Visualization and Visual Analysis of Intermolecular Interactions of Proteins

Rigorous Thesis

Katarína Furmanová

Advisor: prof. Jiří Sochor
Co-Advisor: assoc. prof. Barbora Kozlíková

Brno, Spring 2017

Signature of Thesis Advisor

Declaration

Hereby I declare that this paper is my original authorial work, which I have worked out on my own. All sources, references, and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

<div align="right">Katarína Furmanová</div>

Advisor: prof. Jiří Sochor
Co-Advisor: assoc. prof. Barbora Kozlíková

Acknowledgement

Abstract

Keywords

Contents

PART I

OVERVIEW

1 Introduction

Proteins are highly complex macromolecules that are vital to biochemical processes taking place in each living organism. Weather alone or as a part of multi-unit complexes, they facilitate vast field of functions such as catalysing chemical reactions, transporting molecules across the cells or replication of DNA. In these processes the ability of a protein to interact with other molecules plays a defining role.

Since the proper understanding of protein interactions contributes to advances in medicine, pharmaceutics or even agriculture, the study of interaction patterns of proteins has been at the forefront of biochemical research for decades. Unfortunately, the complexity of protein structures and the necessity for expensive and time consuming in-vitro experiments make the progress in the area slow. Many computational tools aim to support this research by simulating the experiments in-silico and thus reducing the costs. However, these tools can produce a vast amounts of data. For example molecular dynamics simulations can mimic the movement of millions of atoms over a given period of time. It is virtually impossible to identify significant patterns by simply observing such simulation. Another example are the protein-protein docking simulations that predict the possible ways two or more proteins interact together. Here the output often comprises of tens to hundreds of possible conformations that the domain expert needs to analyse individually one by one.

Therefore, visualization and visual analysis tools became inherent part of proteomic research both as guidance during the experiments as well as for validation and analysis of results by the domain experts. The main aim of these tools is to speed up the analysis process by - often interactively - extracting the important features of the data and conveying them in such way, that previously hardly observable patterns and relationships become more prominent. Although much has been done in the field of molecular visualization in the past decades, there are still areas and problems that are not currently addressed.

In this work I will present the current state of the art techniques in the analysis of intermolecular interactions of proteins. I will identify the unsolved problems occurring in the literature, then present the proposed solutions and results that have already been achieved. I will also outline the possibilities for further research.

2 Biochemical Definitions

Although this thesis deals with the research in the field of visualization and visual analysis, it also ventures into the field of biochemistry. It is therefore inevitable to clarify the basic biochemical terms that will occur throughout the thesis and are important for its proper understanding. This section shall provide the reader with all the necessary knowledge.

2.1   Protein Structures

Proteins are complex molecules formed by one or more chains of amino acids. Amino acids are basic building blocks of all living organisms. There are approximately 500 known amino acids, but only 20 standard amino acids are encoded in genetic code. Each of them consists of *carboxyl group* $(-COOH)$, an *amino group* $(-NH_2)$ and a unique *side chain* $(-R)$ that defines its properties. The three groups are connected by a carbon atom, also called an *alpha carbon* C_α. See figure 2.1 a).
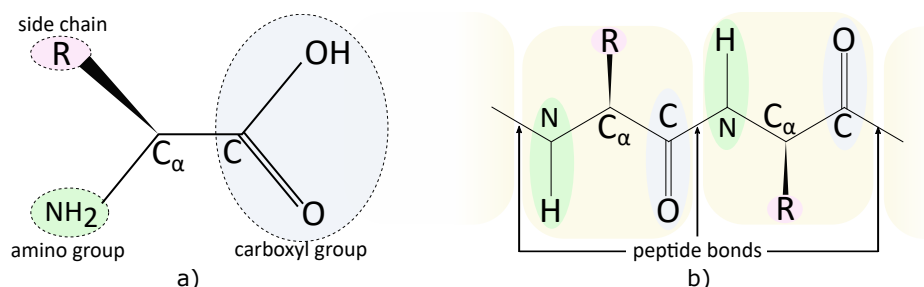

Figure 2.1: a) Illustration of a basic amino acid structure. b) Amino acid residues connected into polypeptide chain. It can be noted, that the amino and carboxyl groups are missing atoms, which were released during the formation of peptide bonds as H_2O molecules.

