Protein folding and aggregation

- Generalities
- Universality *vs* natural selection
  the case of random hetero-polymers
- Folding *vs* aggregation
  the case of the Prion protein (PrP)
  and the role of Cu
- XAS (NMR, EPR) experiments
  data analysis and EXAFS theory
- QM calculations
  DFT and Car-Parrinello dynamics

Generalities

Protein is a complex (and complicated) system

- Many degrees of freedom

  protein: ~ 300 a.a.'s x 10 atoms = ~ 3000 atoms
  solvent: ~ 1000 atoms

  → 3 to 4 times more "active" electrons

- Large range of folding times

  from *μsec's* to *sec's*

  the Levinthal's paradox { too fast for an exhaustive search / too slow for a straight descent to absolute minimum

- Interaction is not short-range

  choice of a phenomenologically acceptable potential in MD
  a Q.M. treatment (DFT, Car-Parrinello) is often needed

- Free-energy landscape looks very corrugated

  many hierarchically organized local minima, separated by high barriers

- System is not living at thermodynamic equilibrium

  flux of energy and matter

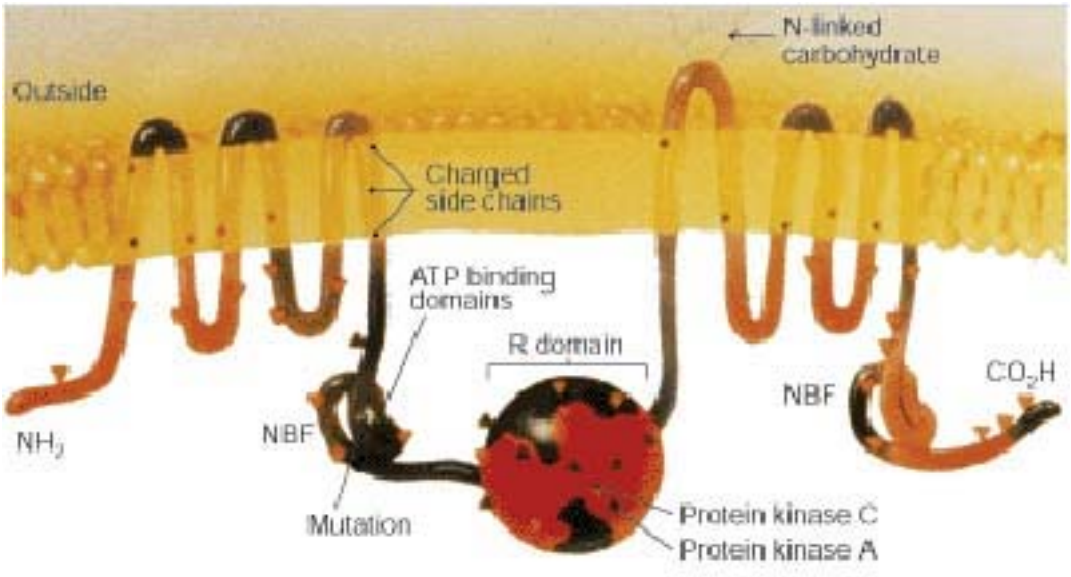
- Even single mutations matter

  though not always

a dramatic case

The CFTR gene is found at the q31.2 locus of chromosome 7, is 230 000 base pairs long, and creates a protein that is 1,480 amino acids long. The most common mutation, ΔF508 is a deletion (Δ) of three nucleotides that results in a loss of the amino acid phenylalanine F at the 508th position on the protein. This mutation accounts for two-thirds of CF cases worldwide and 90 percent of cases in the United States, however, there are over 1,400 other mutations that can produce CF.

There are several mechanisms by which these mutations cause problems with the CFTR protein. ΔF508, for instance, creates a protein that does not fold normally and is degraded by the cell. Several mutations, which are common in the Ashkenazi Jewish population, result in proteins that are too short because production is ended prematurely. Less common mutations produce proteins that do not use energy normally, do not allow chloride to cross the membrane appropriately, or are degraded at a faster rate than normal. Mutations may also lead to fewer copies of the CFTR protein being produced.



Even a single mutation (deletion) can be fatal

Cystic Fibrosis

The protein cannot be crystallized. No full resolution of the critical a.a. 508 region → simulations?

