Biocompounds

www.bic.reapress.com

Biocompd. Vol. 3, No. 3 (2026) 191–217.

Paper Type: Original Article

AI-Driven Strain Engineering for Enhanced Biocompound Production: From Genome Design to Industrial Bioprocessing

Homeira Saheli*

Department of Computer Engineering, Faculty of Electrical and Computer Engineering, Shahrood University of Technology, Shahrood, Iran; homeira.saheli@shahroodut.ac.ir.

Citation:

Received: 26 January 2026 *Revised: 07 February 2026* *Accepted: 13 May 2026*	Saheli, H. (2026). AI-driven strain engineering for enhanced Biocompound production: From genome design to industrial bioprocessing. *Biocompounds, 3*(3), 191-217.

Abstract

The increasing demand for high-value Biocompounds has accelerated the development of engineered microbial cell factories as sustainable platforms for biomanufacturing. However, conventional strain-engineering approaches remain constrained by the complexity of microbial metabolism, nonlinear genotype–phenotype relationships, limited experimental throughput, and difficulties in translating laboratory-scale performance into industrial production. Artificial Intelligence (AI) and Machine Learning (ML) are emerging as powerful approaches for addressing these challenges by enabling data-driven prediction, optimization, and decision-making across multiple stages of Biocompound production. This review provides an integrated overview of AI-driven strain engineering, focusing on the convergence of multi-omics analysis, metabolic modeling, genome engineering, synthetic biology, and bioprocess optimization. The applications of AI in genomics, transcriptomics, proteomics, and metabolomics are discussed in the context of identifying metabolic bottlenecks and predicting genotype–phenotype relationships. Particular attention is given to AI-assisted metabolic pathway discovery, genome-scale metabolic modeling, metabolic flux prediction, precursor and cofactor engineering, CRISPR-based strain design, regulatory-element optimization, and enzyme engineering. The review further examines the use of AI for predicting strain performance, optimizing fermentation conditions, implementing adaptive feeding strategies, monitoring bioprocesses, and facilitating scale-up. The integration of these approaches within Design–Build–Test–Learn (DBTL) cycles is highlighted as a promising framework for accelerating iterative strain development and reducing experimental burden. Finally, current challenges related to data quality, model interpretability, experimental validation, transferability, and industrial implementation are discussed, together with emerging opportunities offered by foundation models, generative AI, digital twins, automated biofoundries, and self-driving laboratories. The convergence of AI and biotechnology is expected to enable increasingly predictive, efficient, and scalable platforms for next-generation Biocompound production.

Keywords: Artificial intelligence, Machine learning, Strain engineering, Metabolic engineering, Multi-omics, Synthetic biology, Biocompound.

1 | Introduction

Biocompounds encompass a broad range of biologically derived molecules, including pharmaceuticals, enzymes, organic acids, pigments, bioactive metabolites, and other high-value compounds with applications

Corresponding Author: homeira.saheli@shahroodut.ac.ir

https://doi.org/10.48313/bic.vi.76

in healthcare, food, agriculture, materials, and industrial biotechnology. The increasing demand for sustainable and economically viable production of these compounds has stimulated the development of microbial cell factories as versatile platforms for biomanufacturing. Compared with conventional chemical synthesis or extraction from natural sources, engineered microorganisms can provide controlled, scalable, and potentially more sustainable routes for producing complex molecules from renewable feedstocks. However, achieving commercially relevant titers, rates, and yields remains challenging because Biocompound biosynthesis is governed by highly interconnected metabolic networks and complex regulatory mechanisms [1–5].

Metabolic engineering and synthetic biology have substantially expanded the ability to redesign microbial hosts to produce value-added compounds. Conventional strain-engineering strategies have employed gene deletion, overexpression, pathway reconstruction, promoter engineering, enzyme modification, and metabolic flux redistribution to redirect cellular resources toward desired products. More recently, CRISPR-based genome editing, dynamic regulatory systems, high-throughput screening, and Genome-Scale Metabolic Models (GEMs) have enabled increasingly systematic manipulation of microbial metabolism. Nevertheless, designing high-performing production strains remains an iterative and resource-intensive process. The large number of possible genetic modifications, nonlinear interactions between metabolic pathways, cellular burden associated with heterologous pathway expression, and limited understanding of genotype phenotype relationships can make experimental optimization slow and difficult [2–7].

The rapid development of Artificial Intelligence (AI) and Machine Learning (ML) provides new opportunities to address these limitations. Modern biotechnological workflows generate increasingly large and heterogeneous datasets from genomics, transcriptomics, proteomics, metabolomics, phenomics, and high-throughput experimentation. AI-based approaches can identify patterns within these multidimensional datasets, predict biological functions and phenotypes, prioritize genetic targets, and support the reconstruction and optimization of metabolic pathways. In microbial production systems, AI has already been explored for genome annotation, enzyme and protein engineering, pathway prediction, metabolic modeling, strain selection, and prediction of production performance. These capabilities create the possibility of replacing parts of conventional trial-and-error strain development with data-driven design and rational prioritization of experimental candidates [8].

An important development in this context is the integration of AI with multi-omics and genome-scale metabolic modeling. Multi-omics datasets provide complementary information regarding the molecular state of engineered cells, whereas metabolic models provide a framework for interpreting how genetic and environmental perturbations influence intracellular fluxes and product formation. AI can strengthen this integration by learning complex relationships that may not be readily captured by purely mechanistic models. Hybrid approaches that combine biological knowledge with ML therefore offer a promising strategy for improving cell-factory design, identifying metabolic bottlenecks, predicting engineering targets, and exploring alternative pathway configurations. Such approaches are particularly relevant when the biological system contains nonlinear interactions or when experimental data are insufficient to establish explicit mechanistic relationships [6–10].

The impact of AI is also extending beyond the initial design of microbial strains toward iterative engineering workflows. The traditional Design–Build–Test (DBT) process is increasingly evolving toward Design–Build–Test–Learn (DBTL) and related predictive workflows, in which experimental results are systematically incorporated into subsequent rounds of design. Within this framework, AI can contribute to candidate generation, experimental prioritization, phenotype prediction, data interpretation, and optimization of successive engineering cycles. The integration of automated experimentation and high-throughput technologies further increases the amount of experimental information available for ML models, creating opportunities for increasingly efficient and partially autonomous strain-development workflows [9].

Importantly, strain engineering represents only one component of successful Biocompound manufacturing. Even a genetically optimized microorganism may exhibit poor performance when transferred from laboratory-scale experiments to industrial fermentation because of changes in oxygen transfer, mixing,

substrate availability, shear conditions, metabolic burden, and process control. Consequently, the future of AI-enabled biomanufacturing requires closer integration between biological design and bioprocess engineering. ML has already been investigated for strain selection, process optimization, monitoring, control, and scale-up, while recent developments in automated biofoundries and intelligent bioprocessing are creating opportunities to connect genetic design with experimental execution and process optimization [10–12].

Despite these advances, the current literature remains distributed across several partially overlapping areas, including AI-assisted microbial production, metabolic modeling, protein engineering, synthetic biology, fermentation optimization, and automated biofoundries. Recent reviews have addressed AI-aided microbial production and have highlighted applications in genome annotation, protein engineering, functional protein design, and pathway prediction. Other recent studies have focused specifically on AI-powered biofoundries, large language models, or AI-enhanced metabolic modeling. However, a comprehensive framework that connects AI-driven strain engineering with multi-omics analysis, metabolic pathway design, genome engineering, iterative DBTL cycles, and subsequent bioprocess optimization for Biocompound production remains particularly valuable [10–17].

Therefore, this review aims to provide an integrated and critical overview of AI-driven strain engineering for enhanced Biocompound production, with particular emphasis on the convergence of AI, multi-omics, metabolic modeling, genome engineering, and synthetic biology. The review examines how AI and ML can support the identification of genetic targets, prediction and optimization of metabolic pathways, design of engineered microbial strains, and evaluation of strain performance. It further discusses the integration of AI-driven strain development with fermentation and industrial bioprocess optimization, highlighting the transition from conventional experimental workflows toward data-driven and iterative DBTL systems. Finally, current limitations, including data quality and availability, model interpretability, biological validation, reproducibility, scalability, and industrial translation, are discussed, followed by emerging opportunities associated with foundation models, generative AI, automated biofoundries, digital twins, and increasingly autonomous biomanufacturing platforms.

2 | Biocompound Production and the Emergence of Artificial Intelligence-Driven Biomanufacturing

Biocompounds represent a diverse class of molecules produced by living organisms and include primary and secondary metabolites, recombinant proteins, enzymes, peptides, pigments, organic acids, antibiotics, vitamins, and other bioactive molecules. Their structural diversity and functional properties have generated considerable interest across pharmaceutical, food, agricultural, cosmetic, and environmental industries. While many Biocompounds can be obtained through extraction from plants, animals, or microorganisms, natural production systems frequently suffer from low product concentrations, seasonal or geographical variability, complex downstream processing, and limited availability of the desired compounds. Chemical synthesis can overcome some of these limitations but may become inefficient or environmentally unfavorable for structurally complex molecules. Consequently, microbial and cell-based production platforms have emerged as important alternatives for sustainable and scalable biomanufacturing [12].

2.1 | Microbial Cell Factories for Biocompound Production

Microbial cell factories provide a flexible platform for converting relatively inexpensive carbon sources into commercially valuable Biocompounds. Bacteria, yeasts, and filamentous fungi have been extensively investigated because of their short generation times, well-characterized genetic systems, and compatibility with industrial fermentation. Among bacterial hosts, Escherichia coli has been widely employed for recombinant protein and metabolite production because of its rapid growth and extensive genetic toolbox. Other microorganisms, including Corynebacterium glutamicum, Bacillus species, and Pseudomonas species, offer specific advantages for amino acid, organic acid, enzyme, and specialized metabolite production. Yeasts such as Saccharomyces cerevisiae and Yarrowia lipolytica have also become important hosts for producing

pharmaceuticals, terpenoids, lipids, organic acids, and other high-value compounds. Filamentous fungi provide additional opportunities for the biosynthesis of complex secondary metabolites and extracellular enzymes [8].

Selecting an appropriate microbial chassis is a critical determinant of production performance. Host physiology, precursor availability, cofactor balance, tolerance to toxic intermediates or products, genetic stability, and compatibility with the desired biosynthetic pathway can strongly influence final productivity. Consequently, the development of microbial cell factories increasingly involves rational modification of native metabolism rather than simple introduction of heterologous genes. Metabolic engineering can redirect carbon and energy toward the desired product by eliminating competing pathways, increasing precursor availability, modifying regulatory networks, and optimizing pathway enzyme expression. However, the high dimensionality of microbial metabolism makes it difficult to predict the consequences of multiple simultaneous genetic modifications [18].

2.2 | From Conventional Metabolic Engineering to Data-Driven Strain Design

Traditional strain development has largely followed an iterative experimental strategy in which candidate genetic modifications are introduced, production performance is experimentally measured, and the results are used to guide subsequent engineering cycles. Although this approach has generated numerous successful microbial production systems, it becomes increasingly inefficient as the number of potential genetic targets increases. A microbial genome may contain thousands of genes and numerous regulatory interactions, while a heterologous biosynthetic pathway can interact with native metabolism at multiple levels. As a result, the number of possible engineering combinations can rapidly become too large for exhaustive experimental testing [12], [14–19].

GEMs have provided an important computational framework for reducing this experimental search space. By representing metabolic reactions and their associated genes within a mathematical network, GEMs can be used to predict metabolic fluxes and identify potential genetic interventions. Methods such as Flux Balance Analysis (FBA), flux variability analysis, and related constraint-based approaches have been applied to identify candidate gene deletions, pathway modifications, and metabolic targets. Nevertheless, mechanistic models can be constrained by incomplete biological knowledge, inaccurate reaction stoichiometry, uncertain kinetic parameters, and limited representation of regulatory and environmental effects [15].

