COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

Volume No: 2, Issue: 3, September 2012, e201209002, http://dx.doi.org/10.5936/csbj.201209002

CSBJ

COMPUTATIONAL APPROACHES FOR RATIONAL DESIGN OF PROTEINS WITH NOVEL FUNCTIONALITIES

Manish Kumar Tiwari [a,†], Ranjitha Singh [a,†], Raushan Kumar Singh [a], In-Won Kim [a], Jung-Kul Lee [a,b,*]

Abstract: Proteins are the most multifaceted macromolecules in living systems and have various important functions, including structural, catalytic, sensory, and regulatory functions. Rational design of enzymes is a great challenge to our understanding of protein structure and physical chemistry and has numerous potential applications. Protein design algorithms have been applied to design or engineer proteins that fold, fold faster, catalyze, catalyze faster, signal, and adopt preferred conformational states. The field of *de novo* protein design, although only a few decades old, is beginning to produce exciting results. Developments in this field are already having a significant impact on biotechnology and chemical biology. The application of powerful computational methods for functional protein designing has recently succeeded at engineering target activities. Here, we review recently reported *de novo* functional proteins that were developed using various protein design approaches, including rational design, computational optimization, and selection from combinatorial libraries, highlighting recent advances and successes.

MINI REVIEW ARTICLE

Introduction

Proteins, polymers of amino acids, are the main building blocks and functional molecules of the cell. They are the most multifaceted macromolecules in living systems and have various important functions, including structural, catalytic, sensory, and regulatory functions. The ability of proteins to cluster together to form well-defined structures comprising amino acid sequences make these numerous roles possible. The collection of data regarding protein sequences is rapidly growing, with approximately 6 million entries in Universal Protein Resource (UniProt) knowledgebase at present [1-8]. To completely understand the function of a protein, knowledge of its three-dimensional structure is essential. Unfortunately, experimental structure determination is only possible for a small fraction of these proteins [2,8-10], with only approximately 2% having experimentally verified structural annotation at present. For the remaining 98%, prediction of the structure is the only alternative. Therefore, the structural characterization of proteins is a major goal in computational biology [1-8].

Advances in molecular modeling have expanded the area of computational protein design, from creating new proteins based on known protein sequences present in nature to designing new proteins that fold into a specific structure or perform a specific function. Before aiming at protein design by using computational methods, one should understand the underlying physical principles governing the folding, stability, and function of a protein. For all these decades, scientists and researchers have been following a perturbation or an alternation-based paradigm in order to determine the functionality of a protein. The method relies on the generation of hundreds and thousands of protein mutants, coupled with selective pressure to identify variants with desired properties. Alternatively, in computational protein design one aims at a design-based paradigm instead of a perturbation-based paradigm. In design-based paradigm, biologists combine design paradigms or methods for problem solving with computational modeling techniques to predict the success of their designs. This paradigm has been effective for the creation and implementation of new ideas and inventions. Design-based paradigm is used for the identification of the boundaries of possible designs and for the elimination of impossible, impractical, inefficient, or otherwise undesirable designs which would have otherwise been difficult to identify using alternation-based paradigm. In a structure-based computational method, a computational or a mathematical framework is constructed by taking into consideration the evolution, function, stability, and functionality of a protein. The designed proteins are then checked experimentally for their specific function. If the designed proteins exhibit all these characteristics, then it can be concluded that the mathematical model or the framework can fundamentally capture the essence of a protein. On the other hand, if the experiments do not work then one can learn from the failures to modify and create a new model, which will ultimately serve the final goal of computationally designing a novel viable protein. Protein design from scratch is thus the most precise way of testing our knowledge on how natural proteins implement their functions.

Engineering proteins with improved functionality or novel applications has been experimentally achieved by screening of large mutant libraries. However, most of these proteins do not provide quantitative design principles and/or comprehend the structural features that support the desired function. Computational protein design has helped overcome these drawbacks. With reliable structural predictions [11-14], protein stability at the desired conditions, and accurate description of intermolecular interactions (protein–protein interactions [15,16] and DNA–protein interactions [17]), the technique of computationally designing proteins has been one of the fast-emerging trends in biotechnology and biomedicine. Furthermore, computational protein design has attained significant breakthrough,

[a]*Department of Chemical Engineering*
[b]*Institute of SK-KU Biomaterials*
Konkuk University, 1 Hwayang-Dong, Gwangjin-Gu, Seoul 143-701, Korea
[†]*These authors contributed equally*

* Corresponding author. Tel.: +82 24503505; Fax: +82 24583504
E-mail address: jkrhee@konkuk.ac.kr (Jung-Kul Lee)

