



Review

Possibilities of Using De Novo Design for Generating Diverse Functional Food Enzymes

Xinglong Wang ^{1,2,†}, Kangjie Xu ^{1,2,†}, Yameng Tan ^{1,2}, Song Liu ^{1,2} and Jingwen Zhou ^{1,2,3,*} 

¹ Engineering Research Center of Ministry of Education on Food Synthetic Biotechnology, School of Biotechnology, Jiangnan University, Wuxi 214122, China

² Science Center for Future Foods, Jiangnan University, Wuxi 214122, China

³ Jiangsu Province Engineering Research Center of Food Synthetic Biotechnology, Jiangnan University, Wuxi 214122, China

* Correspondence: zhoujw1982@jiangnan.edu.cn; Tel./Fax: +86-510-85918310

† These authors contributed equally to this work.

Abstract: Food enzymes have an important role in the improvement of certain food characteristics, such as texture improvement, elimination of toxins and allergens, production of carbohydrates, enhancing flavor/appearance characteristics. Recently, along with the development of artificial meats, food enzymes have been employed to achieve more diverse functions, especially in converting non-edible biomass to delicious foods. Reported food enzyme modifications for specific applications have highlighted the significance of enzyme engineering. However, using direct evolution or rational design showed inherent limitations due to the mutation rates, which made it difficult to satisfy the stability or specific activity needs for certain applications. Generating functional enzymes using de novo design, which highly assembles naturally existing enzymes, provides potential solutions for screening desired enzymes. Here, we describe the functions and applications of food enzymes to introduce the need for food enzymes engineering. To illustrate the possibilities of using de novo design for generating diverse functional proteins, we reviewed protein modelling and de novo design methods and their implementations. The future directions for adding structural data for de novo design model training, acquiring diversified training data, and investigating the relationship between enzyme–substrate binding and activity were highlighted as challenges to overcome for the de novo design of food enzymes.

Keywords: food enzyme engineering; artificial intelligence; enzyme engineering; de novo design



Citation: Wang, X.; Xu, K.; Tan, Y.; Liu, S.; Zhou, J. Possibilities of Using De Novo Design for Generating Diverse Functional Food Enzymes. *Int. J. Mol. Sci.* **2023**, *24*, 3827. <https://doi.org/10.3390/ijms24043827>

Academic Editors: Jose M. Guisan, Yung-Chuan Liu and Antonio Zorroro

Received: 11 January 2023

Revised: 3 February 2023

Accepted: 3 February 2023

Published: 14 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Proteins are widely distributed in all living organisms, and are natural biocatalysts that participate in biological reactions [1–4]. Proteins are widely applied in various industries. They are used for generating food appearance and flavor [5], participating in (bio)-material processing [6], or used as drugs [7] or bio-materials [8]. Food enzymes are a group of proteins that are emerging as additives used for food processing [9]. In recent years, studies have shown that food enzymes not only play a role in traditional food industrial processing, such as in baking, dairy, bean and meat products [10], but have also been advantageous to novel food products processing, such as artificial meat [11]. In addition, several food enzymes show capabilities for degrading allergens [12] or bitter peptides [13], providing food processing convenience. Since the sophisticated performance of food enzymes was approved, the US Food and Drug Administration (FDA) and European Food Safety Authority (EFSA) have issued several statements to standardize the uses of food enzymes, in order to ensure food safety [14]. With the permission for use, the applications of enzymes in food processing continuously increased over the past few years; the reported market size for enzymes (mainly food enzymes) exceeded 17 billion in 2020 [15,16]. However, not all food enzymes can meet their industrial application needs,

partly due to their insufficient stability or low activity, and partly due to adverse processing conditions [17].

Enzyme engineering aims to resolve weak enzyme stability and low activity issues. The enzyme engineering technique can be mainly divided into direct-evolution, rational design, and semi-rational design [18]. Although enzyme engineering still massively relies on high-throughput screening supported by direct evolution and semi-rational design, reports on rational design-based enzyme modification are displaying an increasing trend [19]. One reason is the high labor and experimental cost of high-throughput screening, and it is always hard to find a screening method [20]. Another reason is that the rise in accurate protein modeling and energy computing techniques in recent years has dramatically reduced the difficulty of rational design [21,22]. Therefore, a growing number of researchers are adopting rational design for enzyme evolution. Developing protein design methods has served the need for screening functionalized therapeutic proteins or high-efficiency enzymes [23,24], which aim at satisfying the grand market of commercialized therapeutic proteins [25]. Rational design methods successfully assisted protein evolution [21,26–28]. However, it is difficult to substantially improve a certain function of a given protein by conducting single or several point mutations, or by bringing in highly diversified changes to protein functions.

A breakthrough in the field of rational design is de novo design. De novo design learns features from existing sequences and structures, which can be used to create diversified novel proteins to reinforce the functionality of natural proteins [29,30]. The progression of AI technology and the development of bioinformatics have pushed the rapid development of de novo design techniques to increase the number of proteins that were designed from scratch displaying diverse functions [31–34]. Advances in protein de novo design have provided opportunities for the direct evolution of enzymes. The implementation of de novo design for generating functional proteins, protein binders or industrial enzymes has achieved great success [31–35]. In this review, by describing the function and challenges of food enzymes, the potential and opportunities of using de novo design techniques for functional food enzymes engineering were highlighted. The AI-based protein modeling and de novo design tools, as well as their utility for protein engineering, were reviewed to gain insight into the potential of using these tools for food enzymes engineering.

2. Diverse Functions of Food Enzymes

Food enzymes are widely used in various food processing [19], and are mostly derived from microorganisms. According to enzyme classification, commercial food enzymes can be classified into oxidoreductases, transferases, hydrolases, isomerases and ligases, while proteases take the principal portion (Table 1). In practical uses, food enzymes are mainly used to improve food taste and appearance, or to convert sugar-related reactions [36–38] (Figure 1). For example, in meat product processing, proteases (such as papain) and transglutaminases are used for meat tenderization [39] and improving texture [36], respectively. In dairy products processing, proteases such as lactase can be used for hydrolyzing lactose into galactose and glucose, in order to assist human absorbance [40]; lipase can assist in oil hydrolysis [41]; and esterase can improve the flavor by hydrolyzing esters into acids and alcohol [42]. For winemaking, pectinase can hydrolyze pectin to improve the flavor and color of wine [43], and glucose oxidase can oxidize glucose into gluconic acid and generate hydrogen peroxide, which can improve product quality and enhance the storage period during beer production [44].

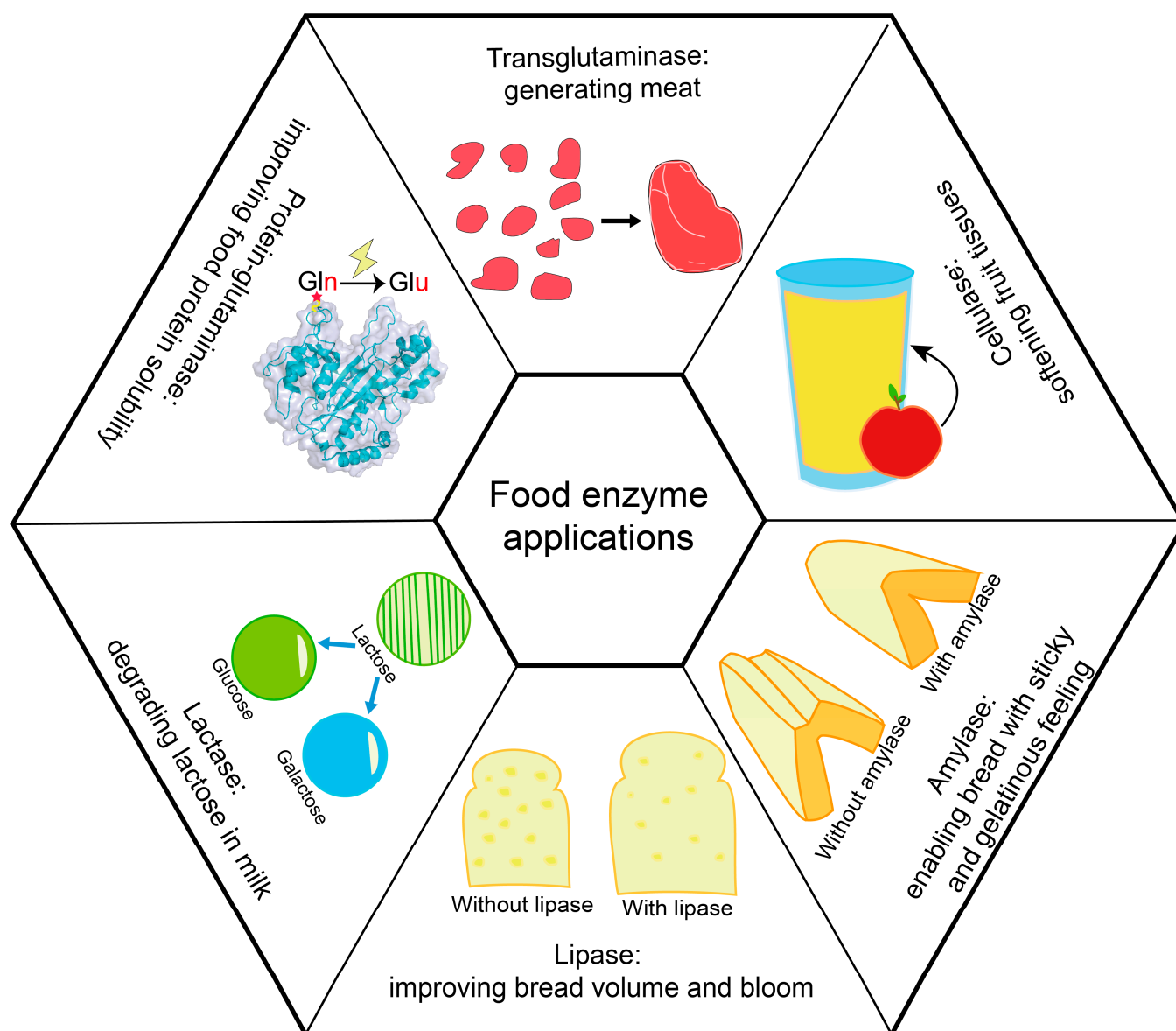


Figure 1. Sample applications of food enzymes. This figure lists six commonly used food enzymes. Transglutaminase (structure obtained from Protein Data Bank, ID: 1IU4, generated by PyMol) catalyzes the formation of heteropeptide bonds between the γ -amide group of the glutamine residue in the protein and the ϵ -amino group of Lys. Cellulase hydrolyzes cellulose to produce glucose and oligomeric fiber. Amylase hydrolyzes the glycosidic bonds inside starch. Lipase hydrolyzes triglycerides to glycerol and fatty acids. Lactase catalyzes the hydrolysis of β -D-galactoside and α -L-arabinoside. Protein-glutaminase catalyzes the deamidation of Glu residues in proteins. Protein structure generated using PyMol (Schrödinger, New York, NY, USA).

Table 1. Typical functions of food enzymes.

