Can Digital Twin Efforts Shape Alternative Food?

Mohamed Helmy, Hosam Elhalis, Md Mamunur Rashid, Kumar Selvarajoo

Declaration of Interest

The authors declare no conflict of interest.

Author credit

Mohamed Helmy: Data Curation, Writing, Visualization, Hosam Elhalis and Md Mamunur Rashid: Writing, Kumar Selvarajoo: Conceptualization, Writing, Reviewing and Editing

Figure 1

Click here to access/download;Figure;Helmy-Elhalis-etal-Figure 1.pdf

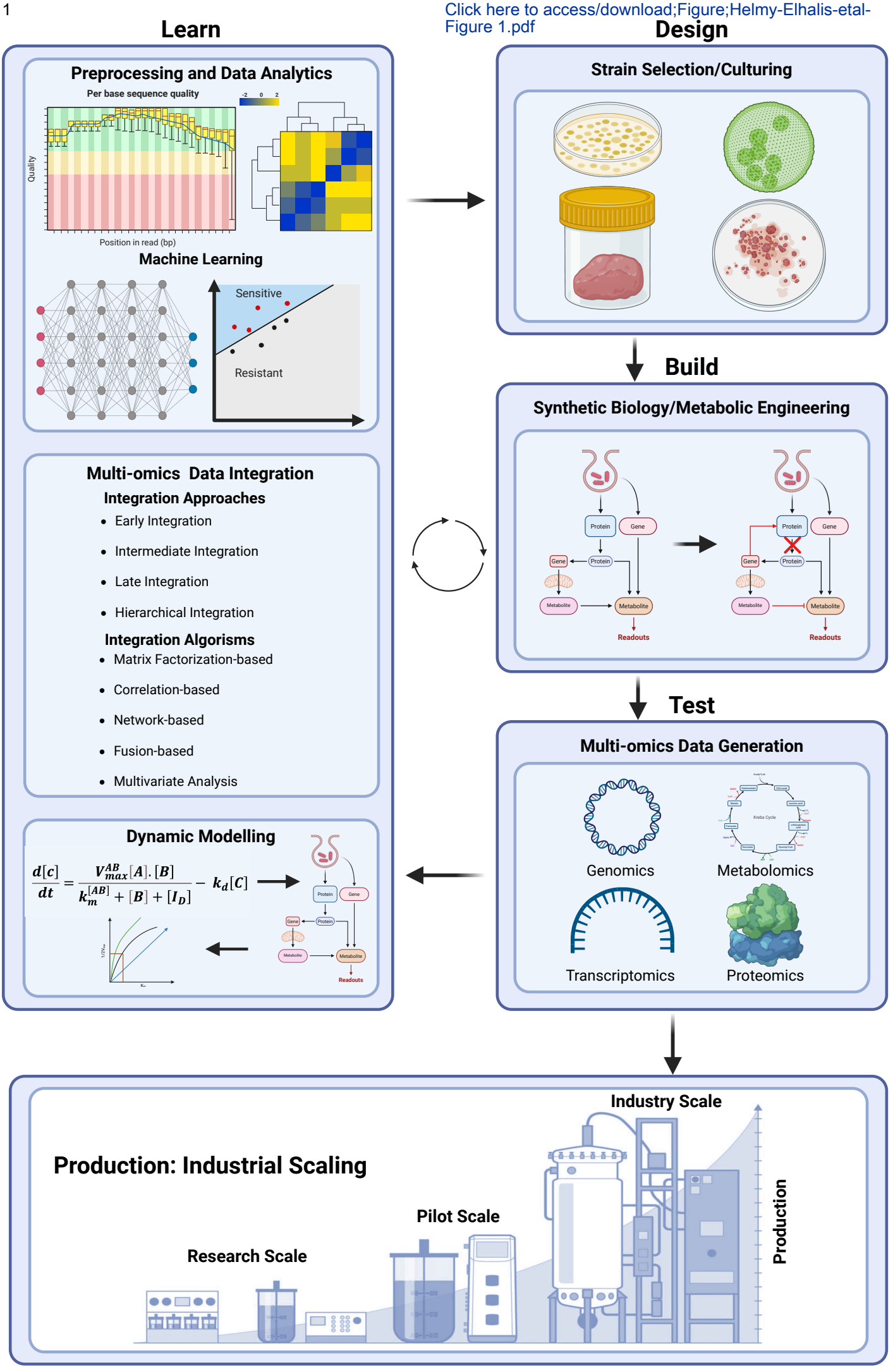

Figure 1

Figure 2

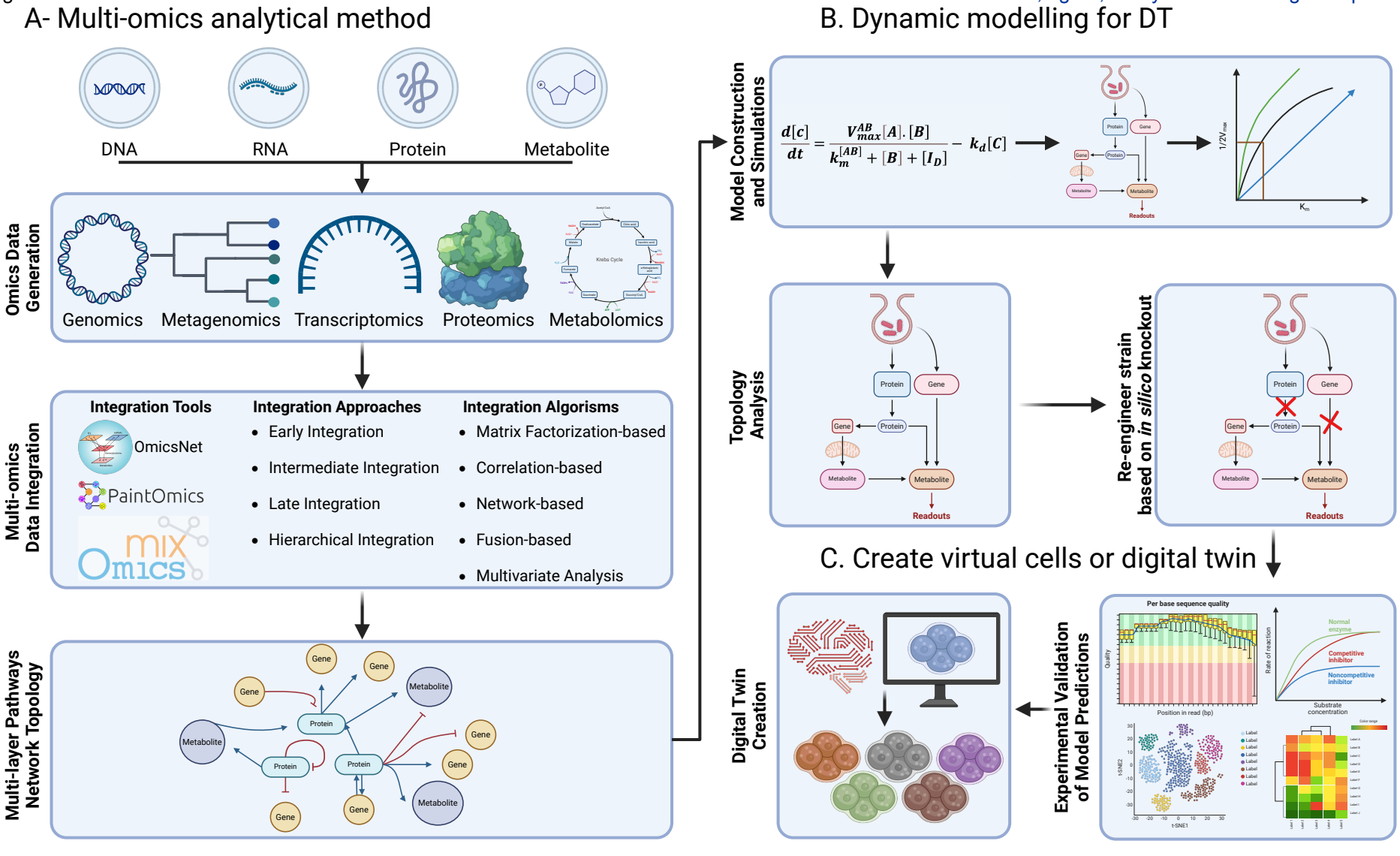

Figure 2

Can Digital Twin Efforts Shape Microorganism-based Alternative Food?

Mohamed Helmy[1,2,3,4], Hosam Elhalis[5], Md Mamunur Rashid[4], Kumar Selvarajoo[4,6,7,#]

[1]Vaccine and Infectious Disease Organization (VIDO), University of Saskatchewan, SK, Canada

[2]Department of Computer Science, Lakehead University, ON, Canada

[3]Department of Computer Science, College of Science and Engineering, Idaho State University, ID, USA

[4]Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore 138671, Singapore

[5]Research School of Biology, Australian National University, Canberra, Australia

[6]Synthetic Biology Translational Research Program and SynCTI, Yong Loo Lin School of Medicine, National University of Singapore (NUS), Singapore 117456, Singapore

[7]School of Biological Sciences, Nanyang Technological University (NTU), Singapore 637551, Singapore

[#]Corresponding Author:
Kumar Selvarajoo
Bioinformatics Institute (BII),
Agency for Science, Technology and Research (A*STAR),
Singapore 138671, Singapore
kumar_selvarajoo@bii.a-star.edu.sg

Abstract

With the continuous increment in global population growth, compounded by post-pandemic food security challenges due to labor shortages, effects of climate change, political conflicts, limited land for agriculture, and carbon emissions control, addressing food production in a sustainable manner for future generations is critical. Microorganisms are potential alternative food sources that can help close the gap in food production. For the development of more efficient and yield-enhancing products, it is necessary to have a better understanding on the underlying regulatory molecular pathways of microbial growth. Nevertheless, as microbes are regulated at multi-omics scales, current research focusing on single omics (genomics, proteomics or metabolomics) independently is inadequate for optimizing growth and product output. Here, we discuss digital twin (DT) approaches that integrate systems biology and artificial intelligence (AI) in analyzing multi-omics datasets, to yield a microbial replica model for *in silico* testing before production. DT models can thus provide a holistic understanding of microbial growth, metabolite biosynthesis mechanisms, as well as identifying crucial production bottlenecks. Our argument, therefore, is to support the development of novel DT models that can potentially revolutionize microorganism-based alternative food production efficiency.

Highlights:

1. Microorganism-based alternative food should be considered as a source for proteins and metabolites as it holds great potential for addressing future food security challenges.
