The EuroBiotech Journal

REVIEW ARTICLE

Medical Biotechnology

AI-Powered Synthetic Biology: Current Situation, Challenges, and Future Perspectives

Izem Olcay Sahin[1], Nuriye Gokce[1], Duygu T. Yildirim[1], A. Baki Yildirim[1], Hilal Akalın[1], Donald Martin[2], Tommaso Beccari[3], Oscar Vicente[4], Iza Radecka[5], Fideline Tchuenbou-Magaia[5], Robert S. Marks[6], Ratnesh Lal[7], Satya Prakash[8], Adam Mechler[9], Mario Petrov Milkov[10], Svetlana Fotkova Georgieva[11], Ilia Iliev[12], Kosi Gramatikoff[13], Ed Judge[14], Milica Markovic[15], Radka Kaneva[16], Galina Aleksieva Yaneva[17], Nadya Vasileva Agova[11], Nikoleta Dobromirova Ivanova[17], Mariya Kiryakova Tsvetkova[17], Ivelin Rosenov Iliev[11], Michel Salzet[18], Kisung Ko[19], Michele Maffia[20], Chiara Coppola[20], Matteo Bertelli[21], Isabelle Fournier[22], Lejla Pojskic[23], Qun Sun[24], Lembit Nei[25], Reynir Armgrisson[26], Gary Henehan[27], Daumantas Matulis[28], Dijana Plaseska-Karanfilska[29], K. Santacruz-Gomez[30], Isabel Belo[31], Štefánia Hrončeková[32], Sehime G. Temel[33,34,35], Greiner-Tollersrud Ole K.[36], Dana Tapaloaga[37], Andreas Janecke[38,39], Ivana Marova[40], Benedetta Spedicato[41], Polona Žnidaršič-Plazl[42], Igor Plazl[42], Ariola Bacu[43], Anargyros N. Moulas[44], Alexander Kilchevsky[45], M.Sait Dundar[46], P. Bartolini[47], Anita Slavica[48], Francisco Fuentes[49], Jose Carlos Lorenzo Feijoo[50], Andrés Izquierdo Romero[51,52,53], Helal Ragab Moussa[54], Amin Hejazi[55], Attya Bhatti[56], Bajram Berisha[57,58,59], Alma Kokhmetova[60], Nikolai Zhelev[61,62], Juraj Krajcovic[63], Viktor Nedovic[64], Alla Salmina[65], Mark Nujiten[66,67], İrem Kalay[68], Majeti Narasimha Vara Prasad[69], Luis Izquierdo Lopez[70], Dominika Vešelényiová[63], Maria Rachele Ceccarini[3,71], Bernard Fioretti[72], Gergana P. Ilieva[73], M. Cerkez Ergoren[74], Noursaid Tligui[75], Serghei Sprincean[76], Saharuddin Bin Mohamad[77], Miklos Kellermayer[78], Esra Arslan Ates[79], Pembe Savas[80], Ivana Márová[81], Martin Koller[82], Hakan Gurkan[83], Erhan Parıltay[79], Ercument Ovalı[84], Sevda Yesim Ozdemir[85], Tanil Kocagoz[86,87], Hilmi Tozkir[88], Gunnur Demircan[89], Eda Tahir Turanli[90,91], Havva Cobanogullari[80], Victor Nedovic[92], Victor Revin[93], Recep Eroz[94], Annika Joy Meitern[95], Munis Dundar[1]

1. Department of Medical Genetics, School of Medicine, Erciyes University, Kayseri, Türkiye
2. Université Grenoble Alpes, CNRS, Grenoble INP, LMGP (UMR 5628), Grenoble, France
3. Department of Pharmaceutical Sciences, University of Perugia, Perugia, Italy
4. Institute for the Conservation and Improvement of Valencian Agrodiversity (COMAV), Universitat Politècnica de València, Valencia, Spain
5. School of Pharmacy and Life Sciences, Faculty of Science and Engineering, University of Wolverhampton, Wolverhampton, UK
6. Avram and Stella Goldstein-Goren Department of Biotechnology Engineering, Ben-Gurion University of the Negev, Beersheba, Israel
7. Neuroscience Research Institute, University of California, Santa Barbara, USA
8. Biomedical Technology and Cell Therapy Research Laboratory, Department of Biomedical Engineering, Faculty of Medicine and Health Sciences, McGill University, 3775 University Street, Montreal, Quebec, H3A 2B4, Canada.
9. Department of Biochemistry and Chemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC, Australia
10. Department of Dental Material Science and Prosthetic Dental Medicine, Faculty of Dental Medicine, Medical University "Prof. Dr. Paraskev Stoyanov", Varna, Bulgaria
11. Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Medical University of Varna, Varna, Bulgaria
12. Department of Biochemistry and Microbiology, Faculty of Biology, University of Plovdiv, Plovdiv, Bulgaria
13. Research Institute, Medical University "Prof. Dr. Paraskev Stoyanov", Varna, Bulgaria.
14. EPAM Systems Ltd, 9 Devonshire Square, 5th Floor, London, EC2M 4YF, UK
15. EPAM Systems d.o.o.. Bulevar Vojvode Misica 17, 5th Floor; 11102, Belgrade, Serbia
16. Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical Faculty, Medical University, Sofia, Bulgaria
17. Department of Biology, Faculty of Pharmacy, Medical University of Varna, Varna, Bulgaria
18. Réponse Inflammatoire et Spectrométrie de Masse - PRISM, University of Lille, Lille, France
19. Department of Medicine, College of Medicine, Chung-Ang University, Seoul, Korea
20. Department of Experimental Medicine, University of Salento, Via Lecce-Monteroni, 73100 Lecce, Italy
21. MAGI'S LAB, Rovereto, Italy
22. Réponse Inflammatoire et Spectrométrie de Masse-PRISM, University of Lille, Lille, France
23. Institute for Genetic Engineering and Biotechnology, University of Sarajevo, Sarajevo, Bosnia and Herzegovina
24. College of Life Sciences, Sichuan University, Chengdu, China
25. School of Engineering, Tallinn University of Technology, Tartu College, Tartu, Estonia
26. Iceland Faculty of Medicine, Division of Life Science, University of Iceland, Reykjavik
27. School of Food Science and Environmental Health, Technological University of Dublin, Dublin, Ireland
28. Department of Biothermodynamics and Drug Design, Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania
29. Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov", Macedonian Academy of Sciences and Arts, Skopje, Macedonia
30. Physics Department, University of Sonora, Blvd Luis Encinas y Rosales S/N, Col. Centro, Hermosillo, Sonora, Mexico.
31. Centre of Biological Engineering, University of Minho, Braga, Portugal
32. Institute of Chemistry, Slovak Academy of Sciences, Bratislava, Slovakia
33. Department of Translational Medicine, Institute of Health Sciences, Bursa Uludag University
34. Department of Medical Genetics, Bursa Uludag University Faculty of Medicine
35. Department of Histology and Embryology, Faculty of Medicine, Bursa Uludag University, Bursa, Turkey
36. Institute of Medical Biology, UiT The Arctic University of Norway, Tromsø, Norway
37. Faculty of Veterinary Medicine, University of Agronomic Sciences and Veterinary Medicine of Bucharest, Bucharest, Romania
38. Department of Paediatrics I, Medical University of Innsbruck, Innsbruck, Austria
39. Division of Human Genetics, Medical University of Innsbruck, Innsbruck, Austria
40. Department of Food Chemistry and Biotechnology, Faculty of Chemistry, Brno University of Technology, Brno, Czech Republic
41. Lab of Clinical Proteomics, PO Fazzi, ASL_ LE, Italy
42. Faculty of Chemistry and Chemical Technology, University of Ljubljana, Ljubljana, Slovenia
43. Molecular Biotechnology Department of Biotechnology, Faculty of Natural Sciences, University of Tirana, Tirana, Albania
44. General Department, University of Thessaly, Larisa, Greece
45. Laboratory of Ecological Genetics and Biotechnology, The National Academy of Sciences of Belarus, Minsk, Belarus
46. Department of Medical Imaging Techniques, Halil Bayraktar Vocational School, Erciyes University, Kayseri, Turkiye
47. European Laboratory for Non-Linear Spectroscopy (LENS), Universitá degli Studi di Firenze, Sesto Fiorentino, Florence, Italy
48. University of Zagreb Faculty of Food Technology and Biotechnology, Zagreb, Croatia
49. Pontificia Universidad Católica de Chile, Facultad de Agronomía e Ingeniería Molecular genetics, Santiago, Chile
50. Department for Plant Breeding and Conservation of Genetic Resources, Bioplant Center, University of Ciego de Ávila Máximo Gómez Báez, Ciego de Avila, Cuba
51. Departamento de Ciencias de la Vida y la Agricultura, Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador
52. Grupo de Investigación en Microbiología y Ambiente GIMA, Universidad de las Fuerzas Armadas, Sangolquí, Ecuador
53. Centro de Nanociencia y Nanotecnología CENCINAT, Universidad de las Fuerzas Armadas, Sangolquí, Ecuador
54. Radioisotope Department, Nuclear Research Center, Atomic Energy Authority, Egypt
55. Department of Food Biotechnology, Agricultural Biotechnology Research Institute of Iran (ABRII), Tabriz, Iran
56. Department of Biomedicine, Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan
57. Animal Biotechnology, Faculty of Agriculture and Veterinary, University of Prishtina Prishtinë, Kosovo
58. Chair of Animal Physiology and Immunology, TUM School of Life Sciences, Munich, Germany
59. Academy of Science of Albania, Tirana, Albania
60. Department of Genetics and Breeding, Institute of Plant Biology and Biotechnology, Almaty, Kazakhstan
61. School of Medicine, University of Dundee, Scotland, UK
62. Medical University, Plovdiv, Bulgaria
63. Institute of Biology and Biotechnology, Faculty of Natural Sciences, University of Ss. Cyril and Methodius, Trnava, Slovakia
64. University of Belgrade, Faculty of Agriculture, Belgrade, Serbia
65. Laboratory of Neurobiology and Tissue Engineering, Russian Center of Neurology and Neurosciences, Moscow, Russia
66. A2M, Health Economics and Valuation, The Netherlands
67. Department of Biotechnology Engineering, Ben-Gurion University, Beersheba, Israel
68. Department of Medical Genetics, University of Health Sciences Sisli Hamidiye Etfal Training and Research Hospital, Istanbul, Turkey
69. Department of Plant Sciences, School of Life Sciences, University of Hyderabad, Hyderabad, Telangana, India
70. Veritas Intercontinental, Madrid, Spain
71. C.I.B., Consorzio Interuniversitario per le Biotecnologie, Trieste, Italy
72. Department of Chemistry, Biology and Biotechnologies, University of Perugia, Perugia, Italy
73. University of Chemical Technology and Metallurgy, Faculty of Chemical Technology, Sofia, Bulgaria
74. Department of Medical Genetics, Near East University Faculty of Medicine, Nicosia, Cyprus
75. Pathology and Veterinary Public Health Department, Institut Agronomique et Vététinaire HASSAN II, Rabat, Morocco
76. Institute of Legal, Political and Sociological Research, State University of Moldova, Chişinău, Moldova
77. Bioinformatics Programme, Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia
78. Department of Biophysics and Radiation Biology, Semmelweis University, Budapest, Hungary
79. Department of Medical Genetics, İstanbul University-Cerrahpaşa, Cerrahpaşa Faculty of Medicine, İstanbul, Türkiye
80. East Mediterranean University, Department of Biological Sciences, Famagusta, Cyprus
81. Department of Biomedical Sciences, Università Degli Studi di Sassari, Sassari, Italy
82. Institute of Chemistry, University of Graz, NAWI Graz, Graz, Austria
83. Trakya Üniversitesi Tıp Fakültesi, Tıbbi Genetik Anabilim Dalı, Edirne, Türkiye
84. Acibadem Labcell Cellular Therapy Laboratory, Istanbul, Turkey
85. Department of Medical Genetics, Uskudar University Faculty of Medicine, Istanbul, Turkey
86. Department of Medical Microbiology, School of Medicine, Acibadem Mehmet Ali Aydinlar University, Istanbul, Türkiye
87. Department of Medical Biotechnology, Graduate School of Health Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Türkiye
88. Department of Medical Genetics, Faculty of Medicine, Tekirdag Namik Kemal University, Tekirdag, Turkey
89. Department of Medical Biology and Genetics, Demiroglu Bilim University, Istanbul, Turkey
90. Department of Molecular Biology and Genetics, Graduate School of Natural and Applied Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Türkiye
91. Department of Molecular Biology and Genetics, Faculty of Engineering and Natural Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Türkiye
92. University of Belgrade, Faculty of Agriculture, Nemanjina, Belgrade, Serbia
93. Department of Biotechnology and Biochemistry, Ogarev Mordovia State University, Saransk, Russia
94. Aksaray University Medical School, Department of Medical Genetics, Aksaray, Turkey
95. School of Engineering, Tallinn University of Technology, Tartu College, Tartu, Estonia

***Corresponding author:**
Munis Dundar
E-mail: dundar@erciyes.edu.tr

DOI: 10.2478/ebtj-2026-0015

Abstract

Synthetic biology has evolved from a set of engineering aspirations to an operationally sophisticated discipline, and artificial intelligence (AI) is its fastest-growing accelerant. This review traces that convergence across six interlocking domains: systems-level biological modeling, de novo protein engineering, metabolic and microbial programming, multi-omics data integration, regulatory element design, and clinical translation. For each domain, we survey established results, integrate findings from 2010–2026 literature, and articulate the trajectories that will define the next decade. Emerging themes include physics-informed neural networks for mechanistically constrained biological modeling, drug design, and federated learning architectures that allow global omics collaboration without centralizing sensitive data, self-driving laboratories that close the Design-Build-Test-Learn loop with minimal human intervention, and large language models that accelerate hypothesis generation from scientific literature. Alongside these opportunities, the review gives equal weight to the governance challenges they create: dual-use risks amplified by generative sequence design, the reproducibility crisis in AI-driven biodesign, and the equitable distribution of autonomous experimentation capacity. The overarching argument is that the promise of synthetic biology, making biological design as deliberate and reliable as any mature engineering discipline, is closer than ever, but will only be realized if technical ambition is matched by scientific rigor, transparent governance, and inclusive access.

Keywords: synthetic biology; artificial intelligence; machine learning; protein engineering; metabolic engineering; multi-omics; self-driving laboratories; biosafety; large language models

Introduction

Biology has historically been an observational discipline that helps us understand what happens in nature, explaining as much as possible the reasons behind these phenomena. Synthetic biology, however, keeps pace with today's technology, going beyond understanding natural processes. Instead of cataloging existing systems, it aims to build new systems using DNA, proteins, and metabolic networks as fundamental elements for deliberate engineering (Khalil & Collins, 2010). The discipline derives its fundamental terminology from engineering, modularity, abstraction, standardization, and applies it to biological cells, aiming to render biological design as methodical as microchip design or software development (A. A. Cheng & Lu, 2012). In reality, the parallel is flawed: Cells exhibit greater noise, heightened context dependence, and significantly less tolerance for design flaws than *in silico* approaches. The disparity between aspiration and reality is, in numerous respects, what propels the discipline onward.

A pivotal moment occurred when the data started to surpass the individuals producing it. Over the past two decades, next-generation sequencing, quantitative proteomics, and extensive metabolomic profiling have inundated laboratories with data at a rate that no team of analysts can manually process (Berger et al., 2016; Gholap et al., 2024). The response indicates a subtle yet significant evolution in biological practice: computational modeling has evolved from a specialized niche to a standard complement to laboratory work, leading to an increasingly blurred distinction between wet-lab experiments and dry-lab simulations (Athanasopoulou et al., 2025; Kacena et al., 2024). At this point, the use of Artificial Intelligence (AI) has become a component that meets and accelerates needs in synthetic biology, opening new horizons (Groff-Vindman et al., 2025; OECD, 2025). Machine learning (ML), specifically, has transitioned from the periphery to the core of this shift. The appeal of these algorithms is apparent; specifically, when confronted with biological datasets characterized by excessive size, inherent noise, or high dimensionality that render conventional statistical approaches inadequate, learning algorithms demonstrate the capacity to discern meaningful patterns (Jiang et al., 2025; Marques et al., 2024; Ozcelik et al., 2024).

These technologies are changing how researchers formulate questions, leading to a more data-driven and algorithmic approach to developing hypotheses (Beardall et al., 2022). Methodologically, AI approaches can be broadly classified into classical ML, deep learning (DL) architectures, and large language models (LLM). Classical ML is typically applied to structured biological datasets for prediction and optimization tasks, such as metabolic flux estimation or phenotype classification. DL methods, particularly convolutional and transformer-based architectures, dominate high-dimensional biological data analysis, enabling sequence- and structure-based modeling of proteins and nucleic acids (Greener et al., 2022; Hein et al., 2025). The practical application of AI in synthetic biology encompasses multiple scales, from molecular design to system-level optimization and automated laboratory workflows. At the molecular level, DL models accurately anticipate the folding of amino acid chains into three-dimensional structures, with a precision comparable to laboratory measurements. This development has enabled the design of proteins and enzymes that are unprecedented in nature, though limitations remain (Jumper et al., 2021). The original AlphaFold 1 software has been iteratively improved and AlphaFold 3 has improved structural prediction of bimolecular interactions (Abramson et al., 2024). More recently, LLM have emerged as powerful tools for knowledge extraction from scientific literature, as well as for generative design tasks in protein engineering and synthetic biology.

At the genome-editing bench, ML assists researchers in selecting CRISPR guide sequences that accurately target specific sites while preserving the integrity of the remaining genome (M. G. Kim et al., 2025). This avoids the complication of unwanted off-target effects. In this framework, the long-context biological foundation model, Evo, was developed as a genera-

tive genomic language model tailored for *de novo* design (Nguyen et al., 2024), with the primary objective of generating complex macromolecular systems, such as functional CRISPR-Cas loci, from scratch. Utilizing an autoregressive architecture, Evo is capable of decoding, predicting, and designing sequences across DNA, RNA, and protein modalities, seamlessly spanning from the molecular to the whole-genome scale (Thomson et al., 2025). In metabolic engineering, analogous algorithms delineate biosynthetic pathways that direct cellular resources towards useful products, biofuels, specialized chemicals, and medication precursors. The algorithms help minimize trial and error compared to conventional methods (J. Chen et al., 2025). In addition, such algorithms can successfully be used to design microbial biopolyesters of pre-defined composition and properties, thus paving the way towards smart bioplastics, which are integratively embedded into the circular bioeconomy of the future (Ligarda-Samanez et al., 2026; Sokra & Meta, 2026). Biofoundry platforms, which are robotic facilities, bring together these elements. They allow the Design Build Test Learn (DBTL) cycle to be repeated thousands of times with little direct experimental involvement (Rai et al., 2024). In terms of translation, integrating patient-specific omics data with predictive models is starting to create digital representations of diseases (Table 1). These representations could eventually help with personalized treatment (Puniya, 2025).

The identical generative algorithms capable of designing beneficial proteins might, theoretically, also generate perilous pathogen sequences. The concern is that AI-assisted biological design tools pose novel risks to human health. The international community, spearheaded by organizations such as the OECD, is currently deliberating on how to regulate this risk without impeding innovation (IGSC, 2024; OECD, 2025). This review traces the development of ML, starting with the atomic coordinates that define a protein's structure and ending with the policy frameworks that govern its safe use (Figure 1). It also evaluates the current state of the field and identifies areas that still need research. To provide a unifying perspective, the convergence between AI and synthetic biology can be interpreted across three interconnected layers: (i) *representation*, where biological systems are encoded into machine-readable formats (e.g., sequences, structures, multi-omics data); (ii) *prediction and design*, where AI models infer patterns and generate novel biological constructs; and (iii) *execution*, where automated platforms and synthetic systems implement these designs in physical or clinical contexts. This layered framework highlights that progress in the field is not merely driven by better algorithms, but by tighter coupling between data, models, and experimental validation pipelines. Throughout this review, these layers serve as an implicit structure to contextualize advancements and identify bottlenecks. A comprehensive synthesis of this functional relationship, which integrates key AI tools—such as Enformer, RFdiffusion, Cello, and ClioMD—across these layers from molecular scales to clinical implementation, is presented in the Master Integration Table (Table 2).

Table 1. Comparison of Traditional Methods and AI-Augmented Approaches

Feature	Traditional Approach	AI-Augmented Approach
Design Process	Trial-and-error, rational design; limited use of natural sequences	Automation of DBTL cycles; physics-informed and generative models
Efficiency	Laboratory studies lasting months; high costs	Rapid *in silico* screening of high-success candidates; self-driving labs
Data Usage	Reductionist; single-layer data analysis	Multi-omics integration; federated learning; hybrid mechanistic-ML models
Error Tolerance	Low tolerance; design flaws often only discovered experimentally	Bayesian uncertainty quantification; digital twins; benchmarked model standards
Governance	Reactive regulation; sequence-homology-based screening	Function-based screening; AI watermarking; agile biosafety frameworks

2. Modeling and Optimization of Biological Systems Using AI

Biological systems are high-dimensional, nonlinear structures in which genetic variations, epigenetic regulations, metabolic networks, and microbial interactions are intricately intertwined. Traditional reductionist approaches often fall short in understanding the integrative dynamics of these systems. The emergence of Big Data driven by advances in omics technologies and high-throughput screening methods has brought AI and ML methods to the forefront of biological research (Volk et al., 2020). AI has become a powerful tool that not only enables the characterization of biological systems but also allows for estimating their future behavior and optimizing them for desired

Table 2. AI-Powered Synthetic Biology: Master Integration Table

Convergence Layer	Model / Tool Name	Biological Scale	Function & Engineering Breakthrough
REPRESENTATION (Encoding Biology)	**Enformer**	Genomic (100kb)	Predicts tissue-specific gene expression and the effects of non-coding variants using long-range DNA interactions.
	Geneformer	Cellular	Pretrained on 30 million single-cell transcriptomes to nominate therapeutic targets via network biology patterns.
	AlphaMissense	Proteomic / Variant	Assigns benign or pathogenic labels to 89% of all possible human missense variants using structural and allele-frequency data.
	PPML-Omics	Multi-omics	Enables privacy-preserving multi-omics collaboration across institutions using federated learning.
PREDICTION & DESIGN (Discovery & Generation)	**AlphaFold 3**	Molecular (All-atom)	Provides joint structure prediction for complexes of proteins, nucleic acids, ligands, and ions.
	ESM-3	Protein	Generates novel functional proteins by jointly processing sequence, structure, and function token streams.
	Evo / Evo 2	Genome Scale	Designs functional CRISPR-Cas loci and synthetic DNA sequences up to the length of a small bacterial chromosome.
	RFdiffusion	Protein Backbone	Constructs *de novo* protein backbones under predefined functional constraints, such as receptor binding sites.
	ProteinMPNN	Protein Sequence	Redesigns protein sequences for specific backbones with high accuracy, recovering native sequences at roughly 52%.
	MODIFY	Enzyme Engineering	Facilitates combinatorial library design by co-optimizing sequence diversity and predicted enzyme fitness.
	ProtGPT2 / ProGen	Protein (LLM)	Demonstrates that autoregressive language models can produce functional, non-natural protein sequences across diverse families.
EXECUTION (Implementation & Automation)	**Cello**	Genetic Circuits	Compiles Verilog-style Boolean logic into DNA; 92% of predicted output states were experimentally validated on the first build.
	ART (Automated Recommendation Tool)	Metabolic Systems	Uses Bayesian machine learning to recommend the next microbial strains to build, reducing trial-and-error in production cycles.
	ClioMD	Clinical Diagnostic	Accelerates screening and variant interpretation for ciliopathies by integrating clinical data with model organism networks.
	CURATE.AI	Clinical / Dosage	Optimizes personalized medicine by determining precise, low-toxicity drug doses based on individual patient responses.
	BioDiscoveryAgent	Laboratory Agent	Boosts gene perturbation experiment efficiency by 46% through automated literature mining and experimental planning.

