TOPICAL REVIEW — 8th IUPAP International Conference on Biological Physics

Computational design of proteins with novel structure and functions*

Wei Yang(杨为)[1,2,3] and Lu-Hua Lai(来鲁华)[1,2,†]

[1] *BNLMS, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, and Peking–Tsinghua Center for Life Sciences at College of Chemistry and Molecular Engineering, Peking University, Beijing* 100871, *China*

[2] *Center for Quantitative Biology, Peking University, Beijing* 100871, *China*

[3] *School of Life Sciences, Tsinghua University, Beijing* 100084, *China*

(Received 15 May 2015; published online 8 December 2015)

Computational design of proteins is a relatively new field, where scientists search the enormous sequence space for sequences that can fold into desired structure and perform desired functions. With the computational approach, proteins can be designed, for example, as regulators of biological processes, novel enzymes, or as biotherapeutics. These approaches not only provide valuable information for understanding of sequence–structure–function relations in proteins, but also hold promise for applications to protein engineering and biomedical research. In this review, we briefly introduce the rationale for computational protein design, then summarize the recent progress in this field, including *de novo* protein design, enzyme design, and design of protein–protein interactions. Challenges and future prospects of this field are also discussed.

Keywords: computational protein design, *de novo* protein design, enzyme design, protein–protein interaction

PACS: 87.14.E–, 87.15.km, 87.14.ej **DOI:** 10.1088/1674-1056/25/1/018702

1. Introduction

Proteins perform a large variety of functions within organisms either alone by themselves or in cooperation with their binding partners. Studies on functions of these biomolecules and the relationship between amino acid sequence, three-dimensional structure, and functions have been prominent biological topics in the last several decades. Recently, with the rapid progress in the fields of biophysics and structural biology,[1] enthusiasm for protein design has motivated researchers to create proteins with novel structure or functions; the success of such work is also a critical test of our understanding of protein folding. These developments are important for both theoretical studies and practical applications. Proteins with specific functions are urgently needed in many fields, such as health care, agriculture, manufacture, and environmental protection. Computational design of proteins is a route to novel proteins that may be used as protein therapeutics, enzymes with special activities, or *de novo* proteins in synthetic organisms.[2,3]

In this review, we briefly discuss the theory of computational protein design and summarize recent progress in this field. The rationale for computational protein design will be introduced first, including the basic methods and strategies. Then, we will describe the research progress on *de novo* protein design. In the next section, design of functional proteins will be discussed, including enzyme design and design of protein–protein interactions. The last section provides conclusions and perspectives.

2. Key elements of computational protein design

The basic assumption in computational protein design is that the structure of a well-folded protein is always in a state of lowest free energy.[4,5] Therefore, identifying the amino acid sequence with lowest energy for a given target structure with particular functions is the most important aim of protein design. Early attempts to build three-dimensional conformations of proteins for a fixed protein backbone structure, with side chain rotamers in a precomposed library, were made by Ponder and Richard.[6] The delicate procedure of using discrete rotamers has effectively reduced the number of conformations to be evaluated for a certain sequence; this development enlarged the search space of sequences greatly. Currently, the basic approach to protein design involves packing of side chains for multiple candidate sequences and evaluating the free energy of each sequence with required structure. Hence, in the past three decades, in order to improve the capabilities and efficiency of computational protein design, most efforts have been devoted to the following two goals: development of sampling methods to explore the sequence space and optimization of free-energy functions to identify the minimum (Fig. 1).[7]

As is the case with the evolution of proteins in nature, selection pressures are needed in the process of computational protein design. The function of free energy is the selection

*Project supported by the National Basic Research Program of China (Grant No. 2015CB910300), the National High Technology Research and Development Program of China (Grant No. 2012AA020308), and the National Natural Science Foundation of China (Grant No. 11021463).

†Corresponding author. E-mail: lhlai@pku.edu.cn

© 2016 Chinese Physical Society and IOP Publishing Ltd

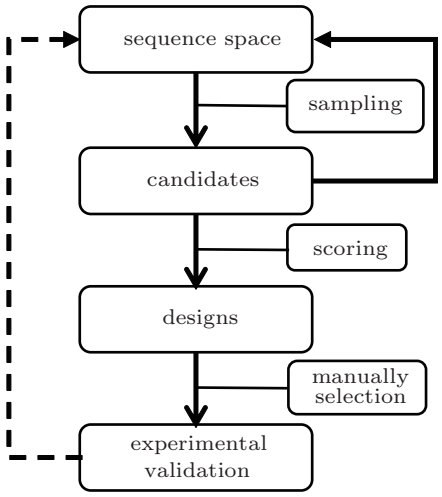


Fig. 1. Flow chart of key steps in computational protein design.