The increasing availability of biological datasets has therefore encouraged a transition toward data-driven and hybrid approaches. Genomic, transcriptomic, proteomic, metabolomic, and phenotypic data can provide complementary information about the molecular state of engineered microorganisms. ML can exploit these datasets to identify nonlinear relationships between genetic modifications and phenotypic outcomes, predict strain performance, and prioritize promising engineering strategies. Rather than replacing mechanistic metabolic models, AI can complement them by incorporating information that is difficult to describe using conventional mathematical formulations [12].

2.3 | The Emergence of Artificial Intelligence-Driven Biomanufacturing

AI has progressively expanded from a computational tool for data analysis into an enabling technology for biological design and bioprocess optimization. In biomanufacturing, AI can potentially contribute to several stages of the production pipeline, beginning with identification of suitable host organisms and genetic targets and extending to metabolic pathway design, strain optimization, fermentation control, and downstream process optimization. This broad applicability has created the concept of AI-driven biomanufacturing, in which biological knowledge, large-scale datasets, computational modeling, automated experimentation, and ML are integrated into a continuous production-development framework [19].

One important advantage of AI-based approaches is their ability to analyze heterogeneous datasets generated from different experimental platforms. For example, genomic information can indicate potential genetic targets, transcriptomic data can reveal changes in gene expression, proteomic data can provide information

about enzyme abundance, and metabolomic measurements can capture changes in intracellular and extracellular metabolites. Integrating these data types can provide a more comprehensive representation of cellular physiology than any individual omics layer. ML models can subsequently be trained to identify relationships between these molecular features and production phenotypes such as titer, yield, productivity, growth rate, and product tolerance.

AI can also facilitate the prioritization of genetic interventions. Instead of experimentally evaluating a large number of possible gene deletions, insertions, promoter modifications, or enzyme variants, predictive models can rank candidate interventions according to their expected impact on the desired phenotype. This capability is particularly important for metabolic engineering, where interactions between pathways may generate unexpected outcomes. AI-assisted prediction can therefore reduce the experimental search space and improve the efficiency of DBTL cycles [20].

2.4 | Integration of Artificial Intelligence with Fermentation and Bioprocessing

The value of an engineered production strain ultimately depends on its ability to perform under controlled and scalable bioprocess conditions. Fermentation parameters such as temperature, pH, dissolved oxygen, agitation, aeration, feeding rate, substrate concentration, and cultivation time can substantially affect microbial growth and product formation. Conventional optimization approaches often investigate these parameters sequentially or through predefined experimental designs. However, the interactions among process variables can be nonlinear and dynamic, making optimization challenging, particularly during scale-up.

ML provides opportunities to model these complex relationships using historical and real-time process data. Predictive models can be used to estimate biomass growth, substrate consumption, metabolite formation, and product concentration, while optimization algorithms can identify process conditions associated with improved productivity. When combined with online sensors and process analytical technologies, AI can also support real-time monitoring and adaptive control of fermentation processes. Consequently, AI-driven biomanufacturing extends beyond the design of a high-performing strain and encompasses the optimization of the biological system and its surrounding process environment [21–23].

2.5 | Toward Closed-Loop Biocompound Manufacturing

The convergence of AI, synthetic biology, metabolic engineering, automation, and bioprocess engineering is enabling a transition from isolated computational predictions toward closed-loop biomanufacturing systems. In a conventional workflow, strain design, experimental testing, and fermentation optimization are often treated as separate stages. In an AI-enabled workflow, information generated during experimental testing can be fed back into computational models, allowing subsequent strain designs and process conditions to be selected based on accumulated evidence [16].

This concept is closely associated with the DBTL framework. During the Design stage, AI and metabolic models can identify candidate genetic modifications and pathway configurations. During Build, synthetic biology and genome-editing technologies can construct the predicted strains. The Test stage generates high-throughput phenotypic and molecular data, while the Learn stage uses ML algorithms to extract patterns and update predictions. Repeated cycles can progressively improve strain performance while reducing the number of low-probability experimental candidates [15].

Such integration has the potential to transform Biocompound production from a predominantly trial-and-error discipline into a more predictive and iterative engineering process. However, successful implementation requires reliable biological datasets, standardized experimental measurements, interpretable computational models, robust experimental validation, and effective integration between computational and laboratory platforms. These requirements establish the foundation for the subsequent sections of this review, which examine in greater detail how AI can be applied to multi-omics analysis, metabolic pathway engineering, genome editing, strain optimization, and industrial bioprocessing [24]. As illustrated in *Fig. 1* and summarized in *Table 1*, AI-driven Biocompound production integrates multi-omics data, ML approaches, metabolic

modeling, genome and strain engineering, and bioprocess optimization to facilitate the development of high-performing microbial cell factories.

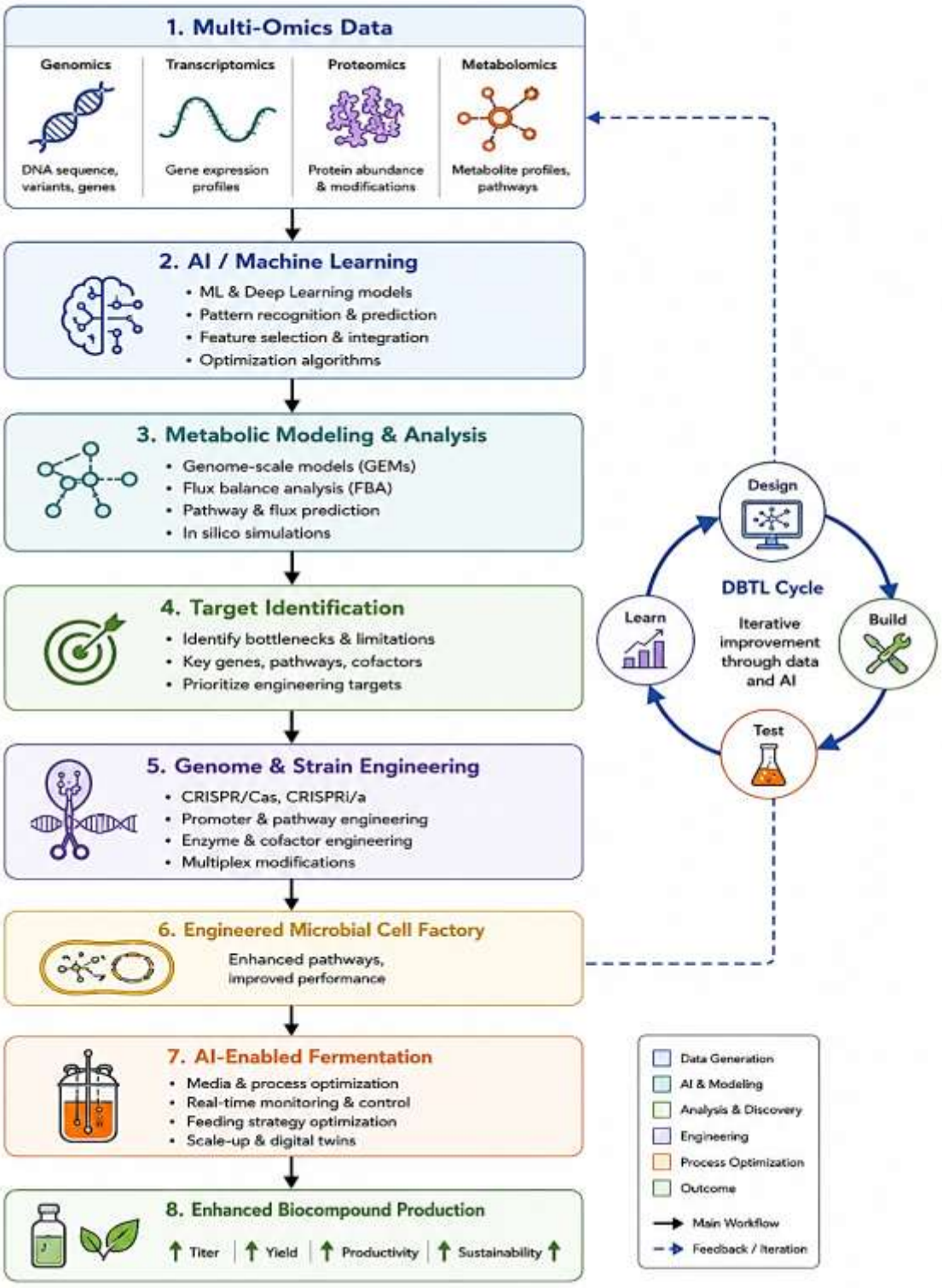

Fig. 1. Integrated framework of AI-driven strain engineering and bioprocess optimization for enhanced Biocompound production.

Table 1. AI and ML approaches for Biocompound production.

AI/ML Approach	Main Input Data	Major Application	Expected Output	Relevance to Biocompound Production
Linear and nonlinear regression	Process and phenotypic data	Production prediction	Titer, yield, productivity	Quantitative prediction of production performance
Random Forest (RF)	Genomic, omics, and phenotypic features	Feature selection and phenotype prediction	Key genes/Features	Identification of important engineering targets
Support Vector Machine	Genomic and molecular features	Strain classification and prediction	High/Low producer classification	Rapid screening of candidate strains
Artificial Neural Networks (ANNs)	Multi-omics and process data	Nonlinear phenotype prediction	Production phenotype	Modeling complex genotype–phenotype relationships
Deep Learning (DL)	Genomic, transcriptomic, proteomic, and sequence data	Pattern and sequence recognition	Functional predictions	Pathway, gene, and regulatory-element discovery
Bayesian Optimization (BO)	Strain and process parameters	Experimental optimization	Optimal parameter combinations	Reduction of experimental screening
Genetic Algorithms (GAs)	Metabolic and genetic design variables	Combinatorial optimization	Optimized strain designs	Identification of multi-gene engineering strategies
Reinforcement learning	Real-time process data	Adaptive process control	Control actions	Dynamic fermentation optimization
Generative AI	DNA, RNA, and protein sequences	Sequence and enzyme design	Novel candidate sequences	Synthetic biology and enzyme engineering
Hybrid mechanistic–AI models	GEMs + omics + experimental data	Constrained prediction and optimization	Biologically consistent predictions	Integration of metabolic knowledge with data-driven learning

3|Multi-Omics Data for Artificial Intelligence-Assisted Strain Engineering

The increasing complexity of microbial cell factories has created a need for comprehensive approaches capable of linking genetic information to cellular behavior and Biocompound production. Although individual omics technologies have provided valuable insights into specific molecular layers, the phenotype of an engineered microorganism is ultimately determined by interactions among its genome, gene expression programs, protein abundance and activity, metabolic state, and environmental conditions. Multi-omics approaches therefore provide a systems-level framework for understanding how genetic modifications propagate through cellular networks and ultimately influence the production phenotype. Integrating multi-omics data with AI and ML offers an opportunity to identify complex genotype–phenotype relationships and improve the rational design of microbial production strains [25].

3.1|Genomics as the Foundation for Artificial Intelligence-Assisted Strain Design

Genomic information provides the fundamental blueprint for identifying genes, regulatory elements, metabolic pathways, and potential engineering targets within a microbial host. Whole-genome sequencing and comparative genomics can reveal genetic differences among strains with distinct production phenotypes, while genome annotation can identify genes associated with biosynthetic pathways, transport systems, stress responses, and metabolic regulation. For engineered microorganisms, genomic analysis can also be used to verify genetic modifications and assess genome stability following repeated engineering cycles [26–28].

AI and ML methods can facilitate the analysis of large genomic datasets by identifying sequence patterns associated with specific phenotypes. Supervised learning approaches can be trained using genotype-phenotype datasets to predict whether particular genes, mutations, or regulatory elements are associated with enhanced production. Unsupervised learning can further identify clusters of genetically similar strains or previously unrecognized relationships among genomic features. DL models can also be applied to DNA sequence analysis and prediction of regulatory activity, potentially supporting the selection of promoters,

terminators, transcription-factor binding sites, and other genetic elements for synthetic biology applications [29].