We expect numerical approaches to be difficult

- Which atoms are going to be bound?

  structure of the potential is not *a priori* known (QM)

- Force computation time grows like NxN

  two-body potential

- The system is very heterogeneous

  the problem is not "embarassingly" parallel

- Dynamics time step is of the order of a *femptosec*

  the system can be followed for very short times

- The system gets easily trapped in metastable states

  the exploration of the system phase-space is far from ergodic

- Energy may not be a good label of the states of the system

  states with largely different 3D-structures can have similar energies
  states with only slightly different 3D-structures can have very different energies

Countless number of approaches

- Geometrical approaches
- Simulated annealing
- Molecular Dynamics
- Monte Carlo simulations
- Simulated tempering and variations thereof
- Multi-canonical simulations
- Effective free-energy profile evaluation
- ...

Different levels of description

- Systems with discretized degrees of freedom
- String of beads
- Detailed atomistic description
  - with effective interaction potentials
  - with *ab initio* potentials
- ...

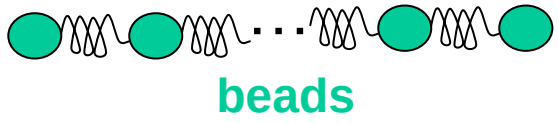
Universality vs natural selection

Iori Marinari
Parisi Struglia

Self-interacting random hetero-polymers

- The complexity of the system is encoded in a certain amount of randomicity of the Hamiltonian

  - $H = \sum_{i=1}^{N} \sum_{i>j} E_{ij}, \qquad N \geq 30$



  - $E_{ij} = k\delta_{i,j+1} r_{ij}^2 + \frac{B}{r_{ij}^{12}} + \frac{\eta_{ij} - A}{r_{ij}^{6}}, \qquad r_{ij}^2 = | \vec{x}_i - \vec{x}_j |^2$

binding repulsive ($B = 2$) it depends on the sign of $\varepsilon - A$

  - η_{ij} uncorrelated random gaussian variables

$$< \eta_{ij} > = 0 \qquad < \eta_{ij}^2 > = \varepsilon$$

- The system is brought to equilibrium at $\beta = 1/k_B T$ under the Boltzmann probability distribution $\propto \exp[-\beta H]$

♦ During the evolution the shape of the chain is continuously monitored and various interesting features are revealed

Coil (open) $\delta >> \rho, \lambda$

Unshaped globule $\delta \geq \lambda$

Frozen well-shaped structures $\delta < \rho, \lambda$

~ folding?

configurations

- $\delta^2_{\alpha\beta} = \frac{1}{N}\sum_{i=1}^{N} | \vec{x}_i^{(\alpha)} - \vec{x}_i^{(\beta)} |^2 \rightarrow$ "distance" between $\{\vec{x}_i^{(\alpha)}\}$ and $\{\vec{x}_i^{(\beta)}\}$

- $\rho = \frac{1}{N_{conf}} \sum_{\alpha} \frac{1}{N} \sum_{i=1}^{N} | \vec{x}_i^{(\alpha)} - < \vec{x}_i^{(\alpha)} > | \rightarrow$ average giration radius

- $\lambda = \frac{1}{N_{conf}} \sum_{\alpha} \frac{1}{N-1} \sum_{i=1}^{N-1} | \vec{x}_i^{(\alpha)} - \vec{x}_{i+1}^{(\alpha)} | \rightarrow$ average link length

I. $\varepsilon = 0$, no randomicity → homo-polymer

- phase transition at $A \approx 2$

coil (open) → **un**-shaped globule (closed)

