Volume 3, Number 3, 2026 Pages 1-25

Bioscience Insights

ISSN: 3080-3489 (Print)
ISSN: 3080-3497 (Online)

BiosciInsights.com

CRISPR/CAS-BASED GENOME EDITING AND SYSTEMS BIOLOGY APPROACHES FOR ENGINEERING HIGH-PERFORMANCE BIOLOGICAL SYSTEMS

Kabeer Ahmad[*1], Badi Uz Zaman[2]

[1]*Department of Biotechnology, Faculty of Health Sciences, COMSATS University Islamabad, Abbottabad Campus, Pakistan*
[2]*Department of Plant Sciences, Quaid-i-Azam University, Islamabad, Pakistan*

[*1]kabeerkhan192@yahoo.com, [2]bzaman@bs.qau.edu.pk

https://doi.org/10.5281/zenodo.24003447

Article Information

Article History

Received:	June 19, 2026
Accepted:	September 07, 2026
Available Online:	October 30, 2026

Keywords:
CRISPR/Cas9, Cas12a, prime editing, base editing, systems metabolic engineering, Design-Build-Test-Learn cycle, genome-scale metabolic models, machine learning, industrial biotechnology, metabolic burden

Copyright @Author
Corresponding Author: *
Kabeer Ahmad

Abstract

The engineering of high-performance biological systems has been fundamentally transformed by the convergence of CRISPR/Cas-based genome editing and systems biology. This review examines the synergistic integration of precision genome editing technologies with computational modeling frameworks to enable predictive, iterative strain optimization. We trace the evolution of programmable nucleases from protein-guided ZFNs and TALENs to RNA-guided CRISPR/Cas9 and Cas12a systems, highlighting recent advances in double-strand break-free precision tools including base editors, prime editors, and recombinase-assisted large-cargo integration platforms. The Design-Build-Test-Learn (DBTL) cycle serves as the central operational framework, wherein genome-scale metabolic models, machine learning algorithms, and artificial intelligence guide target identification, experimental design, and data-driven model refinement. We examine industrial chassis engineering across Escherichia coli, Saccharomyces cerevisiae, Yarrowia lipolytica, and Corynebacterium glutamicum, documenting quantitative production benchmarks including β-thymidine (25.4 g/L), S-adenosyl-L-methionine (13.96 g/L), and ergothioneine (385.70 mg/L). Machine learning platforms such as OptiPrime and PRIDICT are discussed for their capacity to predict mismatch repair evasion and optimize prime editing guide RNA design. Critical bottlenecks including host DNA repair constraints, delivery vector payload limitations, and metabolic burden during scale-up are analyzed. Future directions emphasize MMR-evasive editing strategies, high-capacity adenoviral vectors, split-intein architectures, and biosensor-controlled dynamic regulation circuits that balance cellular maintenance with target product synthesis, establishing a predictive science of industrial strain optimization.

1. INTRODUCTION

The engineering of biological systems for the scalable synthesis of high-value pharmaceuticals, platform chemicals, advanced biofuels, and nutraceuticals is central to modern industrial biotechnology. Historically, classical metabolic engineering relied on intuitive, low-throughput genetic manipulations such as episomal plasmid-driven gene overexpression, random chemical or ultraviolet mutagenesis, and single-gene knockouts (Yang & Lu, 2025). While these early efforts established proof-of-concept microbial production pathways, they were intrinsically constrained by genetic instability, metabolic burden, non-specific physiological off-target effects, and an inability to systematically manipulate complex metabolic networks (Bourgade *et al.*, 2025).

The paradigm has shifted from isolated strain modification to systems metabolic engineering, driven by the integration of synthetic biology, functional multi-omics, computational modeling, and targetable genome editing. Central to this transformation is the Design-Build-Test-Learn (DBTL) cycle framework. The DBTL framework converts empirical biological engineering into a predictive, standardized, and iterative engineering discipline (Abbate *et al.*, 2023).

Within this framework, the Design phase relies on genome-scale metabolic models (GEMs) and machine learning algorithms to identify non-obvious gene targets, precursor bottlenecks, and optimal metabolic flux redistributions across thousands of interconnected biochemical reactions. The Build phase leverages precision genome editing technologies predominantly the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated protein (Cas) systems to physically reconfigure host genomes with nucleotide-level specificity (Chen & Ping, 2023). The Test phase utilizes high-throughput screening (HTS), microfluidic assays, isotopic tracer metabolomics, and genetically encoded biosensors to dynamically assess single-cell states and pathway kinetics. Finally, the Learn phase employs advanced computational architectures, artificial intelligence, and Bayesian optimization to process multi-omics datasets, re-train predictive algorithms, and inform subsequent engineering iterations (Li *et al.*, 2025).

The synergistic convergence of CRISPR/Cas technologies and systems biology principles has redefined bio-manufacturing capacities across model prokaryotes (Escherichia coli, Corynebacterium glutamicum) and eukaryotes (Saccharomyces cerevisiae, Yarrowia lipolytica). By bridging the gap between computational phenotype prediction and physical execution, biological engineers can reprogram global transcriptional networks, optimize metabolic flux toward target molecules, and minimize carbon-divertive pathways, achieving high titers, yields, and productivities (Ojha *et al.*, 2026).

2. Molecular Architecture and Evolution of CRISPR/Cas Technologies

2.1. Transition from Protein-Guided to RNA-Guided Endonucleases

The evolution of targeted gene editing tools spans several distinct technological generations. The earliest programmable nucleases, zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), relied on customizable protein-DNA binding domains fused to the catalytic cleavage domain of the FokI restriction endonuclease (Gaj *et al.*, 2013). Zinc finger nucleases operate as tandem arrays of Cys2His2 zinc finger modules, where each finger recognizes a specific trinucleotide DNA sequence. Transcription activator-like effector nucleases utilize modular DNA-binding domains derived from Xanthomonas species, comprising repeat units of approximately 34 amino acids wherein hypervariable residues at positions 12 and 13 termed repeat variable diresidues (RVDs) dictate single-nucleotide recognition (Saugandhika & Jain, 2024).

Despite their utility, both ZFNs and TALENs require complex, labor-intensive protein engineering for every targeted locus, presenting substantial operational barriers for multiplexed, genome-scale engineering campaigns (White, 2026). The operational architecture shifted with the discovery and adaptation of Class 2 prokaryotic adaptive immune systems, specifically the Class 2 Type II (Cas9) and Type V (Cas12a/Cpf1) endonucleases. Unlike Class 1 systems that require multiprotein effector complexes comprising four to seven subunits, Class 2 systems execute sequence-specific cleavage via a single multi-domain protein guided by a short RNA molecule (Šviković, 2022).

