A 20th Century Physics No-no Problem:

The Reasons of Us Being

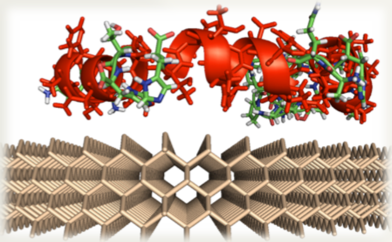
Michael Bachmann

Soft Matter Systems Research Group, The University of Georgia, Athens GA (USA)

Universidade Federal de Mato Grosso, Cuiabá • Universidade Federal de Minas Gerais, Belo Horizonte (Brazil)


Ebola

Global Research



Joint NTZ-SFB/TRR 102 Colloquium, University of Leipzig, 11/27/2014

1785 University of Georgia CSP Center for Simulational Physics

www.smsyslab.org SOFT MATTER SYSTEMS RESEARCH GROUP

SOFT MATTER SYSTEMS RESEARCH GROUP BONSAI Bio-Organic Nucleation and Self-Assembly at Interfaces

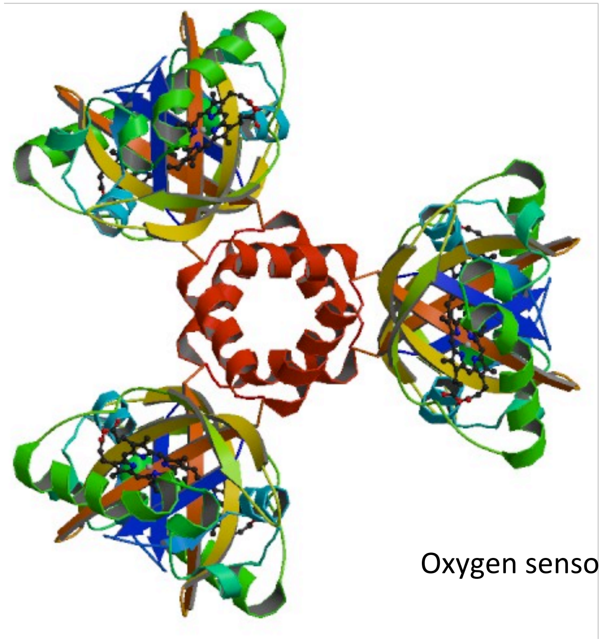
Overview

1. Introduction: Bioproteins

2. Modeling and Simulation

3. Applications

4. Summary



Oxygen sensor

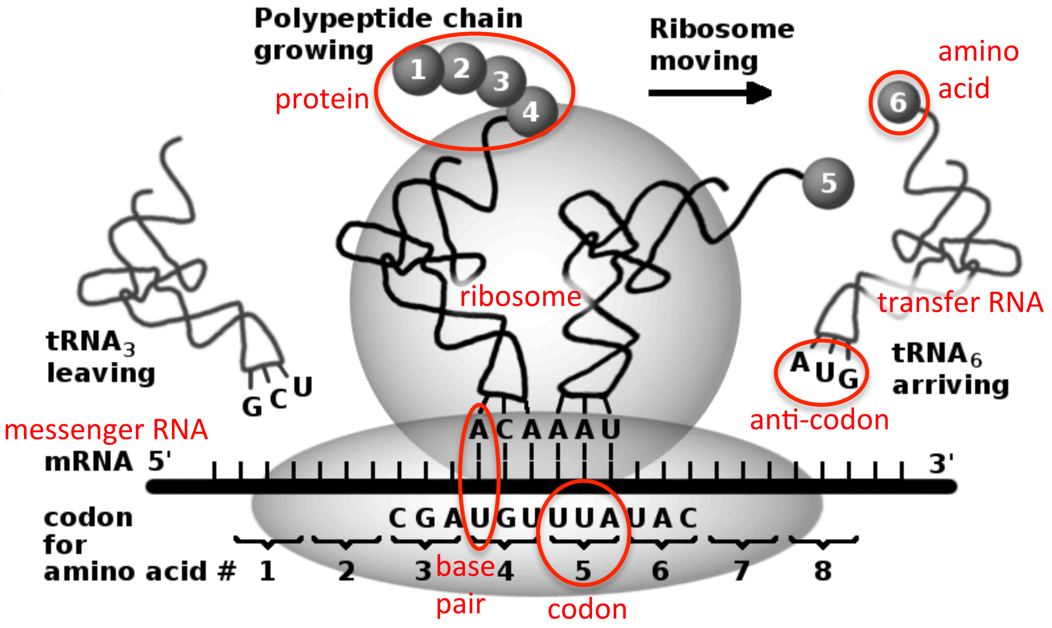
1. Introduction: Bioproteins

Sequence
↓
Conformation
↓
Function

Ribosomal synthesis of amino acid sequences - but not of functional geometric structures!

Proteins: Amino acid sequences encoded in the DNA

- **Polypeptides:**
  $N \sim 20 - 27000$
  amino acids linked by peptide bonds
- **20 amino acids:**
  20^N possible sequences
- **But:** only $\sim 10^6$ functional proteins (human)

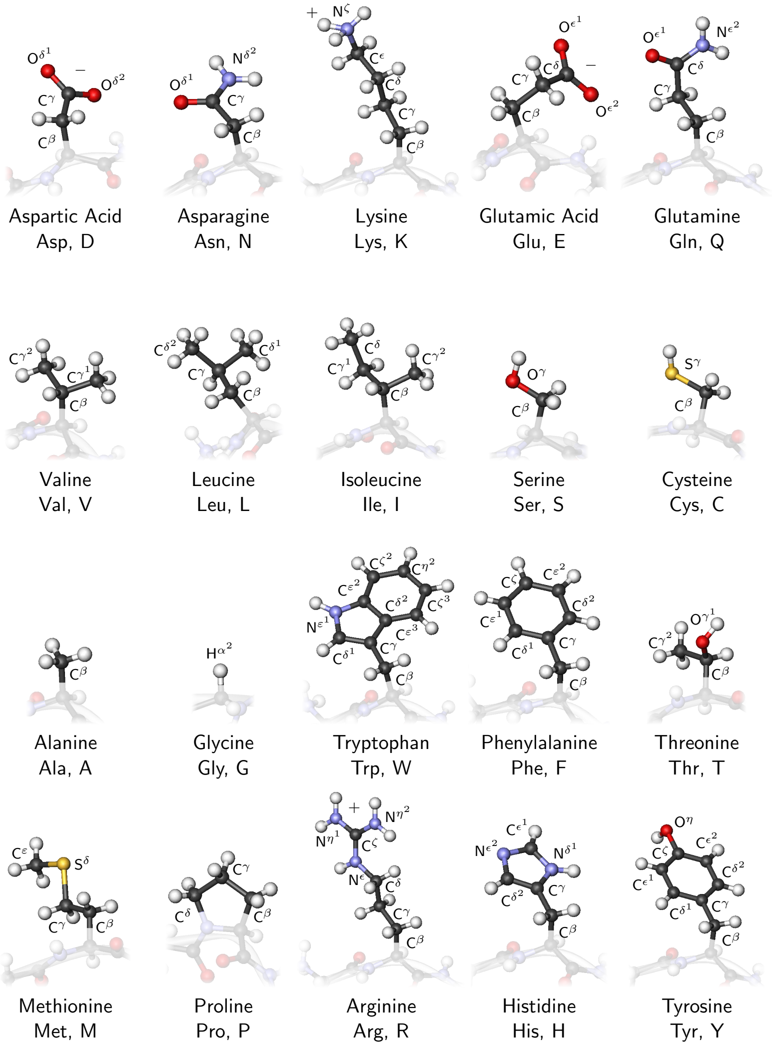


4 base pair combinations, 3 per codon: 64 codes

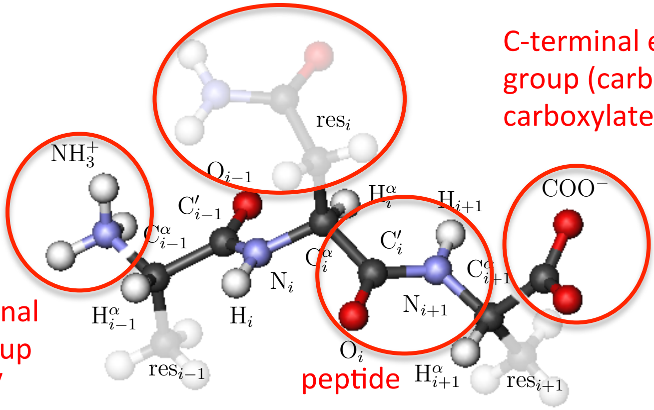
1. Introduction: Bioproteins

Sequence
↓
Conformation
↓
Function

20 typical amino acids side chains



Protein backbone



side chain (residue)

