Review

AI-Assisted Rational Design and Activity Prediction of Biological Elements for Optimizing Transcription-Factor-Based Biosensors

Nana Ding [1,2,*,†] , Zenan Yuan [1,2,†], Zheng Ma [3], Yefei Wu [4] and Lianghong Yin [1,2,*]

1   State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou 311300, China; yuanzenan20001003@163.com
2   Zhejiang Provincial Key Laboratory of Resources Protection and Innovation of Traditional Chinese Medicine, Zhejiang A&F University, Hangzhou 311300, China
3   Zhejiang Provincial Key Laboratory of Biometrology and Inspection & Quarantine, College of Life Sciences, China Jiliang University, Hangzhou 310018, China; mazheng520@163.com
4   Zhejiang Qianjiang Biochemical Co., Ltd., Haining 314400, China; yefei0906@163.com
*   Correspondence: dingnana911128@163.com (N.D.); yinlianghong@zafu.edu.cn (L.Y.)
†   These authors contributed equally to this work.

Abstract: The rational design, activity prediction, and adaptive application of biological elements (bio-elements) are crucial research fields in synthetic biology. Currently, a major challenge in the field is efficiently designing desired bio-elements and accurately predicting their activity using vast datasets. The advancement of artificial intelligence (AI) technology has enabled machine learning and deep learning algorithms to excel in uncovering patterns in bio-element data and predicting their performance. This review explores the application of AI algorithms in the rational design of bio-elements, activity prediction, and the regulation of transcription-factor-based biosensor response performance using AI-designed elements. We discuss the advantages, adaptability, and biological challenges addressed by the AI algorithms in various applications, highlighting their powerful potential in analyzing biological data. Furthermore, we propose innovative solutions to the challenges faced by AI algorithms in the field and suggest future research directions. By consolidating current research and demonstrating the practical applications and future potential of AI in synthetic biology, this review provides valuable insights for advancing both academic research and practical applications in biotechnology.

Keywords: synthetic biology; biological elements; transcription-factor-based biosensor; artificial intelligence; machine learning; deep learning

1. Introduction

Biological elements (bio-elements) are the fundamental building blocks of synthetic biological systems, consisting of molecular sequences with specific functions. These bio-elements include nucleic acids, such as promoters, enhancers, and ribosome binding sites (RBS), as well as proteins, such as transcription factors (TF) and enzymes. The rapid development of synthetic biology has made the rational design of bio-elements and the precise prediction of their activity key areas of research [1–5]. However, the sequences and functions of bio-elements are complex, and their intrinsic relationships remain unclear. Traditional methods for designing bio-elements and predicting their activity often suffer from being time-consuming, costly, unreliable, lacking clear optimization directions, and having insufficient relevant biological theories. Fortunately, the rapid development of machine learning algorithms in artificial intelligence (AI), such as support vector machines, logistic regression, and decision trees, combined with big data, has shown excellent performance in various fields [6–10]. Machine learning possesses strong capabilities to understand biological data and make autonomous decisions, enabling the extraction of hidden features

within biological data that are difficult to obtain through experimental methods [11,12]. Deep learning [13], a new branch of machine learning, can efficiently learn latent patterns in data [14,15]. For example, convolutional neural networks (CNNs, a type of feed-forward neural network that incorporates convolutional calculations and has a deep architecture) and recurrent neural networks (RNNs, a type of neural network designed to take sequence data as input and recursively process it in the order of the sequence) can achieve precise predictions of bio-element activity by learning sequence features [16–20]. Generative adversarial networks (GANs, a type of generative deep learning model that consists of two competing components, a generator and a discriminator, which are trained simultaneously to generate realistic synthetic data) can achieve de novo design of bio-elements through adversarial training between generators and discriminators [21–24]. Therefore, AI-guided de novo design and precise activity prediction of bio-elements present new opportunities for bio-element research.

The rational design and activity prediction of bio-elements are crucial steps in synthetic biology research. Activity prediction results provide reference standards and optimization directions for designing bio-elements with ideal performance. Simultaneously, bio-elements obtained through rational design must be evaluated for reliability by predicting and verifying their performance. Therefore, the rational design and activity prediction of bio-elements are often complementary in the development of synthetic biological systems. However, the design of bio-elements based on the "Design–Build–Test–Learn" (DBTL) cycle often relies on extensive experimental trial and error and still lacks efficient characterization methods. In addition, traditional methods for predicting the activity of bio-elements, which explore the "genotype–phenotype" relationship, are often complex, low in accuracy, and highly dependent on relevant biological mechanisms and theories such as thermodynamics and molecular dynamics [3,25,26]. To address these challenges, AI algorithms have garnered widespread attention from researchers [27–30]. For example, Wang et al. achieved the de novo design of *E. coli* promoters using a GAN model [31]. Jores et al. employed a CNN model to accurately predict promoter activity and rationally designed plant promoters with the desired activity [32]. In these studies, AI algorithms effectively simplified the DBTL exploration process, reduced experimental costs, and provided new directions for the design and optimization of nucleic acid elements. Additionally, deep-learning-based models such as AlphaFold have been used to predict protein structures and characteristics, offering new references for the rational design of protein elements [33,34].

One significant application of de novo-designed bio-elements is their efficient and precise adaptation for optimal response performance in transcription factor-based biosensors (TFBs) [35,36]. TFBs can convert target metabolite concentration signals into fluorescence or expression-level signals of metabolic pathways [37]. They have been widely used in target metabolite concentration detection [38], high-throughput screening [39], direction evolution [40], and dynamic regulation [41,42]. To enhance the robustness and reliability of TFB applications, it is necessary to design TFBs with excellent response performance. The evaluation metrics for TFB response performance include dynamic range, detection range, specificity, and sensitivity. However, bio-elements designed using traditional "DBTL" methods often fail to achieve optimal TFBs, commonly exhibiting issues such as inappropriate dynamic range, low sensitivity, and low specificity [35,43] (Figure 1A). Thus, designing bio-elements that are highly compatible with TFBs to optimize their response performance remains a significant challenge. Previous studies have utilized the rational design of bio-elements, such as promoters, RBS, and TFs, to regulate TFB response performance. For instance, d'Oelsnitz et al. constructed a TFB with two CamR binding sites, doubling its dynamic range [44]. Gong et al. introduced mutations to modify the structure of the transcription factor TrpR, increasing the specificity of TFB for tryptophan by over threefold compared with 5-hydroxytryptophan [45]. However, designing bio-elements using traditional experimental methods is not only time-consuming and labor-intensive but also limits further optimization of TFB response performance. AI algorithms, with their ability to uncover underlying rules, are promising technologies for regulating TFB

response performance (Figure 1B). For example, Ding et al. trained a CNN on large datasets associating RBS and TFB dynamic range, constructing a classification model CLM-RDR to achieve intelligent regulation of TFB dynamic range [46]. This demonstrates that AI algorithms can bypass some complex biological mechanisms to quickly and accurately optimize TFB response performance. Therefore, AI-designed bio-elements can be adapted to achieve TFBs with excellent response performance.
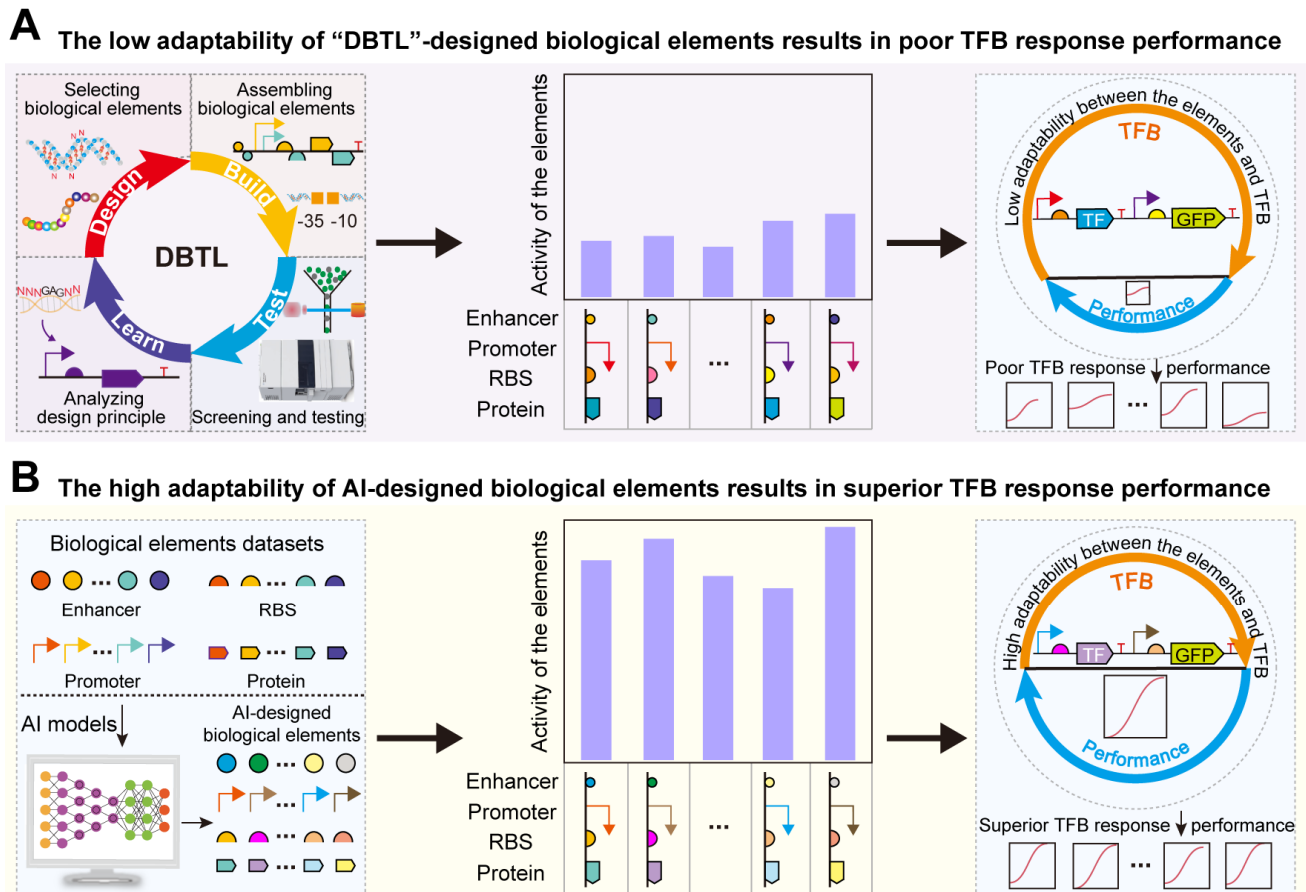
Figure 1. Design of biological elements to regulate the TFB response performance: (**A**) The low adaptability of "DBTL"-designed biological elements results in poor TFB response performances. (**B**) The high adaptability of AI-designed biological elements results in superior TFB response performances.

In this review, we highlight the recent applications and challenges of machine learning and deep learning models in the rational design and activity prediction of bio-elements. We also explore how AI-designed bio-elements can regulate the TFB response performance. By outlining the advantages, adaptability, and biological challenges addressed by AI algorithms, we aim to assist researchers in leveraging AI for bio-element-related research. Furthermore, we propose innovative solutions to overcome these challenges and suggest future research directions. Detailed comparisons of traditional and AI-based methods, their accuracy, and the challenges they address are summarized in Table 1, providing a comprehensive understanding of the current state and future potential of AI in synthetic biology.

Table 1. Challenges and strategies in traditional and AI-based bio-element design and activity prediction.

| Bio-Element	Function	Challenge	Strategy		Accuracy		References
			Traditional	AI	Traditional	AI	
Promoter	Rational design and activity prediction	Small library, vast sequence space	Experimental method	GAN, CNN	ns	0.7	[31]
				DeepSEED (GAN, LSTM)		0.78	[47]
	Activity prediction	High prediction cost and low accuracy	CHIP-seq, RNA-seq	XGBoost		0.88	[48]
			CHIP-seq, RNA-seq	iPro-GAN		0.92	[49]
Enhancer	Rational design and activity prediction	Unclear motif syntax relationships, inadequate compatibility between motifs, and limited applicability	Experimental method	DeepSTARR (CNN), GAN	ns	0.74	[50]
	Activity prediction		CHIP-seq, RNA-seq	iEnhancer-DCLA (CNN, BiLSTM, Attention)		0.83	[51]
RBS	Activity prediction	Demand for larger libraries, cumbersome experimental procedures, and complex thermodynamic analysis data	Experimental method	GPR, Bandit	ns	34% high TIR	[52]
			Ribosome loading, DNA methylation, NGS	CNN		0.927	[53]
Protein	Rational design	Limited sequence space	ns	ProteinGAN (GAN)		0.88	[54]
	Rational design and activity prediction	Limited protein structure types and vast sequence space		WGAN, Rosetta		TM > 0.5	[55]
		Low accuracy and limited accuracy for complex interactions		AlphaFold 2		TM > 0.78	[56]
	Activity prediction			AlphaFold 3		>0.8	[57]
				DLKcat (CNN, GNN)	ns	0.71 (k_{cat})	[58]
	Enzyme catalytic constant prediction	Low accuracy	Experimental method	UniKP (pretrained language models)		0.85 (k_{cat}), 0.73 (k_m), 0.81 (k_{cat}/k_m)	[59]
	Enzyme function prediction	Small and imbalanced datasets		CLEAN (contrastive learning framework)		0.86	[60]

ns, not specified.

2. AI-Based Rational Design and Activity Prediction of Bio-Elements

The rational design and activity prediction of bio-elements are crucial research areas in synthetic biology. AI algorithms have been widely applied to the rational design and precise activity prediction of genetic regulatory elements such as promoters and RBSs (Figure 2), as well as the intelligent design of protein sequences and structures. Combining AI models with traditional biological experiments will enhance the depth of research into bio-elements.