Target site selection by Cas9 is dictated by Watson-Crick base pairing between a synthetic single guide RNA (sgRNA) comprising a 20-nucleotide spacer region and the target genomic locus adjacent to a 5'-

NGG-3' protospacer adjacent motif (PAM) (Yamagata, 2022). Upon PAM recognition and R-loop formation, the HNH and RuvC catalytic domains of Cas9 cleave the target and non-target DNA strands, respectively, generating a blunt-ended double-strand break (DSB). In contrast, Cas12a utilizes a T-rich PAM (5'-TTTN-3'), relies on a shorter single crRNA without requiring a tracrRNA, and employs a RuvC-like catalytic domain, assisted by a Nuc domain to introduce staggered cleavage with four- or five-nucleotide 5' overhangs distal to the PAM site. Staggered cleavage facilitates directionality during homology-directed insertion, making Cas12a a powerful complement to Cas9 in A/T-rich genomic environments (Manchanda & Suneja, 2018).

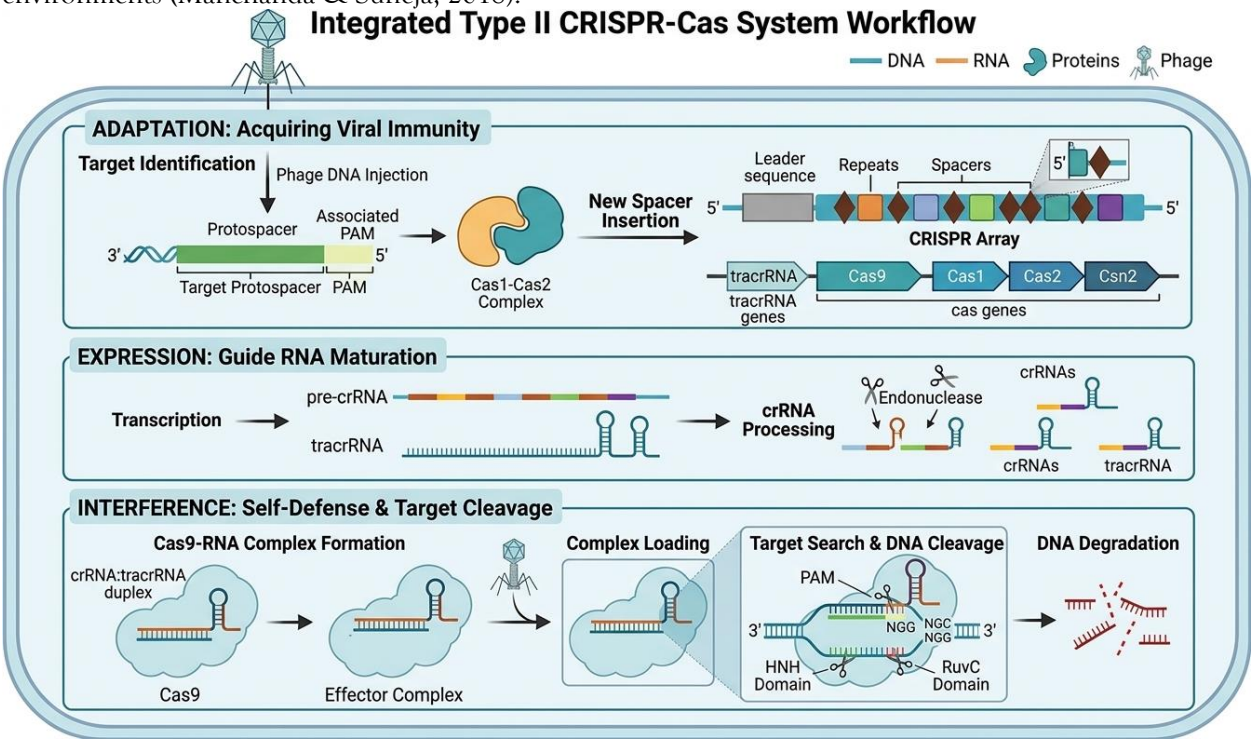


Figure 1. Natural Operational Workflow of Integrated Type II CRISPR-Cas Systems Across Adaptation, Guide RNA Expression, and Interference Stages.

Table 1. Comparative Architecture and Operational Specifications of Major Genome Editing Technologies

Technology	Recognition	Effector	PAM / target site	Editing outcome	Applications
Meganucleases	Structural protein–DNA interface	Intrinsic endonuclease domain	12–40-bp recognition sequence	Double-strand break; often overhangs	Gene targeting; low multiplexing capacity
ZFNs	Engineered zinc-finger arrays	FokI cleavage-domain dimer	Triplet-repeat motifs	Double-strand break; staggered ends	Knockouts and insertions; design-intensive
TALENs	Engineered TALE repeat modules	FokI cleavage-domain dimer	T-base preference	Double-strand break; staggered ends	High-specificity targeting; large assemblies
CRISPR-Cas9	sgRNA–DNA base pairing	HNH and RuvC domains	5′-NGG-3′ PAM	Double-strand break; blunt ends	Knockouts, insertions, multiplex regulation

Technology	Recognition	Effector	PAM / target site	Editing outcome	Applications
CRISPR-Cas12a (Cpf1)	crRNA–DNA base pairing	RuvC nuclease domain	5′-TTTN-3′ PAM	Double-strand break; staggered 5′ overhangs	AT-rich targeting; multiplex crRNA arrays
Base editors (CBE/ABE)	sgRNA-guided nCas9	Deaminase plus nCas9	PAM-dependent; variants available	Single-base transitions; no double-strand break	Precise nucleotide conversion
Prime editors (PE2/PE3/PEmax)	pegRNA-guided nCas9	Reverse transcriptase plus nCas9	Flexible; PAM-dependent	Small substitutions, insertions, and deletions	Search-and-replace editing
Recombinase-assisted PE (PASSIGE/PASTE)	pegRNA-installed landing pad	Prime editor plus serine recombinase	attB/attP landing sites	Precise large-cargo integration	Multi-kilobase pathway insertion

2.2. Expanding the Double-Strand Break (DSB)-Free Precision Toolkit

While classical Cas9 and Cas12a nucleases remain effective for generating target gene knockouts via error-prone Non-Homologous End Joining (NHEJ), the reliance on double-strand breaks introduces significant cellular toxicity, large genomic deletions, chromosome loss, and unpredictable insertion-deletion (indel) profiles (Tou & Kleinstiver, 2023). In industrial strain engineering where metabolic pathway optimization requires fine-tuning gene expression, altering single catalytic residues, or inserting multi-gene cassettes DSB generation often severely reduces cell viability, imposing severe selection bottlenecks during high-throughput screening. To overcome these operational constraints, synthetic biologists developed catalytic null or nickase derivatives of Cas enzymes that decouple target DNA recognition from DSB induction (Qi *et al.*, 2013).

