A Genetic Algorithms approach for Drug Design

Lim Teck Sin

Pembroke College

Oxford University Computing Laboratory

Programming Research Group

8 - 11 Keble Road

Oxford, OX1 3QD

Thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Computation

University of Oxford

FORWARD

The ability to design ligands that can bind targets specifically and selectively is an important goal in drug design. By studying the structures and considering various interactions of the molecules, computers can be used to predict de novo drugs.

One of the approaches utilises hydrophobicity scales to determine the optimum complementary peptide for a given target. Attempts have been made to use Genetic Algorithms (GA) to derive optimal peptides. The preliminary results indicate that the predictions can be used in protein purification [M Chung et al, unpublished results]. Although the molecular basis for interaction between a peptide and its complement are poorly understood, it seems that GA maybe useful in tackling protein-ligand interactions. (Chapter 2)

As the next step, it is proposed to move beyond complementary peptides (a purely two dimensional approach) into predicting three dimensional interactions between a protein and its ligand, using GA techniques.

Due to time constraint, it is proposed that the project will only involve the application of GA to minimise the structure of a ligand. The strategy involves generating a pool of the conformers which have identical chemical formula but different structural conformation. Each of these has a fitness based on energy calculations. The computation will involve van der Waals forces, bond lengths, bond angles and torsion angles and perhaps ionic charges as well. A more stable conformer with lower energy will have better fitness. By varying the conformations via crossover and mutation, and the propagation of fitter conformers, the programs should be able to predict conformers which are stable. The results obtained can then be verified by comparing with data obtained via crystallography database. (Chapter 5)

Before the above three dimensional computation can be done, a number of feasibility studies have to be performed:

a. Investigate a suitable GA implementation to search for structures. Besides being limited by the quality of data, the results obtain will also depend a lot on the quality of the search performed.(Chapter 1, 2 and 4)

b. Investigate if GA can be successfully parallelised. A fast and parallel GA implementation will allow more experiments to be performed and may also improve the accuracy of the results. (Chapter 3)

The ability to minimise conformers properly should pave the way for the identification of suitable ligands to bind targets. It is envisaged that GA can be used to find the optimum target-ligand pairs by creating generations of minimised and identifying those having the best binding fitness with target. A successful parallel implementation of GA will help to accelerate the achieving of this goal.

ACKNOWLEDGEMENT

I would like to thank the Center for Natural Products Research (CNPR) for its funding and interest in the project. The valuable comments of T Hiroshi on GA, R Miller on BSP, C S Low on 3D graphical libraries, Dr Ganesan, Dr T W Tan and Dr S Subbiah on biological aspects of molecular modelling are much appreciated. I am also grateful for the supervision provided by Dr S Muggleton.

CONTENTS