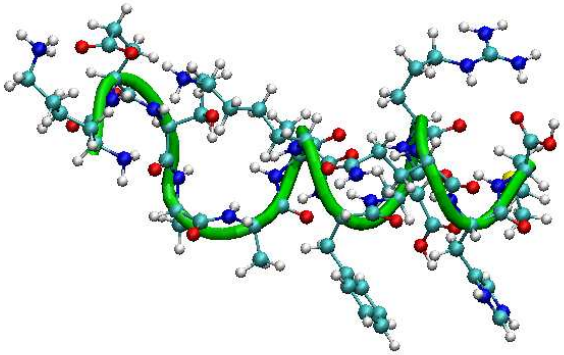
Conformational Transitions of Lattice Heteropolymers

Michael Bachmann

Institut für Theoretische Physik, Universität Leipzig



Workshop LEILAT04, 3 June 2004

UNIVERSITÄT LEIPZIG

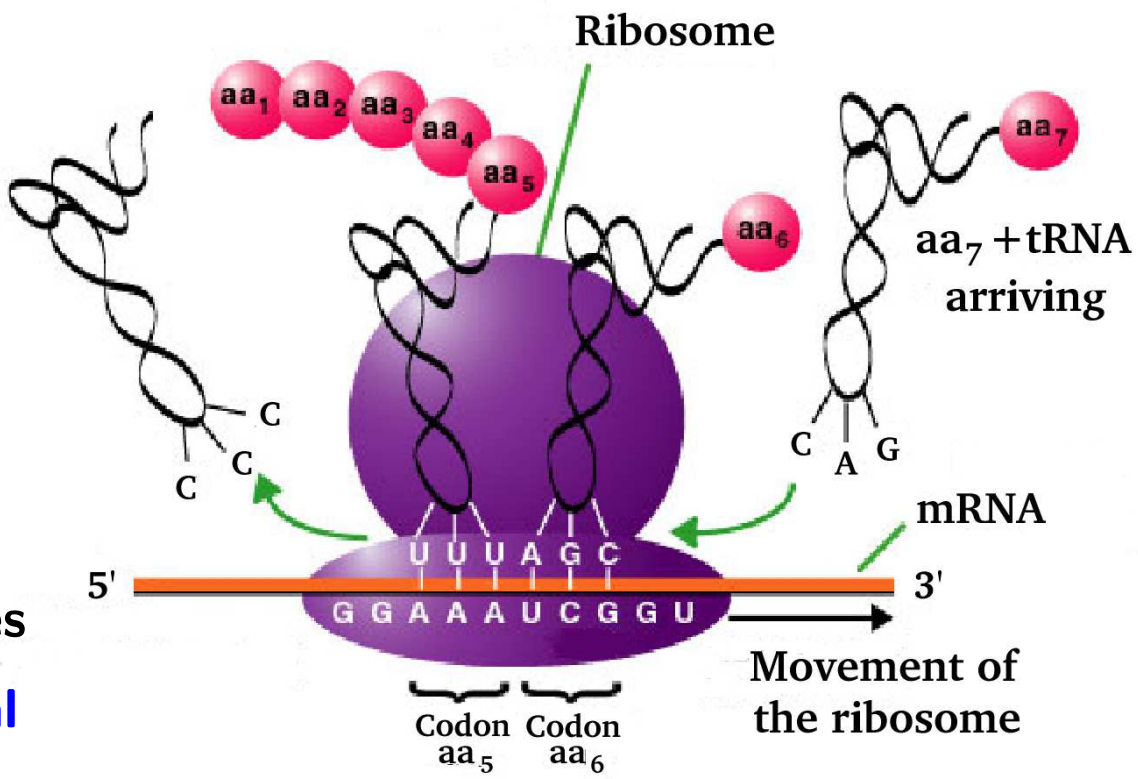
Computational
Quantum Field
Theory

C Q T

Proteins

Sequence
$\Downarrow$
Conformation
$\Downarrow$
Function

Protein synthesis through ribosome.
Sequence, not conformation!!



