Protein folding networks

Erzsébet Ravasz

Zoltán Toroczkai

Gnana Gnanakaran

CNLS/T-13, Los Alamos National Laboratory

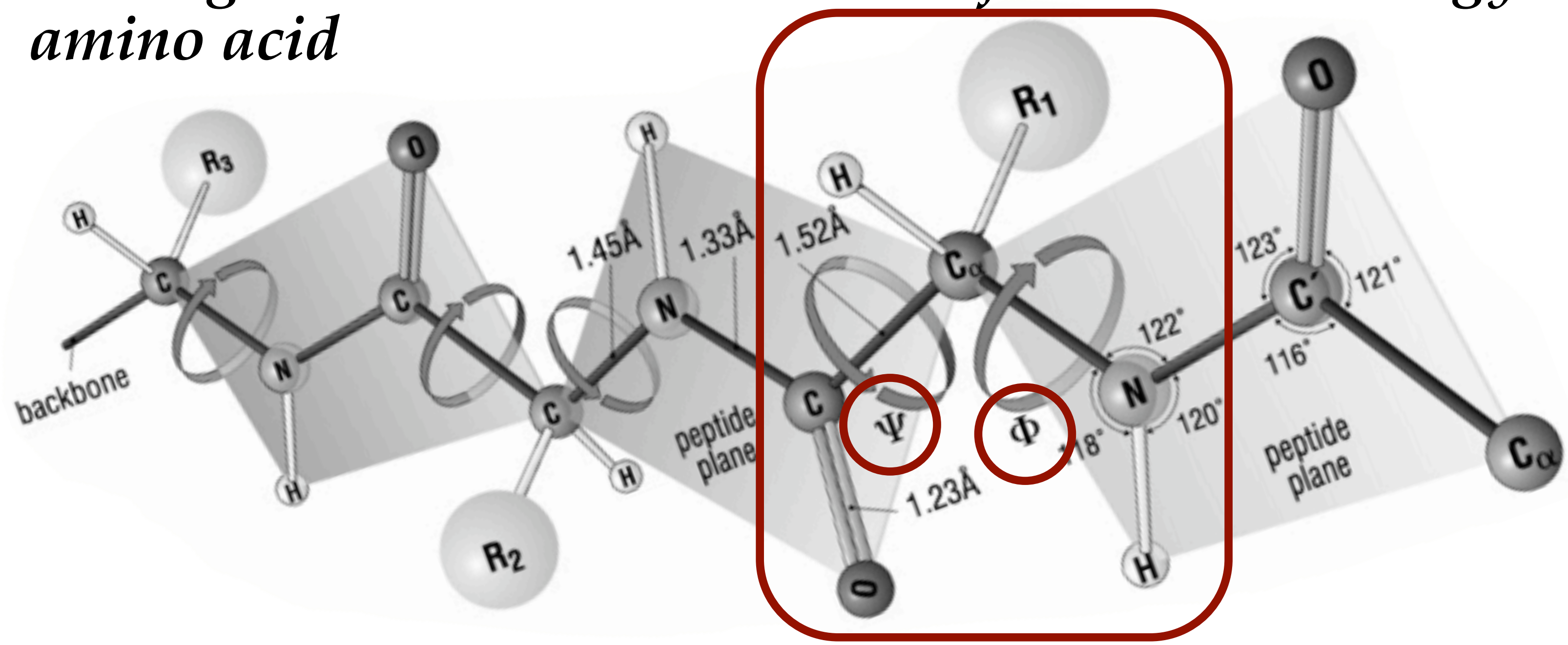
Proteins

- **what are they?**
  - ✓ *chains of amino acids*
  - ✓ *peptide bonds link the backbone*
- **native state**
  - ✓ *unique 3D structure (native physiological conditions)*
  - ✓ *biological function*
  - ✓ *folding times from nanoseconds to minutes*

Motivation

- conformations

  ✓ *2 angles with ~3 local minima of the torsion energy / amino acid*

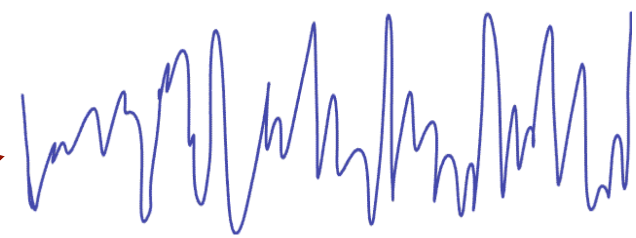


  ✓ *N monomers ⇒ about 10^N different conformations*

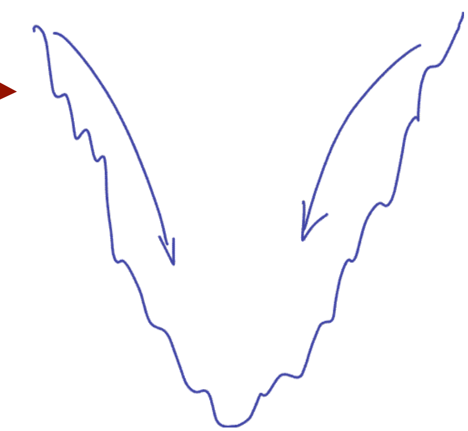
  ✓ *10^{14} conformations / minute (40 monomer protein, experimental data)*

Levinthal type problems

- **Levinthal's paradox, 1969**
  - ✓ ***finding the native state by random sampling is not possible***
  - ✓ ***folding pathways, funnels***



- **An evolutionary paradox?**
  - ✓ ***random amino acids have frustrated landscapes***
  - ✓ ***nature* evolved *funnels***
  - ✓ ***21^{40} possible proteins to choose from***



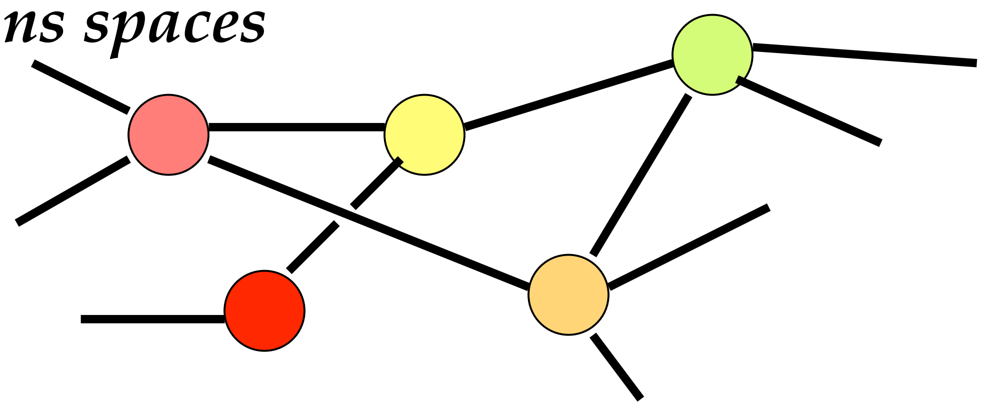
HOW ?

Networks?

