# *In Silico* Folding of Small Proteins

**Ulrich H.E. Hansmann**

John von Neumann Institute for Computing (NIC)
http://www.fz-juelich.de/nic/cbb

Dept. of Physics, Michigan Technological University
http://www.phy.mtu.edu/biophys


***Supported by research grants of the National Science Foundation (CHE-0313618) and the National Institutes of Health (GM62838)***

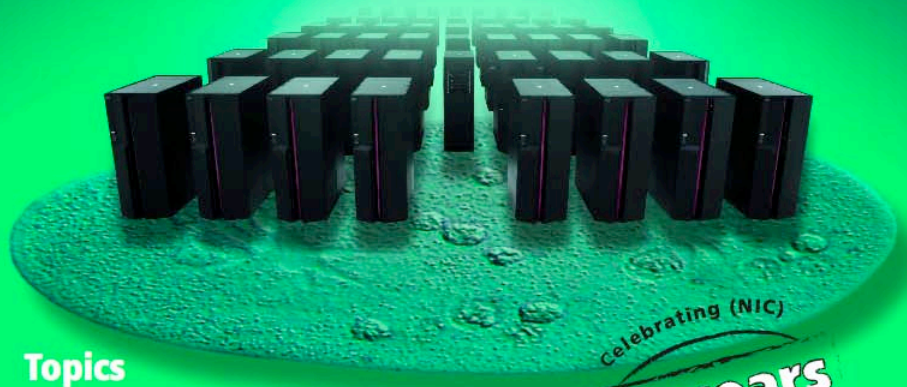
# Co-workers:

- **Everalso Arashiro** (FZJ)
- **Tatjana Eitrich** (FZJ)
- **Jan Meinke** (FZJ)
- **Sandipan Mohanty** (FZJ)
- **Olav Zimmermann** (FZJ)
- **Thomas Neuhaus** (FZJ)
- **Wolfgang Nadler** (MTU)
- **Siegfried Hoefinger** (MTU)

- **Liang Han** (MTU)
- **Parimal Kar** (MTU)
- **Yanjie Wei** (MTU)
- **Xiaolin Xiao** (FZJ)

# From Computational Biophysics to Systems Biology (CBSB07)

**02. to 04. May 2007, Jülich, Germany**



Celebrating (NIC)
20 years
Neumann Institute for Computing

## Topics

- Protein Folding
- Multi Protein Complexes
- Nanostructures
- Cellular systems at the molecular level

## Invited Speakers

M. Cieplak (Pol. Acad. Science, Warsaw)
Ch. Floudas (Prineton)
J. Langowski (DKFZ, Heidelberg)
B. Lesyng (U. Warsaw)
A. Liwo (U. Gdansk)
J. Onuchic (UCSD, La Jolla)
A. Roitberg (UF, Gainesville)
R. Russell (EMBL, Heidelberg)
J. Skolnick (Georgia Tech, Atlanta)
K. Takahashi (tMSI, Berkeley)
D. Thirumalai (UM, College Park)
R. Wade (EML, Heidelberg)

NIC **http:// www.fz-juelich.de/cbsb06**

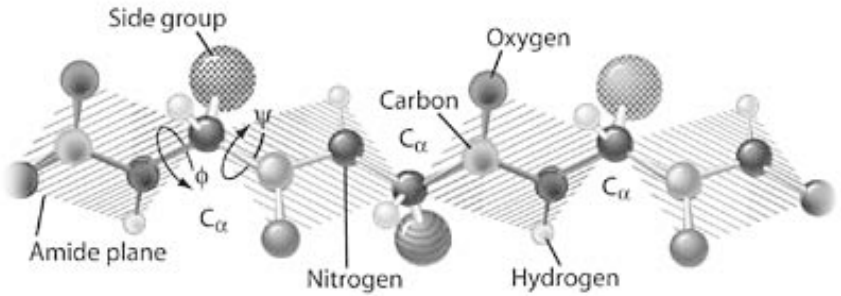
Organizers: U.H.E. Hansmann, J. Meinke, S. Mohanty, O. Zimmermann

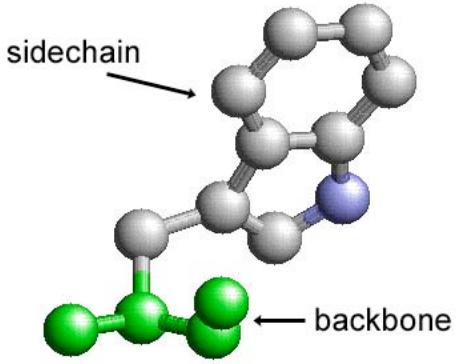
# The Protein Folding Problem

- **> 50,000** different kinds of proteins in human body
- Muscles, antibodies, enzymes, … (“**nanomachines**”)
- Proteins are **polymers** built up from **amino acids**
- The sequence of amino acids is specified in the genome → we know in principal the **chemical composition** of all proteins in the human body
- **Function** of these proteins?
- Sequence – structure – function **relationship**?
  - Understanding (mal)function of enzymes and their role in diseases
  - Design of new drugs

# Biochemistry of Proteins (1)

- Proteins are build up from 20 naturally occurring amino acids  (primary structure)



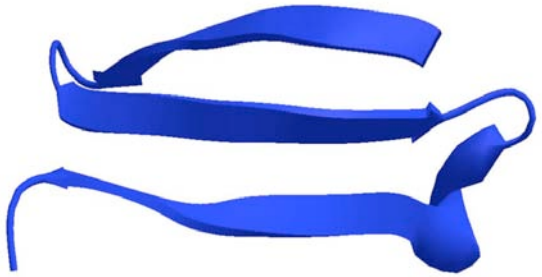




The amino acid **Tryptophan**

# Biochemistry of Proteins (2)

- Proteins form  regular local structures (secondary structure)