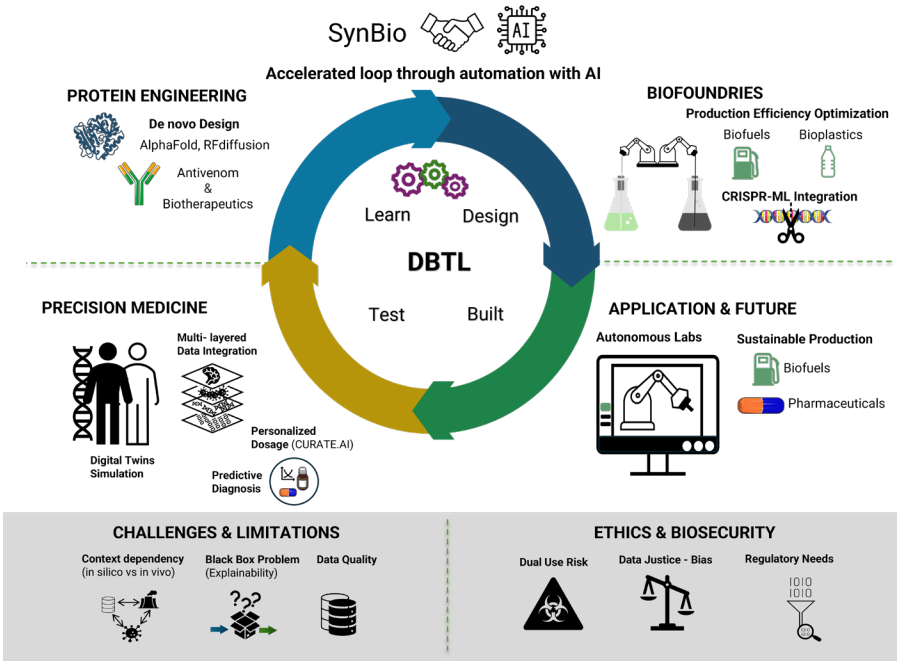

Figure 1. AI and Synthetic Biology Integration. The central DBTL (Design-Build-Test-Learn) cycle illustrates the automated loop driven by AI to streamline biological discovery. (Top Left) Protein Engineering via de novo design (e.g., AlphaFold, RFdiffusion). (Top Right) Biofoundries optimize production efficiency for biofuels and bioplastics through CRISPR-ML integration. (Bottom Left) Precision Medicine leveraging digital twins and personalized dosage platforms (CURATE.AI). (Bottom Right) Future Applications focused on autonomous labs and sustainable production. The lower panels address critical Challenges & Limitations (e.g., Black Box Problem) and Ethics & Biosecurity concerns (e.g., Dual Use Risk, Regulatory Needs). SynBio: Synthetic biology, ML: Machine learning, AI: Artificial intelligence. Some of items are used from Smart servier (Image(s) provided by Servier Medical Art (https://smart.servier.com), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/)).

industrial or clinical goals (Lawson et al., 2021).

A critical challenge in modeling biological systems is how to integrate data from different scales and formats into a single model. For instance, in understanding complex pathologies such as cancer, the representational power of DL can be combined with the ability of pathway modeling to identify interactions. This approach maps the functional relationships between genetic mutations and tissue-level phenotypic changes with much higher accuracy than classical statistical models (Ing et al., 2025). Furthermore, AI-based models can analyze the heterogeneous structure of cancer and the resistance mechanisms determining treatment response by integrating multi-layered omics data with cellular network dynamics, thereby offering a powerful approach for developing personalized biomarkers and targeted therapies within the scope of precision oncology (Nilsson et al., 2025).

A particularly consequential frontier that was underrepresented in earlier reviews is the integration of physics-informed neural networks (PINNs) into biological modeling. Unlike purely data-driven approaches, PINNs embed known physical and biochemical laws, ordinary and partial differential equations governing gene expression, metabolic flux, and cellular dynamics, directly into the neural network training process as loss function constraints. This allows mechanistically sound extrapolation even when experimental data are sparse or noisy (Farea et al., 2025). For example, PINNs have been successfully applied to hybrid modeling of microbial fermentation processes, such as β-carotene production in *Saccharomyces cerevisiae*, where they integrate kinetic models with experimental data to improve prediction accuracy and process optimization (Bangi et al., 2022). Biologically-informed neural networks (BINNs), an extension of PINNs, further adapt this framework so that mechanistic terms themselves are learned from data, removing the need to specify governing equations in advance, a particularly valuable capability for complex gene regulatory networks where causal relationships remain incompletely characterized. In a similar context, BINNs approaches have also been applied to gene regulatory modeling in *S. cerevisiae*, where neural ordinary differential equations (ODEs) frameworks are used to infer transcriptional dynamics from time-resolved expression data while leveraging prior knowledge of regulatory network structure (Hossain et al., 2024).

Experimental data assembly into databases using the FAIR principles, versioning systems, and containing high-quality raw data is an important emerging step to enable the application of AI-ML principles in biology. For example, the Protein-Ligand Binding Database (plbd.org) (Lingė et al., 2023) has enhanced our understanding of the protein – compound recognition and

promoted anticancer drug discovery (Zubrienė et al., 2026).

Beyond physics-informed and biologically informed neural network frameworks, hybrid mechanistic–data-driven modeling approaches are also widely used in systems and synthetic biology. While mechanistic models define the fundamental framework of biological systems, deep neural networks (DNN) fill in the complex parts of the system, those that cannot be fully understood or mathematically expressed, by learning from data. Combining the transparency of mechanistic models with the flexibility of ML results in more accurate, robust, and interpretable models. Used in industrial processes since the 1990s, this method is gaining increasing popularity in systems biology today in areas such as gene function prediction, metabolic networks, and COVID-19 dynamics (Pinto et al., 2023). The resulting data–model–data cycle not only analyzes complex biological data but also transforms it into biologically meaningful information, enabling system-wide repercussions of a single metabolic intervention to be modeled through a globally standardized Systems Biology Markup Language (SBML) infrastructure (DiNuzzo, 2022). Complementary to these approaches, hybrid mechanistic–data-driven models have been successfully implemented in yeast fermentation systems under mixed carbon sources, combining mass-balance and kinetic descriptions with data-driven components to improve process prediction and optimization (Valencia-Velásquez et al., 2025; Yildirim et al., 2026).

The optimization of biosystems is of critical importance, particularly for ensuring the economic sustainability of biotechnological production processes. AI accelerates this process by automating the DBTL loop (Y. Cheng et al., 2023). In a study using *Pseudomonas putida*, by combining laboratory automation and ML cycles, isoprenol production was increased fivefold after just six DBTL cycles (Carruthers et al., 2025). The Automated Recommendation Tool (ART), which combines ML with Bayesian methods to recommend which strains to build next from a small set of past observations, was an early systematic example of this idea and has been validated on metabolic targets including renewable biofuels, fatty acids, and tryptophan production in microbial hosts (Radivojević et al., 2020). These data provide concrete evidence of the significant time and cost advantages AI potentially offers over traditional trial-and-error methods. Causal inference, moving beyond statistical correlation toward a mechanistic understanding of gene regulatory networks, represents a critical emerging need. While correlation-based ML models excel at pattern recognition, they can propose interventions that fail in complex cellular contexts where directionality matters. Coupling causal discovery algorithms with dynamical systems models is expected to produce more robust and generalizable predictions (Maizels & Briscoe, 2026). A critical unresolved challenge lies in distinguishing correlation from causation in biological systems. Recent advances in causal ML, including structural causal models and intervention-aware neural networks, aim to infer directional relationships rather than statistical associations. This is particularly important in gene regulatory networks, where perturbation effects are inherently asymmetric and conditional. Integrating causal inference with experimental perturbation datasets (e.g., CRISPR screens) is expected to significantly improve the reliability of AI-driven biological estimation.

AI has ushered in an era in which biological systems are not merely described, but engineered and optimized. Computer-based, system-wide modeling of cells and molecular networks enables the rapid evaluation of numerous biological scenarios that cannot be tested experimentally. This approach holds the potential to revolutionize the fields of medicine, precision agriculture, and biotechnology (Bhardwaj et al., 2022). While system-level modeling provides the predictive backbone for biological understanding, its true potential is realized when coupled with design-oriented applications such as protein engineering.

3. Protein Engineering: *De Novo* Design

The rapid evolution of AI-driven protein engineering is characterized by several convergent methodological advancements. These include the generative modeling of primary sequences, backbone structures, and atomic coordinates; the adaptation of foundational models to design proteins with specific functional properties; the development of sophisticated protein representations for downstream scoring and fitness prediction of candidate sequences; and the refinement of library design techniques, including synthesis-aware approaches (Listgarten & Jiang, 2026). Collectively, these advancements mark a fundamental shift from the modification of naturally occurring proteins toward a *de novo* design paradigm focused on creating sequences from scratch for specific functions (Zhang et al., 2025). While traditional protein design focuses on enhancing function by introducing mutations within a limited region of naturally occurring protein backbones, modern approaches enable the design of protein structures and functions beyond those that are known to occur in nature, using DL and AI algorithms (Notin et al., 2024; Zhang et al., 2025). Protein structure prediction models like AlphaFold have enabled the fundamental computational infrastructure required for *de novo* protein design through advanced ML approaches that integrate physical principles and biological knowledge with evolutionary signals derived from multiple sequence alignments (Jumper et al., 2021). By using these computational models, researchers can generate protein folds and functional regions that have no natural counterparts in an extremely short time *in silico*.

Despite remarkable progress in structural prediction, a persistent gap remains between protein structure and function. Functional properties such as catalytic efficiency, allosteric regulation, and cellular context dependence are not fully determined by static structures alone. Emerging approaches aim to integrate dynamic simulations, fitness landscapes, and experimental high-throughput screening data into AI models, thereby shifting protein engineering from structure-centric to function-centric paradigms.

Deep network hallucination emerged as a revolutionary

method in *de novo* protein engineering. By feeding random amino acid sequences into structure prediction networks and optimizing the difference between the distance maps predicted by the network and ideal protein distributions, stable, unique, and novel folded proteins are produced (Anishchenko et al., 2021). Controlled diffusion-based models such as RoseTTAFold diffusion (RFdiffusion) have provided a significant methodology that *de novo* constructs the protein backbone under predefined functional constraints, such as receptor binding sites or enzyme active sites (Watson et al., 2023), hereby improving the efficiency of functional motif scaffolding compared to classical approaches, though experimental success rates vary considerably depending on target complexity, and widely used in silico evaluation metrics such as self-consistency refolding pipelines have recognized limitations in reliably predicting experimental outcomes, particularly for designs sharing homology with natural sequences (Korbeld et al., 2026). These diffusion-based networks significantly reduce reliance on classical structure prediction models, enabling the *de novo* design of new folding topologies that are thermodynamically stable but have not yet been identified in the natural protein repertoire (Liu et al., 2024).

The sequence-design step that complements such backbone generation has also been overhauled. ProteinMPNN, a message-passing graph neural network trained on backbones from the Protein Data Bank, recovers native sequences with about 52% accuracy compared with about 33% for the physics-based Rosetta pipeline. It has also been used to redesign Rosetta- and AlphaFold-derived constructs whose initial sequences failed in experimental tests, including single-chain proteins, cyclic symmetric oligomers, designed nanoparticles, and engineered target binders (Dauparas et al., 2022).

To expand this structural-biology infrastructure beyond the protein-only limitations of earlier architectures, recent models have advanced toward modeling full biomolecular assemblies. AlphaFold 3 replaces the original Evoformer with a Pairformer module and adds a diffusion-based output stage. With this design, the model predicts not only protein-only structures but also structures that include nucleic acids, small-molecule ligands, ions, and chemically modified residues, often more accurately than tools designed only for docking or only for nucleic-acid prediction (Abramson et al., 2024). RoseTTAFold All-Atom takes a related but distinct approach: Protein and DNA residues are encoded as sequence-style tokens, while small molecules and other chemical groups are represented as atomic graphs, so that ligands, ions, and covalent modifications are handled inside one network. Its diffusion variant, RFdiffusion All-Atom, has produced experimentally validated proteins that bind a range of cofactors, from the steroid digoxigenin to heme- and bilin-class chromophores-targets that earlier protein-only methods could not reach (Krishna et al., 2024). The practical implication is that the design unit has expanded from the polypeptide chain alone to the full biomolecular assembly.

Protein Language Models (PLMs), which process biological sequences in a manner similar to natural language, have established a new paradigm. Generative models such as ProtGPT2 enable the design of *de novo* protein sequences that exhibit natural protein-like properties but have no evolutionary counterparts in nature (Ferruz et al., 2022). ProGen2 can generate unique sequences that have been experimentally proven to be functional, thanks to statistical rules learned from massive genomic datasets (Nijkamp et al., 2023). Among generative protein language models, ESM-3 differs from earlier sequence-only systems because it represents three protein attributes, amino-acid sequence, three-dimensional structure, and functional annotation, as parallel discrete-token streams that are processed together by a single masked transformer. In a public demonstration, ESM-3 produced a green fluorescent protein, called esmGFP, whose amino-acid sequence is far enough from any known fluorescent protein that the difference would correspond, at the substitution rates inferred for that family, to roughly five hundred million years of natural evolution (Hayes et al., 2025). The earlier ProGen study, although superseded by larger successors, retains its importance as the first experimental demonstration that an autoregressive language model trained on hundreds of millions of protein sequences can produce engineered enzymes with catalytic activity comparable to natural homologues, even when the engineered sequence shares only about thirty percent of its amino acids with any known protein (Madani et al., 2023). Together, these studies mark a shift from models that imitate the natural protein universe to models that explore regions of sequence space that natural evolution has never visited.

Reflecting this evolutionary exploration, a prime example of this shift is the AI-driven development of synthetic antidotes for snake venoms, which offer faster, cheaper, and more stable alternatives to traditional animal-based production (Callaway, 2025). Specifically, *de novo* designed "miniproteins" capable of neutralizing the 3FTx toxin family have emerged as potent biotherapeutics; these structures are smaller than natural antibodies, thermally more stable, and exhibit high binding affinity. Such advancements demonstrate that these computational approaches are transitioning into real-world biomedical breakthroughs (Vázquez Torres et al., 2025).

Three emerging directions deserve explicit attention. First, self-supervised learning on non-coding sequences and metagenomic data promises to expand the training universe for PLMs far beyond the curated protein databases currently used, uncovering functional motifs in the vast dark matter of genomic sequence space. Second, multimodal models that jointly process sequence, structure, functional annotations, and fitness landscape measurements are beginning to outperform unimodal approaches on difficult engineering tasks (Jin et al., 2025). Third, continuous learning frameworks, in which models are iteratively retrained as new experimental structures are deposited in databases such as the Protein Data Bank, represent a paradigm shift from static to living model architectures. This is critical given that experimental protein structure databases are growing faster than any single training cycle can capture (Figure 2).

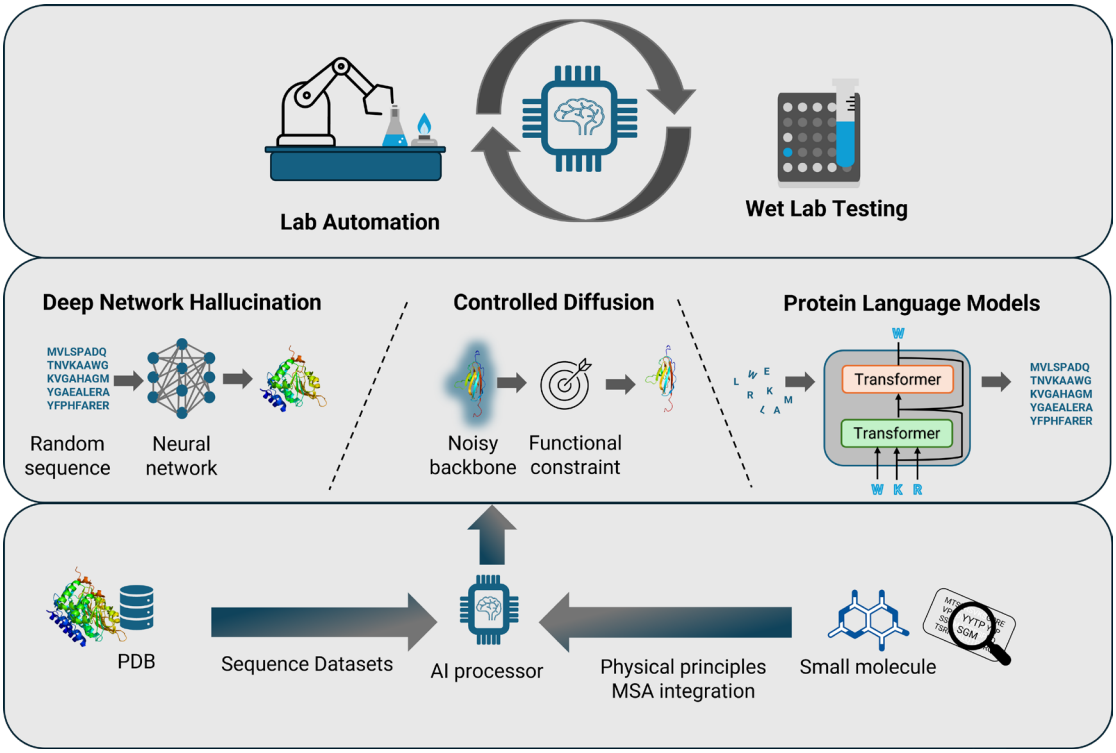

Figure 2. Integration of AI and de novo protein design. The workflow depicts the hierarchy of AI-driven protein engineering, starting from the integration of structural (PDB) and sequence datasets with physical principles. Three primary generative paradigms are highlighted: Deep Network Hallucination for generating novel folds from random sequences; Controlled Diffusion for structural refinement under functional constraints; and Protein Language Models utilizing transformer architectures for sequence optimization. The computational predictions are seamlessly integrated into a closed-loop system featuring Lab Automation and Wet Lab Testing for high-throughput functional validation. PDB: Protein data bank, AI: Artificial intelligence

Advances in multimodal AI approaches supported by laboratory automation may contribute to the faster and more rational generation of protein sequence and structure candidates tailored to specific functional requirements in the future (J. Chen et al., 2025). In this context, implementation of microfluidics that enables high-resolution experimental characterization of biocatalysts, including their activity, kinetics, stability, and structural properties, generating rich and detailed datasets, is of particular benefit. ML can then leverage this information to uncover patterns, perform advanced data analysis, and make predictive models that guide optimization. This technological trend is making it increasingly feasible to view proteins not merely as static biological components, but as functional molecular systems that can be designed to meet specific needs. In the near future, the integration of microfluidics and ML is expected to open new perspectives on enzyme function, evolutionary processes, and catalytic mechanisms, while significantly accelerating the discovery and rational design of highly efficient biocatalysts for diverse industrial applications (Vasina et al., 2023). Modern AI-assisted enzyme engineering, together with advanced reaction engineering strategies, has enabled biocatalysis to operate under demanding conditions such as high temperature, extreme pH, and organic solvent environments. At the same time, progress in enzyme engineering have significantly reduced both the cost and time required to develop enzymes suitable for industrial use (Wahab, 2025). Beyond enzyme engineering, AI is also transforming retrosynthetic planning for chemoenzymatic routes, while improved biocatalyst informatics is speeding up enzyme discovery, selection, and process optimization (Gazis et al., 2026). These advances in protein design naturally extend toward pathway-level manipulation, where engineered enzymes are embedded into cellular systems.

4. Metabolic Engineering and Microbial Programming

Metabolic engineering and microbial programming aim to develop sustainable methods for producing chemicals, fuels, and materials from renewable resources (Cho et al., 2018; Donato et al., 2023). Over time, this field has evolved from classical methods toward more comprehensive, system-level approaches. Today, metabolic engineering not only improves production but also enables the more sustainable and efficient production of a wide range of valuable products, from amino acids to biofuels, from pharmaceuticals to bioplastics and silk-like proteins (Dudeja et al., 2025; S. H. Jeong et al., 2023; X. Ji et al., 2025; Yadav et al., 2025). Metabolic engineering, which integrates systems biology, synthetic biology, and evolutionary engineering, offers an integrative approach that accelerates the design and optimization of metabolic pathways. Through the integration of omics data and computational tools, it enables

the development of microbial cell factories for the efficient production of both natural and synthetic compounds (Lee et al., 2012). The scope of AI-assisted microbial programming extends beyond intracellular pathway optimization to encompass community-level and ecosystem-scale interventions, including microbiome engineering and sustainable agricultural applications. RNAi and microbiome engineering offer highly precise ecosystem management that biotechnologically optimizes soil fertility while suppressing biotic stressors at the molecular level. This synergy is restructuring food security in line with the "probiotic agriculture" vision, freeing agricultural systems from dependence on synthetic chemicals (Yildirim et al., 2026). However, as climate change intensifies the drive to deploy engineered microbial cell factories as powerful biotechnological tools for pollution mitigation, understanding their long-term ecological repercussions remains paramount. Thoroughly evaluating the impacts of these biotechnologically developed and AI-optimized microorganisms on indigenous biodiversity and wider ecosystems is mandatory to ensure that climate mitigation strategies do not inadvertently destabilize natural ecological networks (Celebi et al., 2023).

A turning point in the programming of microorganisms was the discovery of the CRISPR/Cas system, the bacterial adaptive immune system (S. H. Jeong et al., 2023). While canonical CRISPR/Cas9 induces stable genomic changes, CRISPRi/CRISPRa variants enable transient regulation without DNA modification; collectively, these approaches allow researchers to efficiently orchestrate metabolic flux in bacteria and eukaryotes (Cho et al., 2018). System-level modeling for CRISPR-based metabolic engineering now integrates transcriptional interference, activation, and genome-wide regulatory logic into quantitative frameworks that guide multiplex editing strategies (Cardiff et al., 2024).