### 2.1. AI-Assisted Rational Design and Activity Prediction of Promoters

Promoters are short DNA sequences located near gene transcription sites specifically recognized and bound by RNA polymerase to initiate transcription, playing a crucial role in regulating gene expression levels [61]. The rational design and activity prediction of promoters are playing a crucial role in synthetic biology and metabolic engineering. However, promoter design faces challenges such as a small library and a vast potential sequence space [31,62]. To address these challenges, AI models have been employed [63]. For example, Wang et al. achieved the de novo design of *E. coli* promoter sequences using a GAN model (Figure 3A) [31]. They first employed adversarial training between the generator and discriminator in a GAN model to learn and extract the latent features of natural promoters, generating a large number of novel artificial promoter sequences [31]. Finally, they predicted the activity of these artificial promoters based on a CNN model, achieving a prediction accuracy of 0.7, demonstrating the excellent performance of AI-designed artificial promoters [31]. Moreover, Zhang et al. developed the DeepSEED framework based on GAN (generator) and Long Short-Term Memory (LSTM, a predictor, is a specialized type of recurrent neural network designed to effectively capture long-term dependencies in sequential data) models, improving the prediction accuracy of promoter

activity to 0.78 by considering the influence of flanking sequences during the design process (Figure 3B) [47].
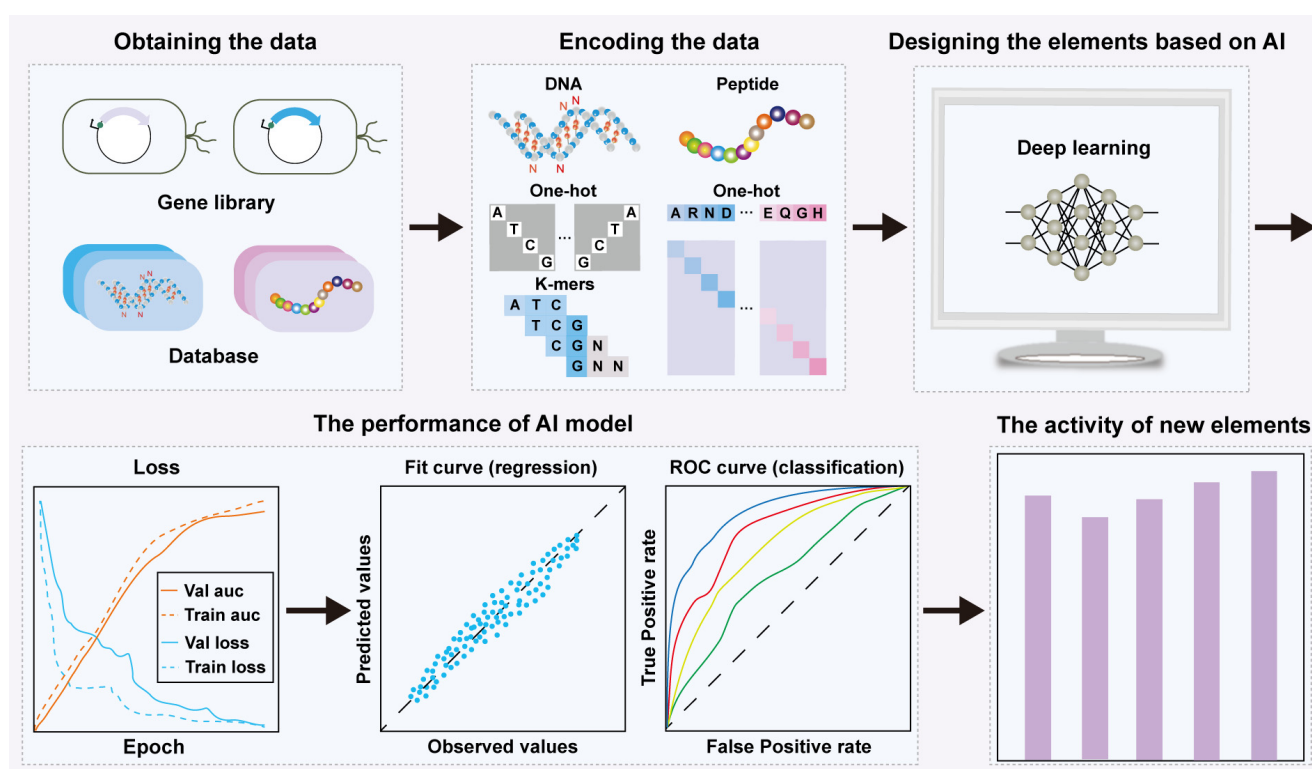
Figure 2. AI-designed biological elements and their activity prediction. Loss: an indicator for evaluating the training process and testing process of the model; Val and Train auc: an index used to evaluate the model accuracy of the validation process and training process, respectively; Val and Train loss: an index used to evaluate the loss of model validation process and training process, respectively; ROC and Fit curve: indicators for evaluating the performance of the classification and regression model, respectively.

Predicting promoter activity also faces challenges such as high costs and low accuracy. To address these issues, Zhao et al. proposed a promoter strength prediction method based on the eXtreme Gradient Boosting model (XGBoost, a powerful machine learning algorithm based on decision trees, optimized for efficiency, accuracy, and handling large datasets through gradient boosting techniques) (Figure 3C) [48]. By constructing a gradient transcription intensity Trc promoter (Ptrc) mutant library dataset based on the mutation–construction–screening–characterization (MCSC) cycle, the XGBoost model was trained, achieving a prediction accuracy of 0.88 [48]. Similarly, Qiao et al. developed the iPro-GAN model, which used spatial data analysis to extract the sequence features based on the Moran model and employed the deep convolutional GAN model to achieve high-precision predictions of promoter transcription intensity, ultimately reaching a prediction accuracy of 0.92 [49]. These examples demonstrate AI algorithms' potential in the rational design and precise activity prediction of promoters.
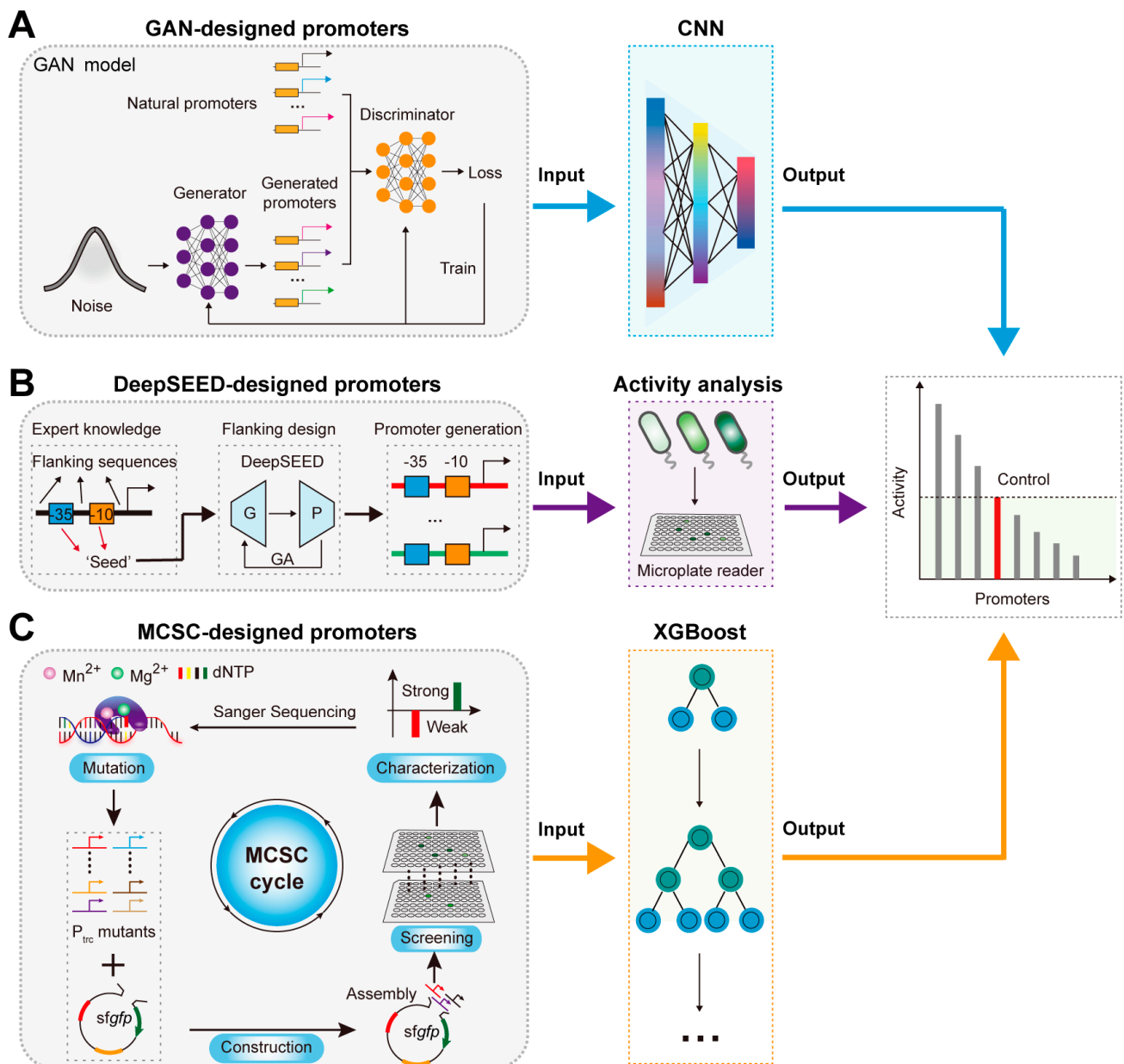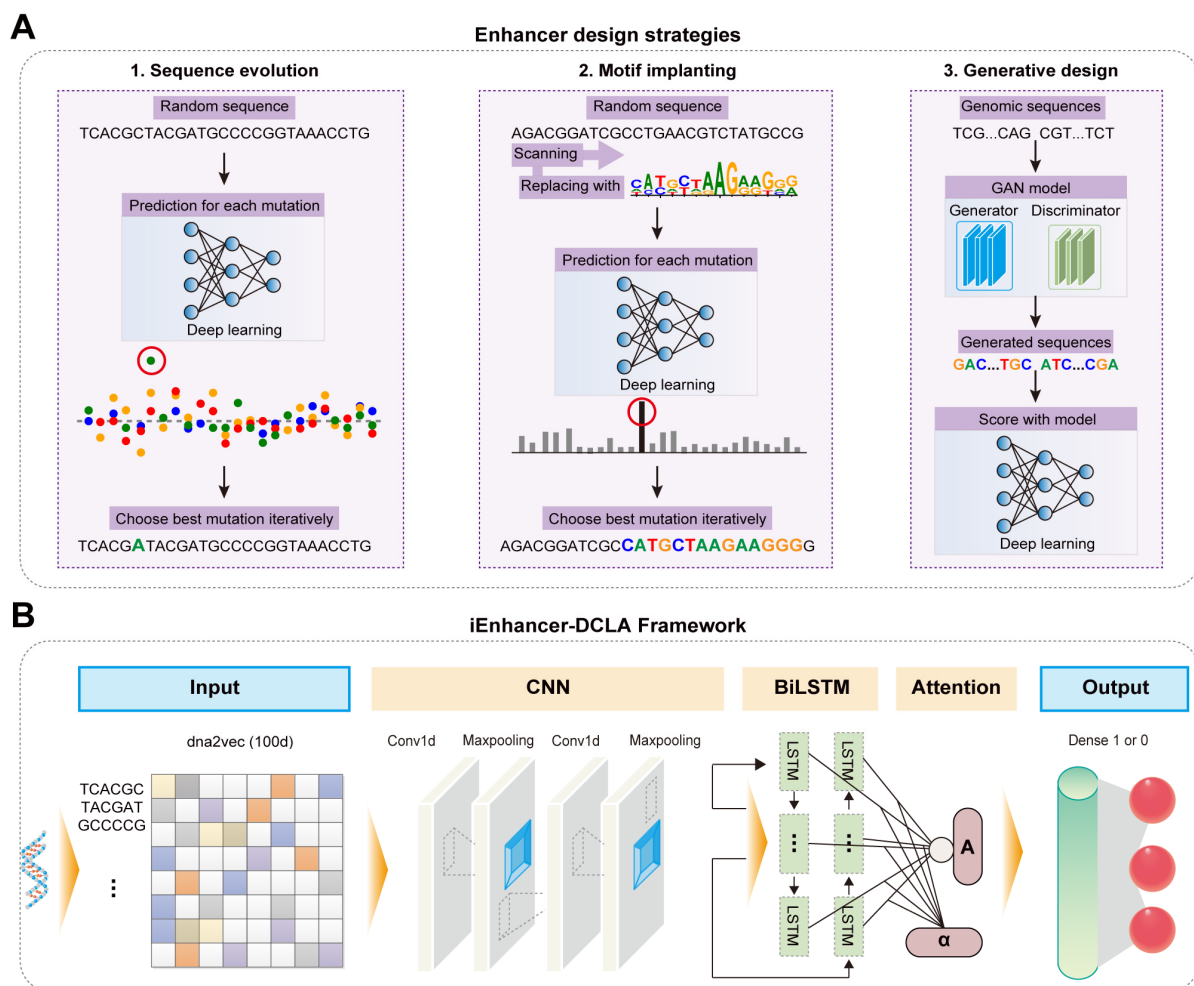
Figure 3. AI–derived promoter design and activity prediction: (**A**) The activity prediction of GAN–designed promoters based on the CNN model. (**B**) The activity prediction of DeepSEED–designed promoters based on activity analysis. G, P, and GA represent the Generator, Predictor, and Genetic algorithm. (**C**) The activity prediction of MCSC–designed promoters based on the XGBoost model.

2.2. AI-Assisted Rational Design and Activity Prediction of Enhancers

Enhancers, typically containing multiple motifs such as transcription factor binding sites (TFBS), are short regions within eukaryotic sequences that can bind to proteins and enhance gene transcription, playing a crucial role in regulating gene expression [64,65]. However, the de novo design and activity prediction of enhancers is challenging due to the unclear relationship between motif syntax and enhancer activity, the inadequate compatibility between motifs, and the limited applicability [66]. To solve these challenges, Almeida et al. developed DeepSTARR, a CNN-based deep learning framework, achieving efficient prediction of enhancer activity in Drosophila S2 cells with a Pearson Correlation Coefficient (PCC, a statistical measure that evaluates the strength and direction of the linear relationship between two variables, often used to assess the performance of a model by comparing predicted values to actual outcomes) of 0.74 [50]. Experimental validation based

on the prediction results revealed some motif syntax rules. Using these rules, they designed novel enhancer sequences with gradient activities ranging from 0.8 to 630 [50]. Taskiran et al. further realized the de novo design of synthetic enhancers based on cell types using various AI algorithms (Figure 4A) [67]. First, they used AI algorithms for directed sequence optimization and the insertion of TFBS into enhancers, analyzing the impact of changes in repressive sites and TFBS on enhancer activity [67]. Then, they compared Drosophila enhancers designed using these strategies with human enhancers generated by a GAN model, demonstrating the applicability of these design strategies to different biological systems [67]. In addition, Liao et al. developed iEnhancer-DCLA, a novel deep learning framework, to predict enhancer activity (Figure 4B) [51]. First, they encoded sequences using data encoding methods such as word embedding, one-hot encoding, and k-mers to determine the most suitable approach for enhancer sequences [51]. Then, they combined algorithms like CNN, LSTM, and attention mechanisms to thoroughly extract sequence features [51]. The model achieved an accuracy of 0.83 in predicting enhancer activity. Thus, AI algorithms have shown significant potential in the rational design and precise activity prediction of enhancers [68,69].

