

Biological Databases and Resources: Their Engineering and Applications in Synthetic Biology

Divya Sindhu*, Saurabh Sindhu

Department of Computer Science,
CRM Jat College, Hisar - 125001, Haryana, India

REVIEW ARTICLE

ABSTRACT: Biological sciences pose a unique set of engineering challenges due to incomplete understanding of natural biological systems. Currently, sequencing of macromolecules i.e., DNA (deoxyribose nucleic acid) and proteins obtained from living cells has provided significant information, which is available in different database repositories. These databases comprising of genomic sequences and amino acid sequences (proteins) are utilized in genetic engineering of biological systems to increase the production of chemicals and pharmaceuticals for improving plant and animal health. Recently, synthetic biology approaches are being employed in rational and high-throughput biological engineering to enhance the production of beneficial chemicals. Recent molecular and bioinformatics tools have enabled to redesign the entire biological cycle, including construction of synthetic DNA inside the cell or replacement of entire genome to create synthetic organisms by utilizing gene libraries, computational tools and interfaces. This review describes the genomic, proteomic and phylogenetic databases, which may be utilized for designing and manipulation of synthetic gene circuits to perform novel functions and desired phenotypes in different ecosystems. In addition, synthetic biology approaches were discussed for designing biological systems for production and release of specific metabolic products. The progress and challenges faced in computational methodology and synthetic biology approaches are discussed for their potential applications in synthetic biology.

KEYWORDS: Biological systems, Biological databases, Computational tools, Synthetic gene circuits, Synthetic biology

<https://doi.org/10.29294/IJASE.9.4.2023.3085-3098> ©2023 Mahendrapublications.com, All rights reserved

INTRODUCTION

Digitalized biological information and databases are doubling after every 12 to 18 months with current advances in genomics and proteomics [1]. Recently, there has been enormous progress in unraveling the complexities of naturally-occurring biological systems, which has provided abundant scientific information in the field of nucleic acid sequences and protein databases for agriculture, biomedical research, synthetic biology (biological and chemical engineering), and metabolic engineering including production of pharmaceuticals and nutraceuticals (Fig. 1) [2–5]. Such interdisciplinary approaches utilize computational and bioinformatics tools to manipulate microorganisms for getting a deeper understanding of the complex biological systems and have linked the multiple networks of fundamental biological discoveries [6,7]. At the same time, synthetic biology has evolved to generate predictable phenotypes of fairly large

networks of molecules and their reactivity, and mechanism is governed by several genes and microbial communities [8–10]. The future of systems biology and synthetic biology will involve engineering of the entire genomes to create synthetic organisms and ecosystems that are capable of performing novel functions and desired phenotypes in different environments [11–13]. Such advances will involve development of new experimental methods, computational approaches and theoretical as well as conceptual frameworks involving multi-scale modeling and data integration [14–18]. Current scientific progress in the biological systems, systematic understanding of complex metabolic regulatory mechanisms and their bioengineering holds enormous potential for improving crop production, human health and eco-friendly environment [19–22]. Biological databases incorporate data from the disciplines of genomics, microarray gene

*Corresponding Author: divya.sindhu91@gmail.com

Received: 15.02.2023

Accepted: 16.04.2023

Published on: 22.05.2023

Divya Sindhu & Saurabh Sindhu

expression, proteomics, phylogenetics and metabolomics in addition to details on the configuration, localization, and function of genes together with commonalities between biological sequences. In biochemical engineering, major components of metabolism could be totally redesigned for more efficient use of asset pools or resources to minimize material drains for a sustainable future [6, 23]. Microbial engineering often utilizes natural databases and computational tools at all degrees of biological organization and functions inside the cell [13,24,25]. Utilizing genome-scale models and optimization of algorithms, metabolic network analysis and designing of biological circuits may be accomplished

[4,26,27]. Additionally, synthetic DNA construct could be transferred in microbial strain or living cell, and these structured DNA sequences could give desired levels of transcription and translation to accomplish enhanced protein production [28–30]. Emerging paradigms for computing in living cells may contribute in development of predictive computational models that could be validated by experimentation and applicable across many living host species [8,31]. Biochemical and computer engineers may offer technical solutions for biosynthesis of recombinant proteins for novel sustainable processes as per ecological and economical needs [32–35].

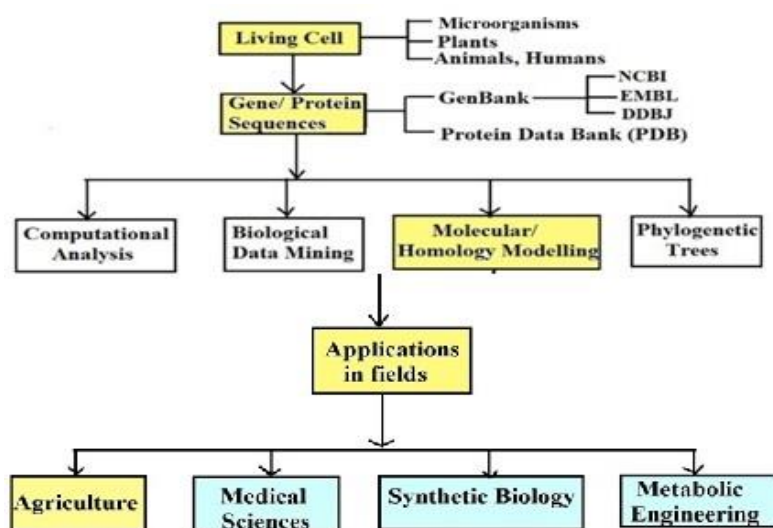


Fig.1. Application of DNA and protein databases for production of novel products and improving efficacy of biological systems

2. NUCLEIC ACID SEQUENCING AND DNA DATABASES

The genomic revolution in the last two decades has provided the ability to sequence a cell's genetic material i.e., deoxyribose nucleic acid (DNA), enabling the effective engineering of biological systems [36]. Nucleic acid constitutes the genetic material of the living organisms and is responsible for transfer of the hereditary information from the parents to the off springs. Nucleic acid exists either as DNA or ribose nucleic acid (RNA; in some viruses). DNA is a polymer of nucleotides, consisting of adenine (A), guanine (G), cytosine (C) and thymine (T) bases. The ability to store billions of nucleotide base sequences is an important feature of the DNA. The hereditary information present in the nucleotide sequences is maintained intact by complex metabolism

involving both DNA replication and repair functions. The different nucleotides i.e., ATP, GTP, CTP and TTP are polymerized by the DNA polymerase enzymes using one of the DNA strand as a template.