C-terminal end group (carboxylic/ carboxylate group)

N-terminal end group (amino/ ammonium group)

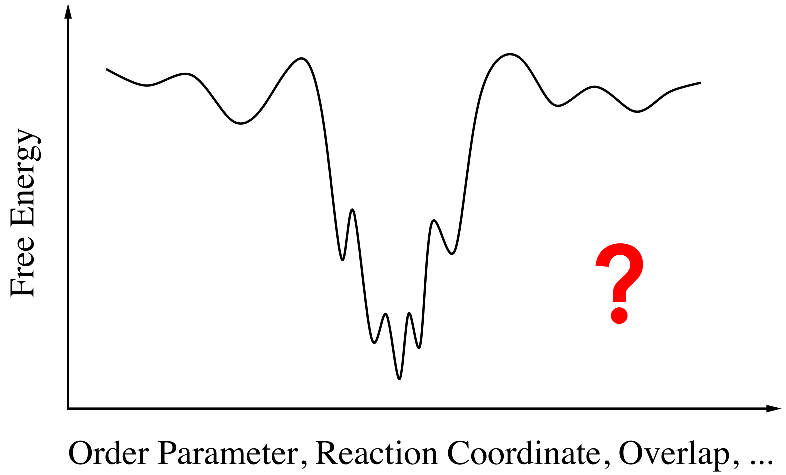
peptide group

1. Introduction: Bioproteins

Sequence
↓
Conformation
↓
Function

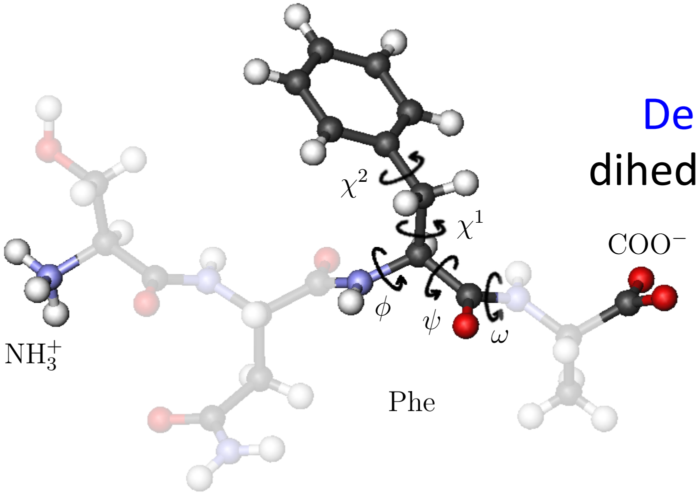
Sequence determines structure.

Free-energy landscape: rugged, funnel-like, with pockets



(which are most appropriate and how many are needed?)

- Spontaneous folding into native conformation (Anfinsen's experiments)
- Folding pathway(s): free-energy driven stochastic search
- Protein folding time scale: milliseconds to seconds
- Metastability



Degrees of freedom: dihedral and side-chain torsion angles

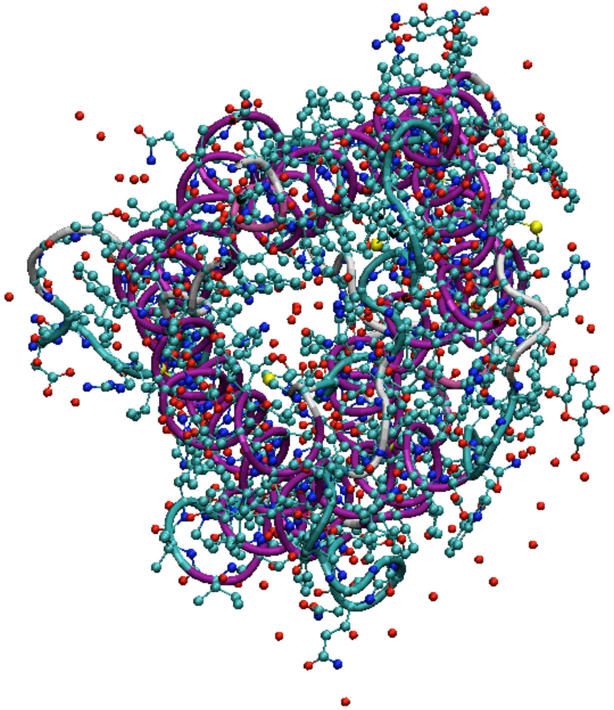
$$\phi, \psi, \omega, \chi^1, \chi^2, \ldots$$

1. Introduction: Bioproteins

Sequence
↓
Conformation
↓
Function

3D protein structure determines functionality.

Example: Aquaporin
transfer of 3,000,000,000 water molecules / sec.



Proteins are involved in virtually all cell processes.

- Transport (channels, pores)
- Cell stability (actin filaments)
- Catalytic activity (enzymes)
- Molecular motors (muscle contraction, DNA polymerase, ATP synthase)
- Epidemic diseases (Alzheimer’s, Huntington’s, virus infections, etc.)

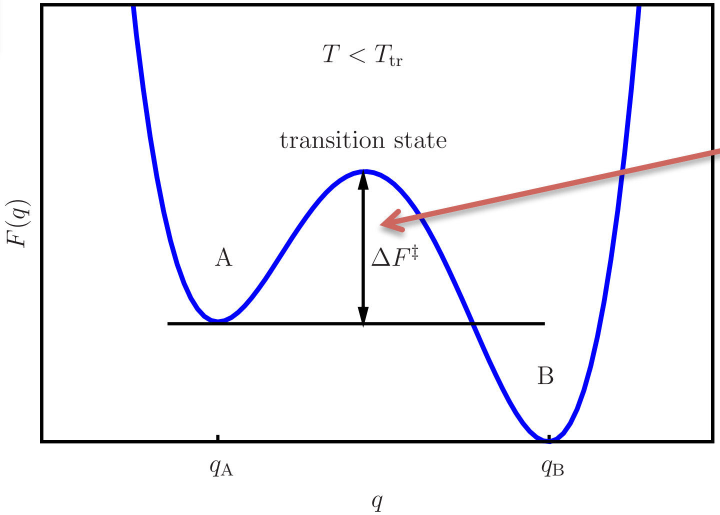
1. Introduction: Bioproteins

Reasons for nature’s preference of macromolecular matter:

- ✓ Small energetic margins allow for complex processes on nanometer scales
- ✓ Small number of degrees of freedom (torsion angles), but large number of constraints (bond lengths, bond angles)
- ✓ Sequential disorder (amino acid sequence), large number of possibilities, easy to encode (RNA)
- ✓ Formation of stable, but flexible molecular structures on mesoscopic scales
- ✓ Preference of finite-size first-order-like transitions (phase separation rather than phase coexistence ➔ entropic suppression: “surface entropy”)
- ✓ Broken long-range symmetries, but long-range cooperativity (amorphous, glasslike) in molecular structures
- ✓ Production of identical macromolecules in huge numbers, statistics
- ✓ Process efficiency (signal exchange, transport, catalysis, etc.)
- ✓ Simple strategies for adaptation, regulation, and genetic evolution

1. Introduction: Bioproteins

The life of a protein



$5 - 10\, k_{\mathrm{B}} T_{\mathrm{room}}$

functional
(after about 1s)

ubiquitination
(prep. recycling)

birth, folding
(synthesis)

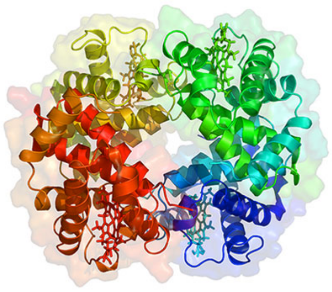
occasional
unfolding/refolding
(at a rate of O(10) s/day)

death, recycling
of amino acids
(after a few weeks)

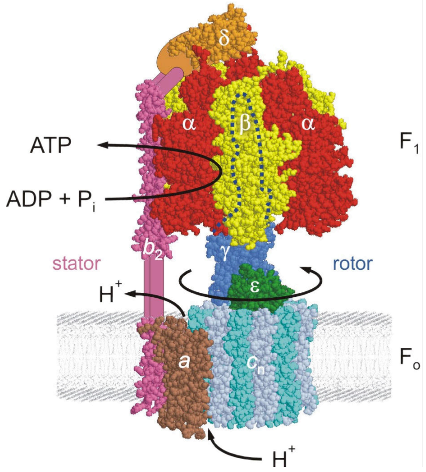
1. Introduction: Bioproteins

Challenges:

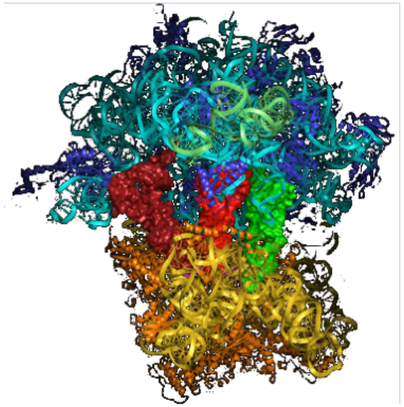
- Single protein: Heterogeneous chain composed of up to 27,000 amino acids
- Geometric structure ⇔ Biological function
- Structure formation process ⇔ Structural phase transition on finite scales?
- But: no thermodynamic limit, no scaling, no (unique) transition points
- Finite-size, surface, and disorder effects



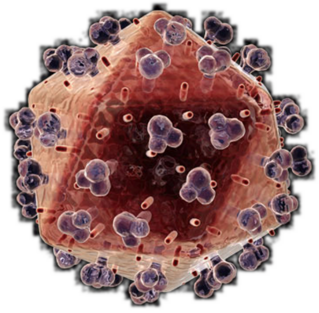
Hemoglobin
Oxygen transport in red blood cells,
4 units, >2,000 atoms



ATP synthase
Production rate 100 ATP/s,
2 units, 40,000 atoms



Ribosome
Protein synthesis,
2 units, 200,000 atoms



Herpes
Icosahedral virus,
10^5-10^6 atoms

1. Introduction: Bioproteins

The Dark Side: Proteomic Diseases

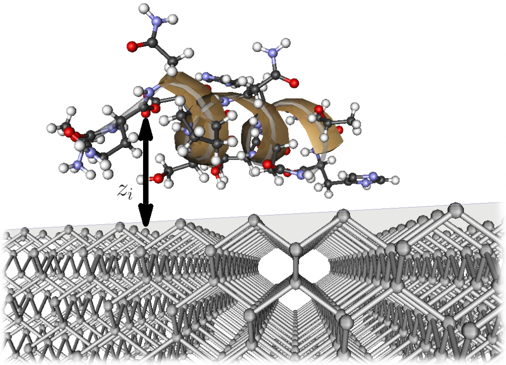
- **A**cquired **I**mmuno-**D**eficiency **S**yndrome (**H**uman **I**mmunodeficiency **V**irus)
  >30 million infected people worldwide (WHO); incurable
- Hepatitis B (**H**epatitis **B** **V**irus)
  2 billion concerned, >350 million chronic (WHO), incurable
- **B**ovine **S**pongiform **E**ncephalopathy (prionic disease, cattle)
  US beef export: 2003: $3.8 billion, 2005: $1.4 billion (USA Today)
- **C**reutzfeld-**J**akob **D**isease (prionic, human)
  one sick per 1 million people; incurable
- **A**lzheimer's **D**isease (amyloid peptide cluster)
  65 years old: 2-3%, at the age of 85: 25-50%
  24 million sick worldwide (AFI e.V.); incurable
- Influenza (virus)
  bird flu, swine flu, ...
- Ebola (virus)
  death rate 50-90% (2014: >1,500), incurable

2. Modeling and Simulation

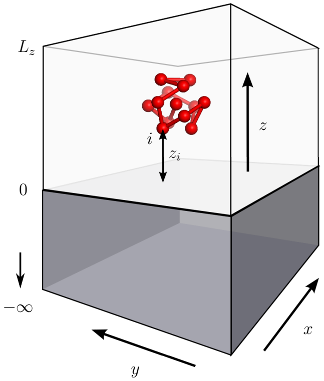
Basic problem: Mesoscopic length scales, 10^1-10^3 nm (proteins)

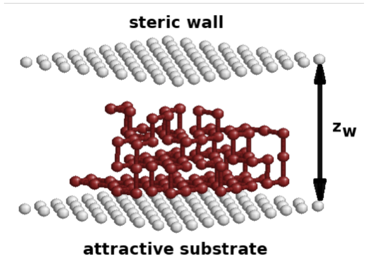
Two major approaches to macromolecular modeling:

- **Semi-empirical, semi-classical all-atom models**
  10^2-10^3 parameters (!)
  attempting to model short- and long-range interactions (including quantum effects) and solvent effects as accurately as possible, aiming at **specific questions** (e.g., short-time dynamics of folding and function: ns to µs)



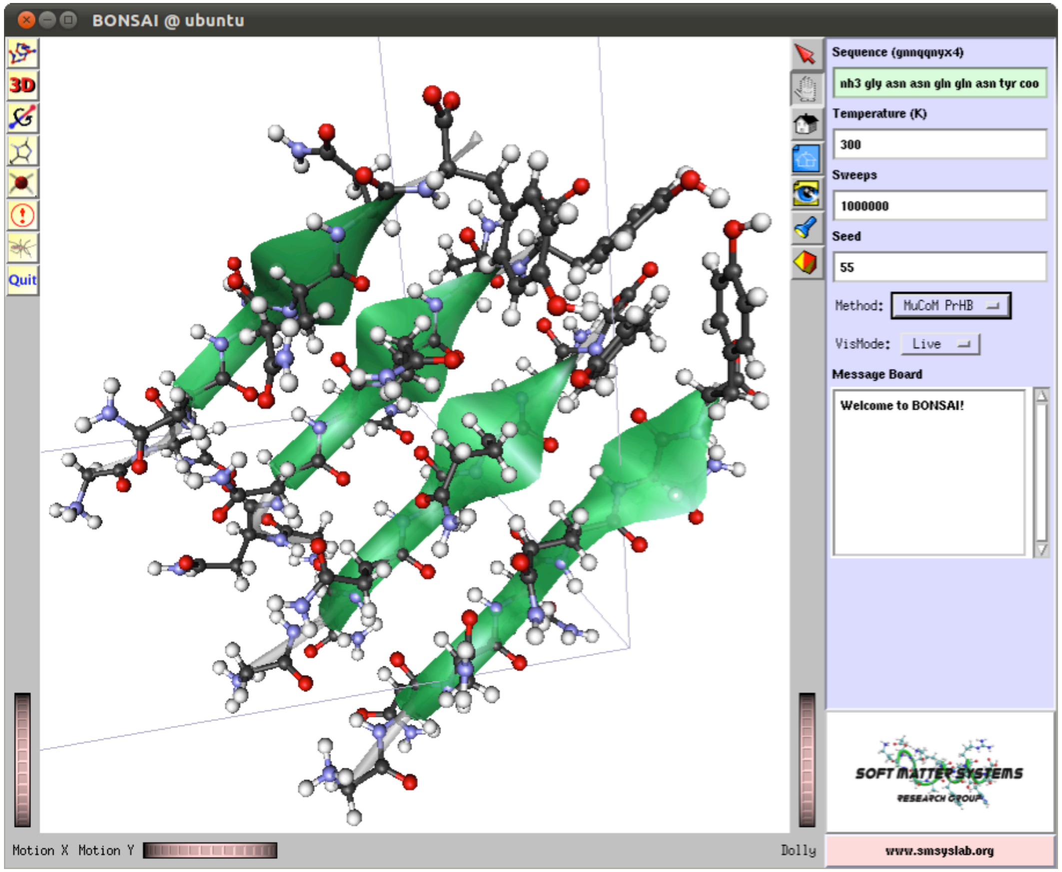
- **Coarse-grained, effective monomeric models**
  < 10 parameters
  extreme abstraction, concentration on **generic properties** such as structural transitions
  → only way to study **complexity**





2. Modeling and Simulation

All-atom models (using simulation packages: CHARMM, Amber, Gromacs, etc.) ...



Sequence (gnnqqnyx4): nh3 gly asn asn gln gln asn tyr coo

Temperature (K): 300

Sweeps: 1000000

Seed: 55

Method: MuCoM PrHB

VisMode: Live

Message Board: Welcome to BONSAI!

