Frontiers in Physics

27 - 29, Sept. 2012

Self Avoiding Growth Walks and Protein Folding

K P N Murthy[†], K Manasa[⋆], and K V K Srinath[†]

[†]School of Physics, [⋆] School of Life Sciences
University of Hyderabad

September 29, 2012

acknowledgement and warning

acknowledgement:

- Thanks to Suneel Singh, V S Ashoka, S V S Nageswara Rao, and Soma Sanyal for the invitation
- Manasa and Srinath acknowledge with thanks the summer fellowship awarded in the year 2012, by the School of Physics, for carrying out this project;
- Monte Carlo simulations were carried out in the Centre for Modeling Simulation and Design(CMSD)

acknowledgement and warning

warning :

- I am going to talk about the following issues :
  - Protein folding
  - Levinthal paradox
  - non-bonded nearest neighbour contact pair
  - athermal to thermal random walk
  - Kinetic Walk - walk that grows faster than it could relax
  - irreversible growth and linear homo/hetero polymers
  - Interacting Growth Walk (IGW)
  - Protein folding - some results

Protein Folding

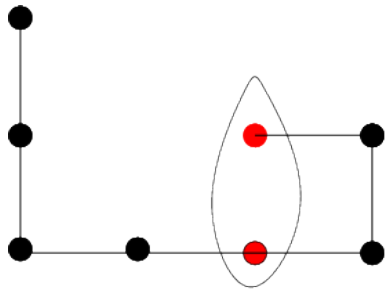
- Protein : non-branching hetero polymer
- monomers are from amongst twenty amino acids
- biological function : intimately related to its unique (?) and thermodynamically stable (meta stable ?) conformation
- A Challenging Problem in biophysics : Levinthal's paradox (1969)
  C Levinthal, "How to fold graciously" Conf. Illinois (1969)
  - a thought experiment
  - astronomical number of possible conformations : order of 3^{300}
  - sequential sampling : requires time, longer than age of the universe to fold to its correct native conformation, even if conformations are sampled at rapid (nanosecond or picosecond) rates.
  - "paradox" : proteins fold spontaneously on a millisecond and often microsecond time scale.
  - This paradox is central to computational approaches to protein structure prediction.

Attempt to resolve Levinthal paradox

- fold step-by-step by considering kinetic growth models - lattice or off-lattice
- speed up the folding by rapidly forming local interaction which in turn determine the next step in folding process
- *i.e.* implement local "equilibration"; do not insisit on global equilibrium
- decide local moves on the basis of local partition function - on the basis of possible energy and entropy changes
- Interacting Growth Walk (IGW) is a kinteic walk that attempts this, within the frame work of lattice models

- Self avoiding walks (SAW) are most suitable for modeling polymer conformations
- SAW is a random walk that does not intersect itself - excluded volume or hard core repulsion
- self avoidance is best modeled by considering walk on a lattice - the random walk can not visit a site it has already visited
- algorithms to generate SAW : blind ant, myopic ant, Boltzmann ant, Kinetic Growth Walks(KGW), Interacting Self avoiding walks *etc*
- We shall consider only Interacting Growth Walks (IGW)
- self avoiding walks are athermal objects - can not define temperature
- define energy - through non-bonded nearest neighbour contact
- athermal to thermal

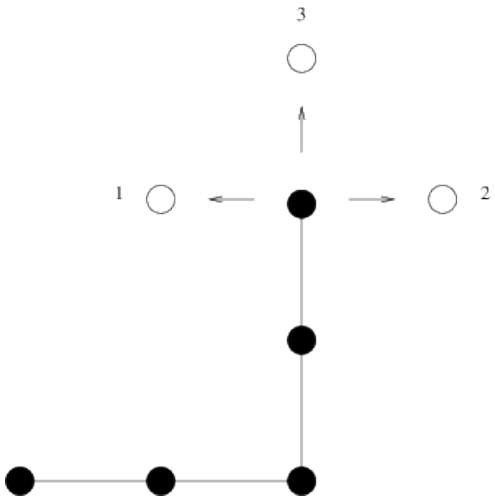
non-bonded Nearest Neighbour (nbNN) contact pair



- The monomers marked in red constitute a non-bonded nearest neighbour pair.
- They occupy nearest neighbour sites on the lattice but are not connected by a bond.
- Each nbNN contact pair carries an energy ϵ.
- The athermal SAW becomes thermal, when we define such contact interaction.

