El titulo de la tesi: In-silico methods to drug discovery

El subtitulo de la tesi: Cancer

Autor: Francisco Martínez Jiménez

TESI DOCTORAL UPF / ANY L'any de la tesi: 2016

DIRECTOR DE LA TESI
Director: Marc A. Marti-Renom Departament Departament: Biomedicine

A mi madre.

Agradecimientos Agraeixo....

Abstract

This is the abstract of the thesis in English. Please, use less than 150 words.

Resum

Vet aqui el resum de la tesi en catala.

Prefaci

Sumari

Índex de figures

Índex de taules

Summary

sencillo

Consta de

tesi-upf.cls `book`

1. Es redissenya la portada \maketitle).

2.

3. Es redefineix `cleardoublepage` para que las paginas en blanco no se numeren

Preambulel

paquets `crop` i `geometry`.

Taules Includes `figure` i un `tabular`

Index

1. lo llamas en preambulo \usepackage{makeidx} \makeindex con esto lo imprimes \printindex con esto lo creas `makeindex`

<div align="center">

CAPÍTOL 1

INTRODUCTION

</div>

1.1   Protein overview

The importance of proteins in biological chemistry is reflected by their name, coined by Jons Jacob Berzelius in 1838, and derived from the Greek word *proteios* that means "of the first rank"[1]. Their presence is so essential that they constitute most of the cell dry mass [1]. They are not only the cell's building blocks, but also they perform nearly all the cell's functions. Some roles of proteins include serving as structural components of cells and tissues (e.g., *keratin* or *collagen*), transmission of information between cells by hormones such as the *insulin* or the *oxytocin*, facilitating the transport and storage of small molecules (e.g., the transport of oxygen by *hemoglobin*) or providing a defense against foreign invaders (e.g., antibodies). Other proteins such as the *actin* and the *myosin* are responsible of muscle contraction and therefore our movement. However, the most fundamental role of proteins is their ability to act as enzymes, which, catalyzes most of the chemical reactions in biological systems.

There is experimental evidence of more than 30,000 human protein products derived from over 17,000 human genes [2].

1.1.1   Protein structure

A protein is a molecule made from a long chain of amino acids linked thorough a covalent peptide bond. Proteins are therefore also known as *polypeptides*. At-

---

[1]The term protein was first used by Gerardus Johannes Mulder, advised by Berzelius, in its publication *Bulletin des Sciences Physiques et Naturelles en Néerlande (1838). pg 104. SUR LA COMPOSITION DE QUELQUES SUBSTANCES ANIMALES*, where he observed that all proteins seemed to have the same empirical formula and came out to the erroneous idea that they might be composed of a single type of very large molecule. Berzelius proposed the name because the material seemed to be the primitive substance of animal nutrition that plants prepare for herbivores.

tached to this repetitive chain are those portions of the amino acids that are not involved in the covalent bond, the **side chains**. Side chains confer the different physico-chemical properties of each of the 20 types of amino acids [3]. The composition of the amino acid sequence determines the function and the structure of a protein. That is because the unique sequence creates a specific pattern of attractive and repulsive forces between amino acids along the polypeptide that leads to a folding process resulting in a specific three-dimensional structure. These forces are usually non-covalent interactions between the side chains of the amino acids. Non-covalent interactions are weaker than covalent ones, allowing the folded structure to certain degree of conformation mobility i.e: to be dynamic. This phenomenon is really important to facilitate the interaction with other molecules as we will explore further in 1.1.3.

Protein structures are complex conformation of atoms organized in a hierarchical manner 1.1. The first level of this hierarchy, referred to as the **primary structure**, is the ordered sequence of amino acids of the polypeptide. Certain segments of these chains, tend to form simple shapes such as helices, strands, turns or loops. These folding patterns are referred to as secondary elements and collectively constitute the **secondary structure** of the protein. The two most frequent type of secondary elements are the α-helixes and the β-sheets [4]. The overall chain tends to fold further into a three-dimensional **tertiary structure**. Contrary to the secondary structure, the tertiary structure folding is driven by interactions from amino acids far apart in the primary sequence. The tertiary structure, is generally the most stable form of the protein, that is, the one that minimizes its free energy [5]. Furthermore, the tertiary structure is also the biologically active form of the protein, and its unfolding usually leads towards partial o total inactivation of the protein. Finally, some proteins are composed by multiple folded chains. In such cases, each folded subunit folds independently and then joins the others forming a biologically active complex. This type of organization is considered as the **quaternary structure**.