The technique used for determination of precise order of nucleotides in a piece of DNA is termed as DNA sequencing. Different methods of sequencing are employed to study arrangement of nucleotides on the genomic DNA. For instance, the chemical cleavage methods developed by Maxam and Gilbert and dideoxy chain termination method developed by Sanger are employed for rapid sequencing of long stretches of DNA [37]. Recently, automated DNA sequencing machines are capable of identifying 10,000 nucleotide base pairs per day and have become commercially available. In the automated systems, detection and analysis of sequencing reactions is carried out by

instruments controlled by computers. In automated SMRT sequencing method, the nucleotides are labelled differentially with fluorescent dyes that will be resolved by photo multiplier and the information is stored in the computer. Second generation sequencing technologies included a combination of a synchronized reagent wash of nucleoside triphosphates (NTPs) with a synchronized optical detection method, and also involved the use of 454 FLX, Solexa, scanning tunneling electron microscope (TEM), fluorescence resonance energy transfer (FRET), single molecule detection and protein nanopores. Sequencing speed and throughput was further increased in "third generation sequencing technologies", which included PacBio SMRT and nanopore sequencing techniques [38]. Thus, next generation sequencing has provided us powerful insights into genetic make-up of the microbial world. Recently, non-canonical nucleobase pairs were developed, which were found to augment the nucleotides present in DNA and RNA [39]. The molecular features needed for informational molecules in biology provided an intellectual framework for technologies to identify alternate genetic systems for life elsewhere in the Universe.

2.1. Nucleic Acid Databases

The Nucleic Acid Database (NDB) was curated by Research Col-laboratory for Structural Bioinformatics (RCSB). It gives user access to software tools and distributes data for extracting information from nucleic acid structures. The database contains tables of primary and derivative information. The primary information includes atomic coordinates, bibliographic references, crystal data, data collection and other structural descriptions. The derivative information is calculated from the primary information and includes chemical bond lengths and angles, virtual bond length and other measures according to various algorithms [40,41]. The experimental data in the NDB have been collected from published literature, as well as from standard crystallographic archive file types [40] and other sources. Several programs have been developed to convert among various file formats [42,43]. Shukla et al. [44] revealed the exhibition of sequence specific structural properties of DNA helix, which could be exploited by DNA-binding proteins to control transcription. Relevant databases are presented and categorized as aids in understanding the

resources that are available to bioinformatics researchers.

2.2. Nucleic acid-base repositories

Specific segments of DNA, which codes for particular protein/polypeptide are termed as genes. Individual person's genome has about three billion nucleotide bases, which possess the capability to encode about 100,000 genes. Interestingly, these coding regions make up only about 10% of our genome. Moreover, some genes are arranged as clusters known as operons and multigene families. The nucleic acid sequences from different viruses, bacteria, fungi and plants have been deposited in the National Center for Biotechnology Information's (NCBI's) GenBank (USA) [45]. Globally, most widely used large biological databank resource on the World Wide Web databases include the NCBI's GenBank, as well as its partners EMBL-Bank (Europe) [2] and the DNA Data Bank of Japan (DDBJ) [46] (Table 1). Other related databases deals at species-oriented databases including TAIR [47] and non-coding RNA sequence databases such as Rfam [48]. The database processing could be done by using the computer programmes and its comparison with the genome of other organisms is carried out by application of bioinformatics tools. Thus, the sequence homology could help in determining the function of particular protein or enzyme encoded by a particular gene.

The biological information encoded contained in various genes is made available by gene expression. The information is transferred from DNA to mRNA by the process of transcription and this information is further translated into the amino acids by making use of the ribosomes. Different amino acids join together through peptide linkages to make various proteins. The sequences of amino acids in various proteins are unique. Arrangement of amino acid sequences in specific proteins determine their primary, secondary and tertiary structures and confers specific functions in living cell [49,50]. The sequences of different proteins are deposited in the protein repositories such as UniProt [51,52] as well as its contributing data repositories viz. Swiss-Prot [53] and the Protein Information Resource [54-56]. Some of the nucleotide data bases along with their respective URLs have been listed in Table 1.

2.3. GenBank, EMBL Bank and DDBJ

GenBank is the genetic sequence database of National Institutes of Health (NIH) and it is an

annotated collection of all publicly available DNA sequences. Three databases were developed separately and the GenBank and EMBL-Bank were launched in 1980 [2, 45]. After the development of DDBJ [46], their collaboration started. These three databases

operate under the direction of the International Nucleotide Sequence Database Collaboration (INSDC) for collecting, maintaining and sharing of nucleotide data. Each database bank caters to the needs of the region in which it is located [57].

Table 1 Various databases and their respective URLs

Database	URL	Database	URL
GenBank	http://www.ncbi.nlm.nih.gov/	dbEST	http://www.ncbi.nlm.nih.gov/dbEST/
EMBL Bank	http://www.ebi.ac.uk/embl/	Rfam	http://rfam.sanger.ac.uk/
DDBJ	http://www.ddbj.nig.ac.jp	RNA STRAND	http://www.rnasoftware.ca/strand/
Ensembl	http://www.ensembl.org/	frnAdb	http://www.ncrna.org/frnadb/
TAIR	http://www.arabidopsis.org/	PIR	http://pir.georgetown.edu/
GeneDB	http://www.genedb.org	SGD	http://www.yeastgenome.org/
BV-BRC	https://www.bv-brc.org/	BIC	http://bic.jhlab.tw/
CottonMD	http://yanglab.hzau.edu.cn/CottonMD/	microbioTA	http://bio-annotation.cn/microbiota
BRAD	http://brassicadb.cn	CyanoOmicsDB	http://www.cyanoomics.cn
TFSyntax	https://tfsyntax.zhaopage.com	Ensembl COVID-19 resource	https://covid-19.ensembl.org

2.3.1. Ensembl Genome Database

The Ensembl database is available as an interactive Website or downloadable as flat files. It is a repository of stable, automatically annotated sequences resulting from the Human Genome Project [58]. Ensembl annotates and predicts new genes, with annotation from the InterPro [59] protein family databases and additional annotations from databases of genetic disease (OMIM) [60], serial analysis of gene expression (SAGE) [61] and gene family [62]. Software for Ensembl is freely available and it is based on relational database models [63].

