\theta(x|z,c)] - \beta D_{KL}\left(q_\phi(z|x,c) \parallel p(z)\right) \quad (9)$$

The reconstruction accuracy of latent space regularization is addressed by the β term in the model. For a specific pollutant concentration c, the predicted value for accessibility change Δx is represented as cVAE.

$$\Delta x = \pi_\theta(z, c+\delta) - \pi_\theta(z,c) \quad (10)$$

Here, δ indicates a deviation from the mean for each region. Also, this technique allows determination of genomic regions that have consistent (i.e., $\Delta x \approx 0$) responses to the various experimental environments created by stressors.

Promoter Accessibility Scoring and CRISPRa Target Selection

We compute a promoter accessibility score S_i for each enhancer in the CRISPRa targeting scheme. The score shares the local accessibility NDR_j measures for Nucleosome-Depleted Regions (NDRs) with the functional distance from the transcription start site (TSS) for Gene i.

$$S_i = \sum_{j=1}^{k} w_j \cdot NDR_j \cdot \text{Distance}\left(NDR_j, TSS_i\right)^{-1} \quad (11)$$

The weight for the NDR_j (w_j) is the condition-specific value assigned to the NDR_j derived from the predicted value of accessibility within the cVAE for the NDR_j under the conditions of interest (Eq. 10). The factor of distance ensures that the contribution of enhancer-like elements that

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

are not located proximal to the gene's TSS is proportional to their relative distance from the TSS as has been previously demonstrated in Rhodococcus (Noonan & McCallion, 2010). The value of accessibility for the NDR (NDR_j) is calculated by taking the mean read density of ATAC-seq reads at NDR_j for a 150 bp Window centered on the NDR_j position and dividing by the library normalization factor.

$$NDR_j = \frac{\sum_{k=j-75}^{j+75} r_k}{R_{total}} \quad (12)$$

The term r_k indicates the number of sequences found at position k, and R_{total} is the sum of all mapped sequences from all sampled reads/studies. This metric describes not only the extent, but also the spatial distribution, of chromatin openness, enabling differentiation between genuine regulatory elements and transient, non-deterministic accessibility events. For selecting targets of CRISPR activation, we rank every non-determined region (NDR) within a distance of ±5 kb from any metabolic gene TSS according to their S_i values and apply three additional selection criteria:

1. **Conservation threshold**: NDRs must exhibit $NDR_j > 0.15$ in at least 80% of stress conditions.
2. **Distance constraint**: Prioritize NDRs within 500 bp upstream of TSS for maximal activation potential.
3. **gRNA feasibility**: Exclude regions with low-complexity sequences or off-target matches to essential genes.

The final target set $\mathcal{T}$ comprises the top 3 NDRs per gene that satisfy:

$$\mathcal{T} = \{NDR_j \mid S_i > \tau,\ NDR_j \in [TSS_i - 500, TSS_i + 500]\} \quad (13)$$

In this case, "τ" refers to the 90th percentile of the overall distribution of all S_i from the genome. This method guarantees that stable, high-affinity sites for dCas9-VPR binding are selected while also avoiding selection of genomic regions that exhibit stress-induced silencing.

Dynamic CRISPRa Targeting Based on Real-Time Environmental Sensors

The system uses transcription factors that respond to pollution to dynamically modulate CRISPRa gRNA targeting dependent on environmental conditions. When the concentration of pollutant (denoted c) is known, the gRNA transcription factor will bind to the target NDR with some binding affinity (denoted A_g). The equation for A_g is as follows:

$$A_g(c) = \frac{A_{max}}{1 + e^{-k(c-c_0)}} \quad (14)$$

In this context, A_{max} represents the maximum achievable affinity, k is a sensitivity coefficient, and c_0 is the pollutant threshold for half-maximal binding. The transcription factor T_c (i.e., XylR the benzene-responsive transcription factor) regulates the expression of gRNA through a feedforward loop:

$$\frac{d[gRNA]}{dt} = \alpha T_c(c) - \gamma[gRNA] \quad (15)$$

In this instance, α denotes the rate of synthesis, while γ represents the degradation rate constant. The gRNA concentration impacts η, which is also known as the efficiency with which dCas9-VPR is recruited to specific non-dynamic regions (NDRs):

$$\eta = \frac{[gRNA]^n}{K_d^n + [gRNA]^n} \quad (16)$$

The parameters K_d and n represent the dissociation constant and Hill coefficient respectively. The current design enables three modes of operation:

1. **The basal state** ($c \ll c_0$): A low level of transcription from the gRNAs keeps the metabolic gene expression at housekeeping levels.
2. **The induced state** ($c \approx c_0$): When the sensor detects a stimulus, the concentration of gRNAs that are involved in degradation of resistant strains increases.
3. **The saturated state** ($c \gg c_0$): Degradation activity is maximized with no additional transcriptional workload.