BONSAI

Bio-Organic Nucleation and Self-Assembly at Interfaces

Monte Carlo Sim Pack

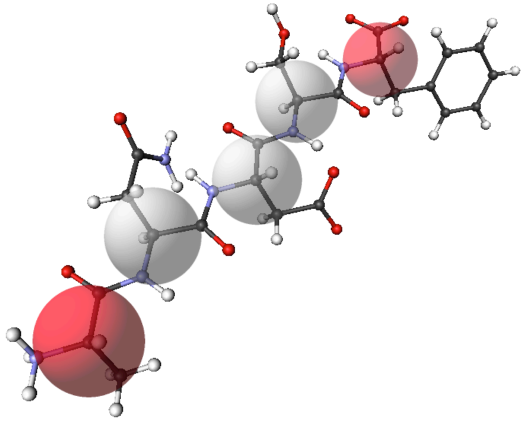
- ⇒ Protein folding
- ⇒ Protein aggregation
- ⇒ Protein adsorption

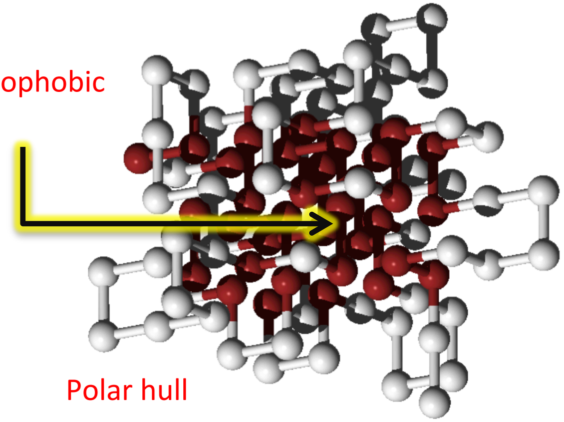
- ⇒ C/C++ modularity
- ⇒ OpenInventor, Coin3D
- ⇒ Multiple sim threads

[www.smsyslab.org/bonsai.html]

2. Modeling and Simulation

... or coarse-grained models for effective non-bonded interactions:

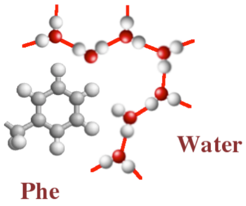




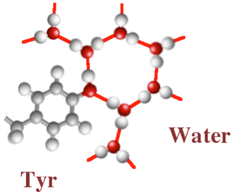
"Hydrophobic-polar" proteins:

20 amino acids ⇒ 2 major groups:

● Hydrophobic (water-repellent)



○ Polar, hydrophilic

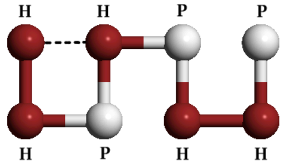


HP model

("Ising model of lattice proteins"):

$$E(\mathbf{X}) = -\varepsilon_{\mathrm{HH}} n_{\mathrm{HH}}(\mathbf{X})$$

$\mathbf{X}$	conformation (coordinates)
$\varepsilon_{\mathrm{HH}}$	H-H energy scale
n_{HH}	number of H-H contacts



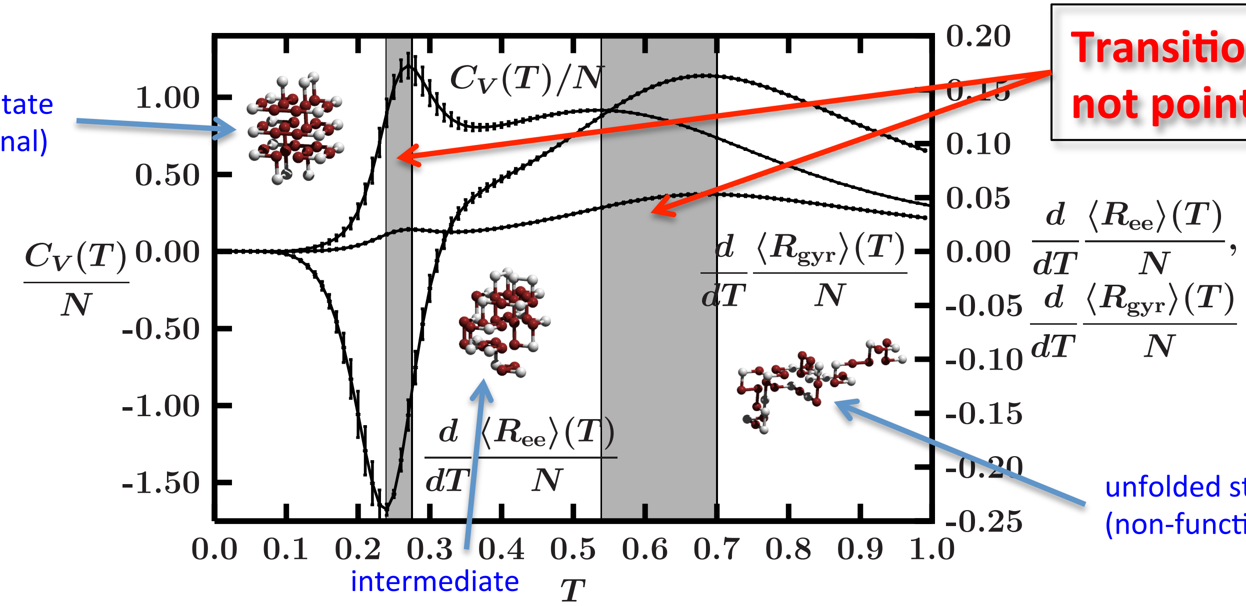
[lattice (HP model): K. F. Lau, K. A. Dill, Macromolecules **22**, 3986 (1989); off-lattice (AB model): F. Stillinger, T. Head-Gordon, C. L. Hirshfeld, Phys. Rev. E **48**, 1469 (1993).]

2. Modeling and Simulation

HP model (and other toy models for proteins)

- cannot address specific questions about proteins
- **but:** some value for studies of more general aspects of macromolecular structure formation

Thermodynamics of protein folding



folded state
(functional)

Transition bands – not points!

intermediate
(non-functional)

unfolded state
(non-functional)

[M.B., W. Janke, Phys. Rev. Lett. **91**, 208105 (2003); J. Chem. Phys. **120**, 6779 (2004).]

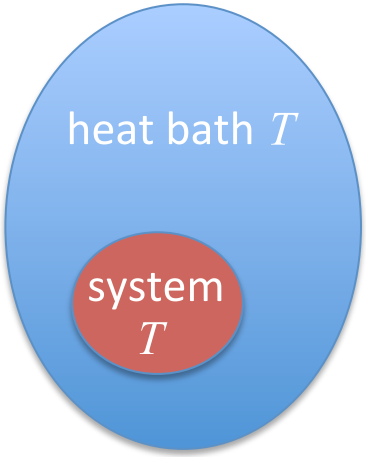
2. Modeling and Simulation

Transition bands: Undesired ambiguity in temperature space

What is temperature?

typically introduced in *macroscopic environments* as:

- an easy-to-control external parameter in experiment
- a canonical equilibrium property of a system in a heat bath
- a constant in Boyle-Mariotte's *ideal gas* law (17th century!!): isothermal process $pV = \text{const}$ (later Clapeyron: $pV = nRT$) ⇒ measured with thermometers



So, asking more precisely: What is the temperature of a finite system?

- can use the microcanonical equilibrium property of two systems in contact:

$$T_1 = \left(\frac{\partial S_1(E_1)}{\partial E_1}\right)^{-1} = \left(\frac{\partial S_2(E_2)}{\partial E_2}\right)^{-1} = T_2$$

S_i microcanonical entropy

E_i energy

Important: Microcanonical temperature is system property!