Figure 4. AI-derived enhancer design and activity prediction: (**A**) The overview of enhancer design strategies based on deep learning. (**B**) The model structure of enhancer activity prediction based on iEnhancer-DCLA. It includes feature representation based on the dna2vec method, two convolutional and maxpooling layers, a bidirectional LSTM network layer, an attention layer, and two fully connected layers. A represents the feature vector after passing through the attention mechanism layer; α represents the importance of the output of the bidirectional LSTM layer.

2.3. AI-Assisted Rational Design and Activity Prediction of RBS

RBS is an untranslated region upstream of the mRNA start codon, recognized and bound by the ribosome to initiate translation. RBS is crucial for translation initiation and gene expression. The rational design of RBS faces challenges due to the increasing demand for larger libraries, cumbersome experimental procedures, and complex thermodynamic analyses [25,70,71]. To address these challenges, Zhang et al. proposed a machine learning-guided DBTL cycle method for designing bacterial RBS, using the Bandit algorithm to design RBS and the Gaussian process regression (GPR, a nonparametric Bayesian regression method that provides probabilistic predictions of the output by assuming a Gaussian process prior over functions, allows it to capture uncertainty and make predictions with confidence intervals) algorithm to predict the translation initiation rate (TIR) of the designed RBS [52]. The method showed that 34% of the designed RBSs had TIR values not lower than the standard RBS, demonstrating AI algorithms' potential in RBS sequence design [52]. Subsequently, Simon et al. used the DNA phenotyping method uASPIre (ultra-deep Sequence-Phenotype Interrelationship acquisition) and Next Generation Sequencing (NGS) technology to obtain large datasets of sequence-activity associations [53]. They then trained a CNN model based on these datasets to achieve a precise prediction of RBS activity with an accuracy of 0.927 [53]. This provided an efficient new approach for the precise prediction of RBS activity.

2.4. AI-Assisted Design of Protein Sequences and Structures and Prediction of Functional Activity

The de novo design of proteins based on deep learning algorithms can generate novel proteins with expected functions and has been widely applied in protein engineering and synthetic biology [72–75]. AI-based protein design can be categorized into structure-based design [76,77] and direct sequence design [57,78–80]. Despite the rapid accumulation of protein sequence and structure data, the limited number of protein structure types and the vast sequence space remain significant challenges. Using limited data to understand protein folding principles and optimize sequences is a key bottleneck. To overcome these challenges, Karimi et al. used a WGAN (an improved version of GAN by using the Wasserstein distance as a loss function, which enhances training stability and generates higher-quality data) model to obtain low-dimensional representations of the protein folding space [55]. Then, they predicted the structures of the generated sequences based on the Rosetta predictor, achieving high TM scores (TM > 0.5), and indicating accurate folding of AI-designed proteins [55]. Recent studies also focus on using AI to explore sequence space and design proteins directly [72,78]. ProteinGAN, developed by Donata et al., explored the complex multidimensional amino acid sequence space and learned the diversity of natural sequences, generating new sequences with natural physical characteristics [54]. Experimental validation showed high matrix similarity (88%) between the generated and natural sequences, with 24% functionality, indicating that ProteinGAN successfully captured the local amino acid relationships in protein sequences [54]. This demonstrates the potential of AI algorithms to rapidly design diverse functional proteins within a limited sequence space.

AlphaFold has revolutionized protein structure prediction based on amino acid sequences [56,78,79,81]. For example, AlphaFold 2, developed by Jumper et al., combines multiple sequence alignments (MSA) and neural networks to achieve near-experimental accuracy in predicting 3D protein structures (Figure 5A) [56]. The model integrates evolutionary, physical, and geometric information, resulting in significantly improved median backbone accuracy as demonstrated in the 14th Critical Assessment of Protein Structure Prediction (CASP14) [82]. AlphaFold 2 achieved a high TM-score of over 0.78 [56]. Despite its advancements, AlphaFold 2 faced challenges such as limited accuracy in predicting complex biomolecular interactions and computational efficiency. To address these challenges, Abramson et al. developed AlphaFold 3, which integrates a diffusion-based architecture to predict the joint structure of complexes, including proteins, nucleic acids, and small molecules, improving accuracy by reducing the complexity of MSA processing and directly

predicting raw atom coordinates through a diffusion module (Figure 5B) [57]. AlphaFold 3 achieved an unprecedented accuracy of over 0.8 across various benchmarks, such as protein–ligand and protein–nucleic acid interactions [57]. Thus, AI algorithms provide a crucial direction for the de novo design of proteins.
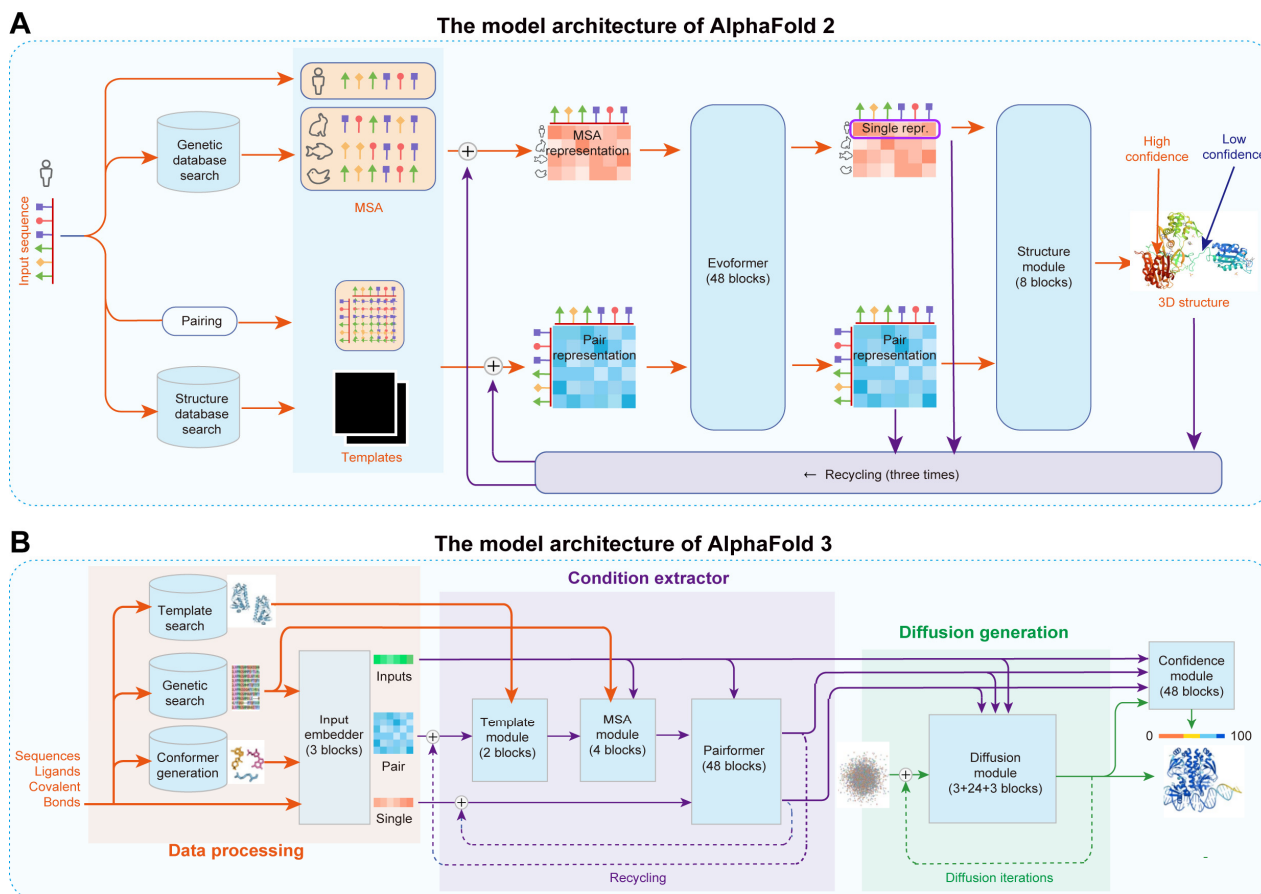
Figure 5. The model architecture of AlphaFold 2 and 3: (**A**) AlphaFold 2 achieves end-to-end prediction of protein 3D structures from protein sequences. The model's framework is divided into three main steps. First, the protein sequence is input into the model and searched against gene and structure databases to obtain homologous sequences and template structures through MSA. Next, the predicted sequence, MSA sequences, and template structures are embedded and processed by the Evoformer module to generate MSA representation and pair representation. Finally, the first row of the MSA representation and the pair representation are used as inputs for the structure module to predict the 3D structure of the protein sequence. repr., representation. (**B**) Accurate structure prediction of biomolecular interactions based on AlphaFold 3. The AlphaFold 3 model consists of three main steps: data processing, condition extraction, and diffusion generation. First, the required structural prediction information is inputted. Based on this input, the system searches and generates data from databases, deriving information from genetic searches, template searches, and conformer generation. The input information and derived reference information are then fed into the Input Embedder, which performs initial encoding to obtain inputs, single representations, and pair representations. Then, a condition extractor with a recycling mechanism is used to integrate template and MSA information into the pair representation through the Template and MSA Module. The Pairformer then merges and refines the single representation and pair representation information. The inputs, single representations, and pair representations are used as conditions for the diffusion model, which constrains and controls the denoising process to generate refined results. Finally, the predicted results are fed into the Confidence Head to predict the confidence level of the structural prediction.

AI algorithms also enhance enzyme activity and function prediction. For enzyme activity prediction, Li et al. developed DLKcat based on CNN and Graph Neural Network (GNN, a deep learning model designed to process and analyze data structured as graphs, consisting of nodes and edges), achieving 0.71 accuracy in predicting the catalytic constant k_{cat} [58]. Yu et al. developed the UniKP framework based on pretrained language models, improving the prediction accuracy of k_{cat} to 0.85 and achieving 0.73 and 0.81 accuracy in predicting the k_m and k_{cat}/k_m, respectively (Figure 6A) [59]. For enzyme function prediction, previous methods often relied on sequence similarity and homology, and some model-based prediction methods were limited by small and imbalanced datasets. To address this problem, Yu et al. developed the CLEAN model using a contrastive learning framework to predict the catalytic functions of different enzymes (Figure 6B) [60]. This model treats the four-digit code of known enzymes as a matrix and uses Euclidean distance to represent functional similarity between different enzymes, ultimately outputting a ranked list of enzyme functions by probability [60]. This approach achieved a precise prediction accuracy of enzyme functions, with a prediction accuracy of over 0.86 [60]. Therefore, AI models provide a new direction for high-precision prediction of enzyme activity and function.
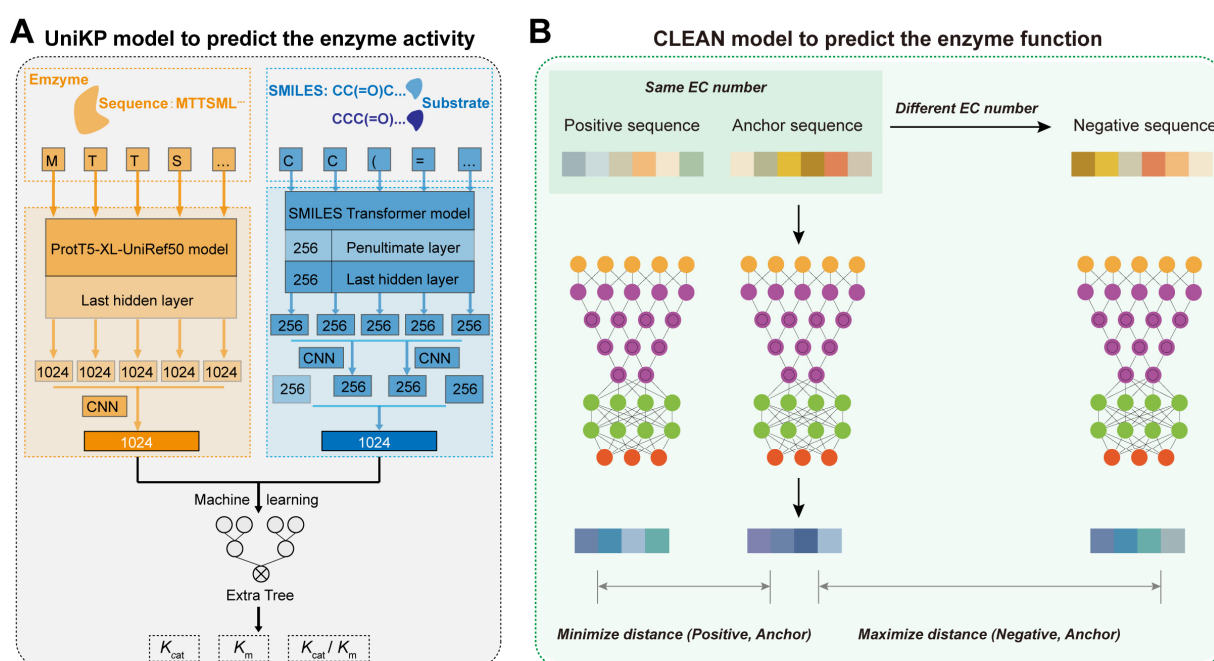
Figure 6. AI-derived prediction of enzyme activity and function: (**A**) The overview of UniKP to predict the enzyme kinetic parameters. First, the pretrained language model, ProtT5-XL-UniRef50, is used to encode enzyme information. Each amino acid is transformed into a 1024-dimensional vector in the last hidden layer. These vectors are then averaged using mean pooling to generate a 1024-dimensional vector representing the enzyme. Next, the pretrained language model, the SMILES transformer, is used to encode substrate information. The substrate structure is converted into a Simplified Molecular Input Line Entry System (SMILES) representation and fed into the SMILES transformer to generate a 1024-dimensional vector. This vector is created by concatenating the average and max pooling of the last layer and the first outputs of the last and penultimate layers. Finally, an interpretable Extra Trees model based on machine learning uses the concatenated representation vectors of the enzyme and substrate as input to predict the k_{cat}, k_m, or k_{cat}/k_m values. (**B**) The overview of CLEAN for the prediction of enzyme function. During training, positive and negative samples are selected based on EC numbers. The input sequences are embedded and processed through a neural network. The warm-colored grid series represents the embeddings of the input sequences from ESM-1b. Similarly, the embeddings obtained from the supervised contrastive learning neural network are depicted in cool colors.