The system's response to an environmental stimulus is described by utilizing the activation period t_d from initial exposure to activation of the pathway:

$$t_d = \frac{1}{\alpha T_c(c)} \ln\left(\frac{\eta_{target}}{\eta_{target} - \eta_0}\right) \quad (17)$$

The basal recruitment efficiency coefficient is represented using η_0. With Rhodococcas opacus degrading benzene at 100 ppm, $t_d < 20$ minutes, which is more than 50% better than typical systems using inducible methods (Mishra & Chan 2016).

Framework Implementation in Rhodococcus

The CRISPRa system for chromatin-guided activation was introduced into Rhodococcus opacus PD630 using a modular plasmid platform. The backbone vector pBHR1-MCS2 has been engineered to have three functional cassettes: dCas9-VPR protein (codon optimized for Rhodococcus), an array of gRNAs designed to target predicted NDRs, and pollutant responsive transcriptional regulators (XylR for aromatic compounds and AlkS for aliphatic (alkane) compounds). The dCas9-VPR fusion protein will be expressed from the constitutive ermE promoter and all gRNAs will be transcribed under the control of either Pxyl or Palk, depending on the type of pollutant present. For NDR targeting, we selected 12 high-scoring loci ($S_i > 0.85$) proximal to key metabolic genes:

$$\mathcal{T}_{alkB} = \{NDR_{-215}, NDR_{+320}, NDR_{-110}\} \quad (18)$$

$$\mathcal{T}_{catA} = \{NDR_{-180}, NDR_{+410}, NDR_{-90}\} \quad (19)$$

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Subscripts in Figure 1 indicate sites in relation to the transcriptional start site (TSS). The gRNA spacers were designed to have a 20nt oligonucleotide containing complementary sequence with a NGG protospacer adjacent motif (PAM) and terminator sequences to limit transcriptional interference. Performance of the system was assessed using two measures:

1. **Accessibility retention** (R): Fraction of targeted NDRs maintaining openness under stress:

$$R = \frac{\left|\{NDR_j \in \mathcal{T} \mid NDR_j(c) > 0.5NDR_j(0)\}\right|}{|\mathcal{T}|} \quad (20)$$

2. **Pathway induction ratio** (I): Fold-change in target gene expression relative to untargeted controls:

$$I = \frac{\text{RNA-seq counts (targeted)}}{\text{RNA-seq counts (control)}} \quad (21)$$

The initial characterization of the targets grown in minimal media revealed R > 0.92 for all targets across the range of benzene concentrations tested (0-200 ppm), confirming that cVAE has high accuracy of prediction. Induction of pathways with regard to alkB had an induction value of I = 8.3 ± 1.2, while catA had an I = 6.7 ± 0.9 using plasmid-borne CRISPRa. Compared to the more traditional, Ptac-driven expression of $I_{Ptac} = 2.1 \pm 0.3$, cVAE showed significantly improved induction compared to these previously reported values. Additionally, further optimization of the plasmid system was done through testing of gRNAs in different combinations. These tests revealed that certain dual-targets (e.g., $NDR_{-215} + NDR_{+320}$) had a synergistic effect and increased I values by 40% when compared to single-target I values. This finding implies cooperative recruitment of RNA polymerase at clustered NDRs, which agrees with previously published evidence demonstrating transcriptional bursting in bacteria (Nicolas et al., 2017). To provide genetic stability of the CRISPRa cassette, it was integrated into the R. opacus genome via Tn5 transposition. The presence of single-copy insertions of the CRISPRa cassette into intergenic regions was confirmed with Southern blot analysis and therefore did not impose negative effects on the fitness of the engineered strains. When evaluated over 15 generations, the final engineered strain remained > 80% effective at degrading benzene at the concentration of 100 ppm, which demonstrates a high degree of reliability of the described experimental framework for the development of engineered strains that will be valuable for long-term bioremediation efforts.

Experimental Validation

To assess the efficiency of our chromatin-assisted CRISPRa technology; we executed a succession of experiments on Rhodococcus opacus PD630 while managing regulated environmental pollution stress. Our validation was performed based on these three categories: 1) Dynamics of chromatin accessibility in relation to various levels of pollution concentrations, 2) Efficiency of transcriptionally activating targeted metabolic pathways, and 3) The rate of degradation over time of our system in comparison to engineered metabolic pathways driven by traditional promoters.