pressure that determines the direction of evolution. Therefore, the desired results can be obtained only by using the function of free energy that can correctly describe the nature of interactions within or between proteins.[8,9] Currently, the function of free energy usually consists of the following three parts: (i) terms that describe packing of atoms, including van der Waals forces and geometric complementarity, (ii) terms that describe the polar interactions, such as hydrogen bonding and electrostatic forces, and (iii) terms that describe hydrophobic effects and desolvation, which mainly affect the buried surface area.[10,11]

As mentioned above, sampling methods or a sampling strategy are another essential element of computational protein design. How long it takes to obtain a target sequence depends on the sampling strategy. Generally, a sampling strategy consists of two parts: degrees of freedom of conformations and the algorithm of sampling. The degrees of freedom of conformations include the translation and rotations of the whole protein as well as conformations of the backbone and side chain of each residue.[12] The degrees of freedom of backbones can be described by dihedral backbone angles, φ and ψ, which should satisfy the Ramachandran plot.[13] If we take the flexibility of a backbone into consideration, the side chain-coupled local motion of the backbone, known as backrub,[14] is often present. Side chain degrees of freedom are represented by rotamers.[15] The degrees of freedom also vary depending on the purposes of design. Once the parameters of the set of degrees of freedom are fixed, the conformation of the protein is predetermined. Sampling methods currently employed in computational protein design can usually be subdivided into two classes: stochastic algorithms and deterministic algorithms. By means of these algorithms, candidate sequences with better properties can be identified. Stochastic methods include the Monte Carlo (MC) methods[16] and genetic algorithms.[17] Exhaustive computations are not required when a problem is solved by stochastic algorithms. The deterministic methods, in contrast, can identify global minima of a system, for example, dead-end elimination[18] and mean-field theory.[19]

There are two major challenges for computational protein design. First, as we mentioned above, the sequence space is enormous (20^N for an N-residue protein)[20] and makes it impossible to evaluate all sequences. Thus, the features of the sequence space cannot be thoroughly analyzed by the sampling strategies that we introduced above. Rather than trying to find all possible solutions in the sequence space, researchers are now paying more attention to the question whether proteins with a particular function or structure can be successfully designed. Nevertheless, access to all possible solutions in the sequence space would surely result in more knowledge about the relation among sequence, structure, and function of a protein. With the increasing number of computational approaches and experimental techniques, more custom-designed proteins are expected to be discovered and tested empirically. On the other hand, more accurate and reasonable functions for scoring are needed,[21] because in the huge sequence space, there is a need for accurate calculations regarding the candidates. Given the large number of possible results with different sequences, the accuracy of functions for scoring is key to the efficiency of computational protein design. Commonly used energy functions are based on physical potential, statistical potential, or a hybrid of both.[22–25] No matter which strategy is employed, as long as the accuracy is improved, computational protein design will most likely be successful.

3. *De novo* protein design

In this field, early famous studies were focused on design of a novel amino acid sequence for desired structure. Many of these proteins were designed rationally, based on the empirical principles of how three-dimensional structures are determined by the amino acid sequence that researchers have deduced. For instance, DeGrado and coworkers designed three-helix and four-helix bundle proteins (Fig. 2(b)).[26,27] Later on, possible sequences can be produced with the help of software that can assemble the side chains of residues automatically. The candidate proteins predicted by the software were then tested empirically. In 1997, Mayo's group successfully designed a $\beta\beta\alpha$ protein motif with a novel sequence, which was designed automatically and validated by empirical experiments (Fig. 2(a)).[28] In 2003, Baker's group designed a 97-residue $\alpha\beta$ protein motif called Top7 (Fig. 2(c)), which had not been found in nature, using the RosettaDesign software.[29] This is the first successful attempt to design proteins with novel folds, and researchers are inspired to explore the large regions of protein structures that have not been found in nature. Our group has designed a stable $\beta\alpha\beta$ motif consisting of two parallel β-sheets connected by an α-helix (Fig. 2(d)). This $\beta\alpha\beta$ motif

has also not been found in nature as a stand-alone protein.[30] In 2012, based on the rules of how secondary-structure patterns translate into a tertiary motif, Bakers' group proposed a set of principles for designing "ideal" proteins.[31] In accordance with these principles, they successfully designed novel proteins with five different folds. These studies not only explore the infinite possibilities of protein structures, but should also help to reveal the process of protein folding.