2.2.1. Transcriptional Regulation via dCas9 and dCas12a (CRISPRi and CRISPRa)

Mutations within the catalytic domains of Cas9 (D10A in RuvC and H840A in HNH) yield dead Cas9 (dCas9), which retains gRNA-guided target DNA-binding affinity without inducing strand cleavage. When dCas9 targets the promoter region or coding strand of a gene, it sterically blocks the binding or elongation of RNA polymerase, causing targeted gene silencing known as CRISPR interference (CRISPRi) (McCarty *et al.*, 2020). Conversely, fusing dCas9 or dCas12a to transcriptional activation domains such as VP64, VPR, or p65 in eukaryotic systems; bacterial CRISPRa instead uses activators such as SoxS or the RNA polymerase omega subunit yields CRISPR activation (CRISPRa) platforms capable of upregulating specific endogenous metabolic enzymes. Combined CRISPRi and CRISPRa multiplexing enables dynamic, non-lethal rewiring of carbon flux by systematically turning down competing metabolic pathways while simultaneously boosting rate-limiting enzymatic steps (Bendixen *et al.*, 2023).

2.2.2. Base Editing Architectures

Base editors fuse a Cas9 nickase (nCas9-D10A) or a miniature Cas effector such as zmBE or Un1Cas12f1 to single-stranded DNA deaminases. Cytosine Base Editors (CBEs) employ cytidine deaminases (such as APOBEC or rAPOBEC1) to catalyze the hydrolytic deamination of target cytosine to uracil, which is subsequently read as thymine during DNA replication, resulting in an overall Cytosine-Guanine to Thymine-Adenine transition. Adenine Base Editors (ABEs) incorporate continuously evolved E. coli TadA variants (such as ABE8e) to deaminate target adenine to inosine, leading to an Adenine-Thymine to Guanine-Cytosine transition (Komor *et al.*, 2016; Gaudelli *et al.*, 2017; Richter *et al.,* 2020). Base editors operate efficiently within a narrow catalytic window of four to eight nucleotides within the protospacer without requiring donor DNA templates or double-strand break formation. This capability allows engineers to introduce precise stop codons, disrupt catalytic sites, or alter promoter elements with high editing efficiency and minimal cytotoxicity (Srinivasa & Escobar, 2025).

2.2.3. Prime Editing Mechanics

Prime editing represents a search-and-replace genome editing technology capable of installing all twelve possible base-to-base transitions and transversions, precise single- or multi-nucleotide insertions, and target deletions without DSBs or donor DNA. A prime editor complex comprises a Cas9 nickase fused to an engineered Reverse Transcriptase (RT) enzyme and a specialized prime editing guide RNA (pegRNA) (Anzalone *et al.*, 2019). The pegRNA is dual-functional: its 5' spacer directs the nCas9 complex to the target genomic locus, while its 3' extension contains a Primer Binding Site (PBS) and a Reverse Transcriptase Template (RTT). Upon target recognition, nCas9 nicks the PAM-containing DNA strand. The liberated 3' single-stranded genomic DNA hybridizes to the pegRNA Primer Binding Site, priming reverse transcription of the edit encoded within the Reverse Transcriptase Template domain (Anzalone *et al.*, 2019).

The evolution of prime editing systems has systematically resolved early efficiency bottlenecks:

- **PE2 Platforms:** Incorporate an engineered Moloney Murine Leukemia Virus reverse transcriptase containing five-point mutations (D200N, T306K, W313F, T330P, L603W) that enhance thermal stability and processivity.
- **PE3 Platforms:** Introduce an auxiliary standard guide RNA that directs nCas9 to nick the unedited non-target strand, biasing cellular DNA repair machinery to replace the unedited strand using the edited strand as a template, thereby elevating edit conversion rates (Anzalone *et al.*, 2019).
- **PEmax Systems:** Incorporate structural codon optimizations and catalytic alterations across both the nCas9 and RT domains, yielding higher editing rates in complex genomic architectures.
- **Engineered pegRNAs (epegRNAs):** Feature structured RNA motifs, such as the evoPreQ1 stem-loop, at the 3' terminus to protect the extension from exonucleolytic degradation, stabilizing the guide complex inside the nucleoplasm (Rambabu *et al.*, 2026).

2.2.4. Large-Cargo Integration via Recombinase-Assisted Systems

While prime editing excels at installing local modifications, its operational efficiency drops when cargo length exceeds 100 base pairs, limiting its ability to directly insert entire multi-gene metabolic operons. To overcome this payload ceiling, synthetic biologists integrated prime editing with site-specific large serine recombinases (LSRs) such as Bxb1 integrase (Yarnall *et al.*, 2023; Pandey *et al.*, 2025). In systems such as Prime Editing-Assisted Site-Specific Integrase Gene Editing (PASSIGE) and Programmable Addition via Site-Specific Targeting Elements (PASTE), dual pegRNAs or twinPE constructs install small, orthogonal attachment sites (attB or attP) into safe-harbor loci. Continuously evolved serine recombinases such as evoBxb1 or eeBxb1, engineered via phage-assisted continuous evolution then catalyze the directional integration of multi-kilobase donor plasmids exceeding 10 to 36 kilobases into the installed landing pads with integration efficiencies exceeding 30% (Pandey *et al.*, 2025).