$P(\delta^2)$ peaked at large δ^2 → small δ^2

II. $\varepsilon \neq 0$, some random interaction → hetero-polymer

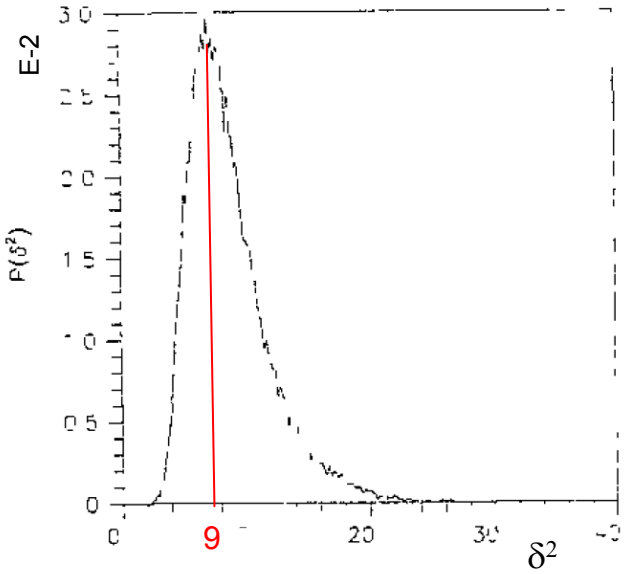
- new phase beyond a critical $\varepsilon_c > A$

well-shaped globule (~ glassy phase in SG ?)

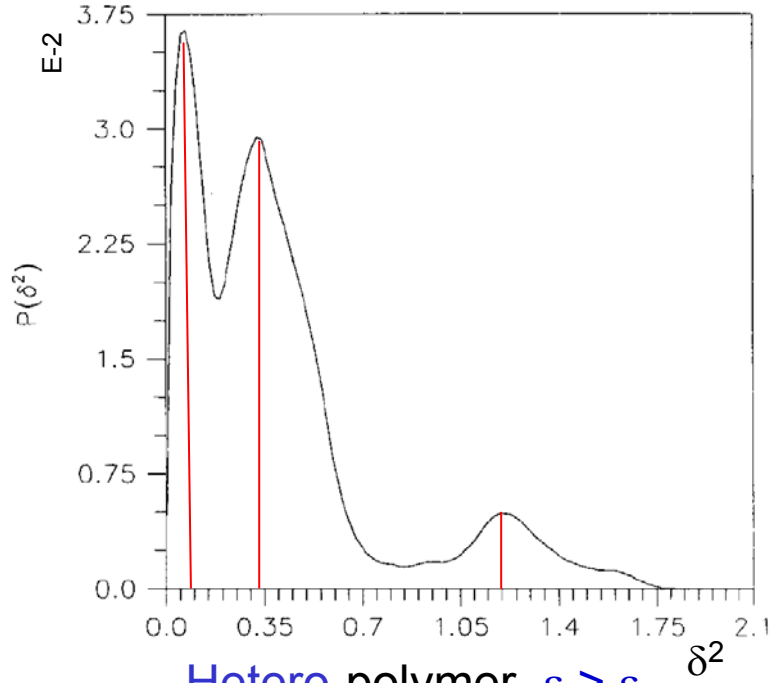
$P(\delta^2)$ is endowed with a lot of structure

Main result →
Sufficiently random hetero-polymers generically fold

Speculation → Perhaps (all the) other a.a. sequences do not fold.
Do they rather aggregate?



