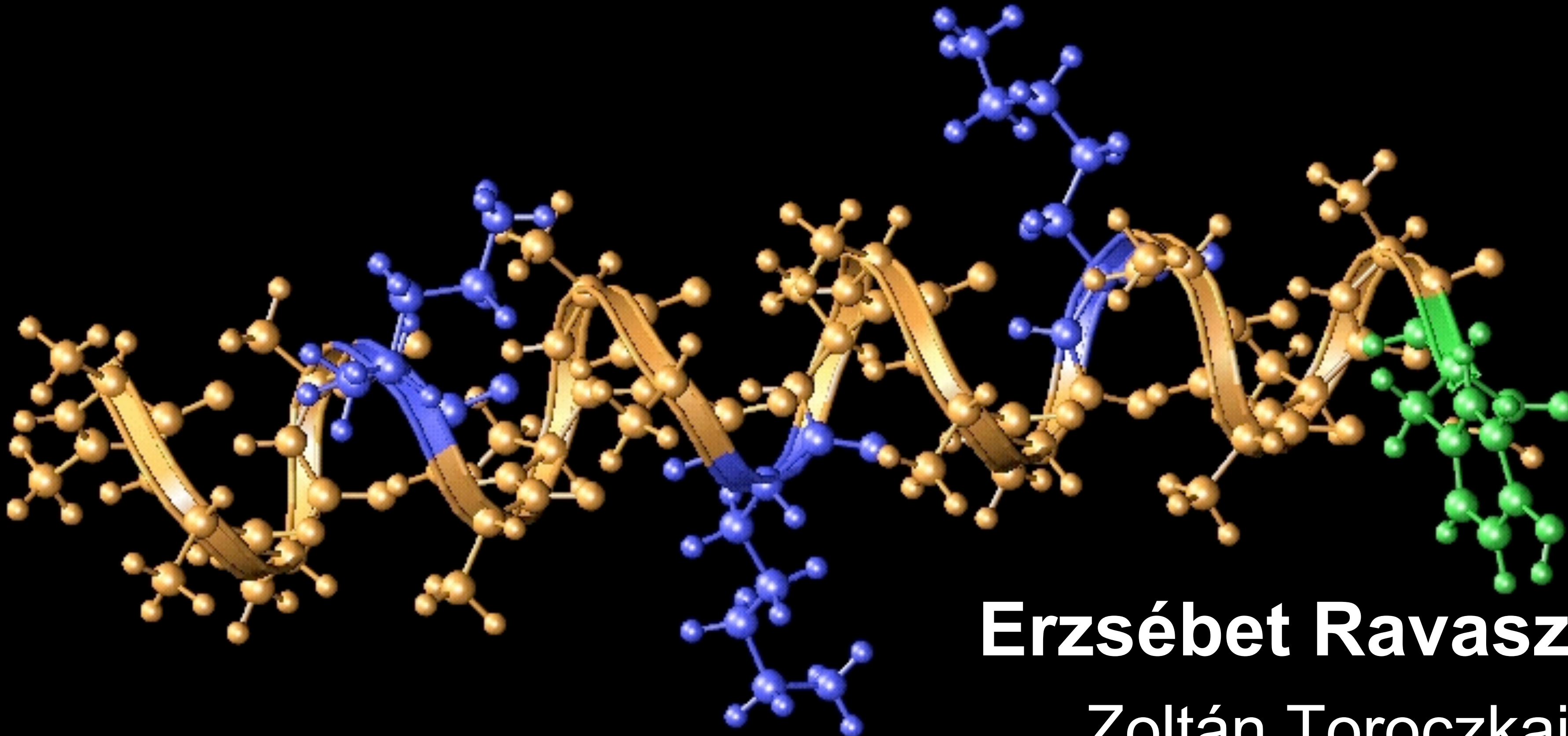
Networks in Protein Folding



Erzsébet Ravasz

Zoltán Toroczkai

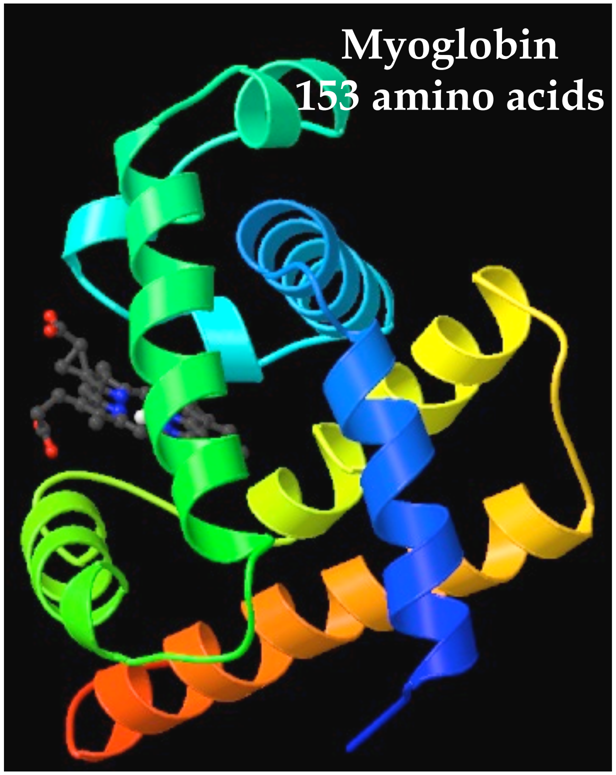
Gnana Gnanakaran

Center for NonLinear Studies (CNLS), Los Alamos National Laboratory

Proteins

Center for Nonlinear Studies

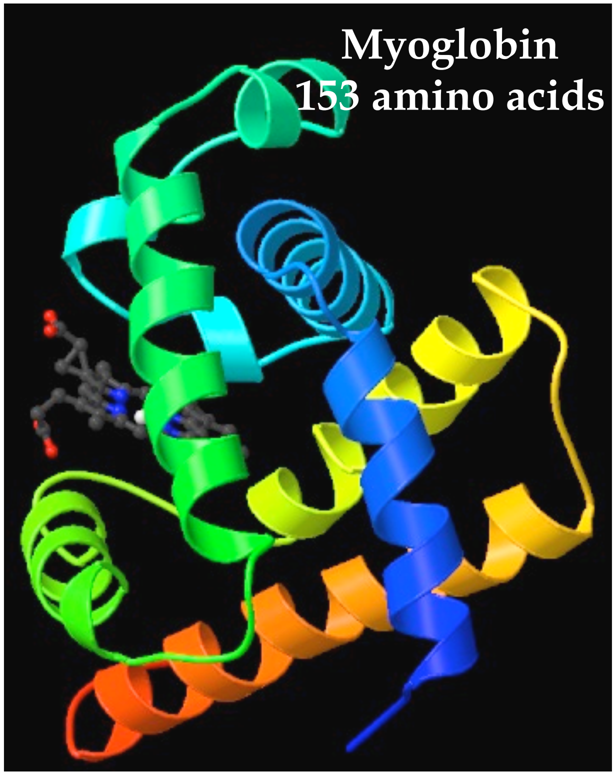
- what are they?
  - ✓ *basic functional units of a cell*
  - ✓ *chains of amino acids (50 – 10^3)*
  - ✓ *peptide bonds link the backbone*



Proteins

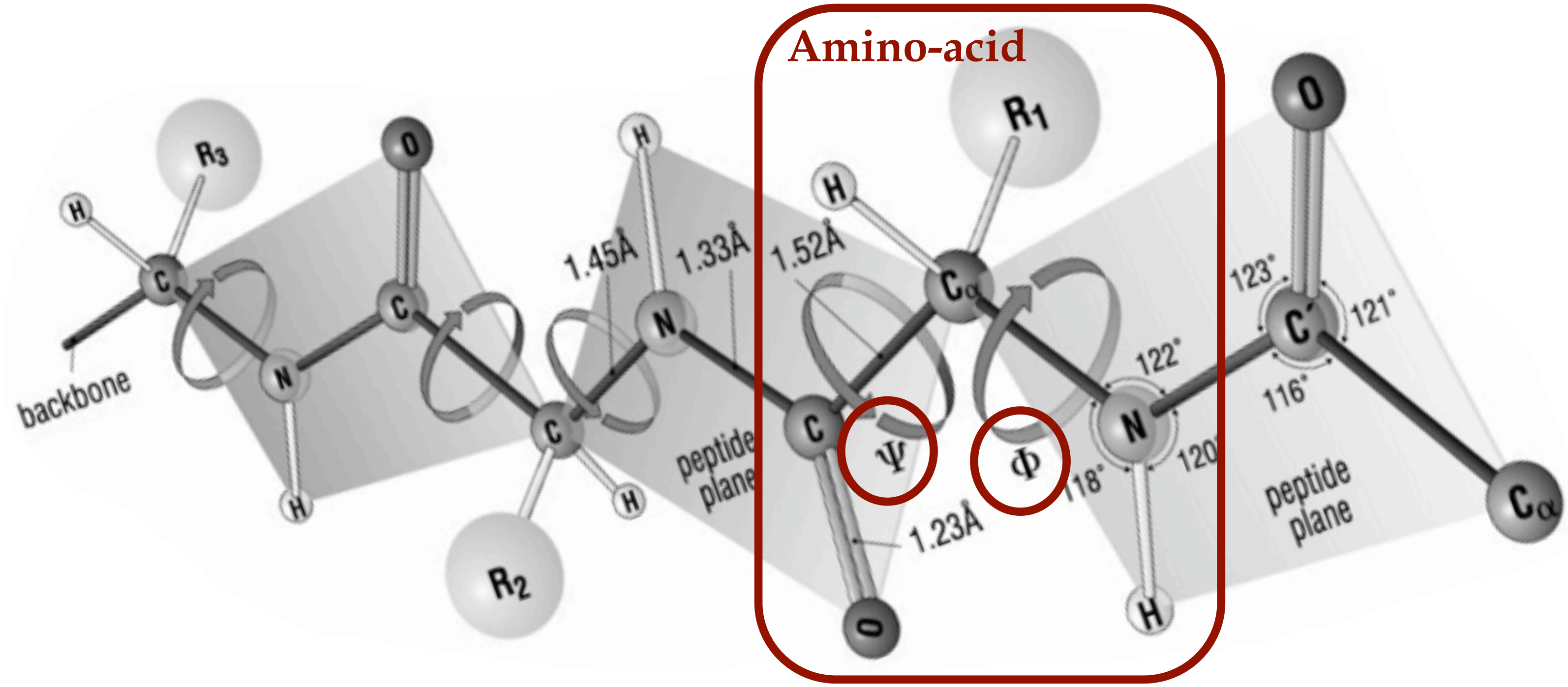
Center for Nonlinear Studies

- what are they?
  - ✓ *basic functional units of a cell*
  - ✓ *chains of amino acids (50 – 10^3)*
  - ✓ *peptide bonds link the backbone*
- native state
  - ✓ *unique 3D structure (native physiological conditions)*
  - ✓ *biological function*
  - ✓ *fold in nanoseconds to minutes*
  - ✓ *about 1000 known 3D structures: X-ray crystallography, NMR*



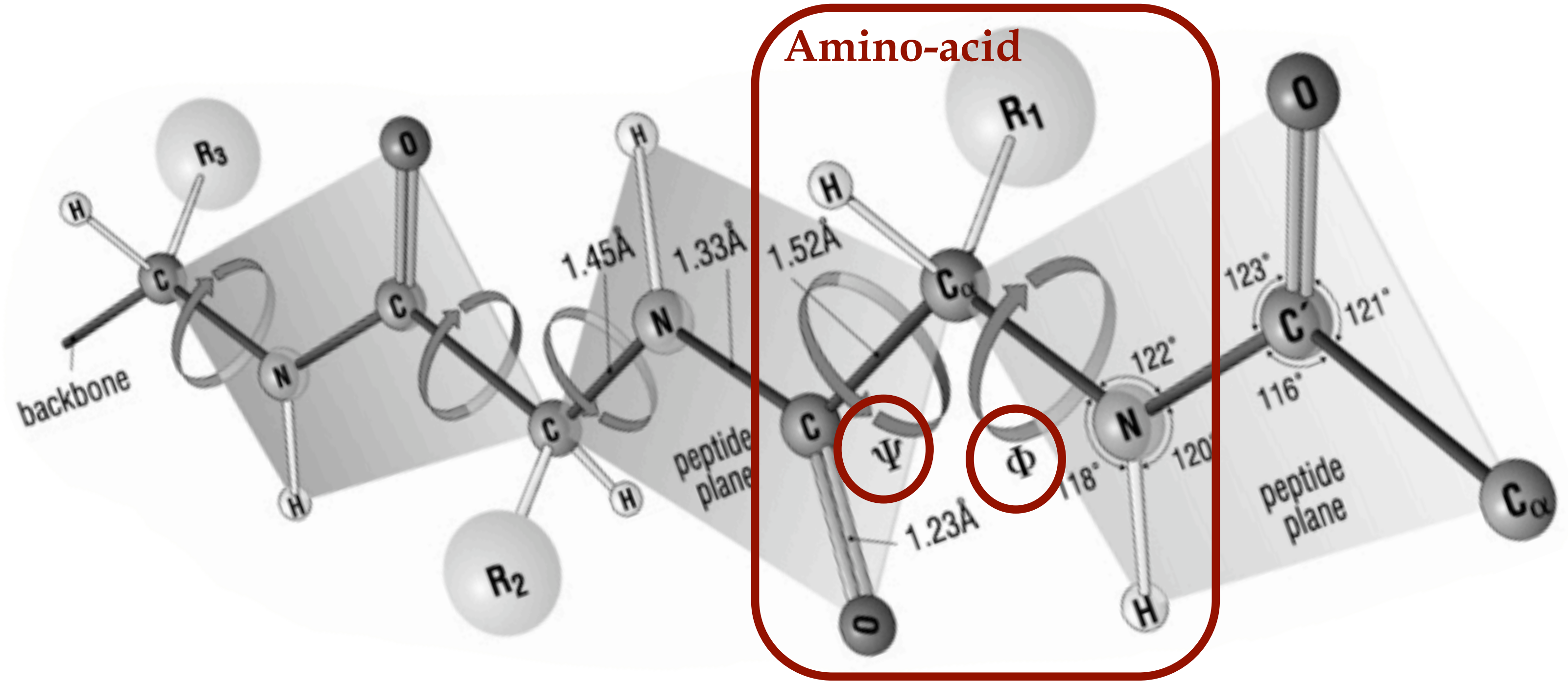
Protein conformations

- **defined by dihedral angles**
  - ✓ *2 angles with 2-3 local minima of the torsion energy*



Protein conformations

- defined by dihedral angles
  - ✓ *2 angles with 2-3 local minima of the torsion energy*



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

Levinthal's paradox

Center for Nonlinear Studies

- **Anfinsen: thermodynamic hypothesis**
  ✓ *native state is at the global minimum of the free energy*

Epstain, Goldberger, & Anfinsen, *Cold Harbor Symp. Quant. Biol.* **28**, 439 (1963)

- **Levinthal's paradox, 1968**

Levinthal, *J. Chim. Phys*. **65**, 44-45 (1968)

✓ *finding the native state by random sampling is not possible*

✓ *40 monomer polypeptide* $\rightarrow 10^{13}$ *conf/s*

$\rightarrow$ *$3 \cdot 10^{19}$ years to sample all*

$\rightarrow$ *universe* $\sim 2 \cdot 10^{10}$ *years old*

Wetlaufer, *P.N.A.S.* **70**, 691 (1973)

Levinthal's paradox

Center for Nonlinear Studies

- **Anfinsen: thermodynamic hypothesis**
  - ✓ *native state is at the global minimum of the free energy*

Epstain, Goldberger, & Anfinsen, *Cold Harbor Symp. Quant. Biol.* **28**, 439 (1963)

- **Levinthal's paradox, 1968**

Levinthal, *J. Chim. Phys*. **65**, 44-45 (1968)

✓ *finding the native state by random sampling is not possible*

✓ *40 monomer polypeptide* $\rightarrow 10^{13}$ *conf/s*

$\rightarrow 3 \cdot 10^{19}$ *years to sample all*

$\rightarrow$ *universe* $\sim 2 \cdot 10^{10}$ *years old*

Wetlaufer, *P.N.A.S.* **70**, 691 (1973)

✓ *nucleation*

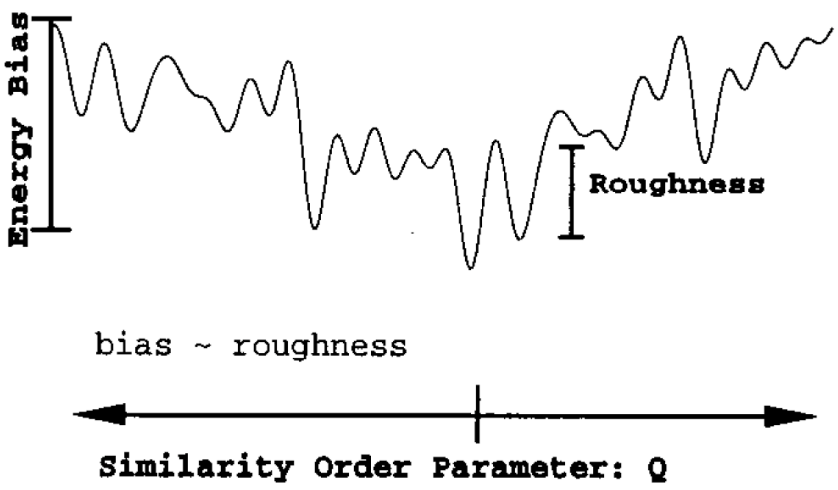
✓ *folding pathways*

Free energy landscapes

Center for Nonlinear Studies

- **Bryngelson & Wolynes, 1987**
  - ✓ *free energy landscape*
  - ✓ *a random hetero-polymer typically does NOT fold*

Bryngelson & Wolynes, *P.N.A.S.* **84**, 7524 (1987))

