How AI Is Transforming Microbial Ingredient Production

Population growth, shrinking arable land and climate change are intensifying pressure on global food systems. Microorganism-based ingredients offer a compelling alternative; many fungi, yeasts and microalgae contain 50–70% protein by dry weight, and modern metabolic engineering now enables microbes to produce specialised food proteins. Fermentation already underpins the production of vitamins, flavours, sweeteners, pigments, enzymes and cosmetic actives, while oleaginous microbes provide sustainable routes to edible and functional oils.

Recent advances in synthetic biology and multi-omics have made microbial platforms more predictable and versatile. Yet major challenges remain in discovering the right ingredients, engineering efficient production pathways and ensuring quality at scale. This article explores how artificial intelligence (AI) can accelerate these steps from ingredient discovery to bioprocess optimisation.

Ingredient Innovation Through AI-Driven Biotechnology: Revolutionising Sourcing and Yield Optimisation

AI-driven approaches tackle ingredient innovation through three complementary strategies:

1. Screening natural microbial diversity to identify organisms that naturally produce compounds of interest.
2. Enzymatic engineering for optimising catalytic processes and discovering novel enzymes.
3. Metabolic pathway engineering for creating entirely new biosynthetic routes through microbial cell factories.

Together, these methodologies are reshaping how industries think about ingredient sourcing, moving from synthetic chemistry to synthetic biology.

1. AI in Natural Ingredient Discovery

To get a picture of what a microorganism can and cannot produce, one needs a detailed mapping of its genes and metabolites. Classical approaches usually rely on similarity searches against previously characterised proteins. Hence, most protein function remain unknown; only 0.3% of the 250 million protein sequences catalogued in the UniProt database have a functional annotation.[1] This leads, down the road, to an incomplete picture of an organism's metabolic potential, that is, what compound can an organism produce. ML models have been proven effective in predicting proteins' enzymatic activity, reaching 87% accuracy (Yu *et al.*). Yu and colleagues demonstrated the value of their model by validating their prediction on halogenases, an enzyme that is difficult to predict but plays an important role in the synthesis of drugs and other bioactives. They reached an accuracy of 86–100%.[2] Prediction of protein structure by AlphaFold[2] and others has further boosted functional annotations through structure similarity annotation.[1] The same principle, inferring biological function directly from sequence data, can be extended beyond enzymes to other metabolic traits, for instance, oils. Transformer models applied to genomic sequences can predict a yeast's propensity to accumulate lipids, enabling the discovery of novel oleaginous yeasts with unique fatty-acid profiles.[3] Some of these profiles could potentially replace up to 74 plant-based oils.

2. AI-Enhanced Enzymatic Engineering: Optimising Nature's Catalysts

Engineering proteins is commonly used in drug manufacturing, agriculture, consumer products and more. Antibodies, for instance, are engineered to enhance their binding and specificity as therapeutics, whereas the stabilities and activities of enzymes can be improved under process conditions to reach more efficient chemical syntheses. AI models integrating sequence, structural and high-throughput experimental data help prioritise variants with enhanced activity, selectivity and robustness.[4] These methods allowed Climax Food to design a Casein analog produced by precision fermentation that replicates the taste and texture of real cheese in plant-based products.[5] Generative AI represents a fundamental shift from analysing and modifying natural proteins to designing them from scratch. Protein language models (e.g., ProteinMPNN6) and diffusion-based networks enable function-first design. Users specify catalytic requirements, and models generate stable, foldable scaffolds beyond evolutionary space. Autonomous platforms coupling large language models with robotic biofoundries require only target function and fitness metrics,[7] unlocking enzymes for non-natural substrates and achieving catalytic efficiencies exceeding directed evolution.

Ginkgo Bioworks showcased the power of generative AI in enzyme engineering through a landmark project on a key enzyme in Central Carbon Metabolism, one that had seen only a two-fold performance gain in over 50 years of traditional research. Using their AI platform Owl, Ginkgo designed and screened iterative libraries beginning with 2,000 variants, and within just four AI-guided generations achieved a whopping 10-fold increase in catalytic efficiency. This breakthrough, validated through extensive activity assays and protein characterisation, shows how generative models can rapidly learn sequence–function relationships and produce highly customised biocatalysts.