Homo-polymer, $\varepsilon = 0$
open phase A = 1.6

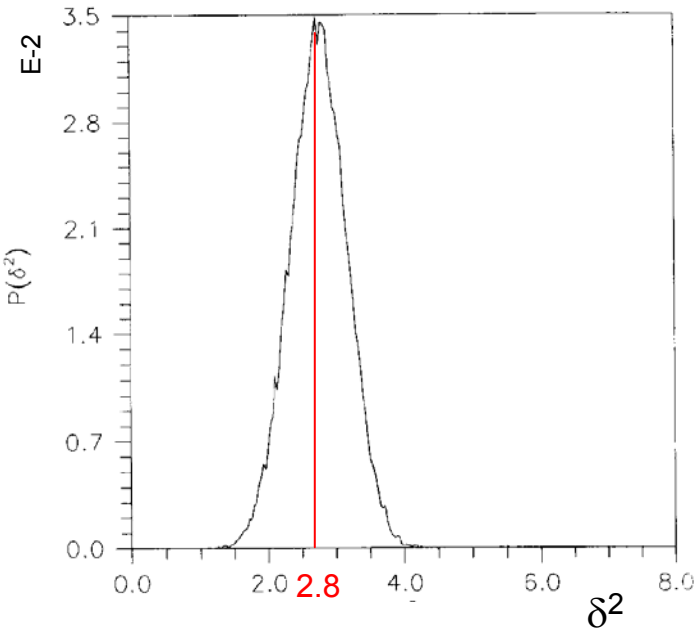


Hetero-polymer, $\varepsilon > \varepsilon_c$
"folded" phase

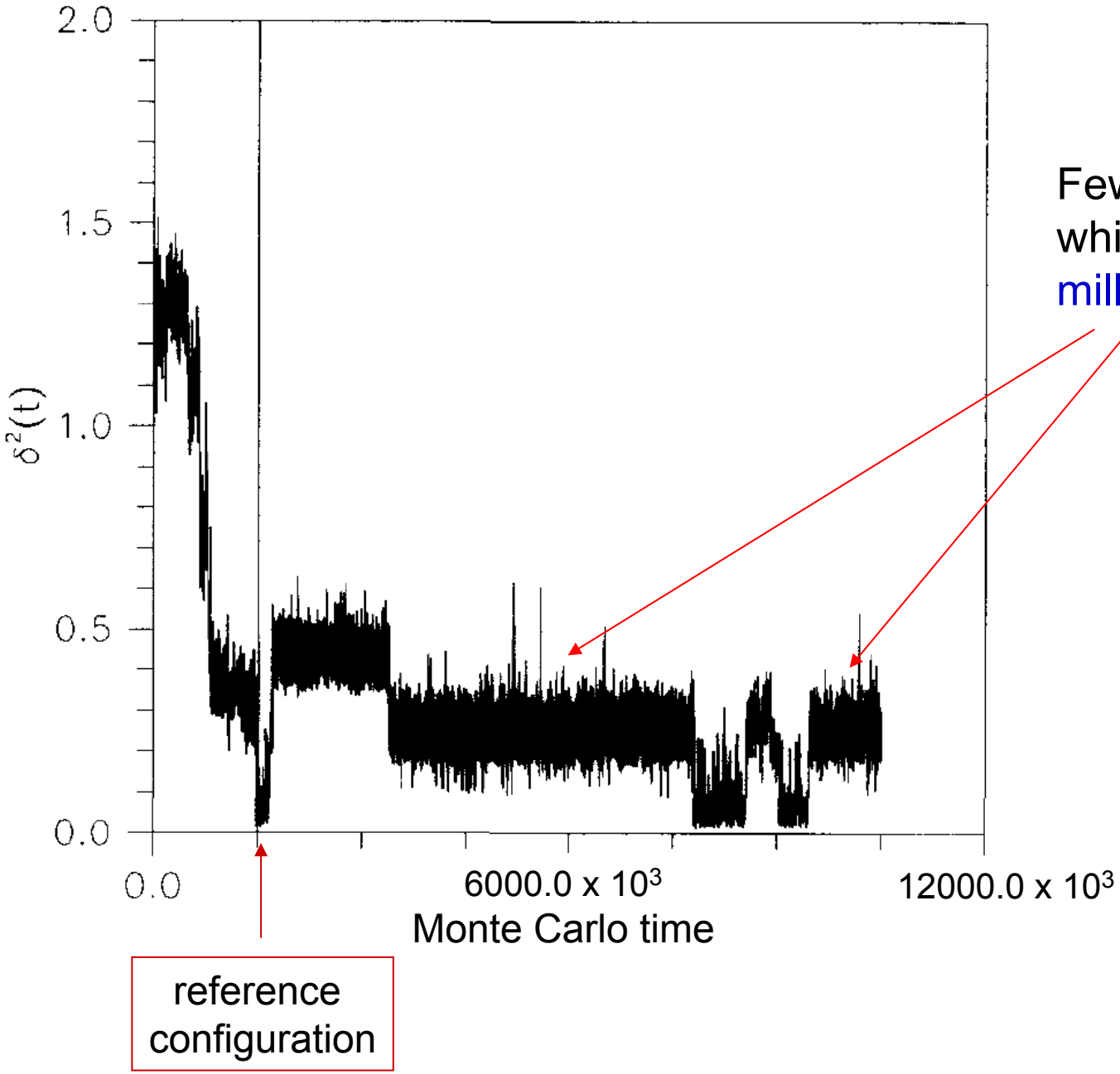
Peaks → macroscopically different folded states