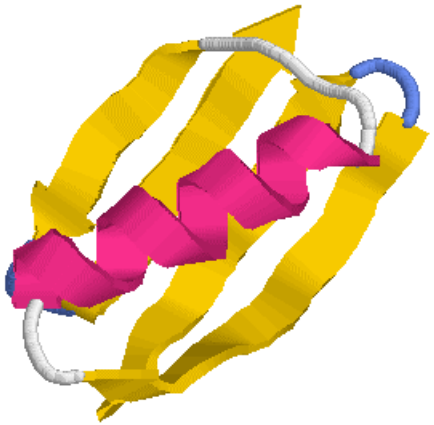
- $\alpha$ - helix

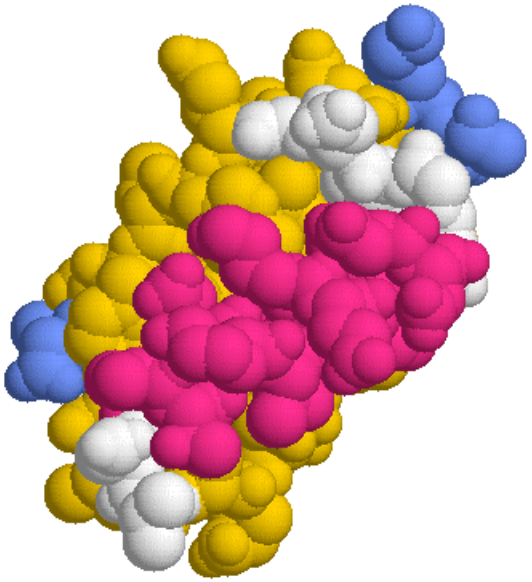


- $\beta$ - sheet

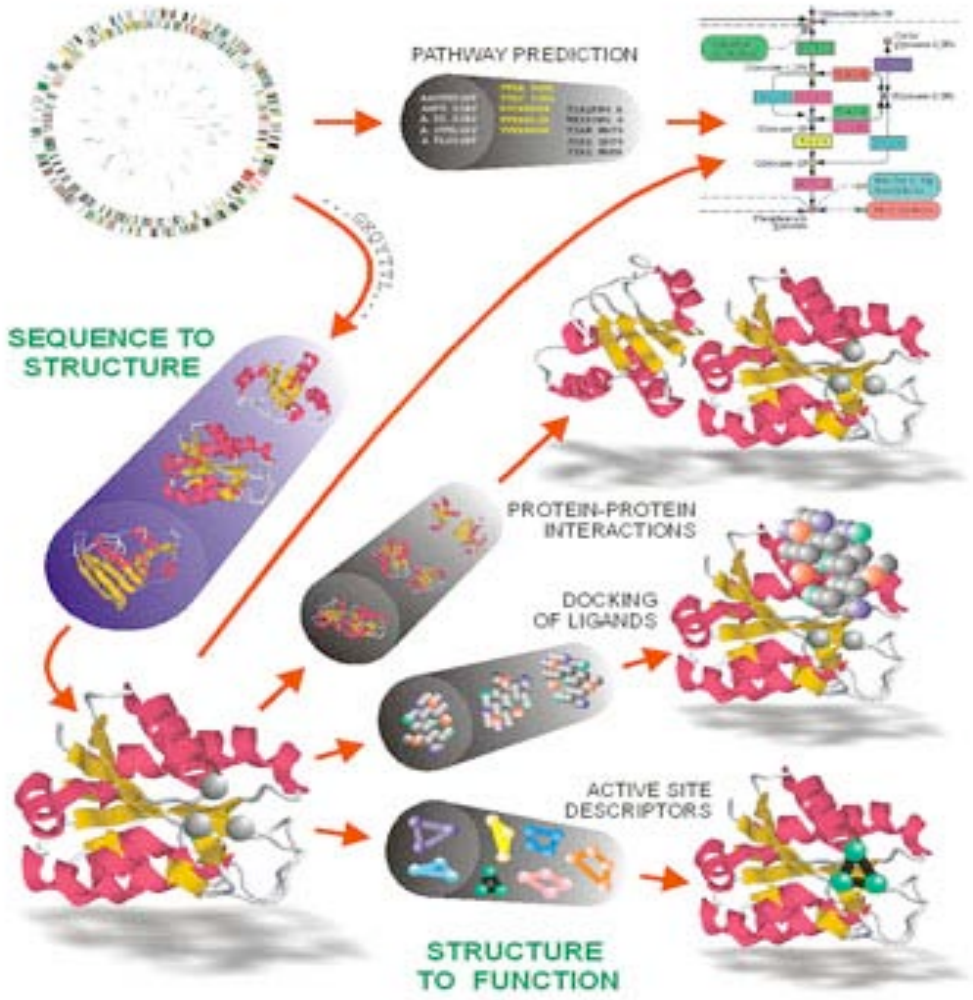
# Biochemistry of Proteins (3)

- The function of proteins depends on their three-dimensional shape (the tertiary structure)