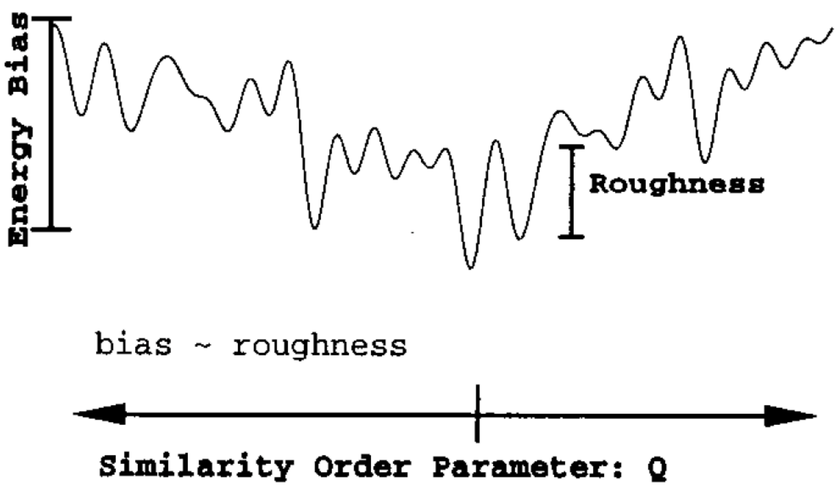


Free energy landscapes

Center for Nonlinear Studies

- **Bryngelson & Wolynes, 1987**
  - ✓ *free energy landscape*
  - ✓ *a random hetero-polymer typically does NOT fold*

Bryngelson & Wolynes, *P.N.A.S.* **84**, 7524 (1987))



Experiment:
— random sequences
— GLU, ARG, LEU
— 80-100 amino-acids
~ 95% did not fold in a stable manner

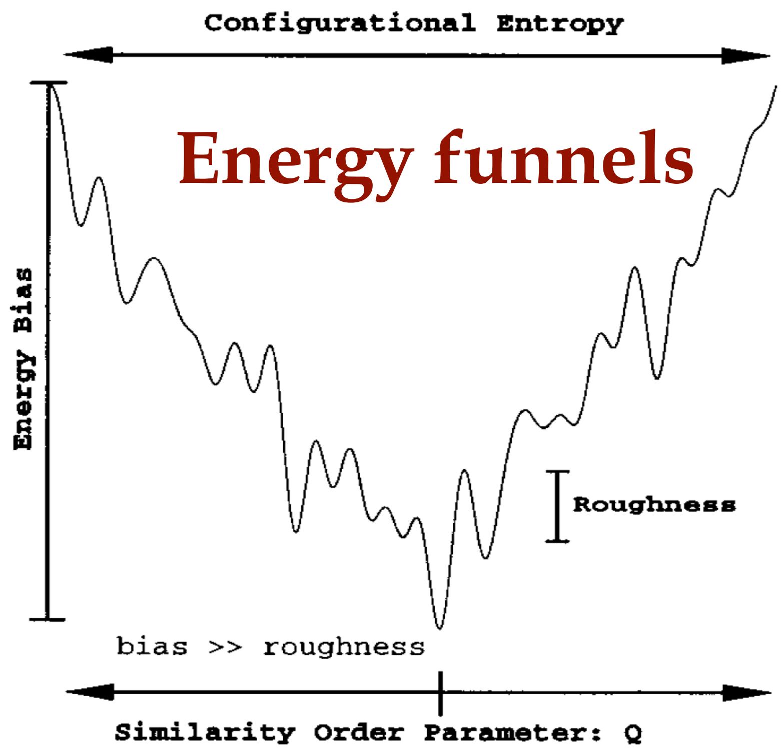
Davidson & Sauer, *P.N.A.S.* **91**, 2146 (1994)

Funnels

Center for Nonlinear Studies

- **Leopold, Mortal & Onuchic, 1992**

Leopold, Mortal & Onuchic, *P.N.A.S.* **89**, 8721 (1992)



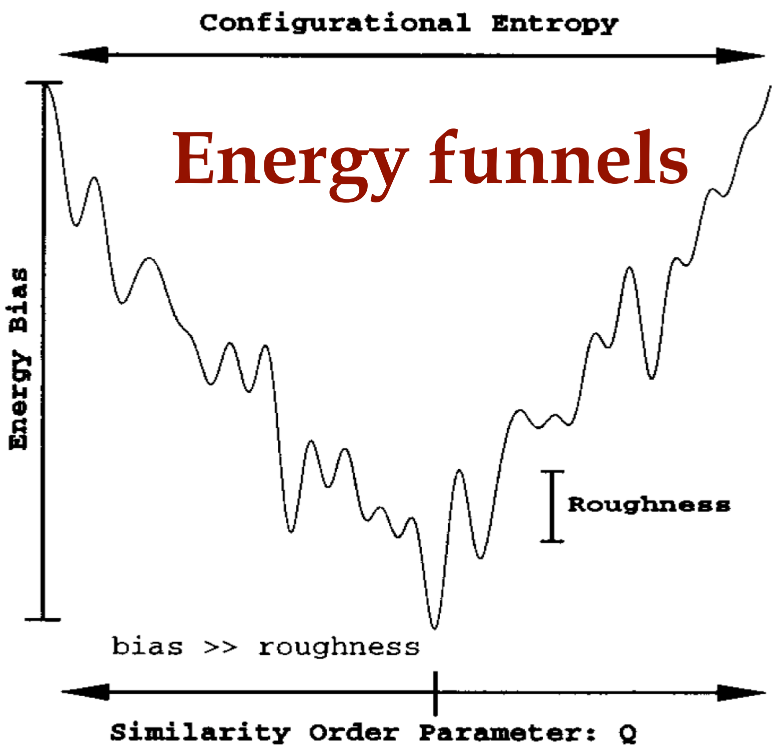
✓ *many folding pathways*

Funnels

Center for Nonlinear Studies

- Leopold, Mortal & Onuchic, 1992

Leopold, Mortal & Onuchic, *P.N.A.S.* **89**, 8721 (1992)



Given any amino-acid sequence: can we tell if it is a good folder?

✓ *experiments (X-ray, NMR)*
✓ *molecular dynamics simulations*
✓ *homology modeling*

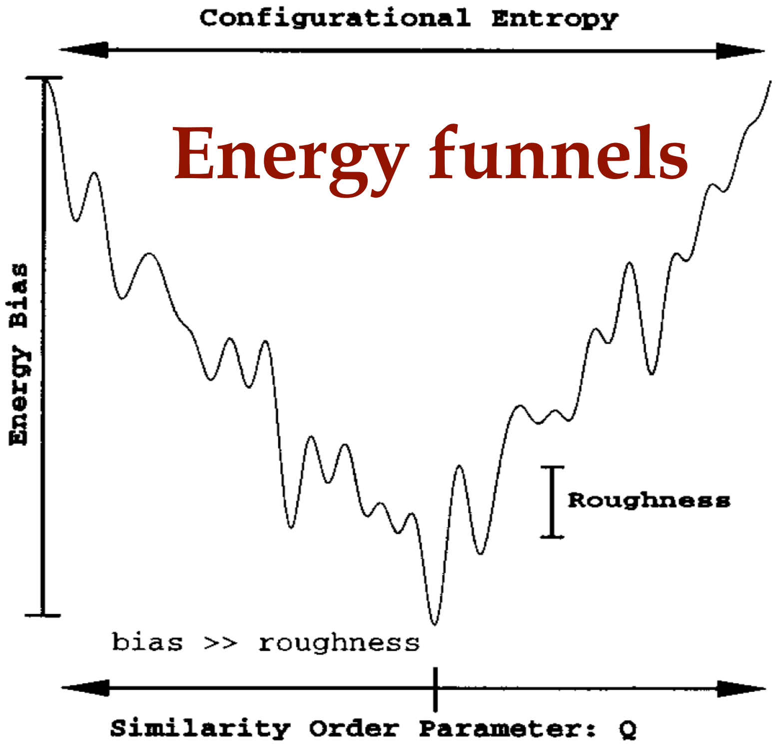
✓ *many folding pathways*

Funnels

Center for Nonlinear Studies

- Leopold, Mortal & Onuchic, 1992

Leopold, Mortal & Onuchic, *P.N.A.S.* **89**, 8721 (1992)



✓ *many folding pathways*

Given any amino-acid sequence: can we tell if it is a good folder?

✓ *experiments (X-ray, NMR)*
✓ *molecular dynamics simulations*
✓ *homology modeling*

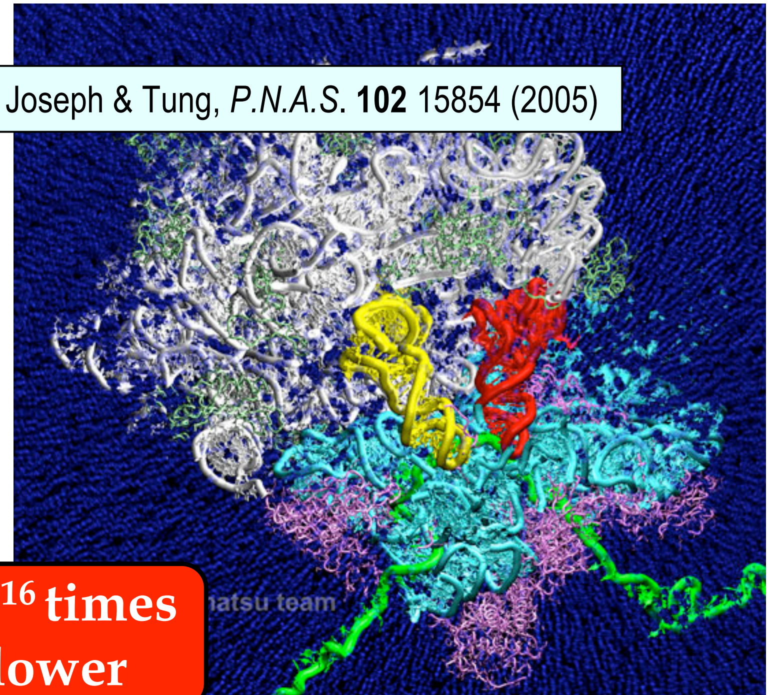
Difficult and slow

Molecular dynamics

Center for Nonlinear Studies

- **State of the art**
  - ✓ *supercomputer* (*LANL*)

Sanbonmatsu, Joseph & Tung, *P.N.A.S.* **102** 15854 (2005)



Ribosome in explicit solvent:
- *targeted MD*
- 2.64×10^{6} atoms (2.5×10^{5} + water)
- Q machine, 768 processors
- 260 days of simulation (event: 2 ns)

~ 10^{16} times slower

- ✓ ***distributed computing*** **(*Stanford,* Folding@home)**

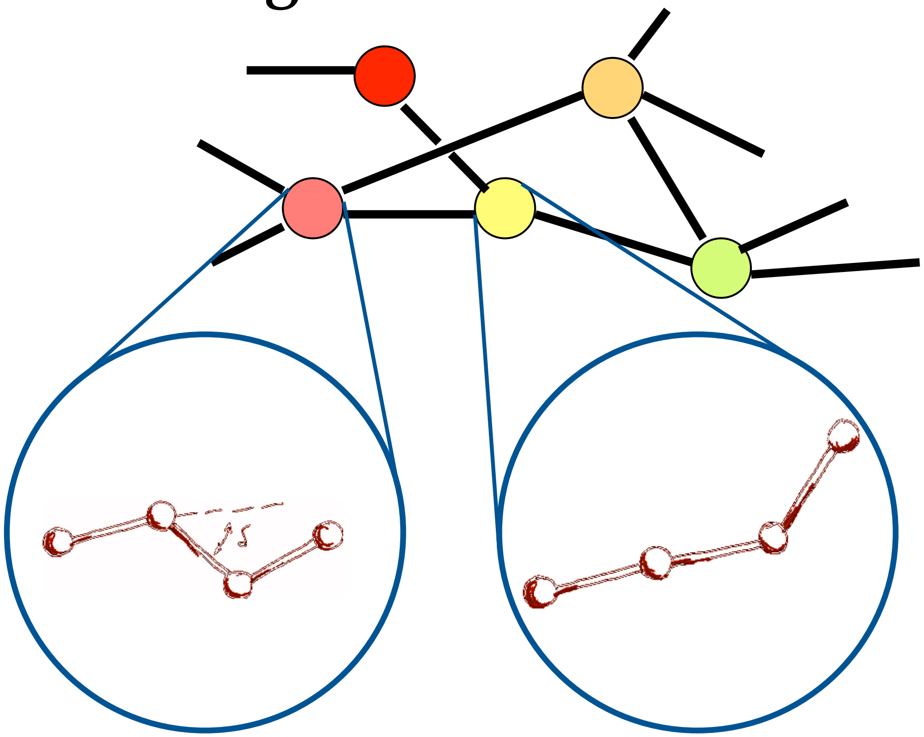
- more than 100,000 CPU's
- simulation of complete folding event
  - » BBA5, 23-residue, implicit water
  - » 10,000 CPU days/folding event (~1μs)

Shirts & Pande, *Science* **290**, 1903 (2000)
Snow, Nguyen, Pande, Gruebele, *Nature* **420**,102 (2002)

Configuration networks

Center for Nonlinear Studies

- Configuration networks



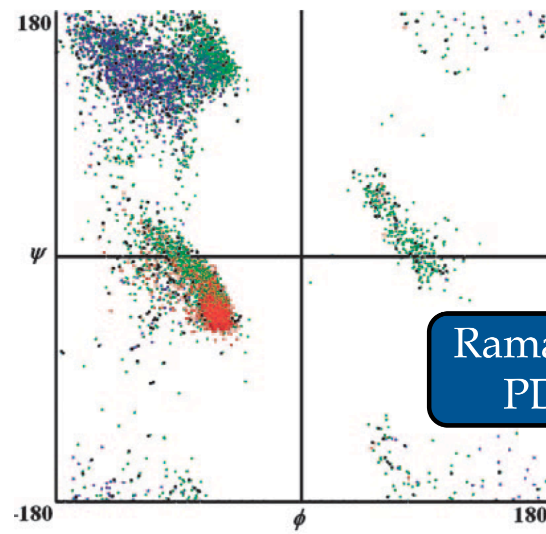
NODE ← *configuration*

LINK ← *change of one degree of freedom (angle)*

Protein conformations

✓ dihedral angles have few preferred values

Ramachandran & Sasisekharan, *J.Mol.Biol.* **7**, 95 (1963)



Ramachandran map
PDB structures

✓ refinement of angle values → continuous case

Why networks?

Center for Nonlinear Studies

- **VERY LARGE:** *100 monomers* → *10^{100} nodes. However:*

Generic features of folding are determined by STATISTICAL properties of the configuration network

Why networks?

Center for Nonlinear Studies

- **VERY LARGE:** *100 monomers → 10^{100} nodes. However:*

Generic features of folding are determined by STATISTICAL properties of the configuration network

✓ *toolkit from network research*

✓ *captures the high dimensionality*

- ♦ degree distribution
- ♦ average distance
- ♦ clustering
- ♦ degree correlations

Albert & Barabási, *Rev. Mod. Phys.* **74,** 67 (2002); Newman, *SIAM Rev.* **45**, 167 (2003)

Why networks?

Center for Nonlinear Studies

- **VERY LARGE:** ***100 monomers → 10^{100} nodes. However:***

Generic features of folding are determined by STATISTICAL properties of the configuration network

✓ ***toolkit from network research***

✓ ***captures the high dimensionality***

- **degree distribution**
- **average distance**
- **clustering**
- **degree correlations**

Albert & Barabási, *Rev. Mod. Phys.* **74,** 67 (2002); Newman, *SIAM Rev.* **45**, 167 (2003)

- faster algorithms to simulate folding events
- pre-screening synthetic proteins
- insights into misfolding

A real example

Center for Nonlinear Studies

• *The Protein Folding Network:* F. Rao, A. Caflisch, *J.Mol.Biol*, 342, 299 (2004)

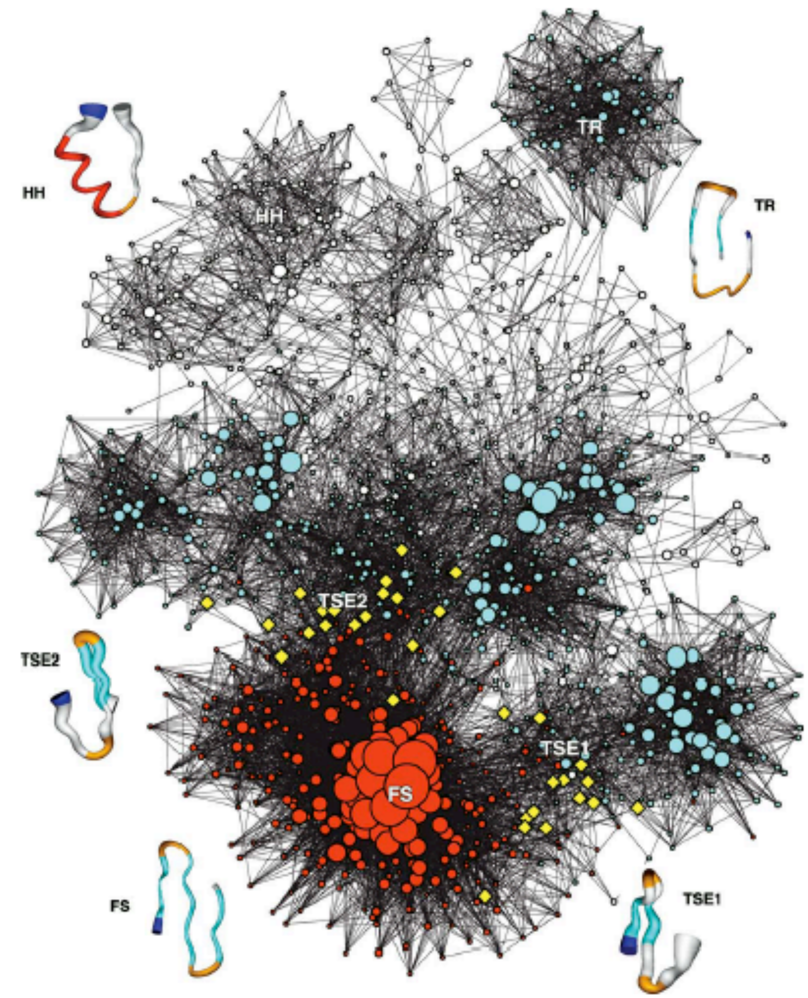
✓ beta3s: *20 monomers, antiparallel beta sheets*