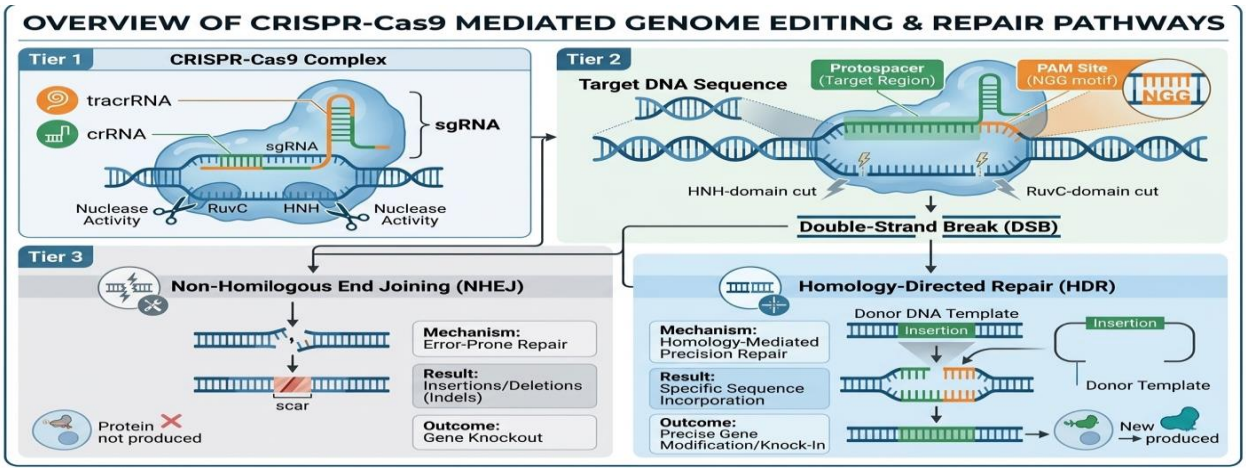


Figure 2. Mechanistic Overview of CRISPR-Cas9 Mediated DNA Double-Strand Cleavage and Endogenous Cellular Repair Pathways (NHEJ vs. HDR).

2.3. Microbial Genome Editing and Host-Specific Implementation

In Escherichia coli, Cas9 cleavage is frequently coupled to λ-Red recombineering: donor-mediated repair creates the desired allele while Cas9 counter-selects unedited cells. CRISPR interference using catalytically inactive Cas9 offers reversible repression without chromosome cleavage (Qi *et al.*, 2013). Cas12a-based bacterial activation and repression expand multiplex regulation options (Schilling *et al.*, 2020), while non-model bacterial gene regulation has been reviewed by Call and Andrews (2022). Signal-responsive CRISPRi switches can support dynamic metabolic control in E. coli (Kim *et al.*, 2025). Yeast editing often exploits efficient homologous recombination (DiCarlo *et al.*, 2013). Base editing and CRISPR-associated transposase systems offer alternative routes to nucleotide substitution and DNA insertion, but their efficiency is host- and locus-dependent (Klompe *et al.*, 2019; Strecker *et al.*, 2019). Mammalian prime-editing performance should not be assumed to transfer directly to industrial microorganisms.

3. Systems Biology Integration into the Design-Build-Test-Learn (DBTL) Framework

The ultimate utility of precision CRISPR editing depends on selecting appropriate genomic targets. Systems biology provides the analytical tools and computational models required to map global metabolic landscapes, enabling targeted genetic interventions through the iterative Design-Build-Test-Learn framework (Jeon *et al.*, 2025).

3.1. Design Phase: In Silico Predictive Modeling and Machine Learning

In the Design phase, computational biologists construct digital representations of the cellular host using Genome-Scale Metabolic Models (GEMs). Utilizing constraint-based reconstruction and analysis (COBRA) techniques, such as Flux Balance Analysis (FBA), GEMs simulate the complete biochemical reaction network of an organism under specific nutrient conditions. Algorithms such as OptKnock, OptForce, and RobustKnock evaluate thousands of simultaneous gene knockouts or overexpressions in silico to identify non-obvious target combinations that couple cell growth directly to target product formation (Ferreira, 2018).

Machine-learning models such as PRIDICT2.0 and OptiPrime predict prime-editing outcomes from pegRNA design and target-context features (Mathis *et al.*, 2024; Hsu *et al.*, 2026). PRIDICT2.0 predicts outcomes across diverse edit types and chromatin contexts, while OptiPrime models mechanistic determinants, including mismatch-repair effects (Mathis *et al.*, 2024; Hsu *et al.*, 2026). The mismatch-repair machinery, including MutSα (MSH2–MSH6) and MutSβ (MSH2–MSH3), recognizes mismatches in prime-editing heteroduplex intermediates and can reduce editing outcomes (Iyer *et al.*, 2006; Ferreira da Silva *et al.*, 2022). Strategy-specific silent or benign changes encoded in the pegRNA reverse-transcription template have been reported to reduce mismatch-repair susceptibility and improve editing in tested contexts (Chen *et al.*, 2021; Hsu *et al.*, 2026).

3.2. Build Phase: High-Throughput Assemblies and Multiplex Genome Reconstruction

In the Build phase, computational designs are converted into physical genetic edits. Automated biofoundries utilize liquid handling platforms and acoustic droplet ejection systems to construct complex guide RNA libraries and donor DNA assemblies. Technologies such as Multiplex Automated Genome Engineering (MAGE) and multiplexed CRISPR systems (CRISPRm) enable simultaneous targeting of dozens of genomic loci in a single transformation step (Bao *et al.*, 2016).

Using arrays of sgRNAs expressed from polycistronic tRNA-gRNA or Csy4-cleavable transcripts, engineered strains can undergo parallel transcript silencing via CRISPRi, gene activation via CRISPRa, and single-nucleotide conversions via base editors. This multiplex capacity enables metabolic engineers to reconfigure entire metabolic networks, bypassing the sequential, single-gene modifications that previously bottlenecked strain development (Ma *et al.*, 2024).

3.3. Test Phase: High-Throughput Phenotyping and Metabolite Analysis

The Test phase measures whether engineered strains exhibit the predicted phenotype (Abbate *et al.*, 2023). Screening can combine microplate cultivation, fluorescence or metabolite-responsive biosensors, flow cytometry, and microfluidic droplet assays to evaluate large strain libraries (Zhang *et al.*, 2023; O'Connor *et al.*, 2024). Promising variants require confirmation by quantitative chromatography or mass spectrometry, followed by shake-flask and bioreactor measurements (O'Connor *et al.*, 2024; Antoniewicz, 2021). Product titer (g/L), yield (g product/g substrate), volumetric productivity (g/L/h), growth rate, substrate uptake, and genetic stability should be recorded together (Abbate *et al.*, 2023; Antoniewicz, 2021). Isotope-tracing experiments and metabolomics can reveal pathway bottlenecks and support refinement of the next Design phase (Antoniewicz, 2021). Measurements must be linked to strain genotype, culture medium, cultivation time, and scale so that results are comparable across DBTL cycles (Abbate *et al.*, 2023).