Protein sequences are encoded in the DNA.

- **Polypeptides:** $N \sim 20 - 4000$ amino acids covalently linked by peptide bonds
- **20 amino acids** $\Rightarrow 20^N$ possible sequences
- but: only $\sim 10^5$ **functional proteins** (human)

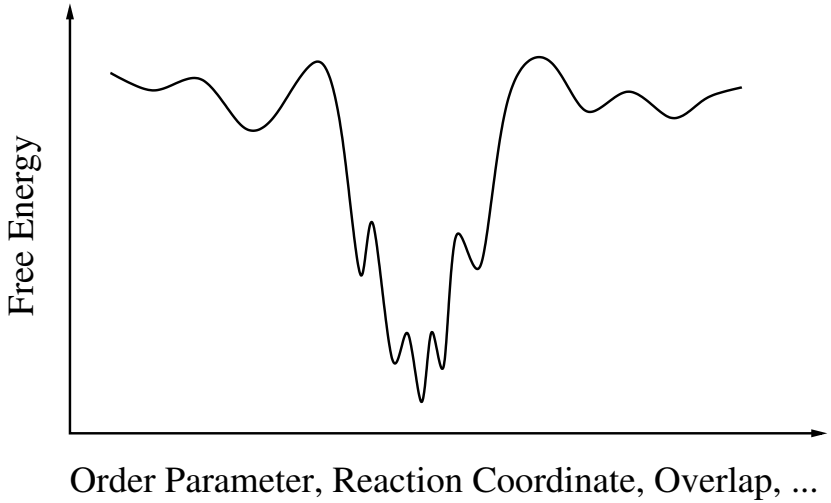
Proteins

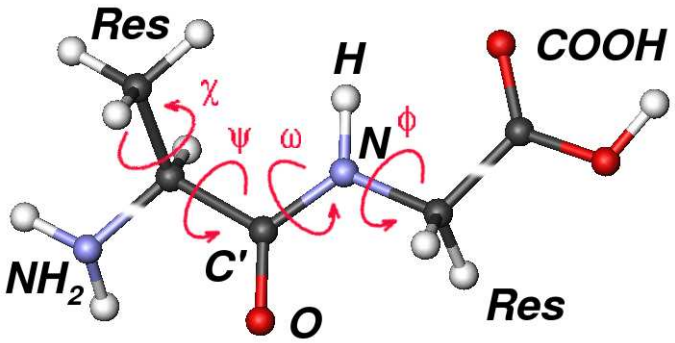
Sequence
⇓
Conformation
⇓
Function

Sequence determines structure.

- **Spontaneous folding** into native conformation (Anfinsen's experiments)
- **Path(s) to the fold:** energy-driven stochastic search
- Time scale of protein folding: milliseconds to seconds
- **Metastability** (function-dependent)

Free energy landscape: rugged, with deep funnel-like global minimum.