2.3.2. Arabidopsis Information Resource

The Arabidopsis Information Resource (TAIR) allows for information retrieval and data analysis pertaining to *Arabidopsis thaliana* genome. *A. thaliana* is a small annual plant belonging to the mustard family and serves as a model for plant genome investigations. The genome of *A. thaliana* is completely sequenced and the database has been designed in a very simple, portable and efficient manner for its efficient utilization by the biologists and biotechnologists [47]. Map Viewer is an innovative aspect of the TAIR Website and it is an integrated visualization tool for viewing genetic, physical and sequence maps for each *Arabidopsis* chromosome. Each component of the map contains a hyperlink to an output page

from the database, which displays all the information related to this component [47].

2.3.3. Saccharomyces Genome Database

Saccharomyces cerevisiae is a baker's and brewer's yeast, and its genome has been completely sequenced. The *Saccharomyces* Genome Database (SGD) provides information for its genes, gene-encoded proteins, the structures and biological functions of known gene products and related literature [64]. The SGD database is not a primary sequence repository, but it is a collection of DNA and protein sequences from existing databases GenBank [45], EMBL-Bank [2], DDBJ [46], protein information resource (PIR) [51] and Swiss-Prot [53]. The sequences have been organized into datasets to make the data more useful and easily accessible.

2.3.4. GeneDB

GeneDB is a genome database for prokaryotic and eukaryotic organisms [65]. It contains genomic data generated from the Pathogen Sequencing Unit (PSU) at the Wellcome Trust Sanger Institute. The GeneDB database stores and frequently updates sequences and annotations. GeneDB also provides a user interface for easy access, visualization, searching and downloading of the data. In addition, the database architecture allows integration of different biological datasets with

the sequences. GeneDB also facilitates the comparisons of species by using structured vocabularies.

2.3.5. dbEST

dbEST database contains sequence data and other information on short, "single-pass" cDNA sequences, or expressed sequence tags (ESTs), generated from randomly selected library clones [66]. dbEST can be accessed using the Web, from NCBI by anonymous FTP or through entries [67]. BLAST sequence search program at the NCBI Website is used to search dbEST nucleotide sequences. dbEST DNA sequences can also be useful for finding novel coding sequences. On the other hand, EST sequences are available in the FASTA format from the /repository/dbEST directory at ftp.ncbi.nih.gov.

3. PROTEIN DATA BANK AND PROTEIN REPOSITORIES

The biological information encoded in various genes is made available by gene expression [68]. The information is transferred from DNA to mRNA by the process of transcription and this information is further translated into the protein by making use of the ribosomes. The arrangement of various amino acids in a protein determines the primary structure of protein. The amino acid sequences of various proteins are deposited in Protein Data Bank (PDB), which is a well-curated data resource, used widely in structural biology and biomedical sciences [69]. PDB was established as a repository for the three dimensional structures of biological macromolecules [3]. The Research Col-laboratory for Structural Bioinformatics (RCSB) maintains the PDB and it allows the user to view the data in plain text. Around 750 gigabytes of data are transferred each month from the website and the RCSB PDB website is accessed by about 250,000 unique visitors per month from 140 countries.

There are three major classes of nucleic acid containing entries in the PDB archive: RNA, DNA, and protein-nucleic acid complexes. More than 7,000 PDB entries contain carbohydrate polymers and/or individual saccharides. Besides the PDB, there are a large number of repositories and databases used in structural biology, chemistry, life sciences and big pharmaceutical companies, where they are crucial in the drug discovery process. PDB allows a wide spectrum of queries through data integration to provide complete information about the features of macromolecular

structures. The PDB collects and integrates external data from scientist's deposition, Gene Ontology (GO) [70], Enzyme Commission, KEGG Pathway Database [71], and NCBI resources [43]. Data integration through data loaders is written in Java, which extracts information from existing databases based on common identification numbers. The PDB also allows data extraction at query run time.

Structural models serve as primary reference data and this reference data resources act as repositories augmented with database functionality (Table 2). The data repository used in structural biology is the earlier version of the PDB [72]. There are five access sites in the PDB repository i.e., the wwPDB site, three data centers (PDBe, PDBj and RCSB PDB), and an NMR specific component, the Biological Magnetic Resonance Data Bank (BMRB) [73]. While the wwPDB site allows data validation and deposition as well as archive download, the remaining sites have more database capabilities that allow for data dissemination. All three PDB data centers utilize the common mmCIF format [74] to store the same underlying structural data, however, the design, information content, and analysis tools of each site are different. In addition to structural information, other information about proteins can be accessed from the Universal Protein Resource (UniProt) [75]. On the other hand, information about protein location, function and interactions can be accessed from Gene Ontology (GO) [70] or Kyoto Encyclopedia of Genes and Genomes (KEGG) [71,76]. Different tools and resources are applied to PDBsum service to obtain an overview of a protein structure [77]. It includes analysis of structural attributes such as protein surfaces, cavities and ligands as well as interaction attributes such as the protein-protein, protein-DNA/RNA and protein-small molecule interactions.

4. PHYLOGENETIC DATABASES

The understanding of genomic and proteomic databases, and to analyze their relatedness is quite crucial to understand biological evolution. Since all biological organisms have developed through the evolutionary process, their patterns, functions, and processes are best analyzed in terms of their phylogenetic histories. The expression of same gene at different timing, in a different tissue or its expression resulting into a whole new function along one phylogenetic branch is usually compared with another. These changes along a

branch affect the biology of all descendant species, thereby leaving phylogenetic patterns everywhere. A detailed mapping between biological data and phylogenetic histories is accomplished to realize the full potential of the data accumulation activities [78,79]. Phylogenetic patterns provide information about the differences observed in affectivity of certain drugs in some species but not in others; and for designing therapies against evolving disease agents such as HIV and influenza.

The need to query data using sets of evolutionarily related taxa has spawned the need to create databases than can serve as repositories of phylogenetic trees. Phylogeny and phylogenetic trees give a picture of the evolutionary history among species, individuals or genes. Therefore, there are at least two distinct goals of a phylogenetic database: archival storage and analysis [80]. Major phylogenetic databases and their respective URLs are provided in Table 2.