Los Alamos
NATIONAL LABORATORY
EST. 1943

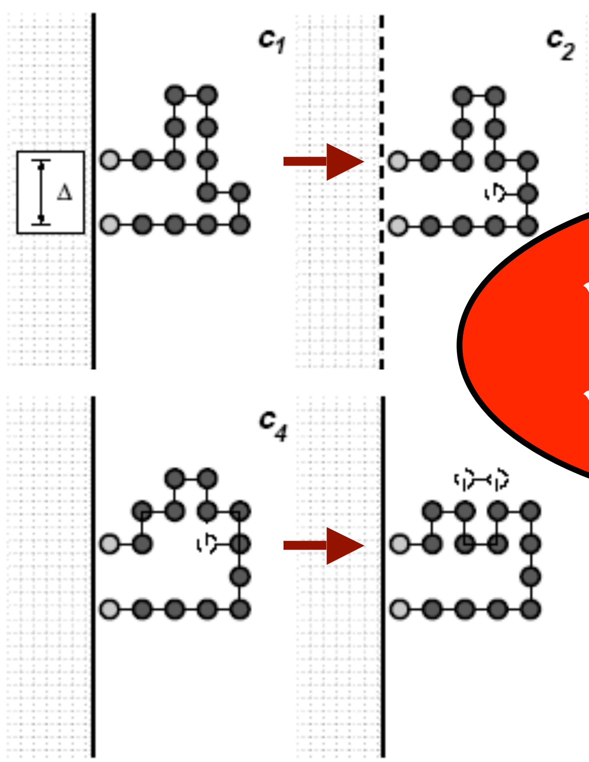
- **Conjecture:**
  - ✓ *generic properties of configuration spaces*
  - ✓ *some properties can explain when/why funnels arise*
  - ✓ *a random non-folding protein is not "far" from a foldable structure*

- **Framework is networks**
  - ✓ *discrete configurations spaces*
  - ✓ *scalable properties*



Configuration networks

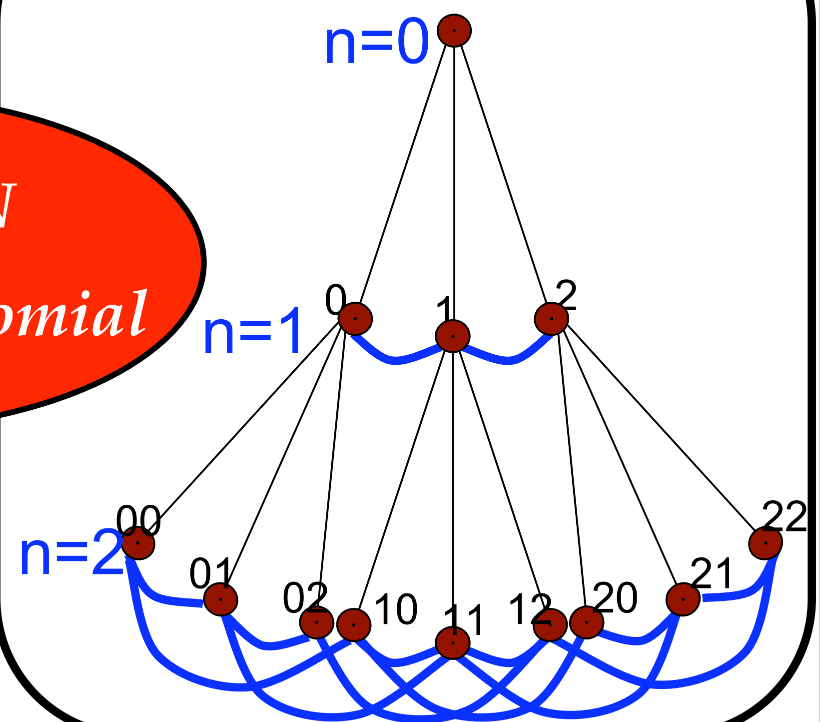
- **Small-world networks and the conformation space of a short lattice polymer chain:** *A. Scala, L.A.N. Amaral and M. Barthélémy, EuroPhys.Lett. 55, 594 (2001)*



✓ $\langle l \rangle \sim \log N$

✓ *P(k) is binomial*

- **Toy robot arm**



How about proteins?

Los Alamos NATIONAL LABORATORY EST.1943

- **The protein Folding Network***: F. Rao, A. Caflisch, JMB, 342, 299 (2004)*

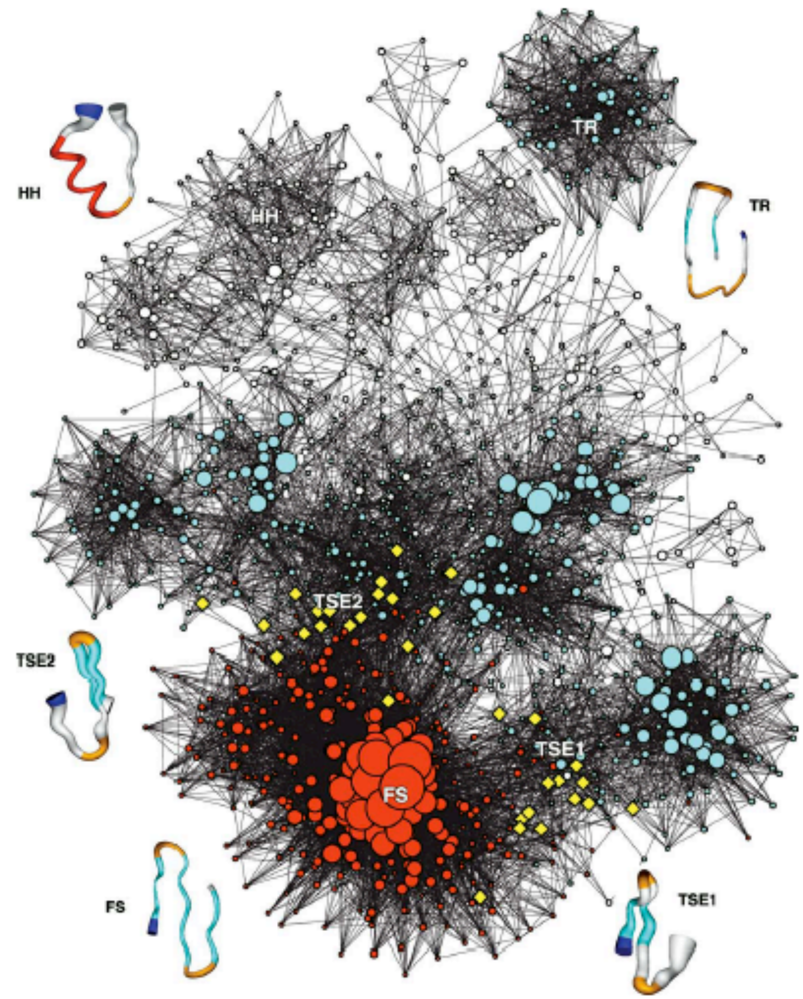
  ✓ **beta3s:** *20 monomers, antiparallel beta sheets*