Not δ-functions → many (only) microscopically different states

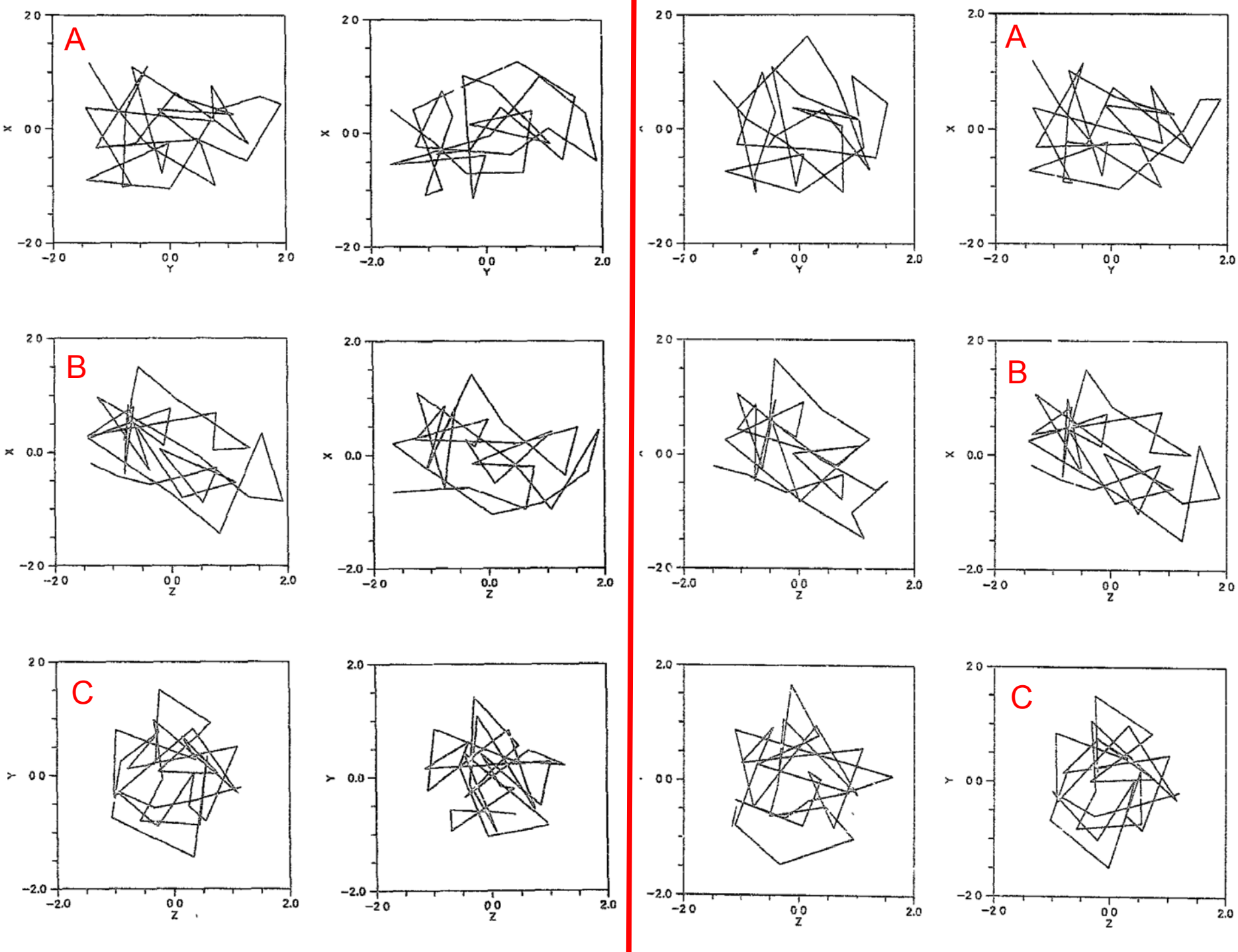
Macrostates are very long living (see next figures)