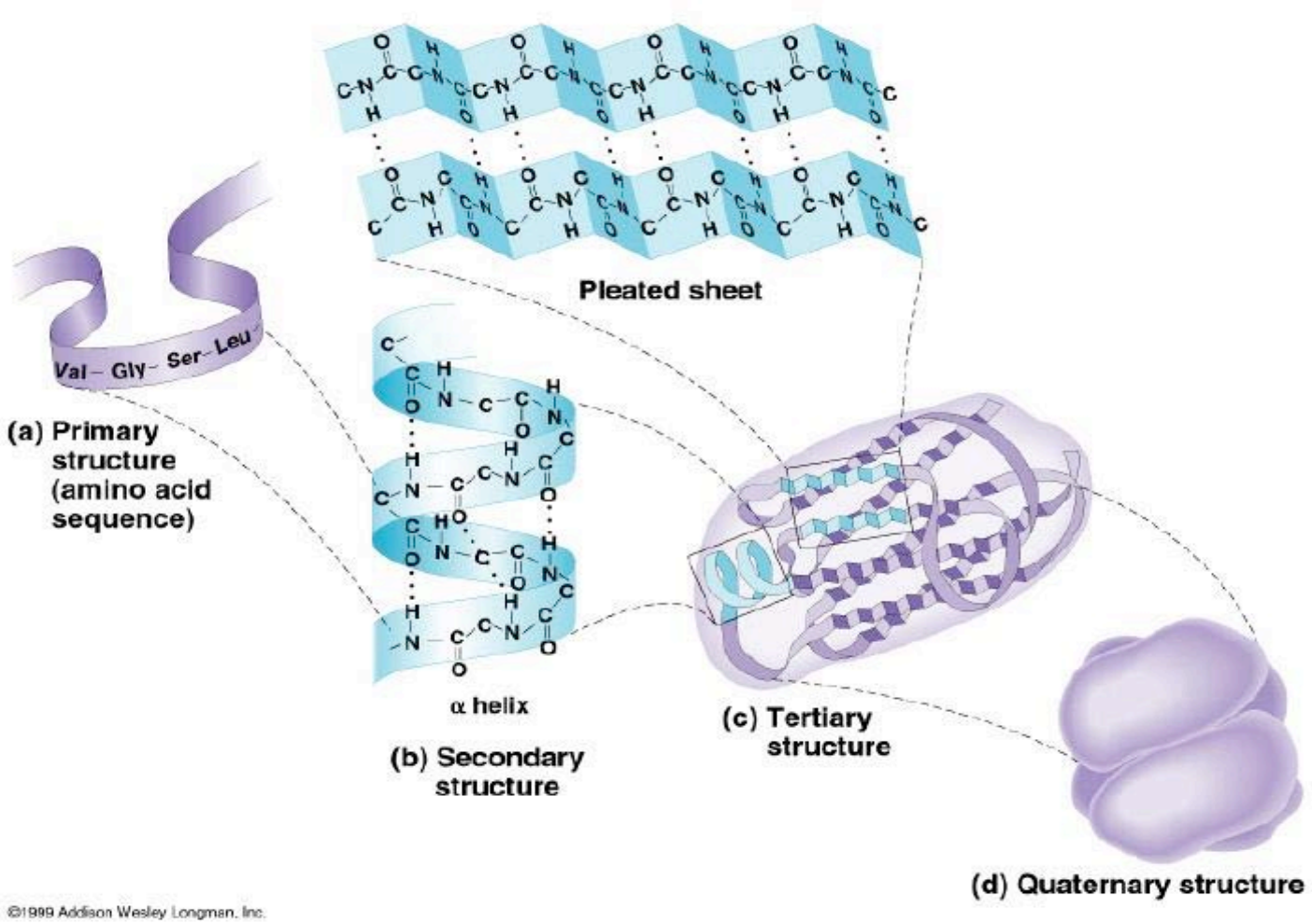


# Protein Science




http://cssb.biology.gatech.edu/skolnick/

# Structure Prediction

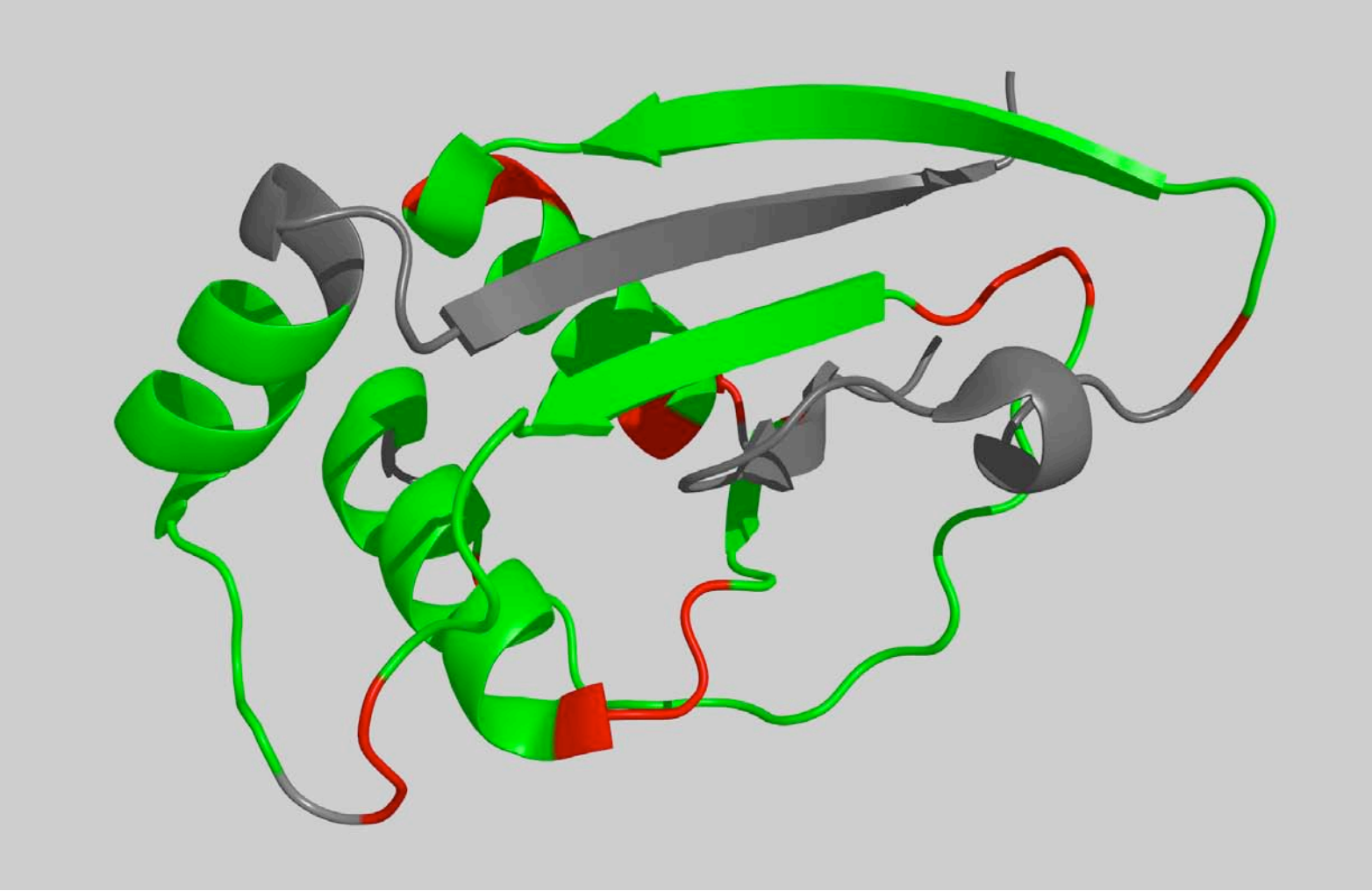


# Dihedral region prediction using SVMs

O. Zimmermann and U.H.E. Hansmann, Bioinformatics **22** (2006) 3009

- **Secondary structure** prediction: helix, sheet, coil
- **Dihedrals** of “coil” residues?
- Additional information for **structure prediction** algorithms.
- **SVM** based multistep method
- **encoding** of sequences as vectors of **amino acid properties** (e.g. volume, hydrophobicity etc.)
- Prediction correct for **~70%** of all residues in protein cores

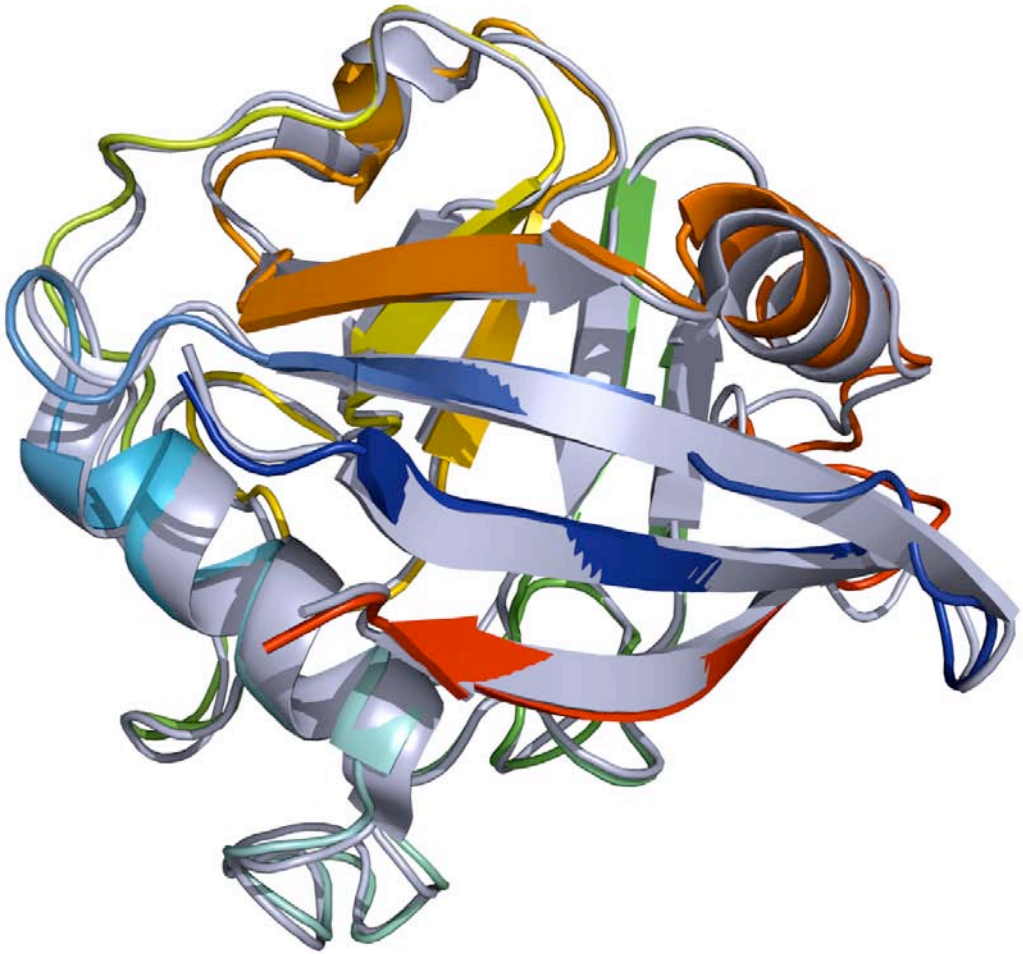
# Example: CASP6 target T0242 (new fold category, PDB:2blkA)