2. Research in this field faces significant challenges due to the highly reductionist nature of the conventional research methodologies, raising the need to develop more innovative approaches.
3. Recent advancements in multi-omics data generation offer novel opportunities to identify and optimize essential biosynthesis pathways or networks for desired food proteins or metabolic products.
4. The integration of data science, machine learning, and systems biology through Digital Twins modeling has potential to revolutionize sustainable alternative food sources by comprehensively identifying and addressing critical biosynthesis bottlenecks.

Introduction

Food security has become a major discussion topic in recent years due to the exponential population growth, drastic climatic changes, and current geopolitical conflicts wars, which collectively disrupt the overall world food production and distribution [1]. Moreover, traditional farming techniques are being criticized for emitting large amounts of harmful greenhouse gasses [2]. It is, therefore, important to explore alternative methods for future food supplementation. Microbial products, derived from bacteria, fungi, and microalgae, can provide promising alternatives to traditional food sources. Microorganisms can produce a wide range of high-value ingredients besides being an excellent source of nutritive proteins [3,4]. Therefore, it is important to explore microorganisms for future food supplementation.

Bacteria, such as *Methylophilus methylotrophus*, *Rhodopseudomonas palustris*, and *Haloarcula* sp., produce 50-80% protein (dry cell weight), and are characterized by small cell sizes and high multiplication [5]. Yeasts like *Saccharomyces cerevisiae* and *Candida tropicalis* have superior nutritional quality and can grow at an acidic pH level, making them excellent sources of protein [3]. Filamentous fungi, such as *Aspergillus niger* and *Fusarium venenatum*, contain up to 63% protein [6]. Microalgae, like *Chlorella sorokiniana* and *Arthrospira platensis* (spirulina), can generate protein levels up to 70% of cell biomass and produce yields 20 to 50 times higher than soybeans [7].

Similarly, microorganisms have already been widely used in the manufacture of natural food ingredients and additives. Among these are food colorants derived from *Monascus* species filamentous fungi, which have been shown to possess antimicrobial and antioxidant properties [8]. Vitamins, terpenoids, steroids, amino acids, lactic acid, functional proteins (texturants),

oligosaccharides, sweeteners, flavors and enzymes are other examples of food ingredients produced successfully by metabolic engineering microorganisms [9].

The extraordinary achievements in molecular biology, biochemistry, real-time monitoring, and data management over the past several decades have led to the widespread usage of microorganisms. However, microbial bioprocesses face challenges from the social and marker perspective such as regulatory issues, safety concerns, sensory attributes, consumer perceptions, and social acceptance. Additionally, limitations such as production optimization and costs are impacting the future of alternative food as a viable option [10]. To overcome these barriers and promote microbial manufacturing processes, further research and innovations are required [11] [12].