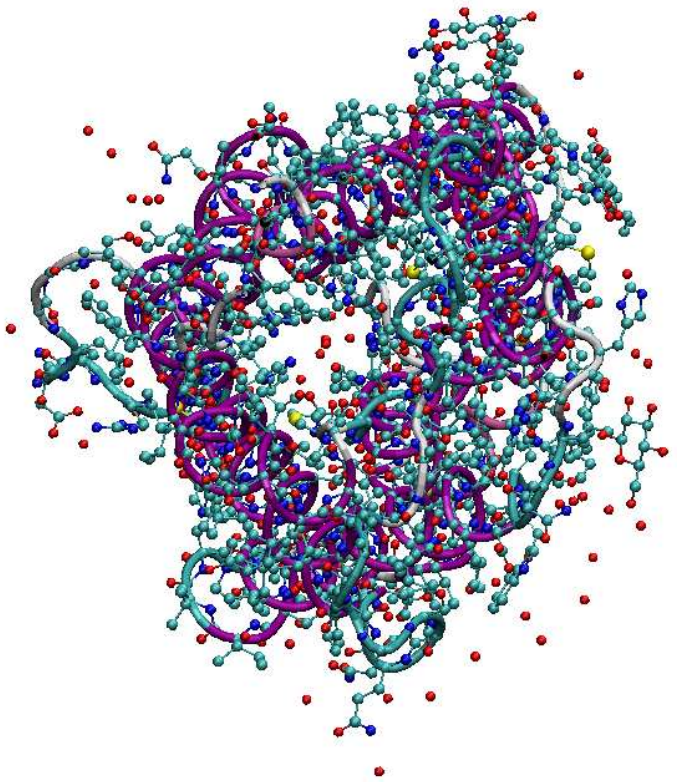
NH_2– . . . –Ala–Gly– . . . –COOH

Proteins

Sequence
⇓
Conformation
⇓
Function

3D structure determines functionality.

Aquaporin:
3 000 000 000 water molecules / sec.



Proteins are involved in almost all cell processes.

- **Transport** (channels, pores)
- **Cell stability** (actin filaments)
- **Catalytic activity** (enzymes)
- **"Molecular motors"**
  (DNA polymerase, ATP synthase)
- ... many more functions

Proteins – Open Questions

General questions:

- **How does nature select the *relevant* sequences?**
- **How does the protein *spontaneously* find the path to the fold?**
- **What is the relationship between *sequence and native conformation*, i.e., how is the structural information encoded in the sequence of the amino acids?**
- **How do these nanoscale machines work?**

The direct folding problem:

Given the sequence: What is the native conformation (= function)?

The inverse folding problem: (... of enormous pharmaceutical interest)

Given the target structure (= function): Which sequences of amino acids fold into this conformation ("drug design")?

Folding kinetics: Computer simulations (in 10 to 20 years...)

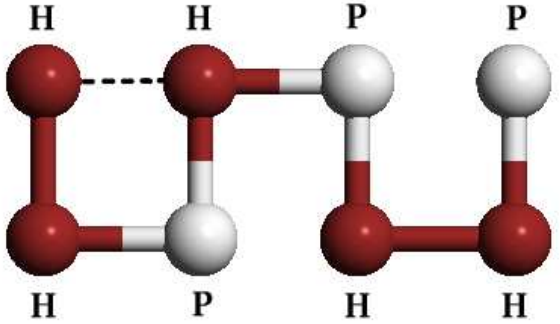
HP Model – The Simplest Model for Heteropolymers

HP lattice proteins:

Lattice heteropolymers with sequence of two types of monomers:

hydrophobic (H)

polar (P)



HP protein folding principle: screening of the hydrophobic core from the (fictitious) aqueous environment by the polar residues

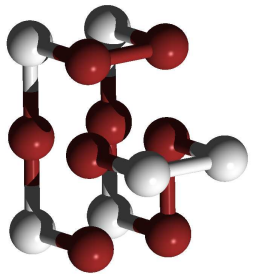
HP model (Dill, 1985); only hydrophobic next-neighbour interaction:

$$E = -\sum_{\langle i,j<i-1\rangle} \sigma_i\sigma_j, \qquad \sigma_i = \begin{cases} 1 & \text{hydrophobic} \\ 0 & \text{polar} \end{cases}$$