2. Modeling and Simulation

A somewhat flaky proof

total density of states (multiplicative?): $g(E_1, E_2) = g_1(E_1)g_2(E_2)$
total entropy (additive?): $S(E) = k_\mathrm{B} \ln g(E) \;\Rightarrow\; S(E_1, E_2) = S_1(E_1) + S_2(E_2)$
total energy (additive?): $E = E_1 + E_2 \;\Rightarrow\; dE = dE_1 + dE_2$

equilibrium: $dS = 0,\ dE = 0$

Thus:
$$\begin{aligned} dS(E_1, E_2) &= \left(\frac{\partial S}{\partial E_1}\right)_{E_2} dE_1 + \left(\frac{\partial S}{\partial E_2}\right)_{E_1} dE_2 \\ &= \frac{\partial S_1}{\partial E_1} dE_1 + \frac{\partial S_2}{\partial E_2} dE_2 = \left(\frac{\partial S_1}{\partial E_1} - \frac{\partial S_2}{\partial E_2}\right) dE_1 \\ &= 0 \end{aligned}$$

and:
$$\frac{\partial S_1(E_1)}{\partial E_1} = \frac{\partial S_2(E_2)}{\partial E_2} \equiv \frac{\partial S(E)}{\partial E} \;\Rightarrow\; T_1(E_1) = T_2(E_2) \equiv T(E)$$

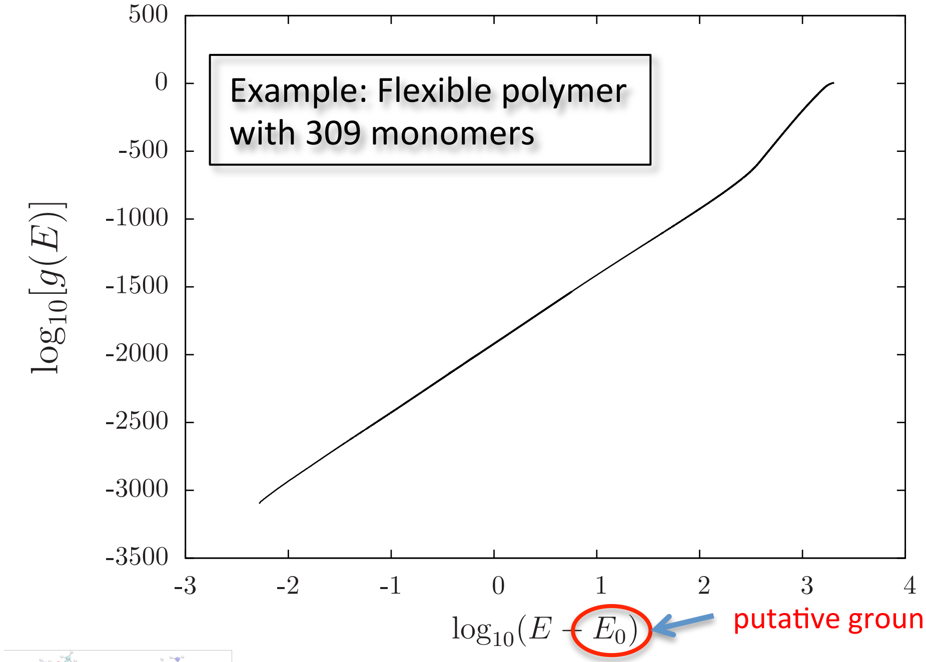
(only valid for dilute gas, neglecting interactions); **we define:** $T(E) := \left(\dfrac{\partial S(E)}{\partial E}\right)^{-1}$

2. Modeling and Simulation

Microcanonical entropy and density of states

Systematic identification and classification of transitions in finite systems:
“Back to the basics”: **Energy and entropy govern cooperative behavior of a system.**

Microcanonical entropy: $S(E) = k_\mathrm{B} \ln g(E)$, where $g(E)$ is the density of states:



Example: Flexible polymer with 309 monomers

putative ground-state energy

$$g(E') = \int \mathcal{D}\mathbf{X}\, \delta(E' - E(\mathbf{X}))$$

No mathematical calculation possible for almost all complex problems

Computational challenge:
Accurate estimation of $g(E)$ over hundreds to thousands of *orders of magnitude*!!

[S. Schnabel, W. Janke, M.B., J. Comput. Phys. **230**, 4454 (2011).]

2. Modeling and Simulation

Microcanonical inflection-point analysis

inverse temperature

$$\beta(E) = T^{-1}(E) = \frac{dS(E)}{dE}$$

$$\gamma(E) = \frac{d^2S(E)}{dE^2}$$

Transitions:

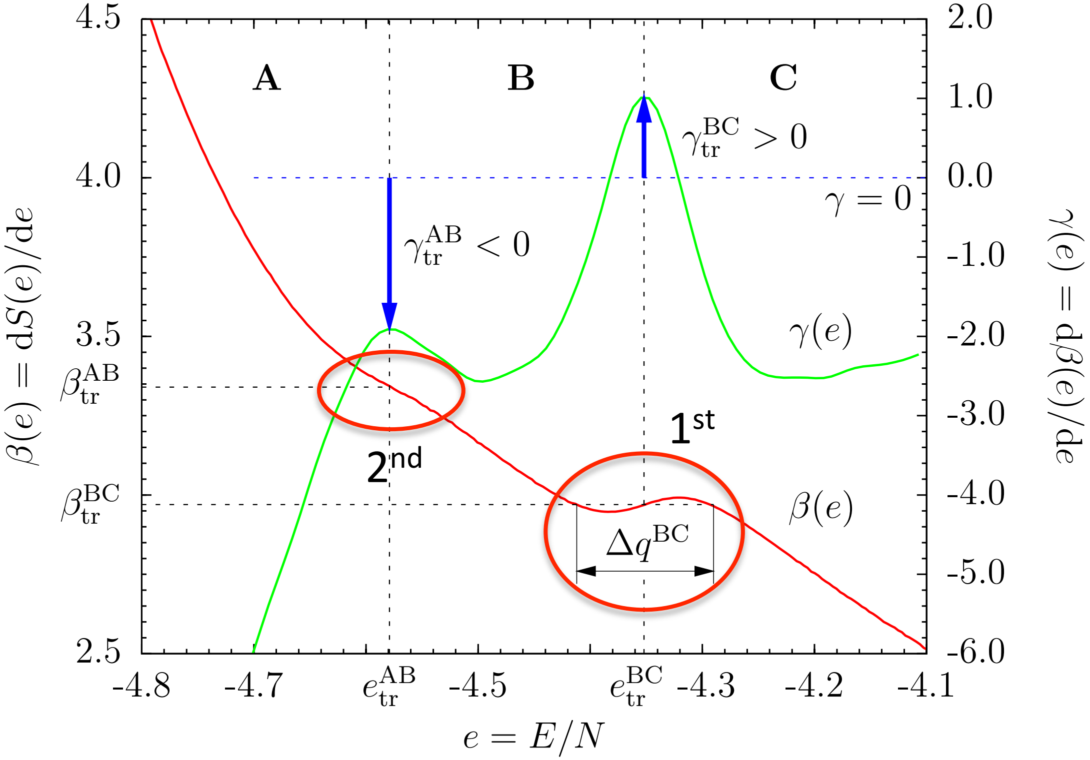
1st order:

$$\gamma_{\rm tr} > 0 \quad (\Delta q_{\rm tr} > 0)$$

2nd order:

$$\gamma_{\rm tr} < 0 \quad (\Delta q_{\rm tr} = 0)$$

$\Delta q_{\rm tr}$ latent heat



Generalization possible: Principle of minimal sensitivity $S^{(n)}(E) = \frac{d^nS(E)}{dE^n} = 0$

[S. Schnabel, D. T. Seaton, D. P. Landau, M.B., Phys. Rev. E **84**, 011127 (2011).]

2. Modeling and Simulation

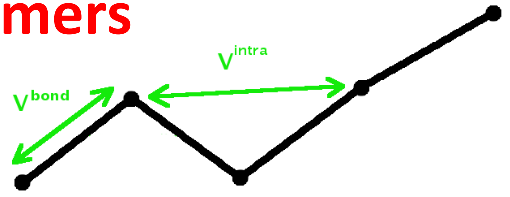
Advanced Monte Carlo simulation techniques:

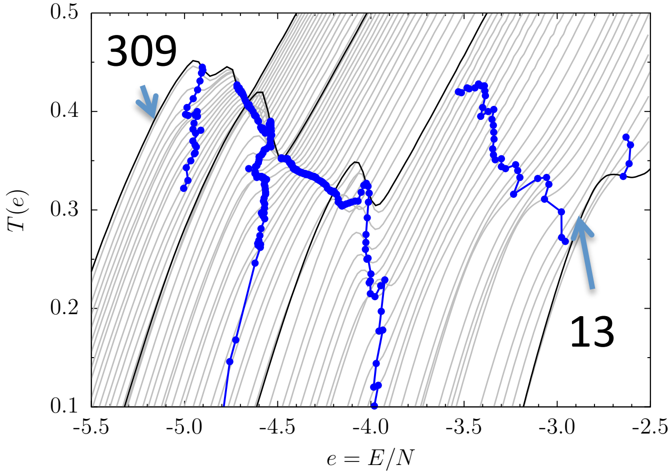
- Replica exchange (Swendsen, Wang, 1986)
- Multiple-histogram reweighting (Ferrenberg, Swendsen, 1989)
- Parallel tempering (Geyer, 1991; Hukushima, Nemoto, 1996)
- Simulated tempering (Marinari, Parisi, 1992; Lyubartsev et al., 1992)
- Multicanonical sampling (Berg, Neuhaus, 1992)
- Wang-Landau Method (Wang, Landau, 2001)
- ...