Here we discuss Digital Twin (DT) modeling to address some of the key challenges. A DT model is generally defined as a virtual model of a process, product, or service that bridges the physical and digital (*in silico*) worlds in real time [13]. DT is becoming increasingly popular in a variety of industries, including medicine, manufacturing, engineering, and aerospace, and playing a vital role in their current revolutions. The DT modeling interest arises due to recent advances in the rapid collection, storage and sharing of data, and the development of computers that can use complex models and algorithms in a reasonable timeframe [29,30]. For example, it has been applied to integrate numerous clinical and molecular multi-omic datasets, and to predict outcomes for pancreatic cancer patients and disease survival [14]. Here, we discuss how DT can be used to integrate multi-omics datasets with machine learning analytics to understand microbial behavior and improve bioprocesses, and to develop an interactive platform between a physical system and its digital replica. As a result, we hope for more accurate predictions and the resultant platform can be adopted in a more sustainable way for microbial manufacturing.

Multi-Omics Data Generation and its Analytics

An indispensable way to deeply investigate an organism's biology and metabolism is the "omics' approach. Today, we can generate large scale omics (e.g., genomics, transcriptomics, proteomics, metabolomics) datasets with ease and within a reasonable cost. This can push the idea of finding alternative proteins to the forefront as it is possible to obtain detailed information relating to the organism's genotype, biosynthetic and metabolic capacities, as well as its potential for increasing protein productivity [15]. By using next-generation sequencing technologies, the study of genome and genetic changes within multiple species becomes affordable, while proteomics focuses on proteins involved in the maintenance of cell structure, organization, and metabolic functions. Metabolomics focuses on metabolites and their metabolic pathways.

The omics analyses provide insight into the phenotypes of microorganisms. Genome and RNA sequencing (DNA-Seq and RNA-Seq) and mass spectrometry (MS) are among the most widely used analytical technologies employed today to study genomics, transcriptomics, proteomics and metabolomics [16]. The generated data can correlate to the sensory, nutritional, and safety of alternative proteins and microbial-based food ingredients, for example, increasing nutritional quality of fermented soybean pastes [17], and assessing the safety or identification of toxic compounds [18]. Even though omics data can be generated swiftly and reliably, if these data are not analyzed systemically and rigorously, the outcome can still be suboptimal.

Although single omics approaches have led to a better understanding of complex biological processes, cells are regulated at a multi-omics level. Thus, a key challenge now is in analyzing these data collectively, rather than individually, to explore the complex signals that are encoded across multiple modalities. In recent years, the integration of multi-omics data has become integral

to the field of food science, representing the most effective approach to gain a deep and holistic understanding of complex traits, molecular interactions, and robust target molecules. To date, multi-omics methods have provided systems-level biological insights, into nutritional markers [19], molecular processes involved in food intake and deficiencies [20], gut/diet-health relationships [21], clustering of samples [22], and the roles of microorganisms in fermented foods [23]. Multi-omics data integration strategies can be classified into four distinct categories, based on the applied mathematical algorithms: early, intermediate, late, and hierarchical integration [24,25]. Nevertheless, multi-omics analytics or machine learning does not provide mechanistic and dynamic understanding of biosynthesis pathways. For example, if the intention is to increase a metabolic output, we need to know how the metabolic fluxes are distributed over time and whether their regulatory bottlenecks prevent optimum output.

Metabolic engineering and design–build–test–learn (DBTL) cycles

For increasing the metabolic flux toward a product of interest, the metabolic engineering field adopts genetic interventions of any known rate-limiting steps or bottlenecks within the metabolic pathways [26–29]. Although this field has shown success in numerous metabolic applications, fundamental limitations do exist. This is due to a significant lack of knowledge regarding the regulation of biosynthesis networks and cellular physiology. To overcome this, metabolic engineers have developed iterative design–build–test–learn (DBTL) cycles for microbial strain optimization (Figure 1). These cycles aim to develop a product strain iteratively, each time incorporating changes from the previous learning cycle. Systems biology, especially the dynamic modeling of cellular networks using differential equations [1], and machine learning methods allow to learn from data and propose new strain designs for the next DBTL cycle.

After proposing the best strain, these need to be tested for bioreactor fermentation processes, from small lab scales of 1–2L to large industrial >1000L bioreactors. As bioreactor fermentation is a complex heterogenous process, several dynamic modelling approaches are used [1]. However, the bioprocess model mostly considers only the extracellular conditions and growth rates for optimization of batch, fed batch or continuous processes [30,31]. The overall design and optimization to improve titer/yield/rate (TYR) of bioprocessing of desired microbial products is essential. Today DBTL is widely used in labs and industries, nevertheless, evaluating its overall effectiveness remains a challenge. As mentioned above, the recent generation of multi-omics datasets warrants a new strategy to improve the current DBTL cycles.

DT Modeling

A range of measuring technologies are first required to generate key data for developing the digital counterpart of the real-world entity [32]. For biology or microbial applications, these can be time-series multi-omics and even imaging datasets (Figure 1). Next, multi-omics machine learning analysis can be performed to elucidate the key genes, proteins or metabolites expressed and their related pathways (Figure 2). With this information, DT models are required to process the complex data and make a virtual representation for *in silico* testing. The models are based on known governing equations, such as biochemical rate laws, or machine learning models to create a digital representation of the physical entity. In this way, it will be likely for combinatorial pathway optimization considering multi-omics regulation.

In an increasingly interconnected, multi-disciplinary and data-driven world, DT approaches, thus, offer a powerful method for innovation and problem-solving. These virtual models are used for in-depth analysis, simulation, and optimization of the physical counterpart,

providing valuable insights and facilitating decision-making in various industries, mostly in manufacturing and infrastructure management [33], although a few attempts have been made for healthcare, agriculture and food research. DT has been applied to production systems to optimize the planning and commissioning of the manufacturing process. It helps to decrease downtime by virtually testing a system before running it and verifying the entire operation in advance [34]. This can be adopted for alternative food research, since having this technology allows in-depth analysis, simulation, prediction and optimization of the *in silico* model before actual experimentation of the physical entity [35] (Figure 1).

For engineering organisms, the models mainly encompass two broad approaches: dynamic modeling and constraint-based modeling [36]. Dynamic models are constructed using differential equations leveraging known biochemical reactions and their kinetics to predict metabolic outcomes, such as to understand key regulatory mechanisms and pathway bottleneck enzymes or reactions. Alternatively, constraint-based models, such as Flux Balance Analysis (FBA), set constraints for reactions and are well-suited for large-scale applications where kinetic laws and parameter values are largely unknown. FBA could efficiently model thousands of metabolites and reactions, making it invaluable for gene contribution analysis and pathway design [37]. These models need to be innovatively extended to incorporate information of gene and protein expressions/activities affecting the metabolic pathways.