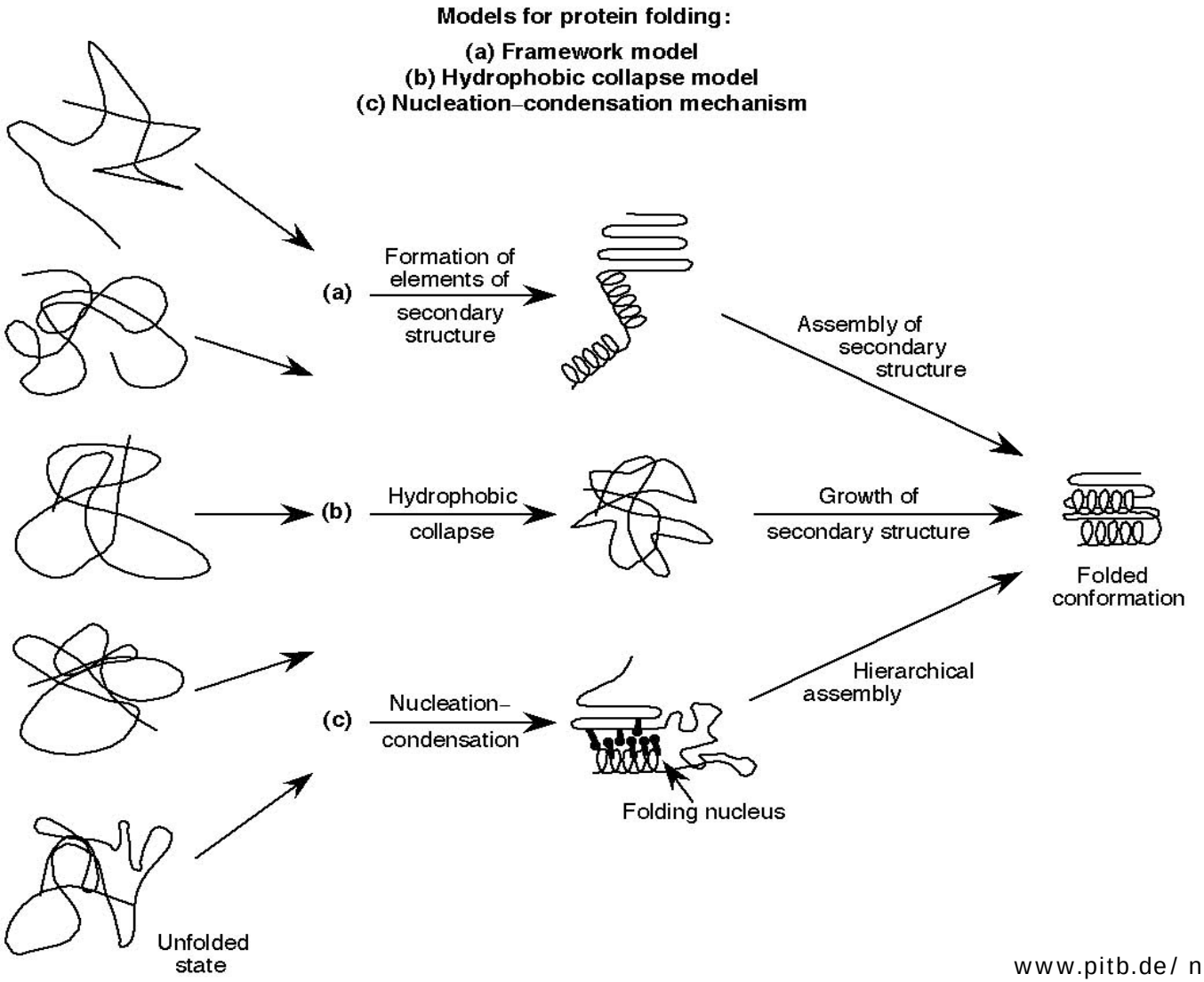
# CASP7 competition

- 207 human expert groups and 48 servers participated.
- 104 targets released, 95 targets assessed
- We submitted 483 models for 97 targets
- # 5 in free-modeling category (19 Targets)

# Example: T0346 (172 residues)



# Models for Protein Folding




www.pitb.de/ nolting/prot00/

# Simulations

- Proteins are only **marginal stable**: ≈ 10 kcal/mol
- **Approximations** necessary
- Interaction with **solvent**?
- **Rough** energy landscape
- **Slow convergence** at room temperature

# Energy Function

- Sum of ***in vacuo*** energy and **solvation** energy
- The ***in vacuo*** energy is modeled by **force fields**
  **Example: ECEPP/2**
  (Nemethy et al., *JPC* **87** (1983) 1883)
- How to model best **protein-water** interaction?

$$E_{tot} = E_{es} + E_{vdW} + E_{hb} + E_{tors}$$

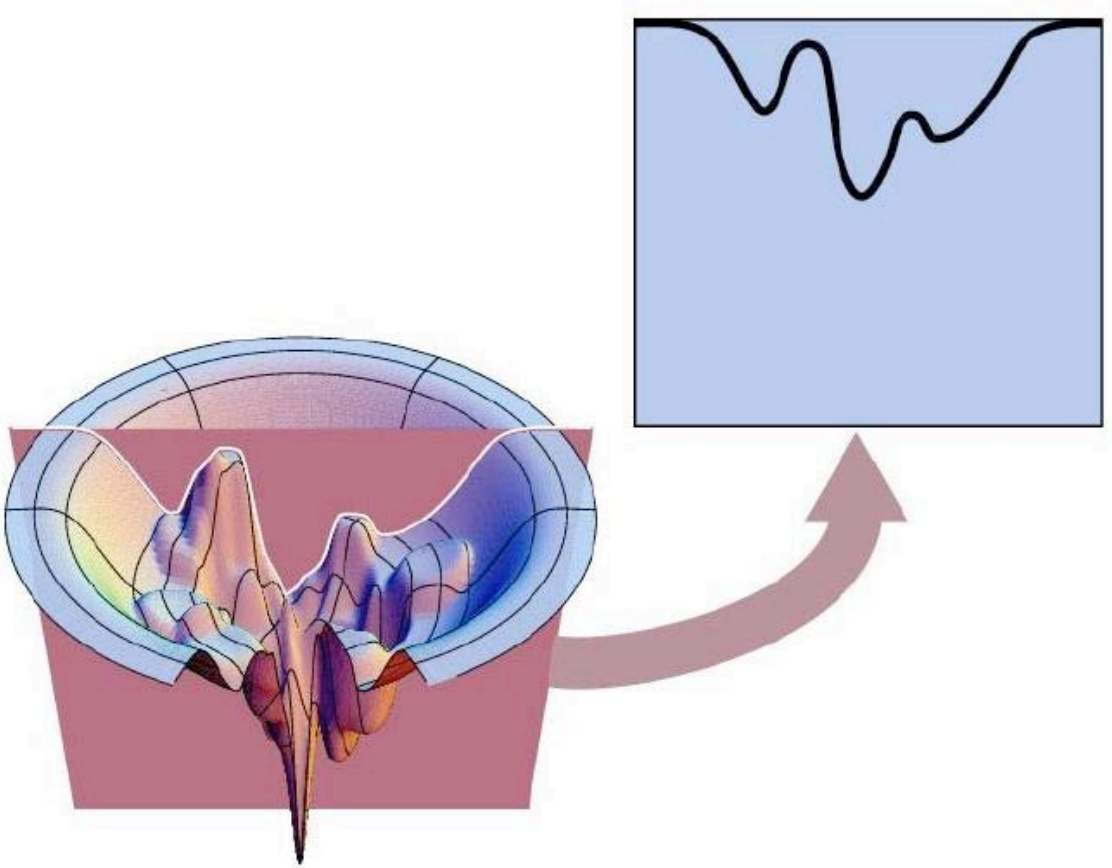
$$E_{es} = \sum_{(i,j)} \frac{332 q_i q_j}{\varepsilon r_{ij}}$$

