

# Hydrophobic-Hydrophilic Patterns and their Effects on the Protein Folding Prediction Process

Trent Higgs                      Bela Stantic                      Md Tamjidul Hoque  
 T.Higgs@griffith.edu.au      B.Stantic@griffith.edu.au      T.Hoque@griffith.edu.au

Institute for Integrated and Intelligent Systems, Griffith University, Queensland, Australia

## Abstract

To enhance the *ab initio* prediction approach we have investigated the influence of a Hydrophobic-Hydrophilic *shape-pattern* to the fragment selection/generation process in order to improve the quality of selected fragments for Protein Folding Prediction .

**Keywords:** Protein Folding Prediction, Fragments, HP Patterns

## 1 Introduction

To improve the performance of the *ab initio* prediction approach numerous protein software prediction suites, ROSETTA [1] and TASSER [3], use protein fragments or sub-conformations. A HP (Hydrophobic-Hydrophilic) pattern model can add an extra layer of sensitivity to the fragment selection process within fragment-based protein prediction software [2]. By presuming a certain HP sub-sequence pattern for a particular sub-conformation it can allow the best sub-conformation to be used at a particular segment of the protein chain in less time, therefore quicken the accuracy and computation time of the prediction process.

## 2 Method and Results

Table 1: Number of valid or non-valid conformations for k-sized sub-conformation, where k = 3 to 10

Length	No of Valid	No of Non-Valid
3	2033682	2539489
4	1639016	3603347
5	1171918	4361272
6	786637	4861816
7	508824	5173235
8	321243	5353717
9	198865	5450304
10	121113	5493685

To observe HP interactions within a k-sized sub-conformation we have analysed the amino acids contained within that sub-conformation, classing them as either H or P, then checking if all of the H amino acids are closer to the core of the protein than the P amino acids. This is done for every k-sized sub-conformation contained within every protein in our database (approximately 24 000). For an example of how this works please see Figure 1.

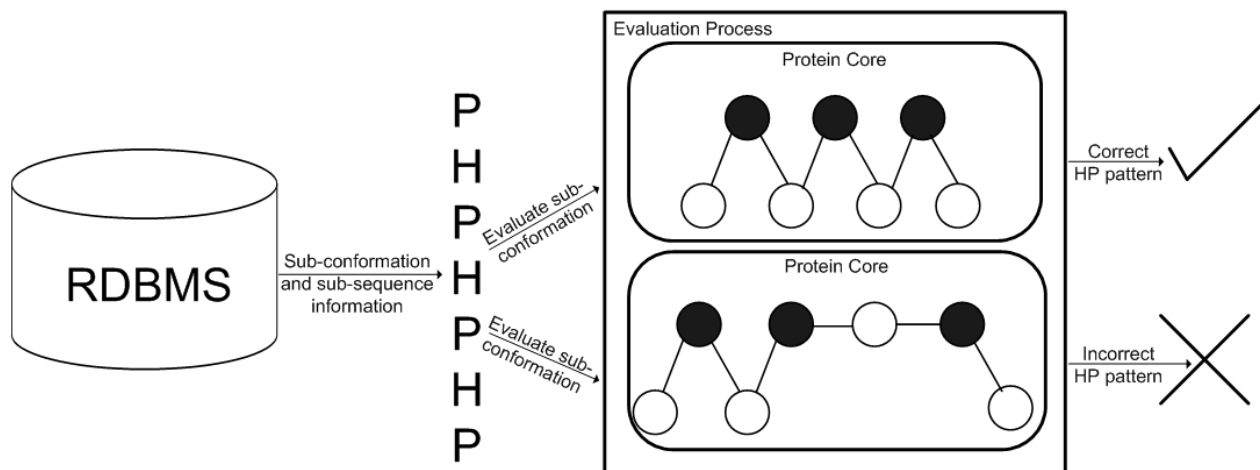


Figure 1: We take a  $k$ -sized sub-conformation ( $A$ ) within a particular protein from our database, in this case  $k=7$  and the sub-sequence is PHPHPHP (H = Hydrophobic, P = Hydrophilic), and check whether all the H (black) amino acids within  $A$  are more towards the core of the protein than the P amino acids (white), if so it is a valid conformation otherwise it is an invalid conformation.

### 3 Discussions

We observe from the experimental results in Table 1 that, as the sub-conformation grows in size there are more non-valid conformations in comparison to valid formations, which leads bi-directionally: (1) to investigate more rigorously on limited size of the sub-conformation to have higher confidence, (2) combinatorial effect is concluding the final folding over the hydrophobic effect. Thus multi-objective optimisation utilising highly likely short sub-conformation and addressing combinatorial effect would be very promising for folding prediction.

Defining a particular *shape-pattern* can add an extra layer of sensitivity to the fragment selection process within fragment-based protein prediction software. In this work we have looked at using a HP pattern approach and discovered that a multi-objective optimisation utilising highly likely short sub-conformations and addressing the combinatorial effect would be needed if a HP *shape-pattern* is to be applied to the prediction process based on the variation of the validity in Table 1.

### Acknowledgements

This research is partly sponsored by ARC (Australian Research Council) grant no DP0557303.

### References

- [1] Baker, D., Prediction and design of macromolecular structures and interactions, *Philosophical Transactions of the Royal Society B*, 361:459–463, 2006.
- [2] Hoque, T., Chetty M., and Dooley, L., A new guided genetic algorithm for 2D hydrophobic-hydrophilic model to predict protein folding, *Proc. IEEE Congress on Evolutionary Computation*, 259-266, 2005.
- [3] Zhang, Y., and Skolnick J., Automated structure prediction of weakly homologous proteins on a genomic scale, *PNAS*, 101(20):7594–7599, 2004.