Chapter 2
Describing Methodology and Applications of an In Silico Protein Engineering Approach

Trias Thireou
Agricultural University of Athens, Greece; National Technical University of Athens, Greece

Kostas Bethanis
Agricultural University of Athens, Greece

Vassilis Atlamazoglou
Agricultural University of Athens, Greece; National Technical University of Athens, Greece

ABSTRACT

This chapter presents a particular cascade of computational steps in order to build a workflow for an in silico protein engineering approach. In this respect, all available information, in order to choose and computationally implement mutations, is described, employed and monitored. Some of the prerequisites of in silico protein engineering are access to various sequence and structure molecular biology databases, software tools for three dimensional molecular visualization and manipulation, sequence and structure alignment and comparison, molecular modelling and molecular docking. The implementation of these steps is demonstrated in the context of performing mutations of particular residues on the ligand pocket of a lipocalin protein family member, to derive the desired ligand binding properties. The example chosen for inclusion introduces the reader to all of the essentials of computational protein engineering experiments. More importantly, it provides insight into understanding and properly interpreting the data produced by these methods.

ASPECTS OF PROTEIN DESIGN & ENGINEERING

Due to the complex nature of the protein-folding problem, the numerous attempts of *de novo* protein design have not lead to a major success. Therefore, more 'realistic' approaches have become accepted towards protein engineering, in order to acquire novel functions (DeGrado W.F., 1997; Desjarlais J.R. and Mayo S. L, 2002)

Indeed there is no need to design a protein sequence from scratch to do protein engineering. A

DOI: 10.4018/978-1-60566-768-3.ch002

Copyright © 2010, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

general strategy is rational design, in which the scientist uses detailed knowledge of the structure and function of the protein to make desired changes. In this sense and in the context of this chapter, we will focus on the concept of scaffolds that can be equipped with artificial binding sites, an approach that has gained recent interest. The term 'scaffold' is being used to describe some kind of natural protein architecture onto which unrelated structural elements can be incorporated and thus new biochemical activities created (Skerra A., 2000).

A few types of protein folds have been examined for this purpose. For example single immunoglobulin (Ig) domains and helical bundles, have been found to be useful for the generation of biomolecules with the ability of binding to other proteins. These scaffolds, however, are hardly capable of complexing small ligands (Skerra A., 2000).

In this respect, the lipocalin protein family appears to be a promising model system. Lipocalins constitute a family of small, robust proteins that typically transport or store biological compounds, which are either of low solubility or are chemically sensitive, including vitamins, steroid hormones, odorants and various secondary metabolites. The artificial lipocalins recognizing specific ligands, termed "anticalins", could provide an alternative to recombinant antibody fragments, with interesting applications in biotechnology and medicine.

Computational approaches offer significant potential for engineering protein structure and function, and can be combined with experimental testing to gain new insights into the fundamental properties of proteins, in rational structure based design. These methods offer the potential to drive structure and function manipulations for implementing *in silico* protein engineering experiments. In this respect, the number of experiments necessary to better understand a protein's function can be significantly reduced.

In this chapter, we present and discuss a particular cascade of computational steps in order to build a workflow for *in silico* protein engineering. In this respect, all available information, in order to choose and implement *in silico* mutations, is described, employed and monitored. Some of the prerequisites of *in silico* protein engineering are access to various sequence and structure molecular biology databases, software tools for three dimensional molecular visualization and manipulation, sequence and structure alignment and comparison, molecular modelling and molecular docking.

An in-depth description of key terms and concepts related to the above mentioned procedure is being described. Additionally the implementation of these steps is demonstrated in the context of performing mutations of particular residues on the ligand pocket of a lipocalin protein family member, to derive the desired ligand binding properties. The example chosen for inclusion introduces the reader to all of the essentials of computational protein engineering experiments. More importantly, it provides insight into understanding and properly interpreting the data produced by these methods.

PROTEIN SCAFFOLDS

As mentioned above, rather than attempting to design from scratch, one can utilize a number of nature provided templates to improve on. At the very least, these templates already fulfill the criteria for stably folding proteins.

Therefore, the idea is to utilize structurally well-defined polypeptide frameworks for the introduction of novel functions, by locally reshaping a part of the protein surface that is thought to be less important for the protein folding process or its stability (Skerra A., 2000).

Hence an ideal protein scaffold should provide a rigid folding unit, which spatially brings together several exposed loops (Ku J, Schultz PG. 1995). Since protein folding is a highly cooperative event, a critical issue is how it can be assessed whether a certain number of amino acid mutations can be

16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/describing-methodology-applications-silico-protein/39602

Related Content

Interactive Data Visualization to Understand Data Better: Case Studies in Healthcare System
Zhecheng Zhu, Bee Hoon Hengand Kiok Liang Teow (2014). *International Journal of Knowledge Discovery in Bioinformatics (pp. 1-10).*
www.irma-international.org/article/interactive-data-visualization-to-understand-data-better/147300

Incorporating Network Topology Improves Prediction of Protein Interaction Networks from Transcriptomic Data
Peter E. Larsen, Frank Collartand Yang Dai (2010). *International Journal of Knowledge Discovery in Bioinformatics (pp. 1-19).*
www.irma-international.org/article/incorporating-network-topology-improves-prediction/47093

Drug Optimization for Cystic Fibrosis Patients Based on Disease Pathways Crosstalk
Shuting Linand Yifei Wang (2021). *International Journal of Applied Research in Bioinformatics (pp. 1-11).*
www.irma-international.org/article/drug-optimization-for-cystic-fibrosis-patients-based-on-disease-pathways-crosstalk/267820

Information Therapy (Ix) Service and Patients' Preference
Vahideh Zarea Gavgani (2013). *Methods, Models, and Computation for Medical Informatics (pp. 117-125).*
www.irma-international.org/chapter/information-therapy-service-patients-preference/73074

The Benefits of Big Data Analytics in the Healthcare Sector: What Are They and Who Benefits?
Andrea Darrel, Margee Hume, Timothy Hardieand Jeffery Soar (2015). *Big Data Analytics in Bioinformatics and Healthcare (pp. 406-439).*
www.irma-international.org/chapter/the-benefits-of-big-data-analytics-in-the-healthcare-sector/121469