$$E_{vdW} = \sum_{(i,j)} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right)$$

$$E_{hb} = \sum_{(i,j)} \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right)$$

$$E_{tors} = \sum_{l} U_l \left(1 \pm \cos\left(n_l \alpha_l\right)\right)$$

# Energy Landscape of Proteins



http://www.dillgroup.ucsf.edu/energy.htm

# To Address Problems of Simulations

## Minimal protein models

- Capture only **predominant** interactions in proteins (chain connectivity… )
- Allow only study of the **general characteristics** of folding
- **Review**: K.A. Dill & H.S. Chan, *Nature Str. Biol.* **4** (1997) 10

## Elaborated simulation techniques

- **Global optimization** techniques
- Evaluating **thermodynamic quantities** requires new **sampling techniques**
- **Review:** U.H. & Y. Okamoto, *Curr. Opp. Str. Biol*, **9** (1999) 177

# New Algorithms for Protein Simulations

- **Generalized-ensemble** techniques
  U.H.E. Hansmann & Y. Okamoto, *JCC* **14** (1993) 1333
- Algorithms relying on **Tsallis-like** weights
  U.H.E. Hansmann & Y. Okamoto, *PRE*, **56** (1997) 2228
- **Stochastic tunneling** and related methods
  W. Wenzel & K. Hamacher, *PRL.* **82** (1999) 3000
  U.H.E. Hansmann, *Eur. Phys. J. B* **12** (1999) 607
- Energy Landscape Paving (**ELP**)
  U.H.E. Hansmann & L. Wille, *PRL.* **88** (2002), 068105
- Multiple Markov Chains (**Parallel Tempering**, REM)
  C.J. Geyer *et al.*, *J. Am Stat Assn* **90** (**431**) (1995) 909;
  U.H.E. Hansmann, *Chem. Phys. Lett.* **281** (1997) 140
  W. Kwak and U.H.E. Hansmann, *Phys. Rev. Lett*. **95** (2005) 138102

# Simulation in Generalized Ensembles:

- **Idea:** choose ensemble that allows **better sampling**
- **Earliest realization:** umbrella sampling
  G.M. Torrie and J.P. Valleau, J. Comp. Phy. **23** (1977) 187
- Re-discovered in the 90's: multicanonical sampling, …
- Energy **barriers** can be crossed → **enhanced** sampling
- Problem: **Weights** are not *a priori* known
- What is the **optimal** ensemble?
- **Review:**
  U.H. & Y. Okamoto, in: D. Stauffer (ed), *Annual Reviews in Computational Physics VI*, World Scientific 1999, p.129

# Multicanonical Ensemble:

B.A. Berg and T. Neuhaus, Phys. Lett. , **B267** (1991) 249

- All **energies** enter with **equal probability**:

$$P_{mu}(E) \propto n(E) w_{mu}(E) = const$$

- The **multicanonical weight** factor has the form:

$$w_{mu}(E) \propto n^{-1}(E)$$

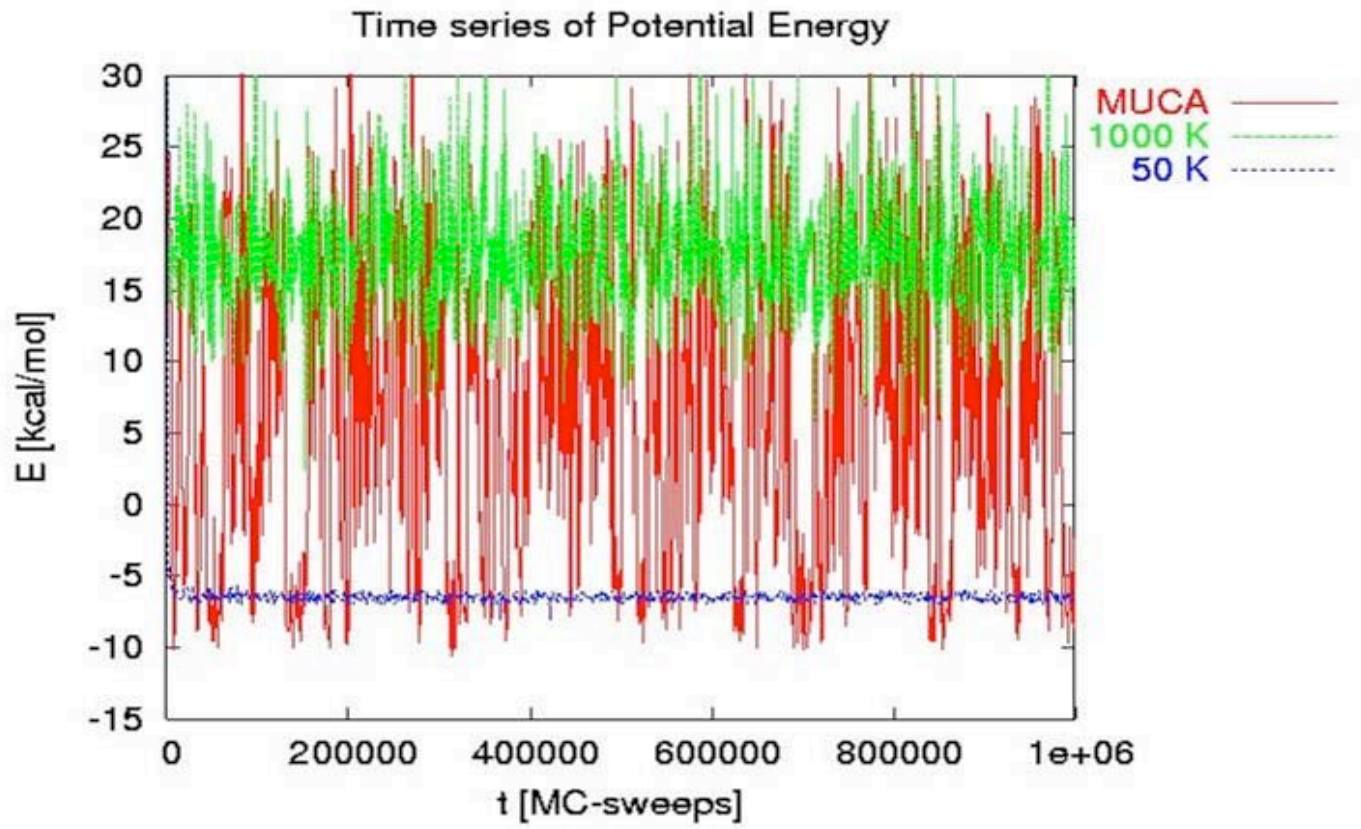
- Connection to the canonical ensemble by **re-weighting** :

$$P_B(T,E) \propto P(E) w_{mu}(E) e^{-E/k_B T}$$

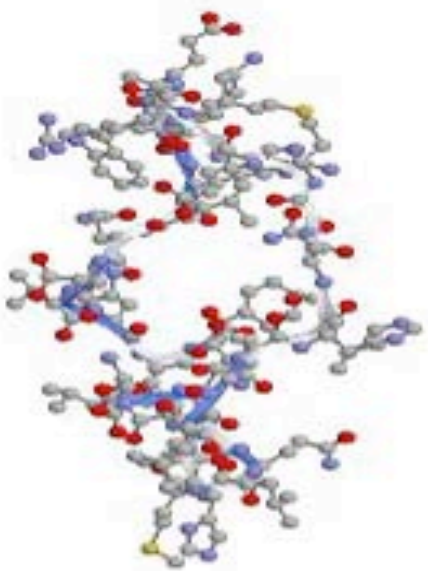
- Expectation values of **physical quantities**:

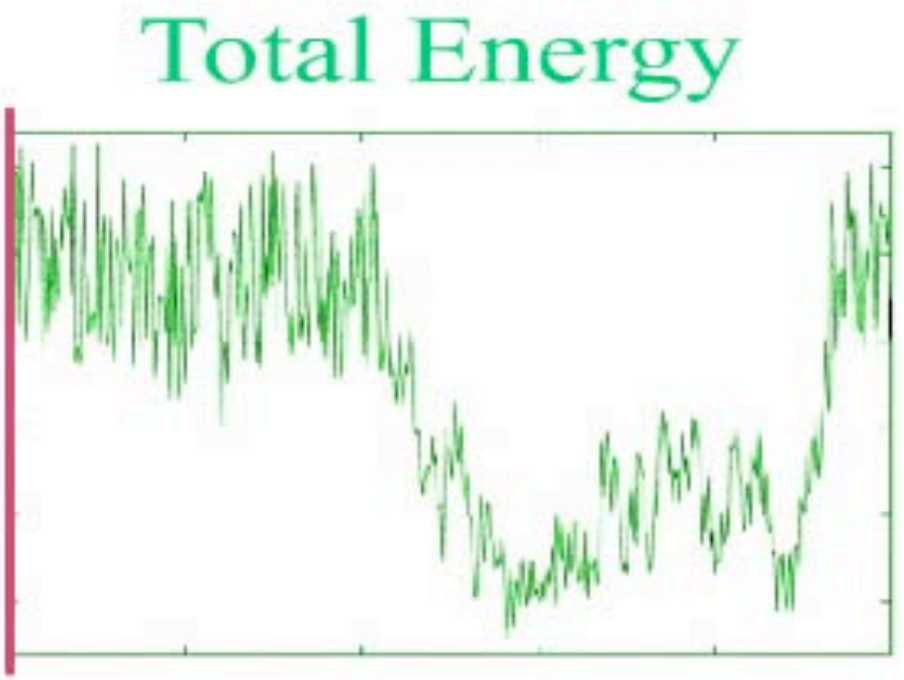
$$< O >= \frac{\int dx\, O(x) w_{mu}^{-1}(E(x)) e^{-E(x)/k_B T}}{\int dx\, w_{mu}^{-1}(E(x)) e^{-E(x)/k_B T}}$$

# Multicanonical Simulation of Met-enkephalin



# Multicanonical simulation of PTH(1-34)





# Energy Landscape Paving

U.H.E. Hansmann, L. Wille, *Phys. Rev. Let.* **88** (2002), 068105;
H.P. Hsu, S.C. Lin, U.H.E. Hansmann, *Act Cryst*. **A 58** (2002) 254

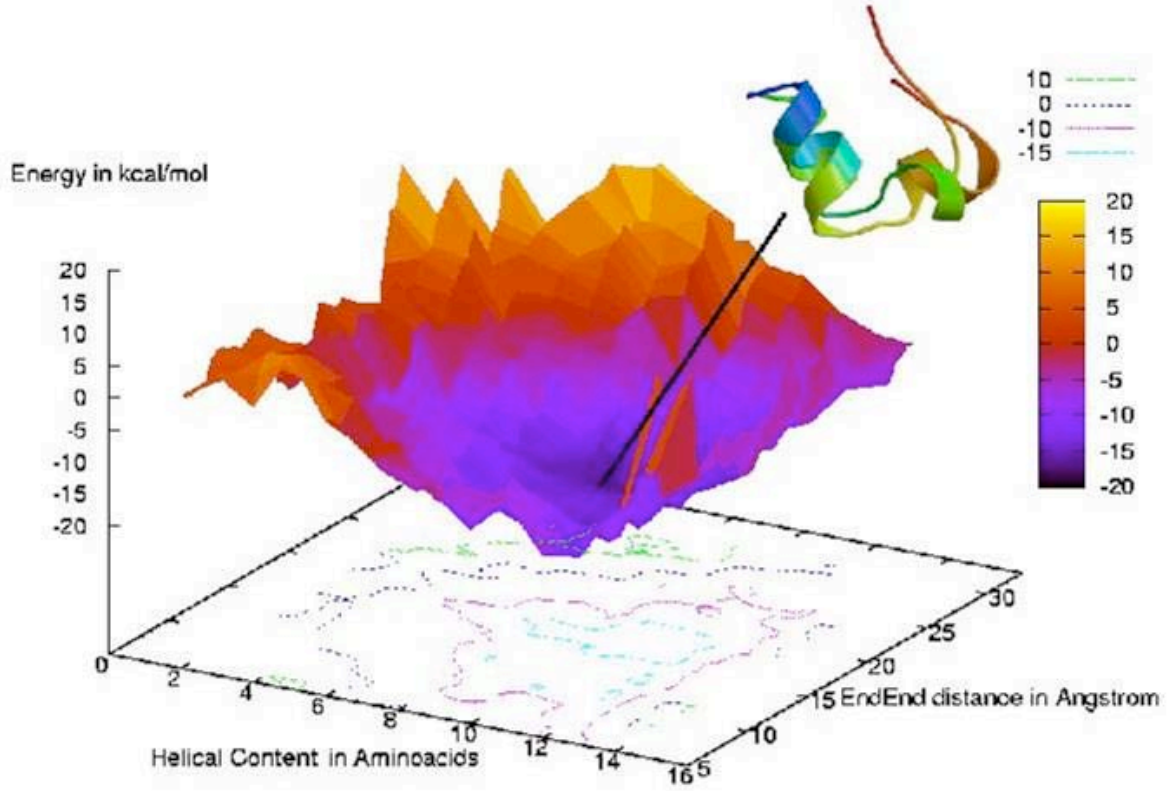
- **ELP** combines ideas from **tabu search** and **energy landscape deformation** approaches.
- Configurations are searched with **time-dependent** weights

$$w(E,q,t) = e^{-(E + f(H(q,t)))/k_B T}$$

- The low temperature *T* leads to drive toward **low energies**.
- The function ***f(H(q,t))*** drives simulation **out** of local minima.
- Often one chooses ***f(H(q,t))= H(q,t)*** or even *f(H(q,t)) = H(E,t)*

# Trp-cage protein (20 residues)

A. Schug, W. Wenzel & U.H.E. Hansmann, *J. Chem. Phys*., **122** (2005) 194711.



# Parallel Tempering (also known as REM)

U.H.E. Hansmann, *Chem. Phys. Lett.*, **281** (1997) 140

- N copies of the molecule at **different temperatures** T
- Parallel tempering uses two kinds of **updates**:

  **1.** Standard MC moves which effect only **single** copy

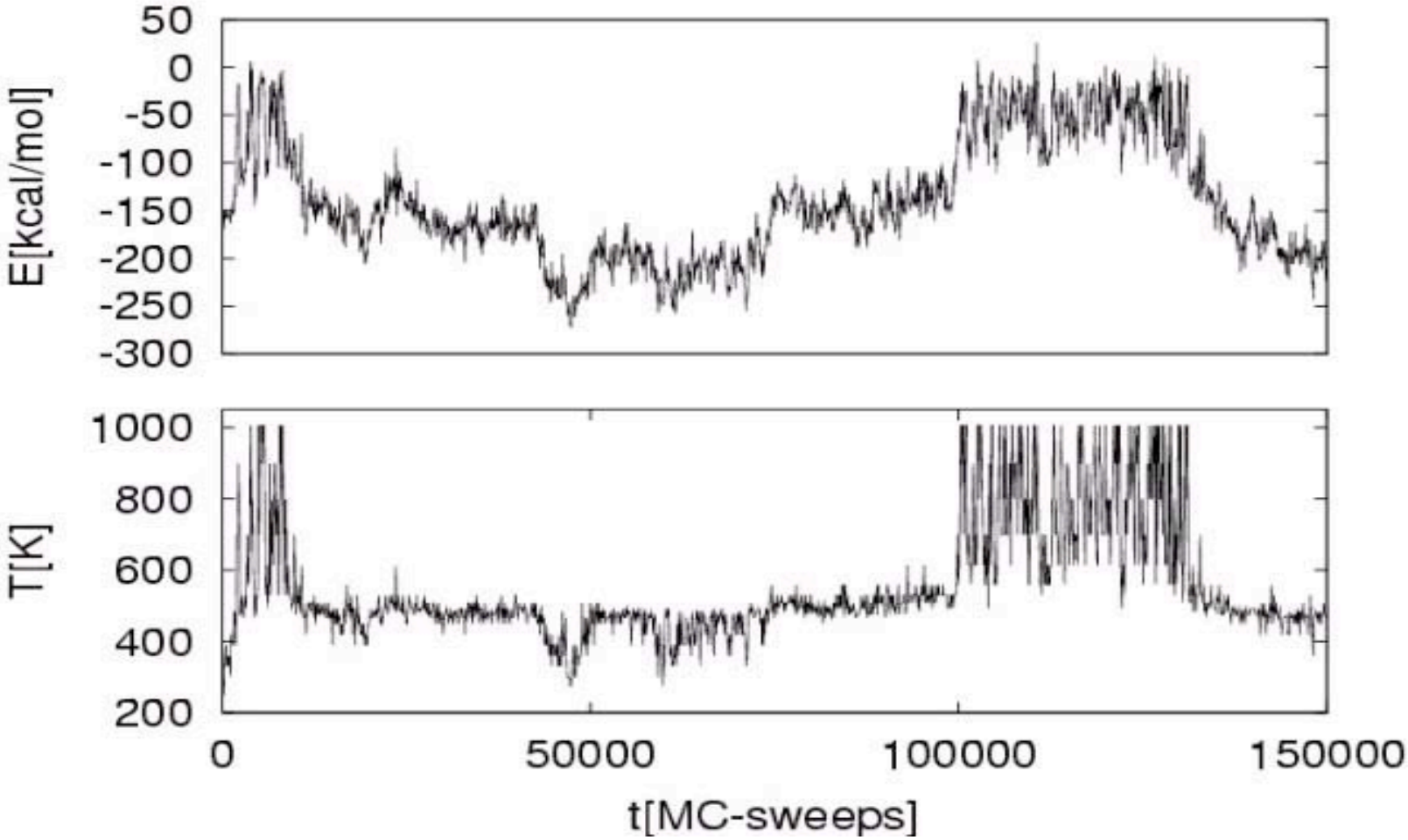
  **2.** **Exchange** of configurations between two copies i and j

$$w(C \rightarrow C') = \min\left(1, \exp\left\{-\frac{E(C_j)}{k_B T_i} - \frac{E(C_i)}{k_B T_j} + \frac{E(C_i)}{k_B T_i} + \frac{E(C_j)}{k_B T_j}\right\}\right).$$

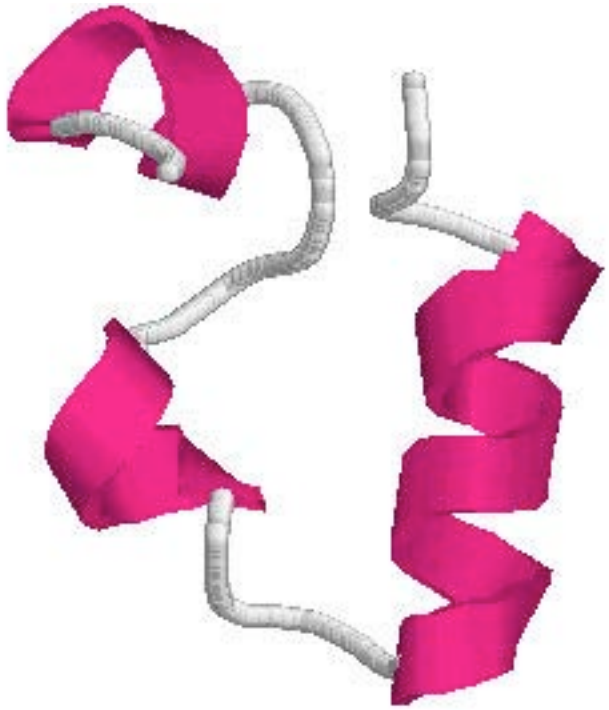
C.J. Geyer *et al., J Am Stat Assn* **90** (431) (1995) 909;
K. Hukushima *et al, J. Phys. Soc (Japan)* **65** (1996) 1604

- No restriction to **Boltzmann weights** or **temperature** ladders!

# Parallel Tempering Simulation of HP-36



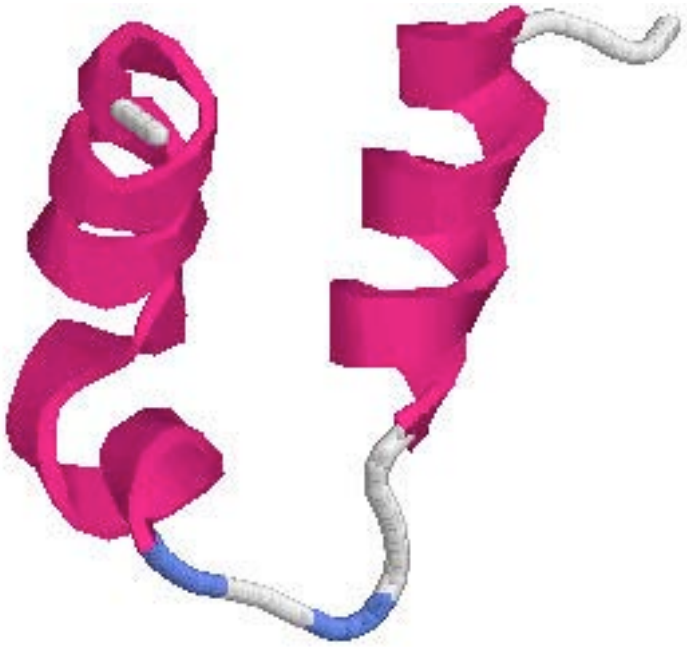
# PDB-structure of HP-36



# Lowest-energy configuration of HP-36

(simulation with solvent-accessible surface area term)