Table 2 Proteomic and Phylogenetic databases and their URLs

Serial no.	Database	URL
Proteomic databases		
1	UniPort	http://www.ebi.ac.uk/uniprot/
2	Swiss-Prot/TrEMBL	http://www.expasy.org/sprot/
3	TmAlphaFold	https://tmalphafold.ttk.hu/ AlphaFold TM protein predictions and assessed
4	COMBATdb	https://db.combat.ox.ac.uk COVID-19 Multi-omics Blood Atlas
5	ProtCAD	http://dunbrack2.fccc.edu/protcad Protein Common Assembly Database
6	ChromLoops	https://3dgenomics.hzau.edu.cn/chromloops Protein-mediated chromatin loops
7	PAT	http://bioinfo.qd.sdu.edu.cn/PAT/ Prokaryotic Antimicrobial Toxin database
Phylogenetic databases		
1	TreeBASE	http://www.treebase.org/
2	TreeFam	http://www.treefam.org/
3	Tree of Life	http://tolweb.org/tree/
4	NCBI Taxonomy	http://www.ncbi.nlm.nih.gov/taxonomy
5	SYSTERS	http://systems.molgen.mpg.de/
6	PANDIT	http://www.ebi.ac.uk/research/goldman/software/pandit/

Due to rapid developments in genomics and proteomics involving novel sequencing technologies, large amounts of biological information is now available in biological databases. Sophisticated computational and bioinformatical analyses using data mining (DM) approaches and phylogenetic profile methods help in understanding the network of biological linkages and predictions of functional interactions between proteins across multiple genomes [49, 81, 82]. The varied applications of phylogenetic tools, algorithms and use of *in silico* methods to study phylogeny of microbes may be considered as a highly reliable and important technique in biological sciences [33, 83, 84] and may replace wet experiments for prediction of evolutionary relationships between two microbial species in a laboratory.

5. ENGINEERING OF BIOLOGICAL DATABASES AND SYNTHETIC BIOLOGY

Living cells of the organisms contain DNA and RNA, proteins, enzymes, carbohydrates, lipids, vitamins and minerals. Various cells operate as highly complex biological computational systems, which sense the surroundings, interrogate the signals and respond to their environment [85]. Synthetic biology involves the construction and manipulation of the biological systems from the minute molecule (individual functional unit) to the functional cellular level [86,87]. The genomic revolution in the last two decades has provided the ability to sequence a cell's genetic DNA, enabling the effective engineering of biological systems. Current advances in DNA synthesis and DNA assembly techniques have made it possible to engineer virus and bacterial genomes as well as

various metabolic pathways to modify, regulate and control cellular behaviour in a desired manner [88, 89]. In addition, a finer control over the expression of particular gene(s) in a given metabolic pathway could be designed and implemented either at the transcription or translation levels for its enhanced production, termed as metabolic engineering [14, 27,30]. Recently, synthetic promoters and synthetic enhancers have been designed to produce a desired level of transcriptional strength with the recent molecular and experimental tools [90]. Moreover, application of specific experimental and computational tools may regulate the expression of specific gene or metabolic pathway at both the transcriptional and translational levels for metabolic engineering applications [28,91]. Synthetic biologists leverage engineering design principles involving manipulations at the genomic level to use the predictability of engineering to control complex biological systems. Moreover, synthetic biology (synbio) endeavours to develop artificial cell-free biological systems through the combination of molecular biology and engineering approaches [86].

Currently, genetic manipulation of various microorganisms i.e., fungi, bacteria, and yeast has been made using various biotechnology and bioinformatics tools/techniques for biological and chemical engineering [4, 9, 92–94]. New computational approaches have been developed for the construction of synthetic pathways using directed evolution of enzymes, for structuring and developing artificial enzymes (for non-natural reactions) and for re-wiring of host metabolism to alter the metabolic flux for synthesis of non-natural chemical products [6,35]. Qiu et al. [95] proposed novel emerging procedures to improve survival and activities of microbial inoculants, improvement in microbial delivery strategies and use of gene editing tools to design and engineer microbial inoculants. In a similar way, multiple biochemical and molecular methodologies were optimized for use in microbiome engineering to enhance beneficial plant-microbiome interactions for improving crop yields [96,97]. In addition, biological engineering for nature-based climate solutions (NbCS) will help in creating a productive, resilient, and proactive “climate-smart agriculture” to mitigate the risks of climate change [98].

In addition to improvement in agriculture production, the engineering of the beneficial

microorganisms may enhance the production of therapeutic and pharmaceuticals small molecules [34], which depends on optimization of biochemical pathway and computational tools developed by metabolic engineers for performing a particular function. This approach was also applied in *Escherichia coli* for ethanol production by using a quorum sensing module for density dependent repression (via a toggle switch) of phosphor-transacetylase (pta), which caused inactivation of a competing acetate-production pathway [99]. However, synthetic circuit decreased the yield and differed in behaviour from predictive models. The development of advanced technologies i.e., multiplex automated genome engineering (MAGE) relied on incorporation of multiple single-strand oligonucleotides introduced via electroporation into daughter cell genomes [100,101]. The application of this technology resulted in 4-5-fold increase in the production of lycopene as well as aromatic amino acid derivatives [102]. Similarly, Kang et al. [4] used genetic circuit-guided population acclimation of a synthetic microbial consortium including *Vibrio* sp. and *Escherichia coli* strains and achieved 4.3-fold increase in 3-hydroxypropionic acid (3-HP) production during a 48 hours fermentation process.

For improving human health and the therapeutic potential, engineering of gut bacteria belonging to Bacteroidetes and Firmicutes was undertaken using different bioinformatics tools [103]. Consequently, synthetic biologists developed a toolkit amenable for engineering of the commensal *Bacteroides thetaiotamicron* and it comprises characterized promoters, ribosome binding site, inducible systems and CRISPRi platform [104]. Similarly, metabolic engineering may be utilized to engineer novel diagnostic and therapeutic strategies for control of cancer and infectious diseases. For instance, *Escherichia coli* was engineered to invade mammalian cells selectively in hypoxic environments [105]. In another study, *Escherichia coli* was engineered to sense the occurrence of bacterium *Pseudomonas aeruginosa*, which has been found to cause infections in the lung, urinary tract, gastrointestinal tract and skin [106]. Furthermore, synthetic biology plays a crucial role in designing and creating genetically engineered living materials (ELMs), comprising cells, microbes, biofilms, and spores, for biotherapeutic applications and represents a new platform for diagnostic and targeted

delivery to treat intractable diseases [107]. Similarly, many enzymes are involved in mediating cellular metabolism that represents drug targets. The anthranilate phosphoribosyl transferase enzyme (involved in catalyzing tryptophan biosynthesis from chorismate) is considered essential for the growth of *Mycobacterium tuberculosis* and could act as a drug target [108]. Such inhibitors may be useful in treating mycobacterial infections, and will address the problem of multi-drug resistance in *M. tuberculosis*.