However, genomic information alone does not necessarily indicate whether a gene is functionally active under a particular production condition. Consequently, genomic predictions must be interpreted in combination with downstream molecular and phenotypic information. This limitation provides a major rationale for integrating genomics with transcriptomics, proteomics, and metabolomics.

3.2|Transcriptomics and Artificial Intelligence-Based Identification of Regulatory Responses

Transcriptomics provides information about gene-expression patterns and enables researchers to investigate how microbial cells respond to genetic modifications, environmental changes, and production demands. RNA sequencing can reveal genes that are upregulated or downregulated during different growth phases or under specific fermentation conditions. In Biocompound production, transcriptomic analysis can help identify regulatory responses associated with precursor availability, pathway activation, metabolic stress, product toxicity, and cellular adaptation [30].

ML Algorithms can analyze transcriptomic datasets to identify expression signatures associated with high-producing and low-producing strains. Feature-selection methods can reduce thousands of gene-expression variables to a smaller subset of candidate genes that provide the strongest predictive information about production performance. These candidate genes can subsequently be evaluated experimentally through gene deletion, overexpression, promoter engineering, or other genome-editing strategies [29].

AI-based analysis may also reveal coordinated expression patterns among genes participating in the same metabolic pathway or regulatory network. Such information can help distinguish individual genes that merely correlate with high production from those that may represent functionally meaningful engineering targets. Nevertheless, transcriptomic correlations should not automatically be interpreted as causal relationships. Experimental validation remains necessary to determine whether modification of an identified gene produces the predicted improvement in Biocompound production.

3.3|Proteomics for Linking Gene Expression to Functional Phenotypes

Proteomics provides information that is closer to cellular function than genomic or transcriptomic measurements because proteins directly execute many metabolic and regulatory processes. Quantitative proteomics can reveal changes in enzyme abundance, transporter levels, stress-response proteins, and regulatory factors following genetic or environmental perturbations. This information is particularly important for metabolic engineering because increased transcript abundance does not necessarily result in proportional increases in functional enzyme levels [28].

AI can be used to integrate proteomic features with metabolic and production data to identify proteins associated with high-performing strains. ML models may rank enzymes according to their contribution to product formation or identify protein abundance patterns associated with metabolic bottlenecks. Furthermore, AI-assisted protein engineering can potentially predict sequence modifications that improve catalytic activity, substrate specificity, stability, or tolerance to industrial conditions [30].

Integrating proteomics with transcriptomics is particularly informative when discrepancies exist between RNA and protein abundance. Such discrepancies may indicate post-transcriptional regulation, protein degradation, translational limitations, or other regulatory mechanisms. Therefore, combining transcriptomic and proteomic information can improve the interpretation of how genetic modifications influence pathway activity.

3.4 | Metabolomics and Artificial Intelligence-Based Mapping of Metabolic States

Metabolomics provides a direct view of the biochemical state of a microbial cell by measuring intracellular and extracellular metabolites. Because metabolites represent the products and intermediates of biochemical reactions, metabolomic data can reveal changes in metabolic flux, precursor availability, accumulation of toxic intermediates, cofactor imbalance, and product-associated stress. For Biocompound production, metabolomics can be particularly valuable for identifying metabolic bottlenecks that are not readily apparent from genomic or transcriptomic data. For example, a pathway may contain sufficient transcript and enzyme abundance but still exhibit poor productivity because of precursor depletion, cofactor limitations, competing reactions, or accumulation of inhibitory intermediates. AI models can analyze complex metabolomic profiles and identify metabolite patterns associated with high- or low-production phenotypes [31].

ML approaches can also combine metabolomic information with fermentation data to predict final product concentration or productivity at earlier stages of cultivation. Early prediction of fermentation outcomes could allow low-performing cultures to be identified before substantial resources are invested in their continued cultivation. Such predictive capabilities may become increasingly valuable when integrated with automated fermentation and high-throughput screening platforms [30].

3.5 | Integrating Multi-Omics Layers for Genotype–Phenotype Prediction

The primary advantage of multi-omics integration is the ability to connect different levels of biological organization. Genomics describes the genetic potential of an organism, transcriptomics captures gene-expression responses, proteomics reflects the abundance of functional molecular machinery, and metabolomics provides information about the biochemical state of the cell. Integrating these layers can therefore provide a more complete representation of the relationship between genotype, cellular state, and production phenotype [32].

AI provides several strategies for integrating heterogeneous omics datasets. Early-integration approaches combine multiple data types into a single feature matrix before model training, whereas late-integration approaches independently analyze individual omics layers and subsequently combine their predictions. Intermediate or representation-learning approaches attempt to transform different data types into shared latent representations that capture biologically meaningful relationships. DL architectures, multimodal learning, network-based approaches, and other advanced ML methods can be particularly useful when the datasets contain large numbers of variables with complex nonlinear interactions.

For strain engineering, the ultimate objective of multi-omics integration is not simply to classify existing strains but to predict actionable interventions. An effective model should ideally identify which genes, enzymes, regulatory elements, or pathway modifications are most likely to improve titer, yield, productivity, or robustness. This transition from descriptive analysis toward predictive and prescriptive modeling represents an important step in the development of AI-driven microbial cell factories [31].

3.6 | Artificial Intelligence-Assisted Identification of Metabolic Bottlenecks and Engineering Targets

A major challenge in microbial strain engineering is determining which molecular components should be modified to improve production. Conventional approaches may rely on known pathway biology or systematic screening of candidate genes. Multi-omics datasets provide an opportunity to identify bottlenecks by examining coordinated changes across molecular layers.

For example, a reduction in precursor metabolites accompanied by increased expression of competing pathways may indicate a metabolic diversion that limits product formation. Similarly, high transcript and protein abundance combined with limited product formation may suggest downstream enzymatic constraints, cofactor limitations, or product inhibition. AI models can integrate these signals and rank potential

intervention points according to their predicted impact. Network-based ML can further incorporate information about gene–gene, protein–protein, metabolite–metabolite, and gene–metabolite interactions. By embedding these relationships into predictive models, it may become possible to identify targets that would be overlooked when each omics dataset is analyzed independently. Such approaches are particularly promising for complex pathways in which multiple small perturbations collectively determine the final production phenotype [32–34].

3.7 | Challenges in Multi-Omics-Based Artificial Intelligence Modeling

Despite its potential, multi-omics integration presents several technical and biological challenges. Omics datasets frequently differ in scale, dimensionality, measurement technology, sampling frequency, and experimental noise. The number of measured variables may greatly exceed the number of experimentally characterized strains, creating a high risk of overfitting. Differences in normalization procedures and batch effects can further complicate integration across experiments or laboratories.

Another important challenge is the interpretation of AI-derived associations. Highly predictive features are not necessarily causal biological determinants, and black-box models may provide limited mechanistic insight. Consequently, combining AI with prior biological knowledge, metabolic networks, and mechanistic models can improve interpretability and reduce biologically implausible predictions. Experimental validation remains essential for confirming whether computationally prioritized targets actually improve strain performance [32].

The development of standardized datasets and interoperable experimental workflows will therefore be critical for advancing AI-assisted strain engineering. High-quality genotype–phenotype datasets generated under well-defined experimental conditions can enable more reliable model training and facilitate comparison among different strains and production systems. In parallel, automated high-throughput experimentation can generate iterative datasets that continuously improve predictive models.

The integration of multi-omics with AI represents a transition from single-layer molecular analysis toward systems-level and predictive strain engineering. By connecting genomic potential with transcriptional regulation, protein function, metabolic state, and production phenotype, AI-enabled multi-omics approaches can help reduce the search space for genetic engineering and guide the design of more efficient microbial cell factories. The next step in this framework is to translate these system-level insights into specific metabolic interventions, which requires the integration of AI with metabolic pathway modeling and optimization. As summarized in *Table 2*, integrating different omics layers provides complementary information that enables AI models to move from individual molecular measurements toward a more comprehensive understanding of cellular metabolism and the identification of actionable strain-engineering targets [35].

Table 2. Multi-omics data and their applications in AI-assisted strain engineering.

Omics Layer	Major Information Obtained	AI/ML Application	Potential Engineering Target	Main Contribution
Genomics	Genes, mutations, gene clusters, regulatory regions	Gene-function prediction and genome mining	Candidate genes and biosynthetic clusters	Identification of genetic potential
Transcriptomics	Gene-expression patterns	Expression-state prediction and pathway analysis	Up/Downregulation targets	Identification of transcriptionally active pathways
Proteomics	Protein abundance and enzyme levels	Enzyme activity and pathway prediction	Rate-limiting enzymes	Evaluation of functional protein capacity
Metabolomics	Metabolite concentrations and metabolic states	Metabolic-state prediction and biomarker identification	Precursor and cofactor pathways	Identification of metabolic bottlenecks
Fluxomics	Intracellular metabolic fluxes	Flux prediction and pathway optimization	Flux redistribution targets	Quantification of pathway activity
Epigenomics	DNA methylation and chromatin-related regulation	Regulatory-state prediction	Regulatory elements	Understanding additional layers of gene regulation
Multi-omics integration	Combined molecular information	Integrated phenotype prediction	Multiple coordinated targets	Comprehensive characterization of cellular state

4 | Artificial Intelligence for Metabolic Pathway Discovery and Optimization

The successful production of a Biocompound in an engineered microorganism depends not only on the presence of the desired biosynthetic genes but also on the ability of the host cell to efficiently supply precursors, cofactors, energy, and reducing equivalents required for product formation. Consequently, metabolic pathway design represents one of the central challenges in microbial cell-factory engineering. Conventional metabolic engineering approaches have traditionally relied on biochemical knowledge, pathway databases, stoichiometric models, and experimental characterization to identify suitable engineering targets. However, the increasing complexity of cellular metabolism and the large number of possible pathway configurations have created a strong need for computational methods that can explore this design space more efficiently. AI, particularly ML, provides new opportunities to combine metabolic models with heterogeneous biological data and identify pathway configurations associated with improved Biocompound production [21].

4.1 | Genome-Scale Metabolic Models as a Framework for Pathway Engineering

GEMs provide a systematic representation of the metabolic capabilities of an organism by connecting genes, enzymes, metabolites, and biochemical reactions within a computational network. These models have become important tools for analyzing microbial metabolism and predicting the effects of genetic and environmental perturbations. By representing the stoichiometric relationships among metabolic reactions, GEMs can be used to investigate the redistribution of metabolic flux following gene deletion, pathway introduction, or changes in nutrient availability [20].

FBA is one of the most widely used constraint-based approaches for analyzing GEMs. In its conventional form, FBA estimates intracellular flux distributions by applying mass-balance constraints and optimizing a predefined objective function, commonly cellular biomass formation. For Biocompound production, the objective function can be modified or constrained to explore conditions that favor synthesis of the target product. Other approaches, including flux variability analysis and related constraint-based methods, can be used to investigate alternative feasible flux distributions and identify reactions that strongly influence the production phenotype.

Despite their utility, GEMs are constrained by the quality and completeness of the underlying biological information. Stoichiometric models generally do not fully capture kinetic regulation, enzyme abundance, transcriptional regulation, metabolite concentrations, or dynamic changes in cellular physiology. In addition, different objective functions can lead to substantially different predicted metabolic states. These limitations have encouraged the development of approaches that combine mechanistic metabolic modeling with data-driven AI methods [18].

4.2 | Artificial Intelligence-Assisted Metabolic Flux Prediction

Metabolic flux represents the rate at which metabolites are converted through biochemical reactions and provides a functional description of cellular metabolism. Experimental flux measurements can be obtained using approaches such as isotope labeling and metabolic flux analysis, but these experiments can be technically demanding and often have limited throughput. AI models can provide complementary approaches for estimating flux distributions from available omics, physiological, and metabolic data [36].