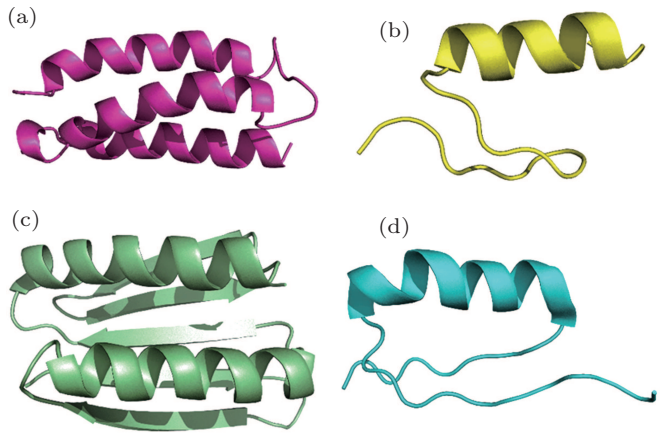


Fig. 2. *De novo* designed proteins. (a) *De novo* designed three-helix bundle protein.[27] (b) *De novo* designed $\beta\beta\alpha$ protein motif.[28] (c) Design of novel $\beta\alpha$ protein motif, TOP7.[29] (d) Design of novel $\beta\alpha\beta$ protein motif.[30]

Recently, novel proteins with more complicated tertiary structures and folds have been designed, which possess some basic functions or potential applications in relevant fields. For instance, Baker's group has designed protein complexes with complicated structures formed by self-assembling. To be more specific, they designed a 13-nm-diameter complex composed of 24 subunits and an 11-nm-diameter complex composed of 12 subunits, using trimeric proteins as building blocks. The elementary subunits were firstly selected by docking them into the target architecture. Then, the researchers designed the self-assembly interface.[32] In addition, right-handed and left-handed antiparallel bundles and a pentameric bundle were designed in 2014.[33] Barrels consisting of an α helix, which are a novel structure, have been designed by Woolfson *et al.*, on the basis of geometric properties and knowledge-based scoring of the helices.[34] Hydrophilic channels are observed in these barrels consisting of five, six, or seven helices. Similar studies were performed by DeGrado's group. They designed dimetal binding sites in a four-helix bundle. They observed a flux of Zn^{2+} and Co^{2+} through this designed structure.[35] In summary, recently, *de novo* designed proteins went beyond simple structures; complicated structures are now obtained using basic structural motifs as building blocks. More proteins with more complicated structure and versatile functions are expected to be explored by means of *de novo* design.

4. Design of functional proteins

4.1. Enzyme design

In various life forms, nearly all metabolic pathways are controlled by enzymes,[36] which can catalyze reactions rapidly and selectively. Due to the fascinating properties of enzymes, for decades, chemists have been dreaming of designing enzymes that catalyze specific chemical reactions. Nowadays, new enzymes can be created by means of bioinformatics, natural evolution,[37] supramolecular chemistry, rational design based on a protein's three-dimensional structure, computational design, and other approaches. Enzymes increase a reaction rate by lowering the activation energy. Pockets in enzymes' active sites help stabilize the transition state in reactions, thus catalyzing the reaction. Therefore, when researchers try to design or optimize enzymes, they engineer the pockets that bind substrates in the original enzyme or an unrelated protein as a template. Quantum mechanical calculations are employed to obtain conformations of these transition states. In the general computational protein design procedure, quantum mechanical calculations are performed to determine the functional groups that will bind to the substrate and stabilize the transition state. Once these catalyzing units are confirmed, proteins from the Protein Data Bank[38] will be screened. Proteins with a scaffold that can accommodate the catalyzing units will be selected as a template and be grafted onto the catalyzing unit. After that, residues surrounding the pockets will be designed to optimize the whole structure, taking stability, geometric complementarity, and other properties into consideration (Fig. 3).[39,40] Since the first computationally designed enzyme (it catalyzed Kemp elimination, which cannot be catalyzed by natural enzymes[41]), other biomolecule catalysts and ligand-binding proteins have been successfully designed, such as retro-aldolases,[42] catalyst for the Diels–Alder reaction,[43] and a protein with phenol oxidase activity.[44] Researchers are now paying more and more attention to this field because of the maturity of the related computational and experimental techniques.[39,45]

In addition to the regular strategies mentioned above, in recent studies, multiple strategies in combination with other computational approaches are utilized for specific purposes or reactions. For instance, Zhan's group has created a cocaine-detoxifying enzyme of high efficiency by computational design.[46] Human butyrylcholinesterase mutants with low catalytic efficiency were selected as a scaffold for their stability and safety.[47,48] On the basis of previous studies, in the process of computational design, researchers first identified a simplified descriptor associated with hydrogen-bonding energy.[49] Then, candidates in a virtual-mutant library, which could maximize the descriptor, were selected. Details of the reaction and free-energy change were then calculated by hybrid quantum mechanical/molecular mechanical (QM/MM) method to evaluate the design. Finally, the catalytic efficiency of the

