

## History of Progress and Challenges in Structural Biology

Sompop Bencharit\*

Department of Prosthodontics, School of Dentistry, and Department of Pharmacology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

### Introduction

In 1953, James Watson and Francis Crick published their work on the molecular structure of DNA double helix after observing the X-ray diffraction results done by Rosalind Franklin and Maurice Wilkins [1,2]. The simple base pairings, A-T and C-G, revolutionized virtually all biological science fields. Understanding the genome of organisms, especially humans, was believed to be the Holy Grail in understanding organism development and disease. This led to the race to publish the whole human genome. The human genome project was completed in 2003 [3,4]. It took about half a century from the discovery of DNA structure to the completion of the human genome. While we understand more about the DNA structure and the genome of several organisms, such as humans, fruit flies, laboratory mice, rice, etc, the understanding of protein structure is still limited.

In concurrence with the discovery of DNA's structure and the completion of the Human Genome Project, the first three-dimensional structure of protein was determined by X-ray crystallography. The limitations of protein structural determination include 1) the complexity of amino acids, 2) the difficulty of protein expression in large quantity for experimental structural studies, 3) the difficulty in phase determination especially for proteins with an unknown structural fold, and 4) difficulty in predicting crystallization feasibility. The complexity of amino acids, 20 amino acids compared to 4 bases in DNA, and the variations of protein structures complicate the generalization of protein structures. John Kendrew and Max Perutz published the very first crystal structures of myoglobin and then hemoglobin [5,6]. Both of these proteins are oxygen transport proteins that can be readily collected in large quantities from nature. Very few if any abundant proteins are left without their structures determined. We are now faced with the challenge of finding a new way to improve expression and purification of unstable proteins or membrane-bound proteins. Several protein expression systems ranging from *E. coli*, yeast, insect, to mammalian systems are currently used to overcome this important problem [7]. Multiple ways of phasing in protein crystallography have been introduced. The more structures we have the easier it becomes. Most structural determination was done using simple molecular replacement [8]. Numerous ways have been proposed to help in generating appropriate construct for protein crystallization. However, crystallizing protein is still one of the most daunting and unpredictable tasks and is the rate-limiting step of structural determination [9].

There have been attempts, historically, to compile most of protein structure in order to predict the folding of proteins with no known structure. This attempt led to structural genomics and structural proteomics [10,11]. The main goal of structural genomics/proteomics projects is to use the high-through-put screening in combination with traditional protein structural biology, X-ray crystallography and NMR, and structural computational modeling to describe the 3-dimensional structure of every protein encoded by a given genome [10,11]. While this is an enormous task that requires significant bioinformatics, it

allows a fast access to novel protein structures. Unfortunately, most of these proteins have no known function and therefore immediate use and publication of the work can be limited [10,11]. It is important to note here that traditional structural biology elucidating the structure of important proteins using X-ray crystallography and NMR along with functional studies is still the standard of publishing the structural work in a high-impact journal.

While learning novel protein folds is essential in moving structural biology forward, it is also important to collect as many experimental structures as possible especially in promiscuous proteins and enzymes that can interact with variety of ligands, substrates or inhibitors [12]. Computational modeling of protein structure can be useful in determining the general fold of the protein [13] and predicting the surface interactive site when combining it with functional studies [14]. However, the tertiary or quaternary structure of multi-meric or multi-domain proteins can be difficult to predict without experimental data. In these cases, structural experiments can be essential in determining the role of embedded structural interactive surfaces that may be exposed or hindered in certain ligand or protein binding partners [15].

In this special issue, we invited the authors to publish their structural studies and review of the protein structures in relation to human diseases. The challenge of protein structural biology is that there is a need for experimental data of novel structural fold and structural data. Coupled with this, there is a need for complementary bioinformatics and computational modeling that will allow further hypothesis generating and understanding of relatively solid crystal/NMR structures. Combination of these technologies and collaboration will, in the near future, lead us to the complete understanding of protein folding and structural prediction that will undoubtedly benefit science and medicine.

### References

1. Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171: 737-738.
2. Watson JD (1968) The double helix: a personal account of the discovery of the structure of DNA.
3. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. *Science* 291: 1304-1351.

\*Corresponding author: Sompop Bencharit, DDS, MS, PhD, FACP, Department of Prosthodontics, School of Dentistry, Department of Pharmacology, School of Medicine, University of North Carolina at Chapel Hill, CB#7450, Chapel Hill, NC 27599-7450, E-mail: [Sompop\\_Bencharit@dentistry.unc.edu](mailto:Sompop_Bencharit@dentistry.unc.edu)

Received July 13, 2012; Accepted July 14, 2012; Published July 18, 2012

Citation: Bencharit S (2012) History of Progress and Challenges in Structural Biology. *J Pharmacogenom Pharmacoproteomics* S4:e001. doi:10.4172/2153-0645.S4-e001

Copyright: © 2012 Bencharit S. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

4. Collins JE, Wright CL, Edwards CA, Davis MP, Grinham JA, et al. (2004) A genome annotation-driven approach to cloning the human ORFeome. *Genome Biol* 5: R84.
5. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, et al. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181: 662-666.
6. Perutz MF (1963) X-ray analysis of hemoglobin. *Science* 140: 863-869.
7. Yokoyama S (2003) Protein expression systems for structural genomics and proteomics. *Curr Opin Chem Biol* 7: 39-43.
8. McCoy AJ (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 63: 32-41.
9. Slabinski L, Jaroszewski L, Rodrigues AP, Rychlewski L, Wilson IA, et al. (2007) The challenge of protein structure determination—lessons from structural genomics. *Protein Sci* 16: 2472-2482.
10. Chandonia JM, Brenner SE (2006) The impact of structural genomics: expectations and outcomes. *Science* 311: 347-351.
11. Norin M, Sundstrom M (2002) Structural proteomics: developments in structure-to-function predictions. *Trends Biotechnol* 20: 79-84.
12. Bencharit S, Edwards CC, Morton CL, Howard-Williams EL, Kuhn P, et al. (2006) Multisite promiscuity in the processing of endogenous substrates by human carboxylesterase 1. *J Mol Biol* 363: 201-214.
13. Uhlik MT, Temple B, Bencharit S, Kimple AJ, Siderovski DP, et al. (2005) Structural and evolutionary division of phosphotyrosine binding (PTB) domains. *J Mol Biol* 345: 1-20.
14. Dibble CF, Horst JA, Malone MH, Park K, Temple B, et al. (2010) Defining the functional domain of programmed cell death 10 through its interactions with phosphatidylinositol-3,4,5-trisphosphate. *PLoS One* 5: e11740.
15. Bencharit S, Cui CB, Siddiqui A, Howard-Williams EL, Sondek J, et al. (2007) Structural insights into fibronectin type III domain-mediated signaling. *J Mol Biol* 367: 303-309.

This article was originally published in a special issue, **OMICS Techniques in Pharmaceutical Sciences** handled by Editor(s), Dr. Clar Lynda R. Williams-DeVane, U.S. Environmental Protection Agency, USA