Output: Density of states

3. Applications: Nucleation

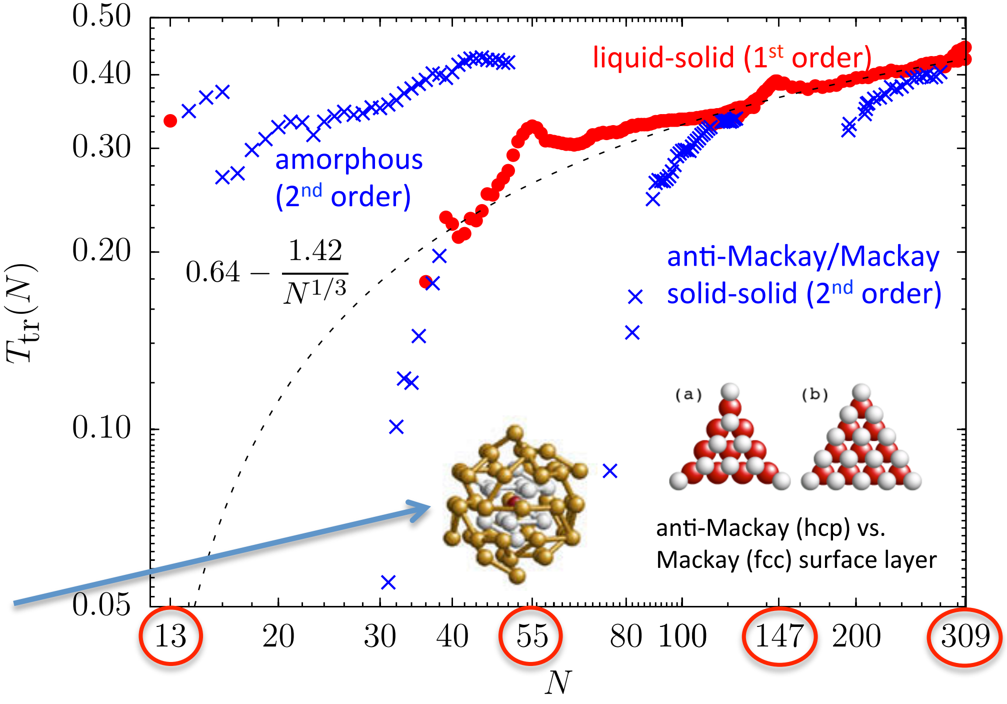
Crystallization of flexible, elastic homopolymers





inflection points in low-energy regions

microcanonical transition temperatures



N = 13 ... 309 monomers

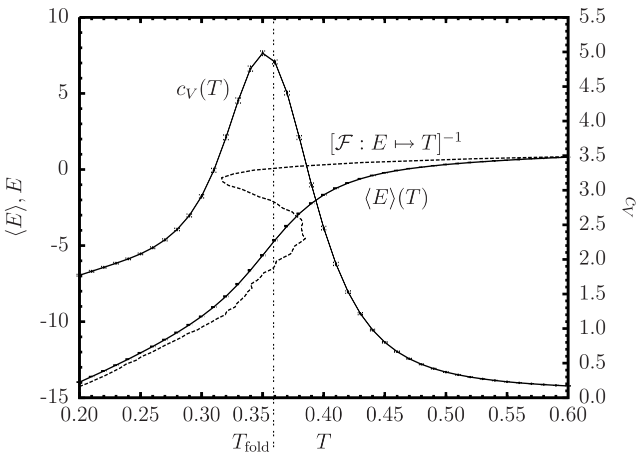
icosahedral geometry preferred; perfect for N=13, 55, 147, 309, ... (“magic” numbers)

[S. Schnabel, D. T. Seaton, D. P. Landau, M.B., Phys. Rev. E **84**, 011127 (2011).]

3. Applications: Protein Folding

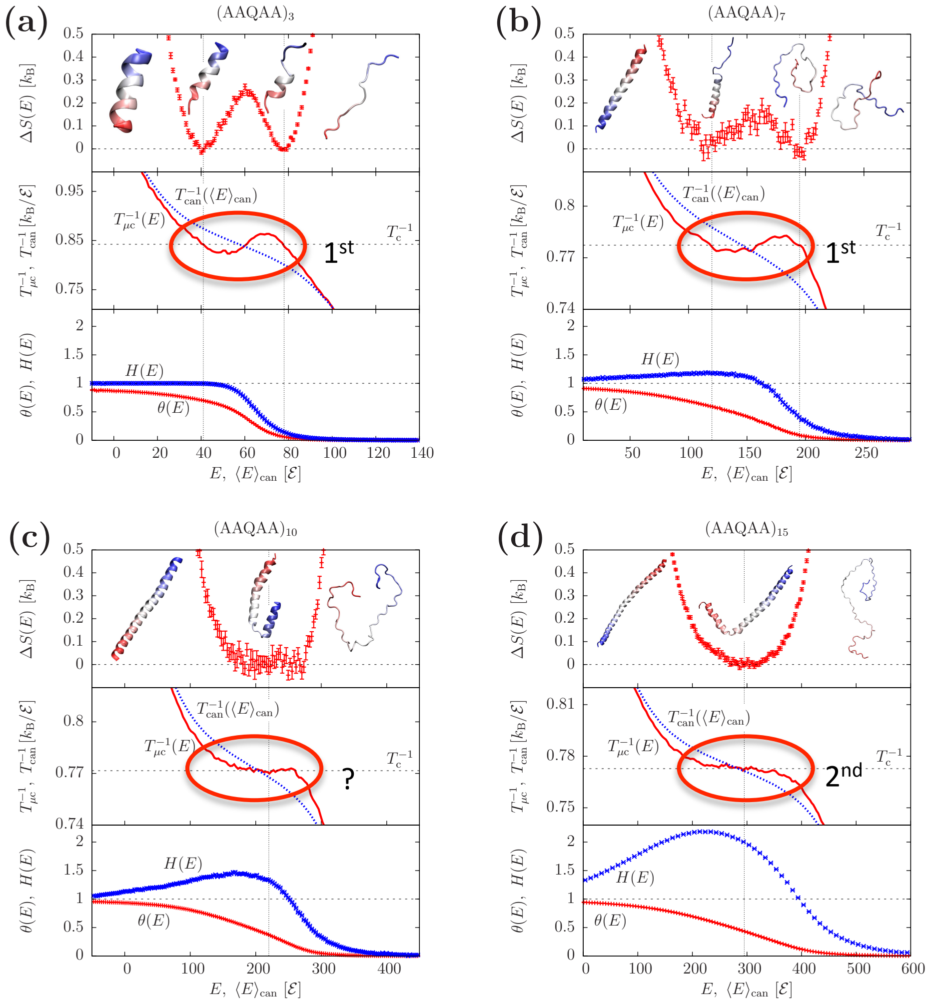
Peptide folding

typical 1st order scenario:



Helix-coil transition of peptide $(\text{AAQAA})_n$ with n=3, 7, 10, 15:

Transition changes from 1st order to 2nd order for longer chains (not preferred in natural proteins)