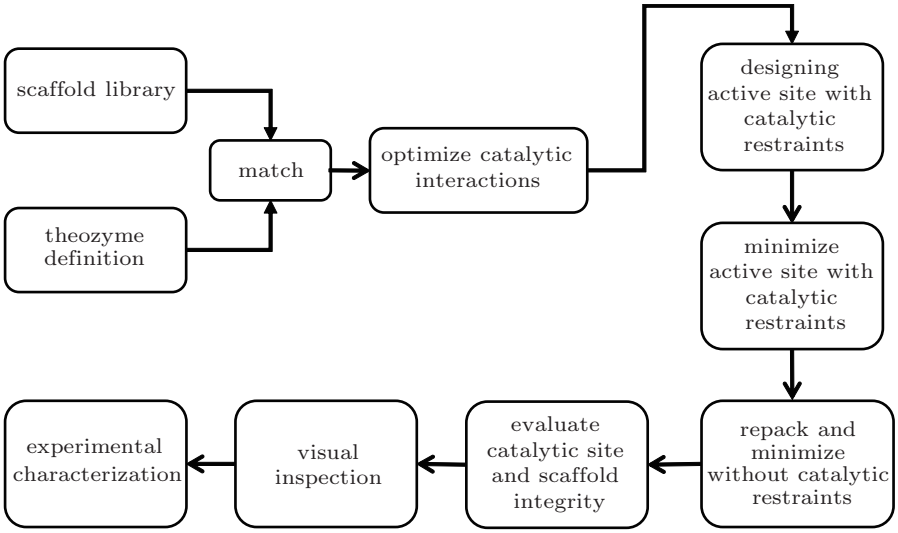


Fig. 3. A general approach for computational enzyme design with Rosetta.[40]

computationally designed E30-6 was comparable to that of acetylcholinesterase, which is known as the most efficient hydrolytic enzyme.[49]

Besides the binding affinity for a substrate, the ability to stabilize transition states, to release the product, and to initiate catalysis (e.g., nucleophilicity of the catalyzing residue) are also key elements that have yet to be improved by computational enzyme design. Rajagopalan and coworkers have designed serine-containing catalytic triads (whose catalyzing units are different from the one previously designed) that are similar to native hydrolases.[50] The structure was eventually solved at 2.3 Å and closely matched the designed model, with the root mean square deviation of the backbone at 0.6 Å. In another study,[49] Khare's group started from zinc enzyme scaffolds and successfully designed an organophosphate hydrolase. This strategy for *de novo* enzyme design from noncognate naturally occurring enzymes is a new approach to exploiting natural enzymes when designing new biocatalysts. Gordon and coworkers successfully designed an α-gliadin peptidase computationally, which can potentially be used as a therapeutic agent for celiac disease.[52] This disease is caused by the proteolytically resistant oligopeptides produced during digestion; these oligopeptides trigger an autoimmune response in patients with celiac disease. Additionally, the enzyme that Gordon's group designed not only degrades the immunogenic α-gliadin peptide but is also resistant to the action of digestive proteases; thus, the novel enzyme may be used as an oral therapeutic agent.

Aside from the common computational tools mentioned above, an online game, Foldit, was developed by University of Washington's Center for Game Science. The objective of this game is to fold a particular protein into well-folded structure. These solutions will then be used for prediction of protein structures and computational protein design. In another study,[53] Eiben's group employed solutions to remodeling of an enzyme guided by a Foldit player that catalyze the Diels–Alder reaction and successfully increase the enzymatic activity (18-fold). These results, once again, show the creativity and intuition of human beings in relation to molecular-scale design problems.

Studies on design of ligand-binding proteins are similar to those on enzyme design. In living creatures, many of the functions of proteins are related to the process of binding to specific molecules, such as amino acids, ions, lipids, or other chemical compounds. Thus, there is a large number of practical applications for custom-designed ligand-binding proteins in medical and biotechnological fields.

Nevertheless, design of proteins with high affinity and selectivity for a particular ligand is still a challenge. In 2013, Bakers' group successfully designed proteins binding to the steroid digoxigenin (DIG), using a similar strategy of computational enzyme design.[54] In the computational design project, the binding sites with idealized interactions of ligand binding were defined first. Then, the ligand and a constellation of interacting residues were placed in scaffolds. Finally, sequences of the binding site were further optimized considering shape complementarity. After directed evolution of the candidates with affinity in the micromolar (mM) range, proteins with picomolar (pM) affinity and specificity toward the derivatives of DIG were obtained. This method can be employed for development of other ligand-binding proteins as biosensors, therapeutics, and diagnostics. In a study by Mill's group, methods for inclusion of noncanonical amino acids into proteins were combined with computational protein design.[55] The method of site-specific insertion of noncanonical amino acids into proteins greatly extends the physical and chemical properties of the proteins available in nature and holds great promise for creation of proteins with novel functions. Some