Homo-polymer, $\varepsilon = 0$
globular phase A = 3.8

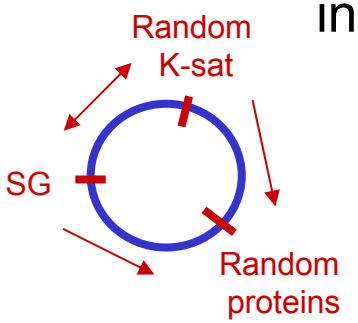


Few metastable states which reappear after millions of MC steps


A
B
C
A
B
C

Comments

- In the "folded" phase the situation displays analogies with what one finds in the glassy phase of SG
  - Many long living, hierarchically organized states at sufficiently large randomicity (frustration)
  - Very long (actually not well defined) correlation times (stretched exponentials: $\propto \exp [- (t/\tau)^{\alpha}]$, $\alpha<1$, "aging")
  - Complexity of protein folding is reflected in the NP-completeness of SG
- Can one make the SG analogy more stringent and useful?
  - Perhaps yes, taking inspiration from results in K-sat problem theory



- Random K-sat problems can be mapped to SG
- Alg's borrowed from SG can help solving Random K-sat problems in polynomial time with probability ~1
- Can a random protein be folded in polynomial time?

- Should we instead move to a more reductionist point of view?

K-sat problems and SG

www.satlib.org

- K-sat problem: M constraints among N boolean variables, $p_1, p_2, \ldots, p_N$
- Constraint: clause among K variables (or their negation, $\neg$)

e.g. $(p_1 \vee \neg p_2) \wedge (p_2 \vee p_3) \wedge (\neg p_1 \vee \neg p_3) \rightarrow$
$[p_1 = t, p_2 = t, p_3 = f]$ or $[p_1 = f, p_2 = f, p_3 = t]$

Conjuntive Normal Form (CNF)

- $K \geq 3 \Rightarrow$ NP-complete problem

K-sat problems	Spin systems
- p_i = true/false	- spin $\Rightarrow \sigma_i$ +1/-1
- clause among a set of p_i	- interaction among a set of spins
- negated / non-negated variables	- coupling J = -1 / +1
- clauses satisfied / violated	- energy = 0 / 1
- # of violated clauses	- total energy H
- 2^N possible ansatz's	- s = 1, 2, …, 2^N possible states

$$P(\sigma,\beta) \propto \exp[-\beta H(\sigma)]$$

- minimal # of violated clauses
- minimum of H → SM at $\beta \rightarrow \infty$ (T = 0)

Random K-sat problem: building the a-th clause, C_a ($a = 1, 2, \ldots, M$)

1) $p_{i1}, p_{i2}, \ldots, p_{iK}$ ($K \geq 3$) are picked up with uniform probability among the N variables $p_1, p_2, \ldots, p_N$

2) variables $p_{i1}, p_{i2}, \ldots, p_{iK}$ are randomly negated

Spin Glass: building the a-th interaction term, E_a ($a = 1, 2, \ldots, M$)

1) $\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{iK}$ ($K \geq 3$) are picked up with uniform probability among the N variables $\sigma_1, \sigma_2, \ldots, \sigma_N$

2) coupling is $J_a = J_{i1} J_{i2} \ldots J_{iK}$ with $J_{ir} = -1$ or $J_{ir} = +1$, according to whether p_{ir} was randomly negated or not.