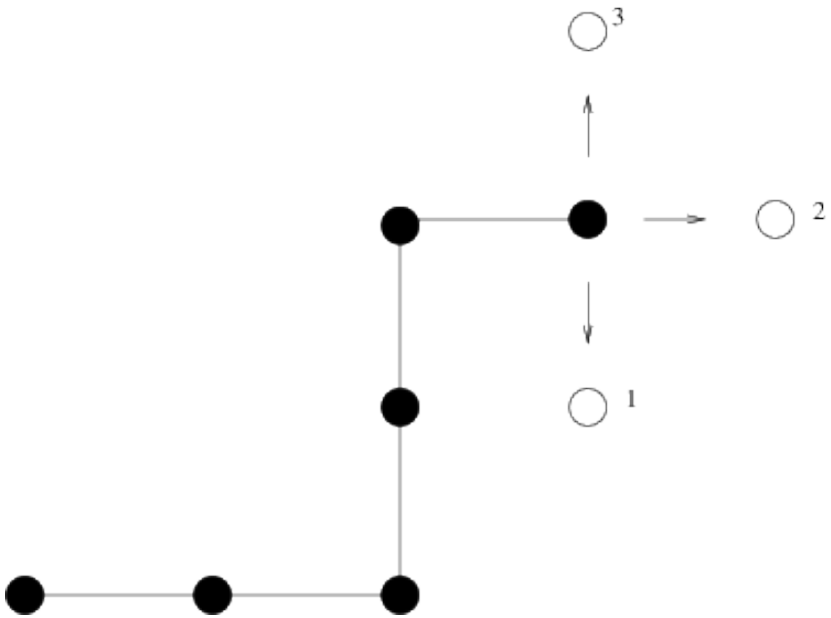
- n : number of non-bonded nearest neighbour contacts in a polymer conformation
- energy $= n \times \epsilon$: where ϵ is the energy per contact.
- ϵ is negative for attractive interaction and positive for repulsive interaction
- each possible step is given Boltzman weight on the basis of change in energy
- a step is randomly selected on the basis Boltzmann weights
- temperature is treated purely a tuning parameter for optimal folding - has no physical significance

IGW algorithm for a linear homopolymer



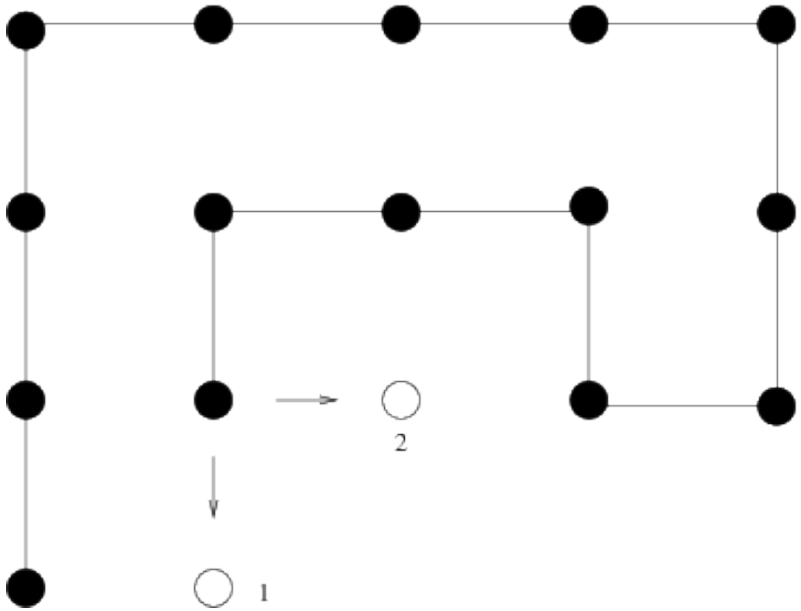
- None of the three moves lead to new nbNN contacts
- hence all the three moves are equally probable
- select one of them randomly

IGW algorithm for a linear homopolymer



- Move-1 leads to one new nbNN contact. Moves - 2 and 3 do not lead to new nbNN contacts
- $Q = \exp(-\beta\epsilon) + 1 + 1$
- $P(1) = \frac{1}{Q}\exp(-\beta\epsilon)$. $P(2) = P(3) = \frac{1}{Q}$

IGW algorithm for a linear homopolymer



- Move-1 leads to one new nbNNcontact.
- Move-2 leads to two new nbNN contacts
- $Q = \exp(-\beta\epsilon) + \exp(-2\beta\ \epsilon)$
- $P(1) = \frac{\exp(-\beta\epsilon)}{Q}$; $P(2) = \frac{\exp(-2\beta\epsilon)}{Q}$