approaches to computational protein design can help to determine the position and conformation of noncanonical amino acids as well as the residues nearby. Mill and coworkers successfully designed a metalloprotein containing the amino acid (2, 2′-bipyridin-5yl) alanine (Bpy-Ala); this protein has an inherent ability to bind diverse metal ions.[55] The eventually designed Bpy-Ala-based metalloprotein can bind divalent cations including Co^{2+}, Zn^{2+}, Fe^{2+}, and Ni^{2+}. In collaboration with Prof. He's group, we have designed a uranyl-binding protein using a computational screening process. The designed protein possesses extremely high affinity and selectivity for uranyl: the dissociation constant is 7.4 femtomoles per liter (fM), with >10000-fold selectivity relative to other metal ions.

4.2. Design of protein–protein interaction

In cells, a wide variety of biological processes are meditated by protein–protein interactions, including metabolic pathways, cell signaling, and cell-to-cell interactions.[56] Hence, modification of protein–protein interactions or design of novel protein–protein interactions is urgently needed for both scientific research and practical applications. For instance, proteins can be designed to bind to viruses or bacteria or to block their infection of other organisms. Proteins that bind to nodes in the biological networks of metabolism can help to return a cell or organism from an abnormal state back to health. Currently, most studies on design of protein–protein interactions are focused on modification of existing protein–protein interfaces, either to increase the affinity or to alter the specificity. A more challenging but more promising goal in this field is to design novel protein–protein interactions, that is, to design proteins targeting novel partners. Such a novel protein can be selected either from libraries of natural proteins or from *de novo* designed sequences, merely on the basis of the structure of the target protein. In the end, together with powerful experimental methods such as directed evolution of proteins, useful molecular tools can be devised by combining computational design with experiments on new functions or new properties.

4.2.1. Redesign of native protein–protein interaction

The most popular approach in the field of protein interface design is redesign of naturally occurring protein interfaces. Affinity of interactions can be increased, or specificity of interactions can be altered. Usually, these modifications of interfaces are achieved by introducing mutations at the protein interface in order to increase the buried hydrophobic surface or to enhance polar interactions.[57,58] For design of proteins that will serve as vaccines, the conformation of conserved epitopes is maintained to ensure induction of the corresponding neutralizing antibodies. Recently, Jardine *et al.*[57] performed computer-aided design of vaccines in combination with in vitro screening. They successfully obtained an immunogen that binds to multiple VRC01-class broadly neutralizing antibodies (bNAbs); this immunogen was validated by crystallography. The VRC01-class bNAbs have previously been isolated from serum of HIV-1-infected patients. On the other hand, the wild-type gp120 proteins cannot induce germline precursors of VRC01-class bNAbs, and the binding affinity is negligible. In contrast, the above-mentioned immunogen of gp120 proteins can not only bind to VRC01-class bNAbs but also activate the germline precursors and promote maturity of VRC01-class B cells. These results are indicative of bright prospects for the possible new vaccine.

4.2.2. Design of unnatural protein–protein interaction

Compared to redesign of a naturally occurring protein interface, creation of novel protein–protein interactions for specific sites of target proteins is obviously a more challenging but rewarding field. Generally, two approaches have been used for computational design of novel protein–protein interactions.[60] The first is protein–protein docking, which is similar to the strategy used during virtual screening of compound libraries[61] when researchers are trying to find new ligands. In the process of protein–protein docking, noninteractive protein scaffolds are docked to the surface of the targeting structure for evaluation of the shape complementarity.[62] Then, the sequences at the interface are further modified to optimize the interaction energy. The second approach utilizes hot spots at a native interface, which dominate the binding energy. After identification of hot spots, these residues are grafted onto unrelated scaffolds that can accommodate their original conformations.[63] Our group has successfully designed novel protein–protein interactions by means of the above strategies. Using the docking method, we screened a protein library of 677 proteins to find proteins that bind tumor necrosis factors (TNFs). Among the six selected proteins, two showed significant binding to TNF in an empirical study.[64] In another project, we grafted the three hot-spot residues at the interface of the human erythropoietin (EPO) and human erythropoietin receptor (EPOR) onto an unrelated protein: rat pleckstrin homology domain of phospholipase C-δ_1 (PLCδ_1-PH). Originally, the wild-type PLCδ_1-PH showed no binding affinity for EPOR, while the mutant showed high affinity for EPOR ($K_D = 24$ nM; Fig. 4(c)).[65] Similarly, novel proteins binding influenza hemagglutinin (HA) were designed, with the hot spot of interfaces recognized by neutralizing antibodies grafted onto two unrelated proteins as scaffolds (Fig. 4(a)).[66] In a study by Procko's group,[65] these two approaches were used simultaneously to design proteins that bind to the active site of lysozyme to inhibit the enzymatic activity. It turns out that the hot-spot grafting performs better and results in a designed protein with micromolar affinity, whereas the proteins

designed by the docking-and-design method failed experimental validation.

