2.3.6 Lattice models

Our discussion of the phases of the designed REM was still largely based on elaborations of thermodynamic arguments. One difficulty is that within the REM the connection between different polymer structures, and the kinetic moves from one structure to another, are not specified. It is precisely the energy barriers encountered along such moves that determine the time scale for reaching equilibrium. These issues can be addressed by adding dynamical rules to the simple lattice models introduced earlier. The rules have to be local (such as pivoting a bond), but sufficient to allow the polymer to arrive from any state to another. [5] In a Monte Carlo (MC) procedure, moves are attempted randomly, and accepted/rejected according to the Metropolis algorithm: If the energy change ΔE is negative the move is always accepted. However, moves that increase energy ($\Delta E > 0$) are still sometimes accepted with probability governed by the Boltzmann weight$e^{-\beta \Delta E}$. While the MC procedure is a far from realistic description of protein dynamics, it simple lattice model does face the issues of speed/stability, and has to overcome the *Levinthal paradox*.

In the 1990s many such simulations were performed, typically on a polymer with 27 sites, which can fold to a variety of compact configurations on a 3x3x3 cubic lattice. While the low energy states were compact the starting (unfolded) configuration could be chosen amongst the much larger number of swollen self-avoiding walks on the square lattice. Each of the 27 sites is labelled with an "amino-acid" a_i, while the energy of a configuration is $E(C) = \sum_{i<j} U(a_i, a_j)\Delta_{ij}(C)$, where $\Delta_{ij}(C)$ is *adjacency matrix* for configuration C, whose elements are 1 or 0 depending on whether or not the non-polymeric pair (ij) is adjacent in configuration C, and $U(a_i, a_j)$ is the matrix of interactions between amino-acids. A random assignment of amino-acids to the sites of the polymers typically does not yield a foldable state; the corresponding polymer at low temperatures usually gets stuck in a swollen state at low temperature. One question is thus how to find a sequence that is foldable.

It is in fact possible to devise a different MC procedure that mimics 'evolution' to design a sequence that has low energy in a specified configuration. For a target "native" structure, characterized by an adjacency matrix $\delta_{ij}(n)$, the goal is to minimize $E_n = \sum U(a_i, a_j)\Delta_{ij}(n)$, over the set of all possible adjacency matrices. More precisely, one needs to minimize $\mathcal{Z}_n \equiv (E_n - E_{ave})/\Sigma$, where E_{ave} is an average energy of all structures, and Σ is the corresponding variance, for a given sequence $\{a_i\}$. In principle, $\mathcal{Z}_n$ has to be obtained by comparing all competing structures with the same number of bonds as the desired configuration, e.g. all compact structures. This is computationally difficult, and in an approximation analogous to REM, we can set $E_{ave} = N\varepsilon_0$ and $\Sigma = N\sigma^2$, as in Eq.(2.61), related to the mean and variance of $U(a_i, a_j)$, respectively. The "evolution" MC then proceeds as follows: start with a random sequence; attempt a move replacing the amino-acid at a site i from a_i to a_i'. Accept or reject the "mutation" according to the change ΔZ_n, with Metropolis probabilities controlled by a "design" temperature T_{des}. An important issue is that the quantities E_{ave} and Σ are not known, and cannot be enumerated. To simplify, it is typically assumed that contacts are random (given by $\langle \Delta_{ij} \approx n/[N(N-1)/2]$) and selected in uncorrelated fashion from possible

[5] It can be shown that a combination (i) *corner flip*, (ii) *crankshaft*, and (iii) *tail flip* moves are sufficient to sample all configurations on a cubic lattice.

interaction pairs, leading to $\Sigma^2 \approx n\langle U(a_i, a_j)^2 \rangle_c$.

A polymer designed by the above procedure, starting from a swollen initial state, does not fold at low temperatures, and typically is again caught in a trap. The same polymer simulated at intermediate temperatures may (or may not) fold. Thus, it is empirically possible to construct foldable sequences in the lattice model that to overcome the Levinthal paradox.

The average folding time obtained in such simulations has a non-monotonic dependence on temperature T; at high temperatures there are too many competing states, while at low temperatures the system is easily caught in a trap. A theoretical model for folding time, which reproduced these trends can be constructed, and is presented in the problem set.