Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Computational design of ligand-binding proteins

Wei Yang[1,2,3] and Luhua Lai[1,2]

Custom-designed ligand-binding proteins with novel functions hold the potential for numerous applications. In recent years, the developments of computational methods together with high-throughput experimental screening techniques have led to the generation of novel, high-affinity ligand-binding proteins for given ligands. In addition, naturally occurring ligand-binding proteins have been computationally designed to recognize new ligands while keeping their original biological functions at the same time. Furthermore, metalloproteins have been successfully designed for novel functions and applications. Though much has been learned in these successful design cases, advances in our understanding of protein dynamics and functions related to ligand binding and development of novel computational strategies are necessary to further increase the success rate of computational protein–ligand binding design.

Addresses
[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

Corresponding author: Lai, Luhua (lhlai@pku.edu.cn)

Current Opinion in Structural Biology 2017, **45**:67–73

This review comes from a themed issue on **Engineering and design**

Edited by **Julia Shifman** and **Niv Papo**

http://dx.doi.org/10.1016/j.sbi.2016.11.021

Introduction

Proteins play essential roles for all organisms and carry out a large variety of functions. Many of these functions are achieved by protein–ligand interactions, such as catalyzing a particular reaction, recognizing small molecules that modulate signal transduction pathways, or regulating transcription by binding to DNA. Thus, by custom designing ligand-binding proteins, the functions of proteins could be broadened, and these novel functions could confer the potential for application as therapeutics, diagnostics, enzymes, biosensors, or tools for synthetic biology and chemical biology research [1,2].

Since the emergence of genetic approaches to study and engineer proteins, several proteins have been engineered to recognize unnatural ligands and to execute modified functions [3]. Although much effort has been directed towards engineering ligand-binding proteins using computational tools, the computational design of ligand-binding proteins has been regarded as an unsolved problem, despite that successful examples of novel enzyme design have been reported [4,5]. In recent years, much progress has been made in this field but many difficulties have yet to be overcome. Protein–ligand binding is the first step in enzyme catalysis. One may wonder that if protein–ligand binding cannot be designed accurately, why novel enzymes can be successfully designed. In fact, binding of a substrate to an enzyme is normally characterized by moderate binding affinities, with a K_m value in the millimolar to micromolar range. To a large extent, the enzyme provides an environment for the reactants to protect them from the solvent (water). In addition to the successful cases of designing novel enzymes using elaborate computational approaches [6,7], there are other examples that construct active enzymes by simply putting catalytic residues in a suitable hydrophobic pocket, though in all the cases the efficacy of the resulting catalysis is still much lower than that of naturally occurring enzymes [8]. To design ligand-binding proteins such as receptors, in most cases much stronger binding is required compared to that of enzyme-substrate complex. Design of ligand binding proteins faces several challenges including sampling of the enormous possible orientations of the ligand with respect to the protein, the large conformational and the sequence space of the protein pocket, and the difficulties in accurately estimating the binding free energies during the course of design. Many of the challenges regarding the design of ligand-binding proteins were discussed in the review paper published in 2013 [5,8]. Despite the challenges that have yet to be overcome, much progress has been made in the past three years. We will discuss computational methods that have been used for successful ligand-binding protein design, their pros and cons, and the potential future directions of the field. Together with the developments in protein–ligand binding theory/simulations and the knowledge-based learning/modeling, we anticipate more reports of successful ligand-binding protein design and their applications in the near future.

Designing proteins for specific ligands

Designing proteins that bind to a given ligand has long been a dream of many chemists. With the accumulated structural information on protein–ligand complexes, experimental data on thermodynamics and kinetics of binding [9], and developments in related theories [10,11], much has been learned about protein–ligand interactions.

These advances have greatly enhanced the development of computational drug design, for which protein–ligand interaction is also a critical aspect [12]. In fact, both ligand-binding protein design and drug design for protein targets involve matching the protein space and the chemical space. Theoretically, a novel protein that specifically binds to a given ligand with high specificity can be created, although it is very difficult to achieve this task in practice. Most of the current design approaches employ naturally occurring protein structures as the starting point for ligand binding design followed by sequence optimization and then experimental directed evolution. The design strategies currently used can be assigned to three main groups.