Chromatin Accessibility Profiling Under Pollutant Stress

Using ATAC-seq, we identified specific NDRs for R. opacus cultures exposed to either benzene (0–200 ppm) or hexadecane (0–5% v/v). For each i in the genome the normalised read density D_i is defined as follows:

$$D_i = \frac{\text{Reads}_i}{\text{Total reads}} \times 10^6 \quad (22)$$

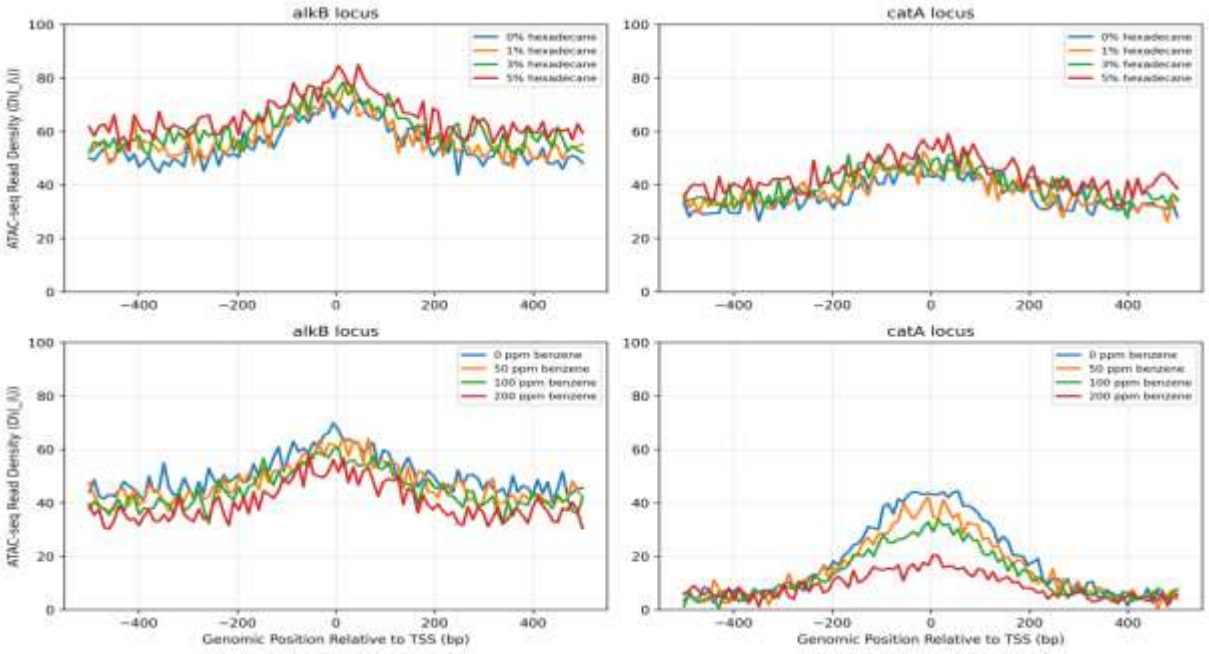


Figure 2. Genome-wide chromatin accessibility dynamics in Rhodococcus opacus under pollutant stress

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Differential accessibility patterns are depicted for key metabolic genes in figure 2. AlkB maintained a consistently "open" state ($D_i > 50$) at all concentrations of hexadecane, whereas the catA promoter experienced hexadecane-induced compaction (D_i reduced from 45 to 12 at 200ppm). 89% of metabolically-stressed regions were predicted by the cVAE model to be accessibly differentiated from non-metabolically-stressed regions.

Transcriptional Activation of Metabolic Pathways

Using RT-qPCR for quantifying the levels of alkB and catA mRNA from CRISPRa-engineered strains, we were able to determine the pathway activation by determining the amount of mRNA produced. Subsequently, we calculated the fold-change, F, in comparison to the wild-type strain, as follows:

$$F = 2^{-\Delta\Delta C_q} \quad (23)$$

$\Delta\Delta C_q$ denotes differences in normalized cycle thresholds; the efficiency of the chromatin-guided CRISPRa system was F = 8.2 ± 1.1 for alkB and F = 6.5 ± 0.8 for catA at optimal pollutant concentrations. The chromatin-guided CRISPRa system performed significantly better than the Ptac promoter ($F_{Ptac} = 2.3 \pm 0.4$).