ML Algorithms can learn relationships between measurable molecular features and experimentally determined or model-derived fluxes. Features such as transcript abundance, protein levels, metabolite concentrations, growth rate, and environmental conditions can be incorporated into predictive models. Once trained, these models can estimate metabolic behavior under conditions that have not been directly characterized experimentally.

An important advantage of AI-based flux prediction is its potential to capture nonlinear relationships that may be difficult to represent using purely stoichiometric models. For example, the relationship between enzyme abundance and pathway flux may depend on substrate availability, cofactor balance, allosteric regulation, competing pathways, and cellular stress. Data-driven models can potentially learn these interactions when sufficiently representative training datasets are available. However, because predicted fluxes can be sensitive to dataset quality and model assumptions, AI-based predictions should ideally be constrained by biochemical knowledge and validated experimentally [31].

4.3 | Identification of Metabolic Bottlenecks

Metabolic bottlenecks occur when one or more reactions, metabolites, enzymes, or regulatory mechanisms limit the conversion of cellular resources toward the desired product. Identifying these bottlenecks is a fundamental objective of strain engineering. Conventional approaches often rely on pathway analysis, flux calculations, or experimental measurement of intermediate metabolites. AI can expand this process by integrating multiple sources of evidence and ranking potential bottlenecks according to their predicted contribution to production limitations.

For example, an ML model may identify a combination of low precursor availability, high activity of a competing pathway, and limited cofactor regeneration as predictors of poor product formation. Instead of treating these factors independently, AI can analyze their interactions and identify combinations of interventions that may produce a larger improvement than modification of any single target.

Feature-importance analysis, interpretable ML methods, and network-based approaches can further assist in identifying biologically relevant targets. When combined with GEMs, these methods can prioritize reactions whose modification is predicted to increase product flux while maintaining acceptable cellular growth. The resulting candidates can then be evaluated experimentally through gene deletion, overexpression, promoter modification, enzyme engineering, or pathway restructuring [33].

As illustrated in *Fig. 2*, AI-assisted metabolic analysis can integrate metabolic-network information and flux predictions to identify precursor limitations, cofactor imbalances, rate-limiting enzymes, and competing pathways, thereby guiding targeted interventions for metabolic pathway optimization.

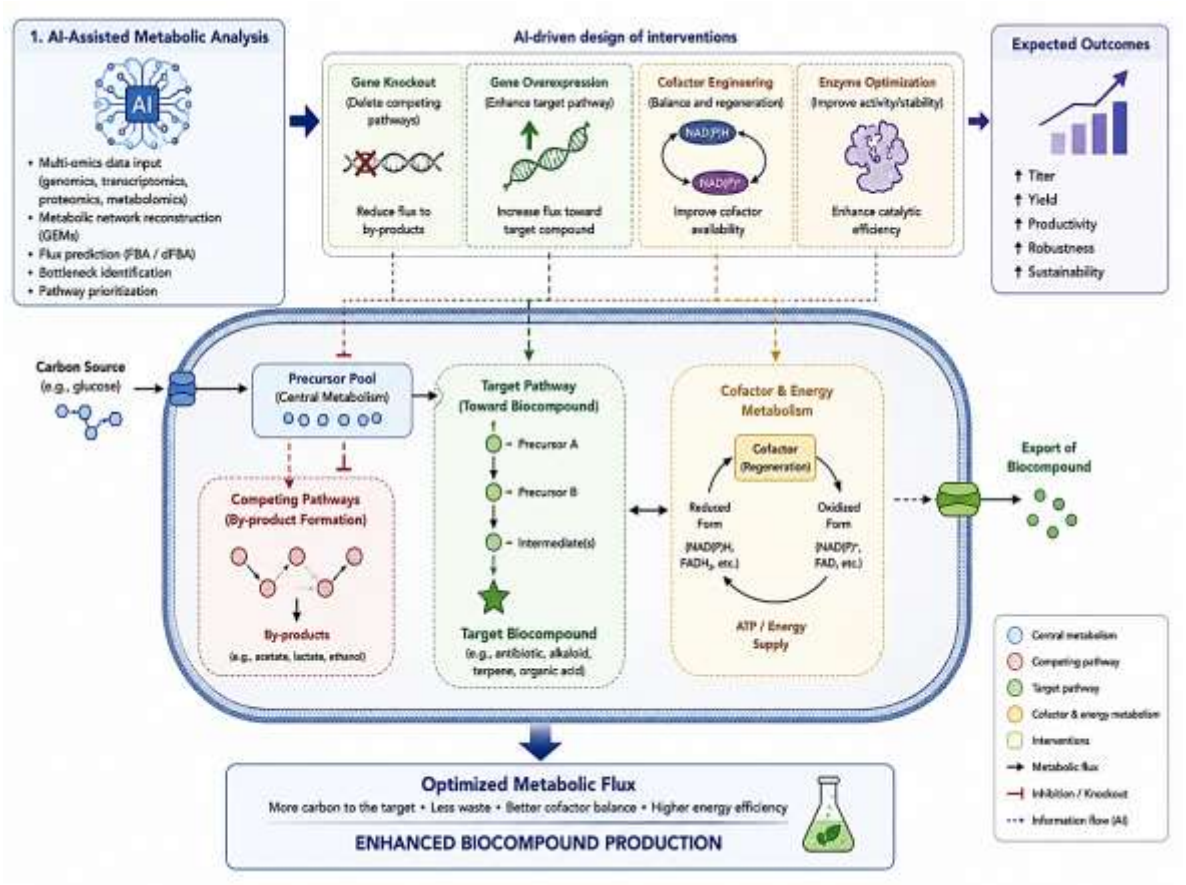

Fig. 2. AI-assisted metabolic pathway engineering for enhanced Biocompound biosynthesis.

4.4 | Artificial Intelligence-Assisted Metabolic Pathway Discovery

The discovery of biosynthetic pathways is another area in which AI can substantially accelerate Biocompound production. Many microorganisms contain biosynthetic gene clusters and metabolic capabilities that remain incompletely characterized. Genome-mining approaches can identify candidate genes and clusters based on sequence similarity, domain composition, genomic organization, and evolutionary relationships. AI and deep-learning methods can extend these approaches by recognizing complex sequence patterns and predicting functional relationships that may not be captured through conventional similarity-based methods.

For novel or poorly characterized Biocompounds, AI-assisted pathway discovery can integrate genomic information with metabolomic and phylogenetic data to predict candidate biosynthetic routes. The predicted pathways can then be reconstructed computationally and evaluated using metabolic models before experimental implementation. This strategy can reduce the number of candidate pathways requiring laboratory validation.

AI can also assist in the selection of heterologous hosts for pathway reconstruction. Different microbial chassis vary in their precursor pools, cofactor availability, intracellular transport systems, tolerance to toxic intermediates, and ability to express heterologous enzymes. A computational model that considers these characteristics can help identify host–pathway combinations with a higher probability of achieving efficient production [21].

4.5 | Precursor and Cofactor Engineering

Improving the supply of pathway precursors is often essential for increasing Biocompound production. Engineered pathways may compete with native metabolic processes for common intermediates, carbon skeletons, ATP, NADH, NADPH, or other cofactors. Excessive diversion of cellular resources toward product formation can also reduce growth and strain stability. Therefore, successful pathway optimization requires balancing product synthesis with host viability.

AI-based models can assist in identifying precursor-supplying reactions and predicting how modifications of upstream metabolism may influence product formation. By integrating metabolic flux information with transcriptomic, proteomic, and metabolomic data, AI can identify pathways that act as major sources or sinks of metabolic resources. Candidate interventions may include enhancing precursor synthesis, reducing competing pathways, modifying transporter activity, or engineering cofactor regeneration [36].

Cofactor engineering is particularly important for redox-intensive Biocompounds. If a biosynthetic pathway requires large quantities of NADPH or NADH, insufficient cofactor availability may constrain product formation even when substrate and enzyme levels are adequate. AI-assisted metabolic models can therefore evaluate alternative cofactor-generation strategies and identify combinations of genetic modifications that improve redox balance without imposing excessive metabolic burden [21–23].

4.6 | Multi-Objective Optimization of Metabolic Pathways

A major challenge in strain engineering is that maximum product formation does not necessarily correspond to maximum cellular fitness. Increasing product flux may reduce biomass formation, increase metabolic burden, or decrease strain stability. Consequently, pathway engineering often involves multiple competing objectives, including product titer, yield, productivity, growth rate, robustness, and resource efficiency.

AI and evolutionary optimization algorithms can be used to explore this multi-dimensional design space. Instead of optimizing a single parameter, multi-objective approaches can identify sets of solutions representing different trade-offs between product formation and cellular performance. GAs, Bayesian Optimization (BO), reinforcement learning, and other optimization strategies can be used to search combinations of gene modifications, pathway configurations, and process conditions [22].

For example, one candidate strain may achieve very high product titer but grow slowly, whereas another may provide moderate production with substantially faster growth. Rather than selecting a single theoretical optimum, multi-objective optimization can identify a Pareto frontier of feasible solutions, allowing researchers to select the strain that best matches the intended industrial application [23].

4.7 | Hybrid Mechanistic–Artificial Intelligence Models

The integration of mechanistic metabolic models with ML algorithms represents an important direction for AI-driven strain engineering. Purely data-driven models can capture complex nonlinear relationships but may require large datasets and can generate biologically implausible predictions outside their training domain. Conversely, mechanistic models incorporate biological constraints but may simplify regulatory and dynamic processes.

Hybrid models seek to combine the strengths of both approaches. Mechanistic models can provide constraints based on stoichiometry, thermodynamics, pathway connectivity, or known biochemical relationships, while AI models can learn residual relationships from experimental data. This combination can improve predictive performance while preserving biological consistency.

Such hybrid approaches are particularly promising when experimental datasets are relatively small, which is common in microbial strain engineering. Prior biological knowledge can reduce the effective search space available to the ML model, potentially improving generalization and reducing overfitting. At the same time, experimental data can be used to refine or correct assumptions embedded within mechanistic models [33].

4.8 | Artificial Intelligence-Guided Pathway Optimization within Design–Build–Test–Learn Cycles

The greatest value of AI-based metabolic pathway modeling may emerge when computational prediction is directly connected to experimental engineering. Within a DBTL framework, metabolic models and AI algorithms can be used during the design stage to identify candidate pathways and genetic interventions. Synthetic biology tools can then construct selected strains, while high-throughput assays generate data regarding growth, metabolite profiles, and product formation. These experimental results can subsequently be incorporated into the learning stage to update the predictive model.

Repeated DBTL cycles can progressively improve the accuracy of pathway predictions and reduce the number of experimental candidates required to identify high-performing strains. Importantly, this process transforms metabolic engineering from a largely sequential activity into an iterative learning system in which each engineering cycle contributes information to the next.

Integrating AI with metabolic modeling therefore has the potential to move pathway engineering from retrospective analysis toward predictive and prescriptive design. Rather than simply asking why a strain produces a particular amount of a Biocompound, future systems may increasingly be capable of predicting which combination of genetic and metabolic interventions is most likely to produce a desired phenotype under defined production conditions [30].

4.9 | Limitations and Future Opportunities

Despite significant progress, AI-assisted metabolic pathway optimization remains constrained by several factors. High-quality genotype–phenotype datasets are relatively scarce compared with the dimensionality of microbial metabolic networks. Experimental conditions may also vary substantially between laboratories, making datasets difficult to combine. Furthermore, metabolic models may contain incomplete pathway annotations, uncertain reaction reversibility, missing transport processes, or inaccurate assumptions regarding cellular objectives.

Another challenge is the biological validation of computational predictions. A model may identify a statistically strong target that fails to improve production when experimentally implemented because of unanticipated regulatory effects, metabolic burden, toxicity, or compensatory cellular responses. Therefore, computational predictions should be regarded as tools for prioritization rather than substitutes for experimental validation.