✓ *MD simulation, implicit water*

✓ *330K, equilibrium folded ↔ random coil*

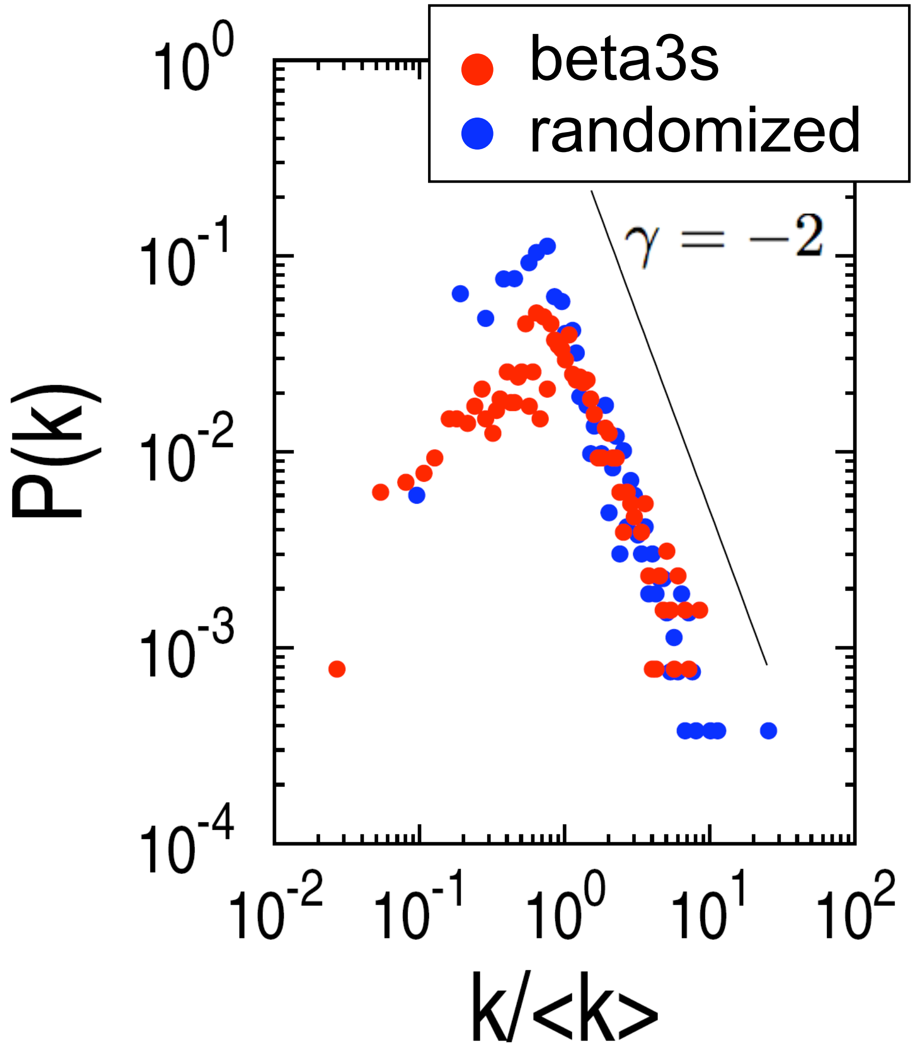
NODE -- 8 letters / AA
(local secondary struct)

LINK -- 2ps transition



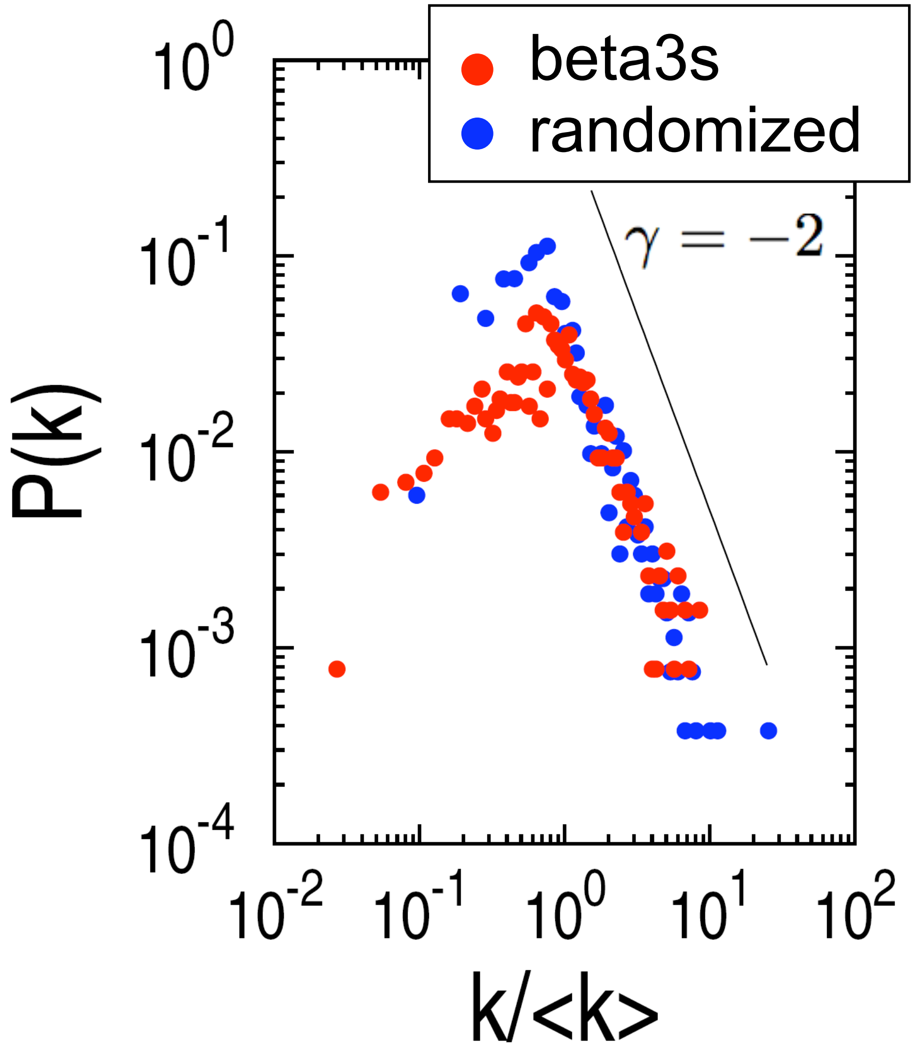
Scale-free network

Center for Nonlinear Studies



Scale-free network

Center for Nonlinear Studies



Many real-world networks are scale free

✓ $P(k) \sim k^{-\gamma}$

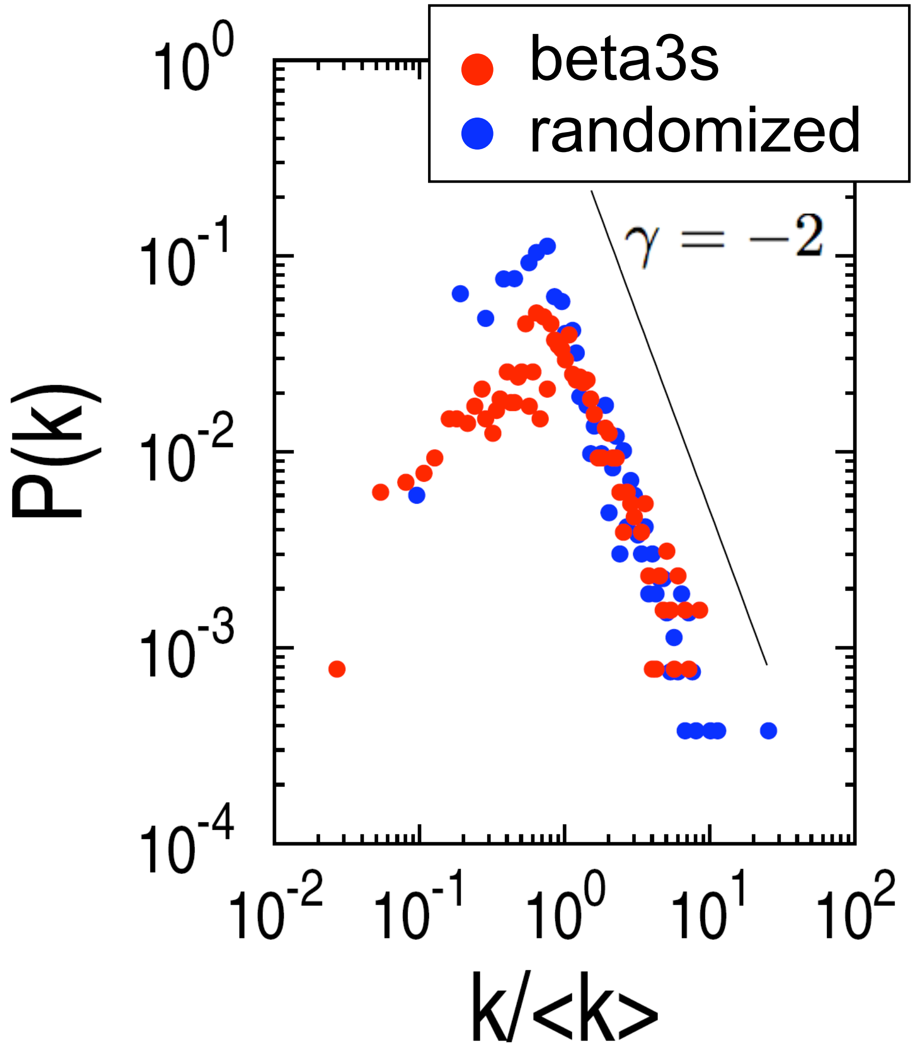
✓ ***hubs***

Barabási & Albert, *Science* **286**, 509, (1999);

- co-authorship (γ=1 - 2.5)
- citations (γ=3)
- sexual contacts (γ=3.4)
- movie actors (γ=2.3)
- Internet (γ=2.4)
- World Wide Web (γ=2.1/2.5)
- Genetic regulation (γ=1.3)
- Protein-protein interactions (γ =2.4)
- Metabolic pathways (γ=2.2)
- Food webs (γ=1.1)

Scale-free network

Center for Nonlinear Studies



Many real-world networks are scale free

✓ $P(k) \sim k^{-\gamma}$

✓ ***hubs***

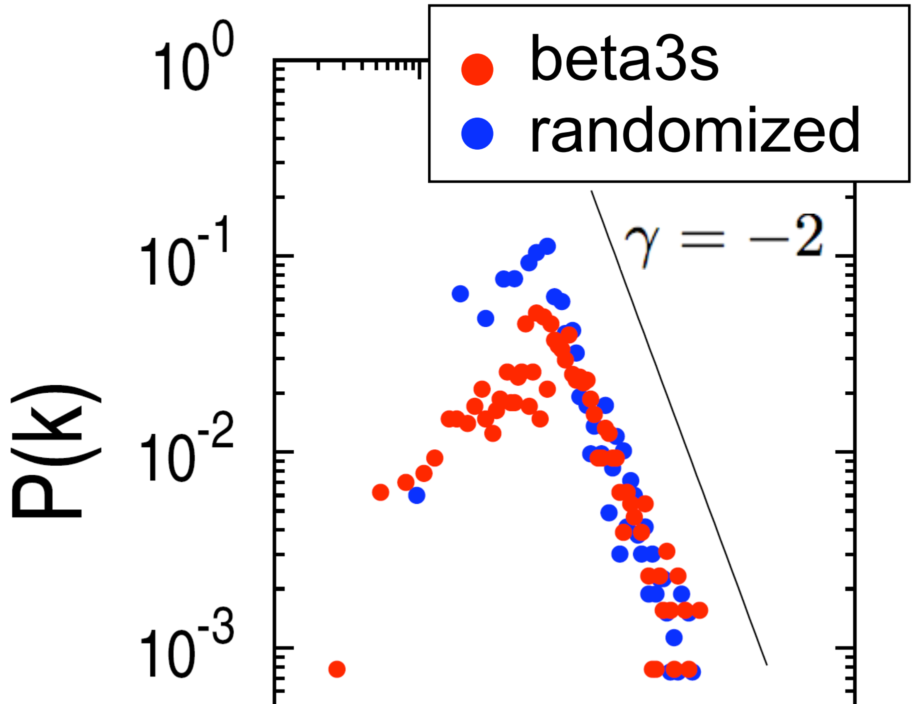
Barabási & Albert, *Science* **286,** 509, (1999);

Many reasons behind SF topology

- cl…
- sexual contacts (γ=3.4)
- movie actors (γ=2.3)
- Internet (γ=2.4)
- World Wide Web (γ=2.1/2.5)
- Genetic regulation (γ=1.3)
- Protein-protein interactions (γ =2.4)
- Metabolic pathways (γ=2.2)
- Food webs (γ=1.1)

Scale-free network

Center for Nonlinear Studies



Many real-world networks are scale free

✓ $P(k) \sim k^{-\gamma}$

✓ *hubs*

Barabási & Albert, *Science* **286**, 509, (1999);

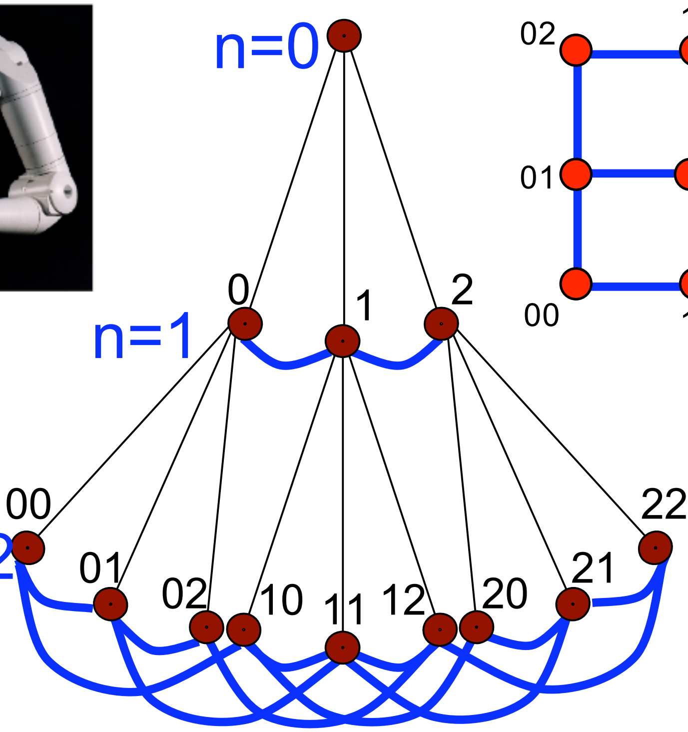
Many reasons behind SF topology

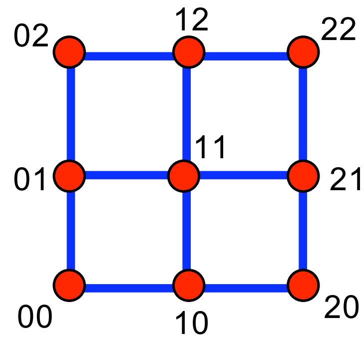
- ♦ cl…
- ♦ sexual contacts (γ=3.4)
- …tors (γ=2.3)
- …=2.4)
- …le Web (γ=2.1/2.5)
- …gulation (γ=1.3)
- …tein interactions (γ =2.4)
- …pathways (γ=2.2)
- …s (γ=1.1)

- Why is the protein network scale free?
- Why does the randomized chain have similar degree distribution?
- Why is γ = - 2 ?