  ✓ *MD simulation, implicit water*

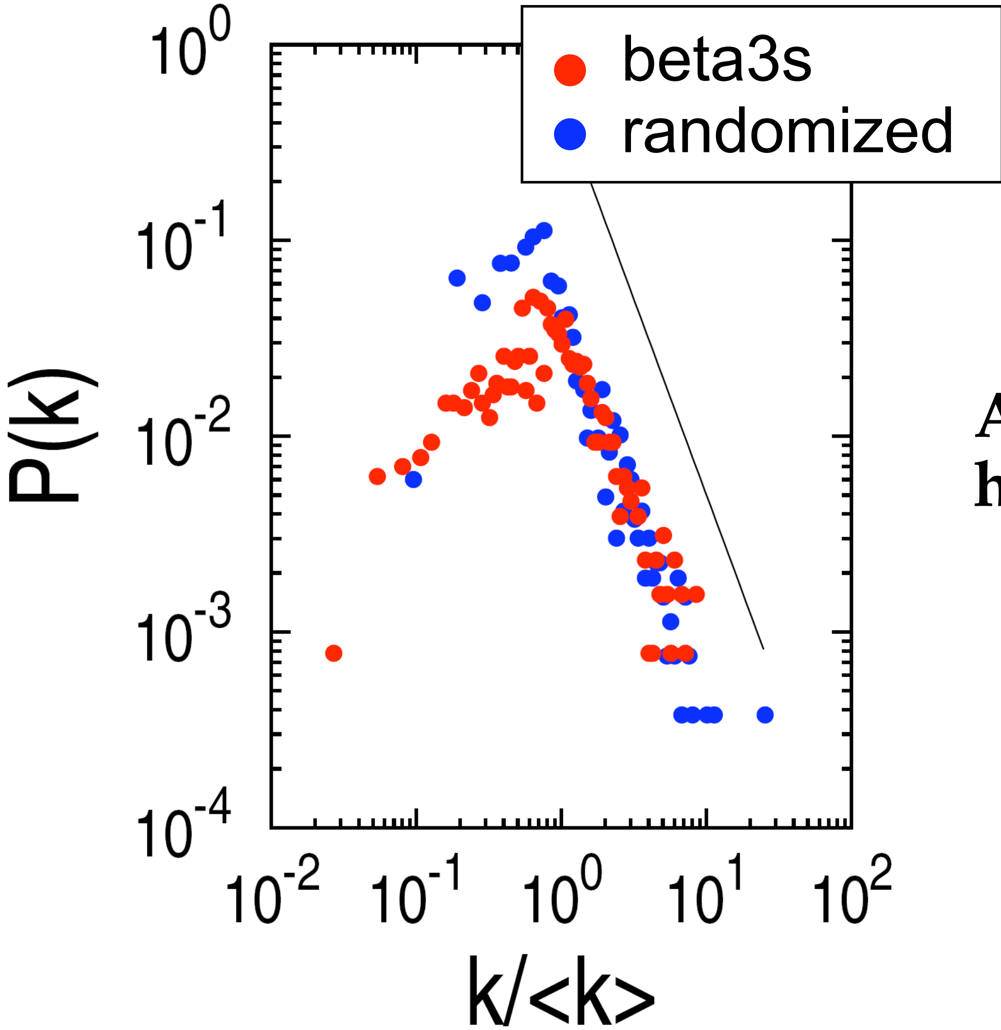
  ✓ *330K, equilibrium folded ↔ random coil*

NODE -- 8 letters / AA
(local secondary struct)

LINK -- 2ps transition



A Scale-free network?!?



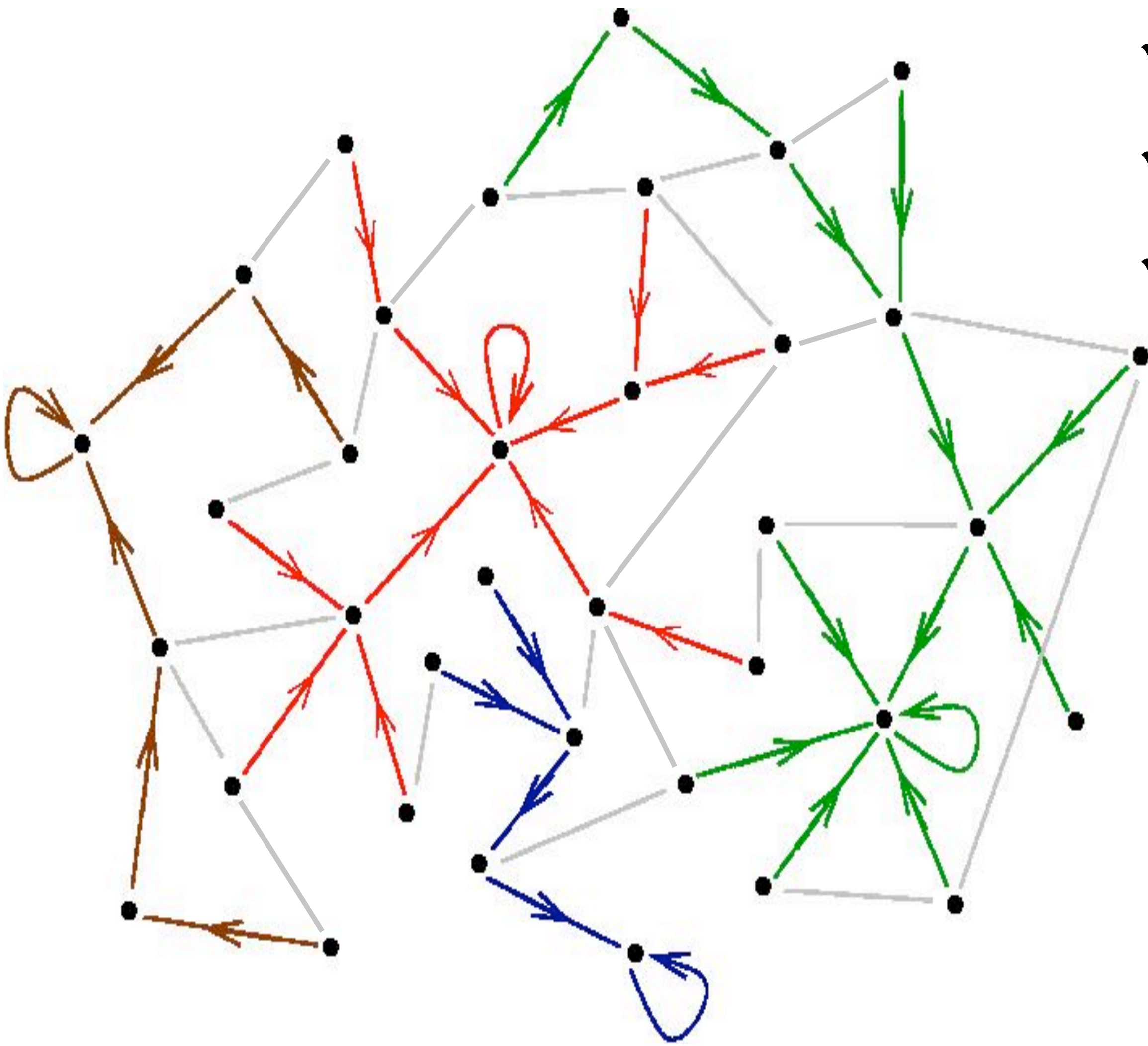
$\gamma = -2$

Are configuration spaces homogeneous or not?