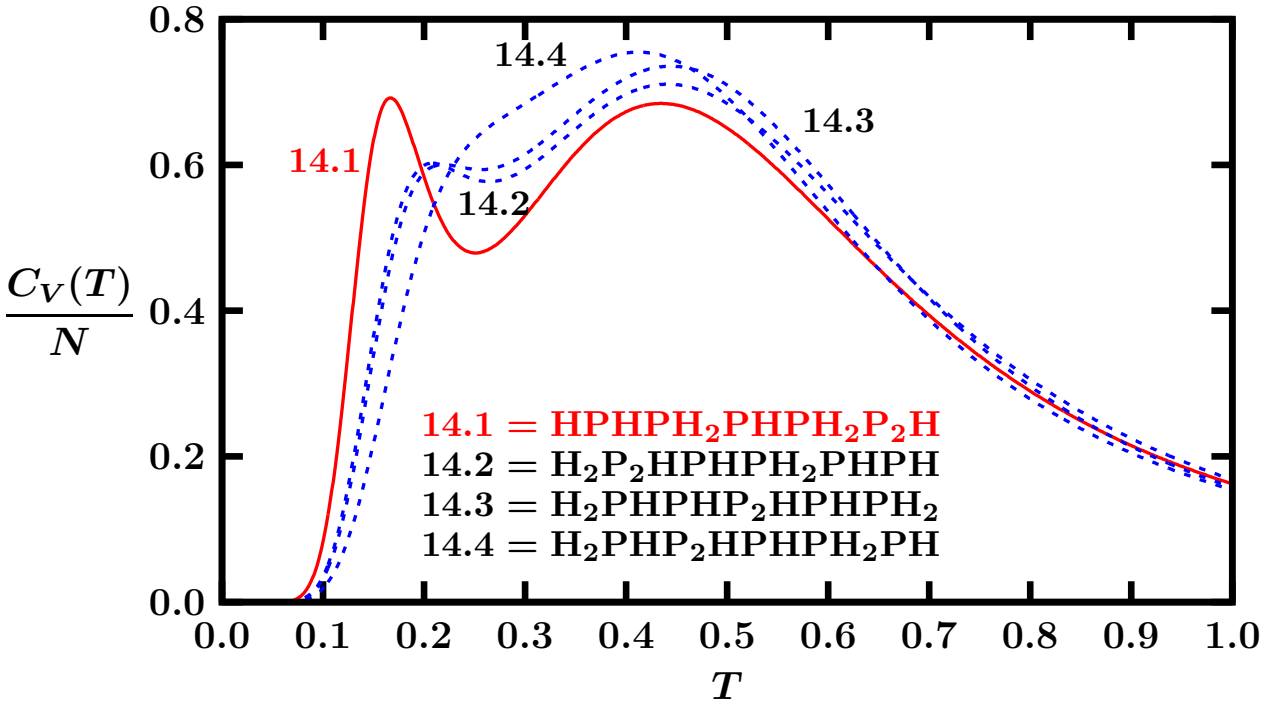
HP Model – The Simplest Model for Heteropolymers

Designing Sequences: Nondegeneracy of the ground-state conformation

Exemplified 14mers on the s.c. lattice: 1 designing, 3 nondesigning
(hydrophobicity $n_H = 8$, energy minimum $E_{\min} = -8$)



Sequence 14.1



Specific heat: pronounced low-temperature peak for the designing 14mer
$\Longrightarrow$ ground state – globule transition [M.B., W. Janke, APP (2003)]

HP Model – The Simplest Model for Heteropolymers

Systematic and exact study on the s.c. lattice:

Enumeration of *all* sequences and conformations with up to 19 monomers

Numbers of all nonredundant (S_N) and *designing sequences* (S_N^{D}):

N	12	13	14	15	16	17	18	19
S_N $(\times 10^3)$	2.1	4.2	8.3	17	33	66	131	263
S_N^{D}	2	0	1	1	1	8	29	47

Number of all (C_N) and *designable* conformations (C_N^{D}):

N	12	13	14	15	16	17	18	19
C_N $(\times 10^6)$	42	199	944	4,469	21,175	100,122	473,730	2,237,724
C_N^{D}	96	0	48	48	48	384	1,344	2,016

Total number of states: $M_N \sim 2^N \times \mu^N N^{\gamma-1}$ ($\mu_{\mathrm{s.c.}} \approx 4.68$, $\gamma \approx 1.16$)