Application of biological databases using synthetic biology aims to design biological systems to a specification that react in a specific manner to an external stimulus to produce/create specific products. For instance, the combination of high-throughput phenotypic data with precision DNA editing by using CRISPR-based tools provided a unique opportunity to link changes in the underlying code to phenotype [109]. Moreover, current use of novel synthetic biology tools such as machine learning-based metabolic modeling, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) derived synthetic biology tools, and synthetic genetic circuits have accelerated our systematic understanding of complex metabolic regulatory mechanisms [22]. These tools are widely used to control the metabolism of microorganisms, manipulate gene expression, and build synthetic pathways for bioproduction in industrial bioprocesses

Using novel synthetic biology tools, semi-synthetic organisms may be created in which increased genetic information may be stored and retrieved. For example, new codon/anti-codon pairs may be created to express proteins in *Escherichia coli* containing amino acids that are not found in nature [13]. Thus, successful production of 'semi-synthetic organisms' has profound implications for alternative metabolic pathways. In addition, novel technologies may help in production of new types of protein-based therapeutics and enzymes for use in the sustainable industrial synthesis of bulk chemicals. Recently, DNA Oligo Libraries (OLs)-based synthetic enhancers were designed and constructed to regulate gene expression by using latest commercial OLs synthesis technology [90]. Authors specifically focussed on synthetic-enhancer-based massively parallel reporter assay, Sort-seq methodologies (e.g. flow cytometry, deep sequencing), and machine learning-based attempts for OL-analysis followed up by validation experiments.

Recently, Benner [39] reviewed the development of non-canonical nucleobase pairs that can augment the nucleotides present in DNA and RNA. In subsequent studies, these non-canonical nucleobase pairs were used to build DNA [44]. Interestingly, synthesized DNA exhibited sequence specific structural properties of DNA helix, which could be exploited by DNA-binding proteins to control transcription.

6. CONCLUSIONS

Significant information about the DNA nucleotide sequences and amino acid sequences of proteins, obtained from different laboratories worldwide, is currently available in different database repositories [1,3]. These biological databases are utilized in genetic engineering of biological systems to increase the production of desired products or chemicals for benefit of plants and animals [4,6]. In addition to utilization of these biological databases, genome-scale models, optimization of algorithms, metabolic network analysis and bioinformatics tools are used in synthetic biology and metabolic engineering to completely redesign the genetic circuits within a living cell for more efficient utilization of resource pools with minimal material drains [26,110]. Molecular tools and strategies have been developed for using in *silico* and mathematical modeling of the biological systems to analyze endogenous biological circuits, with a particular focus on signaling and metabolic pathways for application in systems biology [18,111,112]. In addition, computer simulation allows researchers to build a common framework for designing novel biological networks for metabolic engineering [4,8]. The advantage of the cell-free systems under *in vitro* reactions has further contributed with the analysis, investigation, estimation and elucidation of dynamic interactions between genes and proteins in naturally-occurring systems [29]. Moreover, novel synthetic genetic constructs could be transferred into a microbial strain/living cell, and these designed DNA sequences could provide desired levels of transcription and translation to achieve enhanced protein production [27,28,91].

Thus, synthetic biology approaches include designing of biological systems by manipulation at the genomic level that reacts in a specific manner to the external stimulus to produce specific products. In addition, computational and bioinformatics tools may improve the

predictability of respective mechanisms and techniques for increasing the production and release of specific metabolic product. Current novel technologies involving programmable synthetic protocells, current advances at the interface of hardware and wetware such as solid-phase DNA assembly platforms, or delivery systems in basic bioscience research may act as emerging hubs for automation of biological engineering. Recent powerful approaches involving directed evolution of valuable new enzymes, designing of synthetic genomes using computer-aided design (CAD) technologies, automation of specialized methods for chromosome transfer between microbes, plants, and mammalian cells (such as cell fusion, genome transplantation, or microinjection) and artificial cell research will further help in solving the upcoming challenges of synthetic biology [113–115]. This highly interdisciplinary and international approach of biochemical engineering involving government and private sectors will help in achieving the desired impact in biomedical, pharmaceutical, agricultural and chemical industries.

REFERENCES

- [1] Rigden, D.J., Fernández, X.M., 2023. The 2023 Nucleic acids research database issue and the online molecular biology database collection. *Nucleic Acids Research*, 51, D1–D8.
- [2] Cochrane, G., Aldebert, P., Althorpe, N., Andersson, M., Baker, W., Baldwin, A., Bates, K., Bhattacharyya, S., Browne, P. et al., 2006. EMBL nucleotide sequence database: developments in 2005. *Nucleic Acids Research*, 34, D10–D15.
- [3] Berman, H., 2008. The Protein Data Bank: a historical perspective. *Acta Crystallography A: Foundations of Crystallography*, 64, 88–95.
- [4] Kang, C.W., Lim, H.G., Won, J., Cha, S., Shin, G., Yang J-S., Shung, J., Jung, G.Y., 2022. Circuit-guided population acclimation of a synthetic microbial consortium for improved biochemical production. *Nature Communications*, 13, 6506.
- [5] Sharma, A., 2022. An overview of biological databases used in bioinformatics. *International Journal of Computational Biology and Bioinformatics* 8(2).
- [6] Biz, A., Proulex, S., Xu, Z., Siddartha, K., Mullet Indrayanti, A., Mahadevan, R., 2019. Systems biology based metabolic engineering for non-natural chemicals. *Biotechnological Advances*.
- [7] Roderiguez, P.A., Rothballer, M., Chowdhury, S.P., Nussbaumer, T., Gutjahr, C., Falter-Braun, P., 2019. Systems biology of plant-microbe interactions. *Molecular Plant*, 12, 804–821.
- [8] Zhu, L., Zhu, Y., Zhang, Y., Li, Y., 2012. Engineering the robustness of industrial microbes through synthetic biology. *Trends in Microbiology*, 20(2), 94–101.
- [9] McCarty, N.S., Ledesma-Amaro, R., 2019. Synthetic biology tools to engineer microbial communities for biotechnology. *Trends in Biotechnology*, 37(2), 181–197.
- [10] Richards, N.G.J., Bearne, S.L., Goto, Y., Parker, E.J., 2023. Reactivity and mechanism in chemical and synthetic biology. *Philosophical Transactions of Royal Society B*, 378, 20220023.
- [11] Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.Y., Algire, M.A., et al., 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 329, 52–56.
- [12] Jarboe, L.R., Zhang, X., Wang, X., Moore, J.C., Shanmugam, K.T., Ingram, L.O., 2010. Metabolic engineering for production of biorenewable fuels and chemicals: Contributions of synthetic biology. *Journal of Biomedical and Biotechnology*, 2010, 761042.
- [13] Romesberg, F.E., 2022. Creation, optimization, and use of semi-synthetic organisms that store and retrieve increased genetic information. *Journal of Molecular Biology*, 434, 167331.
- [14] Sindhu, S., Sindhu, D., 2016. Development of computational tools for metabolic engineering. *International Journal of Innovative Research in Computer Communication and Engineering*, 4(5), 9208–9217.
- [15] Tsoi, R., Dai, Z., You, L., 2019. Emerging strategies for engineering microbial communities. *Biotechnol. Adv.*
- [16] Paddon, C.J., Zomorodi, A.R., Segre, D., 2016. Synthetic ecology of microbes: Mathematical models and applications. *Journal of Molecular Biology*, 428, 837–861.