Future developments are likely to focus on more tightly integrated AI–mechanistic frameworks, improved representation of dynamic cellular states, incorporation of spatial and temporal information, and integration with automated experimentation. The emergence of foundation models, generative approaches, and autonomous laboratory platforms may further expand the ability of computational systems to generate and experimentally test increasingly complex metabolic designs.

AI-assisted metabolic pathway discovery and optimization provides a critical bridge between multi-omics information and practical strain engineering. By integrating GEMs, ML, pathway prediction, flux analysis, precursor and cofactor engineering, and multi-objective optimization, AI can help identify and prioritize metabolic interventions that would be difficult to discover through conventional approaches alone. The next challenge is to translate these computationally identified targets into precise genetic modifications, a process increasingly enabled by modern genome-editing and synthetic biology technologies [31–36].

5 | Artificial Intelligence-Assisted Genome and Strain Design

Identifying promising metabolic targets represents only the first step toward developing an efficient microbial cell factory. Computational predictions must ultimately be translated into genetic modifications that alter cellular metabolism in a controlled and stable manner. Recent advances in genome engineering and synthetic

biology have substantially expanded the range of modifications that can be introduced into microbial hosts. At the same time, AI and ML are increasingly integrated into these technologies to improve genetic target selection, predict the effects of genome modifications, and optimize the design of regulatory and biosynthetic elements. The convergence of AI with genome engineering therefore provides a direct connection between computational strain design and experimental construction of high-performing Biocompound-producing microorganisms [32].

5.1 | Artificial Intelligence-Assisted Identification of Genetic Targets

A central challenge in strain engineering is the selection of genetic modifications that improve product formation without severely compromising cellular growth or stability. A microbial genome contains numerous genes and regulatory elements that can potentially influence the target phenotype directly or indirectly. Systematic experimental testing of all possible combinations is therefore impractical, particularly when multiple genes and pathway components interact nonlinearly.

AI-based approaches can reduce this search space by predicting the likely effects of genetic perturbations. ML models can be trained using existing genotype–phenotype datasets to identify genes associated with production performance. Features derived from genomic sequences, gene-expression profiles, protein abundance, metabolic networks, and phenotypic measurements can be integrated to rank candidate targets.

AI-assisted target selection can involve several types of interventions, including gene deletion, gene overexpression, promoter modification, transcriptional repression, activation of alternative pathways, and modification of transport systems. Importantly, the optimal target may not always be a gene directly involved in the biosynthetic pathway. Genes associated with precursor supply, cofactor regeneration, stress responses, membrane transport, protein folding, or cellular tolerance may also influence final product productivity [35].

5.2 | CRISPR-Based Genome Engineering

CRISPR technologies have become central tools for microbial genome engineering because they enable targeted and programmable genetic modification. CRISPR/Cas systems can be used to introduce gene deletions, insertions, substitutions, and regulatory modifications, while multiplexed systems enable simultaneous modification of multiple genomic loci. These capabilities are particularly valuable for metabolic engineering, where coordinated modification of several pathway components may be required to achieve a desired production phenotype.

AI can support CRISPR-based engineering at several stages. Computational models can help select target sequences, predict editing efficiency, and identify potential off-target effects. In addition, ML approaches can analyze experimental editing data to identify sequence and genomic features associated with successful genome modification. This information can subsequently be used to improve the design of future editing experiments.

Integrating AI with CRISPR can also facilitate multiplex genome engineering. Rather than experimentally testing combinations of targets at random, predictive models can prioritize combinations that are expected to produce favorable metabolic outcomes. This capability is especially important for complex Biocompounds whose biosynthesis requires simultaneous manipulation of multiple pathways [37].

5.3 | CRISPRi and CRISPRa for Tunable Metabolic Regulation

Complete gene deletion or constitutive overexpression is not always the optimal strategy for metabolic engineering. Excessive repression may eliminate an essential metabolic function, whereas excessive expression can impose a substantial burden on the host. CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa) provide alternative approaches for tunable control of gene expression.

CRISPRi uses catalytically inactive Cas proteins directed toward specific genomic regions to reduce transcription, whereas CRISPRa enables targeted activation of gene expression. These technologies allow

researchers to modulate metabolic pathways without permanently altering the underlying DNA sequence. Such tunability is particularly useful when pathway components need to be balanced rather than simply switched on or off.

AI can help determine the appropriate level of repression or activation by modeling the relationship between regulatory strength and production phenotype. Instead of evaluating a small number of predefined expression levels, predictive models can guide the selection of combinations of regulatory strengths. This approach may be particularly useful for balancing precursor consumption, pathway flux, and cellular growth [35].

5.4 | Artificial Intelligence-Assisted Promoter and Regulatory Element Design

The expression level and temporal regulation of pathway genes can strongly influence Biocompound productivity. Synthetic biology therefore increasingly relies on libraries of promoters, terminators, ribosome-binding sites, transcription factors, and other regulatory elements with different activities. However, experimentally characterizing large numbers of regulatory sequences can be time-consuming.

ML and deep-learning models can learn sequence–function relationships from experimentally characterized regulatory elements. These models can predict the activity of existing sequences and potentially generate new regulatory elements with desired expression characteristics. AI-assisted promoter design can consequently support more precise control of pathway enzyme abundance and reduce the need for exhaustive experimental screening.

The use of generative models introduces an additional possibility: rather than selecting regulatory sequences from naturally occurring libraries, computational models can propose novel sequences designed to achieve a specific expression profile. Such predictions must nevertheless be experimentally validated because sequence-function relationships can depend strongly on the host organism, genomic context, growth conditions, and interactions with other regulatory elements [31–35].

5.5 | Genetic Circuits and Dynamic Metabolic Control

Static genetic modifications can be insufficient when optimal metabolic behavior changes during different stages of microbial growth. For example, strong pathway expression may be beneficial during the production phase but detrimental during early growth because of metabolic burden. Synthetic biology has therefore introduced dynamic regulatory systems that can change gene expression in response to intracellular or extracellular signals.

Genetic circuits can incorporate biosensors, transcriptional regulators, feedback mechanisms, and inducible systems to dynamically regulate metabolic pathways. AI can help design and optimize these circuits by predicting relationships between sensor inputs, regulatory outputs, and production phenotypes. ML models can also identify circuit architectures that balance growth and product formation.

Dynamic metabolic control is particularly relevant to industrial Biocompound production because it may improve both productivity and strain robustness. By separating biomass accumulation from product synthesis, engineered microorganisms can potentially achieve high cell density before redirecting resources toward the desired product [36–38].

5.6 | Artificial Intelligence-Assisted Enzyme Engineering

The performance of a metabolic pathway depends not only on gene expression but also on the catalytic properties of the enzymes involved. Enzyme activity, substrate specificity, catalytic efficiency, thermal stability, solvent tolerance, and resistance to product inhibition can all affect pathway performance. Protein engineering has traditionally relied on rational design, directed evolution, or combinations of both approaches. AI is increasingly being incorporated into this process through sequence-based prediction, structure prediction, protein-language models, and ML-guided variant selection.

AI models can identify sequence positions likely to influence enzyme activity or stability and can prioritize mutations for experimental testing. Structure-aware approaches can further integrate information about active sites, substrate-binding regions, and protein conformations. Instead of experimentally screening extremely large mutant libraries, AI can be used to select a smaller number of variants with a higher predicted probability of improved performance.

For Biocompound production, enzyme engineering can be particularly important when a pathway contains a rate-limiting enzyme or when the natural enzyme is poorly compatible with the selected microbial host. AI-assisted enzyme optimization can therefore complement metabolic engineering by improving the catalytic properties of individual pathway components [33–36].

5.7 | Multiplex and Combinatorial Strain Engineering

The performance of engineered microorganisms often depends on combinations of genetic modifications rather than individual perturbations. Increasing precursor supply, reducing competing pathways, modifying transport, improving cofactor regeneration, and enhancing product tolerance may need to occur simultaneously. The number of possible combinations can rapidly become enormous, creating a combinatorial optimization problem.

AI can address this challenge by prioritizing combinations of genetic modifications based on predicted interactions. Instead of assuming that the effects of individual modifications are additive, ML models can learn epistatic relationships from experimental datasets. GAs and other evolutionary optimization approaches can then explore combinations of modifications while limiting the number of experimental constructs required.

This approach creates an important connection between AI-based prediction and high-throughput synthetic biology. Automated construction and screening platforms can generate experimental datasets containing the phenotypic effects of different genetic combinations. These data can subsequently be returned to the predictive model, allowing the next generation of strain designs to be selected more intelligently [39]. As summarized in *Table 3*, AI can support a broad range of genome and strain-engineering strategies, including gene deletion and overexpression, CRISPR-based editing, transcriptional regulation, promoter engineering, enzyme optimization, and multiplex pathway modification.

Table 3. AI-assisted genome and strain engineering strategies.

Engineering Strategy	Biological Objective	AI Contribution	Typical Application in Biocompound Production	Major Challenge
Gene knockout	Eliminate competing or undesirable pathways	Prediction of high-impact targets	Redirecting metabolic flux toward product	Essential-gene effects and metabolic compensation
Gene overexpression	Increase pathway capacity	Expression-level prediction	Enhancing precursor or product-forming pathways	Metabolic burden
CRISPR/Cas editing	Precise genome modification	Target and editing-efficiency prediction	Multiplex pathway engineering	Off-target effects and editing efficiency
CRISPRi	Tunable gene repression	Prediction of optimal repression levels	Partial suppression of competing pathways	Context-dependent regulation
CRISPRa	Targeted gene activation	Prediction of activation strength	Activation of biosynthetic pathways	Host- and locus-dependent effects
Promoter engineering	Control gene-expression strength	Sequence–function prediction	Balancing multi-enzyme pathways	Host and genomic-context dependence
Genetic circuits	Dynamic pathway regulation	Circuit design and optimization	Growth–production balancing	Circuit complexity and stability
Enzyme engineering	Improve catalytic performance	Structure and sequence prediction	Optimization of rate-limiting enzymes	Experimental validation
Cofactor engineering	Improve redox balance	Metabolic-state prediction	NADH/NADPH optimization	Cellular redox interactions
Transport engineering	Improve substrate/product transport	Transporter prediction	Product secretion and uptake	Membrane and toxicity effects
Multiplex engineering	Coordinate multiple modifications	Combinatorial optimization	Complex pathway optimization	Epistasis and combinatorial complexity

5.8 | Predicting Strain Fitness, Stability, and Metabolic Burden

High product titer alone does not necessarily indicate a successful production strain. An engineered microorganism must also maintain acceptable growth, genetic stability, and production performance over extended cultivation. Extensive pathway modification can impose metabolic burden by consuming cellular resources, disrupting native metabolism, increasing protein-expression demand, or generating toxic intermediates.

AI models can potentially predict these trade-offs before experimental construction. By incorporating growth data, omics measurements, pathway characteristics, and previous engineering outcomes, predictive models can estimate the likely effects of candidate modifications on both production and cellular fitness. Multi-objective optimization can then be used to identify designs that balance productivity with growth and stability.

Predicting long-term strain stability remains particularly challenging because genetic and physiological adaptation can occur during repeated cultivation. Future AI models will therefore need to incorporate temporal datasets and evolutionary information to better predict how engineered strains behave over extended production processes [28–32].

5.9 | From Computational Design to Experimental Strain Construction

Integrating AI with genome engineering creates a continuous link between computational prediction and experimental implementation. In an AI-enabled DBTL workflow, computational models first generate and rank candidate genetic designs. Genome-editing technologies are then used to construct selected strains, while high-throughput phenotyping measures their growth and Biocompound production. The resulting data are subsequently incorporated into the learning stage to improve future predictions.

This workflow can substantially reduce the number of low-probability designs that require experimental testing. More importantly, each engineering cycle generates new genotype–phenotype information that can improve the predictive capability of the system. Over successive iterations, the computational model can become increasingly specialized for the selected host organism, Biocompound, and production environment.