Thanks to these CRISPR-based systems, model organisms such as *Escherichia coli*, *Streptococcus* sp., *Corynebacterium* sp., and *Bacillus* sp. can be transformed into highly precise cell factories capable of producing amino acids, biofuels, pharmaceutical compounds, succinic acid, and polyhydroxyalkanoate (PHA) biopolyesters (C. Chen & Zheng, 2023; Dudeja et al., 2025; Palit et al., 2025). In the field of PHA biopolyesters, which are auspicious polymeric materials accumulated by numerous predominantly prokaryotic microbial strains as intracellular storage compounds and stress protectants, AI paves new avenues for production process intensification. In this emerging field, AI enhances the metabolic engineering of microbial cell factories, predicts polymer properties needed for specific applications as a consequence of the intramolecular polymer structure, and automates bioreactor control, including pH-value, temperature, dissolved oxygen tension, and substrate supply, thus improving PHA production both quantitatively and qualitatively (Kalia et al., 2025). Artificial Neural Networks and genetic algorithms can be applied to optimize nutrient supply and process parameters during cultivation. Especially for the utilization of inexpensive feedstocks for PHA production, AI helps to tap the wealth of diverse agro-industrial waste streams and even greenhouse gases like CO_2 and methane, a pre-condition to make PHA production cost-efficient (Fukala & Kučera, 2024). Improvement of PHA-producing microbial production strains happens via AI-driven metabolic engineering and CRISPR-based genome editing (Dudeja et al., 2025). Moreover, non-invasive real-time monitoring of intracellular PHA accumulation can be realized via combining AI with advanced analytical sensors, such as 2D fluorescence probes (Guarda et al., 2025) or Fourier-transform infrared (FTIR) spectroscopy (Sola et al., 2025), thus providing viable tools for real-time process fine-tuning and adjustment of cultivation parameters.

Beyond PHA biopolyesters, AI is transforming a broader spectrum of industrial biotechnology applications that merit equal attention. In recombinant protein manufacturing, deep learning models leveraging protein and genomic foundation models have demonstrated accurate prediction of heterologous soluble protein expression in *E. coli*, achieving an AUROC of 0.83 on prospective validation and correctly identifying successfully expressing constructs in 92% of cases, directly reducing the experimental burden of construct screening (Tankhilevich et al., 2026). For glycosylation monitoring, a critical quality attribute in biopharmaceutical manufacturing, ML algorithms such as XGBoost have enabled rapid, automated quantification of mAb N-glycans from fluorescence detector data with R^2 above 0.95, reducing data analysis time by approximately 70% while maintaining accuracy comparable to LC-MS reference methods (Shrivastava et al., 2024). For enzyme engineering, ML-guided cell-free expression platforms have enabled six parallel engineering campaigns to be completed in just one week per compound, with ML-predicted variants demonstrating up to 42-fold improved activity relative to parent enzymes across nine small molecule pharmaceutical targets (Landwehr et al., 2025). In biologics development, including monoclonal antibodies, AI models are contributing to sequence design, affinity optimization, structural prediction, and developability assessment, with the emerging consensus that combining AI-driven computational tools with experimental validation represents the most effective strategy for next-generation biologics development (Ammar et al., 2026). Although much of the published work centres on *E. coli*, the same modelling approaches are being extended to the eukaryotic hosts that dominate biopharmaceutical manufacturing. In *Pichia pastoris*, artificial neural networks combined with response surface methodology have been used to optimize medium composition for recombinant monoclonal antibody production (Das et al., 2025), and language-model-based codon optimization now improves secreted expression titres across diverse protein classes in this host (Narayanan & Love, 2026). For Chinese hamster ovary (CHO) cells, the workhorse of therapeutic antibody production, neural network models trained on historical cultivation runs have identified process conditions that raise viable cell density and product titre beyond standard operating points (Richter et al., 2025). This spread of a common set of predictive tools across bacterial, yeast, and mammalian platforms is one of the clearer trends in the field, and it reflects a growing consensus that

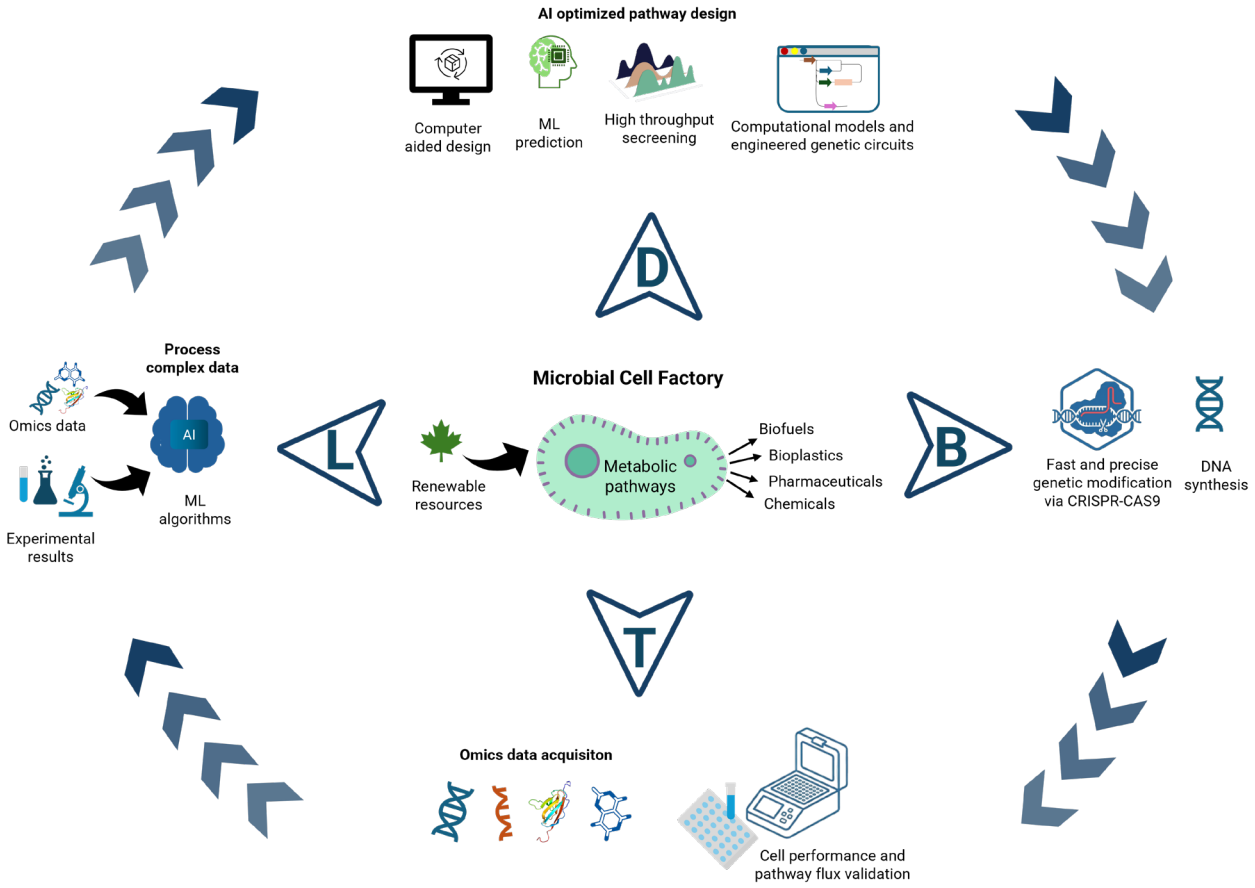

Figure 3. Integration of AI and metabolic engineering and microbial programming. The schematic illustrates the iterative workflow for developing microbial cell factories. (D) Design: AI-optimized pathway design through ML predictions and computational modeling. (B) Build: Fast and precise genetic modification using CRISPR-Cas9 and DNA synthesis. (T) Test: High-throughput omics data acquisition and pathway flux validation. (L) Learn: Integration of experimental results and complex omics data via ML algorithms to inform the next design iteration. The cycle enables the efficient conversion of renewable resources into high-value bioproducts such as biofuels and pharmaceuticals. ML: Machine learning, AI: Artificial intelligence

host-agnostic sequence and process models can be retargeted to whichever expression system a given product requires.

The integration of ML with mechanistic models has transformed metabolic engineering into a more predictable and rapid tool. ML models learn the relationships between protein sequences and their properties and reduce the experimental burden by performing *in silico* fitness predictions. For example, algorithms like MODIFY accelerate the discovery of enzyme functions not found in nature by simultaneously optimizing sequence diversity and predicted fitness values (Ding et al., 2024). ML platforms integrated with cell-free systems enable the functional screening of thousands of enzyme variants, facilitating the parallel optimization of biocatalysts (Landwehr et al., 2025).

Three gaps identified in current practice merit direct discussion. First, spatial metabolic modeling, accounting for the distinct enzymatic milieus within subcellular compartments such as the cytoplasm, mitochondria, and endoplasmic reticulum, remains poorly captured by genome-scale models that treat the cell as a homogeneous reactor. Emerging spatial metabolic frameworks that couple subcellular compartmentalization with flux balance analysis offer a path toward more predictive cell factory design. Second, the deployment of ML-optimized strains into bioreactors at industrial scale remains largely manual; real-time AI control systems that adjust feeding rates, temperature, and aeration in response to online metabolic sensors represent an important engineering frontier. Third, explainable AI (XAI) methods, notably SHAP (SHapley Additive exPlanations) values, which decompose complex model outputs into individual feature contributions based on cooperative game theory, alongside attention visualization and mechanistic feature attribution, are increasingly being applied to metabolic flux predictions. This approach provides the regulatory clarity required for industrial and clinical approval of bioprocesses, allowing researchers to move beyond 'black box' models toward a mechanistic understanding of cellular engineering.

Metabolic engineering and microbial programming integrate various engineering disciplines, including systems biology, synthetic biology, CRISPR-based genome editing, and ML. Thanks to this integration, flexible, data-driven, and highly efficient cellular factories for the production of both natural and synthetic chemicals can be developed (Figure 3) (Cardiff et al.,

2024; K. J. Jeong, 2025; S. H. Jeong et al., 2023; G. B. Kim et al., 2023; Landwehr et al., 2025; Qin et al., 2025; Rai et al., 2024).
While this section is organized around metabolic flux control and microbial consortia, the wider project of programming microorganisms also depends on automated genetic-circuit design. The Cello compiler converts a Verilog-style description of Boolean logic into DNA, using a curated library of insulated transcriptional logic gates. When tested on sixty separate circuits in *E. coli*, the constructs were assembled directly from the software output without further tuning; about 92% of output states behaved as predicted, and forty-five of the sixty circuits passed every line of their truth table on the first build (Nielsen et al., 2016). Cello has since become a reference point for the claim that synthetic biology can be made to resemble electronic design, and it illustrates the conditions for successful design automation: a well-characterized parts library, a formal Boolean abstraction of biological function, and gates that work reliably across genetic contexts.

A second category of advance, distinct from circuit-level automation, involves genome-scale foundation models that learn directly from raw DNA sequences. Evo is a seven-billion-parameter model trained at single-nucleotide resolution on 2.7 million microbial genomes, with a context window of 131 kilobases. Without task-specific fine-tuning, it predicts function across DNA, RNA, and protein sequences, and was used to design matched protein and RNA components of CRISPR-Cas systems within a single end-to-end generation step, an outcome that had not previously been achieved with one language model (Nguyen et al., 2024). Its successor, Evo 2, was trained on 9.3 trillion bases from more than 128,000 whole genomes covering all three domains of life. It scores human disease-associated variants accurately and can generate synthetic DNA sequences whose length approaches that of a small bacterial chromosome (Brixi et al., 2026). For metabolic engineering and microbial programming, the practical consequence is that the design unit no longer needs to be restricted to single enzymes or single pathways, but can extend to whole regulatory cassettes and, in principle, to entire synthetic operons.

4.1. AI-Driven Bioprocess Engineering and Industrial Scale-Up

The translation of AI-optimized microbial strains from laboratory-scale discoveries to commercially viable bioprocesses represents one of the most critical yet underaddressed bottlenecks in the field. While strain engineering and pathway optimization have benefited substantially from machine learning, the subsequent challenges of fermentation development, process intensification, downstream processing, and industrial-scale deployment remain disproportionately reliant on empirical trial-and-error approaches. A key barrier is that strains optimized through laboratory DBTL cycles frequently underperform in stressed industrial bioreactors, where scale-dependent phenomena introduce complex nonlinearities that laboratory-scale models cannot fully capture (Karimi Alavijeh et al., 2024; Yuan et al., 2026). Bridging this gap requires ML-driven fermentation protocols specifically designed to translate laboratory insights into industrial-scale production (H. Li et al., 2025).

Hybrid modelling frameworks, combining mechanistic descriptions of biological systems with data-driven components, have emerged as a particularly effective strategy for bridging this gap under conditions of data scarcity. A joint ML-metabolic modelling approach validated on heterologous peptide production in *E. coli* demonstrated that mechanism-informed predictions can complement experimental design and guide bioprocess development even with small datasets, with applicability spanning both proof-of-concept and industrial settings (Zampieri et al., 2026). At pilot scale, hybrid models integrating mechanistic oxygen transfer descriptions with ML-based soft sensors for viscosity prediction have achieved an R^2 of 0.92 using only online measurements, reducing experimental effort from nine to two fermentations while maintaining predictive accuracy across multiple strains at 550 L scale (Lemperle et al., 2026). For fed-batch processes, deep learning models incorporating sequential spectral features from online Raman spectroscopy have enabled automated glucose feeding control, improving ethanol titers by 3.85% over conventional fed-batch strategies while reducing glycerol byproduct formation (K. Ji et al., 2025).

Beyond monitoring and control of upstream processes, reinforcement learning is beginning to enter industrial bioprocess operation. In what represents, to the best of current knowledge, the first industrial validation of an RL-based control strategy in a bioprocess, a behavior cloning-initialized RL controller for pH regulation in an open photobioreactor reduced control error by 8% compared to conventional PID control and decreased control effort by 54%, demonstrating robustness across 8 days of operation under varying environmental conditions (Gil et al., 2026). Such results illustrate the potential of RL-based methods to handle the inherent nonlinearities and disturbance-prone dynamics of real industrial systems, though broader applicability across diverse fermentation formats remains to be established.

Downstream processing, encompassing cell separation, product recovery, and purification, represents an equally important but comparatively underexplored application domain. In biopharmaceutical manufacturing, machine learning-enhanced design space identification for Protein A chromatography has demonstrated strong performance under high data availability conditions, with transfer learning from synthetic datasets enabling robust design space classification even under limited experimental data availability, thereby reducing wet-lab reliance in early-stage process development (Galeazzi et al., 2025). For monoclonal antibody development specifically, AI models are contributing to sequence design, affinity optimization, structural prediction, and developability assessment, with the emerging consensus that combining AI-driven computational tools with experimental validation represents the most effective strategy for next-generation biologics development (Ammar et al., 2026). For recombinant protein production in *E. coli*, Long Short-Term Memory (LSTM)-based deep learn-

Table 3. Multi-Omics Data Integration Strategies

Integration Approach	Implementation Method	Advantages and Challenges
Alignment-based	Direct matching of data while preserving original structure	Advantages: Practical starting point. Challenges: Scale differences create technical challenges.
Transformation-based	Mapping data into common mathematical space (DL/ML)	Advantages: Uncovers hidden patterns. Challenges: Can reduce biological interpretability.
Network-based	Modeling interactions among biological entities	Advantages: Holistic, multidimensional framework. Challenges: High computational cost.
Federated / Privacy-preserving	Collaborative model training across institutions without data centralization	Advantages: Enables multi-lab pooling; legal compliance. Challenges: Some performance trade-offs *(Malpetti et al., 2025).*
Longitudinal / Time-series	Recurrent neural architectures on time-ordered omics snapshots	Advantages: Captures dynamic cellular states. Challenges: Demands dense sampling; data alignment is challenging.

ing models trained on critical process parameter time series have demonstrated real-time yield estimation capabilities, paving the way for ML-based fermentation control algorithms in upstream biopharmaceutical processes (Bonanni et al., 2023). Beyond chromatography, machine learning is now reaching the other unit operations that make up a purification train. In membrane filtration, Bayesian optimization has been used to design ultrafiltration steps for protein purification with far fewer wet-lab experiments than a full factorial screen would require (Lu et al., 2025), and data-driven models, particularly artificial neural networks and Gaussian Processes, are increasingly utilized to predict flux decline associated with membrane fouling and to enable real-time control of tangential flow filtration through advanced architectures like Model Predictive Control (Oetomo et al., 2025). For crystallization, which offers a lower-cost alternative to preparative chromatography for some products, deep learning trained on large synthetic image sets has enabled non-invasive monitoring of protein crystal growth directly from in-line microscopy (Bischoff et al., 2022). Taken together, these efforts point to a shift from optimizing single steps in isolation toward AI-assisted design and monitoring across the whole downstream workflow, from capture through final polishing.

Digital twin frameworks that couple fermentation kinetics, metabolic modeling, and process control within integrated computational representations offer a particularly promising route toward end-to-end AI-assisted bioprocess development. Machine learning, automation, and large language models can streamline multiscale optimization and digital twin development; however, uncertainty in scale-up performance continues to challenge deployment, requiring shared AI-ready biosystems databases and integrative methods such as transfer learning, reinforcement learning, and Bayesian optimization across the process chain (Yuan et al., 2026). In parallel, AI is transforming industrial enzyme engineering by enabling rapid, data-driven design through models such as AlphaFold2, ProGen, and ESM-2, with hybrid AI-experimental workflows combining predictive modelling with high-throughput screening to accelerate discovery and reduce experimental demand across pharmaceuticals, biofuels, and environmental remediation (Khan & Khan, 2025). Collectively, narrowing the gap between AI-powered strain engineering and AI-enabled industrial deployment will be essential for synthetic biology to realize its commercial and societal potential.

5. Integration of Omics Data and Data Analytics

In recent years, rapid advancements in omics technologies have enabled the integration of data, allowing for the simultaneous measurement of clinical phenotype layers, such as genomic, transcriptomic, proteomic, metabolomic, epigenomic, and even radiomic, spatial omics, and electronic health records, from the same individual or cell, with the aim of holistically integrating biological systems (Nam et al., 2024; Ozcelik et al., 2024). The fundamental challenge in this multi-layered structure lies in the fact that each omics type is generated using different measurement technologies, varying scales, and unique data structures that each have inherent gaps. Conventional methods fall short at capturing complex, often nonlinear relationships among these high-dimensional and heterogeneous data. AI and ML are transforming data analysis by facilitating the extraction of meaningful patterns from high-dimensional datasets and the development of predictive models (McKenzie et al., 2026). A representative case at the single-cell level is Geneformer, a transformer model pre-trained by self-supervi-

sion on about thirty million single-cell transcriptomes. Its attention weights capture patterns of gene-to-gene dependence that broadly resemble a regulatory network, and after downstream fine-tuning the model has been used to suggest therapeutic targets for cardiomyopathy from patient cohorts that are too small for conventional supervised learning (Theodoris et al., 2023). Building upon this foundation, the recently advanced Geneformer v2 expands the pretraining scope to nearly one hundred million single-cell transcriptomes. This significantly upgraded iteration enables specialized downstream fine-tuning for cancer-specific contexts, capturing the complex, aberrant regulatory networks of malignant cells and successfully identifying novel, context-dependent therapeutic targets in oncology where data scarcity previously hindered discovery (H. Chen et al., 2026). A less frequently discussed limitation concerns biological sampling bias in multi-omics datasets. Most publicly available datasets are derived from a limited number of model organisms or populations, leading to skewed representations of biological diversity. This bias can propagate through AI models, reducing their generalizability across species, environments, or underrepresented human populations.

Multi-omics data integration has emerged as one of the most dynamic fields in modern bioinformatics and is fundamentally shaped by three strategic approaches. Alignment-based integration relies on directly aligning data while preserving their original structure, offering researchers a practical starting point yet bringing technical challenges stemming from differences in data scales. Transformation-based integration uses ML and DL techniques to map data into a common mathematical space, successfully uncovering hidden patterns but potentially complicating biological interpretability. Network-based integration models the complex interactions among biological entities, transforming disparate data into a structured and interconnected framework (Nam et al., 2024). Advanced AI models use DNN and representation learning techniques to impute missing omics data and simulate treatment responses in a digital environment (Zack et al., 2025). Table 3 presents an expanded taxonomy of integration approaches.

A critical development that previous reviews of this field have not fully addressed is the rise of federated learning (FL) for multi-omics analysis. The isolated nature of biomedical institutions, governed by patient privacy policies and data-sharing regulations, creates a fundamental tension: the models that would most benefit from large, diverse training sets are the hardest to build when data cannot be centralized. FL resolves this tension by allowing institutions to train models locally and share only model gradients rather than raw data. Privacy-preserving FL frameworks such as privacy-preserving ML method (PPML-Omics) have been demonstrated on RNA sequencing and spatial transcriptomics tasks, achieving near-centralized performance while providing mathematically guaranteed privacy protection (Malpetti et al., 2025). Federated transfer learning with differential privacy has been applied to multi-omics survival analysis in cancer, showing that recruiting additional federated nodes substantially improves model generalization across cancer types (Wen & Li, 2025).

An equally important temporal dimension is emerging through longitudinal omics studies, where repeated biological sampling over time, captured with recurrent neural architectures such as LSTM, which maintain long-range dependencies in sequential data, and transformer-based sequence models, allows dynamic cellular states to be characterized rather than mere static snapshots. This paradigm is particularly relevant for modeling disease progression, drug response trajectories, and microbial community succession. The fusion of spatial transcriptomics with histopathological image analysis, combining gene expression at near-single-cell resolution with tissue morphology, represents a powerful multimodal integration approach that is beginning to transform cancer pathology (McKenzie et al., 2026).

Technical and ethical challenges must be considered during the data integration process. Differing collection protocols across institutions and platforms, batch effects, missing data, and heterogeneous clinical annotations create transparency issues. Therefore, normalization, harmonization, and quality control steps must be performed in a controlled manner to ensure the generalizability of AI models (Zack et al., 2025). Open-source tools and standardized workflows play a critical role in ensuring the reliability of the research process. Ultimately, the primary goal is not merely to store vast amounts of data, but to transform them into a tangible force for improvement that is secure, transparent, and directly benefits the patient (Yetgin, 2025).