- [17] Inomura, K., Deutsch, C., Masuda, T., Prášil, O., Follows, M.J., 2020. Quantitative models of nitrogen-fixing organisms. *Computer Structural Biotechnology Journal*, 18, 3905–3924.
- [18] Sindhu, D., Hooda, E., Sindhu, S., Yadav, S.K., 2021. Development of novel predictive models for estimation of nitrogen fixation under cultural and field conditions using R software. *International Journal of Advance Trends in Computer Science and Engineering*, 10(4), 2704–2713.
- [19] Mueller, U.G., Sachs, J.L., 2015. Engineering microbiomes to improve plant and animal health. *Trends in Microbiology*, 23, 606–617.
- [20] Qiu, Z., Egidi, E., Liu, H., Kaur, S., Singh, B.K., 2019. New frontiers in agriculture productivity: optimised microbial inoculants and *in situ* microbiome engineering. *Biotechnological Advances*.
- [21] Ke, J., Wang, B., Yoshikuni, Y., 2021. Microbiome engineering: Synthetic biology of plant-associated microbiomes in sustainable agriculture. *Trends in Biotechnology*, 39, 244–261.
- [22] Lv, X., Hueso-Gil, A., Bi, X., Wu, Y., Liu, Y., Liu, L., Ledesma-Amaro, R., 2022. New synthetic biology tools for metabolic control. *Current Opinion in Biotechnology*, 76, 102724.
- [23] Takors, R., 2020. Biochemical engineering provides mindset, tools and solutions for the driving questions of a sustainable future. *Engineering Life Sciences*, 20, 5–6.
- [24] Kitada, T., DiAndreth, B., Teague, B., Weiss, R., 2018. Programming gene and engineered-cell therapies with synthetic biology. *Science*, 359(6376), eaad1067.
- [25] Kumar, A., Wang, L., Ng, C.Y., Maranas, C.D., 2018. Pathway design using de novo steps through uncharted biochemical spaces. *Nature Communications*, 9.
- [26] Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M., 2014. Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Molecular System Biology*, 10, 731–736.
- [27] English, M.A., Gayet, R.V., Collins, J.J., 2021. Designing biological circuits: synthetic biology within the operon model and beyond. *Annual Review of Biochemistry*, 90, 221–244.
- [28] Mutalik, V.K., Guimaraes, J.C., Cambray, G., Lam, C., Christoffersen, M.J., 2013. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, 10, 354–360.
- [29] Bradley, R.W., Buck, M., Wang, B., 2016. Tools and principles for microbial circuit engineering. *Journal of Molecular Biology*, 428(5), 862–888.
- [30] Xia, P-F., Foo, J.L., Ling, H., Chang, M.W., 2019. Synthetic genetic circuits for programmable biological functionalities. *Biotechnological Advances*.
- [31] Ma, K.C., Perli, S.D., Lu, T.K., 2016. Foundations and emerging paradigms for computing in living cells. *Journal of Molecular Biology*, 428, 893–915.
- [32] Torres, L., Krüger, A., Csibra, E., Gianni, E., Pinheiro, V.B. 2016. Synthetic biology approaches to biological containment: pre-emptively tackling potential risks. *Essays in Biochemistry*, 60, 395–410.
- [33] Sindhu, S., Sindhu, D., 2017. Data mining and gene expression analysis in Bioinformatics. *International Journal of Computer Science and Mobile Computing*, 6(5), 72–83.
- [34] Sarsaiya, S., Shi, J., Chen, J., 2019. Bioengineering tools for the production of pharmaceuticals: current perspective and future outlook. *Bioengineered* 10:1, 469–492.
- [35] Wang, Y., Ling, C., Chen, Y., Jiang, X., Chen, G.Q., 2019. Microbial engineering for easy downstream processing. *Biotechnological Advances*.
- [36] Rigden, D.J., Fernández, X.M., 2022. The 2022 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Research*, 50, D1–D10.
- [37] Sanger, F., 1988. Sequences, sequences, and sequences. *Annual Review of Biochemistry*, 57, 1–28.
- [38] Shendure, J., Hanlee, J., 2008. Next generation DNA sequencing. *Nature Biotechnology*, 26, 1135–1145.
- [39] Benner, S.A., 2023. Rethinking nucleic acids: From their origins to their applications. *Philosophical Transactions of Royal Society B*, 378, 20220027.