These two general methods, where either naturally occurring protein scaffolds or native binding partners are required, are quite useful. Nevertheless, designing novel proteins that bind to particular targets at a specified site, without the involvement of their naturally occurring binding partners, presents more opportunities. This approach is also more challenging because the folding of the novel proteins must been taken into consideration during the design of novel protein–protein interactions. Various studies on the design of proteins with novel structure and protein–protein interactions have laid a good foundation for this strategy. Recently, our group has successfully designed novel helical peptides targeting TNF and devised a general feasible approach to *de novo* design of helical peptides binding specific proteins (Fig. 4(d)).[68] Most recently, Correia *et al.* implemented a new computational design strategy and successfully grafted an epitope of respiratory syncytial virus onto a novel scaffold.[69] The designed protein can induce potent neutralizing antibodies by mimicking the viral epitope (Fig. 4(b)). In their computational method, first, template topology was selected according to the conformation of the epitope. After that, diverse backbone conformations were folded into the desired topology, and sequences with low energy were selected for further optimization.[69] This study proved feasibility of the epitope-and-scaffold-based vaccine design.

5. Conclusions and perspective

Many successful studies have been conducted in the field of *de novo* protein design, enzyme design, and protein–protein interaction design, but strategies and tools for computational protein design are far from perfect. In studies involving protein design, sequences with high activity can hardly be obtained after only a single round of design. A survey of relevant publications shows that only proteins with low activity can be obtained in the first round of design in most cases.[32,42,69] Thus, the initial sequences from computational design provide only a starting point. In order to obtain proteins with a desired function, empirical studies are needed to optimize the design. How natural proteins evolve into the ones with high affinity or catalytic activity has yet to be explored. Moreover, binding or catalysis that involves dramatic conformational changes still cannot be taken into consideration during computational protein design. Nor can the binding free energy be accurately calculated. To solve these problems, the existing theories and computational approaches need to be improved. With the development of new strategies and methods, together with the maturity of experimental techniques of protein engineering, computational protein design is attracting more and more interest and yielding promising results for practical applications in biomedical research and for development of novel therapeutics.

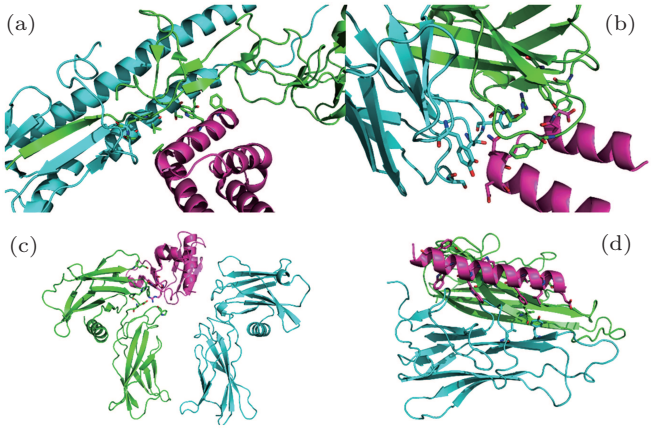


Fig. 4. Computationally designed protein–protein interactions. (a) Design of proteins bind a conserved surface of influenza hemagglutinin (HA).[64] (b) Design of small protein mimicking viral epitope with the capability of inducing potent neutralizing antibodies.[67] (c) Design of non-natural protein–protein interaction by hot-spots grafting.[63] (d) Design of helix peptide targeting TNF with novel sequence and interaction pattern.[66]

Acknowledgement

The authors thank all the past and present members of the Lai lab for their contributions to the protein design related projects.