3. Applications: Protein Folding

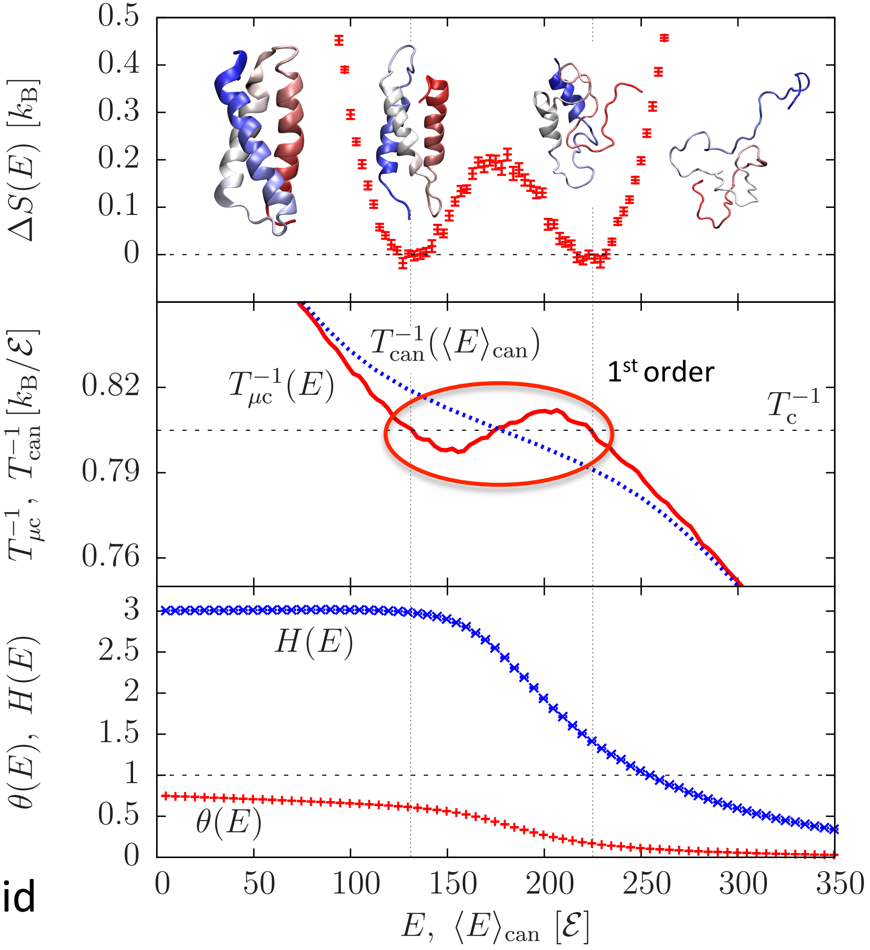
Helix-bundle formation

α3D: three-helix bundle
73 amino acids

clear first-order signal of cooperative helix-coil/ tertiary bundle transition

nucleation due to loss of translational entropy of random/ helical segments: entropic suppression

$H(E)$ average number of helices
$\theta(E)$ helicity of individual amino acid

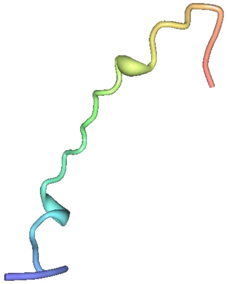


[T. Bereau, M.B., M. Deserno, JACS **132**, 13129 (2010); T. Bereau, M. Deserno, M.B., Biophys. J. **100**, 2764 (2011).]

3. Applications: Protein Aggregation

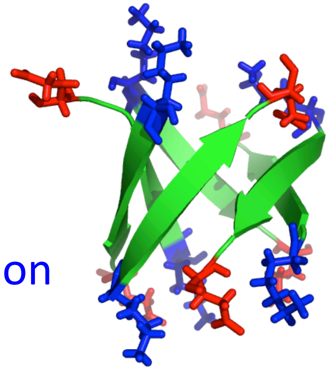
Aggregation of peptides

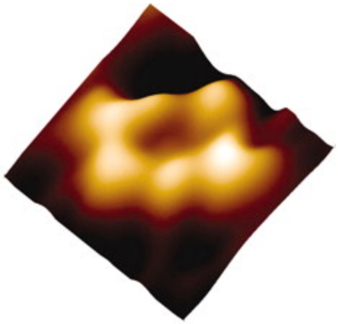
Alzheimer's disease: Toxicity of misfolded proteins (amyloid hypothesis)



Folding of peptide $\mathrm{A}\beta_{1-42}$

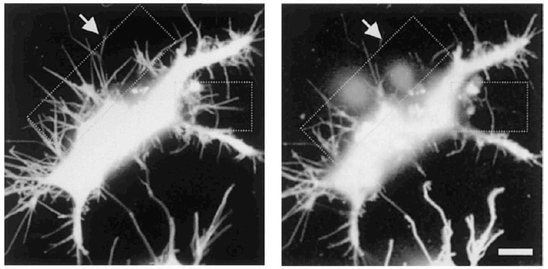
Local unfolding, aggregation





Fusion into neural cell membrane, pore formation

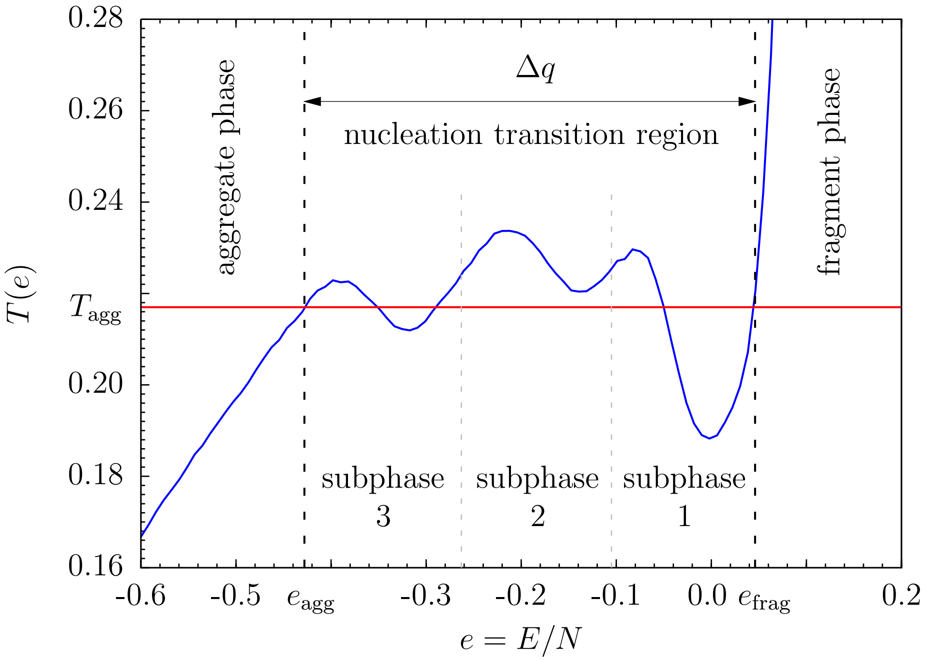
Degeneration of neurons by Ca^{2+} ions

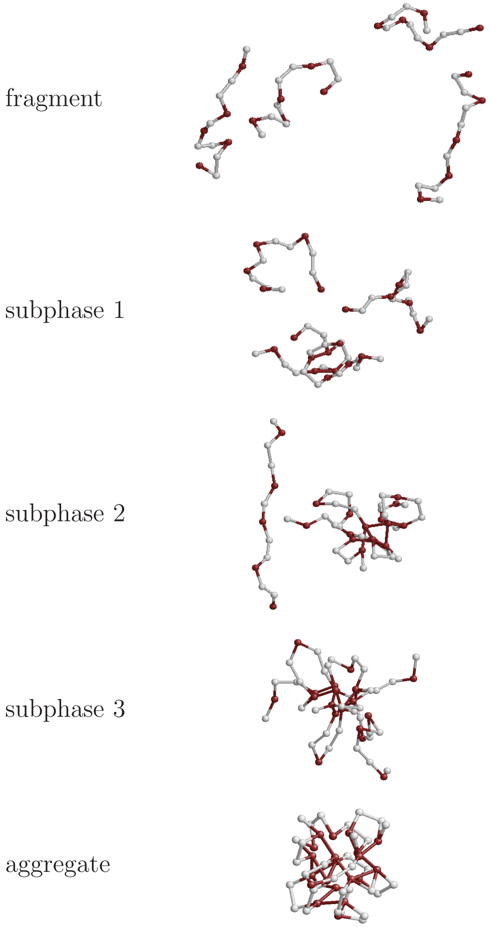


[H. Lin et al., FASEB J. **15**, 2433 (2001); A. Quist et al., PNAS **102**, 10427 (2005); H. A. Lashuel, P. T. Lansbury jr., Quart. Rev. Biophys. **39**, 167 (2006); S. Tomaselli et al., ChemBioChem **7**, 257 (2006); S. Mitternacht, A. Irbäck, Proteins **71**, 207 (2008).]