Enzyme (EC Number)	Catalytic Reaction	Commercial Source
Transglutaminase (EC 2.3.2.13)	Catalyzing the formation of heteropeptide bonds between the γ -amide group of the glutamine residue in the protein and the ϵ -amino group of Lys [36].	<i>Streptomyces mobaraense</i>
Laccase (EC 1.10.3.2)	Catalyzing single-electron oxidation of phenols, aromatic amines, and other electron-rich substrates [45].	<i>Aspergillus oryzae</i> , <i>Mycyceliophora thermophila</i>

Table 1. Cont.

Enzyme (EC Number)	Catalytic Reaction	Commercial Source
Protein-glutaminase (EC 3.5.1.44)	Catalyzing the deamidation of Glu residues of proteins [46].	<i>Chryseobacterium proteolyticum</i>
α -Amylase (EC 3.2.1.1)	Hydrolyzing α -1,4-glycosidic bonds inside starch [38].	<i>Bacillus licheniformis</i> , <i>Thermus Aquaticus</i>
Lactase (EC 3.2.1.108)	Catalyzing the hydrolysis of β -D-galactoside and α -L-arabinoside [40].	<i>Bacillus subtilis</i> , <i>Bifidobacterium bifidum</i>
α -Glucanase (EC 3.2.1.59); β -Glucanase (EC 3.2.1.73)	Hydrolyzing α/β -glucans [47].	<i>Bacillus subtilis</i> , <i>Bacillus amyloliquefaciens</i>
Phosphatidylinositol-specific phospholipase C (EC 3.1.4.11)	Hydrolyzing the phosphodiester bond of phosphatidylinositol to diacylglycerol and water-soluble phosphoinositol [48].	<i>Pseudomonas fluorescens</i>
Polygalacturonase (EC 3.2.1.15)	Catalyzing pectin molecule α -(1,4)-polygalacturonic acid cleavage [49].	<i>Trichoderma reesei</i> , <i>Aspergillus tubingensis</i>
Pectinesterase (EC 3.1.1.11)	Hydrolyzing pectin to produce pectinic acid and methanol [43].	<i>Trichoderma reesei</i> , <i>Aspergillus tubingensis</i>
endo- β -1,4-xylanase (EC 3.2.1.8)	Hydrolyzing xylan molecule β -1,4-glycosidic bonds [50].	<i>Trichoderma reesei</i> , <i>Thermopolyspora flexuosa</i>
Lipases (EC 3.1.1.3)	Hydrolyzing triglycerides to glycerol and fatty acids [41].	<i>Trichoderma reesei</i> , <i>Fusarium oxysporum</i>
4- α -glucanotransferase (EC 2.4.1.25)	Catalyzing the breaking of α -1,4-glycosidic bonds and the transfer of α -glucan residues within or between molecules [51].	<i>Aeribacillus pallidus</i>
Rennin (EC 3.4.4.3)	Hydrolyzing the peptide bond between Phe105-Met106 of κ -casein in milk [52].	<i>Kluyveromyces lactis</i>
Cellulase (EC 3.2.1.4)	Hydrolyzing cellulose to produce glucose and oligomeric fiber [37].	<i>Trichoderma reesei</i>
Glucose isomerase (EC 5.3.1.18)	Catalyzing isomerization of glucose to fructose [53].	<i>Streptomyces</i> , <i>Bacillus subtilis</i>
α -glucosidase (EC 3.2.1.20)	Hydrolyzing the glycosidic bond of the non-reducing end of polysaccharides or converting the α -1,4-glycosidic bond of oligosaccharides into α -1,6-glycosidic bonds [53].	<i>Saccharomycetes</i> , <i>Aspergillusniger</i>
Glucose oxidase (EC 1.1.3.4)	Oxidizing β -D-glucose to become gluconic acid and hydrogen peroxide [44].	<i>Aspergillus niger</i>
Subtilisin (EC 3.4.21.62)	Hydrolyzing proteins into amino acids [54].	<i>Bacillus subtilis</i>
Phytase (EC 3.1.3.8)	Catalyzing the removal of phosphate groups by inositol hexaphosphate [55].	<i>Natuphos</i>

The application of food enzymes in food processing has gradually been recognized with the growing attempts to use food enzymes. Several food enzymes, such as transglutaminases, laccases and lactases, can be used in processing meat, dairy, and bean products [36,40,45]. Traditional food enzymes such as peroxidase, transglutaminases and laccases are well suited for improving meat texture and flavor for novel products such as artificial meat. Recent studies showed that some food and several potential food enzymes (not commercialized) are essential in desensitization and flavoring. There has been research on using aldehyde dehydrogenase [56] and aldehyde oxidase [57] to remove the beany flavor and foul smell of soybean-based artificial meat. Studies on enzymatic degradation of soybean protein allergens have illustrated the critical effect of papain, pepsin, alkaline protease and other commercial proteases for degrading several key protein allergens such as 7S protein α subunits and 11S soybean globulin [12]. However, these proteases showed low degradation efficiency due to their specificity against these protein allergens. The diverse functions of food enzymes support their uses in food processing; however, their

insufficient stability, substrate specificity and activity can directly influence their involved applications, or result in unsatisfactory products after food processing.

3. Challenges for Food Enzyme Engineering

Naturally existing proteins usually show limitations during practical uses as a result of unsatisfactory activity or stability [58,59], which has motivated researchers to find solutions. During food processing, the conditions of high salt, high concentrations of organic solvents or high temperatures can be harsh for food enzymes. These adverse conditions mainly affect the half-life and activity of enzymes. Enzyme engineering aims to provide solutions for insufficient enzyme stability, unsatisfactory specific activity and weak solvent tolerance. Food enzyme dosages have strict standards set by organizations such as the EFSA and FDA [14]. Therefore, strengthening enzyme performance is necessary. In addition, modified enzymes with better environmental tolerance or activity can optimize the processing protocol, and reduce energy consumption by rapidly accomplishing the processing task or avoiding bacterial contamination.

Many studies have attempted to engineer food enzymes using direct evolution or rational design (Table 2) to overcome enzyme property limitations and expand their applications. The initial motivation for food enzyme modification was to meet specific applications, such as engineering the thermostability of microbial transglutaminase to satisfy its application for tofu and fish ball processing, since this processing involves a cool-down process prioritized to enzyme addition [27]. Meanwhile, previous reports suggested that protein thermostability is correlated with various harsh environmental tolerances [60]. Hence, improving enzyme thermostability may positively affect its overall stability under different conditions. More importantly, reinforcing an enzyme's thermostability and activity can extend its potential utilities. For example, modification of substrate specificity can make microbial transglutaminase specifically label one site of human hormone, which supports its potential industrial application for polymer–drug conjugation [61]. The great effort being paid for food enzyme engineering has benefited their applications, such as in engineering the thermostability of phytase that improved its use as an animal feed supplement (usually pre-mixed at 75–95 °C) [55].

Table 2. Enzyme modification and the modification aims.

Enzyme and Source	Effect of Best Variant	Aims and Reference
Transglutaminase (<i>Streptomyces mobaraensis</i>)	T_m and specific activity increased by 3.4 °C and 67.8%.	Processing tofu and fish balls at high temperatures. [27]
Glucoamylase (<i>Talaromyces leycettanus</i>)	T_m and specific activity increased by 9 °C and 305.4%.	Inducing the conversion of starch to glucose at high temperatures. [62]
Alpha-amylase (<i>Rhizopus oryzae</i>)	$t_{1/2}$ (55 °C) increased by 2.55-fold.	Optimizing winemaking protocol. [63]
Cellulase (<i>Penicillium canescens</i>)	$t_{1/2}$ (60 °C) increased by 3.4-fold.	Catalyzing the formation of gentiooligosaccharide at high temperatures. [64]
Serine protease (<i>Pseudomonas aeruginosa</i>)	T_m and specific activity increased by 5 °C and 1.4-fold.	Protease treatment at high temperatures enables fast processing and avoids bacterial contamination. [65]
Lipase (<i>Yarrowia lipolytica</i>)	$t_{1/2}$ (50 °C) increased by 70%.	Optimizing grain and oil-processing protocol. [66]
Endoglucanase (<i>Bipolaris sorokiniana</i>)	Specific activity increased by 1.5-fold.	Enabling rapid food processing. [67]
Phytase (<i>Escherichia coli</i>)	Residual activity improved by 78.4% at 90 °C.	Used as animal feed supplement. [68]
Glucose isomerase (<i>Thermoanaerobacter ethanolicus</i>)	Specific activity increased by 2-fold.	High-fructose corn syrup one-step biosynthesis. [69]
β -glucanases (<i>Bacillus terquilensis</i>)	Improved acidic tolerance, and increased specific activity by 45%.	Serving food mashing process. [70]

T_m : Melting temperature, a temperature point at which protein undergoes a reversible folding or unfolding transition. $t_{1/2}$: Half-life ($t_{1/2}$) for proteins at a specific temperature.

Whether using directed evolution or rational design for enzyme engineering, the resulting mutation sites are limited. Therefore, it is relatively difficult to obtain dramatic reinforcement of the target property using conventional engineering methods compared with screening enzymes from novel organisms, such as finding thermostable enzyme variants from thermophilic bacteria [71], or finding cold adaptive enzymes from deep-sea creatures [72]. In the past five years, achievements in using protein modeling and AI-based energy calculation for the rational design or de novo generation of functional proteins was conspicuous [22,33], especially for using de novo design to screen high-affinity protein binders, including drug proteins and antibodies [31,33]. De novo-designed proteins usually have high thermostability, since they always pursue the lowest energy structures. Meanwhile, de novo design-based novel enzyme screening dramatically reduces the difficulty of searching for potential functional enzymes from the gene library. However, de novo techniques assisting food enzymes engineering have been reported relatively less. De novo design using AI technology may provide more possibilities for expanding enzyme functions (including specific catalytic ability and stability) by generating non-existent proteins naturally. Moreover, recent progress for de novo design protein binders have highlighted the potential of directly generated enzymes with target functions.

4. De Novo Design Inspired by Highly Accurate Protein Modeling

Acknowledging the basis for protein modeling is the key to understanding protein de novo design. The protein evolution reinforced by AI has achieved great success, and led to the protein-based industry entering a new era [58,73]. Currently, AI-based tools have shown a significant impact on proteomics and structural biology, such as annotating proteins [2], searching functional proteins from the bacterial genome [74], predicting enzyme activity [75], modeling protein structures [22] and designing proteins [32]. Protein evolution reinforced by AI has contributed to the protein-based industry, which is entering a new era [58,73]. The pre-trained AI models used for predictions can save computational time and the required computing resources, which have benefited from their efficiency [76]. Machine learning (ML) and deep learning (DL) are critical subfields of AI, while the revolution of DL-based techniques dramatically benefited protein modeling and design in the past five years [22,32,77].

Most of the currently available protein design methods have adopted approaches used for protein modeling [21,26,78]. For instance, one of the examples involves treating the residue–residue contact map of the target protein as an image segmentation task, inspired using Convolutional Neural Network (CNN) [79] for protein folding prediction [80]. This method is commonly used for protein modeling and de novo design by extracting protein features [22,32,77]. Protein sequential arrangement or functional motif decomposition was thought to be like human language, which can be organized to represent certain meanings [81]. Accordingly, Natural Language Processing (NLP) methods originally used for human language translation were applied to resolving protein folding and de novo design tasks, by extracting features from protein sequences [22,82].