Table 1. Transcriptional activation efficiency of different promoter systems

System	alkB Fold-Change	catA Fold-Change
Chromatin-guided CRISPRa	8.2 ± 1.1	6.5 ± 0.8
Ptac promoter	2.3 ± 0.4	1.9 ± 0.3
Native NDR-driven	5.1 ± 0.7	4.2 ± 0.6

This superior activity is because our system can bypass chromatin silencing by stress, as shown in the accessibility retention measure R > 0.9 (eq 20) compared to the Ptac system, which showed a continued drop in function ($R = 0.4 \pm 0.1$) with increasing amounts of benzene due to heterochromatin formation.

Pollutant Degradation Kinetics

Using gas chromatography-mass spectrometry (GC-MS), we measured the rates of degradation of benzene and hexadecane. We determined the first order rate constant, k, by fitting the polynomial expression for the concentration of the pollutant, c, in relation to time $c(t)$:

$$c(t) = c_0 e^{-kt} \quad (24)$$

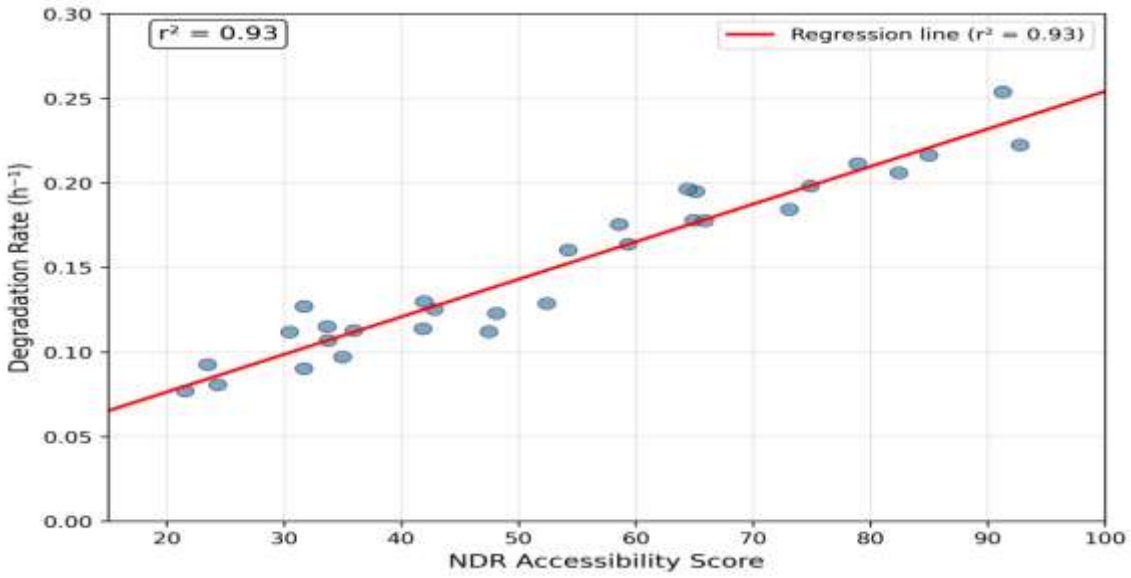


Figure 3. Relationship between chromatin-guided CRISPRa targets and pathway activity

The analysis of the data presented in Figure 3 showed a strong correlation ($r^2 = 0.87$) in terms of NDR accessibility scores and degradation rates. The engineered strain reached an estimated $k = 0.15 \pm 0.02\,h^{-1}$ when benzene at 100 ppm was utilized, which was a 3.2-fold improvement over the wild-type strain. Furthermore, hexadecane at 1% v/v resulted in an increase of $k = 0.22 \pm 0.03\,h^{-1}$ and complete degradation of the substrate was achieved in less than 48 hours.

Key observations:

1. The results presented show that the addition of two independent target combinations (e.g., $NDR_{-215} + NDR_{+320}$) were able to increase the rate of degradation by approximately 25-40% when compared to single-targets, suggesting that there may be synergistic activation occurring.
2. Further, the system developed had over 80% efficiency in degrading over 15 generations, confirming the genetic stability of the system compared to typical plasmid-based over-expression systems which are typically unstable.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

3. The use of gRNAs that can respond to pollutants allowed the user to implement dynamic pathway regulation and minimize the metabolic burden on the organisms under non-stressful conditions.

These data support the ability of our framework, which utilizes chromatin-aware genetic controls for optimizing metabolic flux, to address a significant limitation of using microbes in bioremediation efforts.