The ultimate objective is therefore not simply to automate genome editing but to establish an integrated design system in which AI identifies promising genetic architectures, synthetic biology constructs them, automated experimentation evaluates their performance, and ML uses the resulting data to guide the next design cycle [32].

5.10 | Current Limitations and Future Directions

Despite rapid progress, several challenges limit the widespread application of AI-assisted genome and strain design. Training datasets may be too small or biased toward well-characterized organisms and commonly studied pathways. Genetic modifications can also produce context-dependent effects, meaning that a model trained in one host or culture condition may not generalize to another. In addition, the biological consequences of multiple simultaneous modifications may be difficult to predict because of epistasis, compensatory regulation, and evolutionary adaptation.

Another challenge is the gap between computationally optimized designs and experimentally feasible constructs. A theoretically optimal strain may require genetic modifications that are difficult to implement, genetically unstable, or incompatible with industrial cultivation. Therefore, future AI systems should incorporate not only predicted biological performance but also constraints related to genome-editing feasibility, strain robustness, process compatibility, and manufacturing cost.

The combination of AI, CRISPR-based genome engineering, synthetic regulatory systems, enzyme engineering, and automated high-throughput experimentation nevertheless provides a powerful foundation for next-generation microbial cell factories. As predictive models become more capable and experimental

platforms become increasingly automated, strain engineering may progressively shift from manual optimization toward data-driven, iterative, and partially autonomous design [38].

6 | Artificial Intelligence-Based Prediction of Strain Performance and Biocompound Productivity

The successful construction of an engineered microbial strain does not necessarily result in improved Biocompound production. Genetic modifications can affect multiple aspects of cellular physiology, including metabolic flux, growth, precursor availability, stress responses, and product tolerance. Therefore, predicting strain performance before or during experimental evaluation is an important component of AI-driven strain engineering. ML approaches can integrate genetic, molecular, physiological, and process-level information to predict production outcomes and prioritize promising strains for experimental validation [39].

6.1 | Artificial Intelligence-Based Genotype–Phenotype Prediction

A major objective of AI-assisted strain engineering is to establish relationships between genetic modifications and production phenotypes. Engineered strains may differ in titer, yield, productivity, growth rate, robustness, and product tolerance, while the effects of individual genetic modifications may depend on interactions with other genes and metabolic pathways.

ML models can learn these complex relationships from experimentally characterized strain libraries. Genomic features, gene-expression profiles, protein abundance, metabolite concentrations, pathway characteristics, and cultivation conditions can be used as predictive variables, while production metrics represent the target outputs. Nonlinear models are particularly useful for capturing epistatic interactions in strains containing multiple genetic modifications [37].

6.2 | Prediction of Titer, Yield, and Productivity

Titer, yield, and productivity are among the most important indicators of microbial production performance. AI models can predict these parameters from strain characteristics and experimental conditions, allowing promising candidates to be identified before extensive experimental screening.

An important advantage is the possibility of early prediction. Measurements obtained during the initial stages of cultivation, such as biomass growth, substrate consumption, metabolite profiles, and process variables, can be used to estimate final product formation. Early identification of low-performing strains can reduce cultivation time and experimental costs, while high-performing candidates can be prioritized for further development [35].

6.3 | Artificial Intelligence for Strain Robustness and Product Tolerance

Industrial production exposes microorganisms to stresses that may not be present under laboratory conditions, including high substrate and product concentrations, osmotic stress, oxidative stress, pH variation, temperature fluctuations, and oxygen limitation. Consequently, a strain with high laboratory-scale productivity may not necessarily be suitable for industrial fermentation.

AI can integrate physiological and omics data to identify molecular characteristics associated with stress resistance and product tolerance. Such models can help identify potential engineering targets related to transport systems, membrane properties, stress-response pathways, redox balance, and detoxification mechanisms. Incorporating robustness into strain prediction can therefore improve the likelihood that computationally selected strains will perform under realistic production conditions [37].

6.4 | High-Throughput Screening and Active Learning

The increasing use of automated genome engineering enables the construction of large numbers of candidate strains. However, experimental screening can become a major bottleneck. AI can address this challenge by

ranking candidates according to their predicted performance and selecting the experiments expected to provide the greatest value.

Active-learning and BO approaches are particularly promising because they allow models to select new experiments based not only on predicted performance but also on uncertainty. Experimental results are subsequently incorporated into the model, improving predictions for the next generation of strains. This iterative process creates an efficient feedback loop between computational prediction and experimental validation [34].

6.5 | Explainability and Prediction Uncertainty

Predictive accuracy alone is insufficient for biological applications. Researchers also need to understand why a model predicts that a particular strain or genetic modification will improve production. Explainable AI approaches can identify influential genomic, metabolic, or physiological features and provide insights into the biological factors driving model predictions.

Similarly, uncertainty estimation is essential when datasets are limited or heterogeneous. Bayesian approaches, ensemble models, and other uncertainty-aware methods can distinguish well-supported predictions from those generated in poorly characterized regions of the design space. This information can improve experimental prioritization and reduce the risk of investing resources in unreliable predictions [39]. As summarized in *Table 4*, AI-based approaches enable multidimensional evaluation of engineered strains based on growth, production performance, metabolic efficiency, stability, tolerance, and robustness.

Table 4. AI-assisted evaluation of engineered strain performance.

Performance Parameter	Data / Input	AI Application	Main Outcome
Growth	Growth curves	Prediction	Growth rate
Titer	Product concentration	Regression	Product titer
Yield	Substrate/Product data	Prediction	Conversion efficiency
Productivity	Time-course data	Time-series modeling	Production rate
Metabolic burden	Growth + omics data	ML prediction	Cellular burden
Genetic stability	Genomic + passage data	Classification	Strain stability
Product tolerance	Growth + product levels	Dose–response modeling	Tolerance threshold
Robustness	Process + phenotype data	Predictive modeling	Stable performance
Metabolic efficiency	Flux + metabolite data	Flux prediction	Flux efficiency
Multi-objective performance	Titer, yield, growth	Optimization	Optimal strain

6.6 | Integrating Strain and Process Prediction

Strain performance is strongly influenced by its cultivation environment. The same engineered microorganism can exhibit substantially different productivity under different pH, temperature, oxygen, carbon-source, and feeding conditions. Therefore, future AI models should jointly consider genetic characteristics and bioprocess variables rather than treating strain design and fermentation optimization as independent problems.

Integrated models can predict the performance of a specific strain under defined process conditions and identify combinations of biological and environmental variables that maximize production. This approach is particularly relevant to industrial scale-up, where strains optimized under laboratory conditions may behave differently in larger bioreactors [35].

6.7 | Artificial Intelligence-Based Strain Prediction within the Design–Build–Test–Learn Framework

AI-based strain prediction provides a central link within the DBTL cycle. During Design, AI models prioritize genetic architectures and pathway configurations. During Build, selected strains are constructed using genome-editing and synthetic biology technologies. During Test, high-throughput experiments generate production and phenotypic data. Finally, during Learn, these data are incorporated into the predictive model to improve subsequent strain designs.

Repeated DBTL cycles can progressively reduce the experimental search space and increase the probability of identifying high-performing strains. The ultimate objective is to move from trial-and-error strain development toward predictive and iterative engineering in which computational models continuously learn from experimental outcomes.

AI-based strain-performance prediction can significantly improve the efficiency of microbial cell-factory development by connecting genetic design with measurable production outcomes. However, reliable predictions still depend on high-quality datasets, biologically meaningful features, appropriate model selection, and experimental validation. Its greatest potential is therefore likely to emerge from integration with multi-omics, metabolic modeling, genome engineering, high-throughput screening, and AI-enabled fermentation [38].

7 | Artificial Intelligence-Enabled Fermentation and Bioprocess Optimization

The performance of an engineered microbial strain is strongly influenced by the conditions under which it is cultivated. Even highly optimized strains may exhibit reduced productivity when fermentation conditions are not appropriately controlled. Parameters such as temperature, pH, dissolved oxygen, agitation, aeration, substrate concentration, feeding rate, and cultivation time interact dynamically and can significantly affect microbial growth, metabolic flux, and Biocompound formation. AI and ML provide powerful tools for modeling these complex relationships and optimizing fermentation processes [32].

7.1 | Artificial Intelligence-Based Optimization of Fermentation Conditions

Traditional fermentation optimization often relies on one-factor-at-a-time experiments or statistical design of experiments. Although these approaches can identify important variables, they may be inefficient for highly nonlinear and interacting systems. ML models can simultaneously analyze multiple process variables and identify combinations associated with improved biomass formation, product titer, yield, and productivity. Variables such as pH, temperature, dissolved oxygen, agitation speed, aeration rate, carbon-source concentration, nitrogen availability, and feeding strategy can be incorporated into predictive models. Regression algorithms, Artificial Neural Networks (ANNs), ensemble methods, and BO can then be used to identify process conditions that maximize the desired production objective.

Importantly, AI-based optimization can also account for multiple objectives. For example, maximizing product concentration may reduce growth rate, while increasing aeration may improve oxygen availability but increase energy consumption. Multi-objective optimization can therefore identify operating conditions that balance productivity, yield, process time, and resource consumption [31].

7.2 | Prediction of Fermentation Performance

AI models can be used not only to optimize fermentation conditions but also to predict process outcomes. Historical fermentation data can be used to establish relationships between process variables and biomass concentration, substrate consumption, metabolite formation, and final product titer. Dynamic prediction is particularly valuable for industrial bioprocessing. Measurements collected during fermentation can be continuously incorporated into predictive models, allowing the expected final outcome to be estimated before the process is completed. Such predictions can support early intervention when abnormal process behavior is detected and help prevent the loss of an entire production batch.

AI-based soft sensors can also estimate variables that are difficult or expensive to measure continuously. For example, biomass concentration, metabolic state, or product concentration may be inferred from easily measurable signals such as dissolved oxygen, carbon dioxide evolution, pH, agitation, and off-gas composition [30].

7.3 | Artificial Intelligence for Feeding Strategy and Process Control

Fed-batch fermentation is widely used for industrial production because controlled substrate feeding can prevent substrate inhibition, overflow metabolism, and excessive by-product formation. However, determining an optimal feeding profile can be challenging because microbial nutrient requirements change dynamically during cultivation.

AI can learn the relationship between feeding strategies and production outcomes and subsequently recommend adaptive feeding profiles. Instead of following a fixed feeding schedule, an AI-enabled system can modify substrate addition according to real-time measurements of biomass, oxygen consumption, substrate availability, or product formation.

This adaptive approach may improve substrate utilization and maintain cells in a favorable physiological state for product formation. When combined with automated control systems, AI can potentially enable continuous adjustment of fermentation conditions in response to changes in microbial behavior [39]. As summarized in *Table 5*, AI can be applied across the entire bioprocessing workflow, from medium formulation and fermentation optimization to real-time monitoring, adaptive control, scale-up, and digital-twin development.

Table 5. AI applications in fermentation and bioprocess optimization.