✓ *MD walk is NOT random*

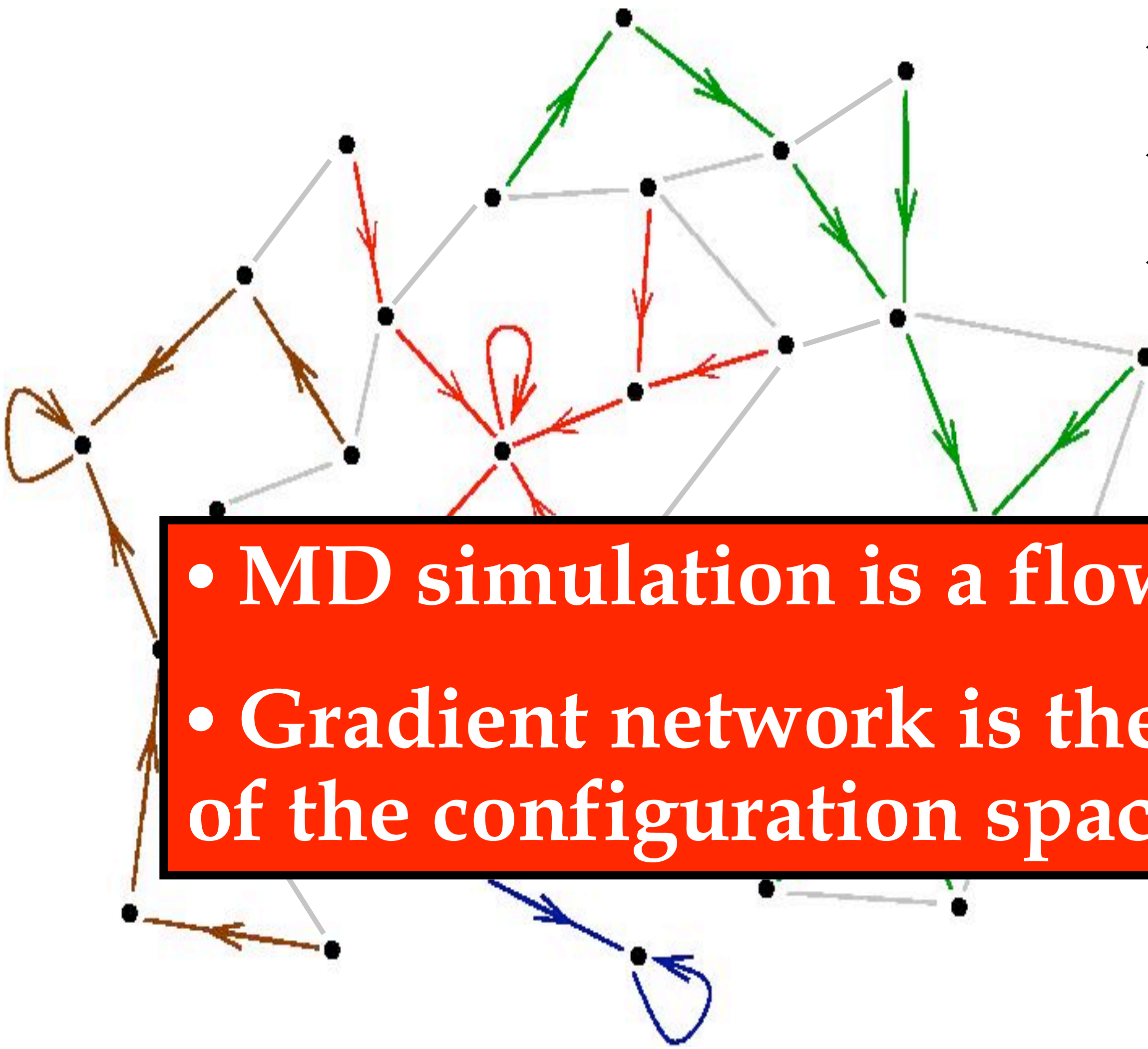
✓ *energy of configurations is a key player*

Gradient flow networks



- ✓ substrate network
- ✓ scalar on nodes
- ✓ gradient flow graph

Gradient flow networks



✓ substrate network

✓ scalar on nodes

✓ gradient flow graph

• MD simulation is a flow network

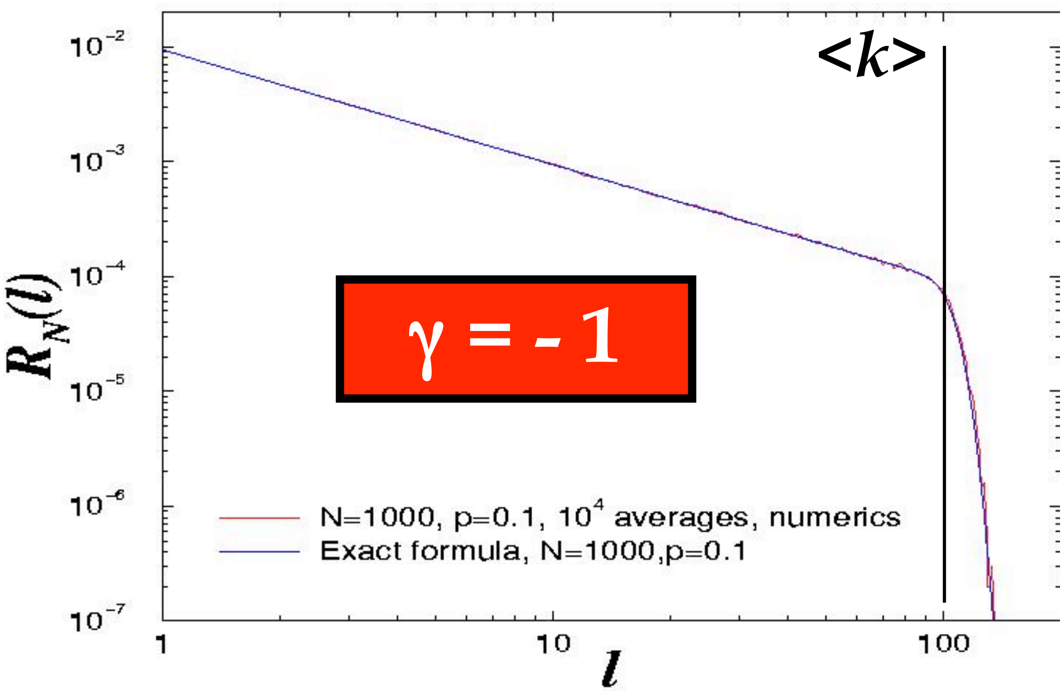
• Gradient network is the backbone of the configuration space at T=0

Why scale-free?