3. Optimizing the TFB Response Performance Based on AI-Designed Biological Elements

The optimization of TFB response performance based on the rational design of bio-elements is a key direction in current research. Bio-elements such as promoters, RBS, and transcription factors are crucial regulatory targets in the study of TFB response performance [35,83,84]. AI algorithms have emerged as a novel method for the rational design of bio-elements to achieve optimal TFB response performance.

3.1. AI-Designed Promoters for Regulating TFB Response Performance

Promoters can regulate TFB response performance by controlling the transcription rate of transcription factors and reporter proteins [36,85]. Promoter engineering has been widely used to regulate the dynamic range and sensitivity of TFBs [44,86,87]. Strategies for fine-tuning TFB response performance include modifying specific sites of transcription factor-responsive promoters and performing site-directed mutation on promoters [44,86,88]. However, traditional promoter engineering to optimize TFB response performance is often time-consuming, labor-intensive, costly, and limited TFB application scope. In addition, the numerous combinations of promoter motifs can affect TFB response performance due to various factors such as the bio-element activity, the metabolism, and the growth of host cells. Thus, traditional trial-and-error methods face significant challenges in regulating TFB response performance.

To address these challenges, Zhou et al. synthesized a large dataset of gradient intensity promoters based on DNA barcode technology to construct a TFB library (Figure 7A) [89]. They characterized the TFB response curves using fluorescence-activated cell sorting (FACS) and NGS sequencing (FACS-seq) technologies [89]. Subsequently, they used the XGBoost machine learning model to achieve accurate genotype-to-phenotype predictions [89]. Finally, they experimentally validated the sequences with superior performance based on the prediction results, obtaining a malonyl-CoA biosensor with a maximum dynamic range of 6.38 [89]. This provides an efficient and cost-effective new approach for regulating and optimizing TFB response performance.

3.2. AI-Designed RBS for Regulating TFB Response Performance

RBS can regulate the TFB dynamic range by modulating the translation levels of transcription factors and reporter genes as well as protein folding [36,46]. Although RBS engineering has been widely applied in research on regulating the TFB dynamic range [44,90,91], obtaining the corresponding TFB dynamic range for different RBSs still relies on time-consuming and costly experimental studies. Moreover, there is a lack of precise prediction techniques to explore the relationship between RBS sequences and TFB dynamic range. To solve these problems, Ding et al. developed a platform using a deep learning model, CNN, for intelligent prediction of TFB dynamic range based on RBS sequences (Figure 7B) [46]. They first obtained large datasets of glucarate biosensor dynamic range and RBS associations using DNA microarray and FACS-seq technology [46,92]. Then, they built the CLM-RDR platform based on a CNN model to accurately predict the glucarate biosensor dynamic range from RBS sequences, achieving a prediction accuracy of 0.86 [46]. The CLM-RDR platform simplified the workload of the DBTL cycle and enabled precise regulation of TFB dynamic range based on AI-screened RBS sequences.
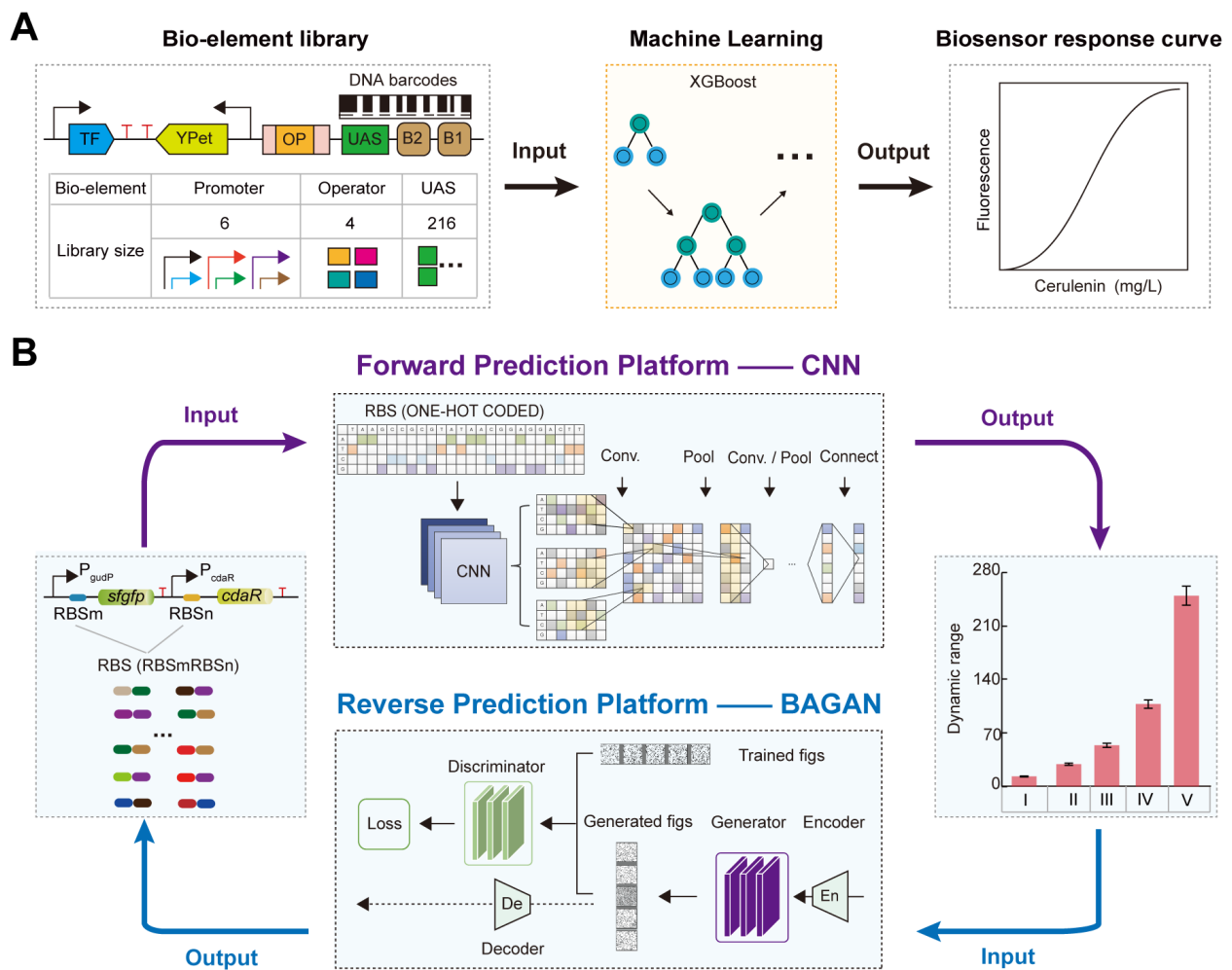
Figure 7. AI-designed bio-elements to fine-tune the TFB response performance: (**A**) The regulation of the cerulenin biosensor response curve based on the XGBoost model. (**B**) The forward–reverse prediction platform to fine-tune the TFB dynamic range. The forward engineering platform to precisely predict the TFB dynamic range based on the CNN model. The reverse engineering platform to rationally design the RBS with the desired TFB dynamic range based on the BAGAN-GP model.

Furthermore, to rationally design RBS with a desired TFB dynamic range, Ding et al. developed a forward and reverse engineering platform for TFB intelligent design based on CNN and GAN-derived models (Figure 7B) [93]. The forward engineering used the Wasserstein GAN model with gradient penalty (WGAN-GP, an improved version of the Wasserstein GAN that incorporates a gradient penalty term to enforce the Lipschitz constraint, leading to more stable training and better quality generated data compared with the original WGAN) to generate a large functional RBS dataset and predicted the TFB dynamic range from the generated RBS using a CNN model, achieving a prediction accuracy of 0.98 [93]. The reverse engineering used the balanced GAN model (BAGAN-GP, a GAN model designed to restore data balance from unbalanced data sets, incorporating a gradient penalty to improve training stability and the quality of generated data) to de novo design RBS sequences based on a given TFB dynamic range, with a design accuracy of 0.82 [93]. These results indicate that deep learning algorithms have become a crucial tool for the rational design of RBS to optimize TFB dynamic range by exploring the relationship between genotype and phenotype.

3.3. AI-Optimized Transcription Factor Regulating the Dynamic Range of TFB

Transcription factors (TFs) are protein molecules that regulate the expression of target genes. Research has shown that the expression levels of TFs and their binding affinity to ligands or target sequences are critical factors affecting TFB response performance. Low TF expression levels can reduce TFB sensitivity and dynamic range, while excessively high TF expression levels can permanently activate or inhibit target gene expression [94]. Additionally, the binding ability of TFs to ligands or DNA affects TFB fluorescence output and dynamic range [85]. Moreover, the regulatory patterns of TFs and metabolites within cells are key factors affecting TFB response performance [35]. Trabelsi et al. constructed a TFB library that included gradient concentrations of the FdeR, the number of binding sites for the activation complex, and plasmid copy numbers [95]. Using a Hill function fitting model, they analyzed the impact of FdeR expression levels and FdeR-ligand binding affinity on TFB response performance, successfully increasing the dynamic range of a naringenin TFB to 60 [95]. This provided new insights for TF design and TFB response performance regulation.

Currently, research on rationally designed TF to regulate TFB response performance has not fully integrated AI algorithms. However, deep learning models have shown excellent performance in analyzing TF characteristics and TFBS interactions [96–98]. Simultaneously, AI algorithms like AlphaFold for designing functional proteins have created new opportunities for constructing desired TF. Thus, using AI algorithms to optimize and design TFs provides new possibilities for regulating TFB response performance.

4. Applications of Optimized TFB

In recent years, TFBs have gained significant attention in the production of compounds within microbial cell factories [35]. The primary applications of TFBs in metabolic engineering include (Figure 8): (1) detection of metabolite concentrations [99]; (2) high-throughput screening of high-yield strains for target metabolites [37,43,100]; (3) directed evolution [37,100,101]; and (4) dynamic regulation of microbial intracellular metabolism [37,102–104]. Optimized TFBs with superior response performance are crucial for enhancing the robustness and reliability of these applications.

4.1. Real-Time Detection of Target Metabolite Concentrations

Real-time detection of intracellular metabolite concentrations is crucial for optimizing cellular biosynthetic processes. Recently, TFBs have been utilized for this purpose (Figure 8A). For example, Baumann et al. developed a TFB in *Saccharomyces cerevisiae* (*S. cerevisiae*) based on the transcription factor War1p, the PDR12 promoter responsive to short-chain and medium-chain fatty acid (SMCFA), and the reporter gene *gfp* [99]. This system allows for easy and rapid detection of SMCFA, serving as an alternative to traditional gas chromatography methods [99]. The TFB exhibited linear responses to hexanoic, heptanoic, and octanoic acids within the concentration ranges of 0.01–2.00 mM (R^2 = 0.98), 0.01–1.50 mM (R^2 = 0.99), and 0.01–0.75 mM (R^2 = 0.99), respectively [99]. Consequently, this TFB has the potential to significantly accelerate the engineering of cell factories for the production of various SMCFAs. However, current detection systems often have limited detection ranges and fail to show concentration-dependent fluorescence changes when target metabolite concentrations exceed millimolar levels, thus restricting their further application in metabolic engineering and synthetic biology [105,106]. Therefore, designing TFBs with the desired performance is crucial for detecting higher metabolite concentrations.
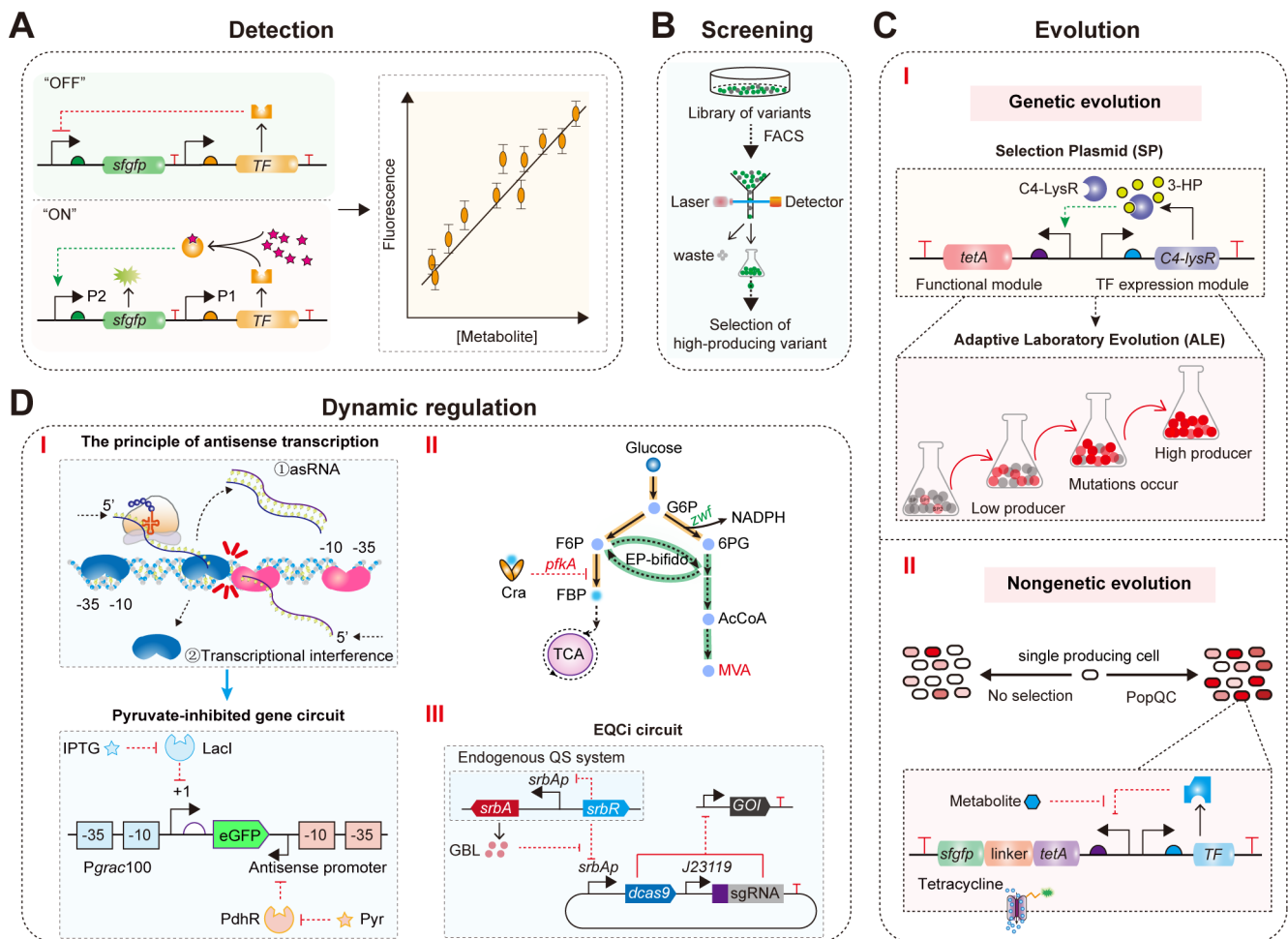
Figure 8. The applications of TFB in metabolic engineering and synthetic biology: (**A**) TFB–derived metabolite concentration detection in real time. The pink stars represent metabolites. (**B**) TFB–derived high-throughput screening of high-titer strain. (**C**) TFB–derived direction evolution of high producer. (**I**) Schematic diagram of using a selection plasmid (SP) as a synthetic biosensor to achieve Adaptive Laboratory Evolution (ALE). The plasmid includes an expression module for the transcription factor C4–LysR and a functional module containing the selection marker *tetA*; (**II**) The design principle of PopQC. PopQC endows nongenetic high–performance cells with a growth advantage, increasing their proportion within the overall population. A metabolite–responsive transcription factor regulates the expression of the tetracycline efflux protein (encoded by tetA). In the presence of tetracycline, high–performance cells outcompete low–performance cells and dominate the population. (**D**) TFB–derived dynamic regulation of microbial intracellular metabolism. (**I**) The schematic diagram of antisense transcription and construction of pyruvate–inhibited gene circuit. The blue and pink irregular patterns represent RNA polymerase initiating transcription from the sense and antisense promoter, respectively. The constitutive promoter (red) at the 3′ end of eGFP suppresses gene expression by triggering RNA polymerase at the sense promoter Pgrac100 (blue). Pgrac100, an IPTG–inducible promoter; LacI, a transcriptional regulator in the *E. coli* lactose metabolism pathway; blue box, the core region of the sense promoter; red box, the core region of the antisense promoter; PdhR, a pyruvate–responsive transcriptional regulator. (**II**) Dynamic control of *pfkA* in EP–bifido strain using a glycolytic flux biosensor for high MVA production. (**III**) The architecture of the genetic circuit based on EQCi. The EQCi genetic circuit is constructed using pSET as the backbone, with the expression of dCas9 driven by the *srbAp* promoter and the transcription of sgRNA targeting the gene of interest (GOI) driven by the strong synthetic promoter J23119 from *E. coli*.