Accurate protein structural modeling is a complex task. A protein consists of amino acids which form its primary structure, and the residue–residue interactions within protein chains drive the formation of its secondary structure, including alpha helixes and beta-strands. A protein's tertiary structure is ultimately formed due to the spatial arrangement of its secondary structures. Experimental techniques such as crystallography [83], nuclear magnetic resonance (NMR) [84] and Cryo-EM [85] have contributed to the Protein Data Bank (PDB) [86] dataset. However, due to the high cost of resolving protein structures, the currently available data comprise less than 200 thousand natural proteins, less than UniProt-deposited (more than 230,000 thousand) natural proteins [1]. In this context, the need for acquiring protein structures has promoted the progress of protein modeling methods.

4.1. Template-Based Protein Modeling

Protein modeling can be either template- or non-template-based. MSA is a well-employed template-based method for structural modeling [87,88]. The use of MSA aims to count the matches, mismatches and gaps of aligned protein sequences for uncovering the coevolutionary features of the target sequence [89] (Figure 2). For template-based protein modeling, the conserved areas revealed by MSA show that the target sequence is aligned to available structural data, and guides local folding to prioritize the whole structural folding. MSA was used very early by MODELLER [88] and SWISS-MODEL [90] to guide structural modeling, and has subsequently been adopted by AI-based tools. AI techniques show significant advantages in extracting protein features and performing predictions (Figure 3). Protein modeling protocols have been optimized using AI-based techniques, such as conducting MSA [91] and generating residue–residue interaction (RRI) [92] networks.

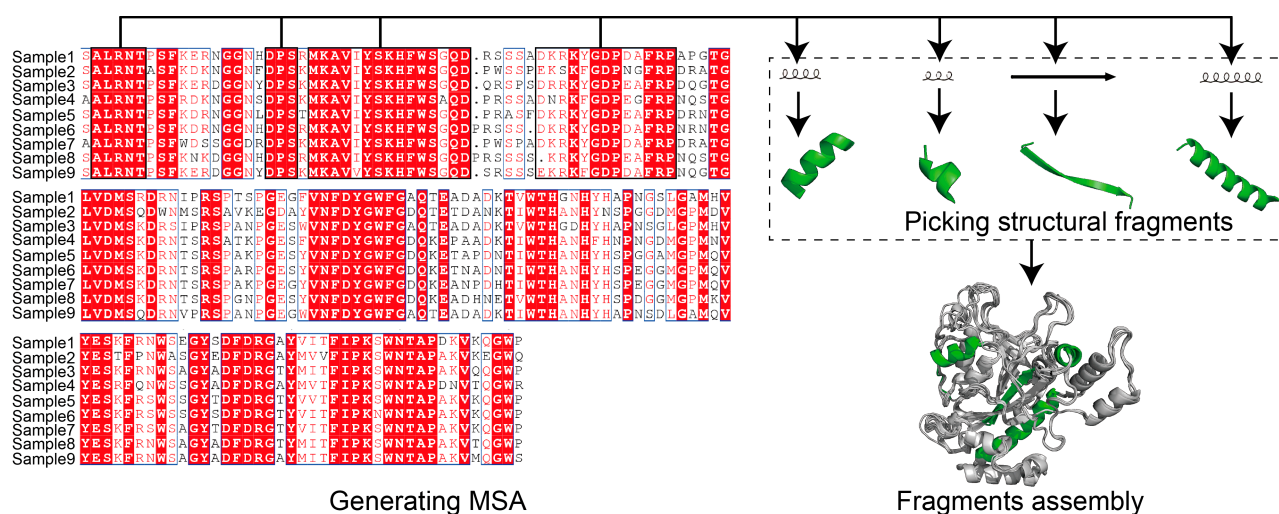


Figure 2. Using MSA for structural modeling. Multiple sequences alignment (MSA) describes aligning protein sequences to highlight the conserved region. For protein modeling, MSA information is used to guide searching of structural fragment templates prior to whole structure folding. Protein sequence alignment was generated using Jalview (Andrew Waterhouse and Geoff Barton’s group, Dundee, Scotland). Protein structure generated using PyMol (Schrödinger, New York, NY, USA).

AlphaFold-2 (AF-2) highlights that the DL model can assist protein modeling as experimentally resolved, with an average error of 1 Å [22]. Protein modeling using MSA is a key to providing better accuracy [93]. Most AI-based protein modeling tools adopted MSA as an initial step, including AF-2 [22], RF [77,87], and I-TASSER. AF-2 learned features from the combination of MSA and RRI networks to guide the structure modeling. AF-2 developed a variant of Transformer [94] named Evoformer, which mainly uses “Attention” mechanism for feature extraction and progression. “Attention” mechanism was developed for language (sequence) translation (from encoder to decoder). It has been used to recognize conserved regions of the input sequence, and to guide the template selection in protein modeling [95]. On the other side, the features of RRI maps that were built depending on amino acids distribution and their steric interaction information (represented by a 2-dimensional distance map) were also extracted using “Attention” mechanism. Solely depending on RRI maps (trained networks), guided protein folding enabled AF-2 to perform non-template-based structure modeling. Generally, AF-2 adopts a combined search function that relies on MSA and RRI information to guide the template matches, followed by the “recycling step” that uses the output structure predicted by the network to match the structural labels, in order to guide model training. The protocol of AF-2 has inspired the invention of RF [77], which focuses on speeding up the prediction time by developing a concise MSA feature extraction step.

Another classic method, “multiple threading alignments” was adopted by I-TASSER [87], although I-TASSER adopted MSA as an initial modeling step. The “threading” method was used to evaluate the fitness of smaller sequential and structural fragments with the template protein backbone. The “threading”-based template search has also been widely used for protein de novo design for accommodating structural fragments, including the methods such as Rosetta match [26] and FixBB [96]. The key for template-based protein modeling is to find reliable local compartments, since fragment assembly is thought of as an early stage of whole protein folding [97].

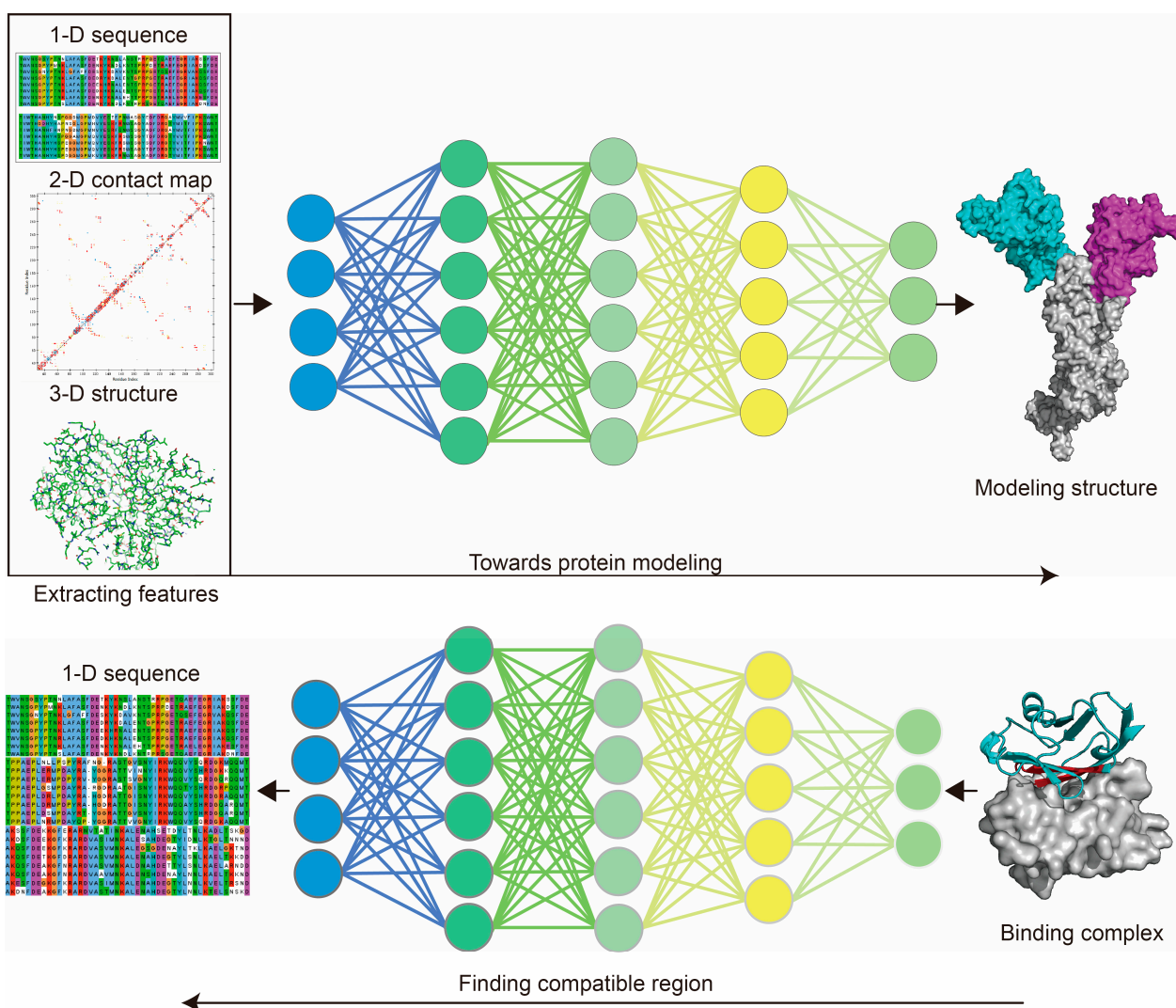


Figure 3. Deep learning used for structural modeling and design. For protein modeling, deep learning (DL) is used to learn sequence information (or MSA features) from the 1-dimensional sequences, residue–residue interaction information from the 2-dimensional contact map, or 3-dimensional structural information based on input structures, through aligning with the target structure (DL labels) to train the model. Protein de novo design takes advantage of protein modeling models by finding the compatible structure (and sequences) of the target block within a given protein to design novel proteins. Protein sequence alignment generated using Jalview. Protein 2-dimensional contact map generated using Discovery Studio 2019. Protein structure generated using PyMol.

4.2. Non-Template-Based Protein Modeling

In classic protein modeling methods, the non-template-based method was mainly stochastic sampling-based. Protein was thought to fold into its energy-minimized state in solution [98]. Monte Carlo (MC) and Markov Chain Monte Carlo (MCMC) simulations

were commonly used to simulate atom movements and interactions, which were used for protein ab initio folding [99–101]. Currently, MC and MCMC methods are commonly used for whole structural optimization, rather than for performing protein folding or for modeling protein regions that are highly disordered [102]. The main drawback of non-template-based protein modeling is the computing speed, since both MC and MCMC use extensive computing resources [76]. In addition, ab initio folding cannot guarantee modeling accuracy either. Therefore, non-template-based protein modeling methods purely relied on MC, and MCMC was gradually abandoned.