6. Promoter Design and Optimization of Gene Expression Using AI

Promoters, the fundamental regulatory elements that determine the location, direction, and strength of gene transcription, occupy a central position in synthetic biology and metabolic engineering (X. Wang et al., 2025). Promoter design and the optimization of gene expression using AI have become key parameters for metabolic engineering, gene therapy, and the design of smart biological circuits. Traditional promoter engineering methods, such as rational design and directed evolution, rely heavily on trial-and-error processes and a limited number of natural sequences. In recent years, the integration of AI and DL techniques has revolutionized the processes of promoter identification, activity prediction, and *de novo* design (Y. Wang et al., 2020; Xia & Huo, 2026). In particular, convolutional neural networks (CNN) and recurrent neural networks (RNN) can predict promoter strength with high accuracy by learning the complex motifs and structural features in promoter sequences (X. Zhou et al., 2026). More recent transformer-based architectures, such as Enformer, extend the effective sequence context of gene-expression predictors to about one hundred kilobases. With this extended context, the model gives more accurate predictions of tissue-specific transcript levels and agrees more closely with high-throughput experimental measurements of how non-coding variants change gene expression (Avsec et al., 2021). Recent studies in *E. coli* have shown

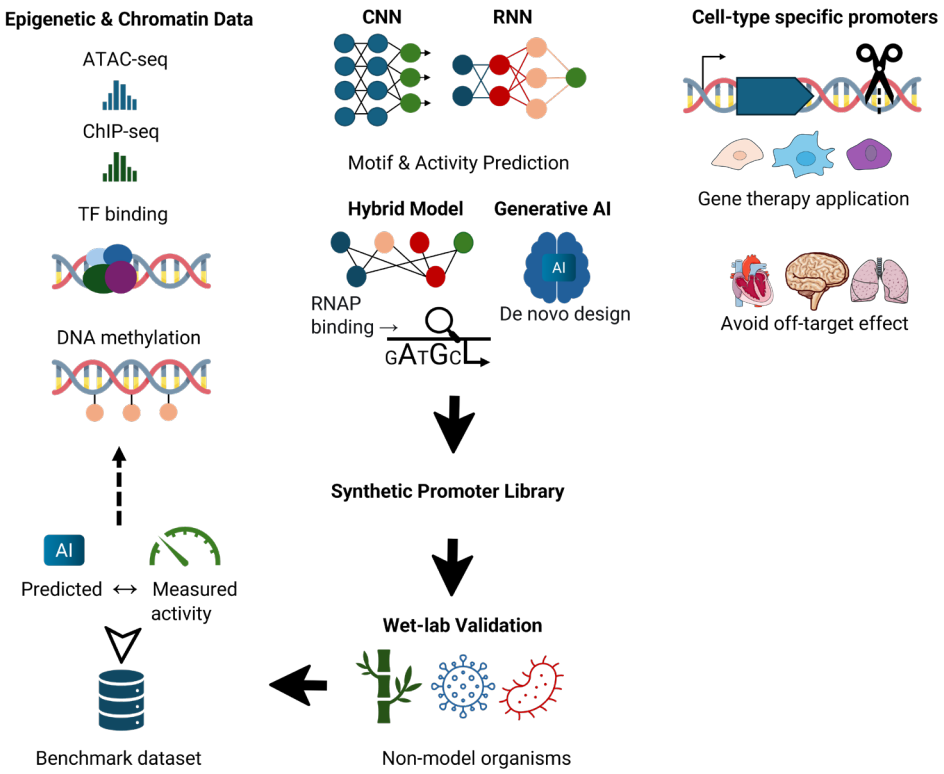

Figure 4. Integration of AI in promoter engineering. The framework integrates multimodal data, including chromatin accessibility, histone modifications, and DNA methylation, into CNN and RNN architectures for motif discovery and activity prediction. Hybrid and generative AI models facilitate the de novo design of synthetic promoter libraries with optimized RNAP binding. The pipeline emphasizes cell-type-specific design for targeted gene therapy to avoid off-target effects, followed by wet-lab validation in non-model organisms to establish robust benchmark datasets. AI: Artificial intelligence, RNAP: RNA polymerase, CNN: Convolutional neural networks, RNN: Recurrent neural networks, ATAC-seq: Assay for transposase-accessible chromatin using sequencing, CHIP-seq: Chromatin immunoprecipitation sequencing. Some of items are used from Smart servier (Image(s) provided by Servier Medical Art (https://smart.servier.com), licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/)).

that AI-guided engineering of promoter −10 and −35 elements enables the rational construction of synthetic promoter libraries with graded expression outputs. Using such frameworks, ML-based models linking RNA polymerase binding energetics to transcriptional activity have achieved expression tuning over a dynamic range of up to 80-fold (X. Zhou et al., 2026).

Generative AI models enable the design of biologically functional and novel promoter sequences from scratch, allowing for the targeting of specific expression levels required for the precise regulation of certain metabolic pathways (Gu et al., 2025; X. Wang et al., 2025). Xu et al. integrated the eGFP reporter gene and the strong sigma70 promoter properties identified via high-throughput screening into DL models, thereby achieving the rational design of novel, high-performance artificial promoters that enable the active and soluble production of industrial-grade enzymes such as microbial transglutaminase in the *E. coli* expression system (K. Xu et al., 2024).

The practical utility of integrating reporter genes with AI-driven promoter design is further demonstrated in environmental and diagnostic biotechnology through bioluminescent bioreporter bacterial sensors. A panel of engineered luminescent bacteria has enabled the production of bioreporters to detect the effect of various toxicities such as genotoxicity, cytotoxicity, membrane toxicity and more. Thanks to their existence whole-cell fiber-optic biosensors have been developed to measure in either water or sediment environmental toxicity in real-time in the field (Carmeli et al., 2025). There are of course limitations, such as 1., measuring mixtures of toxicants within a same sample (Trif et al., 2025); 2., the use of these bioreporters with a single promoter; 3., the need for increased specificity or 4., faster to design; 5., make them more sensitive and 6., easier to interpret in real-world environments. AI could significantly improve them by designing better bioreporter circuits (promoters, regulatory proteins, reporters (like luciferase genes) that 1., predict which genetic designs will produce the strongest or most stable light signal; 2., optimize regulatory networks so bacteria respond only to the target chemical and 3., reduce trial-and-error lab work by simulating thousands of designs *in silico*. ML can analyze experimental data to distinguish weak true signals from random metabolic variation and reduce "background noise" (unwanted light emission). AI could convert raw light data into accurate concentration estimates, correct for temperature, pH, or nutrient effects that distort and detect anomalies or sensor drift over time. This turns

a simple "light/no light" output into better quantitative chemical measurements. Further improvements that will expand what bioreporters can detect are expected by AI that can scan protein and gene databases to identify new bioluminescent systems from unexplored organisms, to suggest novel sensing proteins that respond to specific toxins, or would help identify new bioactive molecules (Trif et al., 2024) and design synthetic regulatory elements that don't exist in nature.

Three directions identified as gaps in current promoter AI research deserve emphasis. First, cell-type-specific promoter design for gene therapy, where precise spatial and cellular expression control is needed to avoid off-target expression in non-target tissues, is an emerging application where AI training datasets incorporating chromatin accessibility (ATAC-seq), histone modification (ChIP-seq), and transcription factor binding data are beginning to enable epigenetic-aware design. Second, epigenetic-aware promoter models that account for chromatin context, DNA methylation, and nucleosome positioning are necessary to bridge the gap between sequence-level approximation and the cellular reality of gene regulation. Third, wet-lab validation of AI-designed promoters in non-model organisms, including industrially relevant bacteria, yeasts, and plant systems, remains a critical bottleneck; systematic benchmark datasets linking estimated and measured promoter activity across diverse hosts will be essential for field progress (Figure 4). Concurrently, these AI-optimized promoter elements and genetic circuits serve as foundational building blocks for the advancement of artificial cells—synthetic constructs designed to mimic the core architecture and functions of biological entities through nanotechnology. By facilitating real-time, biocompatible interactions with living organisms, AI-driven artificial cells are emerging as a potentially groundbreaking field of research and therapy, expanding computational models into functional, controllable biomedical devices (Dundar et al., 2024).

A critical assessment of AI-driven promoter design reveals structural limitations that temper current optimism. Most high-performing models have been trained and validated in model organisms such as *E. coli* and *S. cerevisiae*, and their transferability to industrially or clinically relevant non-model hosts remains largely unvalidated. Furthermore, sequence-level models that ignore chromatin context, DNA methylation state, and nucleosome positioning systematically underrepresent the regulatory complexity of eukaryotic transcription, a gap explicitly identified as a priority for the field. Finally, community-standardized benchmark datasets linking predicted and measured promoter activity across diverse host organisms remain absent, preventing rigorous cross-study comparison of competing design approaches. Closing these gaps will be essential for AI-driven promoter engineering to transition from proof-of-concept demonstrations in model systems to broadly deployable tools for metabolic engineering and gene therapy. As a result, AI-driven data-centric workflows and DNA models optimize gene expression at the transcriptional level, thereby enhancing protein production efficiency in industrial biotechnology and contributing to the development of safer regulators for clinical applications (Liebal et al., 2021; Martin et al., 2021; Xia & Huo, 2026).

7. Clinical and Biomedical Applications of AI-Driven Synthetic Biology

AI-powered synthetic biology holds significant transformation potential in clinical applications. The convergence of these two disciplines offers critical solutions, particularly in disease diagnosis, personalized treatment methods, and novel drug discovery. Traditional methods for regenerative medicine face several obstacles that limit clinical success, including low throughput, high intercellular heterogeneity, and cost (Choudhury et al., 2025). To overcome these challenges, AI-powered synthetic biology has the potential to revolutionize cell therapies by learning and predicting complex and dynamic gene expression patterns during stem cell differentiation (Choudhury et al., 2025). AI accelerates drug discovery by optimizing target identification, predicting protein structures, and performing virtual screening, while enabling personalized medicine through patient-specific simulations and treatment protocol optimization. A meta-analysis covering the period 2015–2025 shows that AI provides strong positive effects on therapeutic efficacy in CRISPR-based epigenetic editing (SMD = 1.67) and guide RNA optimization (SMD = 1.44) (Basarali et al., 2025). DL models increase the potential for safer and more personalized gene editing by minimizing off-target effects. AI plays a significant role in personalized medicine by enabling the development of digital twin frameworks, which are increasingly explored for patient-specific modeling and *in silico* simulation of drug responses (Tudor et al., 2025). On a related axis, AlphaMissense extends AlphaFold2 by fine-tuning it on allele-frequency data from humans and other primates. The trained model assigns a confident benign or pathogenic label to about 89% of every possible single-amino-acid change in the human proteome, and the predictions are released as a precomputed table that clinicians can consult when interpreting variants of uncertain significance (J. Cheng et al., 2023). These systems, alongside dynamic dosing platforms like CURATE.AI, are being explored for optimizing treatment protocols for individual patients (Blasiak et al., 2020). The translational utility of these predictive models is highly pronounced in resolving complex phenotype-genotype correlations that traditional model organisms fail to fully decipher. A prime example of this diagnostic paradigm is ClioMD, an advanced ML platform specifically engineered for heterogeneous genetic disorders like ciliopathies. By integrating large-scale clinical trial data and model organism networks with globally standardized repositories, specialized AI tools like ClioMD accelerate accurate clinical screening, optimize variant interpretation, and deliver automated genetic counseling recommendations (Ergören et al., 2025).

Recent studies enable the development of automated, low-cost, and high-throughput diagnostic systems by integrating synthetic biology with AI and nanotechnology (C. Zhou et al., 2025). AI-supported nanosensors play a transformative role in

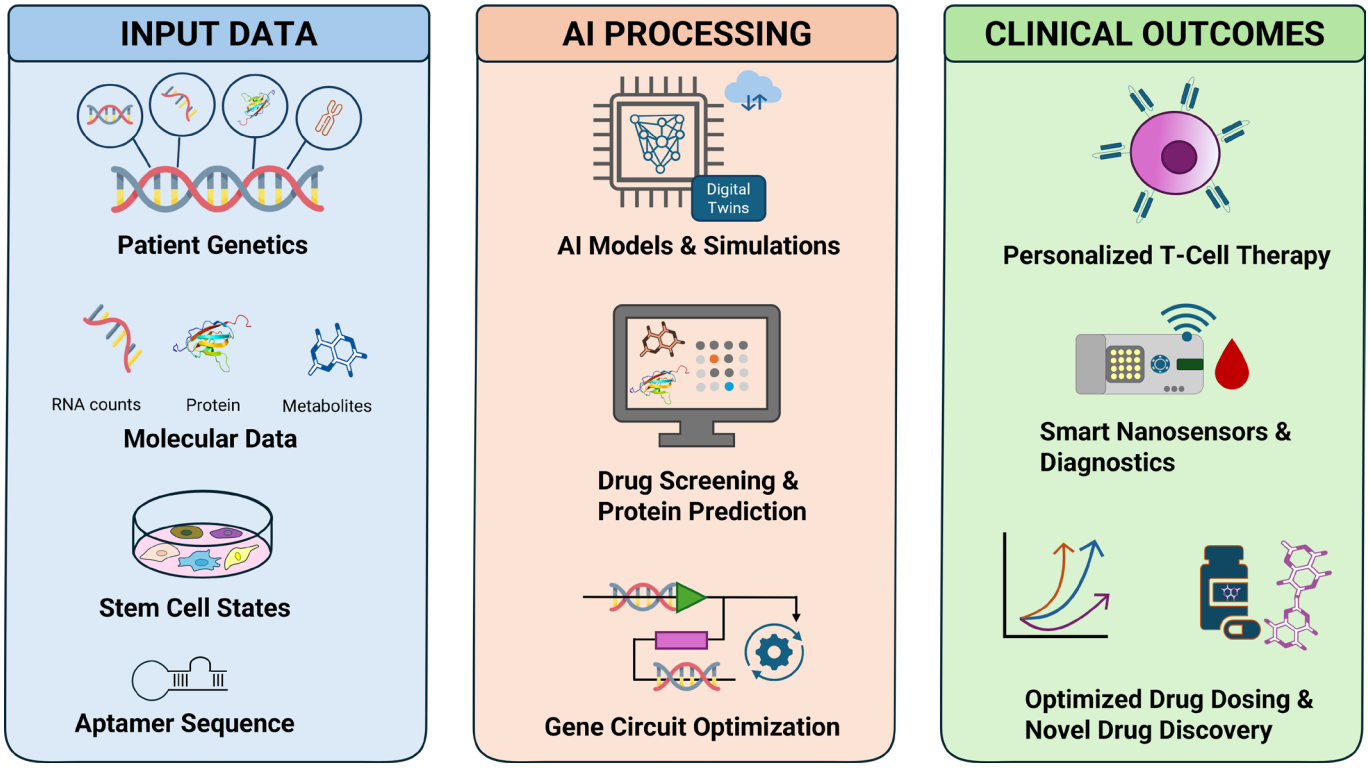

Figure 5. Simplified operational pipeline of AI-integrated synthetic biology in clinical and biomedical applications. The workflow illustrates the systematic transition from Input Data (patient genetics, molecular multi-omics, and stem cell states) through AI Processing (deep learning models, digital twin simulations, and gene circuit optimization) to achieve Clinical Outcomes, such as personalized T-cell therapies, smart wearable diagnostics, and optimized drug dosing platforms. AI: Artificial intelligence.

precision medicine by modernizing biomarker selection, sensor design, performance optimization, and complex data analysis stages in clinical diagnostic processes (Yin, 2025). Engineered biomolecules developed within synthetic biology, such as aptamers, artificial peptides, molecularly imprinted polymers (MIPs), and nucleic acid-based diagnostic elements, form the basis for nanosensors to detect disease-specific biomarkers with high selectivity. In parallel, emerging nanomaterials, including MXenes and novel two-dimensional metallic systems such as goldene, are being increasingly explored to enhance signal transduction, surface chemistry, and analytical sensitivity in AI-driven biosensing platforms (Hussain et al., 2024; Kashiwaya et al., 2024).

Several emerging clinical directions merit forward-looking attention. On-device AI for wearable diagnostics, edge computing models that run inference locally on biosensors without cloud connectivity, address both latency and privacy concerns for real-time health monitoring. The design of regulatory T-cells (Tregs) with ML-optimized synthetic gene circuits represents a promising frontier in autoimmune disease management, extending the ML-guided CAR T-cell engineering paradigm developed by Daniels et al. (2022) to a new therapeutic target class (Daniels et al., 2022). Clinical trial outcome predictions from preclinical synthetic biology data, using AI models trained on cross-species and *in vitro* correlations to forecast translational success, could substantially reduce the high attrition rate of synthetic biology-derived therapeutics moving from bench to bedside (Figure 5). However, the integration of these technologies into clinical applications remains limited due to data security, regulation, and ethical issues (Feizyab et al., 2025; Goktas & Grzybowski, 2025).

However, several translational failures highlight the limitations of current AI-driven approaches. Models trained on preclinical data often fail to generalize to human physiology due to interspecies differences and incomplete biological representations. These discrepancies underscore the need for hybrid validation pipelines combining *in silico*, *in vitro*, and *in vivo* systems.

Such failures are not isolated incidents but reflect structural patterns documented across oncological AI applications more broadly. A comprehensive cross-cancer analysis spanning 18 malignancies found that strong algorithmic performance does not necessarily translate into meaningful patient benefit: most published evidence derives from retrospective, single-institution studies conducted in curated dataset environments that systematically differ from real-world clinical deployment conditions, and diagnostic accuracy has repeatedly failed to serve as a reliable surrogate for patient-centered outcomes such as reduced mortality or procedural harm (Kuchana et al., 2026). Performance disparities across demographic subgroups, disease etiologies, and geographic settings further reveal that current AI systems risk amplifying existing health inequities

rather than resolving them (Kuchana et al., 2026). Beyond algorithmic performance, a deeper structural barrier concerns the absence of an operational definition of trustworthiness in clinical AI: translational gaps markedly arise from conceptual ambiguity around trust, which enables both unintentional misuse and intentional ethics washing by industry stakeholders, leaving "Trustworthy AI" guidelines largely aspirational rather than operationally enforceable (Bürger et al., 2024). Bridging these translational gaps will require prospective trial designs measuring patient-centered endpoints, federated data governance frameworks ensuring demographically inclusive training datasets, and regulatory pathways mandating post-market outcome surveillance as a condition of clinical deployment (Bürger et al., 2024; Kuchana et al., 2026).

7.1 Machine Learning and Synthetic Biology in Parkinson's Disease

Neurodegenerative diseases represent one of the most complex and data-intensive challenges in modern medicine, making them particularly suitable for AI-driven approaches (Young et al., 2024). Among these, Parkinson's disease has emerged as a prominent case study due to its progressive nature, heterogeneous clinical presentation, and the absence of definitive early diagnostic markers (Fan et al., 2025). ML models have demonstrated strong potential in identifying early-stage Parkinson's disease by leveraging multimodal datasets, including neuroimaging, genomic profiles, and increasingly, acoustic biomarkers (Rashidi et al., 2025). Speech impairments, which often manifest in early disease stages, provide a non-invasive and scalable data source. Recent studies employing long-term and short-term acoustic feature extraction combined with ML classification algorithms have achieved promising diagnostic performance, highlighting the feasibility of AI-assisted early screening tools (Serag et al., 2025; Shokrpour et al., 2025).

Importantly, these advances illustrate how data-driven AI models can serve as upstream components within broader synthetic biology pipelines, where early detection signals may inform the design of responsive biological systems. From a synthetic biology perspective, these predictive capabilities open new possibilities for the development of adaptive therapeutic strategies. For instance, engineered cellular systems or gene circuits could, in principle, be designed to detect molecular correlates of neurodegeneration and trigger controlled therapeutic responses. While such integrations remain largely conceptual, they illustrate the broader trajectory toward closed-loop diagnostic–therapeutic systems. Despite these advances, significant challenges remain, including inter-patient variability, limited availability of longitudinal datasets, and the need for clinically validated, interpretable models. Addressing these issues will be critical for translating AI-based diagnostic strategies into practical tools for neurodegenerative disease management.

8. Applications of Artificial Intelligence for Information Storage in Synthetic Biology

DNA data storage represents an emerging interdisciplinary field where synthetic biology and information science converge. With the widespread adoption of digital technologies, global data volumes have experienced explosive growth, necessitating dense and durable alternative media. Due to its ultra-high storage density, long preservation duration, and minimal energy consumption, DNA is regarded as a highly promising future information storage medium (Meiser et al., 2022). The fundamental pipeline involves converting binary data into DNA base sequences (A, T, C, G), which are subsequently synthesized and stored; upon retrieval, high-throughput sequencing technology reads these sequences to decode the original digital data (Meiser et al., 2022). However, due to the inherent biochemical properties of DNA molecules and current technical limitations, these processes are prone to sequence degradation and reading errors, thereby compromising data recovery accuracy. Artificial intelligence, leveraging its robust capabilities in pattern recognition, predictive modeling, and combinatorial optimization, offers novel solutions to these challenges and has become a pivotal force in advancing DNA data storage technologies. During the DNA data encoding phase, factors such as long homopolymers, extreme GC content, and unwanted secondary structures can severely compromise data integrity and DNA stability (Cao et al., 2026; Wu et al., 2023). Consequently, DNA sequence design must strictly adhere to multiple biophysical constraints. The powerful feature extraction capability of CNNs enables them to learn from vast sequence datasets and identify motifs aligned with biological principles. When utilized as intelligent evaluators, these CNN-based models effectively guide the generation and selection of more stable DNA sequences that are easier to synthesize and sequence (Cao et al., 2026; Wu et al., 2023). Similarly, the Graph Convolutional Network integrated with a Self-Attention Mechanism (GCNSA) has been employed to enhance encoding efficiency. Through end-to-end learning, this architecture successfully captures the complex base interactions and spatial constraints of DNA strands to output optimized, constraint-compliant sequences (Cao et al., 2023). Beyond encoding, multiple errors are commonly introduced during DNA synthesis, PCR amplification, and sequencing, with insertion, deletion, and substitution errors being particularly prevalent. Traditional error correction codes—such as Reed-Solomon and Low-Density Parity-Check (LDPC) codes—are limited in efficiency when handling these non-random, asymmetric, and non-Gaussian distributed errors. To address this, researchers have developed a single-stranded representation learning (RSRL) framework. By integrating mathematical error correction coding theory with the biophysical principles of nucleotide interactions, this model significantly reduces the net error rate in raw sequencing data (Cao et al., 2026). Furthermore, bidirectional RNNs (BiRNNs) based on gated recurrent units (GRUs) and LSTM networks have been developed to enable the reliable correction of persistent substitution errors in DNA-encoded data (Srivastava et al., 2024). However, the real-world performance of AI in DNA data storage remains heavily bottlenecked by the availability of high-quality, large-scale, and multimodal experi-

mental datasets. Currently, comprehensive benchmark datasets encompassing error profiles, synthesis biases, and decay metrics across various hardware platforms and storage conditions remain relatively scarce, which limits the generalization and robustness of downstream AI models (Gimpel et al., 2023). Consequently, developing a standardized, open-source, and cross-platform AI benchmark dataset tailored for DNA data storage has emerged as a critical challenge requiring urgent resolution.

9. Ethics, Biosafety, and the Dual-Use Problem in AI Usage

The integration of AI and synthetic biology offers revolutionary innovation potential in medicine, agriculture, and sustainability, but it also brings complex biosafety risks and ethical dilemmas (Groff-Vindman et al., 2025; OECD, 2025). While this technological convergence contributes to the democratization of knowledge by lowering the expertise threshold required for biological engineering, it also increases global security vulnerabilities by enhancing the ability of malicious actors to design dangerous pathogens or toxins (Eskandar, 2026; OECD, 2025). The dual-use risk refers to the fact that models developed for therapeutic purposes can be used for biological weapon design with minimal modifications. Evidence presented by Urbina et al. (2022) showed that a generative AI model designed for drug discovery, when directed to maximize toxicity, generated 40,000 molecules predicted to be highly toxic based on *in silico* scoring and potentially more lethal than known nerve gases in just six hours (Urbina et al., 2022).