There are also other approaches such as transcriptional control and ensemble modeling which can be adopted for DT modeling. Transcriptional control involves modifying gene regulation by manipulating promoter regions, necessitating a deep understanding of regulatory elements [38,39]. Ensemble modeling integrates multiple models or training sets to predict overall

pathway outcomes. This approach is particularly useful when detailed kinetic parameters are lacking [40].

Different modeling strategies can develop DTs for microbial metabolic engineering to improve alternative food production in diverse biotechnological applications [41]. Moreover, they provide a platform for investigating complex systems *in silico*, leading to a better understanding of the system's dynamics and the relationships between its components. Although metabolic engineering offers potential to produce alternative food, and ingredients, many technical challenges remain, including low genetic engineering efficiency and unsafe heterologous genetic components, as well as the high cost of medium components [42]. To overcome these challenges, DT can be used to pre-evaluate proposed genetic engineering tools and methods virtually before testing in the laboratory. It can also be employed to visualize the organism's behavior and metabolites produced under different carbon sources and environmental parameters aiming to select the cheapest and most efficient medium components and incubation conditionals.

By employing sensor technologies with other mathematical and predictive models, DT can bring a real-time virtual view of the alternative food manufacturing process based on their metabolic activities, growth behavior, and incubation conditions. That might help to select the best microorganisms for alternative food production and design novel proteins with improved nutritional value, safety, taste, and texture. DT can also utilize the data collected over time to provide accurate virtual representations of physical objects, including highly dynamic and complex proteins and microbial metabolites from non-conventional sources, concerning their sensory, safety, and nutritional qualities.

DT Application in Food Production and Research

DT is implemented in food production to address challenges such as quality, safety, complexity, cost, losses, and waste [43]. For instance, it is possible to monitor an alternative food manufacturing bioprocess in real-time using DT, which reduces the risk of system failures and optimizes the use of materials and human resources. A recent demonstration showed that sensors and Internet of Things (IoT) technology could be used in DT to collect real-time food supply chain data and facilitate informed decision-making without interfering with actual operations [44].

The application of DT to food production can also provide valuable predictions in normal and unusual circumstances. As an example, DT has been employed to estimate the remaining shelf life of finished products with real-time monitoring and virtual operations such as maintenance [45]. In addition, accurate predictions were obtained using DT after conducting scenario analysis and assessing the impact of possible disruptions during the manufacturing process. Virtual simulations can extend the application of DT beyond evaluating food quality within the existing food system to forecasting the effects of certain future changes [46].

A potential benefit of DT in microbial derived food production is the ability to integrate complex data from a variety of sources, including multi-omics and data generated during microbial growth optimization, scale up, safety and quality assessments. As a result, it may be possible to identify the key regulatory bottleneck or emerging bottlenecks that limit their potential industrial applications. DT can be used to integrate data from multiple sources, facilitating accurate simulations, assessments, and optimizations of layout design and operational processes [47]. Finally, by reducing the costs associated with testing, simulating scenarios, and improving quality, DT enables companies to cutdown costs and improve productivity and quality. Specifically, testing and validating new manufacturing processes and equipment can be accomplished through DT [48].

Meanwhile, operator training can be conducted prior to implementation in the real world, which enhances the efficiency of production through a reduction in human error [49].

The overall benefit of DT approach can leverage off of the recent advancements in omics technologies, genome engineering, synthetic biology, and high-throughput screening of optimal microbial strains for further improvement (Table 1). The DT approach could likely show the optimum biological network configuration for increasing the TYR values of products. This can be extended up to bioprocess modeling scale to offer smart manufacturing processes, as well as a reduction in cost and time, enable quick corrections to be made, optimize and accelerate the production bioprocess, and allow precise and cost-effective decisions to be made.

Challenges of DT

Every *in-silico* approach will have its limitations. Dynamic modeling relies heavily on biochemical parameters, such as reaction kinetics and flux ranges, which can be challenging to ascertain [36,50]. Bottom-up modeling, dependent on experimental data for parameter determination, is hindered by the extensive labor and cost involved in gathering kinetic information for all enzymes in a pathway or network. Moreover, the significant disparities between *in vitro* and *in vivo* experimental data present additional hurdles. The iterative nature of data collection further extends the timeframe required for modeling. Furthermore, as the model scale expands to cover larger networks involved, so does the need for more reliable and reproducible experimental data, leading to heightened expenses and a reduction in accuracy of predictions [51,52].

Top-down approaches leverage time-series metabolomic data to indirectly infer kinetics, flux rates, or metabolite concentrations through causation and correlation networks [51]. Causation networks establish cause-effect relationships between metabolites, while correlation networks employ mathematical and statistical techniques to determine likely connections between enzymes and metabolites [53]. Despite the use of optimization algorithms to estimate model parameters, the intricate, often nonlinear relationships within metabolic models and the variability of parameters hamper the accuracy of these fitting algorithms. Nevertheless, top-down approaches have found success in analyzing simple linear response pathways or mass-action kinetic models with low parameter sensitivity [54,55].

In most modeling strategies, the reliance on high-quality experimental data presents a significant hurdle. Access to diverse data types, including metabolite concentrations, genomic sequences, and gene expression data, is necessary, and while many bioinformatics resources aggregate this data, the challenge lies in identifying the right datasets and analytical approaches. This underscores the growing importance of novel data mining and analytics methods, including artificial intelligence, to harness the wealth of available data effectively [56].

The integration of AI or machine learning (ML) into DT holds immense potential in addressing various challenges within the field. ML, a subset of artificial intelligence, empowers computers to autonomously analyze data based on predefined rules or pattern recognition models. While AI has made substantial strides in different fields, it has yet to realize its full potential in alternative food research [57]. Identifying metabolic pathways, especially when the pathway is poorly understood or the gene responsible needs to be transferred to a model organism for manipulation, is a critical step [58]. ML algorithms plays a pivotal role in the identification of essential enzymes and genes within metabolic pathways. Classifiers like support vector machines,

logistic regression and decision tree-based models have been instrumental in predicting gene essentiality, aiding in the selection of potential drug targets and drug side effect analysis. ML leverages network topology, gene homologies, and gene expression data to make predictions, and its superior accuracy compared with traditional mathematical models has been experimentally validated [59].

Furthermore, the increasing volume of -omics data has necessitated data-driven approaches, where ML methods perfectly fit. Integrating ML with omics data allows for improved predictions and qualitative insights, outperforming traditional kinetic models. As ML continues to evolve, it presents a promising frontier for advancing DT for alternative food research, offering innovative solutions to the challenges of single cell proteins (SCPs, protein-rich microbial products) production and beyond [1,57,60].

