Targeted Peptides and Rosetta on AWS

Tuesday, August 2, 2016 11:24 AM

Introduction

The purpose of this documentation is to track an implementation of the Rosetta molecular engineering software on the AWS cloud with the idea of supporting targeted peptide research undertaken by Gaurav Bhardwaj at the University of Washington Molecular Engineering and Science Institute.

Premise

(Written by a non-molecular biologist) DNA consists of (in humans) about 3 billion base pairs arranged in a helical ladder. A significant task of DNA is to record how to construct proteins, as follows: Each base location can have one of four values (to abuse the terminology, 'nucleotides') abbreviated A, C, G, or T. This is paired across a rung of the ladder to a complementary nucleotide, respectively T, G, C and A. Three bases in a row can be thought of as three digits in base-4; for example AAG or TCA, i.e. a number from 0 to 63. These values found in DNA in turn encode/choose/correspond to one of 20 or so left-handed amino acids found in nature. Notice that having 64 possible values per nucleotide triple over-determines a choice of 20 amino acids; so there is some degeneracy and there are some nucleotide triples that do not encode amino acids.

A sequence of several amino acids -- let us say 20 -- would be encoded by 20 x 3 = 60 nucleotides; and a cell's (ribosome) machinery can bind these amino acids together in a chain; whereupon they can be released into that water environment of the living organism; such as a human cell. This environment includes salts.

Upon release from this machinery the amino acid chain will fold itself into some form according to the laws of chemistry (or if you like: physics) to exist as a small protein called a peptide, typically weighing 2000 Daltons or so. This peptide is available to perform some function in the organism.

Naturally available amino acids have a left-handed chirality; a structure that 'curves to the left' in some sense. Right-handed amino acids can be manufactured in a laboratory and are intrinsically less bio-reactive than natural left-handed amino acids. Right and left-handed amino acids are called respectively D- and L-amino acids.

An interesting idea: Making a peptide using D-amino acids can help that peptide avoid being torn apart by an organism's L-oriented chemistry. The D-amino acids simply don't register.

An interesting idea: Making a therapeutic molecule small and compact could also help keep it stable; and a smaller molecule (a peptide) is more amenable to computational solutions since there are few computing operations needed to evaluate its structure and utility.

An interesting idea: Building a biologically active binding site onto a D-based peptide could be a novel approach to engineering new therapeutic medicines.

Key point: The Rosetta software suite is a multi-faceted set of tools designed to evaluate protein structure and function.

Objective: Implement Rosetta on the AWS cloud; with extensive documentation on this process.

Links

Documentation on Rosetta scripts
https://www.rosettacommons.org/docs/latest/scripting_documentation/RosettaScripts/RosettaScripts
Compiling Rosetta
https://www.rosettacommons.org/docs/latest/build_documentation/Build-Documentation
Overall documents
https://www.rosettacommons.org/docs/latest/Home