Robot arm networks

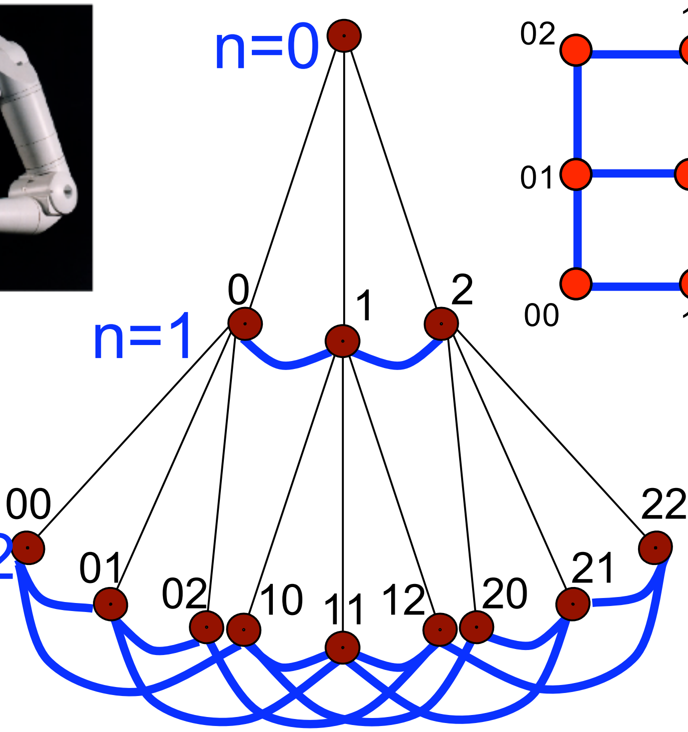
Center for Nonlinear Studies

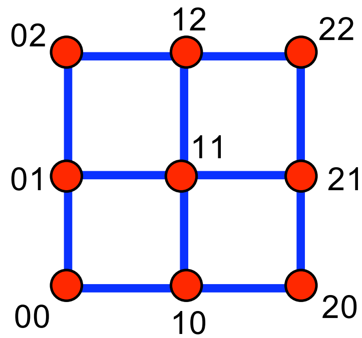


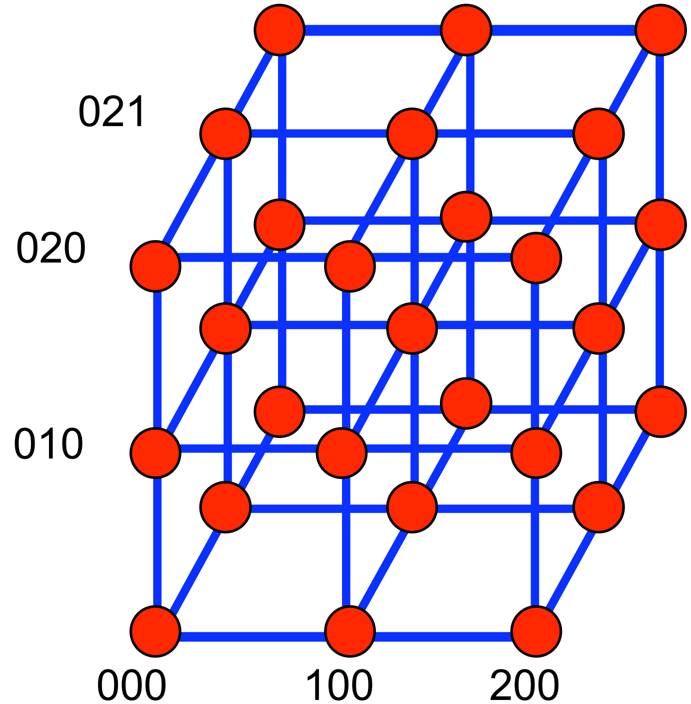


Robot arm networks

Center for Nonlinear Studies





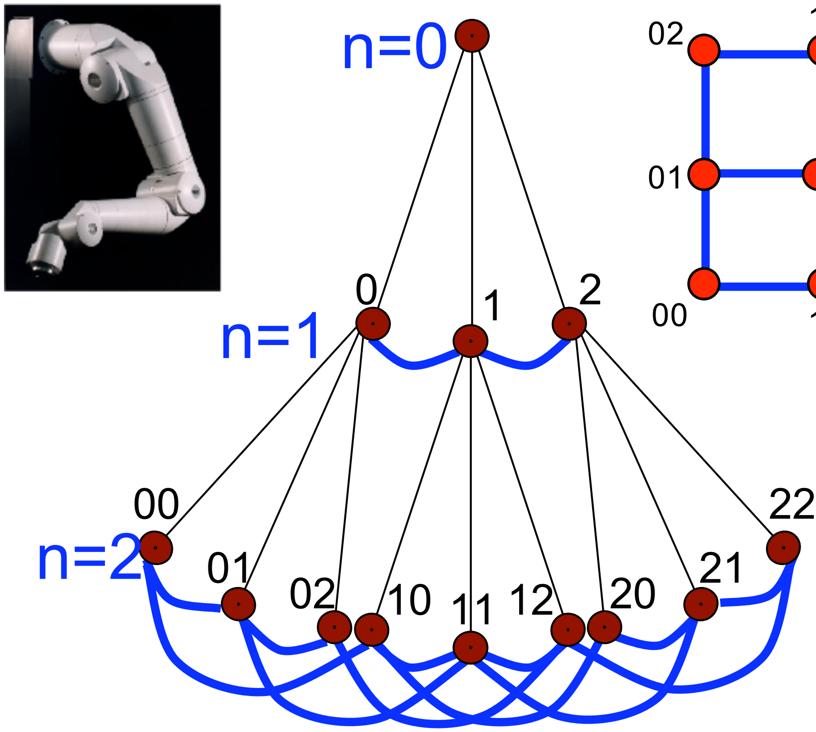


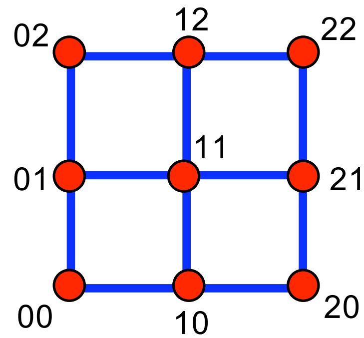
- ✓ *n-dimensional hypercube*
- ✓ *binomial degree distribution*

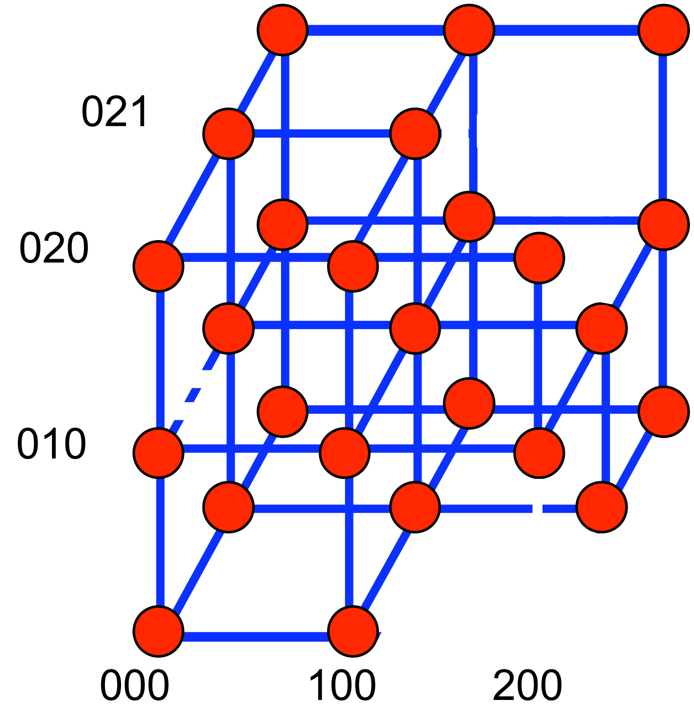
Homogeneous

Robot arm networks

Center for Nonlinear Studies







✓ *n-dimensional hypercube*

✓ *binomial degree distribution*

Homogeneous

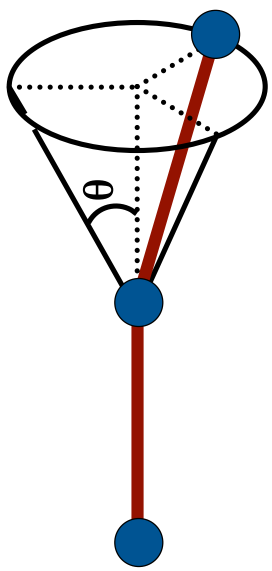
- **Steric constraints?**
  - ✓ *missing nodes*
  - ✓ *missing links*

Swiss cheese

A bead-chain model

- **Beads on a chain in 3D: robot arm model**
  - ✓ *similar to C_α protein models*
  - ✓ *rod-rod angle* **Θ**
  - ✓ *3 positions around axis*

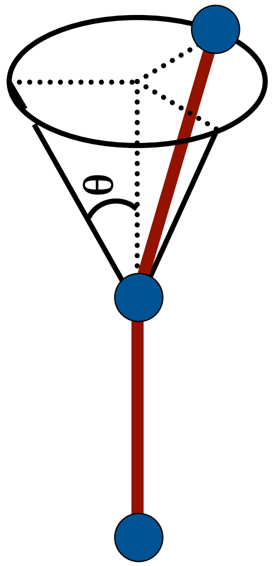
Honeycutt & Thirumalai, *Biopolymers* **32**, 695 (1992)

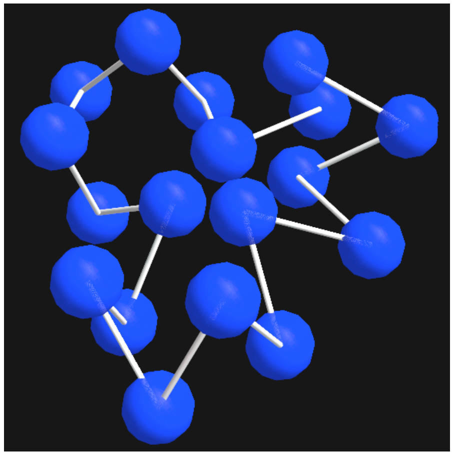


A bead-chain model

- **Beads on a chain in 3D: robot arm model**
  - ✓ *similar to C_α protein models*
  - ✓ *rod-rod angle Θ*
  - ✓ *3 positions around axis*

Honeycutt & Thirumalai, *Biopolymers* **32**, 695 (1992)





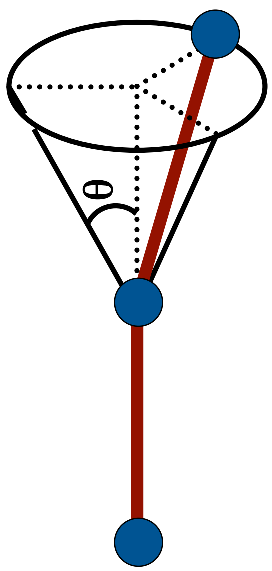
N=18; Θ = 120
221211221211111122

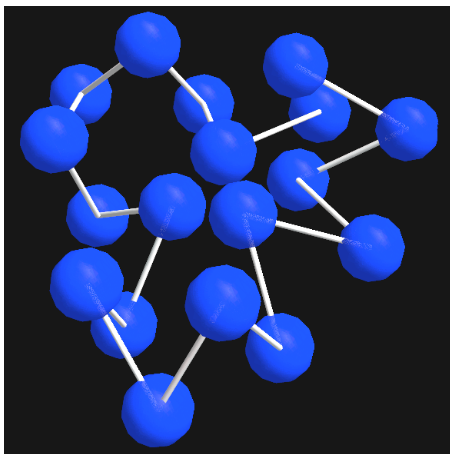
A bead-chain model

Center for Nonlinear Studies

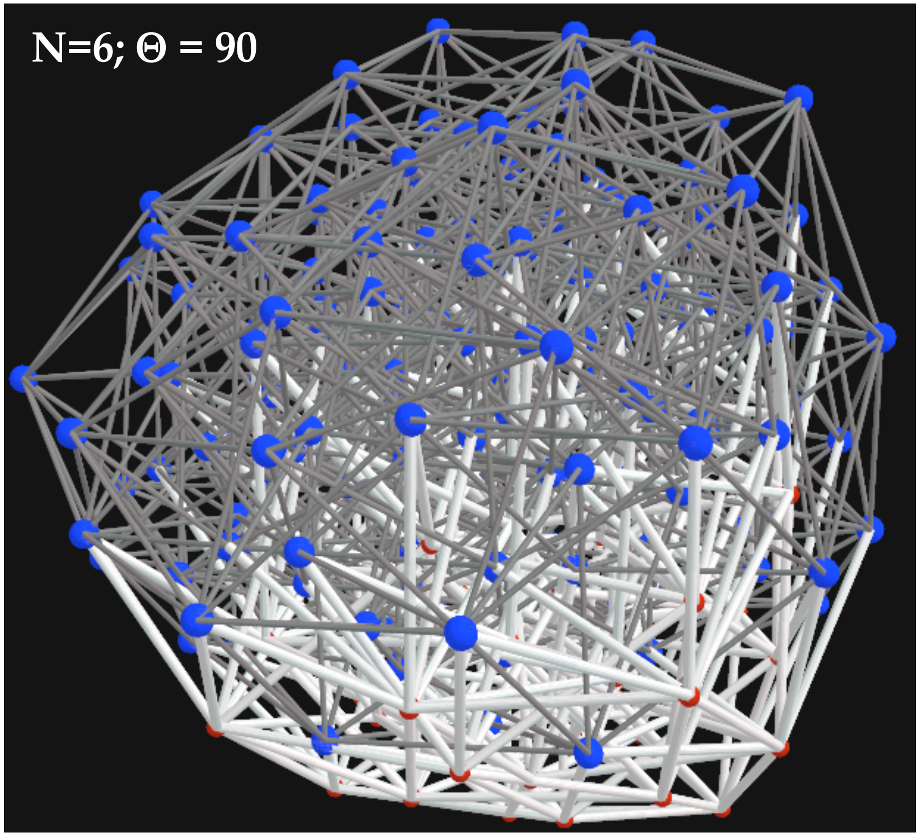
- **Beads on a chain in 3D: robot arm model**
  - ✓ *similar to C_α protein models*
  - ✓ *rod-rod angle Θ*
  - ✓ *3 positions around axis*

Honeycutt & Thirumalai, *Biopolymers* **32**, 695 (1992)





N=18; Θ = 120
221211221212111122



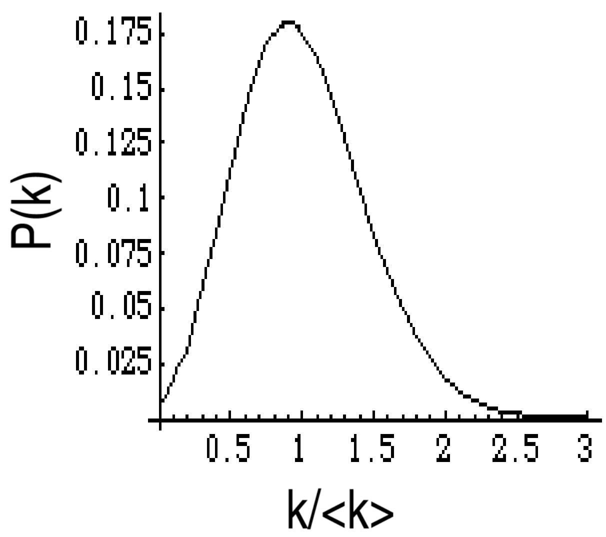
✓ **Homogeneous network**

The “dilemma”

Center for Nonlinear Studies

HOMOGENEOUS

- **from studies of conformation networks**
  - ✓ **bead chain**
  - ✓ **robot arm**



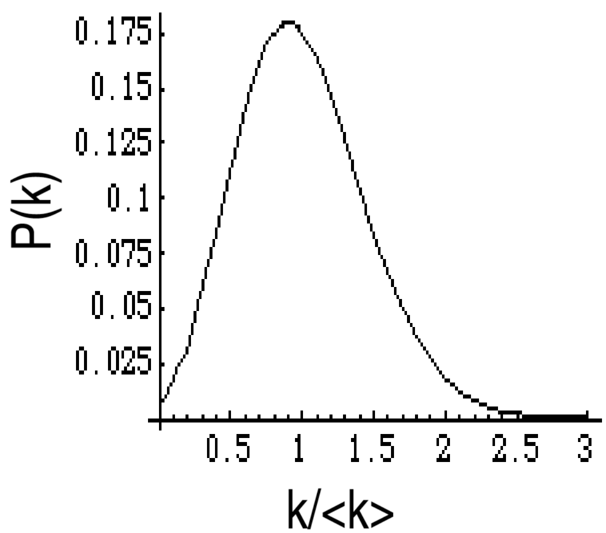
The "dilemma"

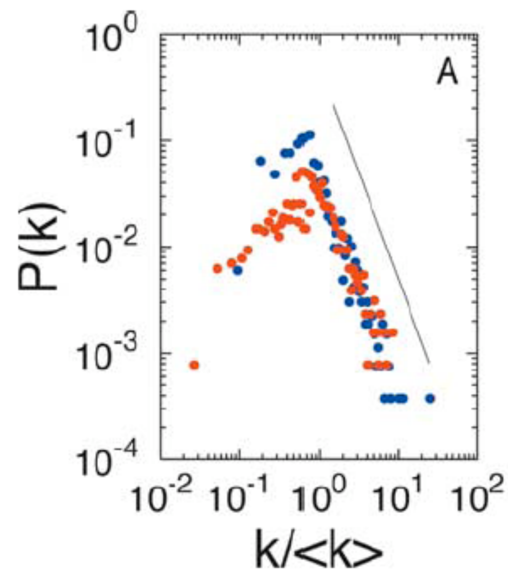
Center for Nonlinear Studies

HOMOGENEOUS ⟷ **SCALE FREE**

- from studies of conformation networks
  - ✓ bead chain
  - ✓ robot arm

- from polypeptide MD simulations
  - ✓ beta3s
  - ✓ randomized version





The "dilemma"

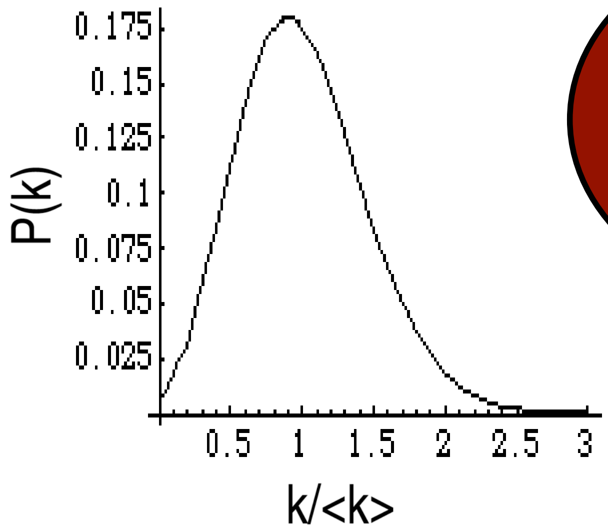
Center for Nonlinear Studies

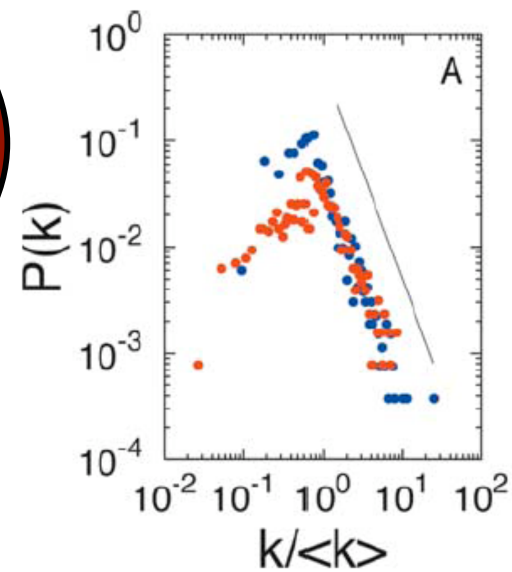
HOMOGENEOUS ⟷ **SCALE FREE**

- from studies of conformation networks
  - ✓ bead chain
  - ✓ robot arm

- from polypeptide MD simulations
  - ✓ beta3s
  - ✓ randomized version

?