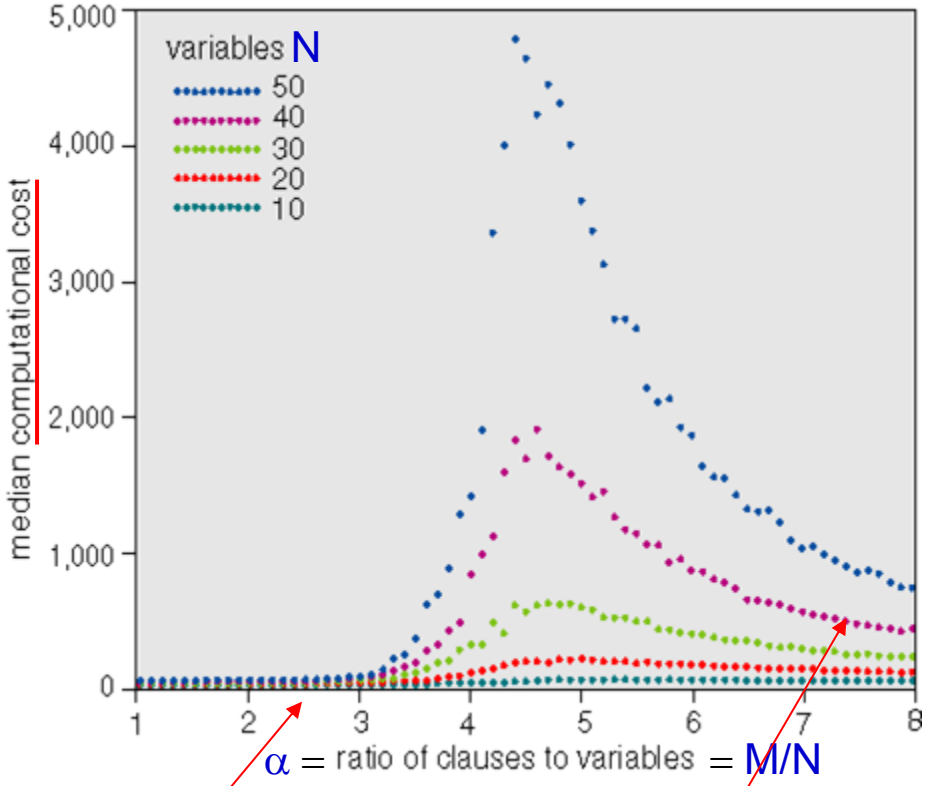
Mézard Monasson
Parisi Zecchina

Some interesting result

1) Emergence of a phase transition as $N \rightarrow \infty$, at a critical value of $\alpha = M/N$

2) Methods developed in SG theory can be used to solve hard K-sat problems (cavity method, decimation alg., …)

3) The average random (not the worst) case can be solved in polynomial time with probability ~ 1

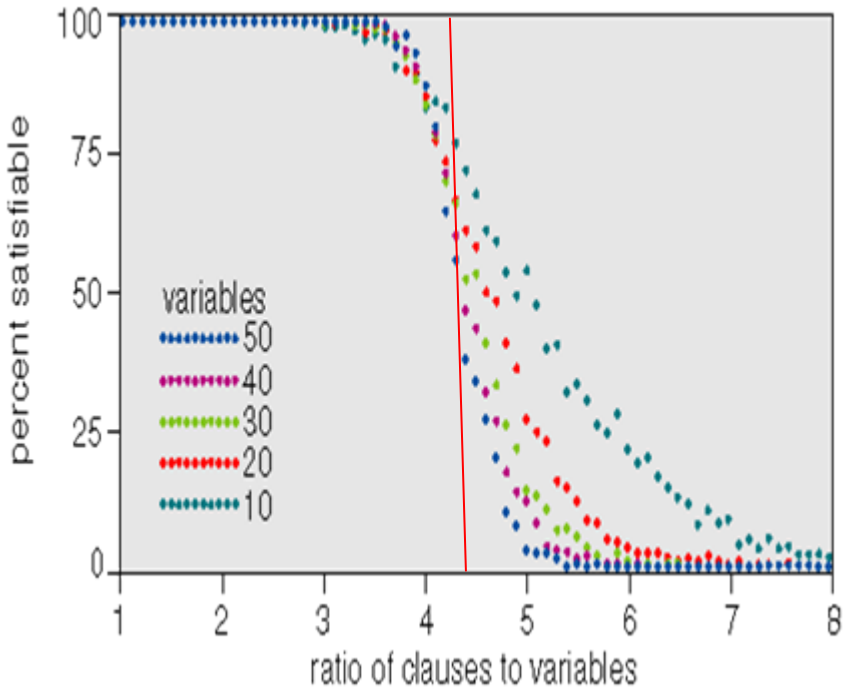


many variables
few clauses
under-constrained
⇓
SATisfiable

not so many variables
compared to # of clauses
over-constrained
⇓
UNSATisfiable

Mitchell Levesque Selman

Hardest problems
around $\alpha_c \approx 4.3$,
where SAT propositions
tend to become UNSAT



Phase transition →
the jump becomes sharper
as N gets larger