A landmark 2025 study published in *Science* demonstrated that AI protein-design tools could rewrite the sequence codes of dangerous proteins, so-called sequence paraphrasing, in ways that preserve their three-dimensional structure and potentially their function, while bypassing current DNA synthesis screening systems that rely on sequence homology to known threats (Wittmann et al., 2025). A cross-sector team, including Microsoft researchers subsequently patched the identified vulnerabilities in commercial screening software, but the underlying challenge remains: Generative models that optimize for structural rather than sequence similarity can produce functional toxins invisible to first-generation screening algorithms. This finding has catalyzed calls for function-based, internationally harmonized biosecurity standards that extend beyond the current paradigm of homology matching (Wittmann et al., 2025).

A complementary technical safeguard gaining momentum is output watermarking, embedding cryptographically traceable signatures into the sequences generated by AI protein design models. Inspired by watermarking frameworks developed for LLM, recent proposals adapt Gumbel sampling methods to bias protein generative models toward statistically identifiable token patterns that remain invisible to biologists and are robust to moderate mutation (Y. Chen et al., 2025). Such watermarks could enable regulatory bodies and synthesis providers to trace AI-generated sequences back to specific model versions or research groups, even for sequences that do not match any entry in existing hazard databases.

On an ethical level, the most fundamental concern is the black-box problem in decision-making processes with AI. This lack of transparency makes it impossible to justify the biological designs proposed by the model, making it difficult to detect erroneous or harmful outcomes (Goktas & Grzybowski, 2025; Groff-Vindman et al., 2025). Furthermore, since the vision of fully autonomous laboratories carries the risk of bypassing human judgment, maintaining the human-in-the-loop principle at critical evaluation points is a legal and moral imperative (OECD, 2025). Another critical issue is data and resource justice; insufficient population representation in training data leads to algorithmic bias, while regional inequalities in access to advanced infrastructure deepen the global technology gap (Eskandar, 2026; OECD, 2025).

Managing these complex risks requires adopting agile and predictive governance models that can adapt to the pace of technology, rather than static rules (Eskandar, 2026; IGSC, 2024). An international consensus is forming around the principles of equity, universality, traceability, availability, robustness, and explainability regarding the use of AI in healthcare through initiatives such as the FUTURE-AI consortium (Lekadir et al., 2025). Therefore, it is vital that high-capacity models undergo red team testing before release, that risk-based incremental access controls are implemented, and that advanced function-oriented screening protocols are aligned with industry standards such as IGCS (Infrastructure Governance and Control Standard) (Eskandar, 2026; IGSC, 2024). Furthermore, as synthetic biology transitions from cellular manipulation to macroscopic tissue-scale architectures, it increasingly intersects with profound bioethical boundaries. This intersection is explicitly demonstrated in the emerging development of artificial placenta (AP) and artificial womb (AW) technologies, which utilize engineered chambers to mimic maternal environments for human embryonic development outside the womb (Medori et al., 2023). While designed as novel medical solutions to mitigate the global infertility crisis and lower preterm infant mortality, they simultaneously provoke urgent questions regarding controllability, legal status, and socio-ethical boundaries that must be resolved prior to public introduction (Medori et al., 2023). Ultimately, equity frameworks specifically designed for low-resource countries, addressing not only access to AI models but also computational infrastructure, training datasets that reflect global biodiversity, and regulatory capacity, represent a foundational justice challenge that the field cannot afford to defer.

10. Methodological and Conceptual Limitations in the Use of AI

The integration of AI deeply influences the macro-level synthesis of research, serving as a powerful enabler for international collaborations and biotechnology data management, while simultaneously introducing distinct systemic challenges. AI is becoming a powerful enabler for the creation of high-quality review papers by international research teams (Abdelrady et

al., 2025; Agho et al., 2025; Bazakos et al., 2026; Ljubešić et al., 2025; Rotter et al., 2020) . It facilitates rapid literature screening across vast, multilingual databases, helping researchers identify relevant studies, detect trends, and reduce bias in evidence synthesis. AI-driven tools can assist in organizing references, summarizing findings, and harmonizing terminology across disciplines, which is particularly valuable in collaborative, cross-country projects (Abdelrady et al., 2025; Agho et al., 2025; Bazakos et al., 2026; Ljubešić et al., 2025; Rotter et al., 2020). Additionally, AI supports consistency in writing style, improves transparency through reproducible workflows, and accelerates the overall drafting process. Importantly, AI also adds significant value in the preparation of manuscripts based on complex environmental datasets linked to biotechnological applications, such as blue biotechnology (Ljubešić et al., 2025; Rotter et al., 2020), agro-biotechnology (Abdelrady et al., 2025; Agho et al., 2025; Bazakos et al., 2026), and medical biotechnology (Bonetti et al., 2023; Feizyab et al., 2025; Kiani et al., 2022; Medori et al., 2023) by enabling advanced data integration, pattern recognition, and predictive modeling. This allows researchers to more effectively link environmental variability with biological responses, optimize resource use, and derive actionable insights for sustainable innovation, ultimately strengthening the scientific rigor and applied relevance of such studies. AI-assisted review papers developed by international research consortia are particularly valuable within the framework of large collaborative projects, where diverse datasets, methodologies, and regional perspectives must be synthesized efficiently. In this context, AI facilitates harmonization of terminology, standardization of data interpretation, and rapid integration of findings across geographically dispersed teams. Importantly, the inclusion of researchers from non-EU countries significantly enhances the scientific robustness and global relevance of such reviews. These contributions bring unique environmental datasets, locally adapted biotechnological practices, and underrepresented ecological contexts—especially critical in fields such as blue biotechnology and agro-biotechnology. Moreover, participation from non-EU partners fosters capacity building, promotes equitable knowledge exchange, and reduces regional bias in scientific narratives. However, disparities in access to AI tools, computational infrastructure, and training may create imbalances in contribution levels, underscoring the need for inclusive frameworks that ensure fair participation and data sovereignty across all partners. Also, the advantages are accompanied by important limitations, including the risk of propagating biases present in training data, potential inaccuracies or oversimplifications in automated summaries, and reduced critical engagement by researchers if over-reliance occurs. Concerns about data privacy, reproducibility of AI-generated outputs, and the need for careful validation of results also highlight the importance of maintaining human oversight and methodological transparency when integrating AI into scientific writing.

These collaborative, macro-level constraints and data-centric bottlenecks directly mirror the micro-level limitations encountered during model development and wet-lab execution. While AI-powered synthetic biology has made significant progress in recent years, many methodological and conceptual limitations still exist in the field. These limitations directly affect both model development processes and the biological validity of the results obtained. One of the most fundamental problems is the uncertainty inherent in biological data. Biological systems are multi-layered structures sensitive to environmental factors, and their highly stochastic behavior limits deterministic modeling approaches and makes the generalizability of AI models difficult (Gangwal et al., 2024). Data quality is another important problem; biological datasets are often incomplete, unbalanced, or contain experimental errors, which can lead to overfitting and model bias (Gholap et al., 2024). Furthermore, most current AI models are described as black boxes, and their inability to explain decision-making mechanisms limits their scientific reliability and clinical applicability (Goktas & Grzybowski, 2025).

Four additional methodological limitations deserve targeted attention. First, the absence of agreed benchmarking standards for biological AI models creates a fragmented literature in which models trained on different datasets, evaluated with different metrics, and applied to different organisms cannot be rigorously compared. Developing community-accepted benchmarks analogous to Critical Assessment of protein Structure Prediction (CASP) for domains such as promoter strength prediction, metabolic pathway optimization, and biosensor design is an urgent priority. Second, uncertainty quantification, the ability of a model to know what it does not know, is underutilized in biological design. Bayesian DL frameworks that express design uncertainty as calibrated probability distributions, rather than point estimates, would allow researchers and regulators to identify when a model is operating outside its domain of reliability. Third, the reproducibility crisis specific to AI-driven biodesign compounds the general reproducibility challenges of experimental biology: different random seeds, software versions, and training set splits can yield substantially different design outputs, undermining comparative assessment across research groups. Fourth, standardized reporting requirements, analogous to CONSORT for clinical trials, are needed for AI model training data, hyperparameter selection, and validation procedures to enable replication and critical appraisal.

Busch et al. (2025) and He et al. (2025) emphasize that LLM used in healthcare should be transparent and explainable (Busch et al., 2025; He et al., 2025). If an AI model suggests a treatment, it is essential for the physician to understand the biological basis of that suggestion (O'Connor et al., 2023). Addressing these limitations is not merely a technical obligation; it is a prerequisite for regulatory acceptance and clinical adoption that would allow the field to fulfill its transformative potential.

While much of the field has focused on model-centric innovation, an emerging paradigm shift emphasizes data-centric AI. In this view, improvements in model performance are achieved not primarily through more complex architectures, but through better data curation, annotation, and standardization.

In synthetic biology, where experimental data are costly and noisy, strategies such as active learning, data augmentation, and automated error correction are becoming increasingly important. Data-centric approaches also highlight the importance of standardized experimental protocols and interoperable data formats. Without these, even the most advanced AI models risk being limited by inconsistencies in training data. The integration of laboratory information management systems (LIMS) with AI pipelines represents a critical step toward creating robust, scalable, and reproducible biological datasets. Ultimately, the shift toward data-centric AI reframes synthetic biology not only as a modeling challenge, but as a data engineering discipline, where the quality, diversity, and structure of data determine the upper bound of innovation.

11. Quantum Computing and Quantum Algorithms in Bioinformatics

Quantum computing represents a fundamentally distinct computational paradigm that exploits quantum mechanical phenomena, including superposition, entanglement, and quantum tunneling to address computationally intractable optimization problems that impose exponential time complexity constraints on classical computing architectures (Marchetti et al., 2022; Pal et al., 2024; Sung et al., 2026). Unlike classical bits that exist in binary states of 0 or 1, quantum bits (qubits) exist in superposition states that enable parallel exploration of exponentially large solution spaces, fundamentally altering the computational complexity landscape for high-dimensional biological optimization problems (Pal et al., 2024; Sung et al., 2026). Quantum entanglement, characterized by non-separable multi-qubit states, provides exponentially scaling computational capacity that classical systems cannot replicate, offering potential advantages for complex biological modeling tasks (Nassir et al., 2025).

Quantum algorithms specifically architected for bioinformatics applications leverage distinct algorithmic primitives that exploit quantum mechanical advantages. Grover's quantum search algorithm provides a quadratic speedup for unstructured database search with direct applications to sequence alignment, motif discovery, and variant calling in genomic datasets (Marchetti et al., 2022; Pal et al., 2024). The Quantum Approximate Optimization Algorithm (QAOA), a variational hybrid quantum-classical framework, has been successfully implemented for protein structure prediction by encoding the conformational search space as optimization problems executable on quantum hardware (Pamidimukkala et al., 2024; Selvaraj et al., 2026). Quantum annealing on adiabatic quantum processors has demonstrated proof-of-concept success in *de novo* genome assembly by mapping sequence reconstruction problems onto quantum-compatible formulations executable on D-Wave quantum annealers (Marchetti et al., 2022; Pal et al., 2024). Recent implementations have achieved protein structure predictions utilizing up to 114 qubits on IBM quantum hardware, successfully capturing the hydrophobic collapse that brings hydrophobic residues to the protein core during folding (Pamidimukkala et al., 2024).

Hybrid quantum-classical workflows represent the most viable implementation strategy on current noisy intermediate-scale quantum (NISQ) devices, which are constrained by limited qubit coherence times, gate fidelities, and connectivity topologies (Sung et al., 2026). Variational quantum eigensolver (VQE) architectures, which partition computational workload between quantum state preparation circuits and classical optimization, have been applied to molecular Hamiltonian simulation for drug discovery, achieving chemical accuracy for small molecules while remaining tractable on NISQ hardware (Gircha et al., 2023; Uttarkar et al., 2026; Uttarkar & Niranjan, 2024). Quantum machine learning (QML) frameworks, including quantum kernel methods and quantum neural networks with parameterized quantum circuits as trainable layers, have demonstrated early proof-of-concept applications in multi-omic data integration, disease classification, and gene regulatory network inference (Matsubara et al., 2025; Nassir et al., 2025; Priyadharshini et al., 2025).

Specific bioinformatics applications where quantum algorithms have demonstrated computational advantages include: protein folding prediction via quantum annealing and gate-based QAOA implementations that navigate exponentially complex conformational landscapes; molecular dynamics simulation acceleration through quantum algorithms for solving time-dependent equations and quantum-enhanced sampling methods; genome sequence assembly and alignment via quantum-accelerated graph traversal algorithms; multi-omic data integration and dimensionality reduction using quantum principal component analysis and quantum support vector machines that exploit quantum kernel advantages for high-dimensional feature spaces characteristic of transcriptomic, proteomic, and metabolomic datasets; and gene regulatory network modeling and causal inference via QML approaches that leverage quantum sampling advantages for probabilistic graphical models (Lappala, 2024; Marchetti et al., 2022; Matsubara et al., 2025; Nassir et al., 2025; Pal et al., 2024; Pamidimukkala et al., 2024; Priyadharshini et al., 2025; Selvaraj et al., 2026).

Despite promising algorithmic developments, practical implementation faces substantial hardware and algorithmic constraints. Qubit decoherence, arising from environmental coupling that induces information loss, fundamentally limits circuit depth and necessitates quantum error correction codes that impose substantial qubit overhead, typically requiring thousands of physical qubits per logical qubit (Marchetti et al., 2022; Sung et al., 2026). Current NISQ devices lack the error-corrected logical qubits required for fault-tolerant quantum computation, constraining practical applications to shallow-depth variational circuits and hybrid algorithms (Carreras et al., 2026; Sung et al., 2026). Algorithm scalability remains challenged by the barren plateau phenomenon in variational quantum algorithms, where optimization gradients vanish exponentially with system size, and by the classical-quantum interface bottleneck in data encoding (Marchetti et al., 2022; Quek et al., 2024). The interdisciplinary expertise gap, requir-

ing simultaneous mastery of quantum information theory, algorithm design, and domain-specific biological knowledge, represents a significant barrier to widespread adoption (Marchetti et al., 2022; Pal et al., 2024).

A rigorous assessment of quantum computing's current role in bioinformatics requires distinguishing between demonstrated computational advantages and speculative future capabilities. At present, demonstrated quantum advantages remain confined to small-scale, narrowly defined problem instances; for genomics specifically, the expected speedup of algorithms such as Grover's search vanishes under realistic assumptions, and quantum advantage is likely achievable only for a specific subset of combinatorial problems that are sufficiently hard for classical methods while requiring relatively limited input variables (Maurizio & Mazzola, 2025). The barren plateau phenomenon in variational quantum algorithms, whereby optimization gradients vanish exponentially with system size, represents a fundamental algorithmic challenge that has not been resolved for biologically relevant problem sizes (Quek et al., 2024). Furthermore, data-loading into quantum states represents an independent bottleneck: for biological datasets with large numbers of real-valued variables, the circuit depth required for data encoding scales with dataset size, often eliminating theoretical speedups before computation begins (Maurizio & Mazzola, 2025). Current NISQ devices remain constrained to shallow-depth circuits insufficient for fault-tolerant execution of quantum algorithms at genomic scale (Carreras et al., 2026; Sung et al., 2026). The practical implication is that the timeline for quantum advantage in bioinformatics remains genuinely uncertain, and progress will depend on co-designing quantum algorithms specifically tailored to biological data structures rather than adapting algorithms originally conceived for other domains (Maurizio & Mazzola, 2025).

As quantum hardware evolves toward fault-tolerant architectures with error-corrected logical qubits, and as quantum algorithm development matures through co-design with biological applications, the integration of quantum computing with AI-driven synthetic biology may unlock transformative capabilities in biomolecular design, system-level simulation, and precision bioengineering that remain intractable for classical computational paradigms (Nassir et al., 2025; Sung et al., 2026).

12. Future Perspectives in Synthetic Biology and AI Integration

The convergence of AI and synthetic biology is approaching an inflection point at which the field begins to redesign not just individual molecules but entire living systems, including the experimental processes used to characterize them. Bibliometric data indicate that significant intellectual resources are being allocated to this domain, particularly in protein design, genomics, and drug discovery (Shi et al., 2023; D. Xu et al., 2022). The next decade will be defined by at least three intersecting technology curves.

First, self-driving laboratories (SDL) are transitioning from proof-of-concept demonstrations to scalable research infrastructure. Autonomous laboratory platforms that couple active learning algorithms with robotic experimentation, high-content imaging, and real-time data analytics can now close the DBTL cycle for defined optimization tasks with minimal human intervention (Tobias & Wahab, 2025); however, most present-day SDLs have been designed primarily for optimization experiments within defined variable spaces, biological applications remain comparatively underrepresented relative to chemistry and materials science, and a fully autonomous Level-5 SDL has not yet been achieved (Tobias & Wahab, 2025). A fully autonomous enzyme engineering platform recently demonstrated 26-fold improvements in industrial *Yersinia mollaretii* phytase activity and 16-fold improvement in *Arabidopsis thaliana* halide methyltransferase activity through AI-guided combinatorial mutagenesis and automated screening (Singh et al., 2025). Plante et al. (2026) present a Total Laboratory Automation framework for synthetic biology in which intelligent Generator, Builder, Analyzer, and Coordinator agents orchestrate the entire life cycle of organism design and construction (Plante et al., 2026). As cloud-based and shared SDL infrastructure becomes more accessible, concerns about the concentration of autonomous experimentation capacity in well-resourced institutions must be addressed through public compute-credit programs and international facility-sharing agreements (Tobias & Wahab, 2025).

Second, LLM are emerging as accelerators of the scientific method in biology, not merely analytical tools but hypothesis generators that can mine decades of literature, synthesize findings across disciplines, and propose testable experimental designs. LLM-based agents such as BioDiscoveryAgent have been shown to boost gene perturbation experiment efficiency by 46% compared to baseline approaches by combining literature retrieval with hypothesis generation and experimental planning (Yang et al., 2025). Dedicated LLM frameworks for synthetic biology, trained on metabolic model databases, standard biological parts registries, and experimental protocol corpora, are beginning to automate DBTL cycle design and decision-making (W. Li et al., 2025). The AI Co-Scientist, a multi-agent system built on LLM, validated a novel hypothesis about bacterial gene transfer mechanisms by autonomously searching literature, critiquing candidate hypotheses, and refining predictions in a generate-debate-evolve loop, a methodology that is potentially applicable to synthetic biology problems where causal mechanisms remain poorly understood (Adam, 2026).

Third, eco-evolutionary AI, designing biological systems for evolvability and long-term ecological performance rather than static function, represents a paradigm shift away from current snapshot optimization. Living systems deployed in real environments evolve; chassis strains engineered to overproduce a metabolite will experience selection pressure to reduce that burden. AI frameworks that model evolutionary trajectories and design genetic circuits robust to selective sweeps, horizontal gene transfer, and environmental fluctuation will be essential for reliable field deployment of synthetic organisms in agriculture, bioremediation, and carbon sequestration appli-

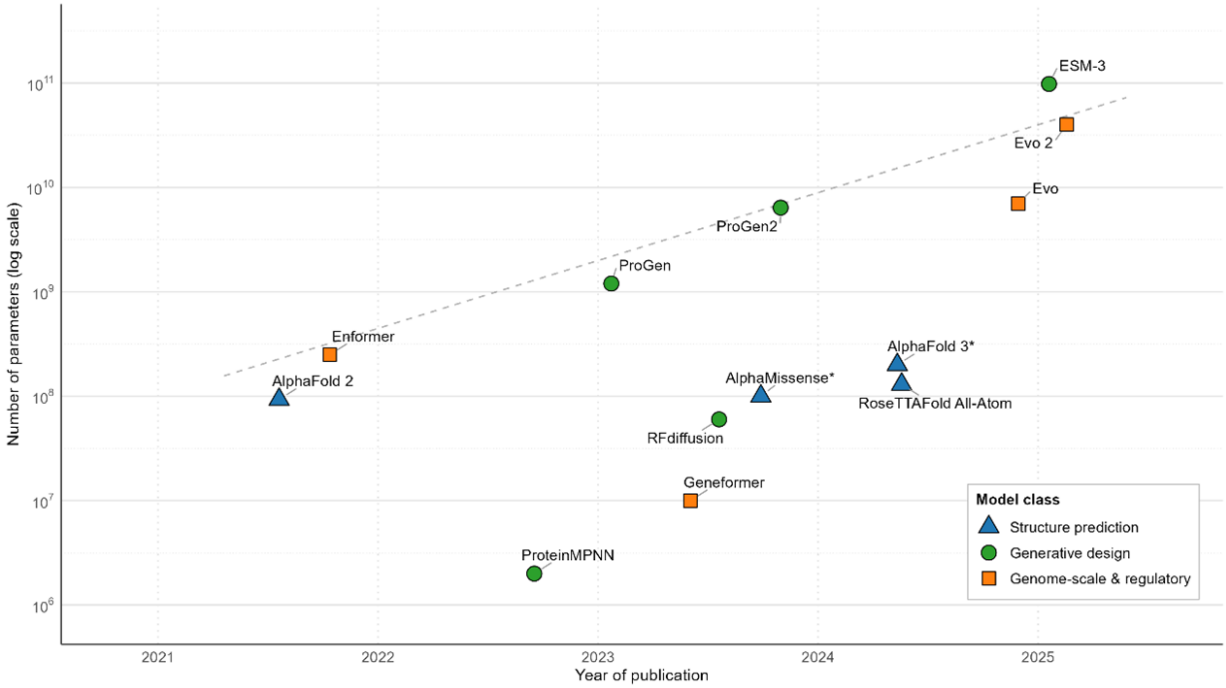

Figure 6. Evolution of AI foundation models for biology, 2021–2025. The figure was generated in R (RStudio) using the ggplot2 and ggrepel packages from a curated data table of the thirteen AI models cited in this review, with marker shape and color encoding the model class and a base-ten logarithmic vertical axis for the parameter count.

cations.

Mirchandani et al. (2025) underscore that future research will concentrate on understanding biological mechanisms via physics-based algorithms and causal reasoning (Mirchandani et al., 2025), a theme that connects the physics-informed modeling renaissance described in Section 2 to the broader vision of mechanistically grounded AI. Industrial scaling and discovery processes will be automated through autonomous laboratories, and ML-driven approaches will minimize the number of experimental iterations in the DBTL cycle (J. Chen et al., 2025). However, this technological progress also brings serious dual-use risks, biosafety, and ethical concerns (Iskuzhina et al., 2025). Unless supported by biosafety and actionable governance frameworks, the digitalization of design processes may create strategic vulnerability against unpredictable unknowns (Eskandar, 2026).