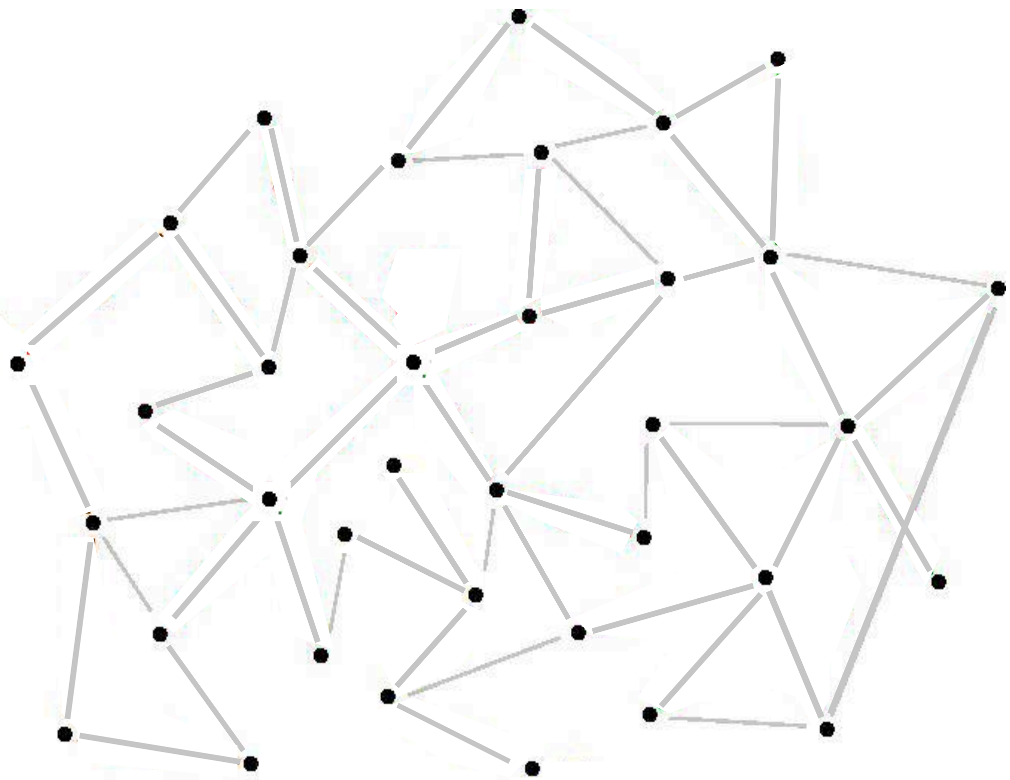


Gradient flow networks

Center for Nonlinear Studies

- ✓ *homogeneous support network*
- ✓ *real numbers on nodes*

Toroczkai & Bassler, *Nature* **428**, 716 (2004);
ArXiv: cond-mat/0408262 (2004)



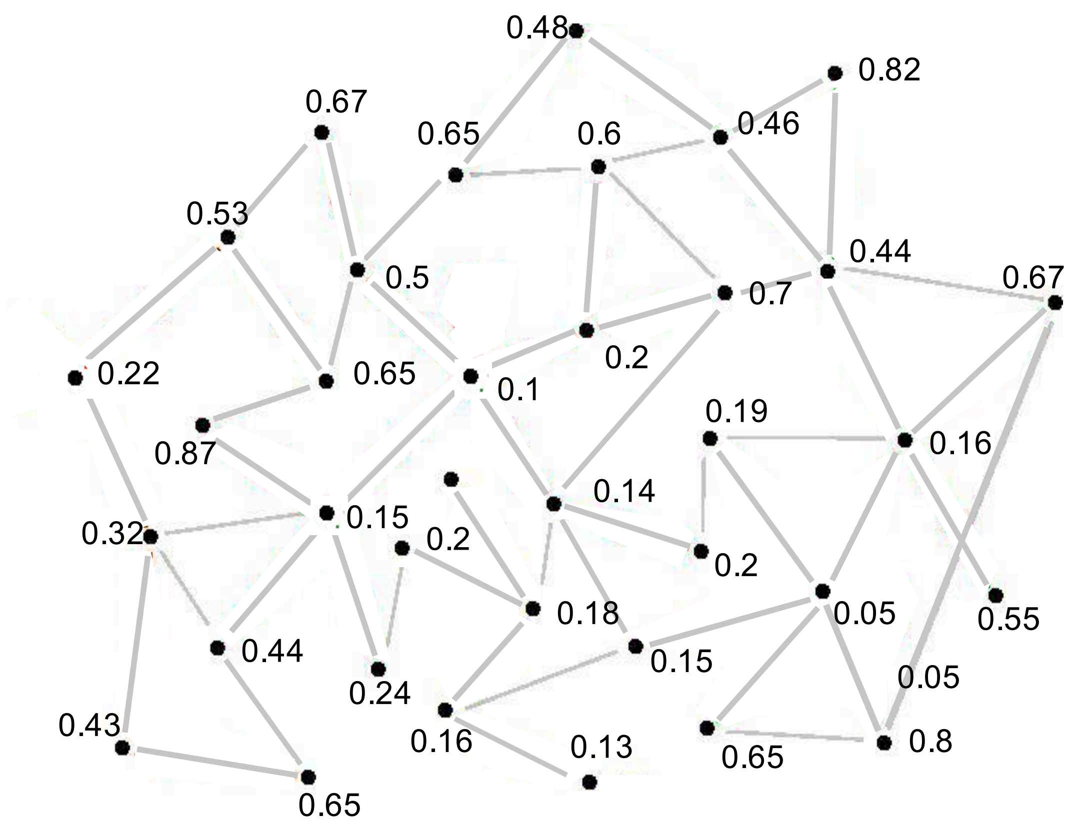
Gradient flow networks

Center for Nonlinear Studies

✓ *homogeneous support network*

✓ *real numbers on nodes*

Toroczkai & Bassler, *Nature* **428**, 716 (2004); ArXiv: cond-mat/0408262 (2004)

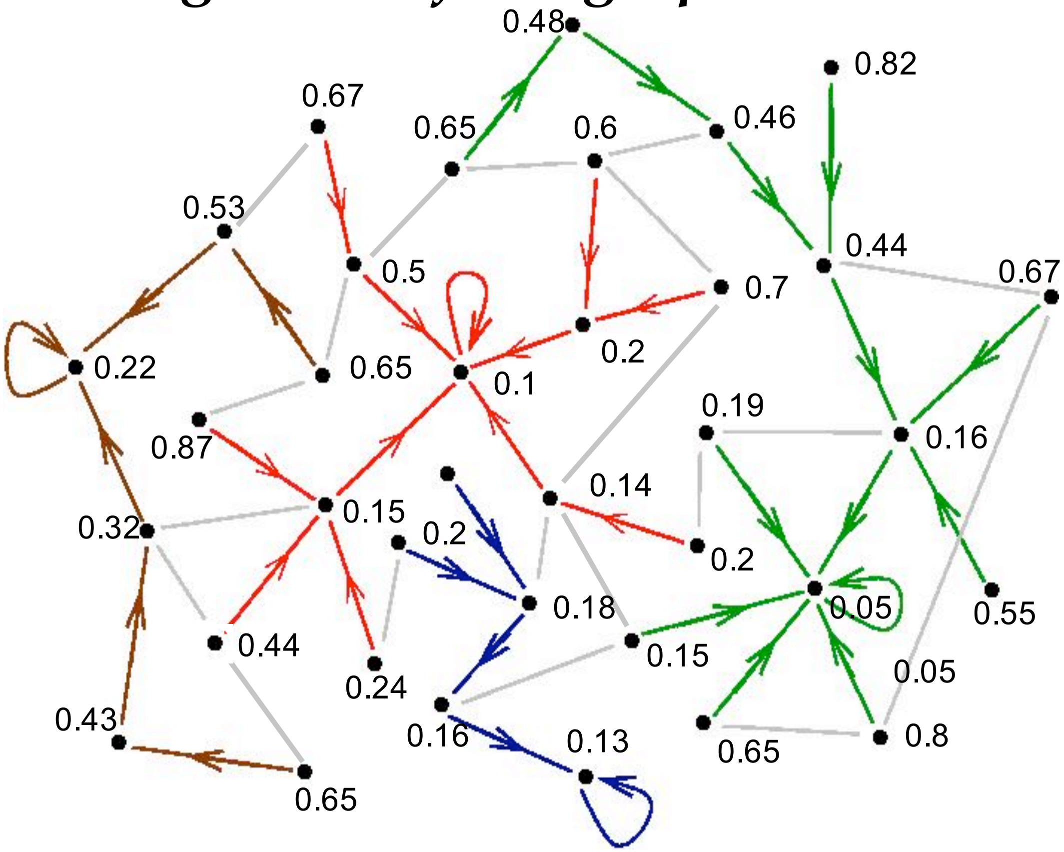


Gradient flow networks

Center for Nonlinear Studies

- ✓ *homogeneous support network*
- ✓ *real numbers on nodes*
- ✓ *gradient flow graph*

Toroczkai & Bassler, *Nature* **428**, 716 (2004); ArXiv: cond-mat/0408262 (2004)

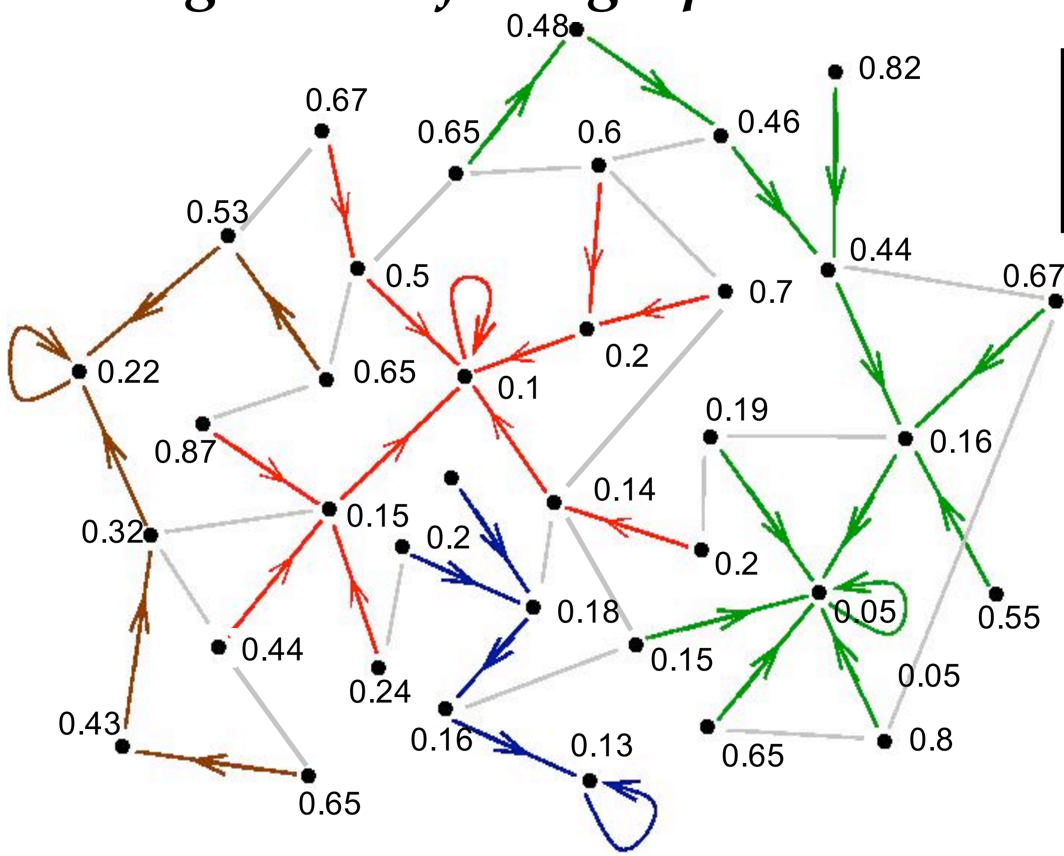


Gradient flow networks

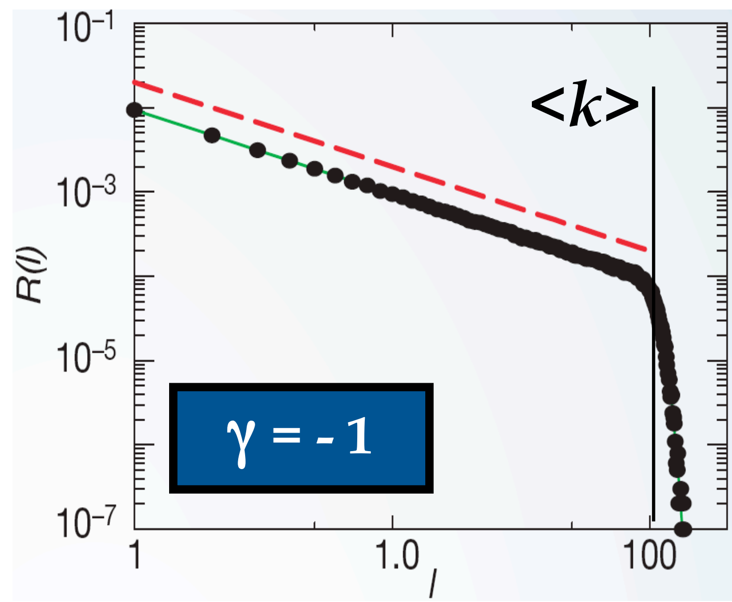
Center for Nonlinear Studies

- ✓ *homogeneous support network*
- ✓ *real numbers on nodes*
- ✓ *gradient flow graph*

Toroczkai & Bassler, *Nature* **428**, 716 (2004); ArXiv: cond-mat/0408262 (2004)

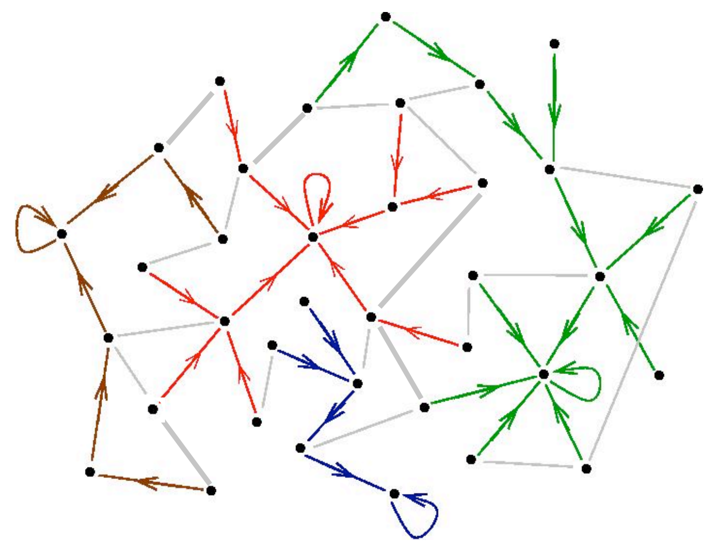


gradient network is scale-free



The energy landscape

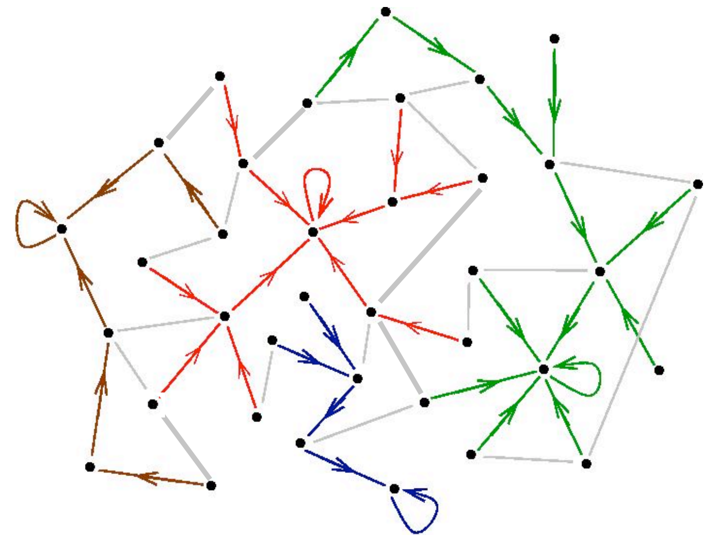
- Energy associated with each node (configuration)
  - ✓ *the gradient network*
    - → *most favorable transitions*
    - → *T=0 backbone of the flow*
  - ✓ *MD simulation*
    - → *tracks the flow network*
    - → biased *walk close to the gradient network*
  - ✓ trees
    - → *basins of local minima*



The energy landscape

Center for Nonlinear Studies

- Energy associated with each node (configuration)
  - ✓ *the gradient network*
    - → *most favorable transitions*
    - → *T=0 backbone of the flow*
  - ✓ *MD simulation*
    - → *tracks the flow network*
    - → biased *walk close to the gradient network*
  - ✓ trees
    - → *basins of local minima*

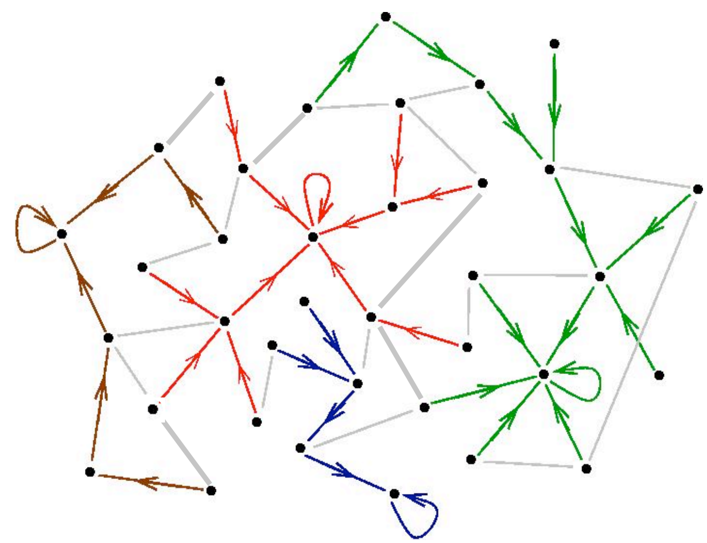


How do we get $\gamma = -2$?