- [40] Grzeskowiak, K., Yanagi, K., Prive, G.G., Dickerson, R.E., 1991. The structure of B-helical C-G-A-T-C-G-A-T-C-G, and comparison with C-C-A-A-C-G-T-T-G-G: the effect of base pair reversal. *Journal of Biological Chemistry*, 266, 8861–8883.
- [41] Babcock, M.S., Olson, W.K., 1993. A new program for the analysis of nucleic acid structure: Implications for nucleic acid structure interpretation. In: Soumpasis, D.M., Jovin, T.M. (eds.), *Computation of Biomolecular Structures*, pp. 65–85. Springer.
- [42] Neudert, G., Klebe, G., 2011. FCONV: format conversion, manipulation and feature computation of molecular data. *Bioinformatics*, 27(7), 1021–1022.
- [43] Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Farrell, C.M., Feldgarden, M., Fine, A.M., Funk, K. et al., 2022. Database resources of the National Center for Biotechnology Information in 2023. *Nucleic Acids Research*.
- [44] Shukla, M.S., Hoshika, S., Benner, S.A., Georgiadis, M.M., 2023. Crystal structures of 'Alternative Isoinformational Engineered' DNA in B-form. *Philosophical Transactions of Royal Society B*, 378, 20220028.
- [45] Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2013. GenBank. *Nucleic Acids Research*, 38, D46–D51.
- [46] Okubo, K., Sugawara, H., Gojobori, T., Tateno, Y., 2006. DDBJ in preparation for overview of research activities behind data submissions. *Nucleic Acids Research*, 34(1), D6–D9.
- [47] Meinke, D.W., Cherry, J.M., Dean, C., Rounsley, S.D., Koornneef, M., 1998. *Arabidopsis thaliana*: A model plant for genome analysis. *Science*, 282, 678–682.
- [48] Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Bateman, A., Nawrocki, E. P., Eddy, S R., Gardner, P.P., 2013. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Research*, 41(D1), D226–D232.
- [49] Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O., Eisenberg, D., 1999. Detecting protein function and protein-protein interactions from genome sequences, *Science*, 285, 751–753,
- [50] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R, Žídek, A., Potapenko, A., et al., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596, 583–589.
- [51] Wu, C.H., Apweller, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., et. al., 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Research*, 34, D187–D191.
- [52] The UniProt Consortium, 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49, D480–D489.
- [53] O'Donovan, C., Martin, M. J., Gattiker, A., Gasteiger, E., Bairoch, A., Apweller, R., 2002. High-quality protein knowledge resource: SWISSPROT and TrEMBL. *Briefings in Bioinformatics*, 3, 275–284.
- [54] Wu, C.H., Huang, H., Nikolskaya, A., Hu, Z.Z., Barker, W.C., 2004. The iProClass integrated database for protein functional analysis. *Computational Biology and Chemistry*, 28, 87–96.
- [55] Wu, C.H., Yeh, L.S., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., Hu, Z.Z., Ledley, R.S., Kourtesis, P. et al., 2003. The protein information resource. *Nucleic Acids Research*, 31, 345–347.
- [56] Galperin, M.Y., 2006. The molecular biology database collection: 2006 update. *Nucleic Acids Research*, 34, D3–D5.
- [57] Lesk, A., 2005. *Database Annotation in Molecular Biology*. John Wiley & Sons.
- [58] Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., et. al., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 916–921.
- [59] Mulder, N.J., Apweller, R., Binns, D., Bradley, P., Das, U., Fleischmann, W., Harte, N., Kanapin, A., et. al., 2005. InterPro, progress and status in 2005. *Nucleic Acids Research*, 33, D201–D205.
- [60] Antonarakis, S.E., McKusick, V.A., 2000. OMIM passes the 1,000-disease-gene mark. *Nature Genetics*, 25(1), 11–17.
- [61] Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W., 1995. Serial analysis

- of gene expression. *Science*, 270, 484–487.
- [62] Enright, A.J., Iliopoulos, I., Kyrpidis, N.C., Ouzounis, C.A., 1999. Protein interaction maps for complete genomes based on gene fusion events. *Nature*, 402, 86–90.
- [63] Hubbard, T., Barker, D., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Pettett, R., Pocock, M., et al., 2002. The Ensembl genome database project. *Nucleic Acids Research*, 30(1), 38–41.
- [64] Cherry, J.M., Adler, C., Ball, C., Chervitz, S.A., Dwight, S. S., Hester, E. T., Jia, Y., Juvik, G., Roe, T., Schroeder, M., Weng, S., Botstein, D., 1998. SGD: *Saccharomyces* genome database. *Nucleic Acids Research*, 26, D73–D79.
- [65] Hertz-Fowler, C., Peacock, C.S., Wood, V., Aslett, M., Kerhornour, A., Mooney, P., Tivey, A., Berriman, M., Hall, N., Rutherford, K., et al., 2004. Gene DB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Research*, 32, D339–D343.
- [66] Boguski, M.S., Lowe, T.M., Tolstoshev, C.M., 1993. dbEST: database for expressed sequence tags. *Nature Genetics*, 4, 332–333.
- [67] Schuler, G.D., Epstein, J.A., Ohkawa, H., Kans, J.A. 1996. Entrez: Molecular biology database and retrieval system. *Methods in Enzymology*, 266, 141–162.
- [68] Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3, 318–356.
- [69] Rose, P.W., Prlic, A., Bi, C., Bluhm, W.F., Christie, C.H., Dutta, S., et al., 2015. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Research*, 43, D345–356.
- [70] Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., et al., 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1), 25–29.
- [71] Kanehisa, M., Goto. S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28, D27–D30.
- [72] Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F. Jr., Brice, M.D., Rodgers, J.R., et al., 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112(3), 535–542.
- [73] Huang, Y.H., Rose, P.W., Hsu, C.N., 2015. Citing a data repository: A case study of the Protein Data Bank. *PLoS One*, 10(8), e0136631.
- [74] Westbrook, J.D., Bourne, P.E., 2000. STAR/mmCIF: an ontology for macromolecular structure. *Bioinformatics*, 16(2), 159–168.
- [75] UniProt Consortium, 2015. UniProt: a hub for protein information. *Nucleic Acids Research*, 43, D204–D212.
- [76] Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., Ishiguro-Watanabe, M., 2022. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*.
- [77] Laskowski, R.A., 2009. PDBsum new things. *Nucleic Acids Research*, 37, D355–D359.
- [78] Sindhu, S., Sindhu, D., Yadav, S.K. 2021. Data mining and phylogenetic analysis of NifH protein of *Azospirillum* strain among nitrogen-fixing bacteria using bioinformatics tools. *International Journal of Computer Sciences and Engineering*, 9(1), 01–10.
- [79] Sindhu, D., Neeru, R., Sindhu, S., Yadav, S.K. 2021. Computational biology: Use of NifA protein amino acid sequences of *Azorhizobium* strain for phylogenetic analysis among nitrogen-fixing organisms. *International Journal of Advanced Research in Science, Engineering and Technology*, 8(3), 16828–16838.
- [80] Wang, J.T.L., Zaki, M.J., Toivonen, H.T.T., Shasha, D., 2005. Data mining in Bioinformatics, Springer.
- [81] Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., Yeates, T.O., 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proceedings National Academy Sciences, USA*, 96, 4285–4288.
- [82] Enright, A.J., Iliopoulos, I., Kyrpidis, N.C., Ouzounis, C.A., Butland, G., Peregrín-Alvarez, J.M., Li, J., Yang, W., et al., 2005. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*, 433, 531–537.
- [83] Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., Sali, A., 2000.