✓ Erdős-Rényi substrate network

✓ i.i.d scalars on nodes

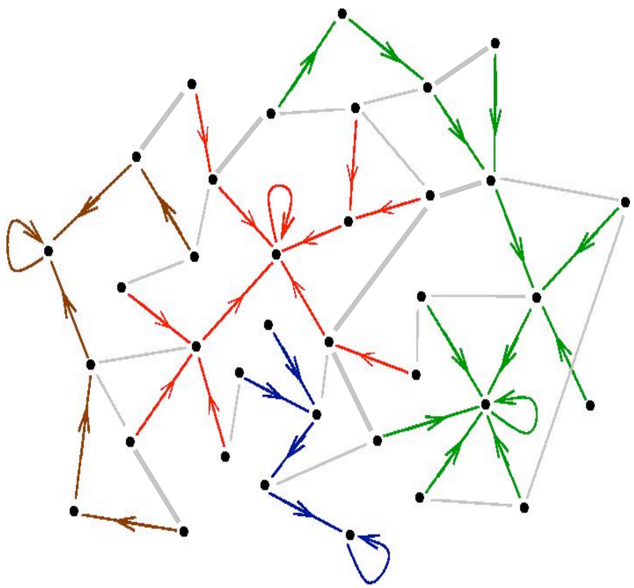
$\Rightarrow$ gradient network is a scale-free tree

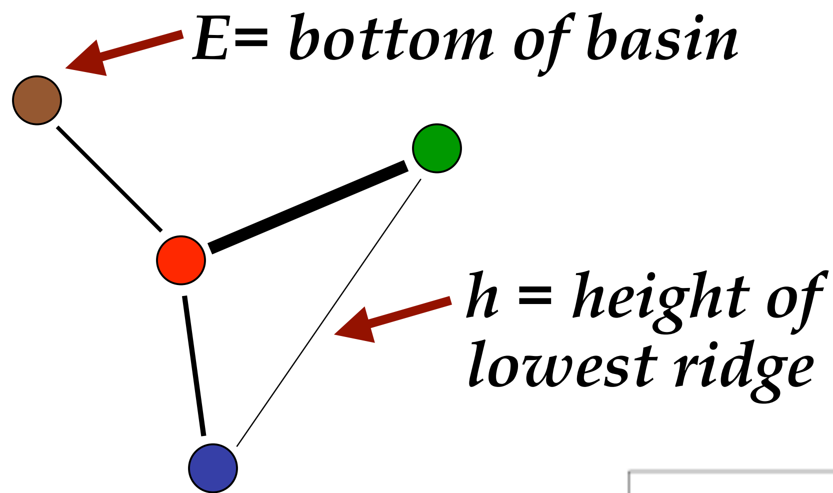


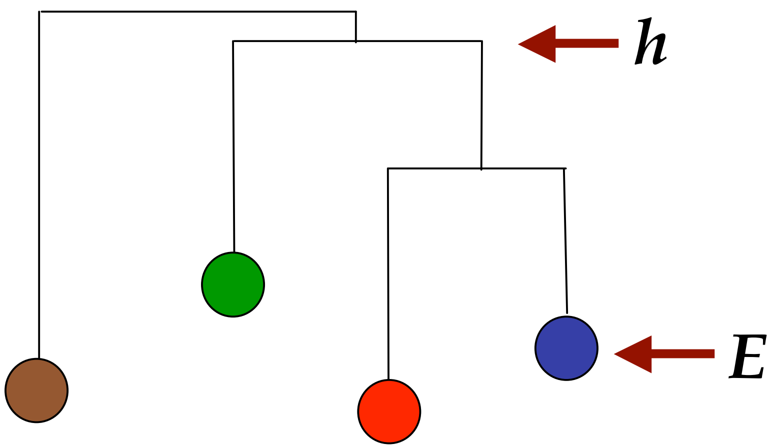
So, how do we get γ = - 2 ?

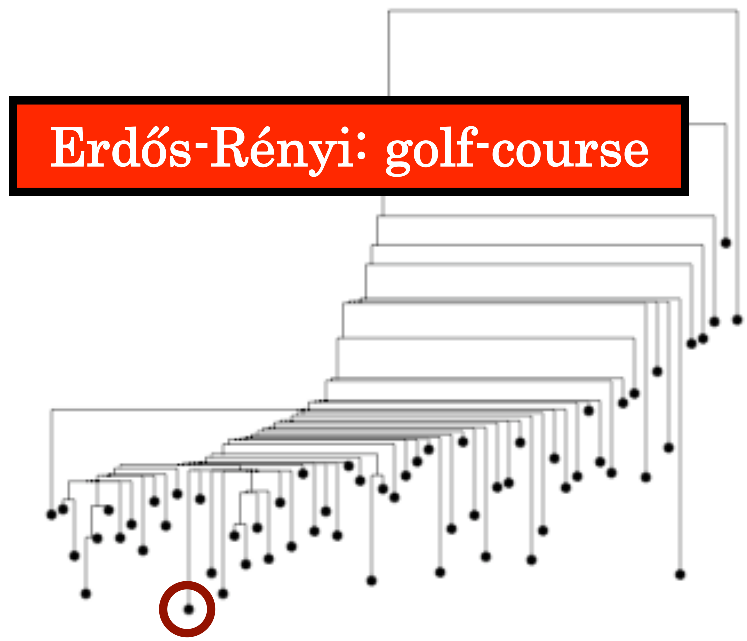
And funnels ?

Energy landscape trees







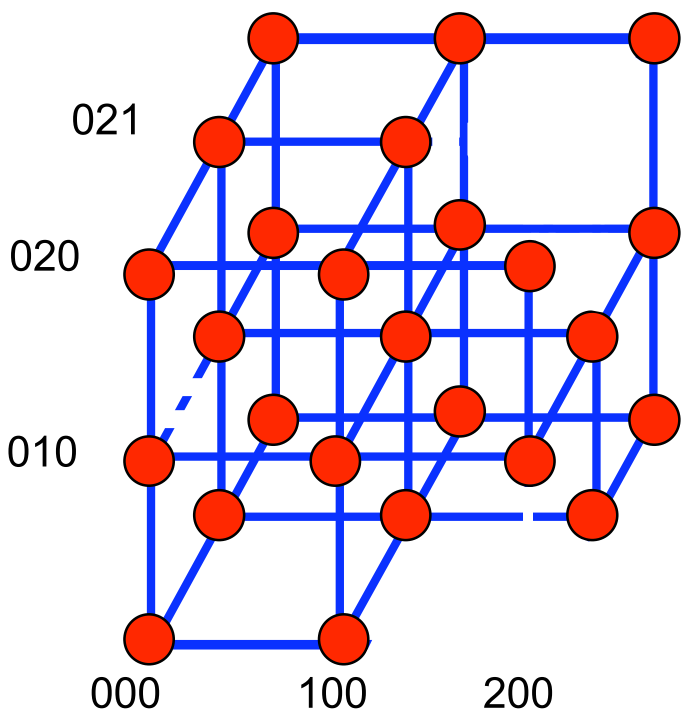


Minimal network model

Los Alamos NATIONAL LABORATORY EST.1943

✓ $\langle l \rangle \sim \log N$

✓ *P(k) is binomial*



- **Lessons from the robot arm:**
  - ✓ *nD hypercube!*
  - ✓ *small world BUT no long range links:*