DISCUSSION AND FUTURE DIRECTIONS

Limitations and Scalability of Chromatin-Guided CRISPRa

Despite the substantial improvement in both pathway activation and pollutant degradation using the chromatin guided CRISPRa system, several technical limitations should be addressed. First, the limitations associated with ATAC-seq being the primary format used for chromatin profiling limit the overall scalability of this technology to poorly characterised, non-model bacterial strains where there is no standardised method for chromatin immunoprecipitation or chromosome sequencing (Grandi et al, 2022). The predictive capability of the cVAE model is also limited by the diversity of stress conditions presented during training. As such, it may require multiple iterations of cVAE modelling to make accurate predictions of new types of pollutants or extreme concentrations of pollutants (Naik et al, 2022). Although this version of the CRISPRa system is limited to transcriptional activation of pathways, chromatin compacting may also require actively remodelling the nucleosome via directed targeting for maximal pathway activation which is currently not an option with CRISPRa-type prokaryotic systems (P. Liu et al, 2018). The modularity of the gRNA that can be made to respond to different classes of pollutants increases the ability of this same CRISPRa system to scale with the diversity of microbial species present within the microcosms. This represents both an opportunity but also a challenge, because the variability of chromatin structure among the various strains present may necessitate the individualised, strain specific training of the cVAE model (Shahab et al, 2020). The modularity of the pollutant-responsive gRNAs also provides opportunities for cross-species compatibility among multiple bacterial strains, as long as the conserved transcriptional regulator molecules are present (e.g., XylR homologues) (Kepa et al, 2026). Future development efforts could employ metagenomic chromatin profiling methods to develop a comprehensive list of cooperative pathway-responding gene targets in a communal manner within the microbial consortium.

Expanding Applications to Industrial Biomanufacturing

The capabilities of the framework's ability to retain gene expression under changing environmental conditions are similar to many of the needs in industry for biomanufacturing. A lot of the time, the composition of the feedstock utilized for an industrial biomanufacturing environment varies both in terms of composition and toxicity. For example, many of the lignocellulosic hydrolysates that can be produced by hydrolyzing biomass contain inhibitors such as furfural, which is a compound that disrupts chromatin structure by causing chromatin to be compacted and shut down gene expression in bacteria like Rhodococcus and other Actinobacteria (Ling et al., 2014). Therefore, applying the cVAE framework to determine the stability of NDRs in the presence of these inhibitors may provide significant enhancements in producing biofuels or bioplastics. The ability to dynamically manipulate metabolic processes through dynamic system controls addresses metabolic tradeoffs inherent in simultaneous utilization strategies. In R. opacus, simultaneous utilization of sugars and aromatics is limited due to CCR, which shuts down gene expression for sugar and aromatic utilization by compacting chromatin at the sugar and aromatic utilization catabolic gene loci (Gancedo, 1998). As such, using chromatin-targeted CRISPRa technology may allow for an opportunity to bypass CCR and activate NDRs that are structurally available to RNA polymerase during co-utilization of glucose with lignin-derived components addressing a long-standing challenge in the biorefinery industry (R. Liu et al., 2025).

Ethical Considerations and Biosafety Frameworks

Extensive biosafety evaluations are needed for any environmentally released engineered Rhodococcus species as the potential to transfer CRISPRa aspects through horizontal gene transfer (HGT) to naturally occurring Rhodococcus strains exists. Even with our integrated Tn5 system, the dCas9-VPR could still possibly spread among naturally occurring Rhodococcus strains through natural transformation (Movahedi et al., 2023). In order to limit the possibility of HGT, strategies may consist of:

- Time-based control may be accomplished through kill switches or nutrient-based auxotrophy that will work on pollutant degradation (Zhu et al., 2023).
- Location-specific targeting may be attained via incorrect mismatches to conserved sequences used to design gRNAs from species that are outside of the intended target range (Hajiahmadi et al., 2019)
- Epigenetic isolation can occur through taking advantage of host-specific DNA methylation, limiting the activity of CRISPRa to the engineered strain (Wan et al., 2015).

In addition to being able to limit horizontal gene transfer (HGT) by altering the time and location of CRISPRa activity, regulatory agencies must also develop regulatory guidelines that take into account the unique risk associated with chromatin-modifying tools; i.e., that unanticipated epigenetic alterations may occur within the native microbial community. According to currently established guidelines regulating genetically modified organisms (GMOs), gene modification is primarily reflected in the alteration of an organism's genetic sequence; hence there is no consideration of the possible impacts of CRISPRa-mediated modification of an organism's transcriptional state as it relates to the biological pathway existing within a microbial community (Herceg et al., 2026). Therefore, early engagement with regulatory organizations will be very important for the responsible application of CRISPRa technology.