3.4. Learn Phase: Artificial Intelligence and Data-Driven Optimization

In the Learn phase, computational models process quantitative phenotyping datasets generated in the Test phase to re-train operational algorithms. Machine learning methods including Random Forests, Convolutional Neural Networks (CNNs), and Gaussian Process Regression identify non-linear relationships between genetic modifications and observed titers (Jebreili & Goli, 2024).

When empirical phenotypic performance deviates from GEM-based predictions, the Learn phase uses constraint-based flux analysis to update flux constraints and intervention targets; enzyme kinetic parameters (Vmax, Km) require separate kinetic models, while enzyme-constrained models may incorporate kcat values. The updated models generate optimized genetic design parameters for the next DBTL iteration. This closed-loop operational architecture systematically narrows the search space, transforming industrial strain optimization into a predictive science (Anis Lahoud *et al.*, 2025).

4. Industrial Chassis Engineering and Quantitative Bio-Manufacturing Benchmarks

Systems metabolic engineering coupled with advanced CRISPR/Cas editing has driven substantial enhancements in production metrics across industrial microbial hosts. The following sub-sections examine specific engineering strategies executed across major host platforms (Saranya, 2026).

4.1. Escherichia coli: Terpenoid, Nucleoside, and Organic Acid Biofoundries

Escherichia coli remains a primary bacterial host due to its rapid doubling time, well-characterized physiology, and extensive genetic toolkit (Singh, 2021).

4.1.1. De Novo β-Thymidine Biosynthesis

β-Thymidine is a primary nucleoside precursor for critical antiviral therapeutics. In a study by Fan *et al.*, an industrial strain was developed through comprehensive systems engineering starting from an orotate-accumulating base strain (Ora14-4). Engineers integrated the pyrFE gene cluster from Bacillus subtilis to direct pyrimidine flux toward de novo deoxyuridine monophosphate (dUMP) synthesis. To eliminate precursor loss, genes involved in β-thymidine catabolism and competing branch points were knocked out using CRISPR/Cas9. The intracellular supply of one-carbon donors was enhanced by upregulating the endogenous folate cycle and the glycine cleavage system. To meet the high demand for reduction equivalents, the Entner-Doudoroff (ED) pathway from Zymomonas mobilis was introduced, augmenting intracellular NADPH pools (Fan *et al.*, 2026). Finally, thymidylate kinase (tmk) expression was placed under the control of a growth-adaptive promoter to minimize metabolic drift during stationary growth. In a fed-batch bioreactor, the engineered strain achieved a final β-thymidine titer of 25.4 g/L and a glucose conversion rate of 0.079 g/g (Fan *et al.*, 2026).

4.2. Saccharomyces cerevisiae and Non-Conventional Yeasts: Complex Molecule Biosynthesis

The budding yeast Saccharomyces cerevisiae and non-conventional oleaginous yeasts such as Yarrowia lipolytica serve as industrial eukaryotic chassis due to their robust tolerance to low pH, high osmotic pressure, and complex eukaryotic membrane-bound enzyme systems (Lacerda *et al.*, 2020).

4.2.1. S-Adenosyl-L-Methionine (SAM) Hyper-Production

The resulting strain (AU18) achieved a SAM titer of 2.46 g/L in shake flasks and 13.96 g/L in a 5-L fermenter under continuous L-methionine feeding (Lv *et al.*, 2024).

4.2.2. Specialized Metabolites and Terpenoids

Metabolic pathways supplying precursor amino acids (histidine, methionine, and cysteine) were rationally engineered, yielding an ergothioneine concentration of 385.70 mg/L (Tetali, 2019).

Table 2. Quantitative Bio-Manufacturing Performance Benchmarks in Systems-Engineered Microbial Platforms

Host	Product	Engineering strategy	Scale / mode	Titer / yield	Reference
Escherichia coli	β-Thymidine	B. subtilis pyrFE insertion; folate/glycine upregulation; NADPH regeneration; dynamic tmk control	Fed-batch fermenter	25.4 g/L; 0.079 g/g glucose	Fan *et al.*, 2026
Escherichia coli	Hydroxytyrosol (HT)	TPL, aadL, KAD, yahK, and GDH enzyme-cascade engineering	5-L bioreactor; fed-batch biotransformation	55.3 mM (≈8.525 g/L)	Xiong *et al.*, 2025
Escherichia coli	Hydroxytyrosol (HT)	EcHpaBC-mediated hydroxylation of exogenous tyrosol	Shake-flask conversion; 4 g/L tyrosol feed	2.02 g/L	Jiang *et al.*, 2026
Escherichia coli	3-Hydroxypropionic Acid (3-HP)	Heterologous pathway under a self-inducing dynamic promoter	1-L bioreactor (Ethanol feed)	3.47 g/L	Wang *et al.*, 2026
Escherichia coli	Ergothioneine (EGT)	egtBDE expression; solubility tags; precursor balancing	Optimized shake flask	385.70 mg/L	Tetali, 2019
Saccharomyces cerevisiae	S-Adenosyl-L-Methionine (SAM)	hxk2, aat1, met17, sam2, and vgb engineering; L-threonine-pathway weakening	5-L fermenter (Continuous L-Met)	13.96 g/L; 2.46 g/L shake flask	Lv *et al.*, 2024

Host	Product	Engineering strategy	Scale / mode	Titer / yield	Reference
Yarrowia lipolytica	Squalene	Mevalonate-pathway deregulation; lipid-droplet expansion	Bioreactor culture	High-density accumulation reported	Xu *et al.*, 2024
Corynebacterium glutamicum	L-Lysine / Specialty Amines	Pentose-phosphate flux enhancement; competing-pathway knockouts	Industrial fermenter	Multi-gram platform metrics	Liu *et al.*, 2022; Nguyen *et al.*, 2015

5. Computational Frontiers and Machine Learning in Genome and Pathway Design

5.1. Algorithmic Optimization of Precision Editors

The parameter space governing genome-editing tools, particularly multi-component platforms such as prime editors, presents significant optimization challenges.

Mechanistic machine-learning models such as OptiPrime and PRIDICT2.0 address the challenge of predicting prime-editing outcomes. OptiPrime was developed using two high-throughput lentiviral libraries—Lib-MMR and Lib-CV—comprising 10,000 and 10,406 pegRNA–target pairs, respectively. It captures determinants of reverse-transcription priming, template hybridization, and DNA-repair interactions (Hsu *et al.*, 2026).