$$d \sim n \text{ but } N = 3^n \Rightarrow d \sim \log N$$

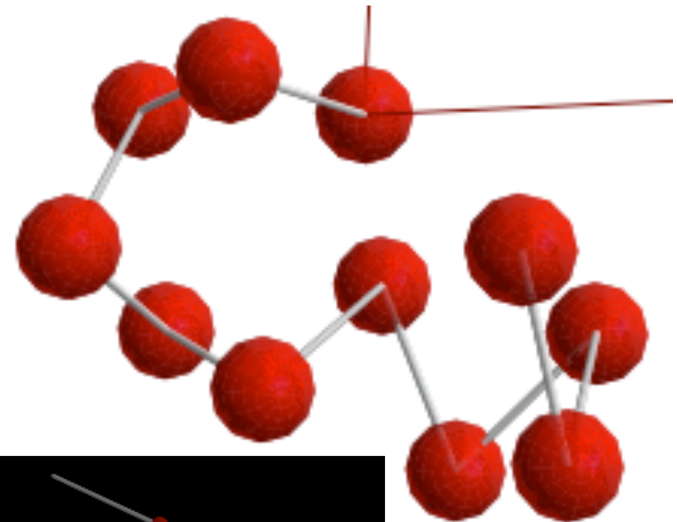
- **Steric constraints?**
  - ✓ *missing nodes*
  - ✓ *missing links*

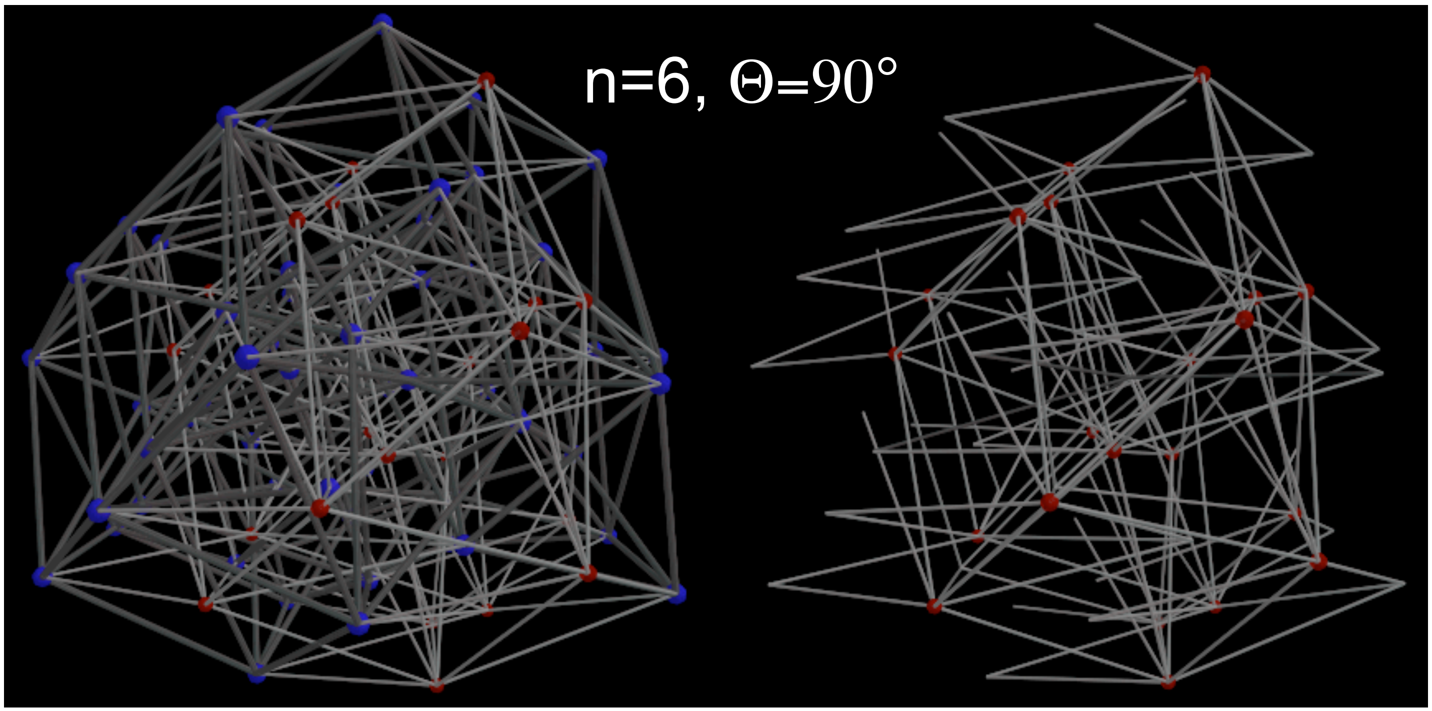
We can model this!

A Bead-Chain Model

Los Alamos
NATIONAL LABORATORY
EST.1943

- The BC robot arm model
  - ✓ *beads and rods in 3D*
  - ✓ *rod-rod angle Θ*
  - ✓ *3 positions around axis*