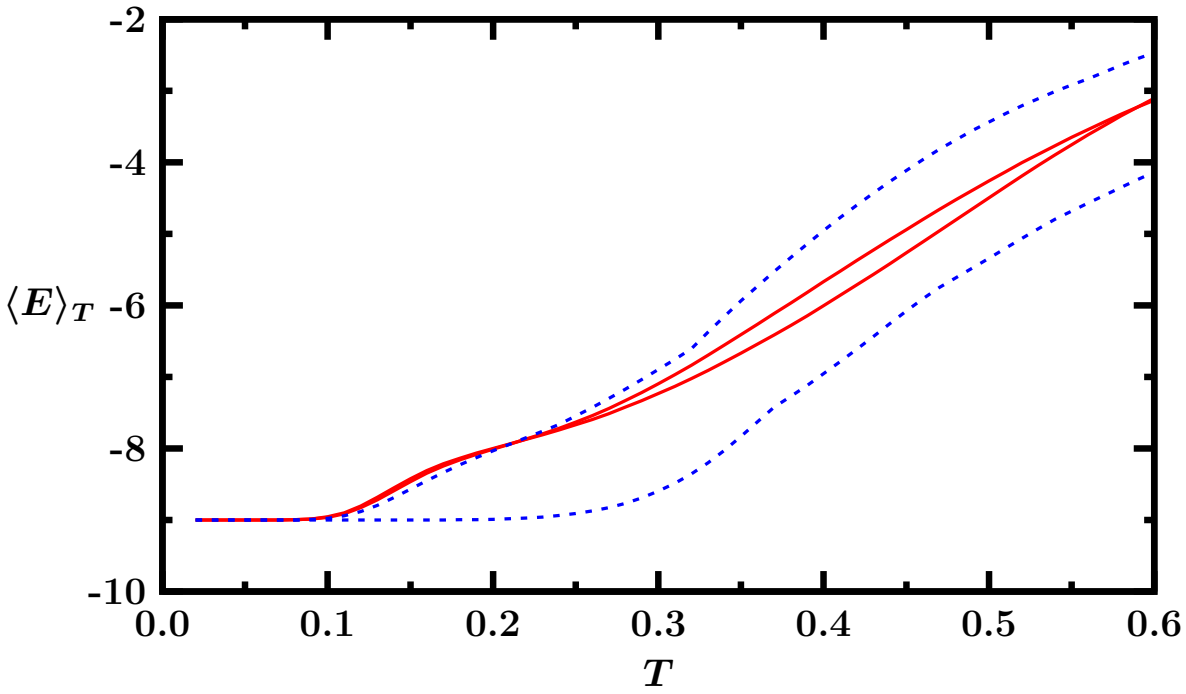
$N = 19$: $M_N = \mathcal{O}(10^{18}) \Rightarrow$ *1.5 CPU years* (AMD Athlon 1.5 GHz)

HP Model – The Simplest Model for Heteropolymers

Example: 527 sequences with 18 monomers (all $E_{\min} = -9$, $m_H = 8$):