For AI-based non-template-based protein modeling, RRI information used to train protein modeling networks brought out the possibility of training non-template-based models. RRI information used for training integrated the model, including AF-2 [22], RF [77], trRosetta [103] and I-TASSER [87]. A dramatic limitation of training networks solely depending on RRI information is the requirement of massive available structural data. However, currently available structural data may not be sufficient for training a solely RRI-based model, due to PDB having many reductant structures or proteins with high identities. Meanwhile, proteins such as membrane proteins only comprise a minor portion of the whole dataset, which can result in biased predictions after training [104]. Therefore, MSA-guided fragments assembly is still necessary for protein modeling, such as for membrane proteins, despite generating MSA in computational resources consumption.

Meta AI developed ESMFold without requiring the MSA step [105]. This model was trained with protein sequential and structural information using the “Attention” mechanism, and passes the input sequence directly to the DL block during prediction progression without an MSA session. This non-template search prediction has enabled fast computing of ESMFold, approximately 60 times faster than AF-2, despite its lower accuracy than AF-2 [105]. Until this review was organized, Meta AI announced that 617 million structures were modeled, and provided public access; meanwhile, the AF-2 database currently provides over 200 million structures. High modeling accuracy can be supportive for subsequent protein design. Meanwhile, building a DL model for protein modeling has inspired the construction of AI-based protein de novo design protocols such as RFdesign [31] and Hallucination [32]. We collectively listed AI-based protein modeling methods by describing their architecture and utilities in Table 3.

Table 3. AI-based protein modeling tools.

Name	Description	Ref
AlphaFold-2	Accurate, structures can be directly downloaded from a public dataset. Slow for protein modeling using source code. Database accessed from: https://alphafold.com/ (accessed on 5 February 2020) https://www.uniprot.org/ (accessed on 5 February 2020)	[22]
ESMFold	Accurate, structures can be directly downloaded from a public dataset. Database accessed from: https://esmatlas.com/about#download_dataset (accessed on 5 February 2020)	[106]
RoseTTAFold	Accurate, support for uploading up to 20 sequences for modeling. Relatively fast for protein modeling using source code. Webserver: https://robetta.bakerlab.org/ (accessed on 5 February 2020)	[77]
I-TASSER	Accurate, support for online uploading modeling tasks and using source code. Webserver: https://zhanggroup.org/I-TASSER/ (accessed on 5 February 2020)	[107]
trRosetta	Accurate, support for online uploading modeling tasks and using source code. Webserver: https://yanglab.nankai.edu.cn/trRosetta/help/ (accessed on 5 February 2020)	[103]
A-Prot	Only support for source code modeling. Source code: https://github.com/arontier/A_Prot_Paper (accessed on 5 February 2020)	[108]
Colossal-AI	Only support for source code modeling. Source code: https://github.com/hpcaitech/ColossalAI (accessed on 5 February 2020)	[109]

5. De Novo Design of Food Enzymes

Protein de novo design refers to the use of rational ways to design novel proteins that do not exist naturally. It is described as solving the inverse problem of protein modeling [110]. De novo design can be used to design sequences compatible with the given structural backbone [111,112], generate a missing block within a given structure [30,31], or “hallucinate” de novo structures with sequences [32] (Figure 4). Previously, physics-based modules such as Rosetta Match [26], Fixbb [96] and Remodel [113] (Table 4) have been successfully implemented to search optimized sequences based on a given structure (Table 4). These physics-based methods adopt structural fragments replacement or insertion. They can also be used for large secondary structure replacement, using Rosetta-based modules such as FunFolDes [114] and LUCS [115]. Physics-based methods have mainly been applied in the past ten years. Users can design blueprints and assemble protein by fragment accommodation attempts using TopoBuilder [34]. However, physics-based de novo design methods require large sampling sizes, which is time-consuming. AI-based methods emerged to enable fast computing, highly accurate structural modeling, and the design of functional enzymes. The successful implementation of AF-2 brought out possibilities for DL to resolve de novo design tasks. Currently, AI-methods such as Recurrent Neural Network (RNN) [116], CNN, Graph Neural Networks (GNN) [117] and Generative Adversarial Nets (GAN) [118] have highly participated in integrating de novo design models (Table 4). Using AI-based techniques for de novo design functional proteins is showing an apparent upward trend [119]. In the following section, we review the progression of de novo designs for food enzyme engineering, and introduce AI-based de novo design methods available for further uses.

Table 4. De novo design methods.

Name	Description	Ref
Match	Physics-based, structural-based, aims at designing de novo functional enzymes using fragment attempts.	[26]
Fixbb	Physics-based, structural-based, fit for short area design.	[96]
Remodel	Physics-based, structural-based, fit for short area design, can be used for protein reassembling.	[113]
FunFolDes	Physics-based, structural-based, blueprint-based whole protein de novo design.	[114]
LUCS	Physics-based, structural-based, fit for designing loop-helix-loop, loop-strand-loop.	[115]
TopoBuilder	Physics-based, structural-based, blueprint-based whole protein de novo design.	[34]
Protein Hallucination	AI-based, de novo design of whole protein structures with compatible sequences based on input sequence.	[32]
RFDdesign	AI-based, “inpainting” module: designing short blocks based on given structure; “hallucination” module: designing short blocks based on a given structure, can be used for designing functional motifs, supporting receptor and donor structure.	[31]
ProteinMPNN	AI-based, fast designing compatible sequences using an input structure.	[33]
DenseCPD	AI-based, only supports uploaded tasks online, online server: http://protein.org.cn/densecpd.html	[120]
ProteinGAN	AI-based, GAN model for real-time generating sequences based on a set of input protein sequences (within the same protein family).	[35]
ProtGPT2	AI-based, pre-trained model for generating sequences based on input sequence.	[121]
Diffusion model-based	AI-based, pre-trained model generating protein structures.	[122]

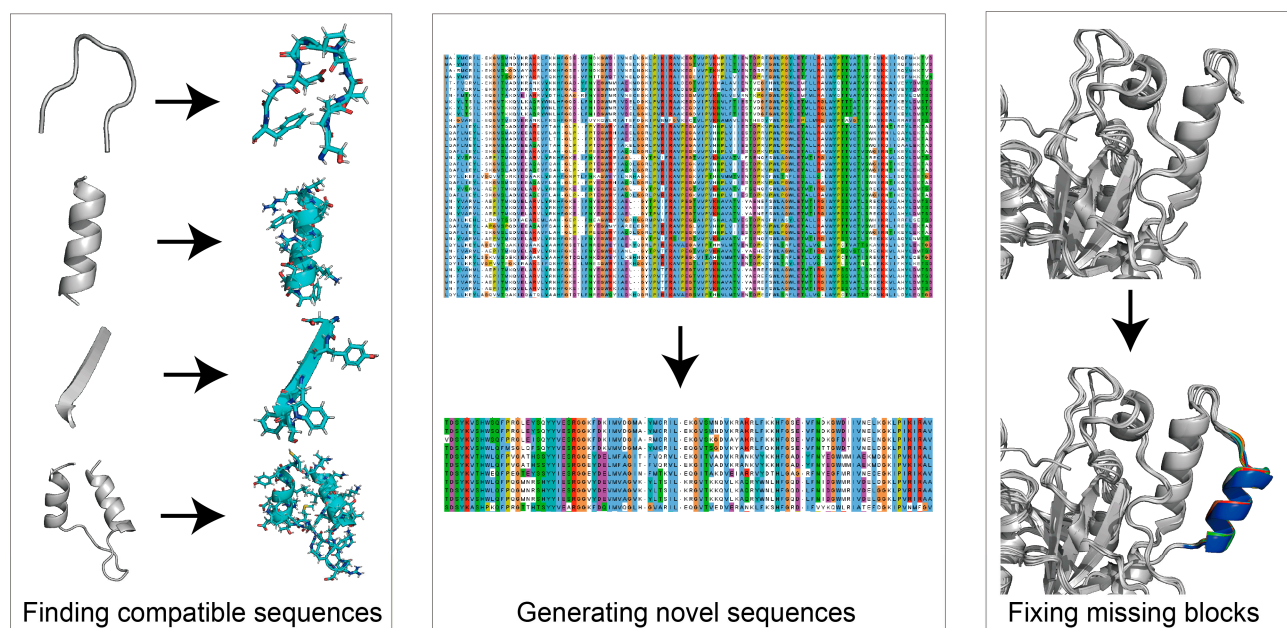


Figure 4. Utilities of protein *de novo* design. Protein *de novo* design can be used for finding compatible sequences of target areas (ProteinMPNN), generating novel sequences based on input sequences (ProteinGAN), or fixing missing blocks of given structures (Hallucination). Protein sequence alignment was generated using Jalview. Protein structure generated using PyMol.

5.1. Current Solutions

Generating functional food enzymes using *de novo* design has been successfully implemented. Note that *de novo* design food enzymes can be traced to 20–30 years ago, such as *de novo* designs of superoxide dismutase-like enzyme [123] and esterase [124]. The initial challenge for enzyme *de novo* design is to sustain its activity, whereas earlier algorithms showed that accommodating catalytic residues in the pocket was difficult and uncontrollable [123,124]. As a result, the functions of *de novo*-designed food enzymes were not satisfactory. Novel algorithms were brought out to overcome this limitation, by focusing on the accommodation of enzyme active sites. The standout algorithm was Rosetta Match, which takes advantage of the structural fragments dataset and fragments' replacement strategy for designing protein functional motifs [78,125]. This algorithm has been used for the *de novo* design of pullulanase [126] and esterase [127]. It should be noted that the combined use of several techniques such as consensus design, MSA and fragments' replacement contributed to the high success rate for designing novel enzymes. Several reports showed that the designed food enzymes notably outperformed native enzymes in terms of their catalytic activity [127] or stability [126]. However, it was reported that *de novo* designs of food enzymes were mostly physics-based, which highly relies on computational resources and empirical factors.

AI-based *de novo* design techniques merged to address these limitations and enable fast design. It was recently reported that ProteinGAN [35] adopted GAN to generate *de novo* protein sequences. ProteinGAN was trained using protein sequence data through the “one-hot” method, in order to convert these sequences into a digital array. The network was built using ResNet [128] (CNN derived) and “Attention” mechanism for extracting features from input sequences (by discriminator), and generating novel sequences (by generator). In the case study, the sequences from the malate dehydrogenase family were used, and the authors showed that 24% of *de novo*-designed sequences were expressed experimentally with enzymatic functions. The initial successes of AI-based techniques for enzyme *de novo* design successes have triggered the rapid development of algorithms. However, many novel AI-based *de novo* design techniques have not been applied to food enzymes, which still show great potential.

5.2. AI-Based De Novo Design Techniques

5.2.1. Protein Hallucination

Protein “Hallucination” [32] was inspired by Google DeepDream, and was introduced in 2020. Hallucination can be used to generate de novo protein structures with compatible sequences, without requesting the protein backbone. Hallucination generates a stabilized structure using a random input sequence, and it adopts the method from trRosetta to describe the 2-dimensional RRI map of the input sequence [103]. An independent test used MCMC for tracking RRI map changes upon mutations, and revealed a sharpened RRI map indicating a stabilized structure. Based on the results, Hallucination employed multiple rounds of iterations by introducing mutations and tracking the RRI map changes to find the optimized structure with a compatible sequence. The prediction efficiency of Hallucination was experimentally characterized by circular dichroism (CD) spectroscopy; there were 62 out of 129 samples solubly expressed that showed the desired 2-dimensional structures as predicted. The expressed proteins showed high thermostability, with an average melting temperature above 70 °C, and the crystallized structures highly assembled the predictions. The released code enables users to input a sequence and adjust the length of the output sequence. However, the limitation is that the user cannot control the motif insertion position, since the output sequence was randomly mutated based on the input sequence.