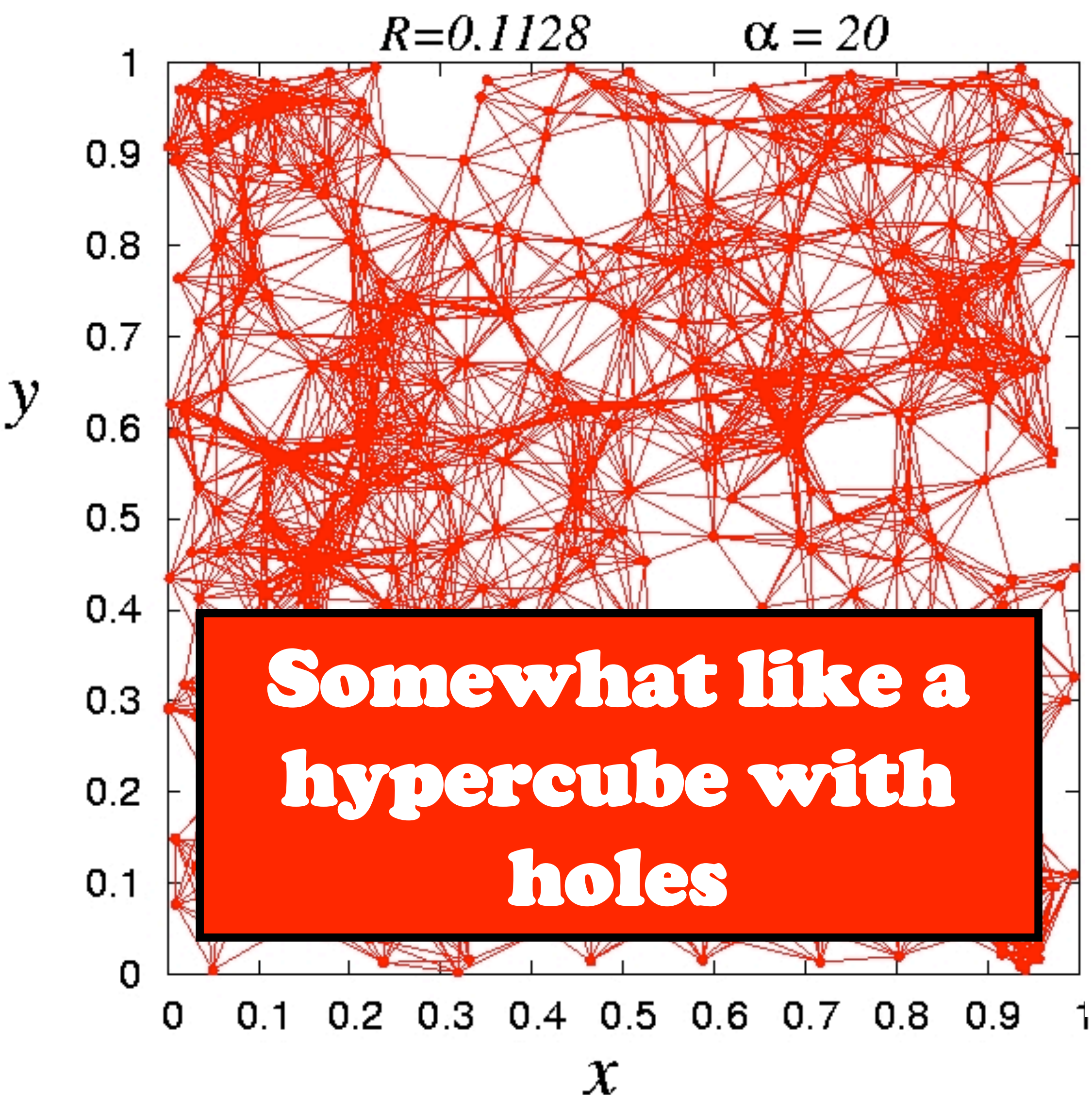


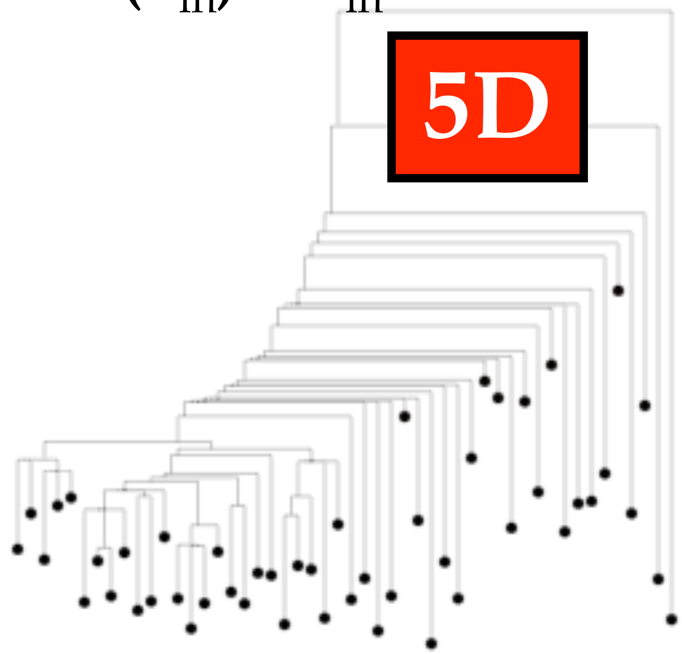
Random geometric nets

- ✓ homogeneous
- ✓ no shortcuts:

  $\langle l \rangle \sim N^{1/d}$

- ✓ $P(k_{in}) \sim k_{in}^{-1}$





N=10000, <k>=200

Where is the energy?

Los Alamos NATIONAL LABORATORY EST.1943

- **Until now:**

  ✓ *energies were independently drawn from the same distribution (any...)*

  ✓ *homogeneity in k*

  ⇒ **slope -1**

  ⇒ **NO funnel**

- **Real systems:**

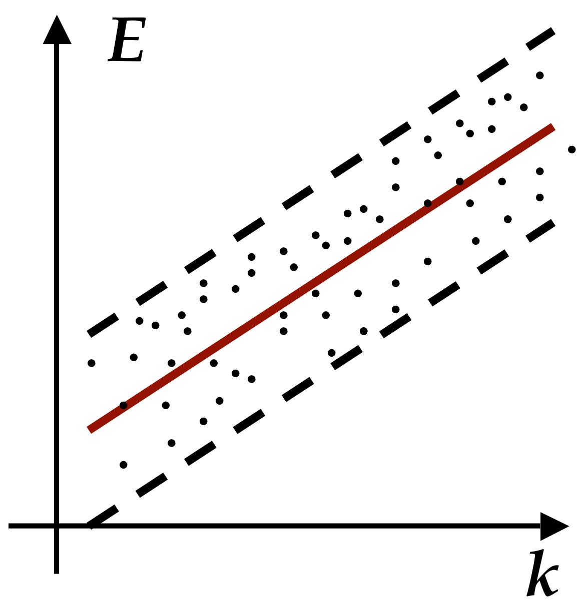
  ✓ *energies have to correlate with properties of the graph*

HOW ?

Attractive potentials

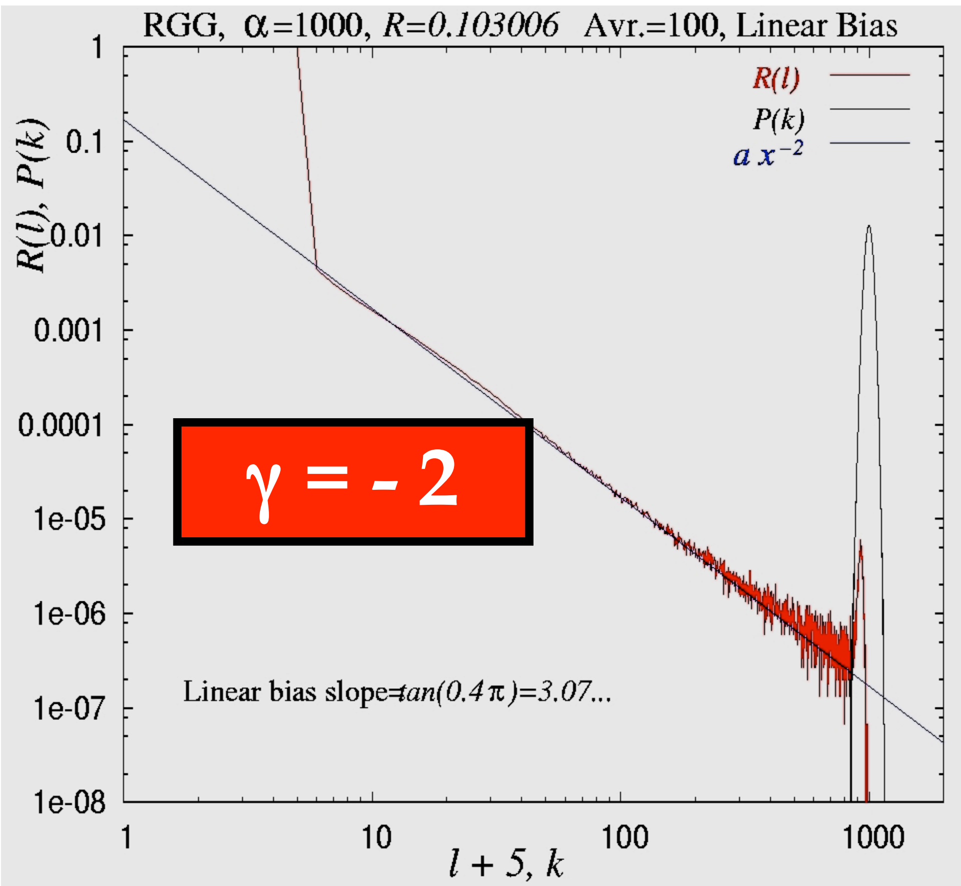
- **Systems with Lennard-Jones like interactions**
  - ✓ ***attractive at medium to long range***
  - ✓ ***repulsive at very short range***
  - ✓ ***the system likes to "clump" (like proteins!)***