525 nondesigning, 2 designing

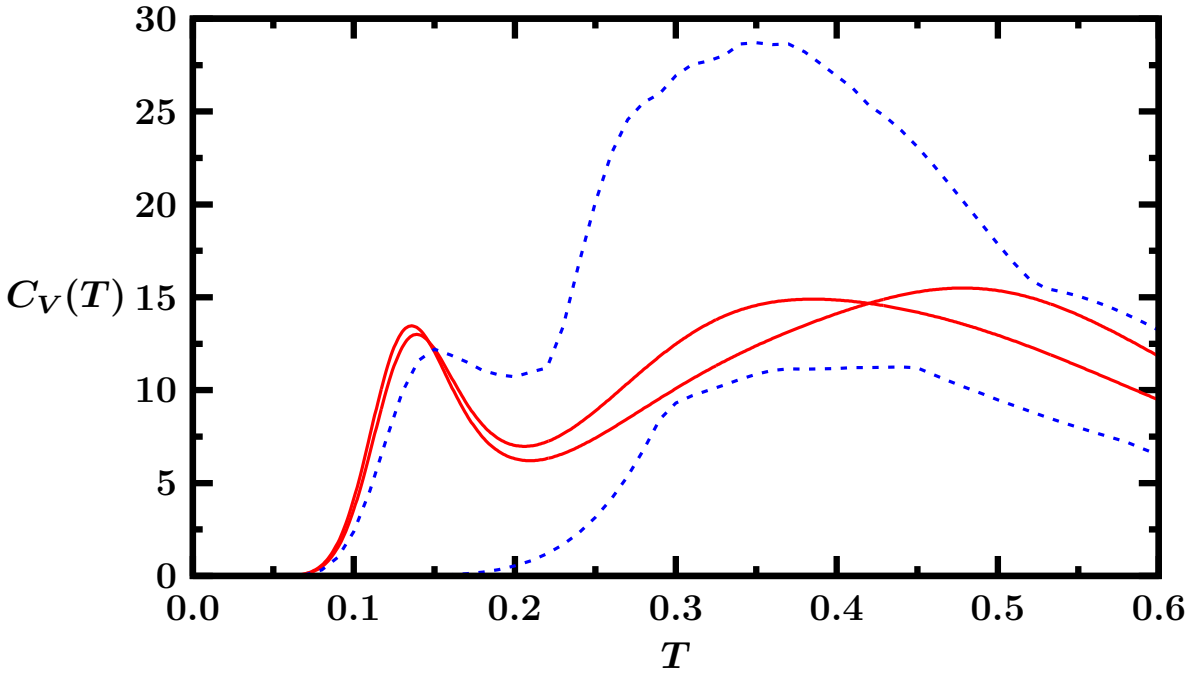
Mean energy:



(solid: designing, dashed: upper and lower bounds for nondesigning sequences)

HP Model – The Simplest Model for Heteropolymers

Specific heat:



(solid: designing, dashed: upper and lower bounds for nondesigning sequences)

designing sequences: *strong* **low-temperature transition**

nondesigning sequences: **rather** *weak* **or no low-temperature transition**

[R. Schiemann, M.B., W. Janke, q-bio/0405009 (2004)]

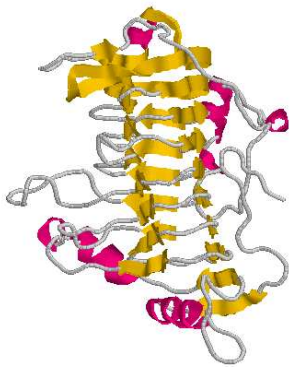
HP Model – The Simplest Model for Heteropolymers

Sequences with $N > 30$ monomers: sophisticated methods required

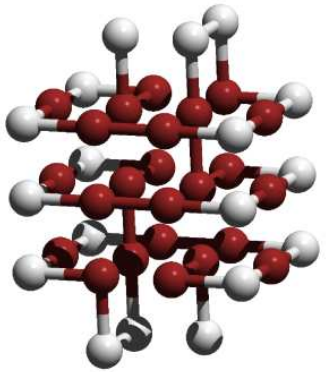
New algorithm: **multicanonical chain growth** [M.B., W. Janke, PRL (2003)]

$\Rightarrow$ Iterative method for the estimation of the density of states

Exemplified 42mer:



Lattice model for parallel β-helix of *pectate lyase C* (Yue, Dill, 1995)