4.2. High-Throughput Screening of High-Titer Strains for Target Metabolites

TFBs are not only used to detect intracellular metabolite concentrations but are also widely applied in the screening of high-titer strains for target metabolites [36,100,107]. TFBs can be used in conjunction with FACS to rapidly screen high-titer strains from extensive libraries by detecting the output signal of fluorescent reporter genes (Figure 8B) [37,83,100]. For example, Kortmann et al. constructed a TFB responsive to L-lysine based on LysG and used it with a FACS screening system to identify pyruvate carboxylase mutants in *C. glutamicum* [108]. This approach improved the ability of *C. glutamicum* to produce L-lysine from glucose [108]. When *C. glutamicum* produces high levels of L-lysine, LysG senses the L-lysine concentration and activates the expression of a fluorescent protein, generating a fluorescence signal [108]. By screening the pyruvate carboxylase mutant library, two mutants that significantly increased L-lysine production in host cells were identified, leading to L-lysine levels increasing by 6% and 14%, respectively [108]. Similarly, Ding et al. developed a TFB responsive to glucaric acid (GA) based on CdaR and used it with FACS to screen for *myo*-inositol oxygenase (MIOX) mutants with high stability and activity, a key rate-limiting enzyme in the GA biosynthesis pathway [92]. This approach increased GA titer to 5.52 g/L in 5 L fermenter cultures, the highest titer reported in *E. coli* to date [92]. These successful cases demonstrate that TFBs can be effectively integrated with mainstream high-throughput screening methods. However, the expression of fluorescent proteins often imposes a metabolic burden on cells, affecting cell growth and potentially leading to bias in FACS screening [109]. Therefore, using antibiotic-resistant genes instead of fluorescent proteins for screening high-titer strains may alleviate this issue.

4.3. Directed Evolution

TFB-mediated directed evolution is a powerful strategy for the efficient production of target metabolites [40,110,111]. Optimized TFBs can enrich high-titer strains by responding to target metabolites and activating or inhibiting downstream gene expression [101,112–114]. For example, Seok et al. developed a synthetic biosensor responsive to 3-hydroxypropionic acid (3-HP) based on the C4-LysR biosensor and the TetA bioselector (Figure 8C(I)) [101]. Using a glycerol-dependent 3-HP production pathway as a model system, they performed adaptive laboratory evolution (ALE) to identify the optimal flux redistribution between the 3-HP biosynthesis pathway and the central carbon metabolism pathway, increasing the 3-HP titer and reducing acetate accumulation by alleviating overflow metabolism [101]. These results demonstrate that whole-genome evolution using synthetic biosensors can lead to effective carbon flux rewiring. Additionally, Shen et al. used a 4-hydroxyphenylacetic acid (4HPAA) biosensor combined with atmospheric and room temperature plasma (ARTP) mutagenesis and ALE to successfully obtain strains with high 4HPAA titer and tolerance [115]. The strains maintained genetic stability after 25 generations of genome shuffling [115]. Ultimately, strain GS-2-4 produced 25.42 g/L 4HPAA in a 2 L fed-batch culture bioreactor [115]. These results indicate that the strain has long-term genetic stability and high production levels, making it a potential candidate for industrial applications. Moreover, Tong et al. constructed a TFB responsive to (2S)-naringenin in *E. coli* based on TtgR [116]. Through directed evolution and saturation mutagenesis, they identified a chalcone synthase (CHS) mutant, *Sj*CHS1^{S208N}, with 2.34-fold increased catalytic activity [116]. Fermentation in a 5 L bioreactor increased the de novo (2S)-naringenin concentration to 2513 $\pm$ 105 mg/L, the highest concentration reported in a stirred batch bioreactor to date [116]. Overall, these directed evolution strategies can be broadly applied to engineer biochemical production pathways without the need for labor-intensive procedures.

In addition, natural heterogeneity caused by nongenetic factors exists between cells at the protein and metabolite concentration levels [117]. Previous studies have shown that genetic heterogeneity in industrial fermentation processes can lead to production burdens due to metabolic load and toxicity, which negatively impact titers [117]. To mitigate the effects of fermentation heterogeneity on metabolite production, Xiao et al. developed

a Population Quality Control (PopQC) system based on the FadR biosensor and TetA bioselector (Figure 8C(II)) [118]. This system continuously enriches high-producing cells and eliminates inefficient ones, resulting in a threefold increase in fatty acid production titer [118]. Similarly, Ding et al. constructed a PopQC system for GA production based on the CdaR biosensor and TetA [92]. High intracellular GA levels trigger the GA biosensor to express TetA, providing a growth advantage to high GA-producing cells under tetracycline selective pressure, ultimately increasing the GA production titer to 5.52 g/L in a 5 L fermenter [92]. Therefore, TFBs are invaluable for studying and controlling metabolic heterogeneity.

4.4. Dynamic Regulation of Microbial Intracellular Metabolism

Using TFBs to dynamically regulate intracellular gene metabolism levels in response to intracellular metabolic states can simulate the naturally occurring metabolic regulatory networks of microbes. This approach can prevent the excessive accumulation of toxic metabolic intermediates and balance the supply of precursors needed for cell growth with the biosynthesis of target metabolites [103,104,119,120]. For example, Zhou et al. designed a TFB based on FdeR and PadR that responds simultaneously to (2S)-naringenin and p-coumaric acid [121]. They used this biosensor to control the synthesis and consumption of malonyl-CoA [121]. Low concentrations of (2S)-naringenin direct malonyl-CoA towards the fatty acid biosynthesis pathway, promoting cell growth [121]. High concentrations of (2S)-naringenin inhibit the fatty acid biosynthesis pathway, slowing cell growth and increasing the availability of malonyl-CoA for producing more (2S)-naringenin [121]. Ultimately, this multilayer dynamic regulatory network increased the titer of naringenin by 8.7-fold [121]. This indicates that dynamic regulation is a promising strategy for fine-tuning metabolic flux in microbial cell factories. However, these synthetic regulatory systems are rarely developed for central carbon metabolites and can only activate or inhibit the expression of target genes, failing to achieve dual-functional dynamic regulation of metabolic pathways. To enable dynamic dual control (activation and inhibition) for central metabolism, Xu et al. constructed a bifunctional pyruvate-responsive biosensor using the PdhR from *E. coli* and an antisense transcription-based "NOT" gate for signal conversion (Figure 8D(I)) [102]. By dynamically upregulating the *ino1* gene and downregulating the *zwf* and *pgi* genes, they increased glucaric acid production from 207 mg/L to 527 mg/L [102]. Additionally, Zhu et al. designed and constructed a bifunctional glycolytic flux biosensor (Figure 8D(II)) [122]. They modified promoters and transcriptional regulators to obtain highly responsive activation and inhibition biosensors for dynamic control of glycolytic flux [122]. Using this biosensor, they upregulated the expression of *zwf*, encoding glucose-6-phosphate dehydrogenase, and downregulated *pfkA*, encoding phosphofructokinase (Figure 8D(II)) [122]. This ultimately increased the mevalonate production titer in *E. coli* to 111.3 g/L in a 1 L fermenter [122]. Thus, bifunctional biosensors are effective tools for dynamically controlling central metabolism in microbial cell factories.

However, the currently available dynamic regulatory elements are very limited, and many metabolites lack specific responsive transcription factors. Therefore, developing convenient, universal, and self-driven dynamic control systems is of great significance for the efficient biosynthesis of target metabolites in microbes. For example, Tian et al. developed a novel dynamic regulation system, EQCi (Endogenous Quorum-sensing (QS) system with Clustered Regularly Interspaced Short Palindromic Repeats interference (CRISPRi)), in *Streptomyces* (Figure 8D(III)) [123]. This system uses a γ-butyrolactone (GBL) signal molecule-responsive promoter to drive the expression of the dCas9, coupling the QS system with gene transcription inhibition technology (CRISPRi) [123]. It allows for fully automated and precise dynamic control of multiple genes in the metabolic pathway [123]. Using the EQCi system, they constructed a rapamycin-producing recombinant strain [123]. By down-regulating key genes in the tricarboxylic acid cycle, fatty acid biosynthesis, and shikimate pathways, they increased the precursor supply for rapamycin biosynthesis, improving the production titer [123]. They then used the EQCi system for combined intervention in

the metabolic flux of the three pathways and fine-tuning of control strength at each node, resulting in an optimized engineered strain with a rapamycin titer of 1836 ± 191 mg/L, approximately 6.6 times higher than the natural strain [123]. This indicates that the EQCi system effectively balanced the metabolic flux distribution between primary metabolism and product biosynthesis (secondary metabolism), providing an efficient and universal optimization strategy for constructing cell factories of important secondary metabolites derived from *Streptomyces*.

5. Conclusions and Perspective

To deepen our understanding of how bio-elements regulate cellular metabolism and apply these insights to metabolic engineering and synthetic biology, researchers have increasingly focused on the rational design and activity prediction of bio-elements, as well as the optimization of TFB response performance using rationally designed bio-elements. The integration of AI technology with bio-element engineering has matured, leading to the widespread application of machine learning and deep learning algorithms in bio-element research. AI algorithms have shown excellent performance in the high-quality design and precise prediction of nucleic acid and protein elements, as well as the efficient optimization of TFB response performance [124,125]. Deep generative models, such as GANs, have emerged as important tools in designing expected nucleic acid and protein sequences due to their ability to de novo generate novel sequences [126]. The remarkable application of AI in the field of bio-elements demonstrates its powerful potential in uncovering biological characteristics and designing biological systems. This paves the way for scalable, automated, engineered, and end-to-end intelligent prediction and design.

AI model training often depends on substantial high-quality, standardized biological data, especially for deep learning models [127]. For example, extensive datasets that associate bio-element sequences with their activity or structure are indispensable in predicting the activity or structure of bio-elements. However, obtaining large amounts of novel high-quality biological data is often costly, and the data frequently contain indistinct features and significant noise, which complicates model performance optimization. Therefore, the lack of high-quality biological data is a significant challenge that must be addressed for the effective application of AI in bio-element research. Future deep learning models may need to extract features deeply and achieve precise predictions and high-quality bio-element generation from relatively small datasets. Additionally, AI algorithms face diverse challenges in model construction and optimization across biological problems and researchers. On one hand, the functionalities and characteristics of the selected models need to match the specific data structures and research objectives. For example, RNN and Transformer models can handle sequence data [128,129], while GCN models excel at processing topological graph structures [130,131]. CNN models are adept at handling prediction tasks, and GAN models are proficient in de novo design of bio-elements. On the other hand, high-performing AI models often rely on the selection of the optimal evaluation metrics and fine-tuning of model parameters. Thus, the deep integration of AI algorithms with bio-element research requires researchers to have a solid understanding of interdisciplinary knowledge.

Due to the high complexity of cellular systems, data feature extraction techniques must be capable of precisely capturing highly ambiguous and low-precision features. AI models must not only achieve high performance in actual prediction and design tasks but also possess sufficient generalization ability and interpretability to handle diverse and challenging tasks. However, many models with strong predictive or generative capabilities still have problems with the biological interpretability of their computational processes and outputs. Therefore, future applications require the integration of AI model mechanisms and biological theories to enhance model performance while improving interpretability. In terms of model interpretability, Zheng et al. developed the NeuronMotif neural network interpretation algorithm, which can learn and summarize gene regulatory sequence coding rules from neurons, providing a method to interpret the pattern recognition of CNN

models [132]. As interdisciplinary fields continue to develop, an increasing number of standardized databases are becoming available on shared platforms. The future emergence of the World Wide Web will further reduce the difficulty of acquiring high-quality data. The advent of deep learning models such as EfficientNet [133], Swin-Transformer [134], and LLMs [135] will further enhance the performance and efficiency of AI algorithms in the design and prediction of bio-elements.

