Writing with amino acids: designing the folding and binding of model proteins
Coluzza, I.

Citation for published version (APA):
Coluzza, I. (2005). Writing with amino acids: designing the folding and binding of model proteins

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Download date: 27 Dec 2018
1 Introduction

Proteins are the ultimate nano-machines. Moreover, they provide the building blocks for complex structures. Many protein-based machines and materials have evolved during the past 4 billion years. For instance proteins can assemble into complex structures that can reach macroscopic sizes. Examples of such structures are actin filaments (Fig. 1.1) and microtubules. These protein-based structural units are responsible for the structure and elastic properties of many cells. Proteins perform specific tasks with high selectivity. The aim of the present thesis is to gain a better understanding of how proteins function, by studying the behavior of simple model proteins.

Proteins are heteropolymers composed of 20 different types of amino acids. Different proteins have different chain lengths. Depending on the amino-acid sequence, some proteins can collapse to form a well-defined “native” conformation whilst others can not. This process of forming a compact, native structure is called folding. Usually, proteins are only biologically active in their native state. The tasks that proteins perform are very diverse; they usually involve the interaction with other proteins or other biomolecules, such as DNA. These interactions are controlled by the same elements that encode for the native structure of the protein itself. Structure and function are, as consequence, strongly correlated, and are directly dependent on the sequence of amino acids along the protein chain. A better understanding of the relation between sequence, structure and function, is crucial for the design of biomolecular materials and molecular machines.

Another way to look at functional proteins is as information carriers. Proteins are an essential part of gene-regulatory networks that control the response of the cell to the environment. A simple way to understand the function of regulatory networks is to picture the stimuli coming from the world outside the cell membrane (e.g. signaling proteins, changes in food concentrations...) as data input in a computer program that will generate a different output, depending on which conditions are satisfied. The code for the program and the design of the computer hardware are written in the DNA. The role of the network is to transmit the signal and translate it from a the chemical language to the expression of proteins that are coded for by the genetic material. Once the signal reaches the DNA, the output is provided in the form of promotion or repression of the expression of specific proteins that will then start a new cascade of reactions to translate back the answer from the genetic code to necessary chemical reactions. This signal-processing property can be achieved only if each element is designed to interact selectively with its molecular partners, otherwise the signal would generate “crosstalk” with potentially deleterious consequences. It is important to remember that such selectivity was designed through the evolution of the amino acids sequence of the protein. An example of the gene-regulatory network of the E-Coli bacterium is shown schematically in Fig. 1.2.
Figure 1.1: One of the most famous example of self assembly in nature, the actin filaments are long chain of the same repeating unit protein. This filaments are the building blocks of the cytoskeleton of many cells, and together with myosin is the most important components of muscular fibers in all animals [1] (http://www.cgl.ucsf.edu/chimera/ImageGallery/entries/actin/actin.html).
Figure 1.2: Gene Regulatory Network (GRN) of TFs in E-coli. This figure illustrates the complexity of the core of the GRN in E-coli, and gives a glimpse of the number of interaction that are involved in a subset of physiological activities of one of the simplest living organisms [2].
I Introduction

There are two ways to approach the modeling of the relation between sequences and function. One is to focus on the properties of specific proteins. Such an approach requires an accurate, atomistic model of the protein and its surrounding medium. The other approach focuses on the more general question how the heterogeneity of a heteropolymer can lead to both folding and specific interactions. In such studies, one can use simpler (cheaper) models to reproduce the specificity of proteins. The reason why we use a simplified model is that the numerical study of conformational changes in proteins tends to be computationally very demanding. The power of computers has only recently reached the level where it becomes feasible to simulate the folding of a single, relatively short, protein. However, for a systematic study of the relation between sequence and conformational change, it is necessary to explore the properties of a great number of molecules with different amino-acid sequences. For such studies, fully atomistic models are not an option. It should be stressed that a protein is not only optimized to fold into a specific structure, or to undergo a specific conformational change (as in a motor protein). Its is also designed to interact strongly with specific substrates but only weakly, if at all, with all other molecules that it encounters as it diffuses through the cell. Clearly these conditions increase the complexity of protein design.

In what follows, we will use lattice heteropolymers to represent the peptide chain. This model is one of the simplest representation of a protein and has been extensively used for the study of folding properties. In Chapter 2 we describe the methods that we developed to artificially evolve the model protein to perform specific, elementary tasks. Once a protein has been “designed” the next step is to study the conformational space of the protein to test if it performs as designed. In Chapter 3 we describe the algorithm that we used to sample the free energy landscape of the heteropolymer even in regions of high free energy. In Chapter 4, we describe the application of our model to the study of conformational changes in proteins induced by a chemical agent. We pay special attention to important role of thermal fluctuations that allow the system to reach state at which the cost of the transition to one structure to the other one is considerably reduced. In Chapter 5 we extend our design technique to include specific binding properties between a protein and a substrate. We study the influence of such substrates on the transition between two conformations and on the folding properties of random domains of proteins (Chapter 6). The final step is the study of a particular class of protein called chaperonins that act as folding catalyzers. We introduce a model for chaperonin action that suggest that protein translocation is a key step by which chaperonins refold misfolded proteins into their native state (Chapter 7).