References

[1] Terwilliger T C, Stuart D and Yokoyama S 2009 *Annu. Rev. Biophys.* **38** 371
[2] Hwang I and Park S 2008 *Drug. Discov. Today Techonol.* **5** e43
[3] Tiwari M K, Singh R, Singh R K, Kim I W and Lee J K 2012 *Comput. Struct. Biotechnol. J.* **2** e201209002
[4] Kuhlman B and Baker D 2000 *Proc. Natl. Acad. Sci. USA* **97** 10383
[5] Samish I, MacDermaid C M, Perez-Aguilar J M and Saven J G 2011 *Annu. Rev. Phys. Chem.* **62** 129
[6] Ponder J W and Richards F M 1987 *J. Mol. Biol.* **193** 775
[7] Pokala N and Handel T M 2001 *J. Struct. Biol.* **134** 269
[8] Mandell D J and Kortemme T 2009 *Nat. Chem. Biol.* **5** 797
[9] Pokala N and Handel T M 2005 *J. Mol. Biol.* **347** 203
[10] Weiner S J, Kollman P A, Case D A, Singh U C, Ghio C, Alagona G, Profeta S and Weiner P 1984 *J. Am. Chem. Soc.* **106** 765
[11] Gordon D B, Marshall S A and Mayo S L 1999 *Curr. Opin. Struct. Biol.* **9** 509
[12] Marshall S A and Mayo S L 2001 *J. Mol. Biol.* **305** 619
[13] Ramachandran G N, Ramakrishnan C and Sasisekharan V 1963 *J. Mol. Biol.* **7** 95
[14] Georgiev I, Keedy D, Richardson J S, Richardson D C and Donald B R 2008 *Bioinformatics* **24** I196
[15] Dunbrack R L 2002 *Curr. Opin. Struct. Biol.* **12** 431
[16] Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H and Teller E 1953 *J. Chem. Phys.* **21** 1087
[17] Desjarlais J R and Handel T M 1999 *J. Mol. Biol.* **290** 305
[18] Desmet J, Demaeyer M, Hazes B and Lasters I 1992 *Nature* **356** 539
[19] Kono H and Doi J 1996 *J. Comput. Chem.* **17** 1667
[20] Zwanzig R, Szabo A and Bagchi B 1992 *Proc. Natl. Acad. Sci. USA* **89** 20
[21] Das R and Baker D 2008 *Annu. Rev. Biochem.* **77** 363
[22] Rohl C A, Strauss C E M, Misura K M S and Baker D 2004 *Method Enzymol* **383** 66
[23] Liang S D, Zhang C, Liu S and Zhou Y Q 2006 *Nucleic Acids Res.* **34** 3698

[24] Xiong P, Wang M, Zhou X Q, Zhang T C, Zhang J H, Chen Q and Liu H Y 2014 *Nat. Commun.* **5** 5330

[25] Li Z X, Yang Y D, Zhan J, Dai L and Zhou Y Q 2013 *Annu. Rev. Biophys.* **42** 315

[26] DeGrado W F, Regan L and Ho S P 1987 *Cold Spring Harb Symp. Quant. Biol.* **52** 521

[27] Walsh S T R, Cheng H, Bryson J W, Roder H and DeGrado W F 1999 *Proc. Natl. Acad. Sci. USA* **96** 5486

[28] Dahiyat B I and Mayo S L 1997 *Science* **278** 82

[29] Kuhlman B, Dantas G, Ireton G C, Varani G, Stoddard B L and Baker D 2003 *Science* **302** 1364

[30] Liang H, Chen H, Fan K, Wei P, Guo X, Jin C, Zeng C, Tang C and Lai L 2009 *Angew. Chem., Int. Ed. Engl.* **48** 3301

[31] Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton T B, Montelione G T and Baker D 2012 *Nature* **491** 222

[32] King N P, Bale J B, Sheffler W, McNamara D E, Gonen S, Gonen T, Yeates T O and Baker D 2014 *Nature* **510** 103

[33] Huang P S, Oberdorfer G, Xu C, Pei X Y, Nannenga B L, Rogers J M, DiMaio F, Gonen T, Luisi B and Baker D 2014 *Science* **346** 481

[34] Thomson A R, Wood C W, Burton A J, Bartlett G J, Sessions R B, Brady R L and Woolfson D N 2014 *Science* **346** 485

[35] Joh N H, Wang T, Bhate M P, Acharya R, Wu Y, Grabe M, Hong M, Grigoryan G and DeGrado W F 2014 *Science* **346** 1520

[36] Lehninger A L, Nelson D L and Cox M M 2005 *Lehninger Principles of Biochemistry*, 4th edn. (New York: W. H. Freeman) p. 190

[37] Lutz S 2010 *Curr. Opin. Biotechnol.* **21** 734

[38] Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E 2000 *Nucleic Acids Res.* **28** 235

[39] Kiss G, Celebi-Olcum N, Moretti R, Baker D and Houk K N 2013 *Angew. Chem., Int. Ed. Engl.* **52** 5700

[40] Richter F, Leaver-Fay A, Khare S D, Bjelic S and Baker D 2011 *PLoS One* **6** e19230

[41] Rothlisberger D, Khersonsky O, Wollacott A M, Jiang L, DeChancie J, Betker J, Gallaher J L, Althoff E A, Zanghellini A, Dym O, Albeck S, Houk K N, Tawfik D S and Baker D 2008 *Nature* **453** 190

[42] Jiang L, Althoff E A, Clemente F R, Doyle L, Rothlisberger D, Zanghellini A, Gallaher J L, Betker J L, Tanaka F, Barbas C F, Hilvert D, Houk K N, Stoddard B L and Baker D 2008 *Science* **319** 1387

