

An Indian-Australian research partnership

Project title: Low resolution models for ab initio protein folding prediction

Project number: IMURA0151

Monash University supervisors: Dr. Madhu Chetty and Prof Mohan Krishnamoorthy

Monash University contact: Dr. Madhu Chetty; Email: madhu.chetty@infotech.monash.edu.au

IITB supervisors: Prof. Pramod Wangikar

IITB contact: Prof Pramod Wangikar; Email: pramodw@che.iitb.ac.in

Research Academy theme/s

1. Advanced computational engineering, computational intelligence, optimisation and simulation
2. Biotechnology and Bioinformatics

The research problem

Real PFP simulation is a very time consuming and expensive task – IBM has started its *Blue Gene* project to have *ab initio* protein folding simulation. The capacity of Blue Gene is 1 petaflop that is, it can perform 10^{15} floating point operations per second. With this very high capacity, it is estimated that to simulate 100 μ sec of real world protein folding, it will require 3 years.

So, we would undertake to investigate our real PFP starting from elementary models. And then move towards formulation of real PFP with every possible optimization.

- a) *Speeding up by optimization of the fitness computation*
- b) *Optimization of GA (or a suitable non-deterministic search algorithm) for elementary HP model*
- c) *Removal of the limitation of GA (e.g. the impact of twins in the population)*
- d) *Computationally expensive real world protein folding prediction*
- e) *Development of fitness function*
- f) *Advancement of GA for larger sequences*

Project aims

The objectives are summarised as

- Implementing novel optimisation strategies to predict the *ab initio* three dimensional structure of a protein from its primary amino acid sequence.
- Developing suitable low resolution (e.g. HP, HPNX, hHPNX models) and high resolution protein models (e.g. all atoms model).
- Incorporating heuristics for making GAs less computationally expensive (more efficient) in Protein Folding Prediction (PFP).
- Extending protein folding problem for drug design.

Expected outcomes

Outcomes of the research will contribute as following:

Protein folding prediction

If it is possible to predict the *ab initio* protein folding – then it will be possible to refine and verify the already known structure. Most importantly, protein folding of the remaining known gene or amino acid sequence can be discovered in inexpensive and accurate manner. *Ab initio* prediction will play a vital role in redacting the membrane protein because it is hard to predict this class of protein using X-ray or NMR due to transparency.

Protein function prediction

Protein structure helps define the function of the protein – as the ab initio protein folding prediction become ready then functionality of protein rooted at the gene or DNA sequence will be evident.

Drug designing

Proteins function by docking. In docking receptor (protein) contains pocket to bind the ligand (drug). Same pathway searching methodologies of PFP is useful for finding the position and orientation of the two molecules being energetically minimized. Thus PFP will be helpful in designing and synthesizing drug.

Combating critical diseases

By being able to design drug, PFP will ultimately help combating diseases mostly without side-effect.

Which of the above Theme does this project address?

The cutting edge project is multidisciplinary project and involves development of advanced computational techniques in bioinformatics domain. It tries to address a problem classified as 'grand challenge' problem in bioinformatics.

How will the project address the Goals of the above Themes?

Due to extraordinarily large number of conformations possible for a given protein sequence, a need is felt for a reliable and robust optimisation search techniques. The project will also involve development of appropriate models for representation of proteins realistically.