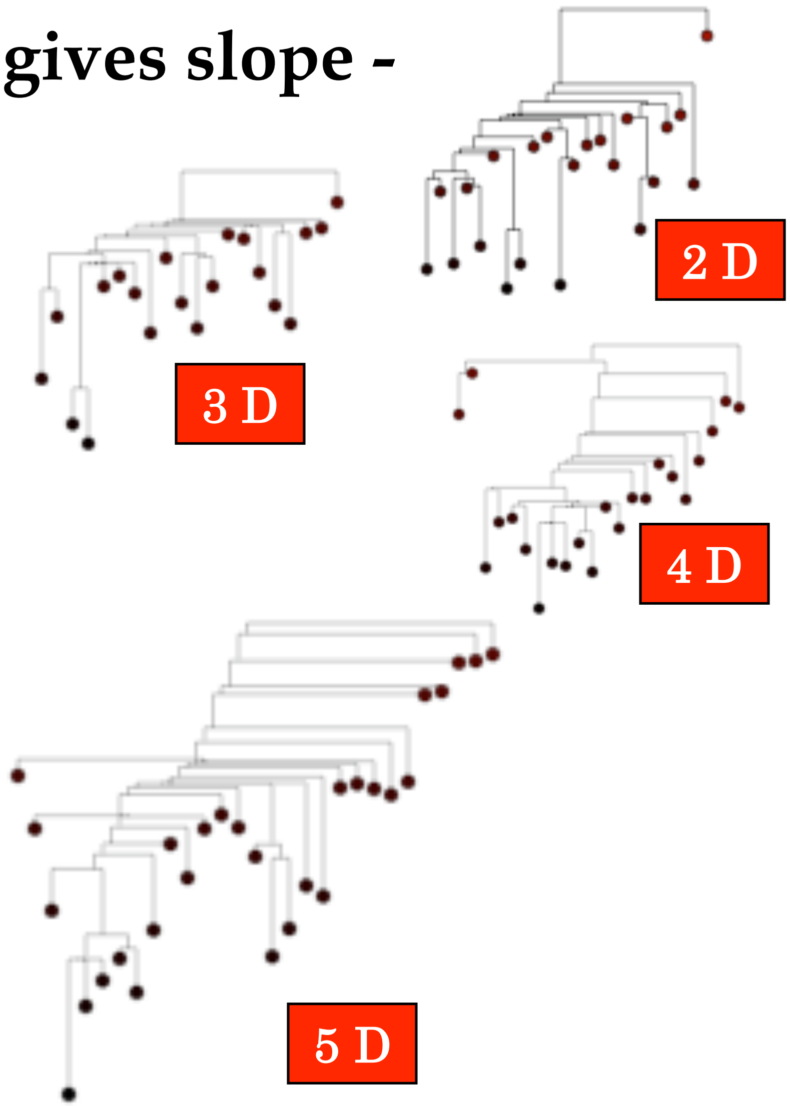
- **Conjecture:**
  - ✓ *small k_{conf} ↔ constrained (folded)*
    ↔ lower energy
  - ✓ *large k_{conf} ↔ loose (random coil)*
    ↔ higher energy



And the winner is

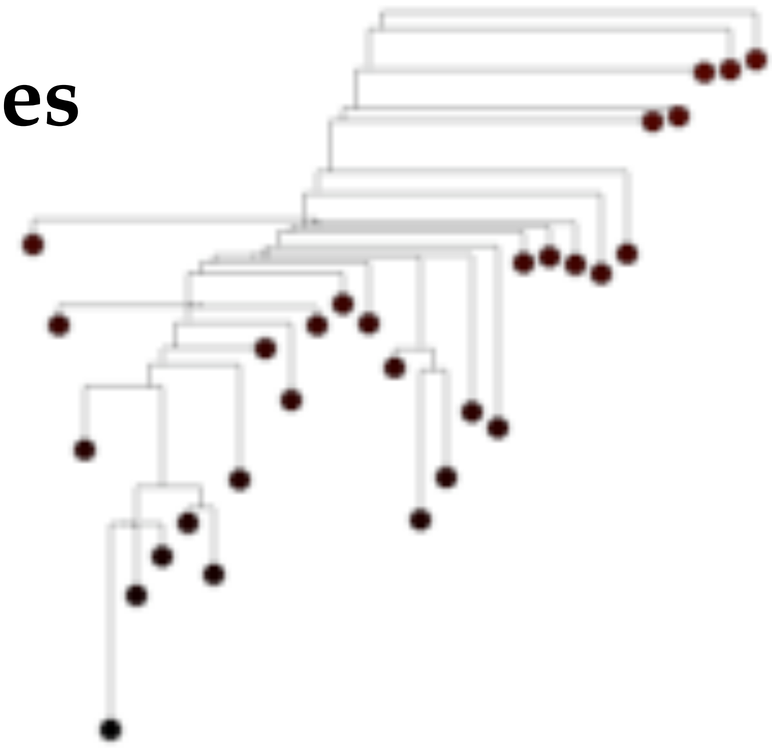
- Random geometric network gives slope -
- AND funnels!





Conclusions

- Swiss cheese model of configuration space:
  - ✓ high D lattice
  - ✓ forbidden subspaces
- Minima at small k
  - ✓ FUNNELS



What's next?

- Are we correct?
  - ✓ *use robot arm measure forbidden subspaces*
  - ✓ *check for funnels: assume and also measure E(k)*
  - ✓ *use MD to measure configuration space and E(k) for real proteins*
  - ✓ *reproduce the MD network using biased walks*
- Analytics
  - ✓ *prove: locally tree-like networks: slope -1 for ANY bias*
  - ✓ *deal with triangles and rectangles*
  - ✓ *solve the RG case (bias and no bias)*


Thank you!