In the future, several key areas will shape the field of AI-assisted rational design and activity prediction of bio-elements for optimizing TFB. Emerging AI technologies, such as deep reinforcement learning [136–138], unsupervised learning techniques [139], and advanced neural network architectures [140], will provide powerful tools for bio-element design. Integrating AI with other scientific disciplines, such as systems biology and bioinformatics, will enhance our understanding of complex biological systems and improve the precision and effectiveness of bio-element design. Potential applications in medicine [141–146], agriculture [147], and environmental science [148,149] will expand as AI-designed enzymes and metabolic pathways revolutionize drug discovery [150] and biomanufacturing processes. Addressing challenges in data acquisition, model interpretability, and ethical considerations will be crucial. Future research should focus on developing standardized data-sharing protocols, enhancing model transparency, and establishing ethical guidelines for AI applications in biology. Generating high-quality data and refining AI models to handle biological complexity will be essential for advancing AI-assisted bio-element design. This includes optimizing TFB, enhancing the robustness and reliability of TFB applications, and ensuring that AI-driven solutions meet the diverse needs of synthetic biology. By tackling these challenges and leveraging the full potential of AI, researchers can significantly advance the field, making AI an indispensable tool in bio-element research and applications.

In conclusion, AI algorithms have already made significant contributions to the field of bio-elements. Although challenges remain in data acquisition, model construction, and optimization, execution of diverse tasks, and model interpretability, machine learning and deep learning methods based on AI algorithms remain indispensable tools. These tools are crucial not only for the rational design and activity prediction of bio-elements but also for optimizing TFB response performance and enhancing the robustness and reliability of TFB applications. By consolidating current research, highlighting innovative AI-driven solutions, and addressing existing challenges, this review demonstrates how AI can transform synthetic biology by improving precision, efficiency, and practicality in bio-element design. These insights offer significant advancements for both academic research and practical applications in biotechnology. Future developments in AI models will continue to drive progress in synthetic biology, ensuring more robust and reliable outcomes.

Abbreviations

AI: Artificial intelligence; Bio-elements: Biological elements; RBS: Ribosome binding site; TF: Transcription factor; CNN: Convolutional neural network; RNN: Recurrent neural network; GAN: Generative adversarial network; DBTL: Design–Build–Test–Learn; TFB: Transcription-factor-based biosensor; LSTM: Long short-term memory; XGBoost: extreme gradient boosting; TFBS: Transcription factor binding site; PCC: Pearson correlation coefficient; TIR: Translation initiation rate; GPR: Gaussian process regression; NGS: Next generation sequencing; GNN: Graph neural network; WGAN: Wasserstein GAN; FACS: Fluorescence-activated cell sorting; WGAN-GP: WGAN with gradient penalty; BAGAN: Balanced GAN; SMCFA: Short-chain and medium-chain fatty acid; GA: Glucaric acid; MIOX: *Myo*-inositol oxygenase; PopQC: Population quality control.

References

1. Portela, R.M.C.; Vogl, T.; Kniely, C.; Fischer, J.E.; Oliveira, R.; Glieder, A. Synthetic core promoters as universal parts for fine-tuning expression in different yeast species. *ACS Synth. Biol.* **2016**, *6*, 471–484. [CrossRef] [PubMed]
2. Weingarten Gabbay, S.; Nir, R.; Lubliner, S.; Sharon, E.; Kalma, Y.; Weinberger, A.; Segal, E. Systematic interrogation of human promoters. *Genome Res.* **2019**, *29*, 171–183. [CrossRef]
3. Reeve, B.; Hargest, T.; Gilbert, C.; Ellis, T. Predicting translation initiation rates for designing synthetic biology. *Front. Bioeng. Biotechnol.* **2014**, *2*, 1. [CrossRef] [PubMed]
4. Avsec, Ž.; Agarwal, V.; Visentin, D.; Ledsam, J.R.; Grabska-Barwinska, A.; Taylor, K.R.; Assael, Y.; Jumper, J.; Kohli, P.; Kelley, D.R. Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* **2021**, *18*, 1196–1203. [CrossRef]
5. Lynch, S.A.; Gill, R.T. Synthetic biology: New strategies for directing design. *Metab. Eng.* **2012**, *14*, 205–211. [CrossRef]
6. Crampon, K.; Giorkallos, A.; Deldossi, M.; Baud, S.; Steffenel, L.A. Machine-learning methods for ligand–protein molecular docking. *Drug Discov. Today* **2022**, *27*, 151–164. [CrossRef]
7. Handelman, G.S.; Kok, H.K.; Chandra, R.V.; Razavi, A.H.; Lee, M.J.; Asadi, H. eDoctor: Machine learning and the future of medicine. *J. Intern. Med.* **2018**, *284*, 603–619. [CrossRef]
8. Guest, D.; Cranmer, K.; Whiteson, D. Deep larning and its application to LHC physics. *Annu. Rev. Nucl. Part. Sci.* **2018**, *68*, 161–181. [CrossRef]
9. Roscher, R.; Bohn, B.; Duarte, M.F.; Garcke, J. Explainable machine learning for scientific insights and discoveries. *IEEE Access* **2020**, *8*, 42200–42216. [CrossRef]
10. Wang, H.; Fu, T.; Du, Y.; Gao, W.; Huang, K.; Liu, Z.; Chandak, P.; Liu, S.; Van Katwyk, P.; Deac, A.; et al. Scientific discovery in the age of artificial intelligence. *Nature* **2023**, *620*, 47–60. [CrossRef] [PubMed]
11. Greener, J.G.; Kandathil, S.M.; Moffat, L.; Jones, D.T. A guide to machine learning for biologists. *Nat. Rev. Mol. Cell Biol.* **2021**, *23*, 40–55. [CrossRef] [PubMed]
12. Carbonell, P.; Radivojevic, T.; García Martín, H. Opportunities at the intersection of synthetic biology, machine learning, and automation. *ACS Synth. Biol.* **2019**, *8*, 1474–1477. [CrossRef] [PubMed]
13. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
14. Merchant, A.; Batzner, S.; Schoenholz, S.S.; Aykol, M.; Cheon, G.; Cubuk, E.D. Scaling deep learning for materials discovery. *Nature* **2023**, *624*, 80–85. [CrossRef] [PubMed]
15. Das, P.; Sercu, T.; Wadhawan, K.; Padhi, I.; Gehrmann, S.; Cipcigan, F.; Chenthamarakshan, V.; Strobelt, H.; dos Santos, C.; Chen, P.Y.; et al. Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. *Nat. Biomed. Eng.* **2021**, *5*, 613. [CrossRef] [PubMed]
16. Li, Z.W.; Yang, W.J.; Peng, S.H.; Liu, F. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6999–7019. [CrossRef]
17. Greg, V.H.; Carlos, M.; Gonzalo, N. A review on the long short-term memory model. *Artif. Intell. Rev.* **2020**, *53*, 5929–5955.
18. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4–24. [CrossRef] [PubMed]
19. Yuan, Q.T.; Chen, K.Y.; Yu, Y.M.; Le, N.Q.K.; Chua, M.C.H. Prediction of anticancer peptides based on an ensemble model of deep learning and machine learning using ordinal positional encoding. *Brief. Bioinform.* **2023**, *24*, bbac630. [CrossRef]
20. Wang, Y.B.; Wu, H.X.; Zhang, J.J.; Gao, Z.F.; Wang, J.M.; Yu, P.S.; Long, M.S. PredRNN: A recurrent neural network for spatiotemporal predictive learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 2208–2225. [CrossRef]
21. Ian, G.; Jean, P.A.; Mehdi, M.; Xu, B.; David, W.F.; Sherjil, O.; Aaron, C.; Yoshua, B. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144.
22. Jin, X.B.; Gong, W.T.; Kong, J.L.; Bai, Y.T.; Su, T.L. PFVAE: A planar flow-based variational auto-encoder prediction model for time series data. *Mathematics* **2022**, *10*, 610. [CrossRef]

23.   Wang, J.M.; Chu, Y.Y.; Mao, J.S.; Jeon, H.N.; Jin, H.Y.; Zeb, A.; Jang, Y.; Cho, K.H.; Song, T.; No, K.T. De novo molecular design with deep molecular generative models for PPI inhibitors. *Brief. Bioinform.* **2022**, *23*, bbac285. [CrossRef] [PubMed]

24.   Zrimec, J.; Fu, X.Z.; Muhammad, A.S.; Skrekas, C.; Jauniskis, V.; Speicher, N.K.; Börlin, C.S.; Verendel, V.; Chehreghani, M.H.; Dubhashi, D.; et al. Controlling gene expression with deep generative design of regulatory DNA. *Nat. Commun.* **2022**, *13*, 5099. [CrossRef]

25.   He, X.; Samee, M.A.H.; Blatti, C.; Sinha, S.; Ohler, U. Thermodynamics-based models of transcriptional regulation by enhancers: The roles of synergistic activation, cooperative binding and short-range repression. *PLoS Comput. Biol.* **2010**, *6*, 1000935. [CrossRef] [PubMed]

26.   Chen, Y.J.; Liu, P.; Nielsen, A.A.K.; Brophy, J.A.N.; Clancy, K.; Peterson, T.; Voigt, C.A. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat. Methods* **2013**, *10*, 659–664. [CrossRef]

27.   Yasmeen, E.; Wang, J.; Riaz, M.; Zhang, L.D.; Zuo, K.J. Designing artificial synthetic promoters for accurate, smart, and versatile gene expression in plants. *Plant Commun.* **2023**, *4*, e1000935. [CrossRef] [PubMed]

28.   Xu, K.J.; Yu, S.Y.; Wang, K.; Tan, Y.M.; Zhao, X.Y.; Liu, S.; Zhou, J.W.; Wang, X.L. AI and knowledge-based method for rational design of *Escherichia coli* sigma70 promoters. *Acs Synth. Biol.* **2024**, *13*, 402–407. [CrossRef] [PubMed]

29.   Shao, B.; Yan, J.W.; Zhang, J.; Liu, L.L.; Chen, Y.; Buskirk, A.R. Riboformer: A deep learning framework for predicting context-dependent translation dynamics. *Nat. Commun.* **2024**, *15*, 2011. [CrossRef]

30.   Krenn, M.; Pollice, R.; Guo, S.Y.; Aldeghi, M.; Cervera-Lierta, A.; Friederich, P.; Gomes, G.D.; Häse, F.; Jinich, A.; Nigam, A.; et al. On scientific understanding with artificial intelligence. *Nat. Rev. Phys.* **2022**, *4*, 761–769. [CrossRef]

31.   Wang, X.; Liu, L.; Li, S.; Wei, L.; Wang, H.; Wang, Y. Synthetic promoter design in *Escherichia coli* based on a deep generative network. *Nucleic Acids Res.* **2020**, *48*, 6403–6412. [CrossRef] [PubMed]

32.   Jores, T.; Tonnies, J.; Wrightsman, T.; Buckler, E.S.; Cuperus, J.T.; Fields, S.; Queitsch, C. Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. *Nat. Plants* **2021**, *7*, 842–855. [CrossRef] [PubMed]

33.   Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; et al. AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50*, 439–444. [CrossRef] [PubMed]

34.   Sajid, S.; Zveushe, O.K.; de Dios, V.R.; Nabi, F.; Lee, Y.K.; Kaleri, A.R.; Ma, L.; Zhou, L.; Zhang, W.; Dong, F. Pretreatment of rice straw by newly isolated fungal consortium enhanced lignocellulose degradation and humification during composting. *Bioresour. Technol.* **2022**, *354*, 127150. [CrossRef] [PubMed]

35.   Pham, C.; Stogios, P.J.; Savchenko, A.; Mahadevan, R. Advances in engineering and optimization of transcription factor-based biosensors for plug-and-play small molecule detection. *Curr. Opin. Biotechnol.* **2022**, *76*, 102753. [CrossRef] [PubMed]

36.   Zhou, G.J.; Zhang, F. Applications and tuning strategies for transcription factor-based metabolite biosensors. *Biosensors* **2023**, *13*, 428. [CrossRef]

37.   Ding, N.; Zhou, S.; Deng, Y. Transcription-factor-based biosensor engineering for applications in synthetic biology. *ACS Synth. Biol.* **2021**, *10*, 911–922. [CrossRef] [PubMed]

38.   Miyake, R.; Ling, H.; Foo, J.L.; Fugono, N.; Chang, M.W. Transporter-driven engineering of a genetic biosensor for the detection and production of short-branched chain fatty acids in *Saccharomyces cerevisiae*. *Front. Bioeng. Biotechnol.* **2022**, *10*, 838732. [CrossRef] [PubMed]

39.   Lu, M.; Sha, Y.; Kumar, V.; Xu, Z.; Zhai, R.; Jin, M. Transcription factor-based biosensor: A molecular-guided approach for advanced biofuel synthesis. *Biotechnol. Adv.* **2024**, *72*, 108339. [CrossRef]

40.   Mao, Y.; Huang, C.; Zhou, X.; Han, R.; Deng, Y.; Zhou, S. Genetically encoded biosensor engineering for application in directed evolution. *J. Microbiol. Biotechnol.* **2023**, *33*, 1257–1267. [CrossRef]

41.   Teng, Y.X.; Gong, X.Y.; Zhang, J.L.; Obideen, Z.; Yan, Y.J. Investigating and engineering an 1,2-propanediol-responsive transcription factor-based biosensor. *ACS Synth. Biol.* **2024**, *13*, 2177–2187. [CrossRef] [PubMed]

42.   Li, C.Y.; Zhou, Y.Y.; Zou, Y.S.; Jiang, T.; Gong, X.Y.; Yan, Y.J. Identifying, characterizing, and engineering a phenolic acid-responsive transcriptional factor from *Bacillus amyloliquefaciens*. *ACS Synth. Biol.* **2023**, *12*, 2382–2392. [CrossRef] [PubMed]

43.   Cheng, F.; Tang, X.L.; Kardashliev, T. Transcription factor-based biosensors in high-throughput screening: Advances and applications. *Biotechnol. J.* **2018**, *13*, 1700648. [CrossRef] [PubMed]