- Comparative protein structure modeling of genes and genomes. Annual Review of Biophysics and Biomolecular Structure, 29, 291–298.
- [84] Sindhu, S., Sindhu, D., 2018. Designing and engineering of biological systems using computation tools. International Journal of Engineering Research and Applications, 8(1), 34–42.
- [85] Phour M, Sehrawat A, Sindhu S.S. and Glick BR (2020) Interkingdom signaling in plant- rhizomicrobiome interactions for sustainable agriculture. Microbiological Research, 241, 126589.
- [86] Ceroni, F., Carbonell, P., François, J.M., Karmella, A., 2015. Synthetic biology for engineering complexity. Frontiers in Bioengineering and Biotechnology, 3, 1–2.
- [87] Cheng, A.A., Lu T.K., 2012. Synthetic biology: An emerging engineering discipline. Annual Review in Biomedical Engineering, 14, 55–78.
- [88] Fong, S.S., 2014. Computational approaches to metabolic engineering utilizing systems biology and synthetic biology. Computational and Structural Biotechnology Journal, 11, 28–34.
- [89] Kelwick, R., MacDonald, J.T., Webb, A.J., Freemont, P., 2014. Developments in the tools and methodologies of synthetic biology. Frontier in Bioengineering and Biotechnology, 2, 60–83.
- [90] Vaknin, I., Amit, R., 2022. Molecular and experimental tools to design synthetic enhancers. Current Opinion in Biotechnology, 76, 102728.
- [91] Covert, M.W., Palsson, B.O., 2002. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. Journal of Biological Chemistry, 277, 28058–28064.
- [92] Yuan, S.F., Alper, H.S., 2019. Metabolic engineering of microbial cell factories for production of nutraceuticals. Microbial Cell Factories, 18.
- [93] Boock, J.T., Freedman, A.J.E., Tompsett, G.A., Muse, S.K., Allen, A.J., Jackson, L.A., Castro-Dominguez, B., Timko, M.T., Prather, K.L.J., Thompson, J.R., 2019. Engineered microbial biofuel production and recovery under supercritical carbon dioxide. Nature Communication, 10.
- [94] Choi, S.S., Katsuyama, Y., Bai, L.Q., Deng, Z.X., Ohnishi, Y., Kim, E.S., 2018. Genome engineering for microbial natural product discovery. Current Opinion in Microbiology, 45, 53–60.
- [95] Qiu, Z., Egidi, E., Liu, H., Kaur, S., Singh, B.K., 2019. New frontiers in agriculture productivity: Optimised microbial inoculants and *in situ* microbiome engineering. Biotechnological Advances. Alori, E.T., Babalola, O.O., 2018. Microbial inoculants for improving crop quality and human health in Africa. Frontiers in Microbiology, 9, 2213.
- [96] Cao, M., Narayanan, M., Shi, X., Chen, X., Li, Z., Ma, Y., 2023. Optimistic contributions of plant growth-promoting bacteria for sustainable agriculture and climate stress alleviation. Environmental Research, 217, 114924.
- [97] Runkle, B.R.K., 2022. Biological engineering for nature-based climate solutions. Journal of Biological Engineering, 16, 7.
- [98] Anesiadis, N., Kobayashi, H., Cluett, W.R., Mahadevan, R., 2013. Analysis and design of genetic circuit for dynamic metabolic engineering, ACS Synthetic Biology, 2, 442–452.
- [99] Isaacs, F.J., Carr, P.A., Wang, H.H., Lajoie, M.J., Sterling, B., Kraal, L., et al., 2011. Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. Science, 333, 348–353.
- [100] Wang, H.H., Kim, H., Cong, L., Jeong, J., Bang, D., Church, G.M., 2012. Genome-scale promoter engineering by co-selection MAGE. Natural Methods, 9, 591–593.
- [101] Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., et al., 2009. Software cells by multiplex genome engineering and accelerated evolution. Nature, 460, 894–898.
- [102] Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., et al., 2010. Human gut microbial gene catalogue established by metagenomic sequencing. Nature, 464, 59–65.
- [103] Mimee, M., Tucker, A.C., Voigt, C.A., Lu, T.K., 2015. Software human commensal bacterium, *Bacteroides thetaiotaomicron*, to sense and respond to stimuli in murine gut microbiota. Cell System, 1, 62–71.

- [105] Anderson, J.C., Clarke, E.J., Arkin, A.P., Voigt, C.A., 2006. Environmentally controlled invasion of cancer cells by engineered bacteria. *Journal of Molecular Biology*, 355, 619–627.
- [106] Saeidi, N., Wong, C.K., Lo, T.M., Nguyen, H.X., Ling, H., et al., 2011. Engineering microbes to sense and eradicate *Pseudomonas aeruginosa*, human pathogen. *Molecular System Biology*, 7, 521–528.
- [107] Rabia, O., Zubair, M.M., Ali, M., Sajid, M.B., Xumeng, H., Meijin, G., Yingping, Z., Jiaofang, H., 2022. Engineered bacteria-based living materials for biotherapeutic applications. *Frontiers in Bioengineering and Biotechnology*, 10.
- [108] Scully, T.W., Jiao, W., Mittelstädt, G., Parker, E.J., 2023. Structure, mechanism and inhibition of anthranilate phosphoribosyl transferase. *Philosophical Transactions of Royal Society B*, 378, 20220039.
- [109] Eslami, M., Adler, A., Caceres, R.S., Dunn, J.G., Kelley-Loughnane, N., Varaljay, V.A., Martin, H.G., 2022. Artificial intelligence for synthetic biology. *Communications of the ACM*, 65(5), 88–97.
- [110] Cameron, D.E., Bashor, C.J., Collins J.J., 2014. A brief history of synthetic biology. *Natural Review Microbiology*, 12, 381–390.
- [111] Tanaka, R.J., Okano, H., Kimura, H., 2006. Mathematical description of gene regulatory units. *Biophysics Journal*, 91, 1235–1247.
- [112] Elizabeth, P., Isaac, L., Kevin, T., 2008. Computational modeling approaches for studying of synthetic biological networks. *Current Bioinformatics*, 3, 1–12.
- [113] Ostrov, N., Beal, J., Ellis, T., Gordon, D.B., Karas, B.J., Lee, H.H., Lenaghan, S.C., Schloss, J.A., et al., 2019. Technological challenges and milestones for writing genomes. *Science*, 366, 310–312.
- [114] De la Torre, D., Chin, J.W., 2021. Reprogramming the genetic code. *Nature Review Genetics*, 22(3), 169–184.
- [115] Gallup, O., Ming, H., Ellis, T., 2021. Ten future challenges for synthetic biology. *Engineering Biology*, 5(3), 51–59.