5.2.2. RFDesign

The accuracy of structural modeling was significantly improved using AF-2 and RF compared with trRosetta. RFDesign took advantage of AF-2 and RF [31], which provided an updated version of protein Hallucination, as well as an “inpainting” module for rebuilding missing sequences and structures by enabling reserved functional regions. The authors attempted to use RF [77] and AF-2 [22] for generating protein features rather than trRosetta [103], and it showed significant advantages. RFDesign hallucination adopted a similar approach as the previous version, to reward those mutations that could stabilize the structure. For the inpainting design module, an iteration method mimics the “recycled” step of AF-2, and RF (RF-Nov05-2021 version) was brought out to continuously introduce mutations to the output structure, followed by overall folding refinement and simultaneously scoring the output structure.

Previously, de novo design proteins were always based on input proteins that could not design proteins with specific functions. To resolve this challenge, RFDesign hallucination [31] developed a combined training loss that scored the repulsive and binding forces between the input protein and its binding partner (can be protein or ligand); those mutations that showed correct binding behavior were retained for iteration. In the case study, the authors used RFDesign hallucination to successfully design the interactive surface of programmed cell death protein 1 (PD-1), and reinforced its binding affinity against programmed cell death ligand 1 (PD-L1). In the released package, users can design specific sites within the input sequence and indicate the receptor file. The protein Hallucination family solved two major challenges in protein de novo design, including generating de novo motifs and designing functional protein binders.

5.2.3. ProteinMPNN

ProteinMPNN [33] was used for the de novo design of compatible sequences based on the input structure backbone. ProteinMPNN was built using a modified Transformer [94] network, which adopted the encoder block to extract features, and the decoder block to translate these features to “readable information”. ProteinMPNN was trained using protein backbone information, including C-alpha atom–atom distance, orientation and backbone dihedral angles, rather than requesting MSA information. The training method contributed to the high capacity for modeling single sequences, despite the fact that AF-2 [22] and RF [77] highly required MSA information for protein folding prediction. ProteinMPNN showed a sequence recovery rate of 52.4%, and could serve for the the design of protein

monomers or cyclic oligomers. In a recent report, ProteinMPNN was successfully used for designing binder protein, highlighting its great potential [129].

5.2.4. DenseCPD

DenseCPD [120] is a sequence design method based on the input structure. This model is trained using protein atomic information. The training structures are prepared by splitting the protein box into small voxels containing only 1 atom (1 Å). DenseCPD learns the atom distribution information from the structures using DenseNet [130] (CNN derived), and predicts the probability of amino acids that build the input protein backbone. This approach displayed higher accuracy than the later released ABACUS-R [131], despite ABACUS-R relying on Transformer to extract more information from both protein sequence and structure. The aim of using DenseCPD is to find the most suitable sequences for the protein backbone, and this model is currently supported only for tasks submitted online.

5.2.5. Unsupervised Learning Methods

The large sample size of deposited protein structural and sequential data promoted the development of unsupervised learning (UL) [132] methods. ProtGPT2 [121] is a pre-trained language-based UL model which generates protein sequences based on training sequences. ProtGPT2 employs a modified Transformer by taking advantage of only the decoder side. This network was trained using UniRef50 [133], which contains approximately 49 million protein sequences that are highly diverse. Compared with ProteinGAN [35], the trained ProtGPT2 model is able to generate de novo sequences within a few seconds based on the user input sequences, which is convenient.

Diffusion model (DM) [134] was recently brought out as a generative network, through adding noises and iterative denoising to recover the targets. DM previously showed state-of-art performance for synthesizing images and generating videos. Namrata and Tudor implemented DM to generate de novo protein design, which aims to build missing areas or find compatible sequences [122]. This model was trained and adopted protein information, including protein full-atom coordinates, protein sequences and amino acid side-chain conformations, and showed reliable performance during the validation tests.

6. Limitations of De Novo Design Techniques and Opportunities for Food Enzyme Engineering

There are still apparent challenges for the generation of functional enzymes using de novo design. Successful implementation of de novo design was achieved with high-affinity protein binders, such as IL-2/IL-15 homologues [135] and ACE2 protein homologues, within 67 days [136]. This evidence showed that the means for designing desired proteins was partially addressed. It is easy to de novo generate thermostable variants of target enzymes, since AI models were trained to output the lowest energy structures [32]. However, enzyme-catalyzed reactions require a certain distance between the catalytic residue of an enzyme, and the residue being catalyzed of a substrate. These distance constraints may not prioritize generating novel enzymes, although we can still use an additional protocol to filter binding complexes that satisfy the distance. For protein binder design, the binding or repulsive forces can be used to indicate the protein–partner binding affinity [137]. However, in enzyme-triggered catalysis, the relationship between enzyme–substrate binding affinity (k_m) and catalytic activity (k_{cat}) are still under debate. Such consequences highlight the fact that much knowledge remains unknown about enzyme-induced catalysis.

The success rate is also a problem for de novo design. The success rate of designing malate dehydrogenase using ProteinGAN only achieved 24% [35]. The DL-models, such as ProteinGAN and DLKcat [75], have mainly been used for extracting information from protein sequences, which means the enzyme structures are not correlated with prediction; however, a lack of structural indication may negatively impact prediction accuracy. Note that DLKcat [75] was specifically used for predicting enzyme k_{cat} (turnover number of an enzyme), which can predict k_{cat} changes toward specific substrates while enzyme sequences

change. Nonetheless, whether using physics-based or AI-based methods, de novo-designed proteins remained with low expression rates, or were difficult to fold [31,138]. Meanwhile, few studies crystallized their predicted structures, which showed noticeable RMSD variations [32,139]. To those who are designing binding partner tasks, minor RMSD differences between designed and actual structures can result in non-binding or bad contacts (such as strong repulsive forces). De novo designing enzymes with desired functions is a huge challenge.

Challenges always coexist with opportunities. De novo design have shown great potential for generating non-naturally existing and diverse functional variants, which provides possibilities for their currently involved and novel applications. A few factors would benefit DL-model architecture for generating various functional food enzymes: firstly, training the model by adding structural information; secondly, collecting more information related to enzymes and their k_{cat} against certain substrates; and thirdly, investigating correlations between complex binding and enzyme activity.

7. Conclusions

This review described the functions and applications of food enzymes, and introduced the need for engineering enzymes to satisfy their applications or expand their utility horizon. The limitations of using conventional enzyme engineering methods are evident, as fewer mutation rates may make it difficult to strengthen the target functions of enzymes. The advances in AI-based protein modeling and de novo design methods were reviewed. The successful implementations of de novo design for functional protein binders highlight the possibilities of using de novo design for functional enzymes. The challenges for the de novo design of enzymes come from the limited knowledge of enzyme–substrate binding behaviors and their correlated activities; the architecture of the models; and insufficient training data. These issues await future investigation.

Author Contributions: Conceptualization, X.W., Y.T., K.X. and S.L.; writing—original draft preparation, X.W.; writing—review and editing, X.W.; supervision, J.Z.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (2021YFC2101400), the Natural Science Foundation of Jiangsu Province (BK20202002) and the Starry Night Science Fund of Zhejiang University Shanghai Institute for Advanced Study (Grant No. SN-ZJU-SIAS-0013).

Data Availability Statement: All data related to this study have been included in the manuscript.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Consortium, T.U. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2020**, *49*, D480–D489. [[CrossRef](#)]
2. Bileschi, M.L.; Belanger, D.; Bryant, D.H.; Sanderson, T.; Carter, B.; Sculley, D.; Bateman, A.; DePristo, M.A.; Colwell, L.J. Using deep learning to annotate the protein universe. *Nat. Biotechnol.* **2022**, *40*, 932–937. [[CrossRef](#)]
3. Lee, J.; Chen, W.-H.; Park, Y.-K. Recent achievements in platform chemical production from food waste. *Bioresour. Technol.* **2022**, *366*, 128204. [[CrossRef](#)]
4. Reshmy, R.; Philip, E.; Sirohi, R.; Tarafdar, A.; Arun, K.B.; Madhavan, A.; Binod, P.; Kumar Awasthi, M.; Varjani, S.; Szakacs, G.; et al. Nanobiocatalysts: Advancements and applications in enzyme technology. *Bioresour. Technol.* **2021**, *337*, 125491. [[CrossRef](#)]
5. Collados, A.; Conversa, V.; Fombellida, M.; Rozas, S.; Kim, J.H.; Arbolea, J.-C.; Román, M.; Perezábad, L. Applying food enzymes in the kitchen. *Int. J. Gastron. Food Sci.* **2020**, *21*, 100212. [[CrossRef](#)]
6. Dixit, M.; Chhabra, D.; Shukla, P. Optimization of endoglucanase-lipase-amylase enzyme consortium from *Thermomyces lanuginosus* VAPS25 using Multi-Objective genetic algorithm and their bio-deinking applications. *Bioresour. Technol.* **2023**, *370*, 128467. [[CrossRef](#)]
7. Katsila, T.; Siskos, A.P.; Tamvakopoulos, C. Peptide and protein drugs: The study of their metabolism and catabolism by mass spectrometry. *Mass Spectrom. Rev.* **2012**, *31*, 110–133. [[CrossRef](#)] [[PubMed](#)]
8. Gagner, J.E.; Kim, W.; Chaikof, E.L. Designing protein-based biomaterials for medical applications. *Acta Biomater.* **2014**, *10*, 1542–1557. [[CrossRef](#)] [[PubMed](#)]