The ultimate frontier of this convergence is the design of bottom-up synthetic cells, marking humanity's transition from a product of evolution into an active designer of artificial life (Aravind Paleri & Hens, 2026; Groff-Vindman et al., 2025). Supported by AI frameworks acting as instruments of augmented cognition (Högberg, 2026), this co-evolution dramatically expands our capacity for knowledge generation (Groff-Vindman et al., 2025; Högberg, 2026). However, this leap introduces severe vulnerabilities (Groff-Vindman et al., 2025; Pannu et al., 2025). Beyond operational constraints, systemic LLM limitations like hallucinations and a lack of true creativity continue to hamper scientific reliability (Groff-Vindman et al., 2025; Peykani et al., 2025). More critically, deploying these systems from methodology design to pathogen biomarker identification amplifies dual-use bioweapons risks, rendering rigorous validation frameworks an absolute biosecurity necessity (Pannu et al., 2025).

Conclusion

AI-powered synthetic biology stands out as one of the fastest-growing and highest-potential fields of modern biotechnology. ML has moved beyond being a promising appendix to synthetic biology; it is now embedded in everyday applications, protein design, genome editing, optimizing metabolic pathways, and interpreting the multi-layered omics data generated by modern experiments. It saves time, reagents, and considerable frustration by reducing a process that once required months of trial and error in the lab to a focused set of candidates selected by an algorithm in many cases.

This review has systematically addressed several rapidly evolving AI application areas: the integration of physics-informed neural networks for biologically constrained modeling; multimodal and continuously learning protein design architectures; explainable AI for metabolic flux estimation and real-time bioprocess control; federated learning for privacy-preserving multi-omics collaboration; longitudinal and spatially resolved omics analytics; epigenetic-aware and cell-type-specific promoter design; on-device and wearable diagnostic AI; AI-enhanced biosecurity through sequence watermarking and function-based screening; and the infrastructure requirements for SDL and LLM-driven hypothesis generation. Addressing these gaps in these areas is not merely an academic exercise; it represents the practical agenda that separates current AI-aug-

mented biology from the self-optimizing, interpretable, and governable discipline that the field aspires to become.

However, alongside all these advances and potentials, significant limitations remain. Training data are often geared towards well-studied organisms and well-characterized pathways, and the most powerful architectures function like black boxes whose internal logic is difficult to oversee. Biological complexity itself presents a fundamental limitation: No model can yet capture the full context dependence of gene expression in a living cell, and designs that perform well *in silico* sometimes fail dramatically *e n*. The ethical dimension is equally challenging. Generative models that can propose novel protein sequences or biosynthetic pathways do not distinguish between beneficial and harmful applications, and governance frameworks designed to manage this dual-use risk, including AI watermarking, function-based DNA screening, and equitable international coordination, are still evolving.

Looking ahead, the trajectory is clear, even if the details are not yet fully defined. Tighter integration of AI with autonomous lab platforms will further shorten the DBTL cycle, and the increasing availability of single-cell and spatial omics data will provide models with richer inputs to learn from. The engineering of organisms for personalized medicine, sustainable biological production, and environmental remediation will benefit as the underlying models become more interpretable, training data more representative, and regulatory environments more consistent. The promise of synthetic biology has always been to make biological design as conscious as engineering in any other field. AI brings us closer to this goal, but the remaining distance requires honest study, rigorous benchmarking, equitable access, and a governance culture that treats biosafety as a design requirement rather than an afterthought.

Ultimately, when returning to the engineering framing with which this review opened, the central question is whether artificial intelligence has delivered on the promise of modularity, abstraction, and standardization that has shaped synthetic biology since its founding. The honest answer is partial. This technological trajectory, characterized by the rapid chronological expansion and scaling of biological foundation models, is systematically synthesized in our architectural mapping (Figure 6). Complementing these views, the Master Integration Table (Table 2) provides the connective structure between scales by mapping representation, design, and execution tools into a single technical framework. At the molecular scale, the combination of all-atom structure prediction, multimodal protein language models, and diffusion-based generative design has effectively industrialized *de novo* protein engineering. At the genome scale, foundation models trained directly on DNA are beginning to do for whole genomes what their predecessors did for protein sequences. At the level of laboratory practice, autonomous-agent architectures have shown, at least in chemistry, that the design-build-test-learn cycle can be closed with little continuous human supervision. What is still missing is the connective tissue between these scales: shared benchmarks that allow competing methods to be compared on equal footing, biosafety-aware generative models that refuse hazardous designs by construction rather than by *post-hoc* filtering, and training corpora that adequately represent non-model organisms and under-explored chemical space. Closing these connective gaps, rather than producing ever larger models, is likely to define the next decade of progress in AI-powered synthetic biology.

Disclaimer

The authors (including Ed Judge) declare that the research, analysis, and opinions presented in this article are entirely their own. The content of this publication does not reflect the views, opinions, or official stance of EPAM Systems, Inc. (EPAM).

Ethical Approval

Not applicable

Funding

The authors received no financial support for the research.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

None.

References

1. Abdelrady, W. A., Mostafa, K., Saeed, M., Elshawy, E. E., Kavas, M., Mladenov, V., Bacu, A., & Zeng, F. (2025). NHX transporters: Molecular mechanisms and applications for enhancing crop resilience to soil salinity in changing environments. Plant Physiology and Biochemistry, 229(Pt C), 110603. https://doi.org/10.1016/J.PLAPHY.2025.110603
2. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C. C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w
3. Adam, D. (2026). The AI co-scientist is here. Nature Medicine, 32(3), 772–775. https://doi.org/10.1038/S41591-026-04275-Z;KWRD
4. Agho, C., Avni, A., Bacu, A., Balazadeh, S., Baloch, F. S., Bazakos, C., Čereković, N., Chaturvedi, P., Chauhan, H., Smet, I. De, Dresselhaus, T., Ferreira, L. J., Fíla, J., Fortes, A. M., Fotopoulos, V., Francesca, S., García-Perez, P., Gong,

W., Graci, S., … Fragkostefanakis, S. (2025). Integrative approaches to enhance reproductive resilience of crops for climate-proof agriculture. Plant Stress, 15, 100704. https://doi.org/10.1016/J.STRESS.2024.100704
5. Ammar, M., Samsonov, M., Gurylina, E., & Bayzigitov, D. (2026). Artificial intelligence advancements in monoclonal antibody development technology. Frontiers in Immunology, 17, 1802038. https://doi.org/10.3389/FIMMU.2026.1802038
6. Anishchenko, I., Pellock, S. J., Chidyausiku, T. M., Ramelot, T. A., Ovchinnikov, S., Hao, J., Bafna, K., Norn, C., Kang, A., Bera, A. K., DiMaio, F., Carter, L., Chow, C. M., Montelione, G. T., & Baker, D. (2021). De novo protein design by deep network hallucination. Nature, 600(7889), 547–552. https://doi.org/10.1038/s41586-021-04184-w
7. Aravind Paleri, V., & Hens, K. (2026). Imagining an ethics for synthetic biology. Frontiers in Genetics, 17, 1746379. https://doi.org/10.3389/fgene.2026.1746379
8. Athanasopoulou, K., Michalopoulou, V. I., Scorilas, A., & Adamopoulos, P. G. (2025). Integrating Artificial Intelligence in Next-Generation Sequencing: Advances, Challenges, and Future Directions. Current Issues in Molecular Biology, 47(6), 470. https://doi.org/10.3390/CIMB47060470
9. Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J. R., Grabska-Barwinska, A., Taylor, K. R., Assael, Y., Jumper, J., Kohli, P., & Kelley, D. R. (2021). Effective gene expression prediction from sequence by integrating long-range interactions. Nature Methods 2021 18:10, 18(10), 1196–1203. https://doi.org/10.1038/s41592-021-01252-x
10. Bangi, M. S. F., Kao, K., & Kwon, J. S. Il. (2022). Physics-informed neural networks for hybrid modeling of lab-scale batch fermentation for β-carotene production using Saccharomyces cerevisiae. Chemical Engineering Research & Design, 179, 415–423. https://doi.org/10.1016/J.CHERD.2022.01.041
11. Basarali, M. K., Daemi, A., Tahiraga, R. G., Özbolat, G., Hooshiar, M. H., Shirazi, M. S. R., & Döğüş, Y. (2025). Artificial intelligence-driven epigenetic CRISPR therapeutics: a structured multi-domain meta-analysis of therapeutic efficacy, off-target prediction, and gRNA optimization. Functional & Integrative Genomics, 25(1), 223. https://doi.org/10.1007/S10142-025-01725-8
12. Bazakos, C., Vidović, M., Radanović, A., Bacu, A., Francesca, S., & Rigano, M. M. (2026). Multi-Level Approaches for Assessing Molecular and Physiological Traits of Drought and Heat Stress Tolerance in Plant Reproductive Development. Physiologia Plantarum, 178(1), e70760. https://doi.org/10.1111/PPL.70760
13. Beardall, W. A. V., Stan, G.-B., & Dunlop, M. J. (2022). Deep Learning Concepts and Applications for Synthetic Biology. GEN Biotechnology, 1(4), 360–371. https://doi.org/10.1089/GENBIO.2022.0017
14. Berger, B., Daniels, N. M., & William Yu, Y. (2016). Computational Biology in the 21st Century: Scaling with Compressive Algorithms. Communications of the ACM, 59(8), 72. https://doi.org/10.1145/2957324
15. Bhardwaj, A., Kishore, S., & Pandey, D. K. (2022). Artificial Intelligence in Biological Sciences. Life (Basel), 12(9), 1430. https://doi.org/10.3390/life12091430
16. Bischoff, D., Walla, B., & Weuster-Botz, D. (2022). Machine learning-based protein crystal detection for monitoring of crystallization processes enabled with large-scale synthetic data sets of photorealistic images. Analytical and Bioanalytical Chemistry, 414(21), 6379–6391. https://doi.org/10.1007/S00216-022-04101-8
17. Blasiak, A., Khong, J., & Kee, T. (2020). CURATE. AI: Optimizing Personalized Medicine with Artificial Intelligence. SLAS Technology, 25(2), 95–105. https://doi.org/10.1177/2472630319890316
18. Bonanni, D., Litrico, M., Ahmed, W., Morerio, P., Cazzorla, T., Spaccapaniccia, E., Cattani, F., Allegretti, M., Beccari, A. R., Del Bue, A., & Martin, F. (2023). A Deep Learning Approach to Optimize Recombinant Protein Production in Escherichia coli Fermentations. Fermentation, 9(6), 503. https://doi.org/10.3390/FERMENTATION9060503
19. Bonetti, G., Donato, K., Medori, M. C., Dhuli, K., Henehan, G., Brown, R., Sieving, P., Sykora, P., Marks, R., Falsini, B., Capodicasa, N., Miertus, S., Lorusso, L., Dondossola, D., Tartaglia, G. M., Cerkez Ergoren, M., Dundar, M., Michelini, S., Malacarne, D., … Bertelli, M. (2023). Human Cloning: Biology, Ethics, and Social Implications. La Clinica Terapeutica, 174(Suppl 2(6)), 230–235. https://doi.org/10.7417/CT.2023.2492
20. Brixi, G., Durrant, M. G., Ku, J., Naghipourfar, M., Poli, M., Sun, G., Brockman, G., Chang, D., Fanton, A., Gonzalez, G. A., King, S. H., Li, D. B., Merchant, A. T., Nguyen, E., Ricci-Tam, C., Romero, D. W., Schmok, J. C., Taghibakhshi, A., Vorontsov, A., … Hie, B. L. (2026). Genome modelling and design across all domains of life with Evo 2. Nature 2026 652:8112, 652(8112), 1349–1361. https://doi.org/10.1038/s41586-026-10176-5
21. Bürger, V. K., Amann, J., Bui, C. K. T., Fehr, J., & Madai, V. I. (2024). The unmet promise of trustworthy AI in healthcare: why we fail at clinical translation. Frontiers in Digital Health, 6, 1279629. https://doi.org/10.3389/FDGTH.2024.1279629
22. Busch, F., Hoffmann, L., Rueger, C., van Dijk, E. H. C., Kader, R., Ortiz-Prado, E., Makowski, M. R., Saba, L., Hadamitzky, M., Kather, J. N., Truhn, D., Cuocolo, R., Adams, L. C., & Bressem, K. K. (2025). Current applications and challenges in large language models for patient care: a systematic review. Communications Medicine, 5(1), 26. https://doi.org/10.1038/s43856-024-00717-2
23. Callaway, E. (2025, January 23). AI-designed proteins tackle century-old problem - making snake antivenoms. Nature, 637(8047), 776. https://doi.org/10.1038/D41586-025-00133-Z
24. Cao, B., Li, X., He, T., Wang, B., Zhou, S., Wu, X., & Zhang, Q. (2026). Learning Structurally Stabilized Representations

for Lossless DNA Storage. Proceedings of the AAAI Conference on Artificial Intelligence, 40(1), 39–47. https://doi.org/10.1609/AAAI.V40I1.36962
25. Cao, B., Wang, B., & Zhang, Q. (2023). GCNSA: DNA storage encoding with a graph convolutional network and self-attention. IScience, 26(3), 106231. https://doi.org/10.1016/J.ISCI.2023.106231
26. Cardiff, R. A. L., Carothers, J. M., Zalatan, J. G., & Sauro, H. M. (2024). Systems-Level Modeling for CRISPR-Based Metabolic Engineering. ACS Synthetic Biology, 13(9), 2643–2652. https://doi.org/10.1021/acssynbio.4c00053
27. Carmeli, G., Paul, A. A., Kristollari, K., Eltzov, E., Batushansky, A., & Marks, R. S. (2025). Whole-Cell Fiber-Optic Biosensor for Real-Time, On-Site Sediment and Water Toxicity Assessment: Applications at Contaminated Sites Across Israel. Biosensors, 15(7), 404. https://doi.org/10.3390/BIOS15070404
28. Carreras, A., Orús, R., & Casanova, D. (2026). Limitations of quantum hardware for molecular energy estimation using VQE. Physical Chemistry Chemical Physics, 28, 2834–2846. https://doi.org/10.1039/d5cp03907j
29. Carruthers, D. N., Kinnunen, P. C., Li, Y., Chen, Y., Gin, J. W., Yunus, I. S., Galliard, W. R., Tan, S., Radivojevic, T., Adams, P. D., Singh, A. K., Sustarich, J., Petzold, C. J., Mukhopadhyay, A., Garcia Martin, H., & Lee, T. S. (2025). Automation and machine learning drive rapid optimization of isoprenol production in Pseudomonas putida. Nature Communications, 16(1), 11489. https://doi.org/10.1038/s41467-025-66304-8
30. Celebi, D., Akalin, H., Yilmaz, M. T., & Dundar, M. (2023). Impacts of Biotechnologically Developed Microorganisms on Ecosystems. The EuroBiotech Journal, 7(4), 196–205. https://doi.org/10.2478/EBTJ-2023-0015
31. Chen, C., & Zheng, P. (2023). Effects of down-regulation of ackA expression by CRISPR-dCpf1 on succinic acid production in Actinobacillus succinogenes. AMB Express, 13(1), 12. https://doi.org/10.1186/S13568-023-01518-X
32. Chen, H., Venkatesh, M. S., Gómez Ortega, J., Mahesh, S. V., Nandi, T. N., Madduri, R. K., Pelka, K., & Theodoris, C. V. (2026). Scaling and quantization of large-scale foundation model enables resource-efficient predictions in network biology. Nature Computational Science, 6(5), 450–463. https://doi.org/10.1038/s43588-026-00972-4
33. 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
34. Chen, Y., Hu, Z., Wu, Y., Chen, R., Jin, Y., Zhan, M., Xie, C., Chen, W., & Huang, H. (2025). Enhancing privacy in biosecurity with watermarked protein design. Bioinformatics, 41(7), btaf141. https://doi.org/10.1093/BIOINFORMATICS/BTAF141
35. Cheng, A. A., & Lu, T. K. (2012). Synthetic biology: An emerging engineering discipline. Annual Review of Biomedical Engineering, 14, 155–178. https://doi.org/10.1146/annurev-bioeng-071811-150118
36. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulyte, A., Applebaum, T., Pritzel, A., Wong, L. H., Zielinski, M., Sargeant, T., Schneider, R. G., Senior, A. W., Jumper, J., Hassabis, D., Kohli, P., & Avsec, Ž. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science, 381(6664), eadg7492. https://doi.org/10.1126/science.adg7492
37. Cheng, Y., Bi, X., Xu, Y., Liu, Y., Li, J., Du, G., Lv, X., & Liu, L. (2023). Machine learning for metabolic pathway optimization: A review. Computational and Structural Biotechnology Journal, 21, 2381–2393. https://doi.org/10.1016/j.csbj.2023.03.045
38. Cho, S., Shin, J., & Cho, B. K. (2018). Applications of CRISPR/Cas system to bacterial metabolic engineering. International Journal of Molecular Sciences, 19(4), 1089. https://doi.org/10.3390/ijms19041089
39. Choudhury, M., Deans, A. J., Candland, D. R., & Deans, T. L. (2025). Advancing cell therapies with artificial intelligence and synthetic biology. Current Opinion in Biomedical Engineering, 34, 100580. https://doi.org/10.1016/J.COBME.2025.100580
40. Daniels, K. G., Wang, S., Simic, M. S., Bhargava, H. K., Capponi, S., Tonai, Y., Yu, W., Bianco, S., & Lim, W. A. (2022). Decoding CAR T cell phenotype using combinatorial signaling motif libraries and machine learning. Science (New York, N.Y.), 378(6625), 1194–1200. https://doi.org/10.1126/SCIENCE.ABQ0225
41. Das, P. K., Sahoo, A., & Veeranki, V. D. (2025). Modeling and Optimization of Recombinant Tocilizumab Production From Pichia pastoris Using Response Surface Methodology and Artificial Neural Network. Biotechnology and Bioengineering, 122(8), 2093–2110. https://doi.org/10.1002/BIT.29024
42. Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Courbet, A., de Haas, R. J., Bethel, N., Leung, P. J. Y., Huddy, T. F., Pellock, S., Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., … Baker, D. (2022). Robust deep learning–based protein sequence design using ProteinMPNN. Science, 378(6615), 49–56. https://doi.org/10.1126/SCIENCE.ADD2187
43. Ding, K., Chin, M., Zhao, Y., Huang, W., Mai, B. K., Wang, H., Liu, P., Yang, Y., & Luo, Y. (2024). Machine learning-guided co-optimization of fitness and diversity facilitates combinatorial library design in enzyme engineering. Nature Communications, 15(1), 6392. https://doi.org/10.1038/s41467-024-50698-y
44. DiNuzzo, M. (2022). How artificial intelligence enables modeling and simulation of biological networks to accelerate drug discovery. Frontiers in Drug Discovery, 2, 1019706. https://doi.org/10.3389/FDDSV.2022.1019706/FULL
45. Donato, K., Medori, M. C., Stuppia, L., Beccari, T.,

Dundar, M., Marks, R. S., Michelini, S., Borghetti, E., Zuccato, C., Seppilli, L., Elsangak, H., Sozanski, G., Malacarne, D., & Bertelli, M. (2023). Unleashing the potential of biotechnology for sustainable development. European Review for Medical and Pharmacological Sciences, 27(6 Suppl), 100–113. https://doi.org/10.26355/EURREV_202312_34694

46. Dudeja, C., Mishra, A., Ali, A., Singh, P. P., & Jaiswal, A. K. (2025). Microbial Genome Editing with CRISPR–Cas9: Recent Advances and Emerging Applications Across Sectors. Fermentation, 11(7), 410. https://doi.org/10.3390/FERMENTATION11070410
47. Dundar, M. S., Yildirim, A. B., Yildirim, D. T., Akalin, H., & Dundar, M. (2024). Artificial cells: A potentially groundbreaking field of research and therapy. The EuroBiotech Journal, 8(1), 55–64. https://doi.org/10.2478/EBTJ-2024-0006
48. Ergören, M. Ç., Senturk, N., Ali, M. S. B., Özcelik, I. Ö., Erol, K. D., Temel, S. G., & Dundar, M. (2025). ClioMD: An artificial intelligence model for ciliopathies. The EuroBiotech Journal, 9(2), 128–137. https://doi.org/10.2478/EBTJ-2025-0011
49. Eskandar, K. (2026). Artificial intelligence and synthetic biology: biosecurity risks, dual-use concerns, and governance pathways. AI and Ethics, 6, 66. https://doi.org/10.1007/S43681-025-00872-9
50. Fan, S. P., Okuzumi, A., Chen, P. S., Tai, Y. C., Kuo, Y. C., Tai, C. H., Li, C. H., Chiang, H. L., Hatano, T., & Lin, C. H. (2025). Parkinson's disease in transition: Genetics, biomarkers, and emerging therapeutics. Journal of the Formosan Medical Association, S0929-6646(25), 00649–7. https://doi.org/10.1016/J.JFMA.2025.12.005
51. Farea, A., Yli-Harja, O., & Emmert-Streib, F. (2025). Using Physics-Informed Neural Networks for Modeling Biological and Epidemiological Dynamical Systems. Mathematics, 13(10), 1664. https://doi.org/10.3390/MATH13101664
52. Feizyab, S., Bonetti, G., Medori, M. C., Micheletti, C., Luca, I. De, Donato, K., Miertuš, J., Dundar, M. S., Vráblová, M., Henehan, G., Brown, R., Marks, R., Miertus, S., Lorusso, L., Tartaglia, G. M., Dundar, M., Michelini, S., Connelly, S. T., Bacu, A., … Bertelli, M. (2025). Artificial Intelligence and Humanoid Robotics: Bioethical Implications of Replacing Human Agency in Healthcare and beyond. The EuroBiotech Journal, 9(3), 238–246. https://doi.org/10.2478/EBTJ-2025-0019
53. Ferruz, N., Schmidt, S., & Höcker, B. (2022). ProtGPT2 is a deep unsupervised language model for protein design. Nature Communications, 13(1), 4348. https://doi.org/10.1038/s41467-022-32007-7
54. Fukala, I., & Kučera, I. (2024). Natural Polyhydroxyalkanoates—An Overview of Bacterial Production Methods. Molecules, 29(10), 2293. https://doi.org/10.3390/MOLECULES29102293
55. Galeazzi, A., Sachio, S., Edwards, E., Hilton, D., & Papathanasiou, M. M. (2025). Machine learning enhanced process design in protein a chromatography. Journal of Chromatography A, 1759, 466193. https://doi.org/10.1016/j.chroma.2025.466193
56. Gangwal, A., Ansari, A., Ahmad, I., Azad, A. K., & Wan Sulaiman, W. M. A. (2024). Current strategies to address data scarcity in artificial intelligence-based drug discovery: A comprehensive review. Computers in Biology and Medicine, 179, 108734. https://doi.org/10.1016/J.COMPBIOMED.2024.108734
57. Gazis, T. A., Wuyts, J., Moutsiou, A., Volpin, G., Ford, M. J., Teixeira, R. I., Wheelhouse, K. M. P., Natho, P., Žnidaršič-Plazl, P., Jost, S., Luisi, R., Benyahia, B., Maes, B. U. W., & Vilé, G. (2026). Towards greener-by-design fine chemicals. Part 1: synthetic frontiers. Chemical Society Reviews, 55(2), 619–674. https://doi.org/10.1039/D5CS00929D
58. Gholap, A. D., Uddin, M. J., Faiyazuddin, M., Omri, A., Gowri, S., & Khalid, M. (2024). Advances in artificial intelligence for drug delivery and development: A comprehensive review. Computers in Biology and Medicine, 178, 108702. https://doi.org/10.1016/J.COMPBIOMED.2024.108702
59. Gil, J. D., Del Rio Chanona, E. A., Guzmán, J. L., & Berenguel, M. (2026). Reinforcement learning meets bioprocess control through behavior cloning: Real-world deployment in an industrial photobioreactor. Engineering Applications of Artificial Intelligence, 164, 113326. https://doi.org/10.1016/J.ENGAPPAI.2025.113326
60. Gimpel, A. L., Stark, W. J., Heckel, R., & Grass, R. N. (2023). A digital twin for DNA data storage based on comprehensive quantification of errors and biases. Nature Communications, 14(1), 6026. https://doi.org/10.1038/s41467-023-41729-1
61. Gircha, A. I., Boev, A. S., Avchaciov, K., Fedichev, P. O., & Fedorov, A. K. (2023). Hybrid quantum-classical machine learning for generative chemistry and drug design. Scientific Reports, 13(1), 8250. https://doi.org/10.1038/s41598-023-32703-4
62. Goktas, P., & Grzybowski, A. (2025). Shaping the Future of Healthcare: Ethical Clinical Challenges and Pathways to Trustworthy AI. Journal of Clinical Medicine, 14(5), 1605. https://doi.org/10.3390/JCM14051605
63. Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. Nature Reviews. Molecular Cell Biology, 23(1), 40–55. https://doi.org/10.1038/S41580-021-00407-0
64. Groff-Vindman, C. S., Trump, B. D., Cummings, C. L., Smith, M., Titus, A. J., Oye, K., Prado, V., Turmus, E., & Linkov, I. (2025). The convergence of AI and synthetic biology: the looming deluge. Npj Biomedical Innovations, 2(1), 20. https://doi.org/10.1038/s44385-025-00021-1
65. Gu, Y., Su, J., Xia, J., Wu, P., Wu, H., Su, Y., Wei, P. J., & Zheng, C. H. (2025). De novo promoter design method based on deep generative and dynamic evolution algorithm. Nucleic Acids Research, 53(16), gkaf833. https://doi.org/10.1093/