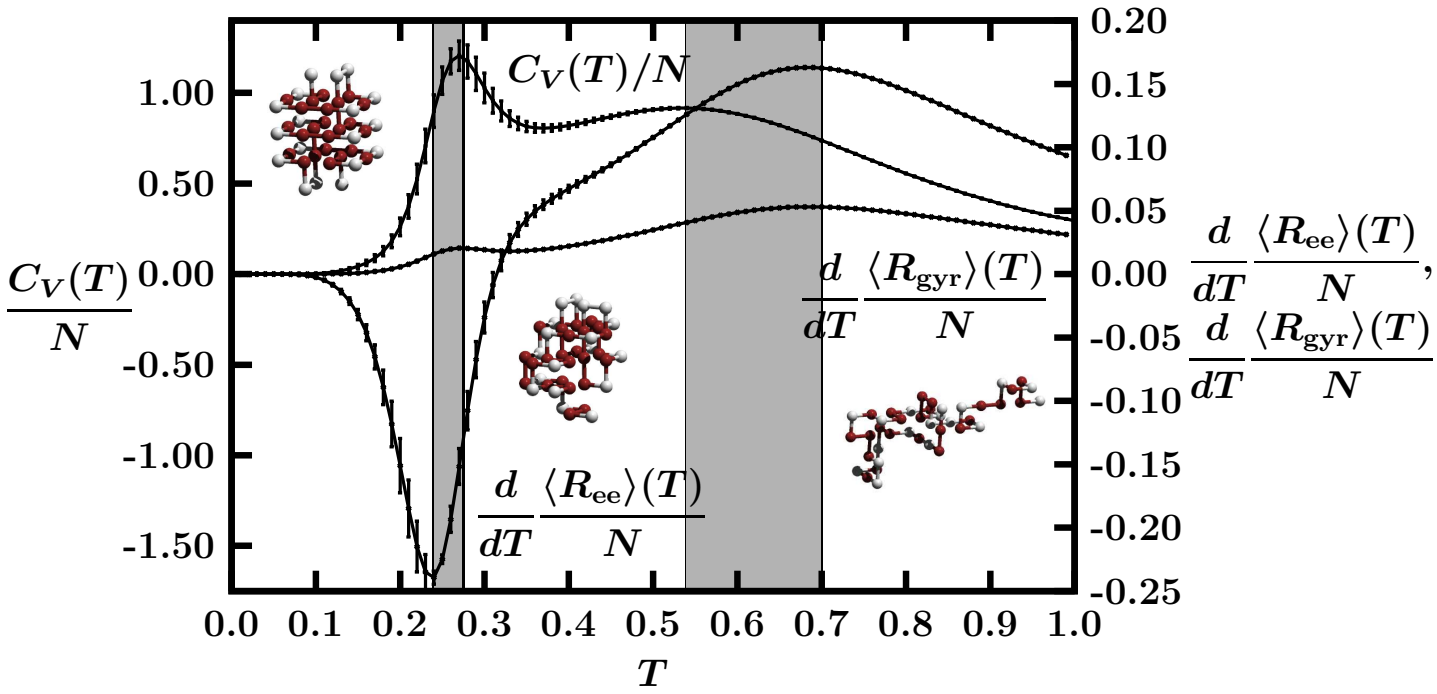


Ground-state properties of the lattice model:

- Minimum energy $E_{\min} = -34$, only 4-fold degeneracy
- Native conformations contain two parallel helices

HP Model – The Simplest Model for Heteropolymers

Energetic and conformational fluctuations:



Three "phases" separated by two conformational transitions: compact hydrophobic core states, maximally compact globules, random coils

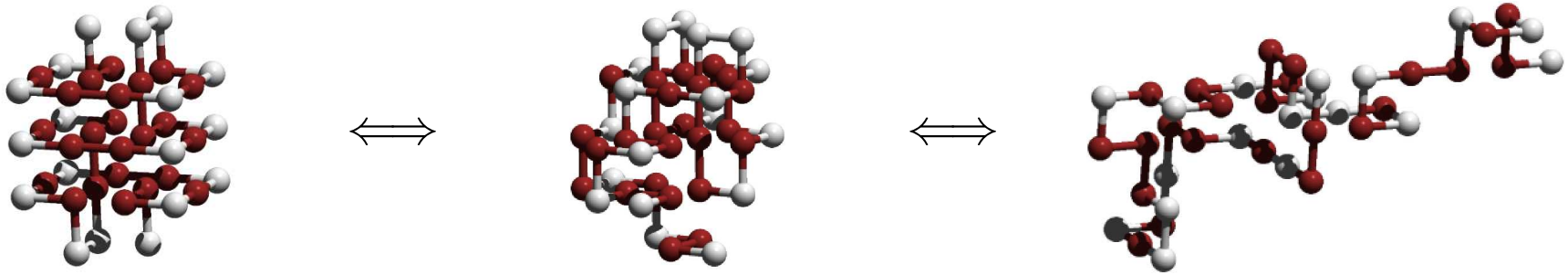
[M.B., W. Janke, JCP (2004)]