Another example of generative AI applied to novel ingredient design is showcased by TastePepAI, a platform for de novo design of multifunctional taste peptides with customisable flavour profiles.[8] The system is built on a tailored variational autoencoder that learns from known sweet, salty and umami peptides while avoiding undesirable profiles such as bitterness. This allows the model to generate entirely new peptide sequences predicted to have specific taste properties. Candidate peptides are then screened for safety using SpepToxPred, an AI toxicity predictor optimised for short food peptides. In experimental validation, the team successfully identified 73 new taste peptides, all of which showed the expected flavour characteristics in electronic-tongue testing and demonstrated excellent biocompatibility. Such AI-accelerated workflows now make it possible to design enzymes for ingredient manufacturing, such as lipases for emollient esters or glycosyltransferases for robust prebiotic oligosaccharides, faster and more precisely.

3. AI-Powered Metabolic Pathway Engineering: Redesigning Production from First Principles

Several graph-based algorithms and other non–machine learning approaches have been successfully used to guide metabolic pathway engineering. These methods rely on curated biochemical reaction databases, such as MetaCyc and KEGG, and have enabled the biosynthesis of a range of target molecules, including vanillin in engineered yeast. Integrating ML into these established retro-biosynthetic workflows shows strong potential to accelerate pathway design by reducing the number of trial-and-error iterations typically required in metabolic engineering.[9] Such hybrid approaches could also unlock the production of molecules previously considered technically or economically infeasible. Recent AI-enabled platforms illustrate this promise. BioNavi-NP, for example, identifies biosynthetic pathways for over 90% of test compounds and achieves 1.7-fold higher accuracy than conventional rule-based tools.[10] These computational advances support more accurate and scalable strain-design pipelines. As a practical demonstration, Moreno-Paz et al. reported a 68% increase in p-coumaric acid titre in yeast after two machine-learning-guided Design–Build–Test–Learn (DBTL) cycles, underscoring how ML can directly enhance microbial production performance.[11]

Manufacturing Optimisation and Quality Control

Optimising and controlling fermentation remains one of the greatest challenges in microorganism-based ingredient production. These steps are essential for reducing variability, lowering production costs and achieving competitiveness as a food source. This challenge is amplified in fermented foods that depend on complex environmental microflora rather than a single organism, making them inherently susceptible to microbial shifts, batch-to-batch inconsistency and safety risks.

1. AI/ML for Real-Time Process Monitoring

The integration of IoT technologies into fermenters has transformed how fermentation is monitored and controlled. Real-time sensors and cloud-based data pipelines now enable high-resolution tracking of key variables with greater precision and scalability.[12] Smart biosensors, including electronic noses and tongues, further strengthen process transparency by detecting biochemical markers, such as glucose or amino acids, at extremely low concentrations. Recent advances in AI-enhanced biosensors improve contaminant detection and overall process reliability. Dynamic regulation of fermentation parameters helps to maintain optimal culture conditions while limiting human intervention. Reinforcement learning approaches have been used for adaptive control strategies that adjust process variables in real-time.

The emergence of deep reinforcement learning (DRL) has strengthened adaptive process control by combining reinforcement learning with deep neural networks, enabling faster and more accurate real-time decision-making. For example, the DRL controller improved penicillin yield by 14%, highlighting its promise for real-world fed-batch fermentation systems.[13]

2. Digital Twins

Digital twins (DTs) are emerging as a powerful addition to this ecosystem. A DT is a virtual representation of a biological process that integrates multi-omics data (genomics, transcriptomics, proteomics, metabolomics) alongside real-time sensor data from bioprocessors. Together, these layers create a dynamic virtual model of organism growth and metabolism under varying conditions. Such systems allow *in silico* testing of genetic modifications, medium compositions or operating conditions, potentially reducing months or years of iterative experimentation.[14] Although digital twins have proven effective in multiple industries, their adoption in biomanufacturing remains early-stage due to biological complexity and stringent regulatory constraints.