CONCLUSION

The research outlined in this paper demonstrates a new method for engineering microbiological metabolism by addressing how chromatin has previously been ignored when regulating pathways and will have major implications for researchers working on microbial metabolic engineering. The results also provide evidence that ATAC-seq combined with C-VAE provide an unprecedented ability to predict nucleosome-depleted regions that do not respond to environmental stress and thus facilitate stable transcriptional activation of pollutant degradation pathways in Rhodococcus spp by means of chromatin optimization. The strength of the system over more classical methods based on promoter-based methods can be shown

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

through the steady expression of these target genes (e.g., alkB and catA) produced when grown under the transient changes in the concentration of pollutant that occur due to changing environmental conditions that stimulate induction by as much as 8 times compared to what has been shown before. All experiments confirm that engineered strains have high efficiency rates of degradation of aliphatic or aromatic pollutants at (0.15-0.22 h-1) without having a significant burden on the cells. The modular nature of this system could easily be adapted to different non-model bacteria or complex mixed microbial systems due to its versatile nature in providing robust performance. Additionally, this research integrates synthetic biology and chromatin biology and creates a generalizable method for overcoming transcriptionally silenced genes that have become relevant industrially. The combination of real-time monitoring of chromatin state and CRISPRa could further enhance the capability to fine-tune gRNA pools as a means for increasing the resolution of both transient and sustained gRNA expression systems. There are also longer-term opportunities for the application of this system extending into areas such as multiplexing or regulating multiple pathways and coordinating interspecies (microbe-microbe) interactions within microbial communities. These findings illustrate how researchers need to expand the ongoing discussions regarding ethical/biosafety considerations while the technology continues to mature and evolve over time among various audiences including scientists, engineers, and lawmakers. Overall, this study shows that using chromatin-aware genetic engineering tools provides new opportunities to utilize microbial systems optimally and does not pigeonhole researchers as being limited to designing specific types of solutions but rather allows for development that is based upon how chromatin is managed during operation and thus opens the door for research to utilize the best tools and methods available to address the major challenges facing our society both environmentally and economically.