9. Ramli, A.N.M.; Hong, P.K.; Abdul Manas, N.H.; Wan Azelee, N.I. Chapter 25—An overview of enzyme technology used in food industry. In *Value-Addition in Food Products and Processing Through Enzyme Technology*; Kuddus, M., Aguilar, C.N., Eds.; Academic Press: Cambridge, MA, USA, 2022; pp. 333–345. [\[CrossRef\]](#)
10. Fernandes, P. Enzymes in food processing: A condensed overview on strategies for better biocatalysts. *Enzym. Res.* **2010**, *2010*, 862537. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Gagaoua, M.; Dib, A.L.; Lakhdara, N.; Lamri, M.; Botineştean, C.; Lorenzo, J.M. Artificial meat tenderization using plant cysteine proteases. *Curr. Opin. Food Sci.* **2021**, *38*, 177–188. [\[CrossRef\]](#)
12. Wang, T.; Qin, G.-X.; Sun, Z.-W.; Zhao, Y. Advances of research on glycinin and β -conglycinin: A review of two major soybean allergenic proteins. *Crit. Rev. Food Sci. Nutr.* **2014**, *54*, 850–862. [\[CrossRef\]](#)
13. Lee, K.D.; Lo, C.G.; Warthesen, J.J. Removal of bitterness from the bitter peptides extracted from cheddar cheese with peptidases from *Lactococcus lactis* ssp. *cremoris* SK111. *J. Dairy Sci.* **1996**, *79*, 1521–1528. [\[CrossRef\]](#)
14. Borchers, A.; Teuber, S.S.; Keen, C.L.; Gershwin, M.E. Food safety. *Clin. Rev. Allergy Immunol.* **2010**, *39*, 95–141. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Enzymes market size to reach \$17.8 bn in 2028, industry trend—Rising demand for high-quality pharmaceutical, food & beverage and personal care products. *Focus Catal.* **2022**, *2022*, 2–3. [\[CrossRef\]](#)
16. Global industrial biotechnology market report 2022: Development of new and improved enzymes and other biocatalysts key for future developments. *Focus Catal.* **2022**, *2022*, 2–3. [\[CrossRef\]](#)
17. Famiglietti, M.; Mirpoor, S.F.; Giosafatto, C.V.L.; Mariniello, L. Enzyme assisted food processing. In *Reference Module in Food Science*; Elsevier: Amsterdam, The Netherlands, 2023. [\[CrossRef\]](#)
18. Madhavan, A.; Arun, K.B.; Binod, P.; Sirohi, R.; Tarafdar, A.; Reshmy, R.; Kumar Awasthi, M.; Sindhu, R. Design of novel enzyme biocatalysts for industrial bioprocess: Harnessing the power of protein engineering, high throughput screening and synthetic biology. *Bioresour. Technol.* **2021**, *325*, 124617. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Ferreira, P.; Fernandes, P.A.; Ramos, M.J. Modern computational methods for rational enzyme engineering. *Chem Catal.* **2022**, *2*, 2481–2498. [\[CrossRef\]](#)
20. Liu, Z.; Chen, H.; Wold, E.A.; Zhou, J. 2.13—Small-molecule inhibitors of protein–protein interactions. In *Comprehensive Medicinal Chemistry III*; Chackalamannil, S., Rotella, D., Ward, S.E., Eds.; Elsevier: Oxford, UK, 2017; pp. 329–353. [\[CrossRef\]](#)
21. Park, H.; Bradley, P.; Greisen, P.; Liu, Y.; Mulligan, V.K.; Kim, D.E.; Baker, D.; DiMaio, F. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* **2016**, *12*, 6201–6212. [\[CrossRef\]](#)
22. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [\[CrossRef\]](#)
23. Daniel, L.; Aikaterini, A.; Simhadri, V.L.; Katagiri, N.H.; Wojciech, J.; Sauna, Z.E.; Chava, K.S. Recent advances in (therapeutic protein) drug development. *F1000Research* **2017**, *6*, 113. [\[CrossRef\]](#)
24. Setiawan, D.; Brender, J.; Zhang, Y. Recent advances in automated protein design and its future challenges. *Expert Opin. Drug Discov.* **2018**, *13*, 587–604. [\[CrossRef\]](#)
25. Dimitrov, D.S. Therapeutic proteins. In *Therapeutic Proteins: Methods and Protocols*; Voynov, V., Caravella, J.A., Eds.; Humana Press: Totowa, NJ, USA, 2012; pp. 1–26. [\[CrossRef\]](#)
26. Kellogg, E.H.; Leaver-Fay, A.; Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins: Struct. Funct. Bioinform.* **2011**, *79*, 830–838. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Wang, X.; Du, J.; Zhao, B.; Wang, H.; Rao, S.; Du, G.; Zhou, J.; Chen, J.; Liu, S. Significantly improving the thermostability and catalytic efficiency of *Streptomyces mobaraensis* transglutaminase through combined rational design. *J. Agric. Food Chem.* **2021**, *69*, 15268–15278. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Khersonsky, O.; Lipsh, R.; Avizemer, Z.; Ashani, Y.; Goldsmith, M.; Leader, H.; Dym, O.; Rogotner, S.; Trudeau, D.L.; Prilusky, J.; et al. Automated design of efficient and functionally diverse enzyme repertoires. *Mol. Cell* **2018**, *72*, 178–186.e175. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Langan, R.A.; Boyken, S.E.; Ng, A.H.; Samson, J.A.; Dods, G.; Westbrook, A.M.; Nguyen, T.H.; Lajoie, M.J.; Chen, Z.; Berger, S.; et al. De novo design of bioactive protein switches. *Nature* **2019**, *572*, 205–210. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Basanta, B.; Bick, M.J.; Bera, A.K.; Norn, C.; Chow, C.M.; Carter, L.P.; Goreshnik, I.; Dimaio, F.; Baker, D. An enumerative algorithm for de novo design of proteins with diverse pocket structures. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 22135–22145. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Wang, J.; Lisanza, S.; Juergens, D.; Tischer, D.; Watson, J.L.; Castro, K.M.; Ragotte, R.; Saragovi, A.; Milles, L.F.; Baek, M.; et al. Scaffolding protein functional sites using deep learning. *Science* **2022**, *377*, 387–394. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Anishchenko, I.; Pellock, S.J.; Chidyausiku, T.M.; Ramelot, T.A.; Ovchinnikov, S.; Hao, J.; Bafna, K.; Norn, C.; Kang, A.; Bera, A.K.; et al. De novo protein design by deep network hallucination. *Nature* **2021**, *600*, 547–552. [\[CrossRef\]](#)
33. Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R.J.; Milles, L.F.; Wicky, B.I.M.; Courbet, A.; de Haas, R.J.; Bethel, N.; et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **2022**, *378*, 49–56. [\[CrossRef\]](#)
34. Harteveld, Z.; Bonet, J.; Rosset, S.; Yang, C.; Sesterhenn, F.; Correia, B.E. A generic framework for hierarchical de novo protein design. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2206111119. [\[CrossRef\]](#)

35. Repecka, D.; Jauniskis, V.; Karpus, L.; Rembeza, E.; Rokaitis, I.; Zrimec, J.; Poviloniene, S.; Laurynenas, A.; Viknander, S.; Abuajwa, W.; et al. Expanding functional protein sequence spaces using generative adversarial networks. *Nat. Mach. Intell.* **2021**, *3*, 324–333. [\[CrossRef\]](#)
36. Strop, P. Versatility of microbial transglutaminase. *Bioconj. Chem.* **2014**, *25*, 855–862. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Awasthi, M.K.; Wong, J.W.C.; Kumar, S.; Awasthi, S.K.; Wang, Q.; Wang, M.; Ren, X.; Zhao, J.; Chen, H.; Zhang, Z. Biodegradation of food waste using microbial cultures producing thermostable α -amylase and cellulase under different pH and temperature. *Bioresour. Technol.* **2018**, *248*, 160–170. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Zhang, Y.; Chen, Y.; Chen, J. The starch hydrolysis and aroma retention caused by salivary α -amylase during oral processing of food. *Curr. Opin. Food Sci.* **2022**, *43*, 237–245. [\[CrossRef\]](#)
39. Shimizu, M.; Sawashita, N.; Morimatsu, F.; Ichikawa, J.; Taguchi, Y.; Ijiri, Y.; Yamamoto, J. Antithrombotic papain-hydrolyzed peptides isolated from pork meat. *Thromb. Res.* **2009**, *123*, 753–757. [\[CrossRef\]](#)
40. Sutay Kocabaş, D.; Lyne, J.; Ustunol, Z. Hydrolytic enzymes in the dairy industry: Applications, market and future perspectives. *Trends Food Sci. Technol.* **2022**, *119*, 467–475. [\[CrossRef\]](#)
41. Salgado, C.A.; dos Santos, C.I.A.; Vanetti, M.C.D. Microbial lipases: Propitious biocatalysts for the food industry. *Food Biosci.* **2022**, *45*, 101509. [\[CrossRef\]](#)
42. Morata, A.; Vejarano, R.; Ridolfi, G.; Benito, S.; Palomero, F.; Uthurry, C.; Tesfaye, W.; González, C.; Suárez-Lepe, J.A. Reduction of 4-ethylphenol production in red wines using HCD+ yeasts and cinnamyl esterases. *Enzym. Microb. Technol.* **2013**, *52*, 99–104. [\[CrossRef\]](#)
43. Kyriakidis, N.B. Use of pectinesterase for detection of hydrocolloids addition in natural orange juice. *Food Hydrocoll.* **1999**, *13*, 497–500. [\[CrossRef\]](#)
44. Ge, L.; Zhao, Y.-s.; Mo, T.; Li, J.-r.; Li, P. Immobilization of glucose oxidase in electrospun nanofibrous membranes for food preservation. *Food Control* **2012**, *26*, 188–193. [\[CrossRef\]](#)
45. Li, X.; Li, S.; Liang, X.; McClements, D.J.; Liu, X.; Liu, F. Applications of oxidases in modification of food molecules and colloidal systems: Laccase, peroxidase and tyrosinase. *Trends Food Sci. Technol.* **2020**, *103*, 78–93. [\[CrossRef\]](#)
46. Yamaguchi, S.; Jeenes, D.J.; Archer, D.B. Protein-glutaminase from *Chryseobacterium proteolyticum*, an enzyme that deamidates glutamyl residues in proteins. *Eur. J. Biochem.* **2001**, *268*, 1410–1421. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Caruso, M.A.; Piermaria, J.A.; Abraham, A.G.; Medrano, M. β -glucans obtained from beer spent yeasts as functional food grade additive: Focus on biological activity. *Food Hydrocoll.* **2022**, *133*, 107963. [\[CrossRef\]](#)
48. Paziak-Domańska, B.; Bogusławska, E.; Więckowska-Szakiel, M.; Kołowski, R.; Różalska, B.; Chmiela, M.; Kur, J.; Dąbrowski, W.; Rudnicka, W. Evaluation of the API test, phosphatidylinositol-specific phospholipase C activity and PCR method in identification of *Listeria monocytogenes* in meat foods. *FEMS Microbiol. Lett.* **1999**, *171*, 209–214. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Long, J.; Li, X.; Xue, L.; Xie, Z.; Jiao, A.; Bai, Y.; Zhou, X.; Chen, L.; Qiu, C.; Xu, X.; et al. Continuous hydrolysis of mango peel pectin for the production of antibacterial pectic oligosaccharides in packed-bed reactor using immobilized polygalacturonase. *Food Biosci.* **2022**, *50*, 102117. [\[CrossRef\]](#)
50. Fernandes de Souza, H.; Aguiar Borges, L.; Dédalo Di Próspero Gonçalves, V.; Vitor dos Santos, J.; Sousa Bessa, M.; Fronja Carosia, M.; Vieira de Carvalho, M.; Viana Brandi, I.; Setsuko Kamimura, E. Recent advances in the application of xylanases in the food industry and production by actinobacteria: A review. *Food Res. Int.* **2022**, *162*, 112103. [\[CrossRef\]](#)
51. Kim, Y.-L.; Mun, S.; Park, K.-H.; Shim, J.-Y.; Kim, Y.-R. Physicochemical functionality of 4- α -glucanotransferase-treated rice flour in food application. *Int. J. Biol. Macromol.* **2013**, *60*, 422–426. [\[CrossRef\]](#)
52. Microbial rennin with enhanced milk-clotting activity. *Trends Food Sci. Technol.* **1997**, *8*, 177. [\[CrossRef\]](#)
53. Kirimura, K.; Cao, W.; Onda, Y.; Yoshioka, I.; Ishii, Y. Selective and high-yield production of ethyl α -D-glucopyranoside by the α -glucosyl transfer enzyme of *Xanthomonas campestris* WU-9701 and glucose isomerase. *J. Biosci. Bioeng.* **2022**, *134*, 220–225. [\[CrossRef\]](#)
54. Suzuki, K.; Nakamura, M.; Sato, N.; Futamura, K.; Matsunaga, K.; Yagami, A. Nattokinase, a subtilisin family serine protease, is a novel allergen contained in the traditional Japanese fermented food natto. *Allergol. Int.* **2022**, *in press*. [\[CrossRef\]](#)
55. Coutinho, T.C.; Tardioli, P.W.; Farinas, C.S. Phytase immobilization on hydroxyapatite nanoparticles improves its properties for use in animal feed. *Appl. Biochem. Biotechnol.* **2020**, *190*, 270–292. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Chiba, H.; Takahashi, N.; Sasaki, R. Enzymatic improvement of food flavor II. Removal of beany flavor from soybean products by aldehyde dehydrogenase. *Agric. Biol. Chem.* **1979**, *43*, 1883–1889. [\[CrossRef\]](#)
57. Maheshwari, P.; Murphy, P.A.; Nikolov, Z.L. Characterization and application of porcine liver aldehyde oxidase in the off-flavor reduction of soy proteins. *J. Agric. Food Chem.* **1997**, *45*, 2488–2494. [\[CrossRef\]](#)
58. Bornscheuer, U.T.; Huisman, G.W.; Kazlauskas, R.J.; Lutz, S.; Moore, J.C.; Robins, K. Engineering the third wave of biocatalysis. *Nature* **2012**, *485*, 185–194. [\[CrossRef\]](#)
59. Wang, X.; Yang, P.; Zhao, B.; Liu, S. AI-assisted food enzymes design and engineering: A critical review. *Syst. Microbiol. Biomanufacturing* **2022**, *3*, 75–87. [\[CrossRef\]](#)
60. Modarres, H.P.; Mofrad, M.R.; Sanati-Nezhad, A. Protein thermostability engineering. *RSC Adv.* **2016**, *6*, 115252–115270. [\[CrossRef\]](#)
61. Buchardt, J.; Selvig, H.; Nielsen, P.F.; Johansen, N.L. Transglutaminase-mediated methods for site-selective modification of human growth hormone. *Pept. Sci.* **2010**, *94*, 229–235. [\[CrossRef\]](#)