Protein Structure Determination
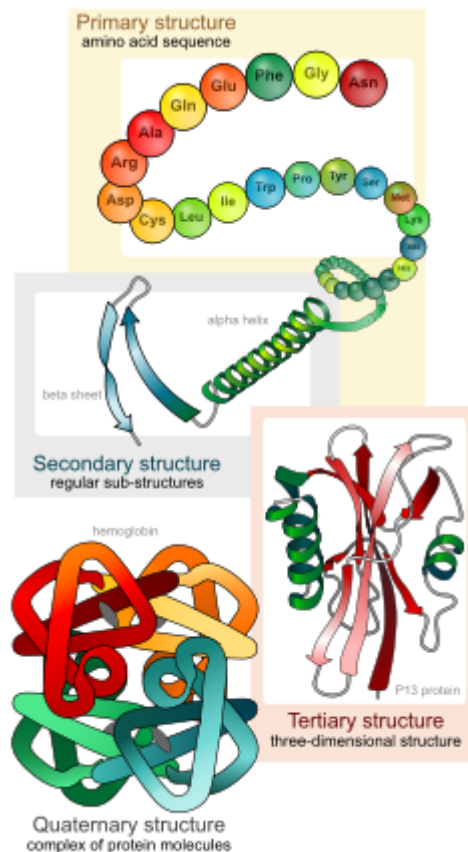
Protein Structure Prediction

4

Figura 1.1: Hierarchical distribution of layers in protein structure

1.1.2 Protein function

The structure of a protein determines its biological function. However, different *regions* of the structure can perform semi-indepent functions from each other. These regions are referred to as **protein domains**. A domain is substructure produced by any part of polypedtide chain that can fold independently into a compact and stable structure [6, 7, 8]. Domains on average contain 80-250 residues [9]. Estimates of the number of domains per protein say that nire than 70% of procaryotik proteins and 80% of eukaryotic proteins include more than one domain [10, 11]. Among this multi-domain proteins, 95% of them contains only two to five protein domains [10]. Domains are not only the basic functional units of proteins, but also the evolutionary units of protein evolution. As proteins have evolved, domains have been modified and combined to build new proteins [12, 13]. Such is the importance of domains in protein evolution, that they have been included in cur-

rent protein classification methods as one of the major classification parameters. Some of these domain classification methods such as SCOP [14] or CATH [15] are purely based on the structure, while others such as Pfam [16] or INTERPRO [17] include information about the function in their classification.

Domains, and consequently proteins, perform its biological activity by interacting with other molecules. Proteins can interact with other proteins, constructing a protein-protein complex, with ions or with small-molecules. The substance that is bound to the *target* protein is called the **ligand**, while the region of the protein where the ligand is binding is called ligand's *binding site*[2].

1.1.3   Protein-Ligand Interactions

The roles played by the ligands are diverse. Table X shows an example of the different functions that a small-molecule ligands can perform in a protein. Binding constants, allosteric and binding-site, induced fit model. Expandir. Imporatante.

1.1.4   Protein-ligand prediction

1.2   Drug discovery

0	0
0	0

Taula 1.1: Prova de taula

1.2.1   subsection

Subsection

1.3   Drug discovery

Second

1.3.1   In-silico methods in drug-discovery

Subsection

---

[2]For simplicity, in this manuscript, unless otherwise indicated, the term ligand will only refer to small molecules ligands, while proteins ligands will be explicit named as protein-protein interactions

1.4 Mycobacterium tuberculosis

1.4.1 Tuberculosis treatments and PPcs

1.5 Drug resistance in cancer

1.5.1 Cancer Treatment and drugs

CAPÍTOL 2

OBJECTIVES

CAPÍTOL 3

NANNOLYZE

CAPÍTOL 4

PREDICTING TARGETS IN MTB

CAPÍTOL 5

DRUG RESISTANCE IN CANCER

Figura 5.1: Example

Bibliografia

[1]   A Kessel i N Ben-Tal. *Introduction to Proteins: Structure, Function, and Motion*. Chapman & Hall/CRC Mathematical and Computational Biology. CRC Press, 2010. ISBN: 9781439810729 (v. la pàg. 3).