During a protein synthesis amino acids are joined together by peptide bonds (covalent bonds), forming polypeptide chains. A peptide bond is formed in a reaction between carboxyl group of one amino acid and amino group of another amino acid (see figure 2.1 b)). As both groups loose atoms that are released as molecule of water during this reaction, the amino acids bonded in polypeptide chains are refereed to as *amino acid residues*.

Each protein contains at least one long polypeptide chain. This sequence of amino acids, connected by rigid peptide bonds, also known as *backbone*, forms *primary structure* of the protein.

Unlike the peptide bonds, the bonds linking the carboxyl and amino groups to the alpha carbon are free to rotate. Based on these rotations and the patterns of hydrogen bonds that form between hydrogen from amino group and oxygen from corboxyl group, the segments of polypeptide chain can take on various 3D formations. The two most common of those are $\alpha -$ *helices* and $\beta -$ *sheets*, which are formed by laterally connected $\beta - strands$. These local formations of polypeptide chain are called *secondary protein structures*. Parts of polypeptide chain with absent secondary structures are called *random coils*. See figure 2.2.
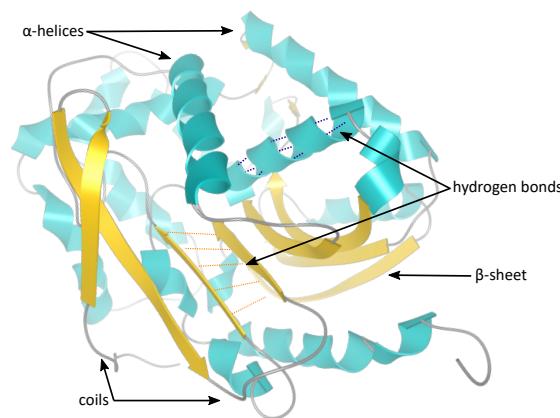

Figure 2.2: Typical secondary structures of protein: $\alpha - helices$ (blue), $\beta - strands$ (orange) forming $\beta - sheet$ and *coils*.

Various side chains of amino acid residues can interact together during the formation of protein. As a result, the secondary structures of the protein are bended and shaped into a unique 3D structure until the protein attains its minimal energy state. This process is called *protein folding* and it results in a *tertiary protein structure*. The tertiary structure defines the complete spatial arrangement of atoms of one polypeptide chain. Interactions between amino acids of multiple polypeptide chains than define their *quaternary protein structure*.

2.2   Properties of Proteins

Previous section described the process of protein attaining its 3D structure. This structure directly influences the way protein is behaving with regards to other molecules and its ability to function properly.

Example of this are the inner voids of the protein. When protein folds, there is naturally some empty space left inside. Depending on the shape of the space we classify four types of inner voids(figure 2.3): *cavities* – void space buried deeply inside the protein, *tunnels* – connecting cavities with surface of the protein, *channels* – passing through the whole protein and *pockets* – shallow protrusions on the surface of the protein.
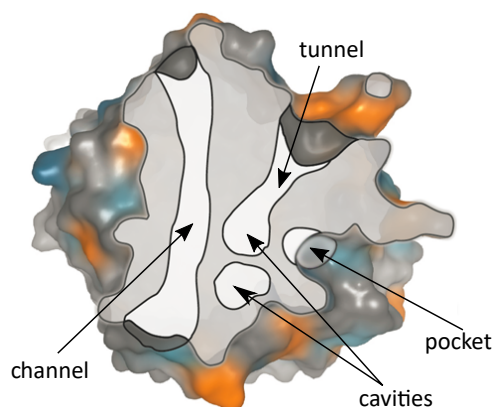

Figure 2.3: Types of inner voids of protein.

These inner voids can significantly influence the reactivity of the protein since they contain *active site*. Active site is a region of reactive amino acids, where other molecules can bind to protein and undergo a chemical reaction. This place is often buried deep inside the protein and its accessibility is thus limited by the size, shape but also physico-chemical properties of the voids leading to it.

2.3   Protein-Ligand Interactions

Ligand

In this thesis I will focus on two types of intermolecular interactions of proteins: a) protein-ligand interactions and b) protein-protein interactions.

Protein-ligand interaction. This type of reaction occurs when a protein reacts with a small molecule called ligand. Here the aim single ligand-protein pair, key lock priciple....

2.4 Protein-Protein Interactions

3 State of the Art

4 Aims of the Thesis

5 Achieved Results

PART II

PUBLICATIONS

6 Author's Publications