62. Tong, L.; Zheng, J.; Wang, X.; Wang, X.; Huang, H.; Yang, H.; Tu, T.; Wang, Y.; Bai, Y.; Yao, B.; et al. Improvement of thermostability and catalytic efficiency of glucoamylase from *Talaromyces leycettanus* JCM12802 via site-directed mutagenesis to enhance industrial saccharification applications. *Biotechnol. Biofuels* **2021**, *14*, 202. [\[CrossRef\]](#)
63. Li, S.; Yang, Q.; Tang, B. Improving the thermostability and acid resistance of *Rhizopus oryzae* alpha-amylase by using multiple sequence alignment based site-directed mutagenesis. *Biotechnol. Appl. Biochem.* **2020**, *67*, 677–684. [\[CrossRef\]](#)
64. Dotsenko, A.S.; Dotsenko, G.S.; Rozhkova, A.M.; Zorov, I.N.; Sinitsyn, A.P. Rational design and structure insights for thermostability improvement of *Penicillium verruculosum* Cel7A cellobiohydrolase. *Biochimie* **2020**, *176*, 103–109. [\[CrossRef\]](#)
65. Ashraf, N.M.; Krishnagopal, A.; Hussain, A.; Kastner, D.; Sayed, A.M.M.; Mok, Y.-K.; Swaminathan, K.; Zeeshan, N. Engineering of serine protease for improved thermostability and catalytic activity using rational design. *Int. J. Biol. Macromol.* **2019**, *126*, 229–237. [\[CrossRef\]](#)
66. Zhang, H.; Sang, J.; Zhang, Y.; Sun, T.; Liu, H.; Yue, R.; Zhang, J.; Wang, H.; Dai, Y.; Lu, F.; et al. Rational design of a *Yarrowia lipolytica* derived lipase for improved thermostability. *Int. J. Biol. Macromol.* **2019**, *137*, 1190–1198. [\[CrossRef\]](#)
67. Aich, S.; Datta, S. Engineering of a highly thermostable endoglucanase from the GH7 family of *Bipolaris sorokiniana* for higher catalytic efficiency. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 3935–3945. [\[CrossRef\]](#)
68. Fakhravar, A.; Hesampour, A. Rational design-based engineering of a thermostable phytase by site-directed mutagenesis. *Mol. Biol. Rep.* **2018**, *45*, 2053–2061. [\[CrossRef\]](#)
69. Jin, L.-Q.; Jin, Y.-T.; Zhang, J.-W.; Liu, Z.-Q.; Zheng, Y.-G. Enhanced catalytic efficiency and thermostability of glucose isomerase from *Thermoanaerobacter ethanolicus* via site-directed mutagenesis. *Enzym. Microb. Technol.* **2021**, *152*, 109931. [\[CrossRef\]](#)
70. Li, Z.; Niu, C.; Yang, X.; Zheng, F.; Liu, C.; Wang, J.; Li, Q. Enhanced acidic resistance ability and catalytic properties of *Bacillus* 1,3-1,4- β -glucanases by sequence alignment and surface charge engineering. *Int. J. Biol. Macromol.* **2021**, *192*, 426–434. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Singh, N.; Mathur, A.S.; Gupta, R.P.; Barrow, C.J.; Tuli, D.K.; Puri, M. Enzyme systems of thermophilic anaerobic bacteria for lignocellulosic biomass conversion. *Int. J. Biol. Macromol.* **2021**, *168*, 572–590. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Bhatia, R.K.; Ullah, S.; Hoque, M.Z.; Ahmad, I.; Yang, Y.-H.; Bhatt, A.K.; Bhatia, S.K. Psychrophiles: A source of cold-adapted enzymes for energy efficient biotechnological industrial processes. *J. Environ. Chem. Eng.* **2021**, *9*, 104607. [\[CrossRef\]](#)
73. Tiwari, M.K.; Singh, R.; Singh, R.K.; Kim, I.-W.; Lee, J.-K. Computational approaches for rational design of proteins with novel functionalities. *Comput. Struct. Biotechnol. J.* **2012**, *2*, e201204002. [\[CrossRef\]](#)
74. Wu, J.; Bao, M.; Duan, X.; Zhou, P.; Chen, C.; Gao, J.; Cheng, S.; Zhuang, Q.; Zhao, Z. Developing a pathway-independent and full-autonomous global resource allocation strategy to dynamically switching phenotypic states. *Nat. Commun.* **2020**, *11*, 5521. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Li, F.; Yuan, L.; Lu, H.; Li, G.; Chen, Y.; Engqvist, M.K.M.; Kerkhoven, E.J.; Nielsen, J. Deep learning-based *kcat* prediction enables improved enzyme-constrained model reconstruction. *Nat. Catal.* **2022**, *5*, 662–672. [\[CrossRef\]](#)
76. Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R.K. Artificial intelligence in drug discovery and development. *Drug Discov. Today* **2021**, *26*, 80–93. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D.; et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Richter, F.; Leaver-Fay, A.; Khare, S.D.; Bjelic, S.; Baker, D. De novo enzyme design using Rosetta3. *PLoS ONE* **2011**, *6*, e19230. [\[CrossRef\]](#)
79. Chen, Q.; Wu, R. CNN is all you need. *arXiv* **2017**, arXiv:1712.09662. [\[CrossRef\]](#)
80. Sheng, W.; Sun, S.; Zhen, L.; Zhang, R.; Xu, J.; Avner, S. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput. Biol.* **2017**, *13*, e1005324. [\[CrossRef\]](#)
81. Ferruz, N.; Höcker, B. Controllable protein design with language models. *Nat. Mach. Intell.* **2022**, *4*, 521–532. [\[CrossRef\]](#)
82. Castro, E.; Godavarthi, A.; Rubinfeld, J.; Givechian, K.; Bhaskar, D.; Krishnaswamy, S. Transformer-based protein generation with regularized latent space optimization. *Nat. Mach. Intell.* **2022**, *4*, 840–851. [\[CrossRef\]](#)
83. Kobe, B.; Guncar, G.; Buchholz, R.; Huber, T.; Maco, B.; Cowieson, N.; Martin, J.L.; Marfori, M.; Forwood, J.K. Crystallography and protein–protein interactions: Biological interfaces and crystal contacts. *Biochem. Soc. Trans.* **2008**, *36*, 1438–1441. [\[CrossRef\]](#)
84. Tang, Y.; Schneider, W.M.; Shen, Y.; Raman, S.; Inouye, M.; Baker, D.; Roth, M.J.; Montelione, G.T. Fully automated high-quality NMR structure determination of small ²H-enriched proteins. *J. Struct. Funct. Genom.* **2010**, *11*, 223–232. [\[CrossRef\]](#)
85. Nygaard, R.; Kim, J.; Mancina, F. Cryo-electron microscopy analysis of small membrane proteins. *Curr. Opin. Struct. Biol.* **2020**, *64*, 26–33. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [\[CrossRef\]](#) [\[PubMed\]](#)
87. Yang, J.; Zhang, Y. Protein structure and function prediction using I-TASSER. *Curr. Protoc. Bioinform.* **2015**, *52*, 5–8. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Webb, B.; Sali, A. Protein structure modeling with modeller. In *Protein Structure Prediction*; Kihara, D., Ed.; Springer: New York, NY, USA, 2014; pp. 1–15. [\[CrossRef\]](#)
89. Pei, J. Multiple protein sequence alignment. *Curr. Opin. Struct. Biol.* **2008**, *18*, 382–386. [\[CrossRef\]](#)

90. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. [\[CrossRef\]](#)
91. Zhang, C.; Zheng, W.; Mortuza, S.M.; Li, Y.; Zhang, Y. DeepMSA: Constructing deep multiple sequence alignment to improve contact prediction and fold-recognition for distant-homology proteins. *Bioinformatics* **2019**, *36*, 2105–2112. [\[CrossRef\]](#)
92. Du, T.; Liao, L.; Wu, C.H.; Sun, B. Prediction of residue-residue contact matrix for protein-protein interaction with Fisher score features and deep learning. *Methods* **2016**, *110*, 97–105. [\[CrossRef\]](#)
93. Yin, R.; Feng, B.Y.; Varshney, A.; Pierce, B.G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci.* **2022**, *31*, e4379. [\[CrossRef\]](#)
94. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv. [\[CrossRef\]](#)
95. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473. [\[CrossRef\]](#)
96. Leaver-Fay, A.; Snoeyink, J.; Kuhlman, B. On-the-Fly rotamer pair energy evaluation in protein design. In *Bioinformatics Research and Applications*; Springer: Berlin/Heidelberg, 2008; pp. 343–354.
97. Haspel, N.; Tsai, C.-J.; Wolfson, H.; Nussinov, R. Reducing the computational complexity of protein folding via fragment folding and assembly. *Protein Sci.* **2003**, *12*, 1177–1187. [\[CrossRef\]](#)
98. Onuchic, J.N.; Luthey-Schulten, Z.; Wolynes, P.G. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600. [\[CrossRef\]](#)
99. Zhang, C.T.; Chou, K.C. Monte Carlo simulation studies on the prediction of protein folding types from amino acid composition. *Biophys. J.* **1992**, *63*, 1523–1529. [\[CrossRef\]](#)
100. Osguthorpe, D.J. Ab initio protein folding. *Curr. Opin. Struct. Biol.* **2000**, *10*, 146–152. [\[CrossRef\]](#) [\[PubMed\]](#)
101. Hansmann, U.H.E.; Okamoto, Y. New Monte Carlo algorithms for protein folding. *Curr. Opin. Struct. Biol.* **1999**, *9*, 177–183. [\[CrossRef\]](#) [\[PubMed\]](#)
102. Ciemny, M.P.; Badaczewska-Dawid, A.E.; Pikuzinska, M.; Kolinski, A.; Kmiecik, S. Modeling of disordered protein structures using Monte Carlo simulations and knowledge-based statistical force fields. *Int. J. Mol. Sci.* **2019**, *20*, 606. [\[CrossRef\]](#) [\[PubMed\]](#)
103. Yang, J.; Anishchenko, I.; Park, H.; Peng, Z.; Ovchinnikov, S.; Baker, D. Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1496–1503. [\[CrossRef\]](#)
104. Bo, C.; Dell’Acqua, F.; Deng, S.; Hsu, D.; Chaintreau, A. Biased programmers? Or biased data? A field experiment in operationalizing AI ethics. In Proceedings of the 21st ACM Conference on Economics and Computation, Budapest, Hungary, 13–17 July 2020.
105. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. Evolutionary-scale prediction of atomic level protein structure with a language model. *bioRxiv* **2022**, 2022, 7. [\[CrossRef\]](#)
106. Weissenow, K.; Heinzinger, M.; Rost, B. Protein language-model embeddings for fast, accurate, and alignment-free protein structure prediction. *Structure* **2022**, *30*, 1169–1177.e1164. [\[CrossRef\]](#)
107. Zheng, W.; Zhang, C.; Li, Y.; Pearce, R.; Bell, E.W.; Zhang, Y. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell Rep. Methods* **2021**, *1*, 100014. [\[CrossRef\]](#)
108. Hong, Y.; Lee, J.; Ko, J. A-Prot: Protein structure modeling using MSA transformer. *BMC Bioinform.* **2022**, *23*, 93. [\[CrossRef\]](#)
109. Bian, Z.; Liu, H.; Wang, B.; Huang, H.; Li, Y.; Wang, C.; Cui, F.; You, Y. Colossal-AI: A unified deep learning system for large-scale parallel training. *arXiv* **2021**, arXiv:2110.14883. [\[CrossRef\]](#)
110. Khoury, G.A.; Smadbeck, J.; Kieslich, C.A.; Floudas, C.A. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol.* **2014**, *32*, 99–109. [\[CrossRef\]](#) [\[PubMed\]](#)
111. Anand, N.; Eguchi, R.; Mathews, I.I.; Perez, C.P.; Derry, A.; Altman, R.B.; Huang, P.-S. Protein sequence design with a learned potential. *Nat. Commun.* **2022**, *13*, 746. [\[CrossRef\]](#) [\[PubMed\]](#)
112. Norn, C.; Wicky, B.I.M.; Juergens, D.; Liu, S.; Kim, D.; Tischer, D.; Koepnick, B.; Anishchenko, I.; Baker, D.; Ovchinnikov, S.; et al. Protein sequence design by conformational landscape optimization. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2017228118. [\[CrossRef\]](#) [\[PubMed\]](#)
113. Huang, P.-S.; Ban, Y.-E.A.; Richter, F.; Andre, I.; Vernon, R.; Schief, W.R.; Baker, D. RosettaRemodel: A generalized framework for flexible backbone protein design. *PLoS ONE* **2011**, *6*, e24109. [\[CrossRef\]](#) [\[PubMed\]](#)
114. Bonet, J.; Wehrle, S.; Schriever, K.; Yang, C.; Billet, A.; Sesterhenn, F.; Scheck, A.; Sverrisson, F.; Veselkova, B.; Vollers, S.; et al. Rosetta FunFolDes—A general framework for the computational design of functional proteins. *PLoS Comput. Biol.* **2018**, *14*, e1006623. [\[CrossRef\]](#)
115. Pan, X.; Thompson, M.C.; Zhang, Y.; Liu, L.; Fraser, J.S.; Kelly, M.J.S.; Kortemme, T. Expanding the space of protein geometries by computational design of de novo fold families. *Science* **2020**, *369*, 1132–1136. [\[CrossRef\]](#)
116. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D: Nonlinear Phenom.* **2020**, *404*, 132306. [\[CrossRef\]](#)
117. Meng, Y.; Zong, S.; Li, X.; Sun, X.; Zhang, T.; Wu, F.; Li, J. GNN-LM: Language modeling based on global contexts via GNN. *arXiv* **2021**, arXiv:2110.08743.
118. Wang, Y. A mathematical introduction to generative adversarial nets (GAN). *arXiv* **2020**, arXiv:2009.00169. [\[CrossRef\]](#)

119. Huang, P.-S.; Boyken, S.E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016**, *537*, 320–327. [[CrossRef](#)] [[PubMed](#)]
120. Qi, Y.; Zhang, J.Z.H. DenseCPD: Improving the accuracy of neural-network-based computational protein sequence design with DenseNet. *J. Chem. Inf. Model.* **2020**, *60*, 1245–1252. [[CrossRef](#)] [[PubMed](#)]
121. Ferruz, N.; Schmidt, S.; Höcker, B. ProtGPT2 is a deep unsupervised language model for protein design. *Nat. Commun.* **2022**, *13*, 4348. [[CrossRef](#)]
122. Anand, N.; Achim, T. Protein structure and sequence generation with equivariant denoising diffusion probabilistic models. *arXiv* **2022**, arXiv:2205.15019.
123. Hellinga, H.W.; Richards, F.M. Construction of new ligand binding sites in proteins of known structure: I. Computer-aided modeling of sites with pre-defined geometry. *J. Mol. Biol.* **1991**, *222*, 763–785. [[CrossRef](#)]
124. Bolon, D.N.; Mayo, S.L. Enzyme-like proteins by computational design. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14274–14279. [[CrossRef](#)]
125. Zanghellini, A.; Jiang, L.; Wollacott, A.M.; Cheng, G.; Meiler, J.; Althoff, E.A.; Röthlisberger, D.; Baker, D. New algorithms and an *in silico* benchmark for computational enzyme design. *Protein Sci.* **2006**, *15*, 2785–2794. [[CrossRef](#)]
126. Bi, J.; Jing, X.; Wu, L.; Zhou, X.; Gu, J.; Nie, Y.; Xu, Y. Computational design of noncanonical amino acid-based thioether staples at N/C-terminal domains of multi-modular pullulanase for thermostabilization in enzyme catalysis. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 577–585. [[CrossRef](#)]
127. Richter, F.; Blomberg, R.; Khare, S.D.; Kiss, G.; Kuzin, A.P.; Smith, A.J.T.; Gallaher, J.; Pianowski, Z.; Helgeson, R.C.; Grjasnow, A.; et al. Computational Design of Catalytic Dyads and Oxyanion Holes for Ester Hydrolysis. *J. Am. Chem. Soc.* **2012**, *134*, 16197–16206. [[CrossRef](#)]
128. Bello, I.; Fedus, W.; Du, X.; Cubuk, E.D.; Srinivas, A.; Lin, T.Y.; Shlens, J.; Zoph, B. Revisiting ResNets: Improved training and scaling strategies. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22614–22627. [[CrossRef](#)]
129. Bennett, N.; Coventry, B.; Goreschnik, I.; Huang, B.; Allen, A.; Vafeados, D.; Peng, Y.P.; Dauparas, J.; Baek, M.; Stewart, L.; et al. Improving de novo protein binder design with deep learning. *bioRxiv* **2022**, *15*, 495993. [[CrossRef](#)]
130. Iandola, F.; Moskewicz, M.; Karayev, S.; Girshick, R.; Keutzer, K. DenseNet: Implementing efficient ConvNet descriptor pyramids. *arXiv* **2014**, arXiv:1404.1869. [[CrossRef](#)]
131. Liu, Y.; Zhang, L.; Wang, W.; Zhu, M.; Wang, C.; Li, F.; Zhang, J.; Li, H.; Chen, Q.; Liu, H. Rotamer-free protein sequence design based on deep learning and self-consistency. *Nat. Comput. Sci.* **2022**, *2*, 451–462. [[CrossRef](#)]
132. Sutskever, I.; Jozefowicz, R.; Gregor, K.; Rezende, D.; Vinyals, O. Towards principled unsupervised learning. *arXiv* **2015**, arXiv:1511.06440. [[CrossRef](#)]
133. Suzek, B.E.; Huang, H.; McGarvey, P.; Mazumder, R.; Wu, C.H. UniRef: Comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **2007**, *23*, 1282–1288. [[CrossRef](#)]
134. Dhariwal, P.; Nichol, A. Diffusion models beat GANs on image synthesis. *arXiv* **2021**. [[CrossRef](#)]
135. Silva, D.-A.; Yu, S.; Ulge, U.Y.; Spangler, J.B.; Jude, K.M.; Labão-Almeida, C.; Ali, L.R.; Quijano-Rubio, A.; Ruterbusch, M.; Leung, I.; et al. De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **2019**, *565*, 186–191. [[CrossRef](#)]
136. Cao, L.; Goreschnik, I.; Coventry, B.; Case, J.B.; Miller, L.; Kozodoy, L.; Chen, R.E.; Carter, L.; Walls, A.C.; Park, Y.-J.; et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* **2020**, *370*, 426–431. [[CrossRef](#)]
137. Wang, B.; Su, Z.; Wu, Y. Computational assessment of protein–protein binding affinity by reversely engineering the energetics in protein complexes. *Genom. Proteom. Bioinform.* **2021**, *19*, 1012–1022. [[CrossRef](#)]
138. Dou, J.; Vorobieva, A.A.; Sheffler, W.; Doyle, L.A.; Park, H.; Bick, M.J.; Mao, B.; Foight, G.W.; Lee, M.Y.; Gagnon, L.A.; et al. De novo design of a fluorescence-activating β -barrel. *Nature* **2018**, *561*, 485–491. [[CrossRef](#)] [[PubMed](#)]
139. Krivacic, C.; Kundert, K.; Pan, X.; Pache, R.A.; Liu, L.; Conchúir, S.O.; Jeliazkov, J.R.; Gray, J.J.; Thompson, M.C.; Fraser, J.S.; et al. Accurate positioning of functional residues with robotics-inspired computational protein design. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2115480119. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.