The energy landscape

Center for Nonlinear Studies

- Energy associated with each node (configuration)
  - ✓ *the gradient network*
    - → *most favorable transitions*
    - → *T=0 backbone of the flow*
  - ✓ *MD simulation*
    - → *tracks the flow network*
    - → biased *walk close to the gradient network*
  - ✓ trees
    - → *basins of local minima*



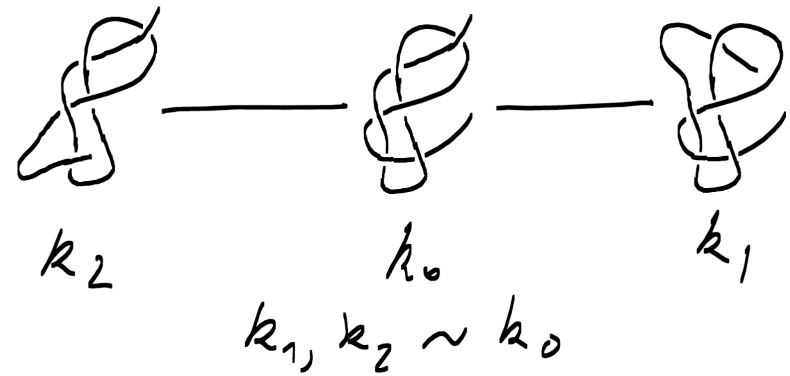
How do we get $\gamma = -2$?

And funnels?

Model ingredients

Center for Nonlinear Studies

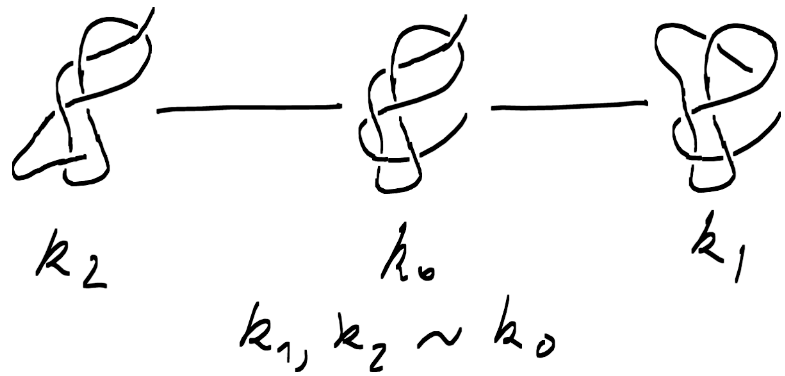
- **A network model of configuration spaces**
  - ✓ ***network topology***
    - → ***homogeneous***
    - → ***degree correlations***



Model ingredients

Center for Nonlinear Studies

- **A network model of configuration spaces**
  - ✓ *network topology*
    - → *homogeneous*
    - → *degree correlations*
  - ✓ *how to associate energies*



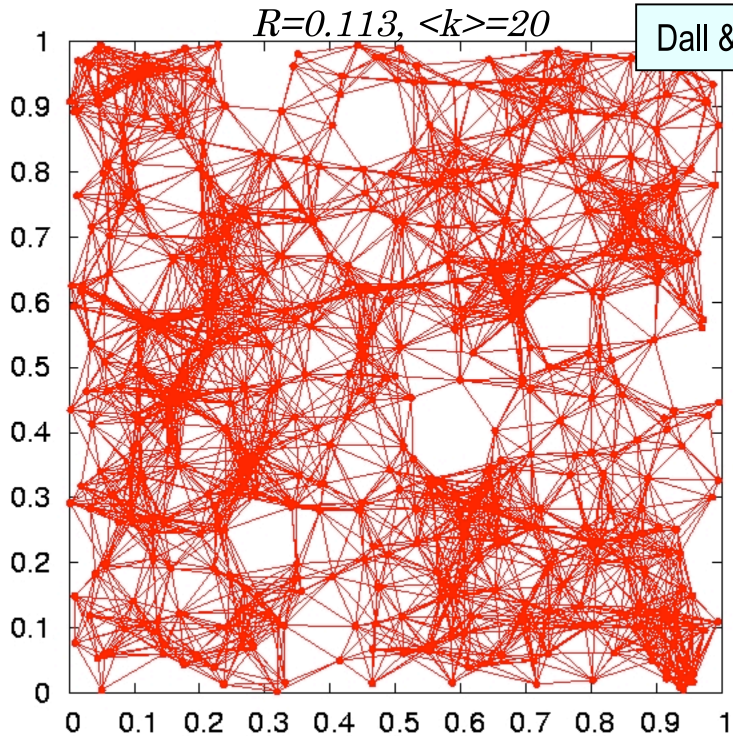
k, E increases

constrained (folded)
small k_{conf}
lower energy

loose (random coil)
large k_{conf}
higher energy

Random geometric graph

- **random geometric graph**



Dall & Christensen, *Phys.Rev.E* **66**, 026121 (2002)

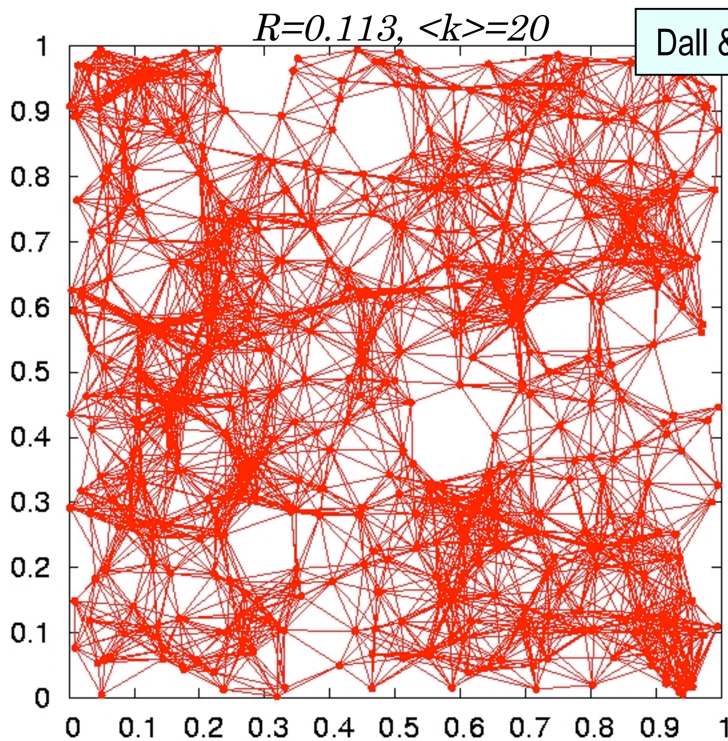
✓ ***in higher D: similar to hypercube with holes***

✓ ***degree correlations***

Random geometric graph

Center for Nonlinear Studies

- random geometric graph

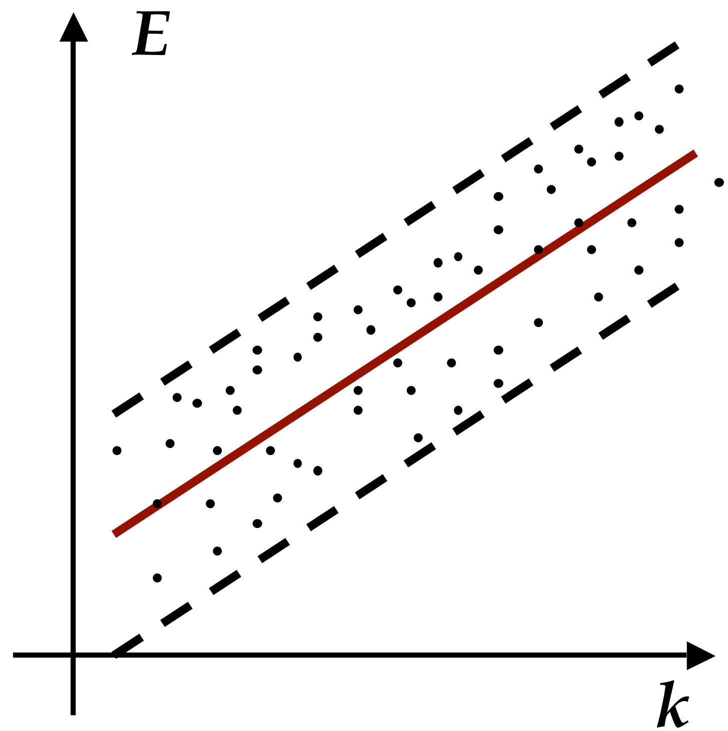


Dall & Christensen, *Phys.Rev.E* **66**, 026121 (2002)

✓ *in higher D: similar to hypercube with holes*

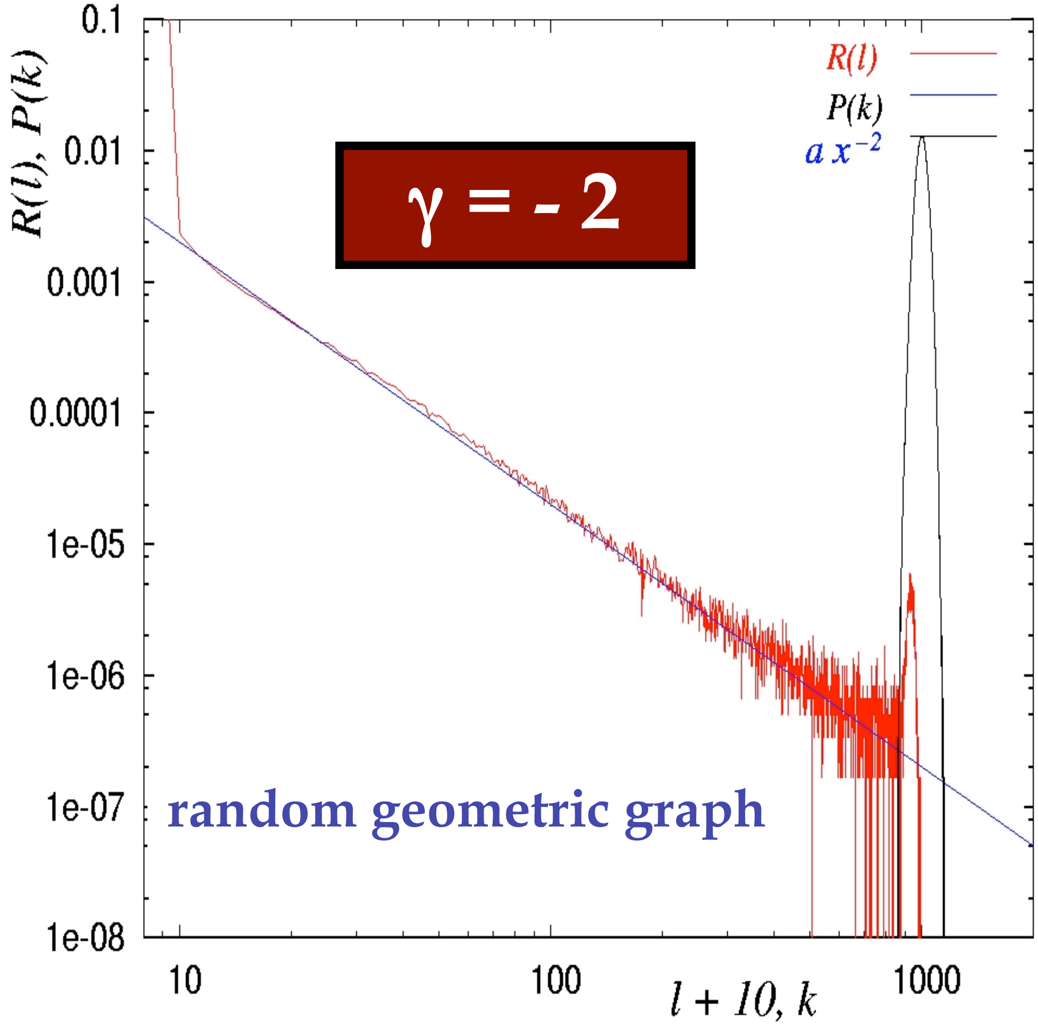
✓ *degree correlations*

- Energy proportional to connectivity



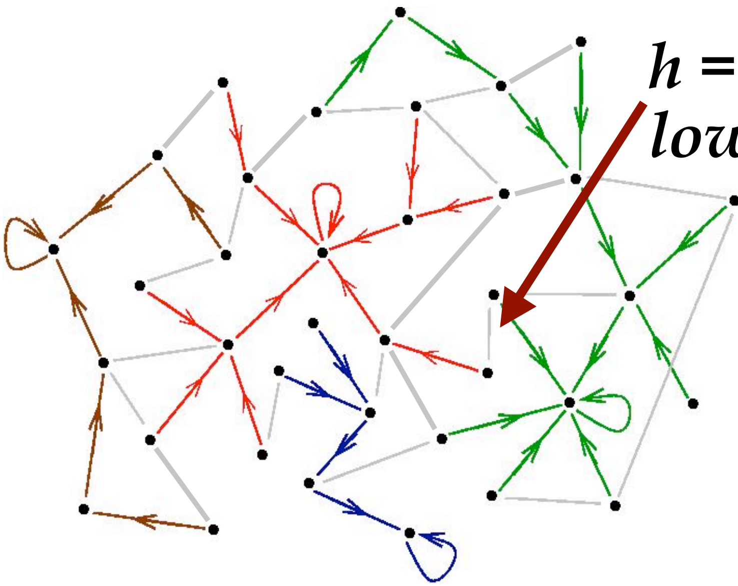
Exponent is - 2

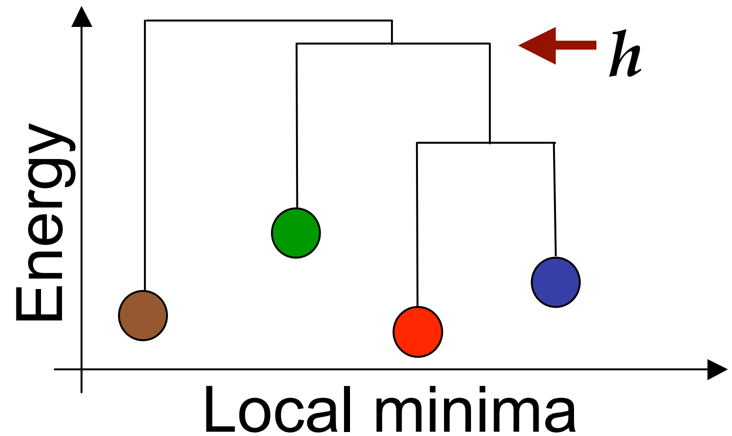
Center for Nonlinear Studies



✓ *monotonic increase of E with k is sufficient*

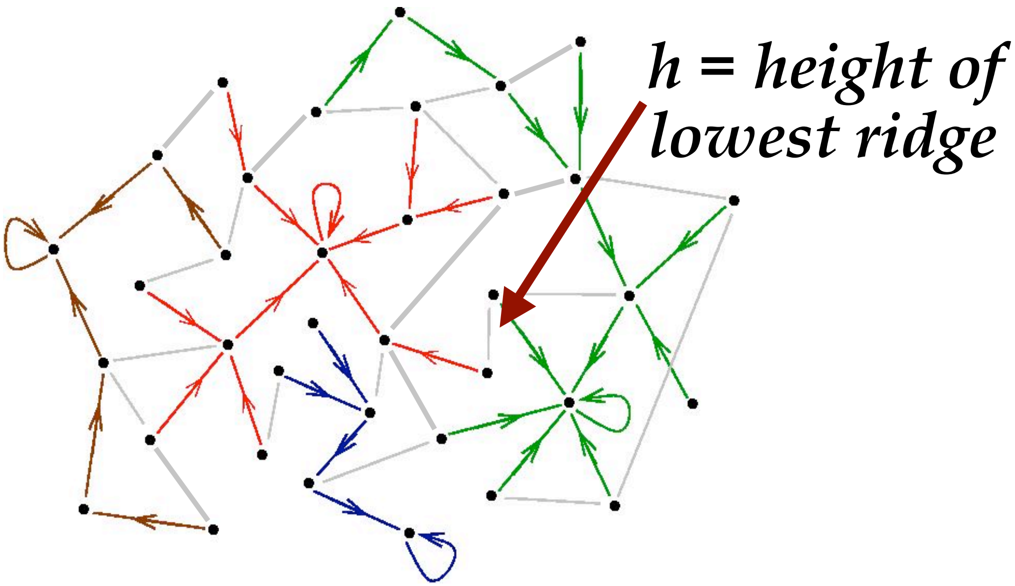
Energy landscape trees

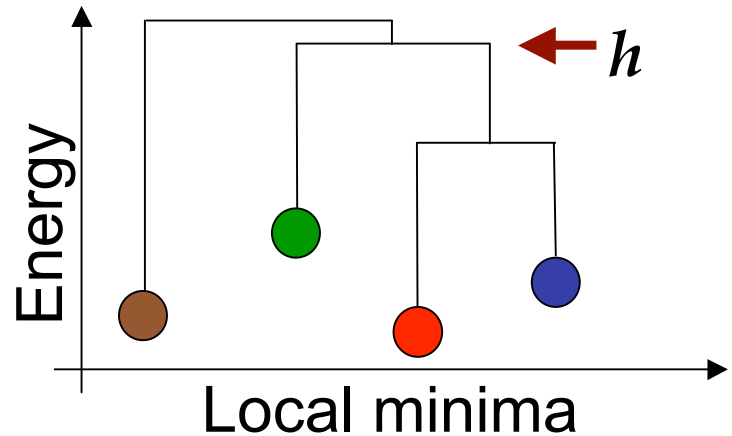




Becker & Karplus, *J.Chem.Phys.* **106**, 149 (1997)