HP Model – The Simplest Model for Heteropolymers

Summary HP model:

- **Two conformational transitions:**

hydrophobic core $\Longleftrightarrow$ **globules (...traps)** $\Longleftrightarrow$ **coils**



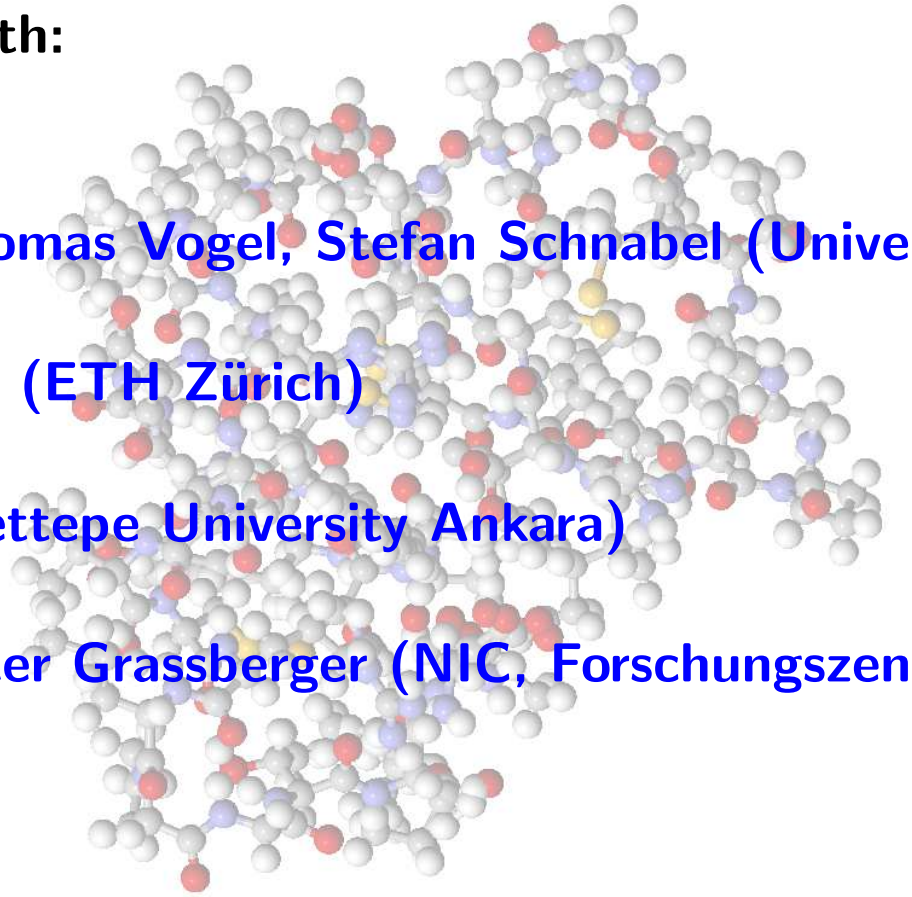
- **Qualitative comparison with natural heteropolymers:**

  $\Rightarrow$ **Hydrophobic effect**
  $\Rightarrow$ **Small number of relevant (*designing*) sequences**
  $\Rightarrow$ **Small number of different *designable* conformations**

- **Quantitative comparison impossible: lattice effects, insufficient volume exclusion, no specific interactions (e.g. hydrogen bonds)**

Thanks for ...

... the collaboration with:



- Wolfhard Janke, Thomas Vogel, Stefan Schnabel (Universität Leipzig)
- Reinhard Schiemann (ETH Zürich)
- Handan Arkın (Hacettepe University Ankara)
- Hsiao-Ping Hsu, Peter Grassberger (NIC, Forschungszentrum Jülich)