Structural Bioinformatics Final Exam
Protein Part

ANTONIO ORTEGA JIMÉNEZ
University of Copenhagen
January 24, 2017

RMSD analysis on identical n-mers in the top100H database

Contents

1 Introduction

The study of protein structure is a hot topic on several fields like medicine and the biotech industry. For that matter, Bioinformatics methods that predict and analyze structure information are constantly being developed by the scientific community. Moreover, since research groups all around the world publish their results on open databases this methods can be easily be applied to real data. One of such databases is the Protein Data Bank, maintained by an international consortium. In other words, bioinformaticians have both open data and free software to try out new ideas.

Do we expect structure similarity in identical sequences of n residues? In this report, a simple *in silico* analysis of protein structure data was run to try to give an answer to this question. The Biopython module, contained in the Python programming language, together with the statistical language R, were used for this purpose.

2 Methods

2.1 Build of the top100H

The top100H database is available in this link http://kinemage.biochem.duke.edu/databases/top100

3 Results

4 Discussion