44.   d'Oelsnitz, S.; Nguyen, V.; Alper, H.S.; Ellington, A.D. Evolving a generalist biosensor for bicyclic monoterpenes. *ACS Synth. Biol.* **2022**, *11*, 265–272. [CrossRef] [PubMed]

45.   Gong, X.; Zhang, R.; Wang, J.; Yan, Y. Engineering of a TrpR-based biosensor for altered dynamic range and ligand preference. *ACS Synth. Biol.* **2022**, *11*, 2175–2183. [CrossRef]

46.   Ding, N.; Yuan, Z.; Zhang, X.; Chen, J.; Zhou, S.; Deng, Y. Programmable cross-ribosome-binding sites to fine-tune the dynamic range of transcription factor-based biosensor. *Nucleic Acids Res.* **2020**, *48*, 10602–10613. [CrossRef] [PubMed]

47.   Zhang, P.; Wang, H.; Xu, H.; Wei, L.; Liu, L.; Hu, Z.; Wang, X. Deep flanking sequence engineering for efficient promoter design using DeepSEED. *Nat. Commun.* **2023**, *14*, 6309. [CrossRef] [PubMed]

48.   Zhao, M.; Yuan, Z.; Wu, L.; Zhou, S.; Deng, Y. Precise prediction of promoter strength based on a de novo synthetic promoter library coupled with machine learning. *ACS Synth. Biol.* **2021**, *11*, 92–102. [CrossRef]

49.   Qiao, H.; Zhang, S.; Xue, T.; Wang, J.; Wang, B. iPro-GAN: A novel model based on generative adversarial learning for identifying promoters and their strength. *Comput. Methods Programs Biomed.* **2022**, *215*, 106625. [CrossRef]

50. de Almeida, B.P.; Reiter, F.; Pagani, M.; Stark, A. DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. *Nat. Genet.* **2022**, *54*, 613–624. [CrossRef]

51. Liao, M.; Zhao, J.P.; Tian, J.; Zheng, C.H. iEnhancer-DCLA: Using the original sequence to identify enhancers and their strength based on a deep learning framework. *BMC Bioinform.* **2022**, *23*, 480. [CrossRef] [PubMed]

52. Zhang, M.; Holowko, M.B.; Hayman Zumpe, H.; Ong, C.S. Machine learning guided batched design of a bacterial ribosome binding site. *ACS Synth. Biol.* **2022**, *11*, 2314–2326. [CrossRef] [PubMed]

53. Höllerer, S.; Papaxanthos, L.; Gumpinger, A.C.; Fischer, K.; Beisel, C.; Borgwardt, K.; Benenson, Y.; Jeschek, M. Large-scale DNA-based phenotypic recording and deep learning enable highly accurate sequence-function mapping. *Nat. Commun.* **2020**, *11*, 3551. [CrossRef] [PubMed]

54. Donatas, R.; Vykintas, J.; Laurynas, K.; Elzbieta, R.; Jan, Z.; Simona, P.; Irmantas, R.; Audrius, L.; Wissam, A.; Otto, S.; et al. Expanding functional protein sequence spaces using generative adversarial networks. *Nat. Mach. Intell.* **2021**, *3*, 324–333.

55. Karimi, M.; Zhu, S.; Cao, Y.; Shen, Y. De novo protein design for novel folds using guided conditional wasserstein generative adversarial networks (gcWGAN). *J. Chem. Inf. Model.* **2019**, *60*, 5667–5681. [CrossRef] [PubMed]

56. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef] [PubMed]

57. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. [CrossRef] [PubMed]

58. Li, F.; Yuan, L.; Lu, H.; Li, G.; Chen, Y.; Engqvist, M.K.M.; Kerkhoven, E.J.; Nielsen, J. Deep learning-based k_{cat} prediction enables improved enzyme-constrained model reconstruction. *Nat. Catal.* **2022**, *5*, 662–672. [CrossRef]

59. Yu, H.; Deng, H.; He, J.; Keasling, J.D.; Luo, X. UniKP: A unified framework for the prediction of enzyme kinetic parameters. *Nat. Commun.* **2023**, *14*, 8211. [CrossRef] [PubMed]

60. Yu, T.; Cui, H.; Li, J.C.; Luo, Y.; Jiang, G.; Zhao, H. Enzyme function prediction using contrastive learning. *Science* **2023**, *379*, 1358. [CrossRef]

61. Busby, S.; Ebright, R.H. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **1994**, *79*, 743–746. [CrossRef]

62. Curran, K.A.; Crook, N.C.; Karim, A.S.; Gupta, A.; Wagman, A.M.; Alper, H.S. Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat. Commun.* **2014**, *5*, 4002. [CrossRef]

63. Huang, Y.K.; Yu, C.H.; Ng, I.S. Precise strength prediction of endogenous promoters from *Escherichia coli* and J-series promoters by artificial intelligence. *J. Taiwan Inst. Chem. Eng.* **2024**, *160*, 105211. [CrossRef]

64. Shlyueva, D.; Stampfel, G.; Stark, A. Transcriptional enhancers: From properties to genome-wide predictions. *Nat. Rev. Genet.* **2014**, *15*, 272. [CrossRef]

65. Spitz, F.; Furlong, E.E.M. Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **2012**, *13*, 613–626. [CrossRef] [PubMed]

66. May, D.; Blow, M.J.; Kaplan, T.; McCulley, D.J.; Jensen, B.C.; Akiyama, J.A.; Holt, A.; Plajzer-Frick, I.; Shoukry, M.; Wright, C.; et al. Large-scale discovery of enhancers from human heart tissue. *Nat. Genet.* **2011**, *44*, 89–93. [CrossRef] [PubMed]

67. Taskiran, I.I.; Spanier, K.I.; Dickmänken, H.; Kempynck, N.; Pančíková, A.; Ekşi, E.C.; Hulselmans, G.; Ismail, J.N.; Theunis, K.; Vandepoel, R.; et al. Cell-type-directed design of synthetic enhancers. *Nature* **2023**, *626*, 212–220. [CrossRef]

68. Wolfe, J.C.; Mikheeva, L.A.; Hagras, H.; Zabet, N.R. An explainable artificial intelligence approach for decoding the enhancer histone modifications code and identification of novel enhancers in *Drosophila*. *Genome Biol.* **2021**, *22*, 308. [CrossRef]

69. Hamamoto, R.; Takasawa, K.; Shinkai, N.; Machino, H.; Kouno, N.; Asada, K.; Komatsu, M.; Kaneko, S. Analysis of super-enhancer using machine learning and its application to medical biology. *Brief. Bioinform.* **2023**, *24*, bbad107. [CrossRef] [PubMed]

70. Peterman, N.; Levine, E. Sort-seq under the hood: Implications of design choices on large-scale characterization of sequence-function relations. *BMC Genom.* **2016**, *17*, 206. [CrossRef] [PubMed]

71. Salis, H.M.; Mirsky, E.A.; Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **2009**, *27*, 946–950. [CrossRef]

72. Ding, W.; Nakai, K.; Gong, H. Protein design via deep learning. *Brief. Bioinform.* **2022**, *23*, 102. [CrossRef] [PubMed]

73. Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R.; Milles, L.; Wicky, B.; Courbet, A.; de Haas, R.; Bethel, N.; et al. Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **2022**, *378*, 49–55. [CrossRef]

74. Watson, J.L.; Juergens, D.; Bennett, N.R.; Trippe, B.L.; Yim, J.; Eisenach, H.E.; Ahern, W.; Borst, A.J.; Ragotte, R.J.; Milles, L.F.; et al. De novo design of protein structure and function with RFdiffusion. *Nature* **2023**, *620*, 1089–1100. [CrossRef] [PubMed]

75. Lin, E.; Lin, C.H.; Lane, H.Y. De novo peptide and protein design using generative adversarial networks: An update. *J. Chem. Inf. Model.* **2022**, *62*, 761–774. [CrossRef]

76. Scalvini, B.; Sheikhhassani, V.; Mashaghi, A. Topological principles of protein folding. *Phys. Chem. Chem. Phys.* **2021**, *23*, 21316–21328. [CrossRef] [PubMed]

77. Baker, D. Protein structure prediction and structural genomics. *Science* **2001**, *294*, 93–96. [CrossRef]

78. Goverde, C.A.; Wolf, B.; Khakzad, H.; Rosset, S.; Correia, B.E. De novo protein design by inversion of the AlphaFold structure prediction network. *Protein Sci.* **2023**, *32*, 4653. [CrossRef]

79. Kosugi, T.; Ohue, M. Design of cyclic peptides targeting protein–protein interactions using AlphaFold. *Int. J. Mol. Sci.* **2023**, *24*, 13257. [CrossRef]

80. Bryant, P.; Elofsson, A. Peptide binder design with inverse folding and protein structure prediction. *Commun. Chem.* **2023**, *6*, 229. [CrossRef]

81. Senior, A.W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Zidek, A.; Nelson, A.W.R.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710. [CrossRef] [PubMed]

82. Moult, J.; Fidelis, K.; Kryshtafovych, A.; Schwede, T.; Topf, M. Critical assessment of techniques for protein structure prediction, fourteenth round. In *CASP 14 Abstract Book*; CASP, 2020; pp. 1–344. Available online: https://www.predictioncenter.org/casp14/doc/CASP14_Abstracts.pdf (accessed on 15 June 2024).

83. Tellechea-Luzardo, J.; Stiebritz, M.T.; Carbonell, P. Transcription factor-based biosensors for screening and dynamic regulation. *Front. Bioeng. Biotechnol.* **2023**, *11*, 1118702. [CrossRef] [PubMed]

84. Hartline, C.J.; Zhang, F. The growth dependent design constraints of transcription-factor-based metabolite biosensors. *ACS Synth. Biol.* **2022**, *11*, 2247–2258. [CrossRef] [PubMed]

85. Mannan, A.A.; Liu, D.; Zhang, F.; Oyarzún, D.A. Fundamental design principles for transcription-factor-based metabolite biosensors. *ACS Synth. Biol.* **2017**, *6*, 1851–1859. [CrossRef]

86. Chen, Y.; Ho, J.M.L.; Shis, D.L.; Gupta, C.; Long, J.; Wagner, D.S.; Ott, W.; Josić, K.; Bennett, M.R. Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. *Nat. Commun.* **2018**, *9*, 64. [CrossRef] [PubMed]

87. Zhao, N.; Song, J.; Zhang, H.; Lin, Y.; Han, S.; Huang, Y.; Zheng, S. Development of a transcription factor-based diamine biosensor in *Corynebacterium glutamicum*. *ACS Synth. Biol.* **2021**, *10*, 3074–3083. [CrossRef]

88. Peters, G.; De Paepe, B.; De Wannemaeker, L.; Duchi, D.; Maertens, J.; Lammertyn, J.; De Mey, M. Development of N-acetylneuraminic acid responsive biosensors based on the transcriptional regulator NanR. *Biotechnol. Bioeng.* **2018**, *115*, 1855–1865. [CrossRef]

89. Zhou, Y.; Yuan, Y.; Wu, Y.; Li, L.; Jameel, A.; Xing, X.; Zhang, C. Encoding genetic circuits with DNA barcodes paves the way for machine learning-assisted metabolite biosensor response curve profiling in yeast. *ACS Synth. Biol.* **2022**, *11*, 977–989. [CrossRef] [PubMed]

90. Kasey, C.M.; Zerrad, M.; Li, Y.; Cropp, T.A.; Williams, G.J. Development of transcription factor-based designer macrolide biosensors for metabolic engineering and synthetic biology. *ACS Synth. Biol.* **2017**, *7*, 227–239. [CrossRef]

91. Greco, F.V.; Pandi, A.; Erb, T.J.; Grierson, C.S.; Gorochowski, T.E. Harnessing the central dogma for stringent multi-level control of gene expression. *Nat. Commun.* **2021**, *12*, 1738. [CrossRef]

92. Ding, N.; Sun, L.; Zhou, X.; Zhang, L.; Deng, Y.; Yin, L. Enhancing glucaric acid production from *myo*-inositol in *Escherichia coli* by eliminating cell-to-cell variation. *Appl. Environ. Microbiol.* **2024**, *90*, e00149-24. [CrossRef]

93. Ding, N.; Zhang, G.; Zhang, L.; Shen, Z.; Yin, L.; Zhou, S.; Deng, Y. Engineering an AI-based forward-reverse platform for the design of cross-ribosome binding sites of a transcription factor biosensor. *Comput. Struct. Biotechnol. J.* **2023**, *21*, 2929–2939. [CrossRef]

94. Xiao, Y.; Jiang, W.; Zhang, F. Developing a genetically encoded, cross-species biosensor for detecting ammonium and regulating biosynthesis of cyanophycin. *ACS Synth. Biol.* **2017**, *6*, 1807–1815. [CrossRef] [PubMed]

95. Trabelsi, H.; Koch, M.; Faulon, J.L. Building a minimal and generalizable model of transcription factor–based biosensors: Showcasing flavonoids. *Biotechnol. Bioeng.* **2018**, *115*, 2292–2304. [CrossRef] [PubMed]

96. Ma, W.; Fu, Y.; Bao, Y.; Wang, Z.; Lei, B.; Zheng, W.; Wang, C.; Liu, Y. DeepSATA: A deep learning-based sequence analyzer incorporating the transcription factor binding affinity to dissect the effects of non-coding genetic variants. *Int. J. Mol. Sci.* **2023**, *24*, 12023. [CrossRef]

97. Han, K.; Shen, L.C.; Zhu, Y.H.; Xu, J.; Song, J.N.; Yu, D.J. MAResNet: Predicting transcription factor binding sites by combining multi-scale bottom-up and top-down attention and residual network. *Brief. Bioinform.* **2022**, *23*, 445. [CrossRef] [PubMed]

98. Quan, L.J.; Sun, X.Y.; Wu, J.; Mei, J.; Huang, L.Q.; He, R.J.; Nie, L.P.; Chen, Y.; Lyu, Q. Learning useful representations of DNA sequences from ChIP-Seq datasets for exploring transcription factor binding specificities. *IEEE-ACM Trans. Comput. Biol. Bioinform.* **2022**, *19*, 998–1008. [CrossRef] [PubMed]

99. Baumann, L.; Rajkumar, A.S.; Morrissey, J.P.; Boles, E.; Oreb, M. A yeast-based biosensor for screening of short- and medium-chain fatty acid production. *ACS Synth. Biol.* **2018**, *7*, 2640–2646. [CrossRef]