(1) The first strategy starts with creating an 'optimal' binding shell around the ligand, followed by searching for suitable protein scaffolds to accommodate the binding residues and experimental optimization (Figure 1a,c) [13••]. Similar strategies have been successfully used in the design of protein-protein interactions and enzymes [7,14,15]. Tinberg et al. used this type of approach to design proteins that bind steroid digoxigenin (DIG) [13••]. The 17 computationally designed proteins were characterized experimentally and two of them exhibited binding affinities in the micromolar range. For optimization, Tinberg and colleagues carried out three rounds of directed evolution, followed by screening, deep sequencing, and introduction of beneficial mutations. The binding affinity of the optimized protein was improved to the picomolar range. The mutations that significantly improve binding are mostly in the second shell of the binding site and affecting protein flexibility, which is rather difficult to model using the present computational design algorithms. In fact, the most successful design for DIG binding protein includes a properly preformed binding site to avoid entropy loss upon binding, which circumvented the flexibility of residues directly interacting with DIG to a certain extent. Ollikainen et al. recently developed a 'coupled moves' approach by incorporating protein backbone flexibility while sampling ligand orientations and conformations [16•].

(2) The second strategy employs computational ligand docking to search for suitable protein structures for the ligand, followed by sequence optimization of the binding pocket and experimental optimization (Figure 1b,c). This could begin with a predefined protein of interest and the placement of the ligand into the pocket using a docking program [17], followed by optimization of the sequences in the pocket using a protein design program [18–20]. Meiler and colleagues tested this idea using the ROSETTA program to generate an endo-1,4-β-xylanase to bind to dipeptides. Unfortunately, the designed xylanase mutants did not bind to the target ligands [21]. Because the binding site was located between two flexible loops and the peptide ligands were also rather malleable, the failure in obtaining binders might have resulted from an unfavorable loss of entropy upon binding. This example demonstrates again that full consideration of the protein and ligand flexibility and accurate scoring functions are vital to successful design.

Although nature often uses large conformational changes such as domain reorientation or loop reshaping for ligand binding, our current computational tools are not yet able to correctly capture these changes. The well-known case of the generation of a ribose-binding protein for binding TNT was probably the first example of an unsuccessful design of this kind [4]. Thus, it is better to avoid proteins with large conformational changes, unless careful consideration of such changes could be achieved during the design. This issue might be circumvented by using reverse docking to search for suitable proteins for the ligand, similar to the computational target identification approach for drugs or natural compounds [22]. After docking green fluorescent protein (GFP)-like chromophores to approximately 3000 *Escherichia coli* protein structures, Povarova et al. experimentally tested four top-scoring candidate proteins. Two of the proteins tested showed sub-micromolar affinity to GFP-like chromophores and fluorogenic behavior [23••].

(3) The third strategy is the *de novo* design of proteins or peptides that bind to a given ligand (Figure 2). Laio et al. proposed an approach that combined multiple computational methods to design peptides with a binding capacity, according to which the conformational space was searched by molecular dynamics and ligand docking, and sequence space was searched by the Monte Carlo method [24]. Because the length of the designed peptides was limited to less than 15 residues, the structural space for the particular sequence could be searched intensively using enhanced molecular dynamics sampling techniques. By iterating between sampling, docking, and mutation introduction, optimal sequences could be obtained. Recently, Carlo et al. designed short peptides that can bind to phenolic compounds with micromolar affinity using this strategy [25]. However, further structure determination is required in order to verify the accuracy of the predictions and more studies are needed to test the efficiency of the approach.

Designing proteins with desired functions by altering specificity or improving binding affinity

An efficient approach of designing ligand-binding proteins with the desired function is to redesign natural ligand-binding proteins with related functions [26]. On one hand, the complex structure of natural ligand-binding