Promising examples still exist, notably in traditional fermented ingredients. In the beer industry, hybrid digital twins combining multiple ML models have been used to predict beer quality throughout fermentation.[15] Similarly, AI-driven modelling in wine fermentation enables dynamic condition adjustment, flavour optimisation and quality forecasting.[16] In algae-based bioactive compound production, AI/ML systems have been used to continuously update operational parameters and predict biomass yields.[16] It is important to recognise that most bioprocess digital twins are not purely ML-driven. Physics-based models, particularly mass-balance ordinary differential equations, remain widely used because they rely on fewer parameters and do not require the extensive datasets needed for ML methods. ML approaches, while powerful, are often limited by the high cost and labour required to generate sufficiently large and diverse experimental data. Hybrid methods combining ML approaches have also demonstrated benefits in fermentation optimisation. For instance, a back-propagation neural network (BPNN) coupled with an adaptive genetic algorithm (AGA) improved Lincomycin fermentation efficiency and increased yield by 8%.[17]

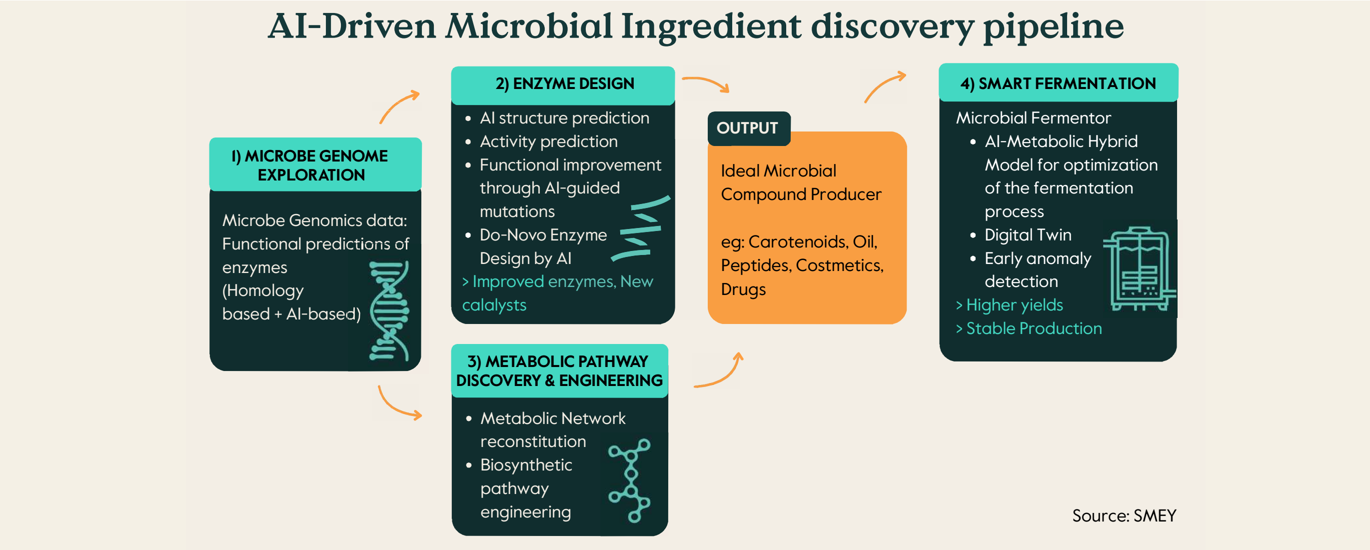
Conclusion

AI is rapidly transforming microbial ingredient innovation, from predicting protein function to designing enzymes, rewiring metabolic pathways and modernising process control. Its performance, however, depends on access to diverse, high-quality biological datasets, while publicly available genomic and functional data tend to be skewed toward a narrow group of model organisms. Initiatives such as IMG/M,[18] the Global Microbial Gene Catalogue[19] and large-scale efforts like BaseData[20] are beginning to close this gap.