Mismatch-repair proteins, including MutSα (MSH2–MSH6) and MutSβ (MSH2–MSH3), recognize mismatches or small insertion–deletion loops in prime-editing heteroduplex intermediates; mismatch repair can suppress prime-editing efficiency in a context-dependent manner (Iyer *et al.*, 2006; Ferreira da Silva *et al.*, 2022). OptiPrime identifies MMR-related sequence determinants and can nominate MMR-evasive silent edits encoded in the reverse-transcription template. Such edits have improved prime-editing efficiency in the reported experimental contexts (Chen *et al.*, 2021; Hsu *et al.*, 2026).

6. Bottlenecks, Delivery Systems, and Future Directions

6.1. Intracellular Repair Machinery and Mismatch Repair Constraints

In most bacteria, canonical non-homologous end joining is absent or inefficient; Cas9-induced double-strand breaks are often lethal and can be used for counter-selection with recombineering. In Saccharomyces cerevisiae, targeted double-strand breaks can enhance donor-mediated homologous recombination (DiCarlo *et al.*, 2013).

6.2. Delivery Vectors and Payload Capacity Challenges

Delivering advanced genome editing machinery into industrial hosts presents significant biophysical hurdles. Prime editors, base editors, and recombinase-assisted platforms exceed the packaging capacity of standard viral vectors, such as Adeno-Associated Virus (AAV), which is limited to less than 4.7 kilobases (Huang *et al.*, 2025).

To overcome packaging bottlenecks, advanced delivery modalities are being implemented:

- **High-Capacity Adenoviral Vector Particles (AdVPs)**: Possess extensive packaging capacities up to 36 kilobases, enabling the delivery of full-length prime editors, dual pegRNAs, and donor templates within a single vector. AdVPs efficiently transduce both cycling and post-mitotic cells without genomic integration, providing a robust delivery platform for large synthetic constructs (Aliciaslan *et al.*, 2025).
- **Split-Intein Delivery Architectures**: Dual-vector delivery platforms split large prime editor proteins into N-terminal and C-terminal fragments fused to fast-splicing split inteins, which reconstitute full-length functional enzymes upon co-entry into the target cell nucleoplasm.

- **Non-Viral Nanoparticles:** Lipid nanoparticles (LNPs) engineered with cell-targeting ligands facilitate transient, non-integrative delivery of Cas mRNA and pegRNA, minimizing off-target activity and cellular toxicity (Chen & Ping, 2023).

6.3. Scale-Up Dynamics and Metabolic Burden Management

Moving engineered strains from laboratory-scale shake flasks to multi-thousand-liter industrial fermenters exposes cells to stresses, including oxygen gradients, nutrient limitation, pH fluctuations, and osmotic pressure. Overexpression of large heterologous pathways can consume ATP, NAD(P)H, and amino acid pools, imposing a metabolic burden that may compromise genetic stability (Hollinshead *et al.*, 2014). In engineered Escherichia coli, sustained production load has been shown to select for non-producing mutants during prolonged fermentation (Rugbjerg *et al.*, 2018).

Mitigating these effects requires integrating dynamic metabolic control networks. Biosensor-controlled CRISPRi circuits dynamically suppress non-essential pathways during high-stress growth phases, balancing cellular maintenance with target product synthesis. Furthermore, integrating growth-adaptive promoters ensures that target gene expression automatically scales with biomass accumulation, stabilizing high-titer production during long industrial fermentation runs (Wu *et al.*, 2016).

7. CONCLUSION

The integration of CRISPR/Cas-based genome editing with systems biology has fundamentally redefined the capabilities of industrial biotechnology, transforming strain engineering from an empirical craft into a predictive, data-driven discipline. This review has demonstrated that the synergistic application of precision editing tools—spanning classical nucleases, DSB-free base and prime editors, and recombinase-assisted large-cargo integration systems—with computational frameworks such as genome-scale metabolic models and machine learning algorithms enables unprecedented control over cellular metabolism. The Design-Build-Test-Learn cycle operationalizes this integration, allowing engineers to systematically identify non-obvious genetic targets, execute multiplexed genomic modifications with nucleotide-level precision, and iteratively refine predictive models through quantitative phenotyping. Quantitative benchmarks across diverse microbial chassis, including record titers of β-thymidine, S-adenosyl-L-methionine, and ergothioneine, validate the practical efficacy of this integrated approach. Nevertheless, significant challenges persist, including host mismatch repair constraints that limit editing efficiency, delivery vector payload restrictions for large multi-component editor complexes, and metabolic burden that triggers genetic instability during industrial scale-up. Addressing these bottlenecks will require continued innovation in MMR-evasive guide RNA design, high-capacity delivery modalities such as adenoviral vectors and split-intein architectures, and dynamic metabolic control circuits that couple target gene expression to biomass accumulation. As machine learning models become increasingly sophisticated in predicting editing outcomes and metabolic flux distributions, the field is poised to achieve fully autonomous, closed-loop strain optimization, accelerating the development of sustainable bio-manufacturing platforms for pharmaceuticals, chemicals, biofuels, and nutraceuticals.

REFERENCES

Abbate, E., Andrion, J., Apel, A., *et al.* (2023). Optimizing the strain engineering process for industrial-scale production of bio-based molecules. Journal of Industrial Microbiology & Biotechnology, 50(1), kuad025. https://doi.org/10.1093/jimb/kuad025

Aliciaslan, M., Erbasan, E., Erendor, F., & Sanlioglu, S. (2025). Prime editing: the next frontier in precision gene therapy. The Journal of Gene Medicine, 27(10), e70040.

Anis Lahoud, A., Khan, A. S., Schaffernicht, E., Trincavelli, M., & Stork, J. A. (2025). Predict-and-Optimize Techniques for Data-Driven Optimization Problems: A Review: A. Anis Lahoud *et al.* Neural Processing Letters, 57(2), 40.

Antoniewicz, M. R. (2021). A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications. Metabolic Engineering, 63, 2–12. https://doi.org/10.1016/j.ymben.2020.11.002

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature, 576(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Bao, Z., Cobb, R. E., & Zhao, H. (2016). Accelerated genome engineering through multiplexing. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 8(1), 5-21.

Bendixen, L., Jensen, T. I., & Bak, R. O. (2023). CRISPR-Cas-mediated transcriptional modulation: The therapeutic promises of CRISPRa and CRISPRi. Molecular Therapy, 31(7), 1920-1937.

Bourgade, B., Xie, H., Lindblad, P., & Stensjö, K. (2025). Development of a CRISPR activation system for targeted gene upregulation in Synechocystis sp. PCC 6803. Communications Biology, 8(1), 772.

