SPECIFICITY AND CONSTRAINTS IN PEPTIDE-PROTEIN BINDINGS IN THE MOUSE PROTEOME

Report for 3rd Year Research Project

February 5, 2016

Dhruv SHARMA

Contents

Part I
Introduction

This report details the work done towards the fulfillment of the requirements of the department of Physics at Ecole Polytechnique. I undertook this research project under the guidance of Dr. Remi Monasson at the Laboratoire de Physique Theoriqu at the Ecole Normale Superieure in Paris.

The aim of the project was to study the interactions between short peptide chains and a specific section of signalling proteins called PDZ Domains. One of the aspects that we study here is the specificities of interactions between peptides and PDZ domains. It is well known that macromolecules such as proteins and enzymes interact in a specific manner with other macromolecules and biomolecules. What interested us over the course of the study are the constraints present in the peptide sequences due to the specificity of their interactions with PDZ domains. We will also have an occasion to understand similar constraints on the PDZ domain sequences.

This report is organized as follows. After a brief introduction to the biological importance of PDZ Domains, we explain the experiments performed by **Insert reference here**. Using these experiments, Stifler et al created a model which is capable of predicting whether a peptide will bind to a PDZ domain given the sequence of the peptide. We shall explain the data that Stifler et al have provided. The first two models that we propose utilise the data provided by Stifler et al.

Once the data presented and the biological context established, we present a first model which seeks to understand the constraints imposed on the peptide sequences under the effect of mutations. We present the results derived from this model and discuss the limitations. A second improved model is then proposed which considers error rates as probabilities. We present some interesting observations on the basis of this model. In particular, we show how certain positions are particularly constrained over all peptides and present a simple way of the calculating the level of constraint.

Finally, to render the study of peptide-PDZ domain specificity complete, we explain how we could integrate the PDZ Domain sequences into the modelization. This is done by a regression method called the *Lasso*. We shall have the chance to present the lasso method in more detail in the relevant section.

We conclude with a summary of our findings and possible directions of further improvements.

1 PDZ Domains

Let us begin by explaining the importance of PDZ Domains, their importance, their structure and how they bind to other macromolecules. PDZ domains are short sections of proteins composed of 80-90 amino acids. PDZ Domains are usually found in signalling proteins where they regulate processes such as the separation of cell membranes. Th

Acknowledgements

I would like to thank Dr. Remi Monasson for guiding me throughout the project. He showed great patience with me and made himself available for me during the project. I am especially grateful also for the numerous conversations surrounding the emerging domain of deep learning, a field which interests us both. I would also like to thank Simona Cocco for her pertinent remarks. Finally I want to thank David, Antoine and Andre for listening to me patiently (more or less) as I struggled to explain my project to them.