As these resources expand, AI will increasingly complement, rather than replace, established biological modelling, enabling faster discovery, more efficient engineering and more reliable manufacturing of food and cosmetic ingredients.

REFERENCES

1. Wang X, Yang P, Zhao B, Liu S. AI-assisted food enzymes design and engineering: a critical review. Systems Microbiology and Biomanufacturing. 2022 Oct 1;3(1):75–87.
2. Yu T, Cui H, Li JC, Luo Y, Jiang G, Zhao H. Enzyme function prediction using contrastive learning. Science. 2023 Mar 31;379(6639):1358–63.
3. Lawrence C. Creating biotech for the post-deforestation era: SMEY launches NOY, a "Neobank of Yeasts" for cultivated oils (Internet). Tech.eu. 2025 (cited 2025 Nov 26). Available from: https://tech.eu/2025/06/26/oil-biotech-for-the-post-deforestation-era-smey-



AI-Driven Microbial Ingredient discovery pipeline

1) MICROBE GENOME EXPLORATION

Microbe Genomics data: Functional predictions of enzymes (Homology based + AI-based)

2) ENZYME DESIGN

- AI structure prediction
- Activity prediction
- Functional improvement through AI-guided mutations
- Do-Novo Enzyme Design by AI

> Improved enzymes, New calalysts

3) METABOLIC PATHWAY DISCOVERY & ENGINEERING

- Metabolic Network reconstitution
- Biosynthetic pathway engineering

OUTPUT

Ideal Microbial Compound Producer

eg: Carotenoids, Oil, Peptides, Costmetics, Drugs

4) SMART FERMENTATION

Microbial Fermentor
- AI-Metabolic Hybrid Model for optimization of the fermentation process
- Digital Twin
- Early anomaly detection