Bioprocess Stage	Key Variables	AI Application	Prediction/Optimization Target	Expected Benefit
Medium optimization	Carbon, nitrogen, minerals	ML-based optimization	Optimal medium composition	Improved growth and production
Batch fermentation	pH, temperature, DO	Predictive modeling	Biomass and product formation	Improved process performance
Fed-batch fermentation	Feed rate and substrate concentration	BO	Optimal feeding profile	Reduced substrate inhibition and by-products
Oxygen management	DO, aeration, agitation	Predictive and adaptive control	Oxygen availability	Improved metabolic activity
Real-time monitoring	pH, DO, CO_2, O_2, biomass	AI-based soft sensors	Unmeasured process variables	Early detection of process deviations
Process control	Multiple online variables	Reinforcement learning/adaptive control	Dynamic control actions	Stable and optimized fermentation
Product prediction	Process trajectories	ML prediction	Final titer and productivity	Early identification of poor batches
Scale-up	Mixing, oxygen transfer, gradients	Hybrid AI–mechanistic models	Large-scale process performance	Improved scale-up reliability
Digital twin	Real-time process + mechanistic data	AI-assisted simulation	Process trajectory and intervention	Virtual process optimization
Industrial production	Integrated strain and process data	Multi-objective optimization	Titer, yield, productivity, cost	Scalable and efficient production

7.4 | Real-Time Monitoring and Intelligent Bioprocess Control

Modern fermentation systems generate large amounts of real-time data through sensors and process analytical technologies. These data can include pH, temperature, dissolved oxygen, pressure, agitation, gas composition, biomass-related signals, and other process variables. AI can transform these measurements into predictive information for monitoring and control. An AI-enabled bioreactor can detect deviations from the expected process trajectory, identify abnormal patterns, and recommend corrective actions. For example, unexpected changes in dissolved oxygen or carbon dioxide evolution may indicate altered metabolic activity or oxygen-transfer limitations. Early detection can allow process parameters to be adjusted before substantial productivity losses occur. Combining AI with advanced process control therefore creates an opportunity to move from passive monitoring to predictive and adaptive fermentation management [36–39].

7.5 | Artificial Intelligence and Scale-Up of Bioprocesses

Scale-up remains a major challenge in industrial Biocompound production. A microbial strain that performs well in a laboratory bioreactor may exhibit substantially different behavior at pilot or industrial scale because of changes in mixing, oxygen transfer, heat transfer, shear stress, and nutrient gradients.

AI can assist scale-up by learning relationships between laboratory and larger-scale process data. Models can integrate variables such as agitation, aeration, power input, oxygen-transfer characteristics, and metabolic indicators to predict process behavior at larger scales.

Rather than simply maintaining identical numerical values for process parameters at different scales, AI-based approaches can help identify the combinations of operating conditions that preserve the desired physiological state of the microorganism. Such an approach can reduce the risk associated with transferring laboratory-optimized processes to industrial production [38].

7.6 | Digital Twins and Intelligent Bioreactors

Integrating AI with mechanistic process models has contributed to the development of digital twins for bioprocessing. A digital twin is a computational representation of a physical fermentation system that can be continuously updated using real-time process data.

For Biocompound production, a digital twin can represent microbial growth, substrate consumption, product formation, and relevant process conditions. AI can improve the predictive capabilities of the digital twin by learning from deviations between simulated and observed behavior.

Such systems could eventually allow researchers to evaluate alternative process strategies computationally before implementing them in the physical bioreactor. Digital twins may therefore support process optimization, fault detection, scale-up, and predictive maintenance while reducing experimental costs [37].

The integration of AI, digital twins, automated experimentation, and real-time bioprocess monitoring provides the foundation for closed-loop biomanufacturing systems in which experimental data continuously update predictive models and guide subsequent design and process decisions, as illustrated in *Fig. 3*.

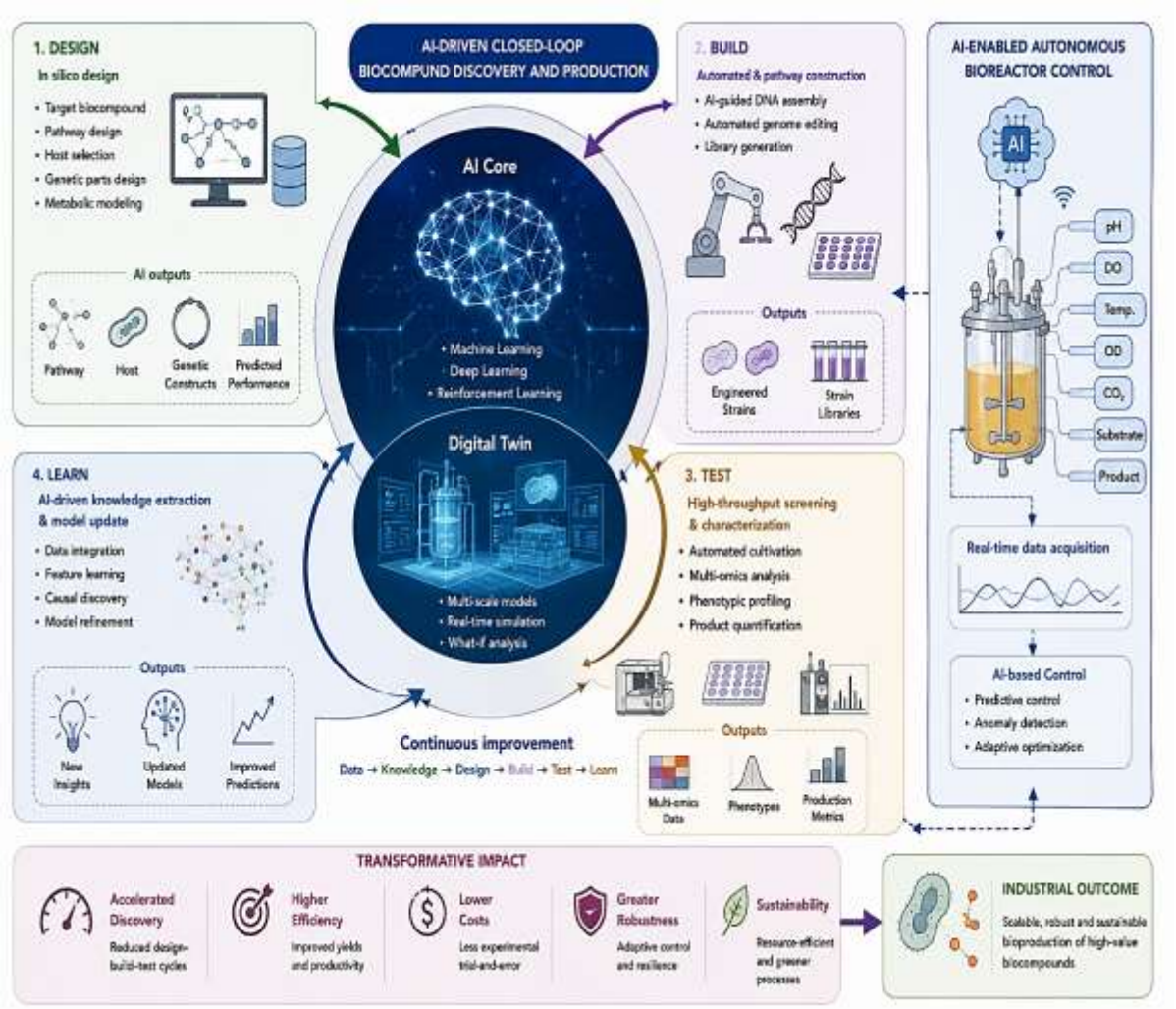

Fig. 3. AI-driven closed-loop biomanufacturing integrating DBTL cycles, digital twins, and autonomous bioreactor control.

7.7 | Integrating Strain Engineering with Fermentation Optimization

An important limitation of treating strain engineering and fermentation optimization as separate activities is that genetic modifications and environmental conditions interact strongly. A genetic modification that improves production under one set of conditions may have little benefit under another. Conversely, an optimized fermentation strategy may compensate for certain limitations of an engineered strain. AI provides an opportunity to optimize these two components jointly. Genetic characteristics, omics data, strain phenotype, and fermentation variables can be incorporated into integrated models that predict production performance under different strain–process combinations. This approach supports a transition from strain optimization followed by process optimization toward simultaneous strain–process optimization. Such integration is particularly relevant for complex Biocompounds in which pathway activity, cellular physiology, and environmental conditions are strongly interconnected [33].

7.8 | Challenges and Future Perspectives

Despite its potential, AI-enabled fermentation optimization faces several challenges. Industrial fermentation datasets may be limited, proprietary, heterogeneous, or generated under different operating conditions. Sensor noise, missing data, process variability, and changes in equipment configuration can also reduce model reliability. Furthermore, models developed at laboratory scale may not generalize directly to industrial reactors. Future developments will likely involve hybrid mechanistic–AI models, improved soft sensors, digital twins, autonomous process control, and real-time learning systems. Integrating AI with automated

bioreactors could enable increasingly adaptive fermentation platforms in which process conditions are continuously optimized according to the physiological state of the production strain. AI-enabled fermentation represents a critical extension of AI-driven strain engineering. While genome engineering determines the biological potential of a microbial cell factory, intelligent bioprocessing determines how effectively that potential can be translated into industrial productivity. Integrating both components is therefore essential for developing efficient, scalable, and sustainable platforms for Biocompound production [38].

8 | Conclusion

AI is increasingly transforming the development of microbial cell factories from a predominantly trial-and-error process into a more predictive, data-driven, and iterative engineering discipline. As reviewed in this article, integrating AI with multi-omics, metabolic modeling, genome engineering, synthetic biology, and bioprocess optimization provides a comprehensive framework for improving the production of high-value Biocompounds [12–15], [17].

At the strain-design level, AI can facilitate the identification of genetic targets, prediction of genotype–phenotype relationships, reconstruction and optimization of metabolic pathways, and selection of promising genome modifications. Integrating genomic, transcriptomic, proteomic, and metabolomic information can provide a more comprehensive understanding of cellular responses and help identify metabolic bottlenecks that may not be apparent from individual datasets. Combined with CRISPR-based genome engineering and synthetic biology, these computational capabilities can accelerate the construction of microbial strains with improved titer, yield, productivity, robustness, and product tolerance. AI also provides important opportunities beyond genetic engineering. ML models can predict strain performance, prioritize candidates for high-throughput screening, estimate uncertainty, and identify relationships between strain characteristics and production outcomes. At the bioprocess level, AI can optimize fermentation conditions, feeding strategies, process control, and scale-up. Integrating real-time sensing, soft sensors, digital twins, and predictive control may further enable adaptive bioprocessing, in which fermentation conditions are continuously adjusted according to the physiological state of the production system. A particularly important development is the integration of these capabilities within DBTL cycles [21–28]. In such systems, AI does not operate as an isolated computational tool but becomes part of a continuous feedback loop connecting biological design, genome construction, experimental testing, and computational learning. Each engineering cycle generates new genotype–phenotype and process-level data, which can subsequently improve the next generation of predictions. The combination of AI with automation and high-throughput experimentation therefore creates the foundation for increasingly efficient and potentially autonomous microbial strain-development platforms. Despite this progress, several challenges must be addressed before AI-driven biomanufacturing can achieve widespread industrial adoption. The availability and quality of genotype–phenotype datasets remain major limitations, particularly for less-characterized microorganisms and uncommon Biocompounds. Differences in experimental protocols, cultivation conditions, data formats, and measurement technologies can also limit model transferability.

Furthermore, highly accurate predictions do not necessarily imply biological causality, emphasizing the continued importance of explainable models, mechanistic constraints, and experimental validation [31-36]. Another important challenge is the transition from laboratory-scale optimization to industrial implementation. Biological systems can behave differently under large-scale conditions because of changes in mixing, oxygen transfer, nutrient gradients, shear stress, and process variability. Consequently, future AI systems should increasingly integrate strain characteristics, metabolic state, and bioprocess variables within unified predictive frameworks rather than optimizing each component independently. Looking forward, the convergence of AI with foundation models, generative approaches, protein-language models, automated biofoundries, digital twins, and self-driving laboratories may substantially expand the capabilities of microbial biomanufacturing. Future platforms could potentially generate candidate genetic designs, predict their metabolic consequences, construct selected strains, perform automated experiments, interpret the resulting data, and initiate subsequent engineering cycles with limited human intervention. Such systems would

represent a transition from AI-assisted biomanufacturing toward increasingly autonomous biological engineering [32–40]. Ultimately, the greatest potential of AI in Biocompound production lies not in replacing experimental biotechnology but in creating a more intelligent interaction between computation and experimentation. By connecting molecular-level information with strain engineering and industrial bioprocessing, AI-driven approaches can shorten development cycles, reduce experimental costs, improve resource efficiency, and accelerate the translation of biological discoveries into scalable production technologies. The successful realization of this potential will depend on the continued integration of high-quality biological data, mechanistic knowledge, advanced AI models, automated experimentation, and rigorous experimental validation.