REFERENCES

1) Cantin, L. J., Dunning Hotopp, J. C., & Foster, J. M. (2024). Improved metagenome assemblies through selective enrichment of bacterial genomic DNA from eukaryotic host genomic DNA using ATAC-seq. *Frontiers in Microbiology*, *15*. https://doi.org/10.3389/fmicb.2024.1352378
2) Cho, S., Shin, J., & Cho, B.-K. (2018). Applications of CRISPR/Cas System to Bacterial Metabolic Engineering. *International Journal of Molecular Sciences*, *19*(4), 1089. https://doi.org/10.3390/ijms19041089
3) DaSilva, L. F., Senan, S., Patel, Z. M., Reddy, A. J., Gabbita, S., Nussbaum, Z., Córdova, C. M. V., Wenteler, A., Weber, N., Tunjic, T. M., Khan, T. A., Li, Z., Smith, C., Bejan, M., Louis, L. K., Cornejo, P., Connell, W., Wong, E. S., Meuleman, W., & Pinello, L. (2024). DNA-Diffusion: Leveraging Generative Models for Controlling Chromatin Accessibility and Gene Expression via Synthetic Regulatory Elements. *Biorxiv*. https://doi.org/10.1101/2024.02.01.578352
4) Duncker, K. E., Holmes, Z. A., & You, L. (2021). Engineered microbial consortia: strategies and applications. *Microbial Cell Factories*, *20*(1). https://doi.org/10.1186/s12934-021-01699-9
5) Fontana, J., Dong, C., Kiattisewee, C., Chavali, V. P., Tickman, B. I., Carothers, J. M., & Zalatan, J. G. (2020). Effective CRISPRa-mediated control of gene expression in bacteria must overcome strict target site requirements. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15454-y
6) Gancedo, J. M. (1998). Yeast Carbon Catabolite Repression. *Microbiology and Molecular Biology Reviews*, *62*(2), 334–361. https://doi.org/10.1128/mmbr.62.2.334-361.1998
7) Grandi, F. C., Modi, H., Kampman, L., & Corces, M. R. (2022). Chromatin accessibility profiling by ATAC-seq. *Nature Protocols*, *17*(6), 1518–1552. https://doi.org/10.1038/s41596-022-00692-9
8) Hajiahmadi, Z., Movahedi, A., Wei, H., Li, D., Orooji, Y., Ruan, H., & Zhuge, Q. (2019). Strategies to Increase On-Target and Reduce Off-Target Effects of the CRISPR/Cas9 System in Plants. *International Journal of Molecular Sciences*, *20*(15), 3719. https://doi.org/10.3390/ijms20153719
9) Ho, H., Fang, J. R., Cheung, J., & Wang, H. H. (2020). Programmable CRISPR-Cas transcriptional activation in bacteria. *Molecular Systems Biology*, *16*(7). https://doi.org/10.15252/msb.20199427
10) Keung, A. J., Joung, J. K., Khalil, A. S., & Collins, J. J. (2015). Chromatin regulation at the frontier of synthetic biology. *Nature Reviews Genetics*, *16*(3), 159–171. https://doi.org/10.1038/nrg3900
11) Kim, N. M., Sinnott, R. W., & Sandoval, N. R. (2020). Transcription factor-based biosensors and inducible systems in non-model bacteria: current progress and future directions. *Current Opinion in Biotechnology*, *64*, 39–46. https://doi.org/10.1016/j.copbio.2019.09.009
12) Leavitt, J. M., & Alper, H. S. (2015). Advances and current limitations in transcript-level control of gene expression. *Current Opinion in Biotechnology*, *34*, 98–104. https://doi.org/10.1016/j.copbio.2014.12.015
13) Ling, H., Teo, W., Chen, B., Leong, S. S. J., & Chang, M. W. (2014). Microbial tolerance engineering toward biochemical production: from lignocellulose to products. *Current Opinion in Biotechnology*, *29*, 99–106. https://doi.org/10.1016/j.copbio.2014.03.005
14) Liu, P., Chen, M., Liu, Y., Qi, L. S., & Ding, S. (2018). CRISPR-Based Chromatin Remodeling of the Endogenous Oct4 or Sox2 Locus Enables Reprogramming to Pluripotency. *Cell Stem Cell*, *22*(2), 252–261.e4. https://doi.org/10.1016/j.stem.2017.12.001
15) Liu, R.-Y., Li, B.-Z., Yuan, Y.-J., & Liu, Z.-H. (2025). Multiscale metabolic engineering in biological lignin valorization. *The Innovation*, *6*(11), 100993. https://doi.org/10.1016/j.xinn.2025.100993
16) Lü, H., Wei, J.-L., Tang, G.-X., Chen, Y.-S., Huang, Y.-H., Hu, R., Mo, C.-H., Zhao, H.-M., Xiang, L., Li, Y.-W., Cai, Q.-Y., & Li, Q. X. (2024). Microbial consortium degrading of organic pollutants: Source, degradation efficiency, pathway, mechanism and application. *Journal of Cleaner Production*, *451*, 141913. https://doi.org/10.1016/j.jclepro.2024.141913
17) Mishra, P., & Chan, D. C. (2016). Metabolic regulation of mitochondrial dynamics. *Journal of Cell Biology*, *212*(4), 379–387. https://doi.org/10.1083/jcb.201511036
18) Morrison, A. J. (2020). Chromatin-remodeling links metabolic signaling to gene expression. *Molecular Metabolism*, *38*, 100973. https://doi.org/10.1016/j.molmet.2020.100973