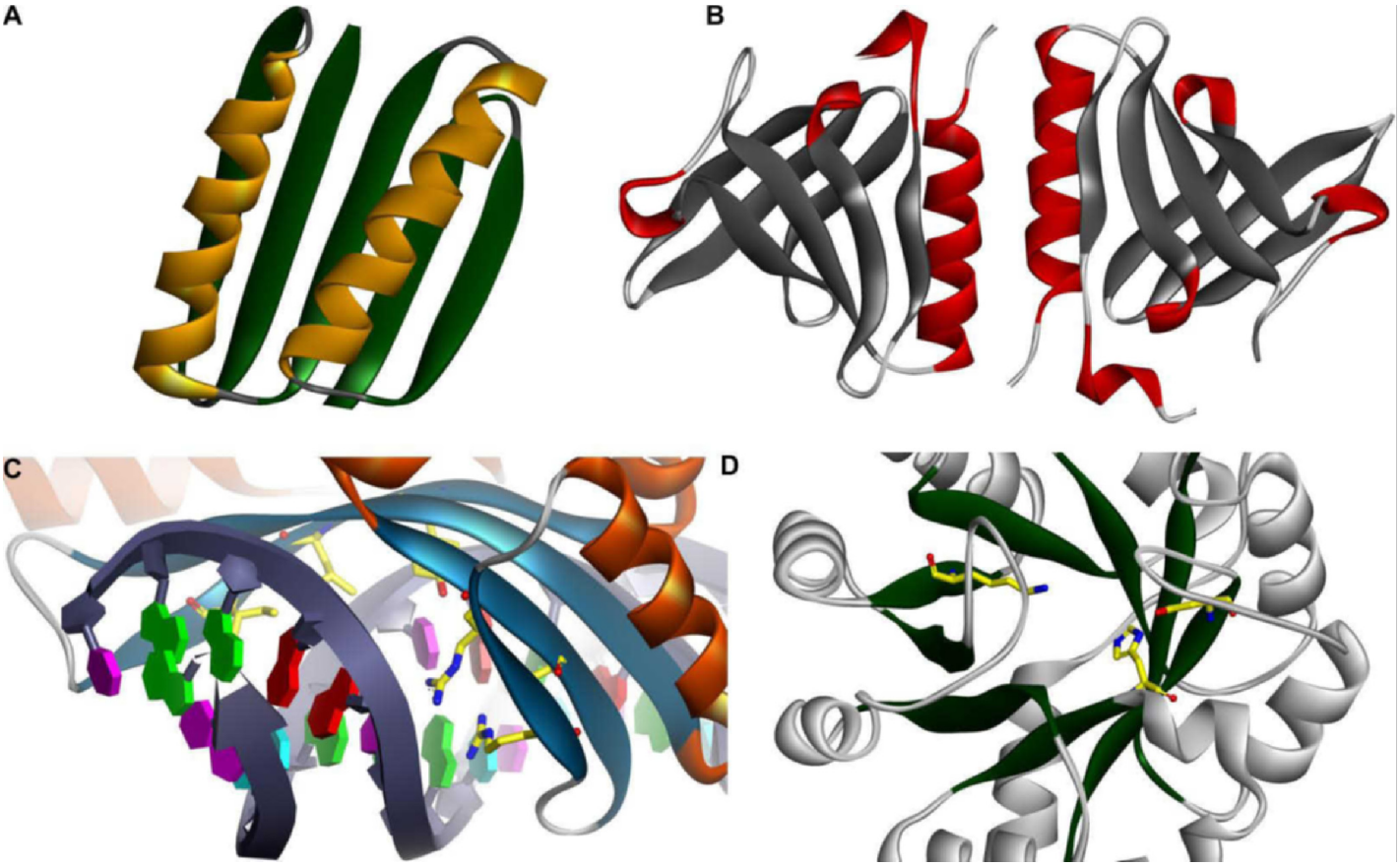

Figure 1. Computationally designed structures and enzymes. (*A*) A novel Top7 globular protein fold with atomic-level accuracy [34]. (*B*) Designed SspB adaptor protein [15]. (*C*) Redesigned endonuclease DNA binding [17]. The redesigned enzyme binds and cleaves the redesigned recognition site ~10,000 times more effectively than does the wild-type enzyme, with a level of target discrimination comparable to the original endonuclease. (*D*) A novel retro-aldol enzyme designed within a TIM-barrel scaffold [20].

for example, in the design of novel biocatalysts [18-21] and biosensors for non-natural molecules, redesign of proteins with improved binding affinity [22], redesign of proteins with greater binding specificity [16,23,24], and design of proteins capable of binding non-biological cofactors [25] (Figure 1).

Enzyme design presents a huge challenge, not only in the *de novo* design of catalysts for which no natural counterparts are known, but also in the design of multipurpose enzymes, which may have a wide range of biotechnological applications in fields, such as industrial organic synthesis and metabolic engineering [26-29]. This review mainly discusses the strengths and recent successes of computational protein design approaches. We also summarize advancements of design methodology and the application of protein design strategies over the past few years. Other recent reviews can provide additional backgrounds and perspective [30-33].

Rational computational design

The creation of biocatalysts from scratch enables scientists and engineers to build synthetic enzymes for a series of different chemical reactions, e.g., retro-aldol reaction [20] and Kemp elimination [21]. It also presents a testing ground for our fundamental understanding of the complexities of protein structure and function. Computational protein design starts with the coordinates of a protein main chain and uses a force field to identify sequences and geometries of amino acids that are optimal for stabilizing the backbone geometry [35]. Even for small proteins, the number of possible sequences far exceeds that which can be thoroughly searched. The development of powerful search algorithms to find optimal solutions has provided a major stimulus to the field [36]. Computational protein design requires correlation of structural predictions and experimental stability. Artificial enzymes have been developed with varying degrees of computational involvement, which includes *de novo* enzymes, where both the protein topology and the active site are built from scratch [20,34,37,38].

De novo active-site design

The introduction of amino acid residues in the form of active site residues into the existing scaffolds is essential for computationally designed enzyme catalysis. These active site residues of the enzymes are responsible for enhancing the chemical reactions by lowering the activation barrier via stabilization of the transition state [39]. Accurate modeling of important forces in the active site requires quantum mechanical (QM) calculations [38]. Potential binding pockets capable of binding tightly to the transition state and retaining the desired geometry of the functional groups are identified within different protein scaffolds. Using geometry-based identification, the transition state is matched with the binding site and the position of the transition state and the catalytic side chains are optimized. Finally, the remaining residues for tight binding of the transition state are