Future Projections

Integrating omics data generation, data analytics, machine learning and systems biology into DT model offers a very promising path for advancing alternative food research that leads to healthier, affordable, and nutritious food sources. However, a major hurdle is the scarcity of high-throughput multi-omics data, especially for microbial strains related to alternative food production, such as SCPs, crucial for holistic organism analysis and pathway discovery. Addressing this requires the generation of comprehensive omics data and the development of online resources, such as meta-databases, for community access. Furthermore, standardized data design is crucial for these resources to enhance their utility [61].

Training machine learning models in metabolic engineering poses another data problem, demanding quantitative data for multiple conditions. Scaling up experiments from lab to industrial

settings introduces significant differences, necessitating specialized data. While databases like LASER and jQMM exist, they often lack the data necessary for AI modeling in DT [62,63]. This shortage calls for an emphasis on producing high-quality quantitative data, fostering collaboration, and data-sharing efforts in the field.

The "black box problem" inherent in AI and machine learning methods raises concerns about the transparency and interpretability of the models, especially in complex biological systems [64]. Addressing this, explainable artificial intelligence (XAI) methods hold promise for making AI outcomes more understandable to humans, facilitating model improvements.

Genome annotation for food-safe and GRAS organisms also demands specific attention, as many remain under-studied or unannotated. Improved ML-based genome annotation and pathway prediction methods are essential for harnessing the full potential of AI in wild and metabolic engineering, even in the absence of genome sequences [36]. Overcoming these challenges will be crucial as we explore the DT efforts to enhance metabolic engineering strategies for alternative food production.

There are several challenges that occur during the production of microbial-based foods and ingredients that are correlated to the metabolic activities of microorganisms and impact yields, sensory, and safety of the final products, as stated above. To address this complexity, the next logical step is to forecast how these components will interact with one another. This would undoubtedly be an intriguing topic for future DT research. Instead of looking at a single prototype, we may eventually have access to a visualization that demonstrates how these diversity interactions occur in greater depth.

Traceability and identification of critical control points (CCPs) are two other aspects that contribute to the food system, including alternative foods and ingredients [65]. With the help of

DT's virtual presentation of the biomanufacturing process in future studies, it is possible to detect CCPs and improve the efficiency of the traceability system. It can also be adapted to effectively predict CCP failure and its corrective actions, thereby improving the safety of alternative food generated products in a more sustainable way.

Future research should also consider waste management systems, how they can be visualized by DT, allowing users to monitor, quickly analyze, and fix problems before they occur. This saves both time and money by decreasing downtime and avoiding costly failures. The collected data overtime by DT can be used to optimize waste management processes and identify opportunities for improvement. Overall, it is an imperative to support the research and development of multi-disciplinary based DT models that have potential for sustainable alternative food production in the foreseeable future.

Conclusion

The integration of Digital Twin (DT) modeling with sensor technologies and predictive models offers a transformative approach to address challenges in food production and research. By leveraging real-time data collection and virtual simulations, DT enables continuous monitoring and optimization of alternative food manufacturing processes, reducing the risk of failures and improving resource utilization. Moreover, DT provides valuable insights into food quality, safety, and shelf-life estimation, facilitating informed decision-making and scenario analysis.

In microbial-derived food production, DT's ability to integrate complex data from various sources, including multi-omics and growth optimization, enhances our understanding of key regulatory bottlenecks and facilitates process optimization. By reducing testing costs and improving productivity and quality, DT accelerates innovation in food manufacturing while

enabling efficient operator training and equipment validation. Furthermore, DT's application extends to bioprocess modeling, offering smart manufacturing processes and enabling quick corrections and precise decision-making in alternative food production.

Despite its potential, DT encounters several challenges that must be addressed to fully leverage its benefits. Dynamic modeling approaches rely heavily on accurate parameters, presenting challenges in parameter determination and experimental data collection. Similarly, top-down modeling approaches face difficulties in accurately inferring kinetics and flux rates due to the complexity and variability of metabolic models. However, the integration of AI and machine learning (ML) into DT offers opportunities to overcome these challenges. ML algorithms play a crucial role in pathway identification, gene essentiality prediction, and omics data analysis, improving the accuracy and efficiency of metabolic engineering strategies. Moreover, explainable AI methods hold promise for enhancing the transparency and interpretability of ML models, addressing concerns about their applicability in complex biological systems.

Looking ahead, we suggest that overcoming challenges related to data scarcity, scalability, and interpretability is essential for advancing DT in alternative food research. Generating comprehensive multi-omics data, developing standardized data resources, and fostering collaboration are critical steps in this direction. Additionally, improving genome annotation and pathway prediction methods for food-safe organisms will unlock the full potential of AI in metabolic engineering. Furthermore, future research should explore the potential of DT in addressing complexities related to microbial-based food production, including interactions between components and waste management. By visualizing biomanufacturing processes, identifying critical control points, and optimizing waste management systems, DT can contribute to improving the efficiency, safety, and sustainability of alternative food production.

Overall, the adoption of DT approaches promises to leverage recent advancements in omics technologies, genome engineering, and synthetic biology to optimize bioprocesses, reduce costs, and enhance the competitiveness of alternative foods and ingredients. With its potential to enhance productivity, quality, and sustainability in food production, DT represents a significant step forward in addressing the challenges faced by the food industry in meeting the growing demand for nutritious and sustainable food sources.

Acknowledgement

The authors thank Derek Smith for critical reading of the manuscript. This work was funded by the Bioinformatics Institute, Agency for Science, Technology and Research, Singapore, and the Vaccine and Infectious Disease Organization, University of Saskatchewan, Canada.

Tables

Table 1: Selected cases of microbial manufacturing process optimization using omics analyses.

Microorganisms	Observations	Omics analyses	Achievements	References
Yarrowia lipolytica	Low lipid production	Metabolomics	Glycerol led to ↑ yield (promote long chain fatty acids synthesis)	[66]
Saccharomyces cerevisiae	Low levels of carotenoids	Transcriptomics	Zinc and copper ions ↑ yield (regulate related genes)	[67]
Streptomyces albulus	pH impact on ε-poly-L-lysine	Transcriptomics	pH shock ↑ yield (prompt key genes and enhance cell respiration)	[68]
Aurantiochytrium Sp.	Temperature effect on generated lipid types	Transcriptomics and lipidomics	↑ Docosahexaenoic and eicosadienoic acid acids at 5ºC and 15ºC, respectively (regulate activity of key enzymes)	[69]
Rhodosporidium toruloides	Oil accumulation during starvation	Genomic, transcriptomic proteomics	Provided different substance to ↑ lipid production under nutritionally limiting medium	[70,71]
Penicillium chrysogenum	Low performance when scale up	Integrating computational fuid dynamics and cellular reaction dynamics	Design a rational strategy for the scale-up process	[72]

Figure Legend

Figure 1. The Design, Build, Test, Learn Iterative Cycle of Metabolic Engineering. The iterative process of metabolic engineering, encompassing four key stages: The Design stage where the problem is identified, and the pathway along with the host organism are selected. The Build stage involves the selection, synthesis, and assembly of components for integration into the host. The Test stage where engineered strains are scrutinized to produce target molecules, such as transcripts, proteins, and metabolites. Finally, the Learn stage entails analyzing the data gathered from testing and utilizing it to inform future iterations of the cycle.