Call, S. N., & Andrews, L. B. (2022). CRISPR-based approaches for gene regulation in non-model bacteria. Frontiers in genome editing, 4, 892304.

Chen, P. J., Hussmann, J. A., Yan, J., Knipping, F., Ravisankar, P., Chen, P. F., Chen, C., Nelson, J. W., Newby, G. A., Sahin, M., Osborn, M. J., Weissman, J. S., Adamson, B., & Liu, D. R. (2021). Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell, 184(22), 5635–5652.e29. https://doi.org/10.1016/j.cell.2021.09.018

Chen, Y., & Ping, Y. (2023). Development of CRISPR/Cas delivery systems for in vivo precision genome editing. Accounts of Chemical Research, 56(16), 2185-2196.

DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., & Church, G. M. (2013). Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Research, 41(7), 4336–4343. https://doi.org/10.1093/nar/gkt135

Fan, W., Li, C., Zhao, R., Shi, T. E., Li, L., Chen, Z., ... & Xu, Q. (2026). Systems Metabolic Engineering of Escherichia coli for High-Titer De Novo β Thymidine Production via Folate Cycle Enhancement and NADPH Regeneration. Journal of Agricultural and Food Chemistry, 74(22), 17220-17236.

Ferreira, A. S. A. (2018). Development of novel strains for the production of biofuels (Doctoral dissertation, Universidade do Minho (Portugal)).

Ferreira da Silva, J., Oliveira, G. P., Arasa-Verge, E. A., Kagiou, C., Moretton, A., Timelthaler, G., Jiricny, J., & Loizou, J. I. (2022). Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. Nature Communications, 13, 760. https://doi.org/10.1038/s41467-022-28442-1

Gaj, T., Gersbach, C. A., & Barbas, C. F., III. (2013). ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology, 31(7), 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature, 551(7681), 464–471. https://doi.org/10.1038/nature24644

Hollinshead, W., He, L., & Tang, Y. J. (2014). Biofuel production: An odyssey from metabolic engineering to fermentation scale-up. Frontiers in Microbiology, 5, 344. https://doi.org/10.3389/fmicb.2014.00344

Hsu, A., Chen, P. J., Li, A. H., Hemez, C. F., Gao, X. D., Terrey, M., Nelson, C., Selvam, V., Cristian, A., McElroy, A. N., Steinbeck, B. J., Mahadeshwar, G. K., Pandey, S., Barsdale, Z., Chen, P. Z., Sousa, A. A., Sakai, H. A., Silverstein, R. A., Morad, I., Krueger, R. K., Shen, M. W., Kleinstiver, B. P., Lutz, C. M., Tolar, J., Blazar, B. R., Osborn, M. J., & Liu, D. R. (2026). Mechanistic machine learning for prediction of prime editing outcomes. Nature Biotechnology. https://doi.org/10.1038/s41587-026-03261-7

Huang, J., Li, J., Xu, X., & Li, K. (2025). Adeno-associated virus vectors in retinal gene therapy: challenges, innovations, and future directions. Biomolecules, 15(7), 940.

Iyer, R. R., Pluciennik, A., Burdett, V., & Modrich, P. L. (2006). DNA mismatch repair: functions and mechanisms. Chemical reviews, 106(2), 302-323.

Jebreili, S., & Goli, A. (2024). Optimization and computing using intelligent data-driven approaches for decision-making: optimization applications, 90(4), 1-12.

Jeon, S., Sohn, Y. J., Lee, H., Park, J. Y., Kim, D., Lee, E. S., & Park, S. J. (2025). Recent advances in the Design-Build-Test-Learn (DBTL) cycle for systems metabolic engineering of Corynebacterium glutamicum. Journal of Microbiology, 63(3), e2501021.

Jiang, W., Chen, Z., Fan, W., Li, C., Li, L., Yu, Z., Li, X., & Xu, Q. (2026). Systems metabolic engineering for hydroxytyrosol production in Escherichia coli. Applied and Environmental Microbiology, 92(5), e02455-25. https://doi.org/10.1128/aem.02455-25

Kim, S. K., Woo, S. G., Park, J. H., Lee, S. G., & Lee, D. H. (2025). Modular and signal-responsive transcriptional regulation using CRISPRi-aided genetic switches in Escherichia coli. Journal of Biological Engineering, 19(1), 56.

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. Nature, 571(7764), 219–225. https://doi.org/10.1038/s41586-019-1323-z

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature, 533(7603), 420–424. https://doi.org/10.1038/nature17946

Lacerda, M. P., Oh, E. J., & Eckert, C. (2020). The model system Saccharomyces cerevisiae versus emerging non-model yeasts for the production of biofuels. Life, 10(11), 299.

Li, Z., Khan, W. U., Bai, G., Dong, C., Wang, J., Zhang, Y., ... & Chen, F. (2025). From Code to Life: The AI-Driven Revolution in Genome Editing. Advanced Science, 12(30), e17029.

Liu, J., Xu, J.-Z., Rao, Z.-M., & Zhang, W.-G. (2022). Industrial production of L-lysine in Corynebacterium glutamicum: Progress and prospects. Microbiological Research, 262, 127101. https://doi.org/10.1016/j.micres.2022.127101

Lv, Y., Chang, J., Zhang, W., Dong, H., Chen, S., Wang, X., ... & Xu, P. (2024). Improving microbial cell factory performance by engineering SAM availability. Journal of Agricultural and Food Chemistry, 72(8), 3846-3871.

Ma, Y., Zhang, Z., Jia, B., & Yuan, Y. (2024). Automated high-throughput DNA synthesis and assembly. Heliyon, 10(6).

Manchanda, P., & Suneja, Y. (2018). Genome editing for crop improvement: status and prospects. Biotechnologies of Crop Improvement, Volume 3: Genomic Approaches, 75-104.

Mathis, N., Allam, A., Tálas, A., Kissling, L., Benvenuto, E., Schmidheini, L., Schep, R., Damodharan, T., Balázs, Z., Janjuha, S., Ioannidi, E. I., Böck, D., van Steensel, B., Krauthammer, M., & Schwank, G. (2024). Machine learning prediction of prime editing efficiency across diverse chromatin contexts. Nature Biotechnology, 43(5), 712–719. https://doi.org/10.1038/s41587-024-02268-2

McCarty, N. S., Graham, A. E., Studená, L., & Ledesma-Amaro, R. (2020). Multiplexed CRISPR technologies for gene editing and transcriptional regulation. Nature communications, 11(1), 1281.