100. Yu, W.; Xu, X.; Jin, K.; Liu, Y.; Li, J.; Du, G.; Lv, X.; Liu, L. Genetically encoded biosensors for microbial synthetic biology: From conceptual frameworks to practical applications. *Biotechnol. Adv.* **2023**, *62*, 108077. [CrossRef]

101. Seok, J.Y.; Han, Y.H.; Yang, J.S.; Yang, J.; Lim, H.G.; Kim, S.G.; Seo, S.W.; Jung, G.Y. Synthetic biosensor accelerates evolution by rewiring carbon metabolism toward a specific metabolite. *Cell Rep.* **2021**, *36*, 109589. [CrossRef]

102. Xu, X.; Li, X.; Liu, Y.; Zhu, Y.; Li, J.; Du, G.; Chen, J.; Ledesma-Amaro, R.; Liu, L. Pyruvate-responsive genetic circuits for dynamic control of central metabolism. *Nat. Chem. Biol.* **2020**, *16*, 1261–1268. [CrossRef]

103. Su, B.; Lai, P.; Deng, M.R.; Zhu, H. Design of a dual-responding genetic circuit for high-throughput identification of L-threonine-overproducing *Escherichia coli*. *Bioresour. Technol.* **2024**, *395*, 130407. [CrossRef]

104. Zhao, X.; Wu, Y.; Feng, T.; Shen, J.; Lu, H.; Zhang, Y.; Chou, H.H.; Luo, X.; Keasling, J.D. Dynamic upregulation of the rate-limiting enzyme for valerolactam biosynthesis in *Corynebacterium glutamicum*. *Metab. Eng.* **2023**, *77*, 89–99. [CrossRef]

105. Rogers, J.K.; Church, G.M. Genetically encoded sensors enable real-time observation of metabolite production. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2388–2393. [CrossRef] [PubMed]

106. Rogers, J.K.; Guzman, C.D.; Taylor, N.D.; Raman, S.; Anderson, K.; Church, G.M. Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Res.* **2015**, *43*, 7648–7660. [CrossRef]

107. Mitchler, M.M.; Garcia, J.M.; Montero, N.E.; Williams, G.J. Transcription factor-based biosensors: A molecular-guided approach for natural product engineering. *Curr. Opin. Biotechnol.* **2021**, *69*, 172–181. [CrossRef] [PubMed]

108. Kortmann, M.; Mack, C.; Baumgart, M.; Bott, M. Pyruvate carboxylase variants enabling improved lysine production from glucose identified by biosensor-based high-throughput fluorescence-activated cell sorting screening. *ACS Synth. Biol.* **2019**, *8*, 274–281. [CrossRef]

109. Huang, R.; Chen, H.; Upp, D.M.; Lewis, J.C.; Zhang, Y.-H.P.J. A high-throughput method for directed evolution of NAD(P)$^+$-dependent dehydrogenases for the reduction of biomimetic nicotinamide analogues. *ACS Catal.* **2019**, *9*, 11709–11719. [CrossRef] [PubMed]

110. Trivedi, V.D.; Mohan, K.; Chappell, T.C.; Mays, Z.J.S.; Nair, N.U. Cheating the cheater: Suppressing false-positive enrichment during biosensor-guided biocatalyst engineering. *ACS Synth. Biol.* **2022**, *11*, 420–429. [CrossRef]

111. Nasr, M.A.; Martin, V.J.J.; Kwan, D.H. Divergent directed evolution of a TetR-type repressor towards aromatic molecules. *Nucleic Acids Res.* **2023**, *51*, 7675–7690. [CrossRef]

112. Du, H.; Liang, Y.; Li, J.; Yuan, X.; Tao, F.; Dong, C.; Shen, Z.; Sui, G.; Wang, P. Directed evolution of 4-hydroxyphenylpyruvate biosensors based on a dual selection system. *Int. J. Mol. Sci.* **2024**, *25*, 1533. [CrossRef]

113. Liang, Y.; Luo, J.; Yang, C.; Guo, S.; Zhang, B.; Chen, F.; Su, K.; Zhang, Y.; Dong, Y.; Wang, Z.; et al. Directed evolution of the PobR allosteric transcription factor to generate a biosensor for 4-hydroxymandelic acid. *World J. Microbiol. Biotechnol.* **2022**, *38*, 104. [CrossRef] [PubMed]

114. Chen, D.; Xu, S.; Li, S.; Tao, S.; Li, L.; Chen, S.; Wu, L. Directly evolved AlkS-based biosensor platform for monitoring and high-throughput screening of alkane production. *ACS Synth. Biol.* **2023**, *12*, 832–841. [CrossRef] [PubMed]

115. Shen, Y.P.; Pan, Y.; Niu, F.X.; Liao, Y.L.; Huang, M.; Liu, J.Z. Biosensor-assisted evolution for high-level production of 4-hydroxyphenylacetic acid in *Escherichia coli*. *Metab. Eng.* **2022**, *70*, 1–11. [CrossRef] [PubMed]

116. Tong, Y.; Li, N.; Zhou, S.; Zhang, L.; Xu, S.; Zhou, J. Improvement of chalcone synthase activity and high-efficiency fermentative production of (2S)-naringenin via *in vivo* biosensor-guided directed evolution. *ACS Synth. Biol.* **2024**, *13*, 1454–1466. [CrossRef] [PubMed]

117. Rugbjerg, P.; Sommer, M.O.A. Overcoming genetic heterogeneity in industrial fermentations. *Nat. Biotechnol.* **2019**, *37*, 869–876. [CrossRef] [PubMed]

118. Xiao, Y.; Bowen, C.H.; Liu, D.; Zhang, F. Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis. *Nat. Chem. Biol.* **2016**, *12*, 339–344. [CrossRef] [PubMed]

119. Guan, A.; He, Z.; Wang, X.; Jia, Z.J.; Qin, J. Engineering the next-generation synthetic cell factory driven by protein engineering. *Biotechnol. Adv.* **2024**, *73*, 108366. [CrossRef] [PubMed]

120. Zhang, F.; Carothers, J.M.; Keasling, J.D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **2012**, *30*, 354–359. [CrossRef]

121. Zhou, S.; Yuan, S.F.; Nair, P.H.; Alper, H.S.; Deng, Y.; Zhou, J. Development of a growth coupled and multi-layered dynamic regulation network balancing malonyl-CoA node to enhance (2S)-naringenin biosynthesis in *Escherichia coli*. *Metab. Eng.* **2021**, *67*, 41–52. [CrossRef]

122. Zhu, Y.; Li, Y.; Xu, Y.; Zhang, J.; Ma, L.; Qi, Q.; Wang, Q. Development of bifunctional biosensors for sensing and dynamic control of glycolysis flux in metabolic engineering. *Metab. Eng.* **2021**, *68*, 142–151. [CrossRef] [PubMed]

123. Tian, J.; Yang, G.; Gu, Y.; Sun, X.; Lu, Y.; Jiang, W. Developing an endogenous quorum-sensing based CRISPRi circuit for autonomous and tunable dynamic regulation of multiple targets in *Streptomyces*. *Nucleic Acids Res.* **2020**, *48*, 8188–8202. [CrossRef] [PubMed]

124. Dhakal, A.; McKay, C.; Tanner, J.J.; Cheng, J.L. Artificial intelligence in the prediction of protein-ligand interactions: Recent advances and future directions. *Brief. Bioinform.* **2022**, *23*, bbab476. [CrossRef] [PubMed]

125. Burley, S.K.; Bhikadiya, C.; Bi, C.X.; Bittrich, S.; Chao, H.Y.; Chen, L.; Craig, P.A.; Crichlow, G.V.; Dalenberg, K.; Duarte, J.M.; et al. RCSB Protein Data Bank (RCSB.org): Delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res.* **2023**, *51*, D488–D508. [CrossRef]

126. Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.; Yeo, J.; et al. AlphaFold protein structure database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **2023**, *52*, D368–D375. [CrossRef] [PubMed]

127. Yang, Y.Q.; Tang, X.; Zhang, X.R.; Ma, J.J.; Liu, F.; Jia, X.P.; Jiao, L.C. Semi-supervised multiscale dynamic graph convolution network for hyperspectral image classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 6806–6820. [CrossRef] [PubMed]

128. Chavez, M.R.; Butler, T.S.; Rekawek, P.; Heo, H.; Kinzler, W.L. Chat generative pre-trained transformer: Why we should embrace this technology. *Am. J. Obstet. Gynecol.* **2023**, *228*, 706–711. [CrossRef]

129. Lee, G.S.; Kim, S.; Bae, S. Efficient design method for a forward-converter transformer based on a KNNGRUDNN model. *IEEE Trans. Power Electron.* **2023**, *38*, 73–78. [CrossRef]

130. Bhatti, U.A.; Tang, H.; Wu, G.L.; Marjan, S.; Hussain, A. Deep learning with graph convolutional networks: An overview and latest applications in computational intelligence. *Int. J. Intell. Syst.* **2023**, *2023*, 8342104. [CrossRef]

131. Yu, J.C.; Xu, T.Y.; Rong, Y.; Bian, Y.T.; Huang, J.Z.; He, R. Recognizing predictive substructures with subgraph information bottleneck. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 1650–1663. [CrossRef]

132. Wei, Z.; Hua, K.; Wei, L.; Ma, S.; Jiang, R.; Zhang, X.; Li, Y.; Wong, W.H.; Wang, X. NeuronMotif: Deciphering cis-regulatory codes by layer-wise demixing of deep neural networks. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2216698120. [CrossRef] [PubMed]

133. Marques, G.; Agarwal, D.; de la Torre Díez, I. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Appl. Soft Comput.* **2020**, *96*, 106691. [CrossRef] [PubMed]

134. Zhang, J.; Qin, Q.; Ye, Q.; Ruan, T. ST-Unet: Swin transformer boosted U-Net with cross-layer feature enhancement for medical image segmentation. *Comput. Biol. Med.* **2023**, *153*, 106516. [CrossRef] [PubMed]

135. Singhal, K.; Azizi, S.; Tu, T.; Mahdavi, S.S.; Wei, J.; Chung, H.W.; Scales, N.; Tanwani, A.; Cole-Lewis, H.; Pfohl, S.; et al. Large language models encode clinical knowledge. *Nature* **2023**, *620*, 172–180. [CrossRef] [PubMed]

136. Gao, Q.; Schweidtmann, A.M. Deep reinforcement learning for process design: Review and perspective. *Curr. Opin. Chem. Eng.* **2024**, *44*, 101012. [CrossRef]

137. Wang, D.; Gao, N.; Liu, D.R.; Li, J.N.; Lewis, F.L. Recent progress in reinforcement learning and adaptive dynamic programming for advanced control applications. *IEEE-Caa J. Autom. Sin.* **2024**, *11*, 18–36. [CrossRef]

138. Kaufmann, E.; Bauersfeld, L.; Loquercio, A.; Mueller, M.; Koltun, V.; Scaramuzza, D. Champion-level drone racing using deep reinforcement learning. *Nature* **2023**, *620*, 982–987. [CrossRef]

139. Chen, Y.B.; Mancini, M.; Zhu, X.T.; Akata, Z. Semi-supervised and unsupervised deep visual learning: A Survey. *Ieee Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 1327–1347. [CrossRef] [PubMed]

140. Min, B.N.; Ross, H.; Sulem, E.; Ben Veyseh, A.P.; Nguyen, T.H.; Sainz, O.; Agirre, E.; Heintz, I.; Roth, D. Recent advances in natural language processing via large pre-trained language models: A Survey. *ACM Comput. Surv.* **2024**, *56*, 1–40. [CrossRef]

141. Xue, Q.; Miao, P.; Miao, K.; Yu, Y.; Li, Z. An online automatic sorting system for defective Ginseng Radix et Rhizoma Rubra using deep learning. *Chin. Herb. Med.* **2023**, *15*, 447–456. [CrossRef]

142. Klauschen, F.; Dippel, J.; Keyl, P.; Jurmeister, P.; Bockmayr, M.; Mock, A.; Buchstab, O.; Alber, M.; Ruff, L.; Montavon, G.; et al. Toward explainable artificial intelligence for precision pathology. *Annu. Rev. Pathol.-Mech. Dis.* **2024**, *19*, 541–570. [CrossRef] [PubMed]

143. Tong, L.; Shi, W.Q.; Isgut, M.; Zhong, Y.S.; Lais, P.; Gloster, L.; Sun, J.M.; Swain, A.; Giuste, F.; Wang, M.D. Integrating multi-omics data with EHR for precision medicine using advanced artificial intelligence. *IEEE Rev. Biomed. Eng.* **2024**, *17*, 80–97. [CrossRef] [PubMed]

144. Zhou, B.C.; Qian, Z.H.; Li, Q.Y.; Gao, Y.; Li, M.H. Assessment of pulmonary infectious disease treatment with Mongolian medicine formulae based on data mining, network pharmacology and molecular docking. *Chin. Herb. Med.* **2022**, *14*, 432–448. [CrossRef] [PubMed]

145. Yang, Z.Y.; Zeng, X.X.; Zhao, Y.; Chen, R.S. AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduct. Target. Ther.* **2023**, *8*, 115. [CrossRef] [PubMed]

146. Chen, R.J.; Wang, J.J.; Williamson, D.F.K.; Chen, T.Y.; Lipkova, J.; Lu, M.Y.; Sahai, S.; Mahmood, F. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat. Biomed. Eng.* **2023**, *7*, 719–742. [CrossRef]

147. Zhang, W.D.; Li, Z.X.; Li, G.H.; Zhuang, P.X.; Hou, G.J.; Zhang, Q.; Li, C.Y. GACNet: Generate adversarial-driven cross-aware network for hyperspectral wheat variety identification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5503314. [CrossRef]

148. Moghadam, P.Z.; Chung, Y.G.; Snurr, R.Q. Progress toward the computational discovery of new metal-organic framework adsorbents for energy applications. *Nat. Energy* **2024**, *9*, 121–133. [CrossRef]

149. Bi, K.F.; Xie, L.X.; Zhang, H.H.; Chen, X.; Gu, X.T.; Tian, Q. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **2023**, *619*, 533–538. [CrossRef]

150. Nussinov, R.; Zhang, M.Z.; Liu, Y.L.; Jang, H. AlphaFold, allosteric, and orthosteric drug discovery: Ways forward. *Drug Discov. Today* **2023**, *28*, 103551. [CrossRef]