Figure 2. Workflow for multi-omics data integration, next-generation dynamic model and digital twin development. The process involves collecting raw individual omics data (e.g., genomics, metagenomics, proteomics, and metabolomics) for specific phenotypes, which is followed by a proper quality control process to remove less informative features. Next, the researcher chooses one of the integration strategies from the listed options: the early integration method, merging all omics datasets to create a single large matrix of multi-omics data; the intermediate integration method allows the analysis of multi-omics datasets together, focusing on finding common latent states that can reveal the underlying biological mechanisms; Late integration analyzes each data set separately and later integrates its result into a joint model, and Hierarchical integration leveraging prior knowledge of the regulatory relationships between the different omics layers to extract the most relevant information. Various supervised or unsupervised machine learning methods are subsequently employed to analyze and select the most distinctive features for each phenotype of interest and map them into biological pathways. From an empirical standpoint, the most current data integration algorithms can distinguish into six categories: matrix factorization-based, correlation-based, network-based, Bayesian, fusion-based, and multivariate analysis [73,74]. The use of one or more of these algorithms depends upon the specific research objectives and their relevance to the different biological challenges addressed by multi-omics studies. Indeed, these algorithms have been incorporated into various tools, including web-based platforms such as OmicsNet, PaintOmics, Asterics and others, and development environments like mixOmics, as outlined by [75]. The resultant data is used for the construction of a dynamics model using the multi-omics pathways network topology (considered as input) obtained from the previous steps. Next, the metabolic network topology is adjusted to fit the time-series multi-omics data of molecular profiles and *in silico* knockouts are used to increase the yield of desired proteins and to guide the re-engineering of the microbial strain. Finally, the predictions of the dynamic models are validated experimentally, and a digital twin is created to be used to achieve better understanding of the processes involved in alternative food production.

References

1. *Smith DJ, Helmy M, Lindley ND, Selvarajoo K: **The transformation of our food system using cellular agriculture: What lies ahead and who will lead it?** *Trends Food Sci Technol* 2022, **127**:368–376.
   This article highlights the recent progress in the area of alternative protein, cultured meat and nutrition and provide a comprehensive assessment of the global situation.
2. Holka M, Kowalska J, Jakubowska M: **Reducing Carbon Footprint of Agriculture—Can Organic Farming Help to Mitigate Climate Change?** *Agriculture 2022, Vol 12, Page 1383* 2022, **12**:1383.
3. Jach ME, Serefko A, Ziaja M, Kieliszek M: **Yeast Protein as an Easily Accessible Food Source**. *Metabolites* 2022, **12**.
4. Nasseri AT, Rasoul-Amini S, Morowvat MH, Ghasemi Y: **Single cell protein: Production and process**. *Am J Food Technol* 2011, **6**:103–116.
5. Higgins C, Muralidhara B, Wittung-Stafshede P: **How Do Cofactors Modulate Protein Folding?** *Protein Pept Lett* 2005, **12**:165–170.
6. Nasseri AT, Rasoul-Amini S, Morowvat MH, Ghasemi Y: **Single cell protein: Production and process**. *Am J Food Technol* 2011, **6**:103–116.
7. Chronakis IS, Madsen M: **Algal proteins**. *Handbook of Food Proteins* 2011, doi:10.1533/9780857093639.353.
8. Vendruscolo F, Tosin I, Giachini AJ, Schmidell W, Ninow JL: **Antimicrobial Activity of Monascus Pigments Produced in Submerged Fermentation**. *J Food Process Preserv* 2014, **38**:1860–1865.
9. Hanlon P, Sewalt V: **GEMs: genetically engineered microorganisms and the regulatory oversight of their uses in modern food production**. *Crit Rev Food Sci Nutr* 2021, **61**:959–970.
10. Das D, Pandit S: **Industrial Biotechnology**. 2021, doi:10.1201/9780367822415.
11. Aidoo R, Kwofie EM, Adewale P, Lam E, Ngadi M: **Overview of single cell protein: Production pathway, sustainability outlook, and digital twin potentials**. *Trends Food Sci Technol* 2023, **138**:577–598.
12. Sirohi R, Joun J, Choi HI, Gaur VK, Sim SJ: **Algal glycobiotechnology: omics approaches for strain improvement**. *Microb Cell Fact* 2021, **20**:1–10.
13. Sharma A, Kosasih E, Zhang J, Brintrup A, Calinescu A: **Digital Twins: State of the art theory and practice, challenges, and open research questions**. *J Ind Inf Integr* 2022, **30**:100383.
14. Osipov A, Nikolic O, Gertych A, Parker S, Hendifar A, Singh P, Filippova D, Dagliyan G, Ferrone CR, Zheng L, et al.: **The Molecular Twin artificial-intelligence platform integrates multi-omic data to predict outcomes for pancreatic adenocarcinoma patients**. *Nat Cancer* 2024, doi:10.1038/S43018-023-00697-7.
15. Guarnieri MT, Pienkos PT: **Algal omics: Unlocking bioproduct diversity in algae cell factories**. *Photosynth Res* 2015, **123**:255–263.
16. Guarnieri MT, Nag A, Yang S, Pienkos PT: **Proteomic analysis of Chlorella vulgaris: Potential targets for enhanced lipid accumulation**. *J Proteomics* 2013, **93**:245–253.

