Modelling Cotranslational Protein Folding

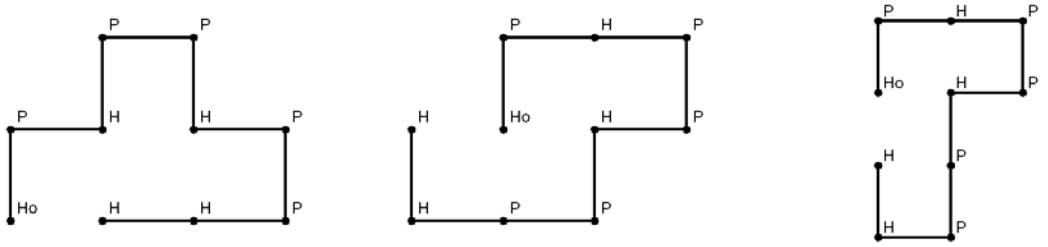
Accurate prediction of the three-dimensional structure of a protein remains a considerable challenge. It is acknowledged in the literature [1] that proteins can fold rapidly as they are being translated, so called "cotranslational folding". Yet state-of-the-art protein folding prediction methods do not use this sequential aspect; in the latest CASP meeting (December 2004) none of the chosen methods exploited the sequential nature of folding. The purpose of this paper is to explore the consequences of following a sequential route to the final fold.

"HP lattice models" [2] have proven a useful tool in the first stage of this research. Such models have been successfully used in recent years to explore and model folding principles, predicated on the assumption that protein folding is ruled by hydrophobic collapse. They consider sequences involving only two types of monomer (hydrophobic H and polar P), with monomer positions restricted to either a two- or three-dimensional lattice. Hydrophobic residues placed adjacent to each other in space but not in sequence correspond to a contact and register an energy fall of one unit; all other adjacencies register a zero fall.

We now define sequential folding, in segments of length *s,* of an ordered HP string of length *n,* in which it is possible to overcome energy barriers of height *d*; this aims to capture the essence of cotranslational folding. The first *s* monomers are laid down, locating them in a configuration with minimum energy, at the same time retaining all configurations within energy *d* of this minimum. We then add *s* monomers to all these partial configurations, retaining that with minimum energy and all within energy *d* of the minimum. This procedure is repeated until all monomers are used. A configuration with minimum final energy is termed a "sequential folding".

We have used HP models to investigate the difference between the minimum energy state of a sequential folding and the globally minimum energy state. A difference in these two end states will be found if nature is incapable of pushing a partially formed protein back over a sufficiently high free energy barrier. We have explored the influence on this difference of *n*, *d* and *s*. An example will now be described.

Recent studies [3] provide HP sequences and their known unique global lowest energy conformation. Such a conformation is shown, for the sequence $H_oPHPPHPPHH$, in (a) below. This has four contacts, so energy level of −4. Two minimum energy sequentially generated conformations exist (using $d = 0$ and $s = 1$), each with three contacts, so with minimum energy of −3. These are shown in (b) and (c).



(a) (b) (c)

Conformation (a) can be obtained sequentially if we continue to extrude one monomer at a time (so $s = 1$) but raise the surmountable energy level to one ($d = 1$). We observe that as the length *n* of the chain increases, the surmountable energy barrier and the number of monomers extruded at each step must be increased in order to sequentially reach the globally minimum energy configuration.

[1] Baldwin, T. O., *Nature Cell Biology* 1, E154 - E155 (1999).

[2] Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., Chan, H.S. *Protein Science* 4: 561-602 (1995).

[3] Irback A., Troein C., *Journal of Biological Physics* 28, 1- 15 (2002)