[2]   Min-sik Kim et al. **?**A draft map of the human proteome**?** A: (2014). DOI: 10.1038/nature13302 (v. la pàg. 3).

[3]   B. Alberts. *Molecular Biology of the Cell: Reference edition*. Molecular Biology of the Cell: Reference Edition v. 1. Garland Science, 2008. ISBN: 9780815341116 (v. la pàg. 4).

[4]   Wolfgang Kabsch i Christian Sander. **?**Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features**?** A: *Biopolymers* 22.12 (1983), pàg. 2577 - 2637. ISSN: 1097-0282. DOI: 10.1002/bip.360221211 (v. la pàg. 4).

[5]   K A Dill. **?**Dominant forces in protein folding.**?** A: *Biochemistry* 29.31 (1990), pàg. 7133 - 7155. ISSN: 0006-2960. DOI: 10.1021/bi00483a001 (v. la pàg. 4).

[6]   Jane S. Richardson. *Advances in Protein Chemistry Volume 34*. Vol. 34. 1981, pàg. 167 - 339. ISBN: 9780120342341. DOI: 10.1016/S0065-3233(08)60520-3 (v. la pàg. 5).

[7]   Extracellular Proteins That i Modulate Cell-matrix Interactions. **?**Extracellular Proteins That**?** A: 266.23 (1991), pàg. 15 - 18 (v. la pàg. 5).

[8]   D B Wetlaufer. **?**Nucleation, rapid folding, and globular intrachain regions in proteins.**?** A: *Proceedings of the National Academy of Sciences of the United States of America* 70.3 (1973), pàg. 697 - 701. ISSN: 0027-8424. DOI: 10.1073/pnas.70.3.697 (v. la pàg. 5).

[9]   Suhail A Islam et al. **?**Identification and analysis of domains in proteins presented to identify domains in proteins**?** A: 8.6 (1995), pàg. 513 - 525 (v. la pàg. 5).

[10]   Jung-Hoon Han et al. **?**The folding and evolution of multidomain proteins.**?** A: *Nature reviews. Molecular cell biology* 8.4 (2007), pàg. 319 - 330. ISSN: 1471-0072. DOI: `10.1038/nrm2144` (v. la pàg. 5).

[11]   Cyrus Chothia et al. **?**Evolution of the protein repertoire.**?** A: *Science (New York, N.Y.)* 300.5626 (2003), pàg. 1701 - 3. ISSN: 1095-9203. DOI: `10.1126/science.1085371` (v. la pàg. 5).

[12]   Christine Vogel et al. **?**Structure, function and evolution of multidomain proteins**?** A: *Current Opinion in Structural Biology* 14.2 (2004), pàg. 208 - 216. ISSN: 0959440X. DOI: `10.1016/j.sbi.2004.03.011` (v. la pàg. 5).

[13]   G Apic, J Gough i S a Teichmann. **?**Domain combinations in archaeal, eubacterial and eukaryotic proteomes.**?** A: *Journal of molecular biology* 310.2 (2001), pàg. 311 - 325. ISSN: 0022-2836. DOI: `10.1006/jmbi.2001.4776` (v. la pàg. 5).

[14]   Alexey G. Murzin et al. **?**SCOP: A structural classification of proteins database for the investigation of sequences and structures**?** A: *Journal of Molecular Biology* 247.4 (1995), pàg. 536 - 540. ISSN: 00222836. DOI: `10.1016/S0022-2836(05)80134-2` (v. la pàg. 6).

[15]   Ca Orengo et al. **?**CATH - a hierarchic classification of protein domain structures**?** A: *Structure* March (1997), pàg. 1093 - 1109. ISSN: 09692126. DOI: `10.1016/S0969-2126(97)00260-8` (v. la pàg. 6).

[16]   A Bateman et al. **?**The Pfam protein families database**?** A: *Nucleic Acids Research* 28.1 (2002), pàg. 276 - 280. ISSN: 0305-1048 (Print) 0305-1048 (Linking). DOI: `gkd038[pii]` (v. la pàg. 6).

[17]   Sarah Hunter et al. **?**InterPro: The integrative protein signature database**?** A: *Nucleic Acids Research* 37.SUPPL. 1 (2009), pàg. 211 - 215. ISSN: 03051048. DOI: `10.1093/nar/gkn785` (v. la pàg. 6).