Authors' Contributions

The author was responsible for all stages of the research and manuscript preparation and approved the final version.

Data Availability

All data are included in the text.

Funding

This research was not supported by any specific grant from funding bodies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

There are no competing interests to declare.

Consent for Publication

The author confirms consent for the publication of this work

Ethics Approval and Consent to Participate

This article does not contain any studies with human participants performed by the author.

References

[1] Maslej, N., Fattorini, L., Perrault, R., Gil, Y., Parli, V., Kariuki, N., ... & Oak, S. (2025). *Artificial intelligence index report*. https://doi.org/10.48550/arXiv.2504.07139

[2] Gursoy, D., & Cai, R. (2025). Artificial intelligence: An overview of research trends and future directions. *International journal of contemporary hospitality management*, *37*(1), 1–17. https://doi.org/10.1108/IJCHM-03-2024-0322

[3] Gama, F., & Magistretti, S. (2025). Artificial intelligence in innovation management: A review of innovation capabilities and a taxonomy of AI applications. *Journal of product innovation management*, *42*(1), 76–111. https://doi.org/10.1111/jpim.12698

[4] Svennberg, E., Han, J. K., Caiani, E. G., Engelhardt, S., Ernst, S., Friedman, P., ... & Zorzi, A. (2025). State of the art of artificial intelligence in clinical electrophysiology in 2025: A scientific statement of the European heart rhythm association (EHRA) of the ESC, the heart rhythm society (HRS), and the ESC working group on E-cardiology. *Europace*, *27*(5), EUAF071. https://doi.org/10.1093/europace/euaf071

[5] Eaton, S. E. (2025). Global trends in education: Artificial intelligence, postplagiarism, and future-focused learning for 2025 and beyond-2024-2025 werklund distinguished research lecture. *International journal for educational integrity*, *21*(1), 12. https://doi.org/10.1007/s40979-025-00187-6

[6] Zha, D., Bhat, Z. P., Lai, K. H., Yang, F., Jiang, Z., Zhong, S., & Hu, X. (2025). Data-centric artificial intelligence: A survey. *ACM computing surveys*, *57*(5), 1-42. https://doi.org/10.1145/3711118

[7] Mageed, I. A. (2025). *Ad educationem in millionibus annorum antea: The profound impact of artificial intelligence on education. Zenodo preprints*. https://doi.org/10.5281/zenodo.16789605

[8] McDonald, N., Johri, A., Ali, A., & Collier, A. H. (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. *Computers in human behavior: Artificial humans*, *3*, 100121. https://doi.org/10.1016/j.chbah.2025.100121

[9] Ooi, K. B., Tan, G. W. H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., ... & Wong, L. W. (2025). The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of computer information systems*, *65*(1), 76-107. https://doi.org/10.1080/08874417.2023.2261010

[10] Zhang, K., Yang, X., Wang, Y., Yu, Y., Huang, N., Li, G., ... & Yang, S. (2025). Artificial intelligence in drug development. *Nature medicine*, *31*(1), 45-59. https://doi.org/10.1038/s41591-024-03434-4

[11] Derakhshan, A., Teo, T., & Khazaie, S. (2025). Investigating the usefulness of artificial intelligence-driven robots in developing empathy for English for medical purposes communication: The role-play of Asian and African students. *Computers in human behavior*, *162*, 108416. https://doi.org/10.1016/j.chb.2024.108416

[12] Yim, I. H. Y., & Su, J. (2025). Artificial intelligence (AI) learning tools in K-12 education: A scoping review. *Journal of computers in education*, *12*(1), 93–131. https://doi.org/10.1007/s40692-023-00304-9

[13] Faiyazuddin, M., Rahman, S. J. Q., Anand, G., Siddiqui, R. K., Mehta, R., Khatib, M. N., ... & Sah, R. (2025). The impact of artificial intelligence on healthcare: A comprehensive review of advancements in diagnostics, treatment, and operational efficiency. *Health science reports*, *8*(1), e70312. https://doi.org/10.1002/hsr2.70312

[14] Xu, T., & Baghaei, S. (2025). Reshaping the future of sports with artificial intelligence: Challenges and opportunities in performance enhancement, fan engagement, and strategic decision-making. *Engineering applications of artificial intelligence*, *142*, 109912. https://doi.org/10.1016/j.engappai.2024.109912

[15] Chai, T., Tao, Y., Zhao, C., & Chen, X. (2025). Hierarchical metabolic engineering for rewiring cellular metabolism. *FEMS microbiology reviews*, *49*, fuaf047. https://doi.org/10.1093/femsre/fuaf047

[16] Yadav, J., Marwah, H., & Kumar, C. (2025). Synthetic biology and metabolic engineering paving the way for sustainable next-gen biofuels: A comprehensive review. *Energy advances*, *4*(10), 1209–1228. https://doi.org/10.1039/d5ya00118h

[17] Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The journal of strategic information systems*, *34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

[18] Ren, Y., Celinska, E., Cai, P., & Zhou, Y. J. (2025). Combing directed enzyme evolution with metabolic engineering to develop efficient microbial cell factories. *Chem & bio engineering*, *2*(8), 449–459. https://doi.org/10.1021/cbe.5c00002

[19] Woolston, B. M., Edgar, S., & Stephanopoulos, G. (2013). Metabolic engineering: Past and future. *Annual review of chemical and biomolecular engineering*, *4*, 259–288. https://doi.org/10.1146/annurev-chembioeng-061312-103312

[20] Cheng, X., Han, M., Zhang, G., Qiao, W., Wang, L., & Ma, Y. (2026). Metabolic engineering strategies for flavonoid biosynthesis in saccharomyces cerevisiae: Current advances and future perspectives. *Biotechnology and bioengineering*, 123(8), 2162–2179. https://doi.org/10.1002/bit.70248

[21] Zhou, Y., Liu, Y., Sun, H., & Lu, Y. (2025). Creating novel metabolic pathways by protein engineering for bioproduction. *Trends in biotechnology*, *43*(5), 1094–1103. https://www.cell.com/trends/biotechnology/abstract/S0167-7799(24)00308-1

[22] Ramakrishnan, P., Sundaram, T., Lahiri, D., Nag, M., & Bhattacharya, D. (2025). Genetic engineering and modulation of metabolic pathways. In *Introduction to metabolic engineering and application* (pp. 295–330). Springer. https://doi.org/10.1007/978-3-031-93189-5_12

[23] Shi, S., Chen, Y., & Nielsen, J. (2025). Metabolic engineering of yeast. *Annual review of biophysics*, *54*(1), 101–120. https://doi.org/10.1146/annurev-biophys-070924-103134

[24] Chen, J., Singh, N., Lu, J., Lane, S. T., & Zhao, H. (2025). Artificial intelligence-powered biofoundries for protein engineering and metabolic engineering. *Current opinion in biotechnology*, *96*, 103380. https://doi.org/10.1016/j.copbio.2025.103380

[25] 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://www.cell.com/trends/biotechnology/abstract/S0167-7799(16)00044-5

[26] Lee, J., Yu, H. E., & Lee, S. Y. (2025). Metabolic engineering of microorganisms for carbon dioxide utilization. *Current opinion in biotechnology*, *91*, 103244. https://doi.org/10.1016/j.copbio.2024.103244

[27] Wang, K., Song, X., Cui, B., Wang, Y., & Luo, W. (2025). Metabolic engineering of Escherichia coli for efficient production of ectoine. *Journal of agricultural and food chemistry*, *73*(1), 646–654. https://doi.org/10.1021/acs.jafc.4c07640

[28] Zhang, G., & Wang, J. (2025). Artificial intelligence-driven metabolic engineering is applied to the development of active ingredients in Traditional Chinese Medicine. *BIO web of conferences* (Vol. 174, p. 03013). EDP Sciences. https://doi.org/10.1051/bioconf/202517403013

[29] Zhao, M., Pan, X., Arsalan, A., Zabed, H. M., Guo, L., Zhang, C., & Qi, X. (2026). Metabolic engineering and synthetic biology-driven strategies to harness microbial production of Adipic acid: Current status and future direction. *ACS synthetic biology*, *15*(3), 893–914. https://doi.org/10.1021/acssynbio.5c00869

[30] Domenzain, I., Lu, Y., Wang, H., Shi, J., Lu, H., & Nielsen, J. (2025). Computational biology predicts metabolic engineering targets for increased production of 103 valuable chemicals in yeast. *Proceedings of the national academy of sciences*, *122*(9), e2417322122. https://doi.org/10.1073/pnas.2417322122

[31] Sokra, I., Somaly, S., & Meta, H. (2026). CRISPR-Cas9-Based genome editing in microbial biotechnology: Advances in metabolic engineering, fermentation systems, and industrial applications. *Journal of agriculture and technology*, *2*(1), 154–166. https://doi.org/10.5281/zenodo.17994411

[32] Panagiotidou, E., & Theodosiou, E. (2026). Systems metabolic engineering: An integrated approach for future environmental biotechnology. *Biotechnology for the environment*, *3*(1), 8. https://doi.org/10.1186/s44314-026-00042-z

[33] Singh, A., & Negi, P. S. (2025). Biotechnological application of health-promising bioactive compounds. In *Biotechnological intervention in production of bioactive compounds: Biosynthesis, characterization and applications* (pp. 73–94). Springer. https://doi.org/10.1007/978-3-031-76859-0_5

[34] Moss, A., Peh, J. H., Segaran, T. C., Lananan, F., Kari, Z. A., Wei, L. S., ... & Azizi, M. N. (2025). Bioactive compounds in aquaculture. *Aquaculture international*, *33*(6), 414. https://doi.org/10.1007/s10499-025-01985-y

[35] Trujillo-Cayado, L. A., Sánchez-García, R. M., García-Domínguez, I., Rodríguez-Luna, A., Hurtado-Fernández, E., & Santos, J. (2025). Emerging trends in sustainable biological resources and bioeconomy for food production. *Applied sciences*, *15*(12), 6555. https://doi.org/10.3390/app15126555

[36] Szabo, K., Varvara, R.-A., Ciont, C., Macri, A. M., & Vodnar, D. C. (2025). An updated overview on the revalorization of bioactive compounds derived from tomato production and processing by-products. *Journal of cleaner production*, *497*, 145151. https://doi.org/10.1016/j.jclepro.2025.145151

[37] Marzban, N., Psarianos, M., Herrmann, C., Schulz-Nielsen, L., Olszewska-Widdrat, A., Arefi, A., ... & Sturm, B. (2025). Smart integrated biorefineries in bioeconomy: A concept toward zero-waste, emission reduction, and self-sufficient energy production. *Biofuel research journal*, *12*(1), 2319-2349. https://doi.org/10.18331/BRJ2025.12.1.4

[38] Gallego, I. (2025). Biocompounds of commercial interest from freshwater and marine phytoplankton. In *The role of plankton in freshwater and marine ecology* (pp. 1-23). IntechOpen. https://doi.org/10.5772/intechopen.1008500

[39] Khan Jani, F. (2025). Design and chemical engineering of polymeric biomaterials: Functionalization, crosslinking, and characterization strategies for advanced 3D bioprinting. *Biocompounds*, *2*(3), 151–167. https://doi.org/10.48313/bic.vi.42

[40] Laribi, A., Zieniuk, B., Bouchedja, D. N., Hafid, K., Elmechta, L., & Becila, S. (2025). Valorization of olive mill wastewater via Yarrowia Lipolytica: Sustainable production of high-value metabolites and Biocompounds—A review. *Fermentation*, *11*(6), 326. https://doi.org/10.3390/fermentation11060326