17. Sun X, Lyu G, Luan Y, Yang H, Zhao Z: **Metabolomic study of the soybean pastes fermented by the single species Penicillium glabrum GQ1-3 and Aspergillus oryzae HGPA20**. *Food Chem* 2019, **295**:622–629.
18. Gao Y, Hou L, Gao J, Li D, Tian Z, Fan B, Wang F, Li S: **Metabolomics Approaches for the Comprehensive Evaluation of Fermented Foods: A Review**. *Foods 2021, Vol 10, Page 2294* 2021, **10**:2294.
19. Afshari R, Pillidge CJ, Dias DA, Osborn AM, Gill H: **Biomarkers associated with cheese quality uncovered by integrative multi-omic analysis**. *Food Control* 2021, **123**:107752.
20. Sundekilde UK, Yde CC, Honore AH, Caverly Rae JM, Burns FR, Mukerji P, Mawn MP, Stenman L, Dragan Y, Glover K, et al.: **An Integrated Multi-Omics Analysis Defines Key Pathway Alterations in a Diet-Induced Obesity Mouse Model**. *Metabolites* 2020, **10**.
21. Belda I, Cueva C, Tamargo A, Ravarani CN, Acedo A, Bartolomé B, Moreno-Arribas MV: **A multi-omics approach for understanding the effects of moderate wine consumption on human intestinal health**. *Food Funct* 2021, **12**:4152–4164.
22. Dong TS, Mayer EA, Osadchiy V, Chang C, Katzka W, Lagishetty V, Gonzalez K, Kalani A, Stains J, Jacobs JP, et al.: **A Distinct Brain-Gut-Microbiome Profile Exists for Females with Obesity and Food Addiction**. *Obesity (Silver Spring)* 2020, **28**:1477.
23. Shi H, An F, Lin H, Li M, Wu J, Wu R: **Advances in fermented foods revealed by multi-omics: A new direction toward precisely clarifying the roles of microorganisms**. *Front Microbiol* 2022, **13**.
24. *Picard M, Scott-Boyer MP, Bodein A, Périn O, Droit A: **Integration strategies of multi-omics data for machine learning analysis**. *Comput Struct Biotechnol J* 2021, **19**:3735–3746.
    This mini review focuses on challenges and existing multi-omics integration strategies and pay special attention to machine learning applications.
25. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E: **Using machine learning approaches for multi-omics data analysis: A review**. *Biotechnol Adv* 2021, **49**.
26. Kim GB, Kim WJ, Kim HU, Lee SY: **Machine learning applications in systems metabolic engineering**. *Curr Opin Biotechnol* 2020, **64**:1–9.
27. Choi KR, Jang WD, Yang D, Cho JS, Park D, Lee SY: **Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering**. *Trends Biotechnol* 2019, **37**:817–837.
28. Bailey JE, Sburlati A, Hatzimanikatis V, Lee K, Renner WA, Tsai PS: **Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes**. *Biotechnol Bioeng* 2002, **79**:568–579.
29. Stephanopoulos G, Aristidou AA, Nielsen JHøiriis: **Metabolic engineering : principles and methodologies**. [date unknown],
30. Henriques D, Balsa-Canto E: **The Monod Model Is Insufficient To Explain Biomass Growth in Nitrogen-Limited Yeast Fermentation**. *Appl Environ Microbiol* 2021, **87**:1–22.
31. Rathore AS, Mishra S, Nikita S, Priyanka P: **Bioprocess Control: Current Progress and Future Perspectives**. *Life (Basel)* 2021, **11**.

32. Boschert S, Rosen R: **Digital twin-the simulation aspect**. *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and Their Designers* 2016, doi:10.1007/978-3-319-32156-1_5/FIGURES/4.
33. Yao JF, Yang Y, Wang XC, Zhang XP: **Systematic review of digital twin technology and applications**. *Vis Comput Ind Biomed Art* 2023, **6**.
34. Guerra-Zubiaga D, Kuts V, Mahmood K, Bondar A, Nasajpour-Esfahani N, Otto T: **An approach to develop a digital twin for industry 4.0 systems: manufacturing automation case studies**. *Int J Comput Integr Manuf* 2021, **34**:933–949.
35. Krupitzer C, Noack T, Borsum C: **Digital Food Twins Combining Data Science and Food Science: System Model, Applications, and Challenges**. *Processes 2022, Vol 10, Page 1781* 2022, **10**:1781.
36. *Helmy M, Smith D, Selvarajoo K: **Systems biology approaches integrated with artificial intelligence for optimized metabolic engineering**. Metab Eng Commun 2020, 11:e00149.
The article reviews the latest attempts of combining systems biology and AI in metabolic engineering research and highlight how this alliance can help overcome the current challenges.
37. Bordbar A, Monk JM, King ZA, Palsson BO: **Constraint-based models predict metabolic and associated cellular functions**. *Nature Reviews Genetics 2014 15:2* 2014, **15**:107–120.
38. Shukal S, Chen X, Zhang C, S S, X C, C Z: **Systematic engineering for high-yield production of viridiflorol and amorphadiene in auxotrophic Escherichia coli**. 2019, **55**:170–178.
39. Curran KA, Crook NC, Karim AS, Gupta A, Wagman AM, Alper HS: **Design of synthetic yeast promoters via tuning of nucleosome architecture**. *Nat Commun* 2014, **5**:1–8.
40. Tran LM, Rizk ML, Liao JC: **Ensemble Modeling of Metabolic Networks**. *Biophys J* 2008, **95**:5606–5617.
41. Kotu V, Deshpande B: **Data Mining Process**. *Predictive Analytics and Data Mining* 2015, doi:10.1016/B978-0-12-801460-8.00002-1.
42. Balagurunathan B, Ling H, Choi WJ, Chang MW: **Potential use of microbial engineering in single-cell protein production**. *Curr Opin Biotechnol* 2022, **76**.
43. Huang Y, Ghadge A, Yates N: **Implementation of digital twins in the food supply chain: a review and conceptual framework**. *Int J Prod Res* 2024, doi:10.1080/00207543.2024.2305804.
44. Pan YH, Qu T, Wu NQ, Khalgui M, Huang GQ: **Digital Twin Based Real-time Production Logistics Synchronization System in a Multi-level Computing Architecture**. *J Manuf Syst* 2021, **58**:246–260.
45. Semeraro C, Lezoche M, Panetto H, Dassisti M: **Digital twin paradigm: A systematic literature review**. *Comput Ind* 2021, **130**:103469.
46. **Defraeye T, Shrivastava C, Berry T, Verboven P, Onwude D, Schudel S, Bühlmann A, Cronje P, Rossi RM: **Digital twins are coming: Will we need them in supply chains of fresh horticultural produce?** *Trends Food Sci Technol* 2021, **109**:245–258.

This article describes physics-based and data-driven digital twins, and provides good examples for maximizing shelf life, reduce food losses and predict postharvest food quality evolution.