3. Applications: Protein Aggregation

Example: Aggregation of 4 peptide chains



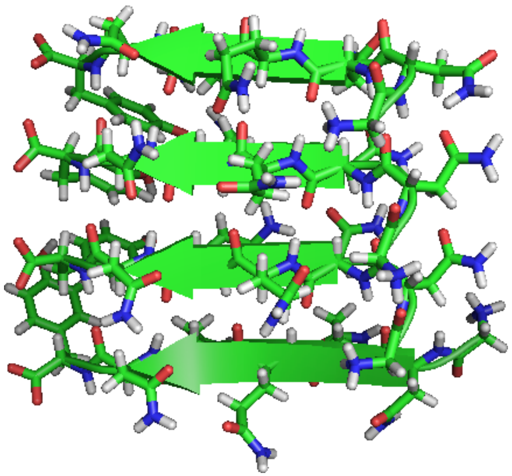


- hierarchy of subphase transitions
- latent heat does not vanish (1^{st} order transition)

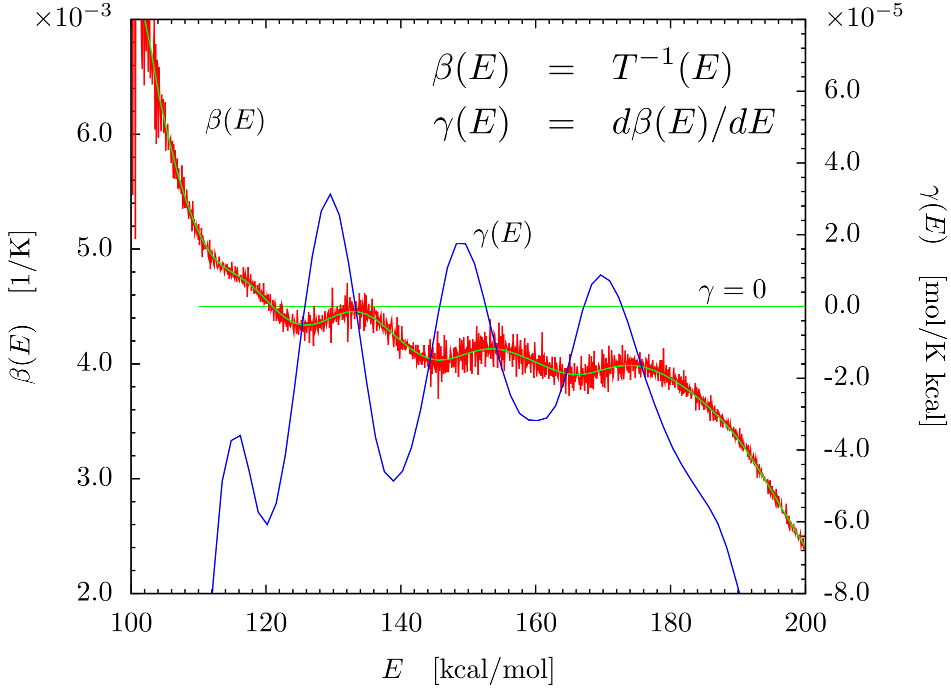
[C. Junghans, M.B., W. Janke, Phys. Rev. Lett. **97**, 218103 (2006); J. Chem. Phys. **128**, 218103 (2008); C. Junghans, W. Janke, M.B., Comp. Phys. Commun. **182**, 1937 (2011).]

3. Applications: Protein Aggregation

Protein-protein aggregation: GNNQQNY (4 chains)



lowest-energy state: β - sheet

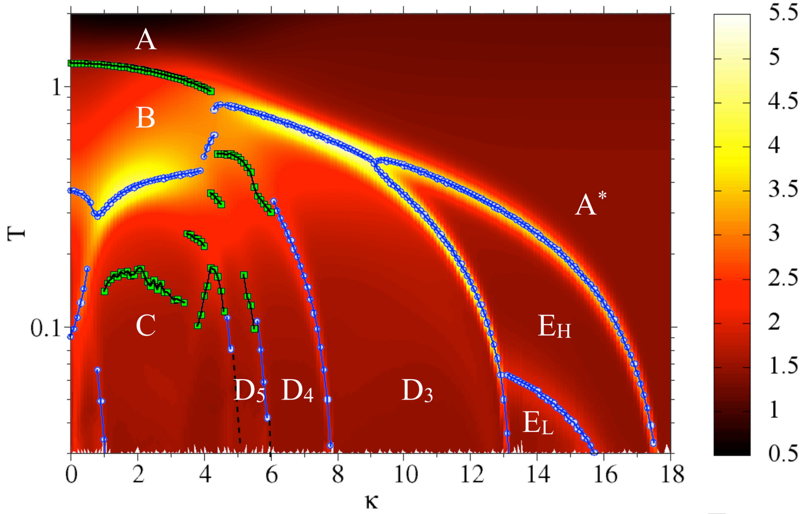


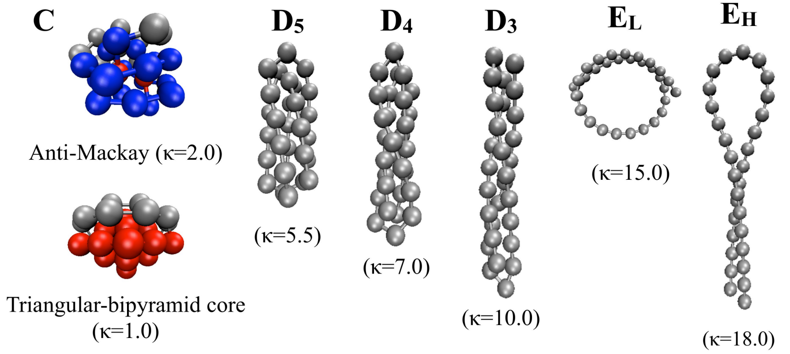
- same behavior: hierarchy of subphase transitions
- *O*(10) CPU years using parallel tempering Monte Carlo

[M.B., *Thermodynamics and Statistical Mechanics of Macromolecular Systems* (Cambridge Univ. Press, 2014).]

3. Applications: Structural Phase Diagrams

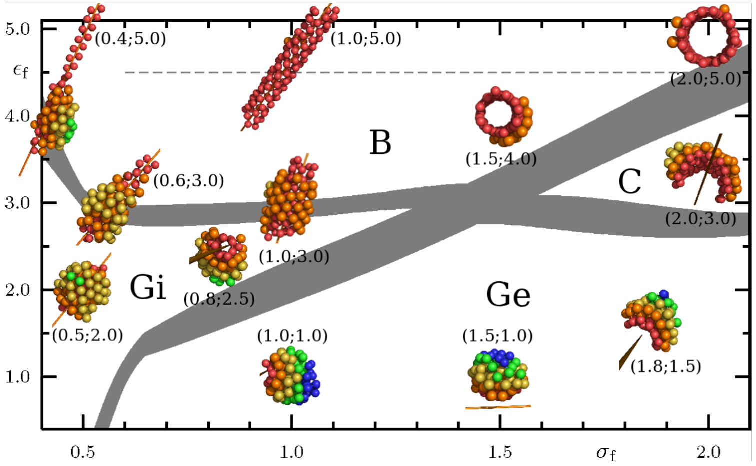
Semiflexible polymers





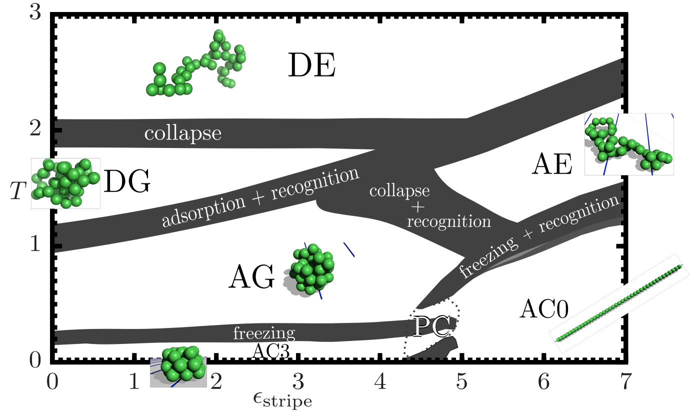
[D. T. Seaton, S. Schnabel, D. P. Landau, M.B., Phys. Rev. Lett. **110**, 028103 (2013).]

Adsorption at nanowires



[T. Vogel, M.B., Phys. Rev. Lett. **104**, 198302 (2010).]

Pattern recognition



[M. Möddel, W. Janke, M.B., Phys. Rev. Lett. **112**, 148303 (2014).]

4. In the End...

Summary:

- Structural transitions of macromolecules exhibit features resembling thermodynamic phase transitions.
- Transitions are dominated or at least influenced by finite-size and surface effects.
- Studies require sophisticated algorithmic approaches and advanced computational resources (which is why it is a 21st century problem!).
- Novel or rediscovered statistical physics methods enable efficient and systematic analyses.
- Studies of transitions on mesoscopic scales help understand phase transitions in macroscopic systems even better.

Ads:

NSF

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

- Website: www.smsyslab.org
- M.B., *Thermodynamics and Statistical Mechanics of Macromolecular Systems* (Cambridge Univ. Press, 2014)