nar/gkaf833

66. Guarda, E. C., Galinha, C. F., Pasculli, G., Duque, A. F., & Reis, M. A. M. (2025). Monitoring Polyhydroxyalkanoates (PHA) Production by Mixed Microbial Cultures Using 2D Fluorescence Spectroscopy: Impact of Operating Conditions. Natural Sciences, 5(4), e70030. https://doi.org/10.1002/ntls.70030
67. Hayes, T., Rao, R., Akin, H., Sofroniew, N. J., Oktay, D., Lin, Z., Verkuil, R., Tran, V. Q., Deaton, J., Wiggert, M., Badkundri, R., Shafkat, I., Gong, J., Derry, A., Molina, R. S., Thomas, N., Khan, Y. A., Mishra, C., Kim, C., … Rives, A. (2025). Simulating 500 million years of evolution with a language model. Science, 387(6736), 850–858. https://doi.org/10.1126/science.ads0018
68. He, Y., Huang, F., Jiang, X., Nie, Y., Wang, M., Wang, J., & Chen, H. (2025). Foundation Model for Advancing Healthcare: Challenges, Opportunities and Future Directions. IEEE Reviews in Biomedical Engineering, 18, 172–191. https://doi.org/10.1109/RBME.2024.3496744
69. Hein, Z. M., Guruparan, D., Okunsai, B., Che Mohd Nassir, C. M. N., Ramli, M. D. C., & Kumar, S. (2025). AI and Machine Learning in Biology: From Genes to Proteins. Biology, 14(10), 1453. https://doi.org/10.3390/BIOLOGY14101453
70. Högberg, A. (2026). Becoming human in the age of AI: cognitive co-evolutionary processes. Frontiers in Psychology, 16, 1734048. https://doi.org/10.3389/fpsyg.2025.1734048
71. Hossain, I., Fanfani, V., Fischer, J., Quackenbush, J., & Burkholz, R. (2024). Biologically informed NeuralODEs for genome-wide regulatory dynamics. Genome Biology, 25(1), 127. https://doi.org/10.1186/S13059-024-03264-0
72. Hussain, T., Chandio, I., Ali, A., Hyder, A., Memon, A. A., Yang, J., & Thebo, K. H. (2024). Recent developments of artificial intelligence in MXene-based devices: from synthesis to applications. Nanoscale, 16(38), 17723–17760. https://doi.org/10.1039/D4NR03050H
73. IGSC. (2024). IGSC-Harmonized-Screening-Protocol-v3.0-1.
74. Ing, A., Andrades, A., Cosenza, M. R., & Korbel, J. O. (2025). Integrating multimodal cancer data using deep latent variable path modelling. Nature Machine Intelligence, 7(7), 1053–1075. https://doi.org/10.1038/s42256-025-01052-4
75. Iskuzhina, L., Turaev, Z., Rozhin, A., Romanov, A., Skomorokhova, E., Ishmukhametov, I., & Rozhina, E. (2025). Artificial intelligence in biology and medicine. The Science of Nature, 112(6), 80. https://doi.org/10.1007/S00114-025-02029-4
76. Jeong, K. J. (2025). Advancing microbial engineering through synthetic biology. Journal of Microbiology (Seoul, Korea), 63(3), e2503100. https://doi.org/10.71150/jm.2503100
77. Jeong, S. H., Lee, H. J., & Lee, S. J. (2023). Recent Advances in CRISPR-Cas Technologies for Synthetic Biology. Journal of Microbiology (Seoul, Korea), 61(1), 13. https://doi.org/10.1007/S12275-022-00005-5
78. Ji, K., Yu, X., Chen, L., Wang, Y., Guo, Z., Chen, B., Li, Q., Li, Z., Zhang, H., Wang, G., Zhuang, Y., & Ruan, Y. (2025). Data-Augmented Deep Learning Algorithm for Accurate Control of Bioethanol Fermentation Using an Online Raman Analyzer. Biotechnology and Bioengineering, 122(9), 2366–2376. https://doi.org/10.1002/BIT.29040
79. Ji, X., Li, Y., Wang, J., Wang, G., Ma, B., Shi, J., Cui, C., & Wang, R. (2025). Silk Protein Gene Engineering and Its Applications: Recent Advances in Biomedicine Driven by Molecular Biotechnology. Drug Design, Development and Therapy, 19, 599–626. https://doi.org/10.2147/DDDT.S504783
80. Jiang, J., Li, Y., Cao, S., Shan, Y., Liu, Y., Fei, T., Yu, Y., Feng, Y., Li, Y., Li, Y., & Yuan, J. (2025). Artificial intelligence in bioinformatics: a survey. Briefings in Bioinformatics, 26(6), bbaf576. https://doi.org/10.1093/BIB/BBAF576
81. Jin, S., Wu, Q., Fu, G., Lu, D., Wang, F., Deng, L., & Nie, K. (2025). Breaking Evolution's Ceiling: AI-Powered Protein Engineering. Catalysts, 15(9), 842. https://doi.org/10.3390/catal15090842
82. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2
83. Kacena, M. A., Plotkin, L. I., & Fehrenbacher, J. C. (2024). The Use of Artificial Intelligence in Writing Scientific Review Articles. Current Osteoporosis Reports, 22(1), 115–121. https://doi.org/10.1007/S11914-023-00852-0
84. Kalia, V. C., Singh, R. V., Gong, C., & Lee, J. K. (2025). Toward Sustainable Polyhydroxyalkanoates: A Next-Gen Biotechnology Approach. Polymers, 17(7), 853. https://doi.org/10.3390/POLYM17070853
85. Karimi Alavijeh, M., Lee, Y. Y., & Gras, S. L. (2024). A perspective-driven and technical evaluation of machine learning in bioreactor scale-up: A case-study for potential model developments. Engineering in Life Sciences, 24(7), e2400023. https://doi.org/10.1002/ELSC.202400023
86. Kashiwaya, S., Shi, Y., Lu, J., Sangiovanni, D. G., Greczynski, G., Magnuson, M., Andersson, M., Rosen, J., & Hultman, L. (2024). Synthesis of goldene comprising single-atom layer gold. Nature Synthesis, 3(6), 744–751. https://doi.org/10.1038/s44160-024-00518-4
87. Khalil, A. S., & Collins, J. J. (2010). Synthetic biology: applications come of age. Nature Reviews Genetics, 11(5), 367–379. https://doi.org/10.1038/nrg2775
88. Khan, M. F., & Khan, M. T. (2025). AI-Driven Enzyme Engineering: Emerging Models and Next-Generation Biotechnological Applications. Molecules, 31(1), 45. https://doi.org/10.3390/MOLECULES31010045

89. Kiani, A. K., Pheby, D., Henehan, G., Brown, R., Sieving, P., Sykora, P., Marks, R., Falsini, B., Capodicasa, N., Miertus, S., Lorusso, L., Dondossola, D., Tartaglia, G. M., Ergoren, M. C., Dundar, M., Michelini, S., Malacarne, D., Bonetti, G., Dautaj, A., … Bertelli, M. (2022). Ethical considerations regarding animal experimentation. Journal of Preventive Medicine and Hygiene, 63(2S3), E255–E255. https://doi.org/10.15167/2421-4248/JPMH2022.63.2S3.2768
90. Kim, G. B., Choi, S. Y., Cho, I. J., Ahn, D. H., & Lee, S. Y. (2023). Metabolic engineering for sustainability and health. Trends in Biotechnology, 41(3), 425–451. https://doi.org/10.1016/j.tibtech.2022.12.014
91. Kim, M. G., Go, M. J., Kang, S. H., Jeong, S. H., & Lim, K. (2025). Revolutionizing CRISPR technology with artificial intelligence. Experimental & Molecular Medicine, 57(7), 1419–1431. https://doi.org/10.1038/s12276-025-01462-9
92. Korbeld, K. T., Viliuga, V., & Fürst, M. J. L. J. (2026). Limitations of the refolding pipeline for de novo protein design. Protein Science, 35(6), e70613. https://doi.org/10.1002/PRO.70613
93. Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G. R., Morey-Burrows, F. S., Anishchenko, I., Humphreys, I. R., McHugh, R., Vafeados, D., Li, X., Sutherland, G. A., Hitchcock, A., Neil Hunter, C., Kang, A., Brackenbrough, E., Bera, A. K., … Baker, D. (2024). Generalized biomolecular modeling and design with RoseTTAFold All-Atom. Science, 384(6693), eadl2528. https://doi.org/10.1126/science.adl2528
94. Kuchana, S. K., Repalle, U. K., Alahari, N. V., Kondamuri, M., Manduva, S. K., Vanguru, R. V., Gorle, S. A., & Alahari, S. K. (2026). Artificial Intelligence in Oncology: A Comprehensive Cross-Cancer Translational Readiness Analysis Across 18 Malignancies. Cancers, 18(10), 1543. https://doi.org/10.3390/CANCERS18101543
95. Landwehr, G. M., Bogart, J. W., Magalhaes, C., Hammarlund, E. G., Karim, A. S., & Jewett, M. C. (2025). Accelerated enzyme engineering by machine-learning guided cell-free expression. Nature Communications, 16(1), 865. https://doi.org/10.1038/s41467-024-55399-0
96. Lappala, A. (2024). The next revolution in computational simulations: Harnessing AI and quantum computing in molecular dynamics. Current Opinion in Structural Biology, 89, 102919. https://doi.org/10.1016/j.sbi.2024.102919
97. Lawson, C. E., Martí, J. M., Radivojevic, T., Jonnalagadda, S. V. R., Gentz, R., Hillson, N. J., Peisert, S., Kim, J., Simmons, B. A., Petzold, C. J., Singer, S. W., Mukhopadhyay, A., Tanjore, D., Dunn, J. G., & Garcia Martin, H. (2021). Machine learning for metabolic engineering: A review. Metabolic Engineering, 63, 34–60. https://doi.org/10.1016/j.ymben.2020.10.005
98. Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S., & Lee, S. Y. (2012). Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nature Chemical Biology, 8(6), 536–546. https://doi.org/10.1038/nchembio.970
99. Lekadir, K., Frangi, A. F., Porras, A. R., Glocker, B., Cintas, C., Langlotz, C. P., Weicken, E., Asselbergs, F. W., Prior, F., Collins, G. S., Kaissis, G., Tsakou, G., Buvat, I., Kalpathy-Cramer, J., Mongan, J., Schnabel, J. A., Kushibar, K., Riklund, K., Marias, K., … Starmans, M. P. A. (2025). FUTURE-AI: International consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ, 388, e081554. https://doi.org/10.1136/BMJ-2024-081554
100. Lemperle, M., Ramin, P., Kager, J., Cassells, B., Stocks, S., & Gernaey, K. V. (2026). Hybrid modeling for industrial fermentation processes with an "Intra-Batch Experimental Design." Journal of Industrial Microbiology & Biotechnology, 53, kuag014. https://doi.org/10.1093/JIMB/KUAG014
101. Li, H., Wu, Y., Qin, D., Xie, J., Huang, W., Chen, R., Liu, H., Wang, Y., & Zhao, D. (2025). Phase-driven rewiring in Escherichia coli enhances coenzyme Q10 biosynthesis via temporal and energetic coordination. Applied Microbiology and Biotechnology, 109(1), 248. https://doi.org/10.1007/S00253-025-13619-7
102. Li, W., Mao, Z., Xiao, Z., Liao, X., Koffas, M., Chen, Y., Ma, H., & Tang, Y. J. (2025). Large language model for knowledge synthesis and AI-enhanced biomanufacturing. Trends in Biotechnology, 43(8), 1864–1875. https://doi.org/10.1016/J.TIBTECH.2025.02.008
103. Liebal, U. W., Köbbing, S., Netze, L., Schweidtmann, A. M., Mitsos, A., & Blank, L. M. (2021). Insight to Gene Expression From Promoter Libraries With the Machine Learning Workflow Exp2Ipynb. Frontiers in Bioinformatics, 1, 747428. https://doi.org/10.3389/fbinf.2021.747428
104. Ligarda-Samanez, C. A., Huamán-Carrión, M. L., Palomino-Rincón, H., Taipe-Pardo, F., Moscoso-Moscoso, E., Cabel-Moscoso, D. J., Garcia-Espinoza, A. J., Calderón Huamaní, D. F., Romero Plasencia, J. M., Martinez-Hernandez, J. A., Luciano-Alipio, R., & Apaza-Cruz, J. (2026). Sustainable Biopolymers for Environmental Applications: Advances and Future Perspectives Toward a Circular Economy. Polymers, 18(5), 618. https://doi.org/10.3390/POLYM18050618
105. Lingė, D., Gedgaudas, M., Merkys, A., Petrauskas, V., Vaitkus, A., Grybauskas, A., Paketurytė, V., Zubrienė, A., Zakšauskas, A., Mickevičiūtė, A., Smirnovienė, J., Baranauskienė, L., Crossed D sign apkauskaitė, E., Dudutienė, V., Urniežius, E., Konovalovas, A., Kazlauskas, E., Shubin, K., Schiöth, H. B., … Matulis, D. (2023). PLBD: protein–ligand binding database of thermodynamic and kinetic intrinsic parameters. Database, 2023, baad040. https://doi.org/10.1093/DATABASE/BAAD040
106. Listgarten, J., & Jiang, H. (2026). How artificial intelligence is reengineering protein engineering. Science, 392(6794), 159–166. https://doi.org/10.1126/SCIENCE.AEC8444
107. Liu, Y., Wang, S., Dong, J., Chen, L., Wang, X., Wang, L.,

Li, F., Wang, C., Zhang, J., Wang, Y., Wei, S., Chen, Q., & Liu, H. (2024). De novo protein design with a denoising diffusion network independent of pretrained structure prediction models. Nature Methods, 21(11), 2107–2116. https://doi.org/10.1038/s41592-024-02437-w

108. Ljubešić, Z., Gligora Udovič, M., Overlingė, D., Grgurević, F., Akgül, F., Bacu, A., Díaz-Marrero, A. R., Drakulović, D., Fazi, S., Gaudêncio, S. P., Kolda, A., Novoveska, L., Safarik, I., Sousa, J. R., Thomas, O. P., Reddy, M. M., Varese, G. C., Vasquez, M. I., Makovec, T., & Rotter, A. (2025). Exploring marine microbial diversity: an overview of representative sampling strategies. Frontiers in Marine Science, 12, 1597865. https://doi.org/10.3389/fmars.2025.1597865
109. Lu, Q., Zhang, H., Fan, R., Wan, Y., & Luo, J. (2025). Machine learning-based Bayesian optimization facilitates ultrafiltration process design for efficient protein purification. Separation and Purification Technology, 363, 132122. https://doi.org/10.1016/j.seppur.2025.132122
110. Madani, A., Krause, B., Greene, E. R., Subramanian, S., Mohr, B. P., Holton, J. M., Olmos, J. L., Xiong, C., Sun, Z. Z., Socher, R., Fraser, J. S., & Naik, N. (2023). Large language models generate functional protein sequences across diverse families. Nature Biotechnology, 41(8), 1099–1106. https://doi.org/10.1038/s41587-022-01618-2
111. Maizels, R. J., & Briscoe, J. (2026). Gene regulatory networks: from correlative models to causal explanations. Nature Reviews. Genetics. https://doi.org/10.1038/S41576-026-00939-1
112. Malpetti, D., Scutari, M., Gualdi, F., van Setten, J., van der Laan, S., Haitjema, S., Lee, A. M., Hering, I., & Mangili, F. (2025). Technical and legal aspects of federated learning in bioinformatics: applications, challenges and opportunities. Frontiers in Digital Health, 7, 1644291. https://doi.org/10.3389/FDGTH.2025.1644291/TEXT
113. Marchetti, L., Nifosì, R., Martelli, P. L., Da Pozzo, E., Cappello, V., Banterle, F., Trincavelli, M. L., Martini, C., & D'Elia, M. (2022). Quantum computing algorithms: getting closer to critical problems in computational biology. Briefings in Bioinformatics, 23(6), bbac437. https://doi.org/10.1093/bib/bbac437
114. Marques, L., Costa, B., Pereira, M., Silva, A., Santos, J., Saldanha, L., Silva, I., Magalhães, P., Schmidt, S., & Vale, N. (2024). Advancing Precision Medicine: A Review of Innovative In Silico Approaches for Drug Development, Clinical Pharmacology and Personalized Healthcare. Pharmaceutics, 16(3), 332. https://doi.org/10.3390/PHARMACEUTICS16030332
115. Martin, D. K., Vicente, O., Beccari, T., Kellermayer, M., Koller, M., Lal, R., Marks, R. S., Marova, I., Mechler, A., Tapaloaga, D., Žnidaršič-Plazl, P., & Dundar, M. (2021). A brief overview of global biotechnology. Biotechnology & Biotechnological Equipment, 35(1), 354–363. https://doi.org/10.1080/13102818.2021.1878933
116. Matsubara, T., Machida, S., Owusu, S. P. K., Asakura, A., Hashimoto, H., Matsuoka, M., & Nagasaki, M. (2025). QTFPred: robust high-performance quantum machine learning modeling that predicts main and cooperative transcription factor bindings with base resolution. Briefings in Bioinformatics, 26(6), bbaf604. https://doi.org/10.1093/bib/bbaf604
117. Maurizio, A., & Mazzola, G. (2025). Quantum Computing for Genomics: Conceptual Challenges and Practical Perspectives. PRX Life, 3(4), 047001. https://doi.org/10.1103/h49j-bsc6
118. McKenzie, M., Irac, S. E., Chen, Z., Moradi, A., Jenner, A., Nguyen, Q., & Rashidieh, B. (2026). Integrative spatial omics and artificial intelligence: transforming cancer research with omics data and AI. Seminars in Cancer Biology, 119, 65–82. https://doi.org/10.1016/j.semcancer.2026.01.002
119. Medori, M. C., Bonetti, G., Donato, K., Dhuli, K., Henehan, G., Brown, R., Sieving, P., Sykora, P., Marks, R., Falsini, B., Capodicasa, N., Miertus, S., Lorusso, L., Dondossola, D., Tartaglia, G. M., Tartaglia, G. M., Ergoren, M. C., Dundar, M., Michelini, S., … Bertelli, M. (2023). Bioetics Issues of Artificial Placenta and Artificial Womb Technology. La Clinica Terapeutica, 174(Suppl 2(6)), 243–248. https://doi.org/10.7417/CT.2023.2494
120. Meiser, L. C., Nguyen, B. H., Chen, Y. J., Nivala, J., Strauss, K., Ceze, L., & Grass, R. N. (2022). Synthetic DNA applications in information technology. Nature Communications, 13(1), 352. https://doi.org/10.1038/s41467-021-27846-9
121. Mirchandani, I., Khandhediya, Y., & Chauhan, K. (2025). Review on Advancement of AI in Synthetic Biology. Methods in Molecular Biology, 2952, 483–490. https://doi.org/10.1007/978-1-0716-4690-8_26
122. Nam, Y., Kim, J., Jung, S. H., Woerner, J., Suh, E. H., Lee, D. G., Shivakumar, M., Lee, M. E., & Kim, D. (2024). Harnessing Artificial Intelligence in Multimodal Omics Data Integration: Paving the Path for the Next Frontier in Precision Medicine. Annual Review of Biomedical Data Science, 7(1), 225–250. https://doi.org/10.1146/annurev-biodatasci-102523-103801
123. Narayanan, H., & Love, J. C. (2026). Pichia-CLM: A language model-based codon optimization pipeline for Komagataella phaffii. Proceedings of the National Academy of Sciences of the United States of America, 123(8), e2522052123. https://doi.org/10.1073/PNAS.2522052123
124. Nassir, N., Hashmi, M. A., Raji, K. G., Jamalalail, B., Maksymowsky, A., Scherer, S. W., Alsheikh-Ali, A., & Uddin, M. (2025). Quantum computing and the implementation of precision medicine. Npj Genomic Medicine, 10(1), 80. https://doi.org/10.1038/s41525-025-00537-w
125. Nguyen, E., Poli, M., Durrant, M. G., Kang, B., Katrekar, D., Li, D. B., Bartie, L. J., Thomas, A. W., King, S. H., Brixi, G., Sullivan, J., Ng, M. Y., Lewis, A., Lou, A., Ermon, S., Baccus, S. A., Hernandez-Boussard, T., Ré, C., Hsu, P. D., & Hie, B. L. (2024). Sequence modeling and design from

molecular to genome scale with Evo. Science, 386(6723). https://doi.org/10.1126/science.ado9336

126. Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., & Voigt, C. A. (2016). Genetic circuit design automation. Science, 352(6281). https://doi.org/10.1126/science.aac7341
127. Nijkamp, E., Ruffolo, J. A., Weinstein, E. N., Naik, N., & Madani, A. (2023). ProGen2: Exploring the boundaries of protein language models. Cell Systems, 14(11), 968-978.e3. https://doi.org/10.1016/j.cels.2023.10.002
128. Nilsson, A., Meimetis, N., & Lauffenburger, D. A. (2025). Towards an interpretable deep learning model of cancer. Npj Precision Oncology, 9(1), 46. https://doi.org/10.1038/s41698-025-00822-y
129. Notin, P., Rollins, N., Gal, Y., Sander, C., & Marks, D. (2024). Machine learning for functional protein design. Nature Biotechnology, 42(2), 216–228. https://doi.org/10.1038/s41587-024-02127-0
130. O'Connor, S., Yan, Y., Thilo, F. J. S., Felzmann, H., Dowding, D., & Lee, J. J. (2023). Artificial intelligence in nursing and midwifery: A systematic review. Journal of Clinical Nursing, 32(13–14), 2951–2968. https://doi.org/10.1111/JOCN.16478
131. OECD. (2025). SYNTHETIC BIOLOGY, AI AND AUTOMATION: A FORWARD-LOOKING TECHNOLOGY ASSESSMENT. https://www.oecd.org/en/publications/synthetic-biology-ai-and-automation_12158721-en.html
132. Oetomo, B., Luo, L., Qu, Y., Discepola, M., Kentish, S. E., & Gras, S. L. (2025). Controlling tangential flow filtration in biomanufacturing processes via machine learning: A literature review. Digital Chemical Engineering, 14, 100211. https://doi.org/10.1016/J.DCHE.2024.100211
133. Ozcelik, F., Dundar, M. S., Yildirim, A. B., Henehan, G., Vicente, O., Sánchez-Alcázar, J. A., Gokce, N., Yildirim, D. T., Bingol, N. N., Karanfilska, D. P., Bertelli, M., Pojskic, L., Ercan, M., Kellermayer, M., Sahin, I. O., Greiner-Tollersrud, O. K., Tan, B., Martin, D., Marks, R., … Dundar, M. (2024). The impact and future of artificial intelligence in medical genetics and molecular medicine: an ongoing revolution. Functional & Integrative Genomics, 24(4), 138. https://doi.org/10.1007/S10142-024-01417-9
134. Pal, S., Bhattacharya, M., Lee, S. S., & Chakraborty, C. (2024). Quantum Computing in the Next-Generation Computational Biology Landscape: From Protein Folding to Molecular Dynamics. Molecular Biotechnology, 66(2), 163–178. https://doi.org/10.1007/s12033-023-00765-4
135. Palit, P., Minkara, M., Abida, M., Marwa, S., Sen, C., Roy, A., Pasha, M. R., Mosae, P. S., Saha, A., & Ferdoush, J. (2025). PlastiCRISPR: Genome Editing-Based Plastic Waste Management with Implications in Polyethylene Terephthalate (PET) Degradation. Biomolecules, 15(5), 684. https://doi.org/10.3390/BIOM15050684
136. Pamidimukkala, J. V., Bopardikar, S., Dakshinamoorthy, A., Kannan, A., Dasgupta, K., & Senapati, S. (2024). Protein Structure Prediction with High Degrees of Freedom in a Gate-Based Quantum Computer. Journal of Chemical Theory and Computation, 20(22), 10223–10234. https://doi.org/10.1021/acs.jctc.4c00848
137. Pannu, J., Bloomfield, D., MacKnight, R., Hanke, M. S., Zhu, A., Gomes, G., Cicero, A., & Inglesby, T. V. (2025). Dual-use capabilities of concern of biological AI models. PLoS Computational Biology, 21(5), e1012975. https://doi.org/10.1371/journal.pcbi.1012975
138. Peykani, P., Ramezanlou, F., Tanasescu, C., & Ghanidel, S. (2025). Large Language Models: A Structured Taxonomy and Review of Challenges, Limitations, Solutions, and Future Directions. Applied Sciences, 15(14), 8103. https://doi.org/10.3390/app15148103
139. Pinto, J., Ramos, J. R. C., Costa, R. S., & Oliveira, R. (2023). A General Hybrid Modeling Framework for Systems Biology Applications: Combining Mechanistic Knowledge with Deep Neural Networks under the SBML Standard. AI (Switzerland), 4(1), 303–318. https://doi.org/10.3390/ai4010014
140. Plante, M., Champie, A., Michaud, F., & Rodrigue, S. (2026). Toward full automation in synthetic biology: A progressive conceptual framework integrating robotics and intelligent agents. SLAS Technology, 36, 100378. https://doi.org/10.1016/J.SLAST.2025.100378
141. Priyadharshini, M., Raju, B. D., Banu, A. F., Kumar, P. J., Murugesh, V., & Rybin, O. (2025). A quantum machine learning framework for predicting drug sensitivity in multiple myeloma using proteomic data. Scientific Reports, 15(1), 26553. https://doi.org/10.1038/s41598-025-06544-2
142. Puniya, B. L. (2025). Artificial-intelligence-driven Innovations in Mechanistic Computational Modeling and Digital Twins for Biomedical Applications. Journal of Molecular Biology, 437(17), 169181. https://doi.org/10.1016/J.JMB.2025.169181
143. Qin, K., Liu, F., Zhang, C., Deng, R., Fernie, A. R., & Zhang, Y. (2025). Systems and synthetic biology for plant natural product pathway elucidation. Cell Reports, 44(6), 115715. https://doi.org/10.1016/j.celrep.2025.115715
144. Quek, Y., Stilck França, D., Khatri, S., Meyer, J. J., & Eisert, J. (2024). Exponentially tighter bounds on limitations of quantum error mitigation. Nature Physics, 20(10), 1648–1658. https://doi.org/10.1038/s41567-024-02536-7
145. Radivojević, T., Costello, Z., Workman, K., & Garcia Martin, H. (2020). A machine learning Automated Recommendation Tool for synthetic biology. Nature Communications, 11(1), 4879. https://doi.org/10.1038/s41467-020-18008-4
146. Rai, K., Wang, Y., O'Connell, R. W., Patel, A. B., & Bashor, C. J. (2024). Using machine learning to enhance and accelerate synthetic biology. Current Opinion in Biomedical Engineering, 31, 100553. https://doi.org/10.1016/J.COBME.2024.100553
147. Rashidi, M., Arima, S., Stetco, A. C., Coppola, C., Musarò,

D., Greco, M., Damato, M., My, F., Lupo, A., Lorenzo, M., Danieli, A., Maruccio, G., Argentiero, A., Buccoliero, A., Donzella, M. D., & Maffia, M. (2025). Prediction of Parkinson Disease Using Long-Term, Short-Term Acoustic Features Based on Machine Learning. Brain Sciences, 15(7), 739. https://doi.org/10.3390/BRAINSCI15070739

148. Richter, J., Wang, Q., Lange, F., Thiel, P., Yilmaz, N., Solle, D., Zhuang, X., & Beutel, S. (2025). Machine Learning-Powered Optimization of a CHO Cell Cultivation Process. Biotechnology and Bioengineering, 122(5), 1153–1164. https://doi.org/10.1002/BIT.28943
149. Rotter, A., Bacu, A., Barbier, M., Bertoni, F., Bones, A. M., Cancela, M. L., Carlsson, J., Carvalho, M. F., Cegłowska, M., Dalay, M. C., Dailianis, T., Deniz, I., Drakulovic, D., Dubnika, A., Einarsson, H., Erdoğan, A., Eroldoğan, O. T., Ezra, D., Fazi, S., … Vasquez, M. I. (2020). A New Network for the Advancement of Marine Biotechnology in Europe and Beyond. Frontiers in Marine Science, 7, 524577. https://doi.org/10.3389/FMARS.2020.00278/TEXT
150. Selvaraj, A. A., Jayawant, M., Kayalvizhi, N., Krishnan, M., Puvaneshvari, N., Santhosh Kumar, A. W., & Rameshkumar, N. (2026). Exploring quantum frontiers in protein structure prediction: techniques, challenges, and opportunities. Methods, 251, 94–112. https://doi.org/10.1016/J.YMETH.2026.04.006
151. Serag, I., Azzam, A. Y., Hassan, A. K., Diab, R. A., Diab, M., Hefnawy, M. T., Ali, M. A., & Negida, A. (2025). Multimodal diagnostic tools and advanced data models for detection of prodromal Parkinson's disease: a scoping review. BMC Medical Imaging, 25(1), 103. https://doi.org/10.1186/S12880-025-01620-5/TABLES/1
152. Shi, J., Bendig, D., Vollmar, H. C., & Rasche, P. (2023). Mapping the Bibliometrics Landscape of AI in Medicine: Methodological Study. Journal of Medical Internet Research, 25(1), e45815. https://doi.org/10.2196/45815
153. Shokrpour, S., MoghadamFarid, A. M., Bazzaz Abkenar, S., Haghi Kashani, M., Akbari, M., & Sarvizadeh, M. (2025). Machine learning for Parkinson's disease: a comprehensive review of datasets, algorithms, and challenges. Npj Parkinson's Disease, 11(1), 187. https://doi.org/10.1038/s41531-025-01025-9
154. Shrivastava, A., Nikita, S., & Rathore, A. S. (2024). Machine learning tool as an enabler for rapid quantification of monoclonal antibodies N-glycans using fluorescence detector. International Journal of Biological Macromolecules, 271, 132694. https://doi.org/10.1016/J.IJBIOMAC.2024.132694
155. Singh, N., Lane, S., Yu, T., Lu, J., Ramos, A., Cui, H., & Zhao, H. (2025). A generalized platform for artificial intelligence-powered autonomous enzyme engineering. Nature Communications, 16(1), 5648. https://doi.org/10.1038/s41467-025-61209-y
156. Sokra, I., & Meta, H. (2026). Bioplastics in the Circular Bioeconomy: Production Pathways, Biodegradation Mechanisms, and Environmental Implications. Journal of Agriculture and Technology, 2(2), 23–35. https://doi.org/10.6084/m9.figshare.30986584
157. Sola, P. V., Montenegro, J. F., Moreira, S. I., Lorenzo, F. R., Vandervorst, P., González, E. P., Lorenzo, M. P., Couñago, I. P., Landín, S. M., Castilla, L. H., Quinteiro, J. I., Álvarez Rodríguez, J. A., Algarra, B. A., Flo, E. G., & García, J. (2025). PHB in cyanobacteria: analyzing production through images processing and FT-IR techniques. New Biotechnology, 89, 119–129. https://doi.org/10.1016/j.nbt.2025.07.003
158. Srivastava, S., Benerjee, K. G., & Banerjee, A. (2024). Efficient Bidirectional RNNs for Substitution Error Correction in DNA Data Storage. 2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN), 434–439. https://doi.org/10.1109/ICMLCN59089.2024.10625179
159. Sung, J. Y., Park, S. min, & Cheong, J. H. (2026). Quantum computing for biotechnological innovation: transformative potential in biomedicine and drug discovery. Trends in Biotechnology, S0167-7799(26), 00008–00009. https://doi.org/10.1016/J.TIBTECH.2026.01.008
160. Tankhilevich, E., Martinez Cuesta, S., Barrett, I., Berg, C., Holmberg Schiavone, L., & Leach, A. R. (2026). RP3Net: a deep learning model for predicting recombinant protein production in Escherichia coli. Bioinformatics, 42(1), btag003. https://doi.org/10.1093/BIOINFORMATICS/BTAG003
161. Theodoris, C. V., Xiao, L., Chopra, A., Chaffin, M. D., Al Sayed, Z. R., Hill, M. C., Mantineo, H., Brydon, E. M., Zeng, Z., Liu, X. S., & Ellinor, P. T. (2023). Transfer learning enables predictions in network biology. Nature 2023 618:7965, 618(7965), 616–624. https://doi.org/10.1038/s41586-023-06139-9
162. Thomson, T., Li, G., Strilchuk, A., Cui, H., Wang, B., & Li, B. (2025). Harnessing artificial intelligence to advance CRISPR-based genome editing technologies. Nature Reviews Genetics, 27(3), 212–230. https://doi.org/10.1038/s41576-025-00907-1
163. Tobias, A. V., & Wahab, A. (2025). Autonomous 'self-driving' laboratories: a review of technology and policy implications. Royal Society Open Science, 12(7), 250646. https://doi.org/10.1098/RSOS.250646/235354
164. Trif, C., Le, S. Q., Vunduk, J., Parcharoen, Y., & Marks, R. S. (2025). A novel bioluminescent bacteria-based method coupled with dynamic time warping for detecting and differentiating copper and mercury in water. Water Research, 286, 124230. https://doi.org/10.1016/J.WATRES.2025.124230
165. Trif, C., Vunduk, J., Parcharoen, Y., Bualuang, A., & Marks, R. S. (2024). Bioluminescent Whole-Cell Bioreporter Bacterial Panel for Sustainable Screening and Discovery of Bioactive Compounds Derived from Mushrooms. Biosensors, 14(11), 558. https://doi.org/10.3390/BIOS14110558
166. Tudor, B. H., Shargo, R., Gray, G. M., Fierstein, J. L., Kuo, F.

H., Burton, R., Johnson, J. T., Scully, B. B., Asante-Korang, A., Rehman, M. A., & Ahumada, L. M. (2025). A scoping review of human digital twins in healthcare applications and usage patterns. Npj Digital Medicine, 8(1), 587. https://doi.org/10.1038/s41746-025-01910-w

167. Urbina, F., Lentzos, F., Invernizzi, C., & Ekins, S. (2022). Dual use of artificial-intelligence-powered drug discovery. Nature Machine Intelligence, 4(3), 189–191. https://doi.org/10.1038/s42256-022-00465-9
168. Uttarkar, A., & Niranjan, V. (2024). Quantum synergy in peptide folding: A comparative study of CVaR-variational quantum eigensolver and molecular dynamics simulation. International Journal of Biological Macromolecules, 273, 133033. https://doi.org/10.1016/J.IJBIOMAC.2024.133033
169. Uttarkar, A., Niranjan, V., Saxena, A., & Kumar, V. (2026). QuPepFold: A python package for hybrid quantum-classical protein folding simulations with CVaR-optimized VQE. PloS One, 21(2), e0342012. https://doi.org/10.1371/JOURNAL.PONE.0342012
170. Valencia-Velásquez, J., Yaker-Moreno, H. A., Martínez-Guerrero, A., Ibáñez-Espinel, F., Pérez-Correa, J. R., & Caicedo-Ortega, N. H. (2025). Advancing hybrid modeling of Saccharomyces cerevisiae fermentation with mixed carbon sources and urea in a mini-stirred tank reactor. Bioprocess and Biosystems Engineering, 48(11), 1919–1937. https://doi.org/10.1007/S00449-025-03222-5
171. Vasina, M., Kovar, D., Damborsky, J., Ding, Y., Yang, T., deMello, A., Mazurenko, S., Stavrakis, S., & Prokop, Z. (2023). In-depth analysis of biocatalysts by microfluidics: An emerging source of data for machine learning. Biotechnology Advances, 66, 108171. https://doi.org/10.1016/J.BIOTECHADV.2023.108171
172. Vázquez Torres, S., Benard Valle, M., Mackessy, S. P., Menzies, S. K., Casewell, N. R., Ahmadi, S., Burlet, N. J., Muratspahić, E., Sappington, I., Overath, M. D., Rivera-de-Torre, E., Ledergerber, J., Laustsen, A. H., Boddum, K., Bera, A. K., Kang, A., Brackenbrough, E., Cardoso, I. A., Crittenden, E. P., … Baker, D. (2025). De novo designed proteins neutralize lethal snake venom toxins. Nature, 639(8053), 225–231. https://doi.org/10.1038/s41586-024-08393-x
173. Volk, M. J., Lourentzou, I., Mishra, S., Vo, L. T., Zhai, C., & Zhao, H. (2020). Biosystems Design by Machine Learning. ACS Synthetic Biology, 9(7), 1514–1533. https://doi.org/10.1021/ACSSYNBIO.0C00129
174. Wahab, W. A. A. (2025). Review of research progress in immobilization and chemical modification of microbial enzymes and their application. Microbial Cell Factories, 24(1), 167. https://doi.org/10.1186/S12934-025-02791-0
175. Wang, X., Xu, K., Huang, Z., Lin, Y., Zhou, J., Zhou, L., & Ma, F. (2025). Accelerating promoter identification and design by deep learning. Trends in Biotechnology, 43(12), 3071–3087. https://doi.org/10.1016/j.tibtech.2025.05.008
176. Wang, Y., Wang, H., Wei, L., Li, S., Liu, L., & Wang, X. (2020). Synthetic promoter design in Escherichia coli based on a deep generative network. Nucleic Acids Research, 48(12), 6403–6412. https://doi.org/10.1093/nar/gkaa325
177. Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Hanikel, N., Pellock, S. J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S. V., … Baker, D. (2023). De novo design of protein structure and function with RFdiffusion. Nature, 620(7976), 1089–1100. https://doi.org/10.1038/s41586-023-06415-8
178. Wen, G., & Li, L. (2025). Federated transfer learning with differential privacy for multi-omics survival analysis. Briefings in Bioinformatics, 26(2), bbaf166. https://doi.org/10.1093/BIB/BBAF166
179. Wittmann, B. J., Alexanian, T., Bartling, C., Beal, J., Clore, A., Diggans, J., Flyangolts, K., Gemler, B. T., Mitchell, T., Murphy, S. T., Wheeler, N. E., & Horvitz, E. (2025). Strengthening nucleic acid biosecurity screening against generative protein design tools. Science, 390(6768), 82–87. https://doi.org/10.1126/science.adu8578
180. Wu, W., Xiang, L., Liu, Q., & Yang, K. (2023). Deep Joint Source-Channel Coding for DNA Image Storage: A Novel Approach With Enhanced Error Resilience and Biological Constraint Optimization. IEEE Transactions on Molecular, Biological, and Multi-Scale Communications, 9(4), 461–471. https://doi.org/10.1109/TMBMC.2023.3331579
181. Xia, Y., & Huo, Y. X. (2026). Controlling gene expression using AI designed Cis-regulatory elements. Biotechnology Advances, 87, 108802. https://doi.org/10.1016/j.biotechadv.2026.108802
182. Xu, D., Liu, B., Wang, J., & Zhang, Z. (2022). Bibliometric analysis of artificial intelligence for biotechnology and applied microbiology: Exploring research hotspots and frontiers. Frontiers in Bioengineering and Biotechnology, 10, 998298. https://doi.org/10.3389/FBIOE.2022.998298
183. Xu, K., Yu, S., Wang, K., Tan, Y., Zhao, X., Liu, S., Zhou, J., & Wang, X. (2024). AI and Knowledge-Based Method for Rational Design of Escherichia coli Sigma70 Promoters. ACS Synthetic Biology, 13(1), 402–407. https://doi.org/10.1021/acssynbio.3c00578
184. 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
185. Yang, T., Xiao, Y., Bao, Z., Hao, J., & Peng, J. (2025). The rise and potential opportunities of large language model agents in bioinformatics and biomedicine. Briefings in Bioinformatics, 26(6), bbaf601. https://doi.org/10.1093/BIB/BBAF601
186. Yetgin, A. (2025). Revolutionizing multi-omics analysis with artificial intelligence and data processing. Quantitative Biology, 13(3), e70002. https://doi.org/10.1002/qub2.70002
187. Yildirim, D. T., Yildirim, A. B., Martin, D., Beccari, T.,

Vicente, O., Radecka, I., Tchuenbou-Magaia, F., Marks, R., Lal, R., Prakash, S., Mechler, A., Milkov, M. P., Georgieva, S. F., Iliev, I., Kaneva, R., Gokce, N., Yaneva, G. A., Agova, N. V., Ivanova, N. D., … Dundar, M. (2026). Integrated Biotechnological Strategies for Planetary Health and Ecosystem Restoration. The EuroBiotech Journal, 10(2), 81–102. https://doi.org/10.2478/EBTJ-2026-0009

188. Yin, S. (2025). Artificial Intelligence-Assisted Nanosensors for Clinical Diagnostics: Current Advances and Future Prospects. Biosensors, 15(10), 656. https://doi.org/10.3390/BIOS15100656
189. Young, A. L., Oxtoby, N. P., Garbarino, S., Fox, N. C., Barkhof, F., Schott, J. M., & Alexander, D. C. (2024). Data-driven modelling of neurodegenerative disease progression: thinking outside the black box. Nature Reviews Neuroscience, 25(2), 111–130. https://doi.org/10.1038/s41583-023-00779-6
190. Yuan, S., Xu, V., Muddana, C., Sureshkumar, P., & Tang, Y. J. (2026). From design–build–test–learn cycles to AI-driven digital twins for bioprocess scale-up in the Genesis Mission era. Current Opinion in Biotechnology, 100, 103516. https://doi.org/10.1016/J.COPBIO.2026.103516
191. Zack, M., Stupichev, D. N., Moore, A. J., Slobodchikov, I. D., Sokolov, D. G., Trifonov, I. F., & Gobbs, A. (2025). Artificial Intelligence and Multi-Omics in Pharmacogenomics: A New Era of Precision Medicine. Mayo Clinic Proceedings: Digital Health, 3(3), 100246. https://doi.org/10.1016/j.mcpdig.2025.100246
192. Zampieri, G., Sandner, V., Verma, S., Kraemer, J., Lennon, C., Occhipinti, A., McCreath, G., & Angione, C. (2026). Bioprocess optimisation via joint machine learning and metabolic modelling. Metabolic Engineering, 96, 113–128. https://doi.org/10.1016/j.ymben.2026.03.004
193. Zhang, G., Liu, C., Lu, J., Zhang, S., & Zhu, L. (2025). The Role of AI-Driven De Novo Protein Design in the Exploration of the Protein Functional Universe. Biology, 14(9), 1268. https://doi.org/10.3390/biology14091268
194. Zhou, C., Jiang, F., Chen, W., Nugen, S. R., & Huang, C. (2025). Synthetic biology meets diagnostics: Engineering biosensing platforms for rapid and accurate pathogen and viral detection. Biosensors and Bioelectronics, 290, 117946. https://doi.org/10.1016/J.BIOS.2025.117946
195. Zhou, X., Ding, N., Zhou, S., & Deng, Y. (2026). AI-Guided Design and Predictive Modeling of Synthetic Escherichia coli Promoters through Comprehensive −10/−35 Box Engineering. ACS Synthetic Biology, 15(2), 728–739. https://doi.org/10.1021/ACSSYNBIO.5C00765
196. Zubrienė, A., Kurtenoka, M., Paketurytė-Latvė, V., Leitans, J., Manakova, E., Žvirblis, M., Kazaks, A., Eimonta, V., Tars, K., Gražulis, S., Petrauskas, V., Matulienė, J., Dudutienė, V., Shubin, K., & Matulis, D. (2026). Achieving femtomolar affinities in structure-based drug design. European Biophysics Journal, 55(1), 55–62. https://doi.org/10.1007/S00249-025-01812-5