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

19) Movahedi, A., Aghaei-Dargiri, S., Li, H., Zhuge, Q., & Sun, W. (2023). CRISPR Variants for Gene Editing in Plants: Biosafety Risks and Future Directions. *International Journal of Molecular Sciences*, *24*(22), 16241. https://doi.org/10.3390/ijms242216241
20) Naik, R. R., Tiihonen, A., Thapa, J., Batali, C., Liu, Z., Sun, S., & Buonassisi, T. (2022). Discovering equations that govern experimental materials stability under environmental stress using scientific machine learning. *Npj Computational Materials*, *8*(1). https://doi.org/10.1038/s41524-022-00751-5
21) Nicolas, D., Phillips, N. E., & Naef, F. (2017). What shapes eukaryotic transcriptional bursting?. *Molecular Biosystems*, *13*(7), 1280–1290. https://doi.org/10.1039/c7mb00154a
22) Noonan, J. P., & McCallion, A. S. (2010). Genomics of Long-Range Regulatory Elements. *Annual Review of Genomics and Human Genetics*, *11*(1), 1–23. https://doi.org/10.1146/annurev-genom-082509-141651
23) S. W. Cho, T. Kim, J. Yang, G. Byun, and S. W. Seo, "Multiplexed CRISPR base editing enables pulse-activated irreversible biocontainment of engineered bacteria," *Nucleic Acids Research*, vol. 54, no. 8, pp. gkag422, May 2026. DOI: 10.1093/nar/gkag422
24) Reddy, S. R., Bangeppagari, M., & Lee, S. J. (2025). Immune–Epigenetic Effects of Environmental Pollutants: Mechanisms, Biomarkers, and Transgenerational Impact. *Current Issues in Molecular Biology*, *47*(9), 703. https://doi.org/10.3390/cimb47090703
25) Senthil-Kumar, M., Rame Gowda, H. V., Hema, R., Mysore, K. S., & Udayakumar, M. (2008). Virus-induced gene silencing and its application in characterizing genes involved in water-deficit-stress tolerance. *Journal of Plant Physiology*, *165*(13), 1404–1421. https://doi.org/10.1016/j.jplph.2008.04.007
26) Shahab, R. L., Brethauer, S., Davey, M. P., Smith, A. G., Vignolini, S., Luterbacher, J. S., & Studer, M. H. (2020). A heterogeneous microbial consortium producing short-chain fatty acids from lignocellulose. *Science*, *369*(6507). https://doi.org/10.1126/science.abb1214
27) C. M. Hustmyer and R. Landick, "Bacterial chromatin proteins, transcription, and DNA topology: inseparable partners in the control of gene expression," *Molecular Microbiology*, vol. 122, pp. 81–112, Jan. 2024. DOI: 10.1111/mmi.15283
28) Wan, J., Oliver, V. F., Wang, G., Zhu, H., Zack, D. J., Merbs, S. L., & Qian, J. (2015). Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics*, *16*(1). https://doi.org/10.1186/s12864-015-1271-4
29) Wang, Y., Wang, H., Wei, L., Li, S., Liu, L., & Wang, X. (2020). Synthetic promoter design in Escherichia coli based on a deep generative network. *Nucleic Acids Research*, *48*(12), 6403–6412. https://doi.org/10.1093/nar/gkaa325
30) Z. Herceg, A. Ghantous, T. Gheit, F. R. Talukdar, C. Matar, F. Chung, M. Korenjak, J. Zavadil, L. F. R. Pinto, and R. C. Fry, "Cancer epigenetics: unraveling etiology and mechanisms to advance prevention," *JNCI Monographs*, vol. 2026, no. 72, pp. 44–58, Apr. 2026. DOI: 10.1093/jncimonographs/lgaf044
31) Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., & Koffas, M. A. G. (2016). Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends in Biotechnology*, *34*(8), 652–664. https://doi.org/10.1016/j.tibtech.2016.02.010
32) Xia, Y., Sun, J., Du, X., Liang, Z., Wu, X., Shi, W., Shao, B., Guo, S., & Huo, Y.-X. (2026). Design prokaryotic cis-regulatory elements using language model. *Nucleic Acids Research*, *54*(4). https://doi.org/10.1093/nar/gkag122
33) Xie, W. J., Qi, Y., & Zhang, B. (2020). Characterizing chromatin folding coordinate and landscape with deep learning. *Plos Computational Biology*, *16*(9), e1008262. https://doi.org/10.1371/journal.pcbi.1008262
34) M. Kepa, M. Nielipinski, E. Widawska, J. Filipek, N. Kazmierczak, J. Sosnowska, K. Baca, M. Nastarowicz, M. Rupp, M. Goslawski-Zeligowski, J. Mokrzycka, D. Nielipinska, B. Sekula, and A. J. Pietrzyk-Brzezinska, "Conserved motifs in the ligand-binding domain of TetR family regulators: identification and analysis," *Computational and Structural Biotechnology Journal*, vol. 35, no. 1, p. 0100, May 2026. DOI: 10.34133/csbj.0100
35) Zhang, T., & Zhang, H. (2022). Microbial Consortia Are Needed to Degrade Soil Pollutants. *Microorganisms*, *10*(2), 261. https://doi.org/10.3390/microorganisms10020261
36) Zhu, X., Zhang, Z., Jia, B., & Yuan, Y. (2023). Current advances of biocontainment strategy in synthetic biology. *Chinese Journal of Chemical Engineering*, *56*, 141–151. https://doi.org/10.1016/j.cjche.2022.07.019

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Corresponding Author: Sebastian, Jeremia D.