IGW algorithm for a hetero polymer

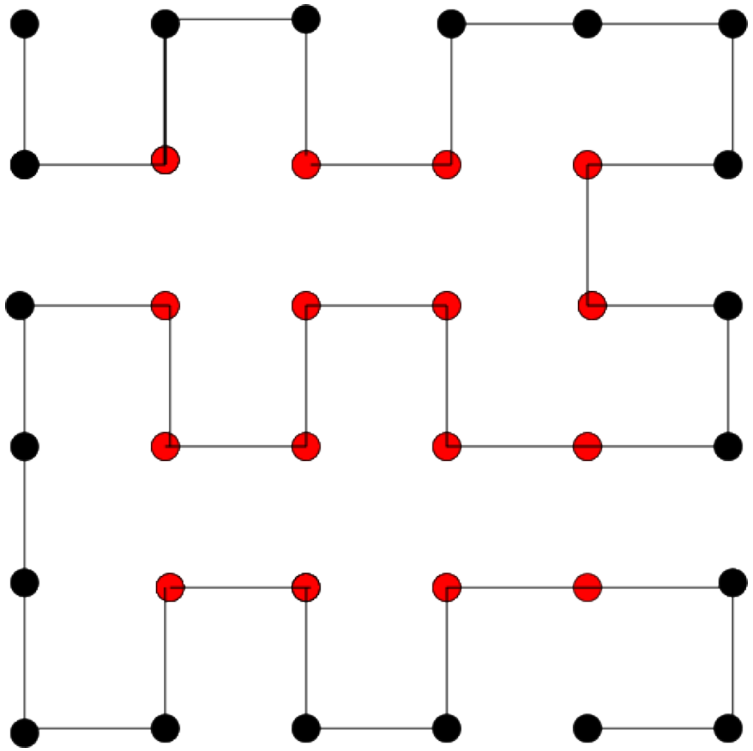
- I have illustrated IGW growth rules for a linear homo polymer
- a protein is a linear hetero polymer
- we coarse grain the amino acids and put them into two categories Hydrophoebic H and Polar P.
- $\epsilon_{HH}, \epsilon_{PP}, \epsilon_{HP}$ denote the energy associated with nbNN Contact made by H, H, P, P and H, P respectively.
- H and H would like to come close for expelling water from the interior Hence we take $\epsilon_{HH} = \epsilon < 0$.
- P and P or H and P do not have any such preference. We take $\epsilon_{PP} = \epsilon_{HP} = 0$
- Carry out IGW growth exactly the way described earlier with appropriate for nbNN contact energies.

Results

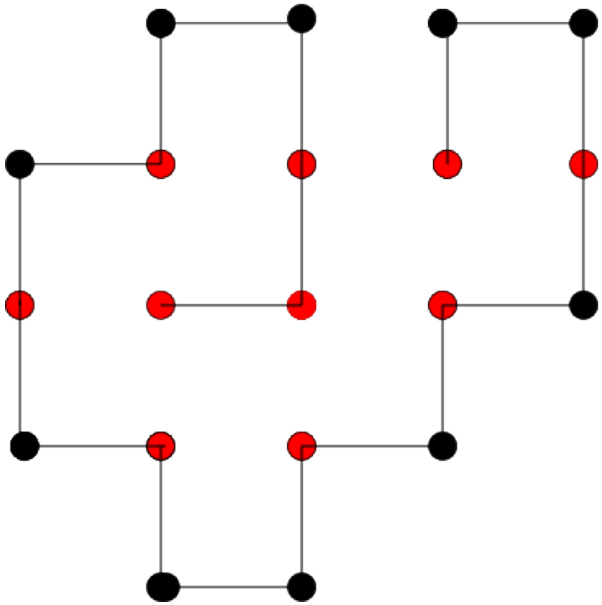
Sequences considered are

- $H-H-H-P-P-H-P-H-P-H-P-P-H-P-H-P-H-P-P-H$
- $P-P-P-H-H-P-P-H-H-P-P-P-P-P-P-H-H-H-H-H-H-H-P-P-H-H-P-P-P-P-H-H-P-P-H-P-P$

Results



Results



Results

- Ten benchmark sequences are given in K Yue, Proc. Natl. Acad. Sci. USA **92**, 325 (1995) have been taken up for folding
- All the sequences have 48 Monomers
- We also present the results obtained by Yue *et al* and
- U Bestola, H Fruenken, E Gerstner, P Grassberger, and W Nadler Struc. Func. Genetics **32**, 52 (1998)

Results on Benchmark sequences : 1 - 5

Sequence	$-E_{min}$(Reported)	$-E_{min}$ (Ours)
1	31,32	31
2	32,34	32
3	31,34	32
4	30,33	30
5	30,32	30

Results on benchmark sequences : 6 - 10

Sequence	$-E_{min}$(Reported)	$-E_{min}$(Ours)
6	30,32	30
7	31,32	31
8	31,31	30
9	31,34	31
10	33,33	31

L Toma and S Toma, Protein Sci. **5**, 147 (1996)

Sequence	$-E_{min}$(Reported)	$-E_{min}$ (Ours)
Toma and Toma - 1	34	33
Toma and Toma - 2	42	41

- Interacting Growth Walks in three dimension
- folding of Benchmark protein sequences employing IGW algorithm
- Study of performance of algorithm for various values of β
- and

- **Thanks**