[43] Siegel J B, Zanghellini A, Lovick H M, Kiss G, Lambert A R, St Clair J L, Gallaher J L, Hilvert D, Gelb M H, Stoddard B L, Houk K N, Michael F E and Baker D 2010 *Science* **329** 309

[44] Faiella M, Andreozzi C, de Rosales R T, Pavone V, Maglio O, Nastri F, DeGrado W F and Lombardi A 2009 *Nat. Chem. Biol.* **5** 882

[45] Nanda V and Koder R L 2010 *Nature Chem.* **2** 15

[46] Zheng F, Xue L, Hou S, Liu J, Zhan M, Yang W and Zhan C G 2014 *Nat. Commun.* **5** 3457

[47] Zheng F, Yang W, Xue L, Hou S, Liu J and Zhan C G 2010 *Biochemistry* **49** 9113

[48] Xue L, Ko M C, Tong M, Yang W, Hou S, Fang L, Liu J, Zheng F, Woods J H, Tai H H and Zhan C G 2011 *Mol. Pharmacol.* **79** 290

[49] Quinn D M 1987 *Chem. Rev.* **87** 955

[50] Rajagopalan S, Wang C, Yu K, Kuzin A P, Richter F, Lew S, Miklos A E, Matthews M L, Seetharaman J, Su M, Hunt J F, Cravatt B F and Baker D 2014 *Nat. Chem. Biol.* **10** 386

[51] Mills J H, Khare S D, Bolduc J M, Forouhar F, Mulligan V K, Lew S, Seetharaman J, Tong L, Stoddard B L and Baker D 2013 *J. Am. Chem. Soc.* **135** 13393

[52] Gordon S R, Stanley E J, Wolf S, Toland A, Wu S J, Hadidi D, Mills J H, Baker D, Pultz I S and Siegel J B 2012 *J. Am. Chem. Soc.* **134** 20513

[53] Eiben C B, Siegel J B, Bale J B, Cooper S, Khatib F, Shen B W, Players F, Stoddard B L, Popovic Z and Baker D 2012 *Nat. Biotechnol.* **30** 190

[54] Tinberg C E, Khare S D, Dou J Y, Doyle L, Nelson J W, Schena A, Jankowski W, Kalodimos C G, Johnsson K, Stoddard B L and Baker D 2013 *Nature* **501** 212

[55] Mills J H, Khare S D, Bolduc J M, Forouhar F, Mulligan V K, Lew S, Seetharaman J, Tong L, Stoddard B L and Baker D 2013 *J. Am. Chem. Soc.* **135** 13393

[56] Jones S and Thornton J M 1996 *Proc. Natl. Acad. Sci. USA* **93** 13

[57] Stranges P B and Kuhlman B 2013 *Protein Sci.* **22** 74

[58] Kortemme T and Baker D 2004 *Curr. Opin. Chem. Biol.* **8** 91

[59] Jardine J, Julien J P, Menis S, Ota T, Kalyuzhniy O, McGuire A, Sok D, Huang P S, MacPherson S, Jones M, Nieusma T, Mathison J, Baker D, Ward A B, Burton D R, Stamatatos L, Nemazee D, Wilson I A and Schief W R 2013 *Science* **340** 711

[60] Zhang C S and Lai L H 2011 *Biochem. Soc. Trans.* **39** 1382

[61] Shoichet B K 2004 *Nature* **432** 862

[62] Zhang C S and Lai L H 2011 *J. Comput. Chem.* **32** 2598

[63] Zhang C S and Lai L H 2012 *Proteins: Struct., Funct., Bioinf.* **80** 1078

[64] Zhang C S, Tang B, Wang Q and Lai L H 2014 *Proteins: Struct., Funct., Bioinf.* **82** 2472

[65] Liu S, Liu S, Zhu X, Liang H, Cao A, Chang Z and Lai L 2007 *Proc. Natl. Acad. Sci. USA* **104** 5330

[66] Fleishman S J, Whitehead T A, Ekiert D C, Dreyfus C, Corn J E, Strauch E M, Wilson I A and Baker D 2011 *Science* **332** 816

[67] Procko E, Hedman R, Hamilton K, Seetharaman J, Fleishman S J, Su M, Aramini J, Kornhaber G, Hunt J F, Tong L, Montelione G T and Baker D 2013 *J. Mol. Biol.* **425** 3563

[68] Zhang C, Shen Q, Tang B and Lai L 2013 *Angew. Chem. Int. Ed. Engl.* **52** 11059

[69] Correia B E, Bates J T, Loomis R J, *et al.* 2014 *Nature* **507** 201