Nguyen, A. Q. D., Schneider, J., Reddy, G. K., & Wendisch, V. F. (2015). Fermentative production of the diamine putrescine: System metabolic engineering of Corynebacterium glutamicum. Metabolites, 5(2), 211–231. https://doi.org/10.3390/metabo5020211

Ojha, D., Kulkarni, R., & Rode, A. B. (2026). Employing synthetic biology to modulate metabolic flux: Current strategies and applications. Journal of Bioscience and Bioengineering.

O'Connor, E., Micklefield, J., & Cai, Y. (2024). Searching for the optimal microbial factory: high-throughput biosensors and analytical techniques for screening small molecules. Current Opinion in Biotechnology, 87, 103125. https://doi.org/10.1016/j.copbio.2024.103125

Pandey, S., Gao, X. D., Krasnow, N. A., McElroy, A., Tao, Y. A., Duby, J. E., Steinbeck, B. J., McCreary, J., Pierce, S. E., Tolar, J., Meissner, T. B., Chaikof, E. L., Osborn, M. J., & Liu, D. R. (2025). Efficient site-specific integration of large genes in mammalian cells via continuously evolved recombinases and prime editing. Nature Biomedical Engineering, 9(1), 22–39. https://doi.org/10.1038/s41551-024-01227-1

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell, 152(5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022

Rambabu, I., Baskar, G., Suliman, M., Saeed, M., Radhakrishnan, M., Balu, R., & Palaniyandi, T. (2026). Engineered CRISPR-Cas systems for transcriptional regulation and precision molecular diagnostics: advances, challenges, and emerging microfluidic integration. Clinica Chimica Acta, 121145.

Richter, M. F., Zhao, K. T., Eton, E., Lapinaite, A., Newby, G. A., Thuronyi, B. W., Wilson, C., Koblan, L. W., Zeng, J., Bauer, D. E., Doudna, J. A., & Liu, D. R. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nature Biotechnology, 38(7), 883–891. https://doi.org/10.1038/s41587-020-0453-z

Rugbjerg, P., Myling-Petersen, N., Porse, A., Sarup-Lytzen, K., & Sommer, M. O. A. (2018). Diverse genetic error modes constrain large-scale bio-based production. Nature Communications, 9, 787. https://doi.org/10.1038/s41467-018-03232-w

Saranya, N. (2026). Synthetic Biology Strategies for Sustainable and Carbon-Neutral Industrial Bioprocessing Using Genome-Scale Metabolic Modeling. Frontiers in Life Sciences Research, 10-20.

Saugandhika, S., & Jain, N. (2024). Evolution of Transgenic Technology: From Random Transgenesis to Precise Genome Editing. In Sustainable Agriculture Reviews: Animal Biotechnology for Livestock Production 4 (pp. 85-129). Cham: Springer Nature Switzerland.

Schilling, C., Koffas, M. A., Sieber, V., & Schmid, J. (2020). Novel prokaryotic CRISPR-Cas12a-based tool for programmable transcriptional activation and repression. ACS synthetic biology, 9(12), 3353-3363.

Singh, V. (Ed.). (2021). Microbial cell factories engineering for production of biomolecules. Academic Press.

Srinivasa, M. A., & Escobar, M. (2025). CRISPR-based transcriptional regulation: technologies, applications, and future directions. DNA, 5(4), 57.

Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., & Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. Science, 365(6448), 48–53. https://doi.org/10.1126/science.aax9181

Šviković, S. (2022). CRISPR Cas: From Bacterial Adaptive Immunity to the Swiss Army Knife of Drug Discovery. Genome Editing in Drug Discovery, 19-46.

Tetali, S. D. (2019). Terpenes and isoprenoids: a wealth of compounds for global use. Planta, 249(1), 1-8.

Tou, C. J., & Kleinstiver, B. P. (2023). Recent advances in double-strand break-free kilobase-scale genome editing technologies. Biochemistry, 62(24), 3493-3499.

Wang, Y., *et al.* (2026). Physiological transcriptome analysis of Escherichia coli grown on ethanol. Green Carbon. https://doi.org/10.1016/j.greenca.2026.03.008

White, A. A. (2026). EVOLUTION-GUIDED REDESIGN OF RNA-GUIDED THERAPEUTICS: ANCESTRAL CAS9 FOR IMMUNE EVASION AND PRECISION RNAI FOR REPEAT DISORDERS.

Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., & Koffas, M. A. (2016). Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. Trends in biotechnology, 34(8), 652-664.

Xiong, T., Li, X., Liu, W., Yue, H., Liu, J., Bai, D., Li, W., & Fan, G. (2025). Multienzyme cascade for synthesis of hydroxytyrosol via engineered Escherichia coli. Scientific Reports, 15, 471. https://doi.org/10.1038/s41598-024-84624-5

Xu, M., Yang, N., Pan, J., Hua, Q., Li, C. X., & Xu, J. H. (2024). Remodeling the homologous recombination mechanism of Yarrowia lipolytica for high-level biosynthesis of squalene. Journal of agricultural and food chemistry, 72(17), 9984-9993.

Yamagata, M. (2022). Programmable proteins: target specificity, programmability and future directions. SynBio, 1(1), 65-76.

Yang, L., & Lu, Q. (2025). Beyond cutting: CRISPR-driven synthetic biology toolkit for next-generation microalgal metabolic engineering. International journal of molecular sciences, 26(15), 7470.

Yarnall, M. T. N., Ioannidi, E. I., Schmitt-Ulms, C., Krajeski, R. N., Lim, J., Villiger, L., Zhou, W., Jiang, K., Garushyants, S. K., Roberts, N., Zhang, L., Vakulskas, C. A., Walker, J. A., Kadina, A. P., Zepeda, A. E., Holden, K., Ma, H., Xie, J., Gao, G., Foquet, L., Bial, G., Donnelly, S. K., Miyata, Y., Radiloff, D. R., Henderson, J. M., Ujita, A., Abudayyeh, O. O., & Gootenberg, J. S. (2023). Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. Nature Biotechnology, 41(4), 500–512. https://doi.org/10.1038/s41587-022-01527-4

Zhang, Z., Guo, Q., Wang, Y., & Huang, H. (2023). High-throughput screening of microbial strains in large-scale microfluidic droplets. Frontiers in Bioengineering and Biotechnology, 11, 1105277. https://doi.org/10.3389/fbioe.2023.1105277