> Higher yields
> Stable Production

Source: SMEY

launches-noy-a-neobank-of-yeasts-for-cultivated-oils/

4. Orsi E, Schada von Borzyskowski L, Noack S, Nikel PI, Lindner SN. Automated *in vivo* enzyme engineering accelerates biocatalyst optimization. Nature Communications (Internet). 2024 Apr 24;15(1):3447. Available from: https://www.nature.com/articles/s41467-024-46574-4
5. BabyBel maker Bel Group, biotech Climax Foods partner to refine plant-based cheese with AI (Internet). FoodNavigator-USA.com. 2023 (cited 2025 Nov 27). Available from: https://www.foodnavigator-usa.com/Article/2023/04/12/babybel-maker-bel-group-biotech-climax-foods-partner-to-refine-plant-based-cheese-with-ai/
6. Jin S, Wu Q, Fu G, Lu D, Wang F, Deng L, et al. Breaking Evolution's Ceiling: AI-Powered Protein Engineering. Catalysts. 2025 Sep 2;15(9):842.
7. Singh N, Lane S, Yu T, Lu J, Ramos A, Cui H, et al. A generalized platform for artificial intelligence-powered autonomous enzyme engineering. Nature Communications (Internet). 2025 Jul 1;16(1). Available from: https://www.nature.com/articles/s41467-025-61209-y
8. Yue J, Li T, Ouyang J, Xu J, Tan H, Chen Z, et al. TastepepAI: An artificial intelligence platform for taste peptide de novo design. PLoS Computational Biology. 2025 Oct 16;21(10):e1013602–2.
9. Guillaume Gricourt, Meyer P, Duigou T, Jean-Loup Faulon. Artificial Intelligence Methods and Models for Retro-Biosynthesis: A Scoping Review. ACS Synthetic Biology. 2024 Jul 24;13(8):2276–94.
10. Zheng S, Zeng T, Li C, Chen B, Coley CW, Yang Y, et al. Deep learning driven biosynthetic pathways navigation for natural products with BioNavi-NP. Nature Communications (Internet). 2022 Jun 10 (cited 2023 Jun 1);13(1):3342. Available from: https://www.nature.com/articles/s41467-022-30970-9
11. Moreno-Paz S, van, Elif Eliana, Zwartjens P, Gosiewska S, Santos, et al. Machine Learning-Guided Optimization of p-Coumaric Acid Production in Yeast. ACS synthetic biology. 2024 Mar 28;13(4):1312–22.
12. Yee CS, Nur Asyiqin Zahia-Azizan, Muhamad, Mohd A, Raja Balqis Raja-Razali, Muhammad Ameer Ushidee-Radzi, et al. Smart Fermentation Technologies: Microbial Process Control in Traditional Fermented Foods. Fermentation (Internet). 2025 Jun 5;11(6):323–3. Available from: https://www.mdpi.com/2311-5637/11/6/323
13. Li H, Qiu T, You F. AI-based optimal control of fed-batch biopharmaceutical process leveraging deep reinforcement learning. Chemical Engineering Science. 2024 Jun;292:119990.
14. Helmy M, Elhalis H, Rashid MM, Selvarajoo K. Can digital twin efforts shape microorganism-based alternative food? Current Opinion in Biotechnology (Internet). 2024 Jun (cited 2025 May 20);87:103115. Available from: https://doi.org/10.1016/j.copbio.2024.103115
15. Colomba Dazzarola, Tighe R, Pérez-Correa JR, Saa PA. Toward a digital twin for beer quality control: development of a digital model integrating industrial process data and model-based fermentation descriptors. Journal of Food Engineering. 2025 Jul 1;112726–6.
16. Khosravi P, Cosimo D'Aiello. Enhancing Wine Fermentation: The Role of AI-Driven Predictive Modeling in Flavor Optimization. Communications in computer and information science. 2025 Jan 1;64–83.
17. Guo B, Lu X, Jiang X, Shen XL, Wei Z, Zhang Y. Artificial Intelligence in Advancing Algal Bioactive Ingredients: Production, Characterization, and Application. Foods. 2025 May 17;14(10):1783–3.
18. Li Z, Faiza Atique, Shahzad M, Rehman KU. Intelligent Control Strategy Based on Back-Propagation Neural Network with Adaptive Genetic Algorithm for Lincomycin Fermentation Process. Industrial Biotechnology. 2022 Mar 30;18(2):98–105.
19. JGI IMG Integrated Microbial Genomes & Microbiomes (Internet). Doe.gov. 2018. Available from: https://img.jgi.doe.gov/
20. Ivica Letunic. GMGC: Global Microbial Gene Catalog (Internet). Embl.de. 2020 (cited 2025 Nov 27). Available from: https://gmgc.embl.de/
21. Vince O, Oldach P, Pereno V, Leung MHY, Greco C, Minto-Cowcher G, et al. Breaking Through Biology's Data Wall: Expanding the Known Tree of Life by Over 10x using a Global Biodiscovery Pipeline. BioRxiv. 2025 Jun 14

Julie Rojas

Julie Rojas, PhD, is a computational biologist and Scientist AI at SMEY and holds a doctorate from the Ludwig-Maximilians-Universität München (Germany). She develops machine learning models to predict lipid production and profiles in wild yeast strains for sustainable ingredient discovery. With over six years of experience in bioinformatics and multi-omics analysis, she combines genomics, lipidomics and statistical modelling to accelerate microbial innovation in oils, enzymes and fermentation-based bioprocesses.

Email: julie.r@smey.cc

Heykel Trabelsi

Heykel Trabelsi, PhD, is a Product Manager at SMEY and holds a doctorate in Biotechnology from the Université catholique de Louvain (Belgium). His work focuses on microbial strain engineering for the production of high-value biosourced ingredients. At SMEY, he contributes to the development of a novel yeast neobank (NOY), enabling lipid production from non-GMO yeasts and the design of tailored oils and fats.

Email: heykel.t@smey.cc