47. Maheshwari P, Kamble S, Kumar S, Belhadi A, Gupta S: **Digital twin-based warehouse management system: a theoretical toolbox for future research and applications**. *International Journal of Logistics Management* 2023, **ahead-of-print**.
48. Soori M, Arezoo B, Dastres R: **Digital twin for smart manufacturing, A review**. *Sustainable Manufacturing and Service Economics* 2023, **2**:100017.
49. Yin Y, Zheng P, Li C, Wang L: **A state-of-the-art survey on Augmented Reality-assisted Digital Twin for futuristic human-centric industry transformation**. *Robot Comput Integr Manuf* 2023, **81**:102515.
50. *Kim OD, Rocha M, Maia P: **A Review of Dynamic Modeling Approaches and Their Application in Computational Strain Optimization for Metabolic Engineering**. *Front Microbiol* 2018, **9**:1690.
    The article describes approaches developed to undertake issues regarding the mathematical formulation and the underlying optimization algorithms, and that address the phenotype prediction byincluding available kinetic rate laws of metabolic processes.
51. Cuperlovic-Culf M: **Machine Learning Methods for Analysis of Metabolic Data and Metabolic Pathway Modeling**. *Metabolites* 2018, **8**:4.
52. *Selvarajoo K, Tomita M, Tsuchiya M: **Can complex cellular processes be governed by simple linear rules?** *J Bioinform Comput Biol* 2009, **7**:243–268.
    The article highlights that differential equation together with simple linear rules govern the dynamics perturnation-response behavior of biological networks in an ensemble of cells.
53. Srinivasan S, Cluett WR, Mahadevan R: **Constructing kinetic models of metabolism at genome-scales: A review**. *Biotechnol J* 2015, **10**:1345–1359.
54. Helmy M, Gohda J, Inoue J, Tomita M, Tsuchiya M, Selvarajoo K: **Predicting Novel Features of Toll-Like Receptor 3 Signaling in Macrophages**. *PLoS One* 2009, **4**:e4661.
55. Selvarajoo K: **Macroscopic law of conservation revealed in the population dynamics of Toll-like receptor signaling**. *Cell Communication and Signaling* 2011, **9**:1–7.
56. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD: **The MetaCyc database of metabolic pathways and enzymes-a 2019 update**. *Nucleic Acids Res* 2020, **48**:D445--D453.
57. Helmy M, Elhalis H, Liu Y, Chow Y, Selvarajoo K: **Perspective: Multiomics and Machine Learning Help Unleash the Alternative Food Potential of Microalgae**. *Advances in Nutrition* 2023, **14**:1–11.
58. Yeo HC, Selvarajoo K: **Machine learning alternative to systems biology should not solely depend on data**. *Brief Bioinform* 2022, **23**:1–6.
59. Nandi S, Subramanian A, Sarkar RR: **An integrative machine learning strategy for improved prediction of essential genes in Escherichia coli metabolism using flux-coupled features**. *Mol Biosyst* 2017, **13**:1584–1596.

60. Khanijou JK, Kulyk H, Bergès C, Khoo LW, Ng P, Yeo HC, Helmy M, Bellvert F, Chew W, Selvarajoo K: **Metabolomics and modelling approaches for systems metabolic engineering**. *Metab Eng Commun* 2022, **15**:e00209.
61. Helmy M, Crits-Christoph A, Bader GD: **Ten Simple Rules for Developing Public Biological Databases**. *PLoS Comput Biol* 2016, **12**:e1005128.
62. Winkler JD, Halweg-Edwards AL, Gill RT: **The LASER database: Formalizing design rules for metabolic engineering**. *Metab Eng Commun* 2015, **2**:30–38.
63. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al.: **KBase: The United States department of energy systems biology knowledgebase**. *Nat Biotechnol* 2018, **36**:566–569.
64. Zednik C: **Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence**. *Philos Technol* 2019, doi:10.1007/s13347-019-00382-7.
65. Yu Z, Jung D, Park S, Hu Y, Huang K, Rasco BA, Wang S, Ronholm J, Lu X, Chen J: **Smart traceability for food safety**. *Crit Rev Food Sci Nutr* 2022, **62**:905–916.
66. Yun EJ, Lee J, Kim DH, Kim J, Kim S, Jin YS, Kim KH: **Metabolomic elucidation of the effects of media and carbon sources on fatty acid production by Yarrowia lipolytica**. *J Biotechnol* 2018, **272–273**:7–13.
67. Su B, Li A, Deng MR, Zhu H: **Transcriptome Analysis Reveals a Promotion of Carotenoid Production by Copper Ions in Recombinant Saccharomyces cerevisiae**. *Microorganisms* 2021, **9**:1–10.
68. Pan L, Chen X, Wang K, Mao Z: **Understanding high ε-poly-L-lysine production by Streptomyces albulus using pH shock strategy in the level of transcriptomics**. *J Ind Microbiol Biotechnol* 2019, **46**:1781–1792.
69. Song Y, Hu Z, Xiong Z, Li S, Liu W, Tian T, Yang X: **Comparative transcriptomic and lipidomic analyses indicate that cold stress enhanced the production of the long C18-C22 polyunsaturated fatty acids in Aurantiochytrium sp**. *Front Microbiol* 2022, **13**.
70. Zhu Z, Zhang S, Liu H, Shen H, Lin X, Yang F, Zhou YJ, Jin G, Ye M, Zou H, et al.: **A multi-omic map of the lipid-producing yeast Rhodosporidium toruloides**. *Nat Commun* 2012, **3**.
71. Wang Y, Zhang S, Zhu Z, Shen H, Lin X, Jin X, Jiao X, Zhao ZK: **Systems analysis of phosphate-limitation-induced lipid accumulation by the oleaginous yeast Rhodosporidium toruloides**. *Biotechnol Biofuels* 2018, **11**.
72. Tang W, Deshmukh AT, Haringa C, Wang G, van Gulik W, van Winden W, Reuss M, Heijnen JJ, Xia J, Chu J, et al.: **A 9-pool metabolic structured kinetic model describing days to seconds dynamics of growth and product formation by Penicillium chrysogenum**. *Biotechnol Bioeng* 2017, **114**:1733–1743.
73. Das T, Andrieux G, Ahmed M, Chakraborty S: **Integration of Online Omics-Data Resources for Cancer Research**. *Front Genet* 2020, **11**.
74. Subramanian I, Verma S, Kumar S, Jere A, Anamika K: **Multi-omics Data Integration, Interpretation, and Its Application**. *Bioinform Biol Insights* 2020, **14**.
75. **Herráiz-Gil S, del Carmen de Arriba M, Escámez MJ, León C: **Multi-omic data integration in food science and analysis**. *Curr Opin Food Sci* 2023, **52**:101049.

This review provides an up-to-date critical overview of some of the multi-omic approaches, advantages, and limitations.

Point-to-point response letter.

C1:Please discuss how DT can be applied to food through specific examples and analyze its mechanisms in detail.
R1: We added a new section entitled "DT Application in Food Production and Research" where we discussed several applications in detail with examples.

C2: It is suggested that the author should discuss the conclusion of the review in depth and put forward his own views rather than just a superficial summary.
R2: We expended the conclusion. For our forward view, we have a section entitled "Future Projections" where we discuss our forward view. Also, we added this to the new extended conclusion.

C3: Some grammatical errors and sentence structure issues still need to be addressed.
R3: An independent colleague , Dr Derek Smith, an native Englishman has read, checked and corrected all language issues. His name is now acknowledged in the manuscript.

The authors thank the reviewer for his/her comments and suggestions.