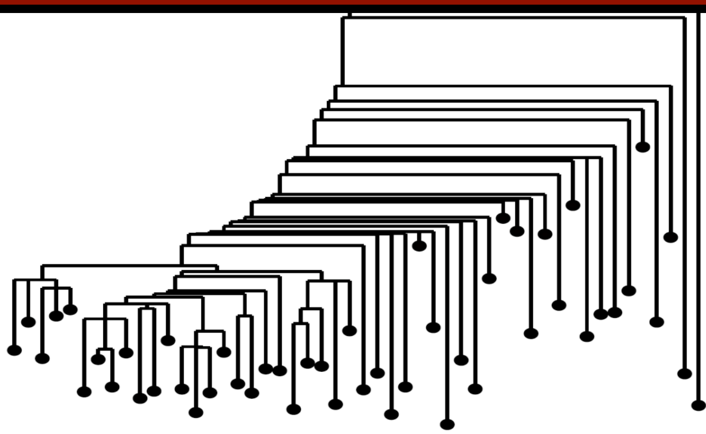
Energy landscape trees



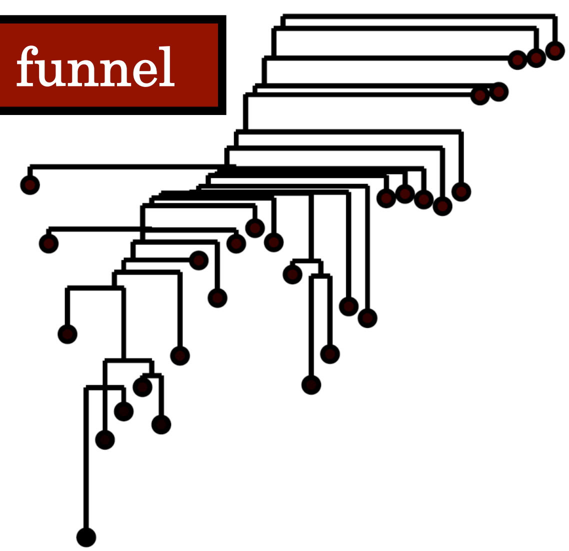


Becker & Karplus, *J.Chem.Phys.* **106**, 149 (1997)

Random energies: golf-course



E ~ k: funnel



RGG, N=5000

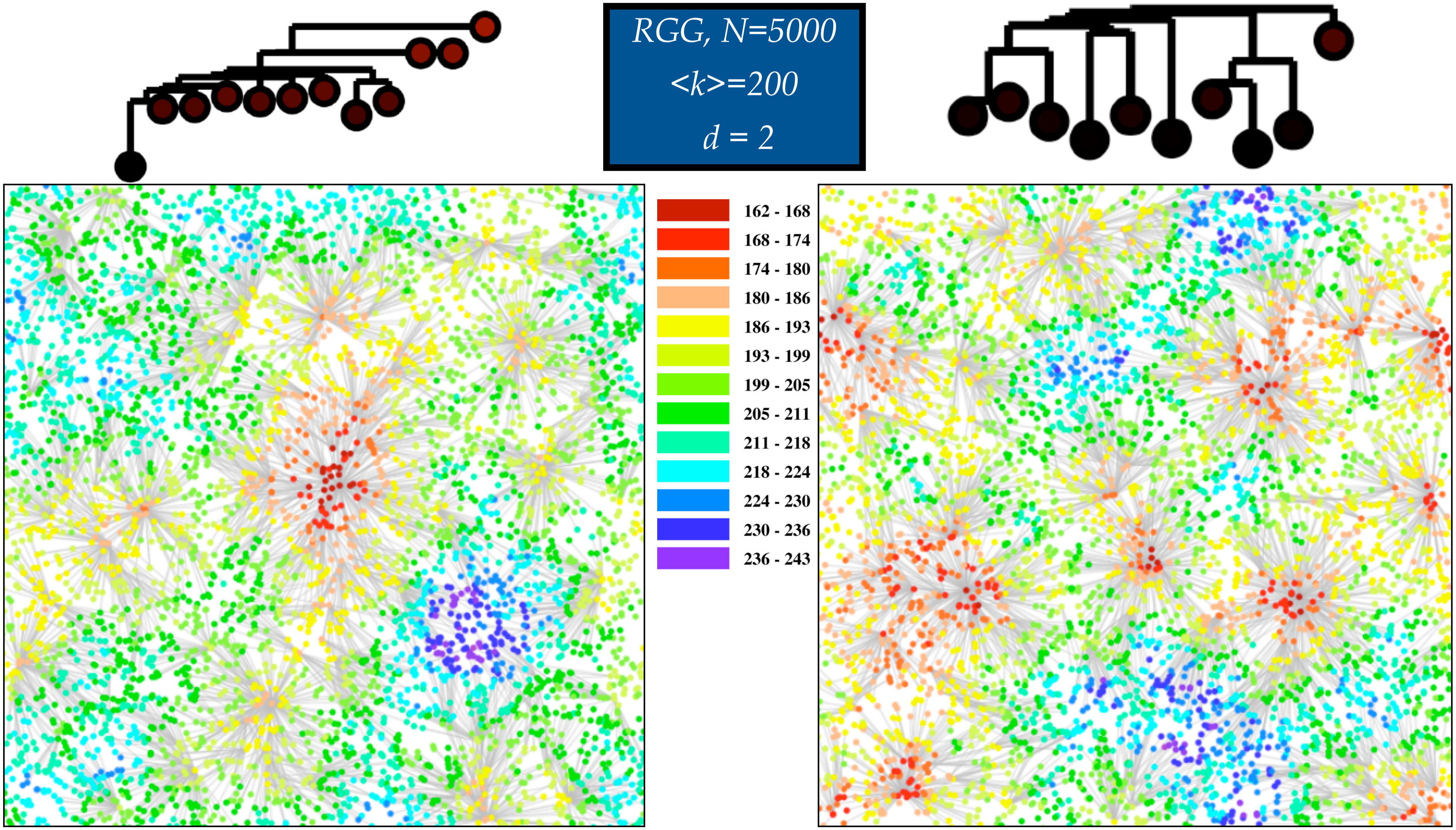
<k>=200

d = 5

Topological funnels

Center for Nonlinear Studies

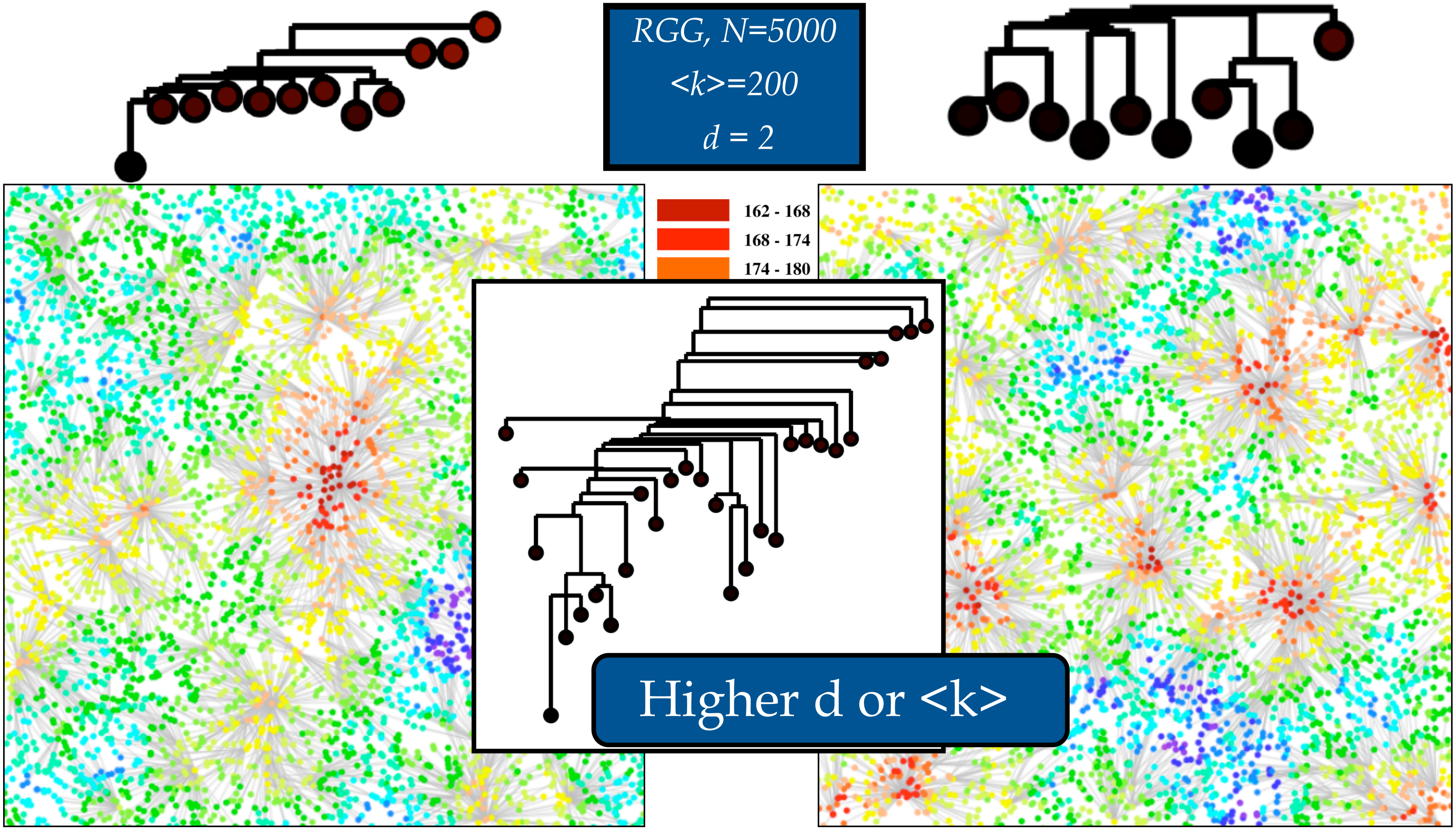
- landscapes on RG graphs with E ~ k



Topological funnels

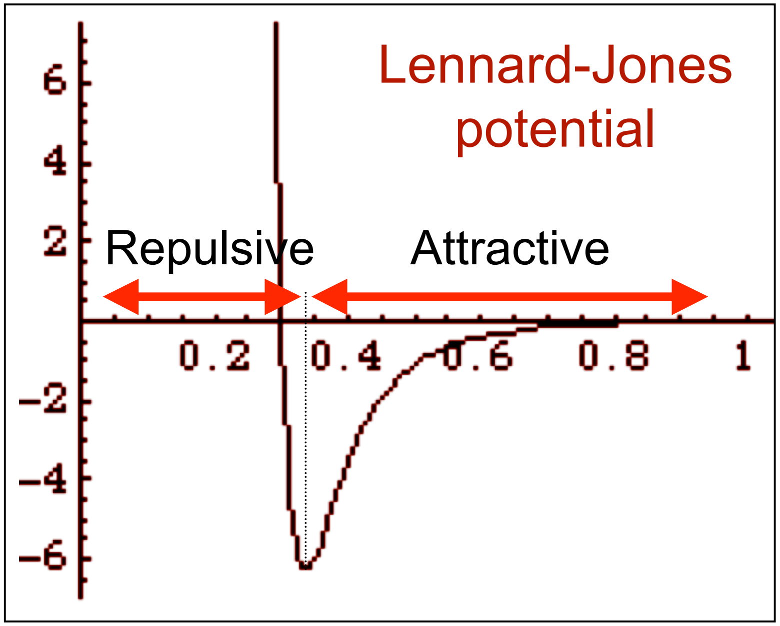
Center for Nonlinear Studies

- landscapes on RG graphs with E ~ k



Bead-chain model

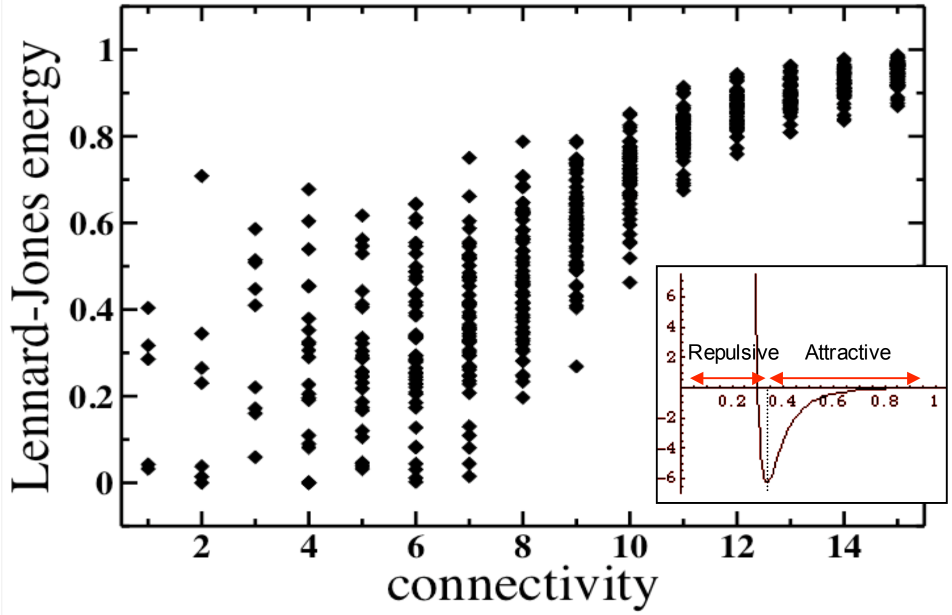
- **more realistic model: bead-chain**
  - ✓ *configuration network*
    - → *excluded volume*
  - ✓ *energy: Lennard-Jones*



Bead-chain model

Center for Nonlinear Studies

- more realistic model: bead-chain
  - ✓ *configuration network*
    - → *excluded volume*
  - ✓ *energy: Lennard-Jones*
  - ✓ *topological funnel*

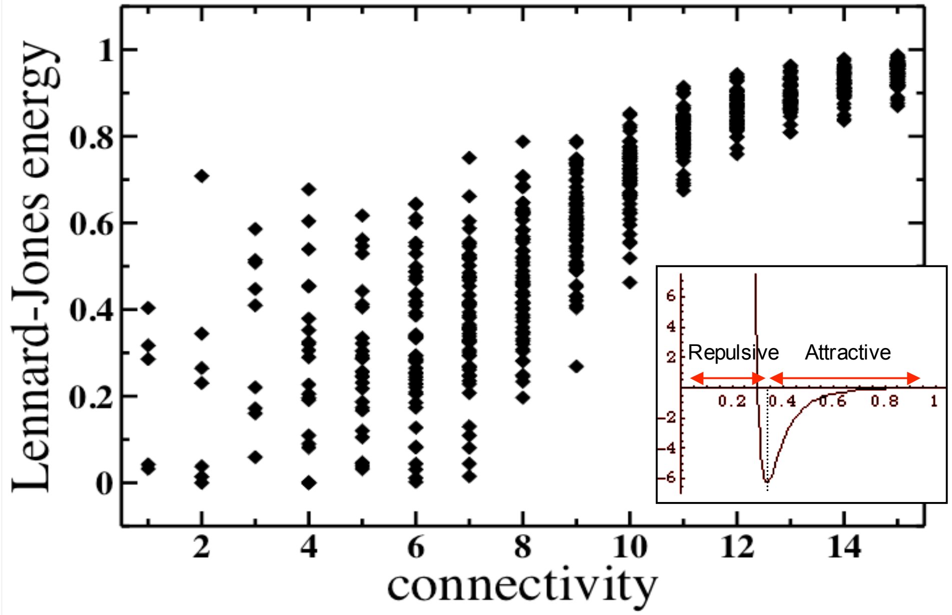


Bead-chain model

Center for Nonlinear Studies

- more realistic model: bead-chain
  - ✓ *configuration network*
    - → *excluded volume*
  - ✓ *energy: Lennard-Jones*
  - ✓ *topological funnel*

NO FUNNEL

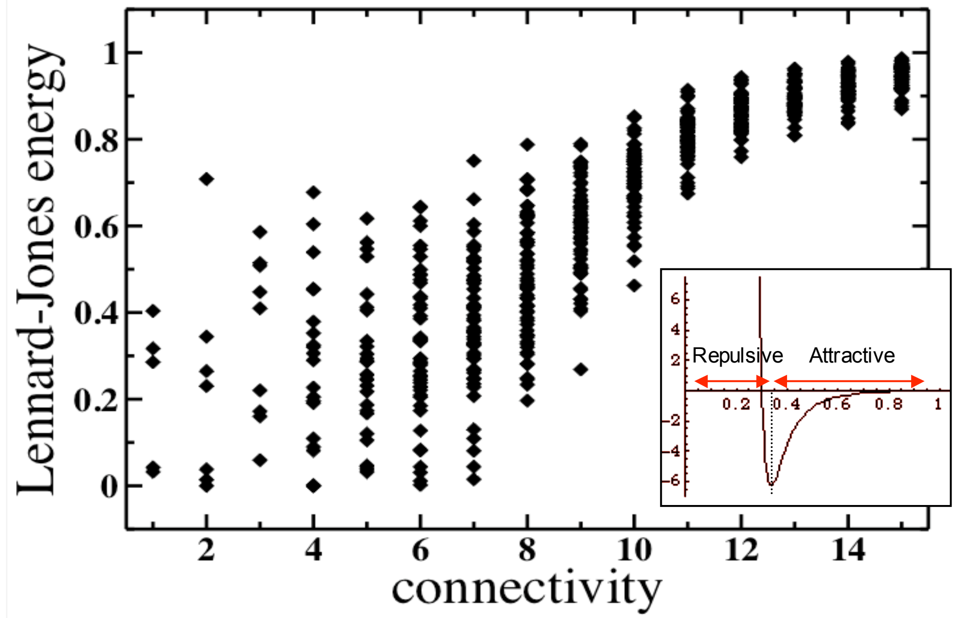


Bead-chain model

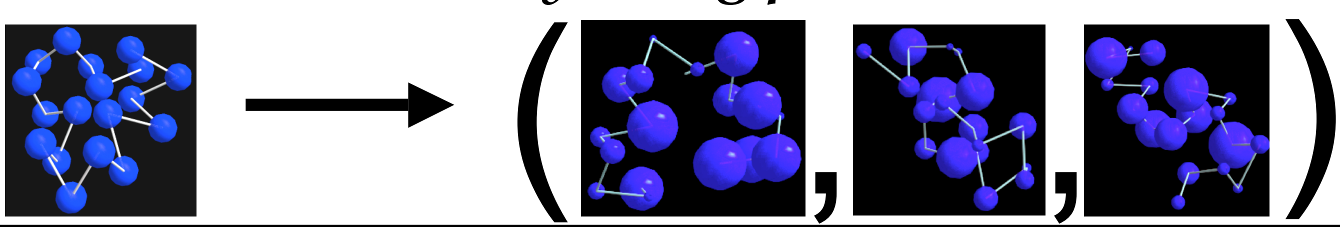
Center for Nonlinear Studies

- more realistic model: bead-chain
  - ✓ *configuration network*
    - → *excluded volume*
  - ✓ *energy: Lennard-Jones*
  - ✓ *topological funnel*

NO FUNNEL

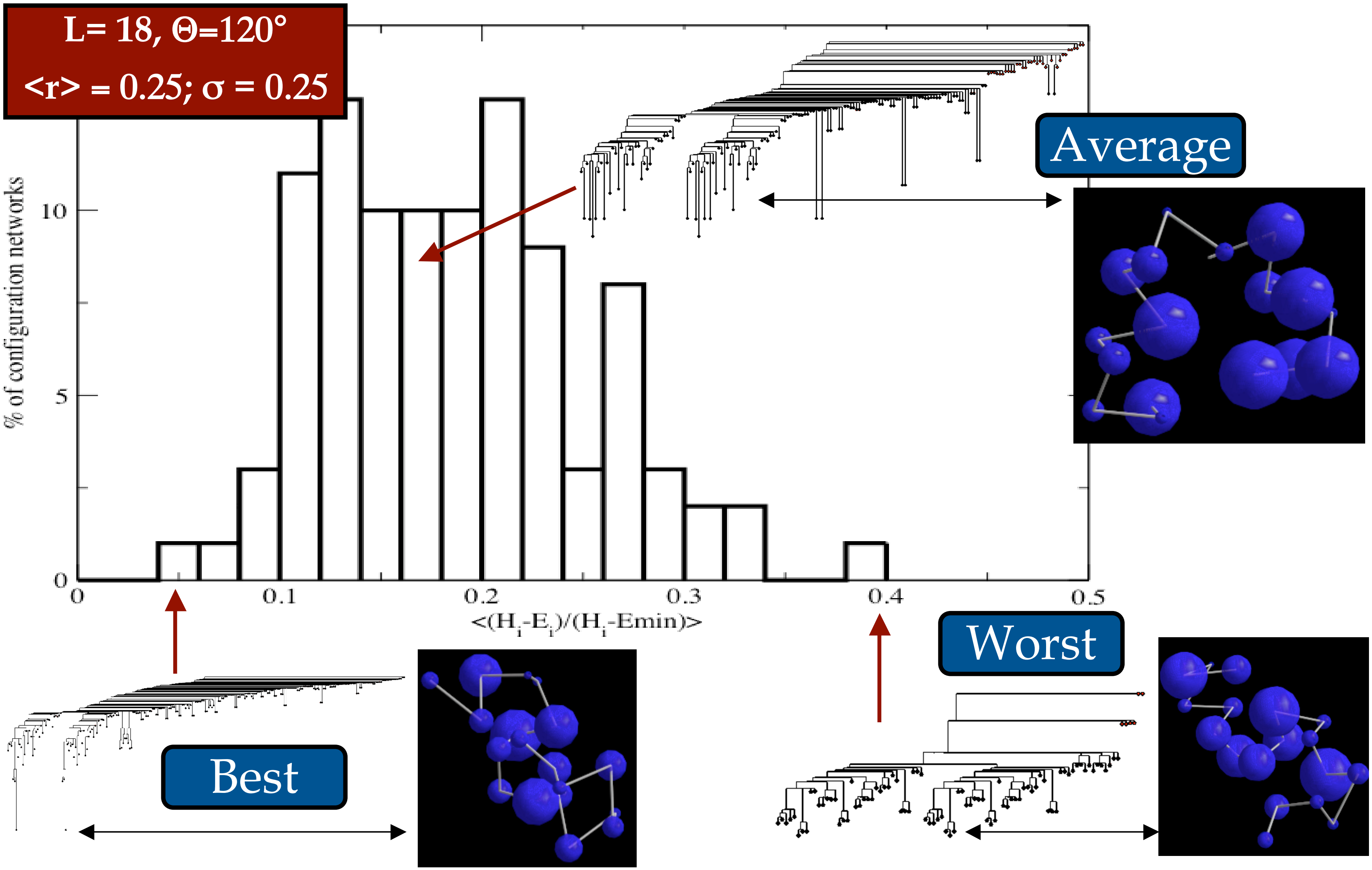


- changing bead sizes (Gaussian distribution)
  - ✓*amino-acid sizes vary along protein chains*



Funnels in the BC model

Center for Nonlinear Studies



Mechanisms for funnels

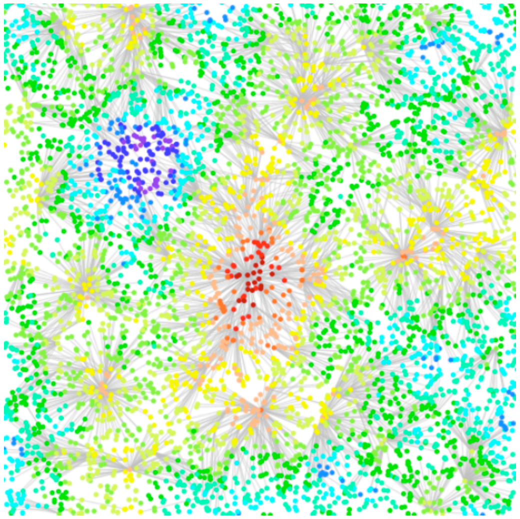
Center for Nonlinear Studies

Topological funnel:

✓ *one low-k basin*

✓ *one way make a tight knot*

✓ *energy follows k*

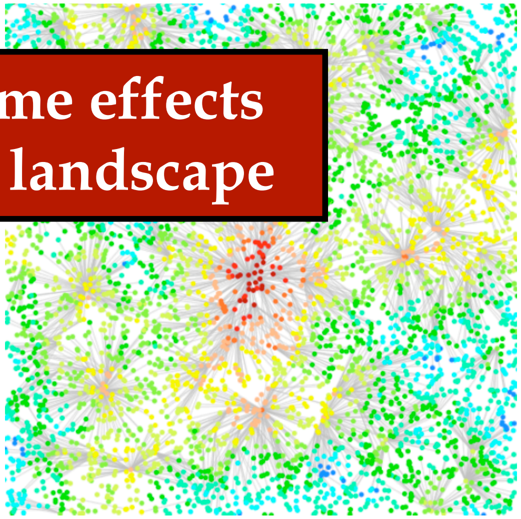


Mechanisms for funnels

Center for Nonlinear Studies

Topological funnel:
- ✓ *one low-k basin*
- ✓ *one way make a tight knot*
- ✓ *energy follows k*

excluded volume effects determine the landscape

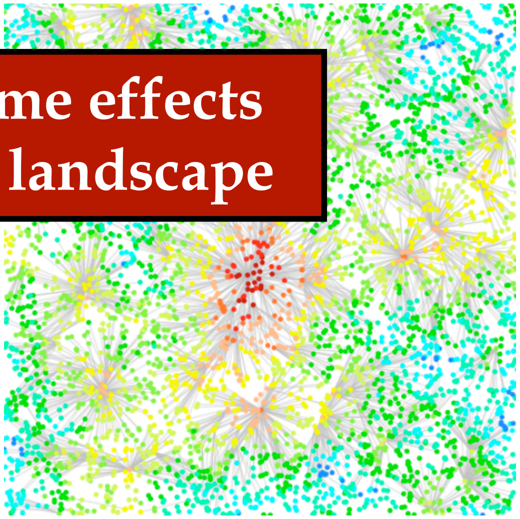


Mechanisms for funnels

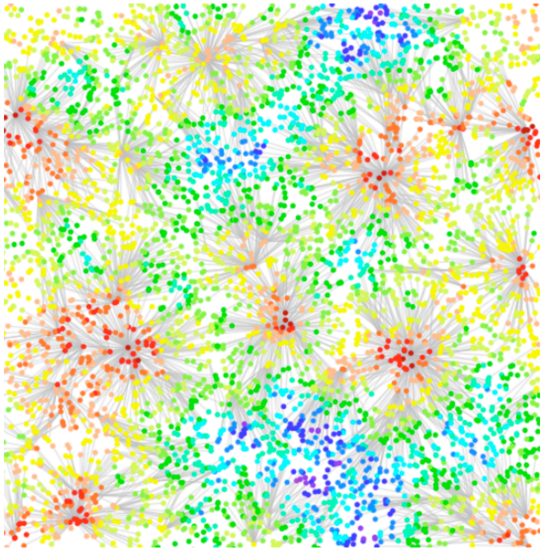
Center for Nonlinear Studies

Topological funnel:
- ✓ *one low-k basin*
- ✓ *one way make a tight knot*
- ✓ *energy follows k*

excluded volume effects determine the landscape



Energy-based funnel:
- ✓ *several tight knots*
- ✓ *many low-k basins*
- ✓ *funnel ⇔ energy is LARGE for most knots*



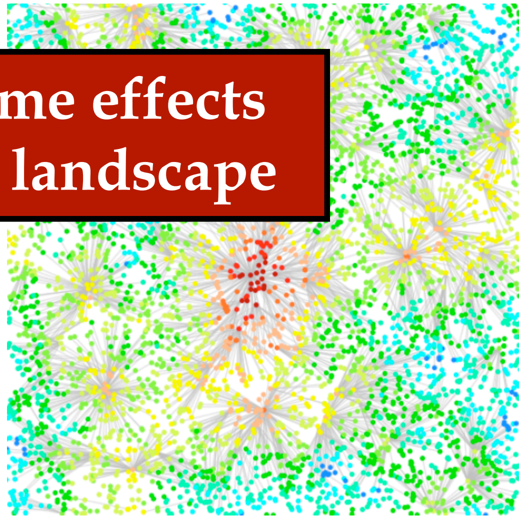
Mechanisms for funnels

Center for Nonlinear Studies

Topological funnel:

- ✓ *one low-k basin*
- ✓ *one way make a tight knot*
- ✓ *energy follows k*

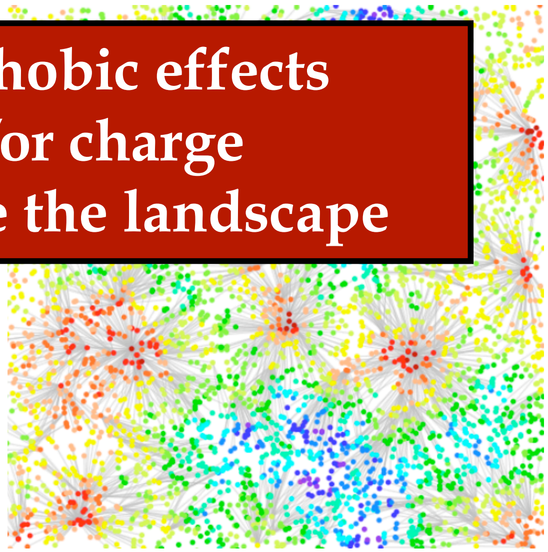
excluded volume effects determine the landscape



Energy-based funnel:

- ✓ *several tight knots*
- ✓ *many low-k basins*
- ✓ *funnel ⇔ energy is LARGE for most knots*

hydrophobic effects and/or charge determine the landscape

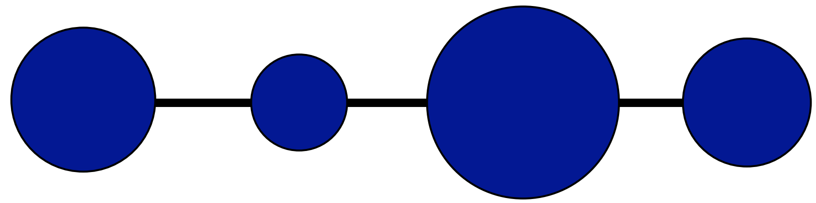


Ongoing work

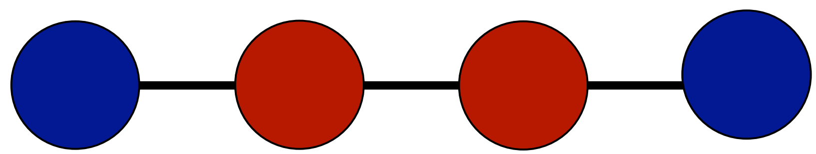
Center for Nonlinear Studies

- **which mechanism is (more) used in nature?**

  ✓ *statistical properties of sizes on BC chains → funnels*



  ✓ *BC model with charges → learn about the energy-only funnels*



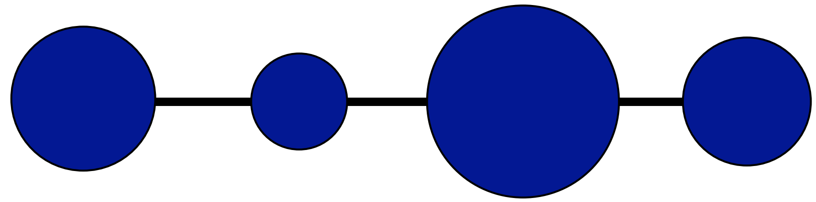
  ✓ *look for these rules in real proteins*

Ongoing work

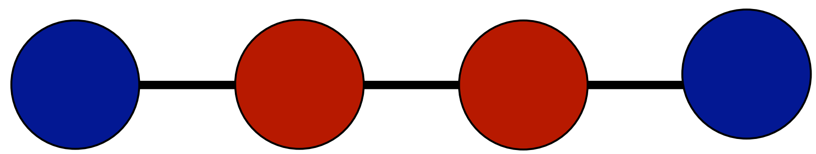
Center for Nonlinear Studies

- **which mechanism is (more) used in nature?**

✓ *statistical properties of sizes on BC chains → funnels*



✓ *BC model with charges → learn about the energy-only funnels*



✓ *look for these rules in real proteins*

Statistical rule
amino-acid sequence ↔ folding

Conclusions

Center for Nonlinear Studies

- **New framework to look at protein energy landscapes**

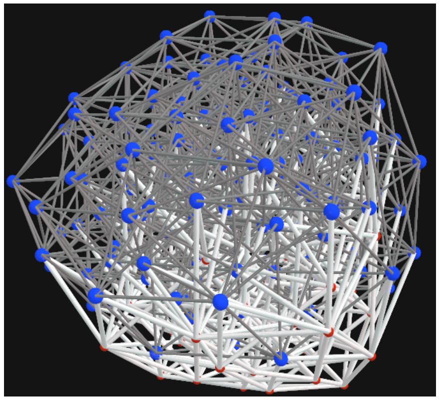
  ✓ ***basic topology: "Swiss-cheese" networks***

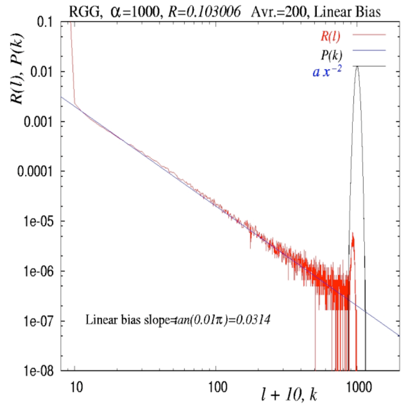
  ✓ **folding:** ***biased random walk on the configuration network***

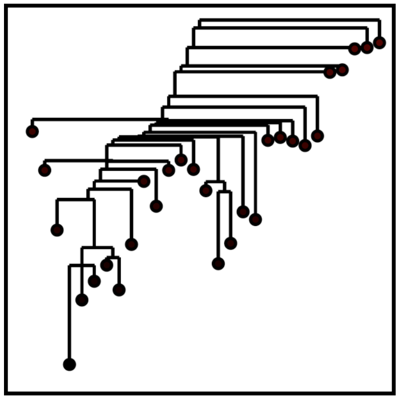
  → ***leads to scale-free flow networks (MD)***

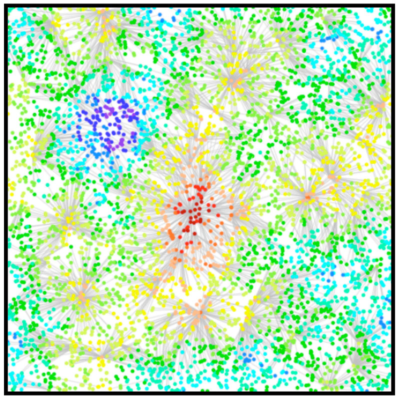
  ✓ ***funnel formation can be modeled***

  → ***strong dependence on network topology***









Thank you!