:

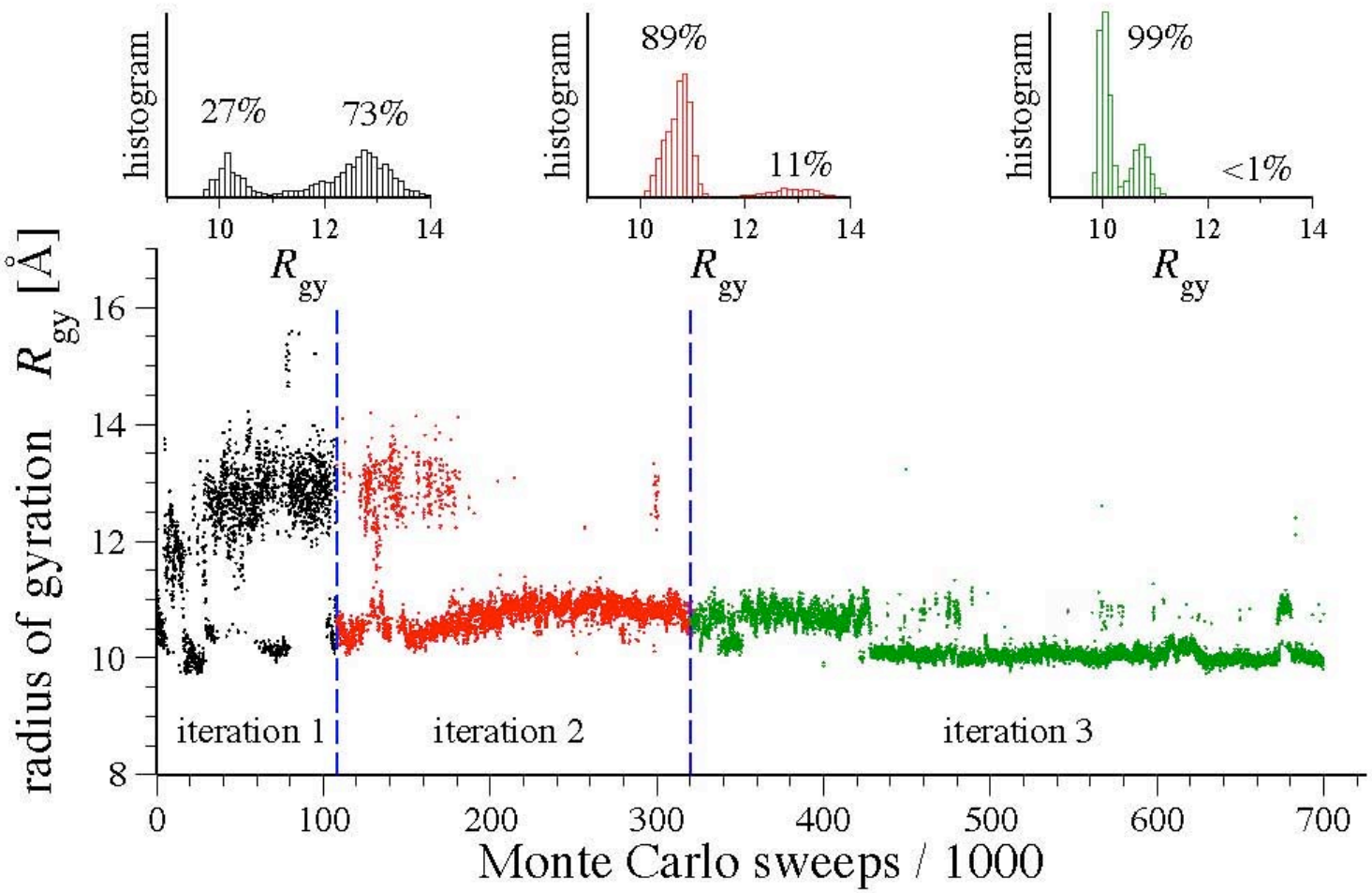


# Low-energy configuration of HP-36

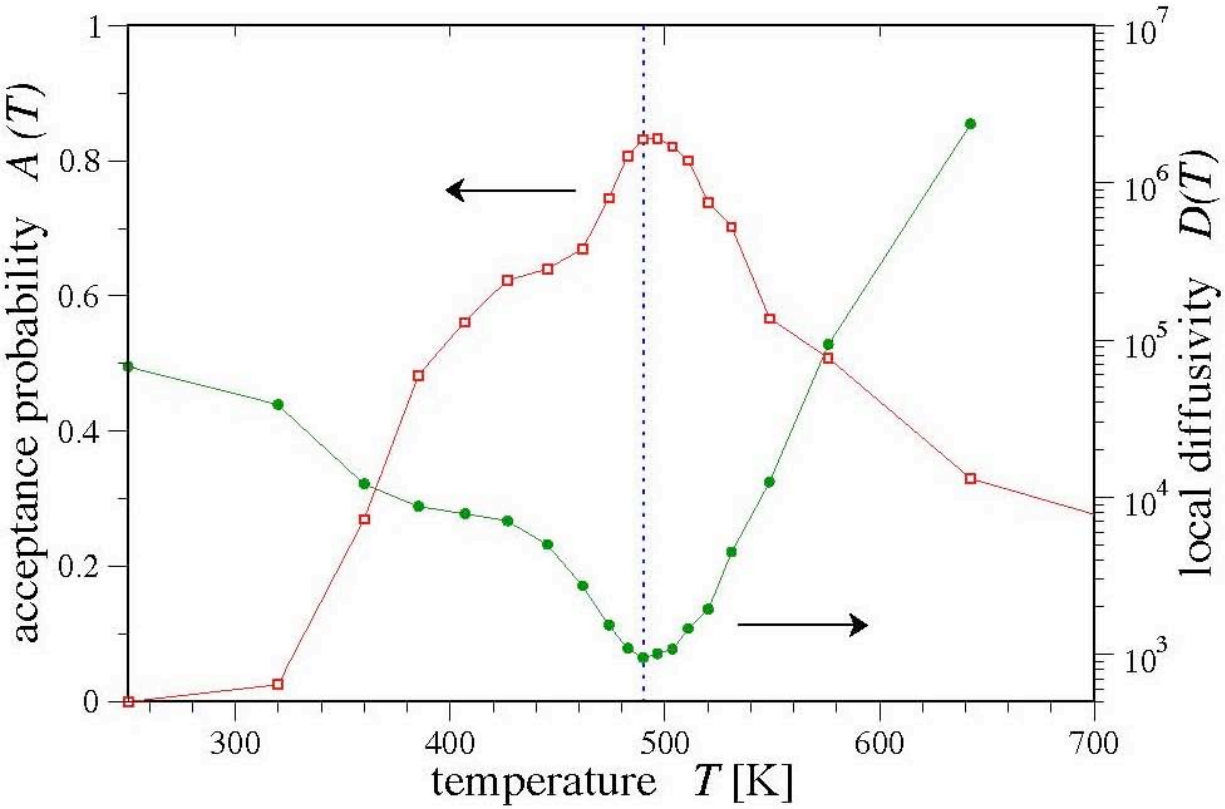
(simulation with solvent-accessible surface area term)

# Time series of RGY

S.Trebst, M. Troyer and U.H.E.Hansmann, J. Chem. Phys. 124 (2006) 174903

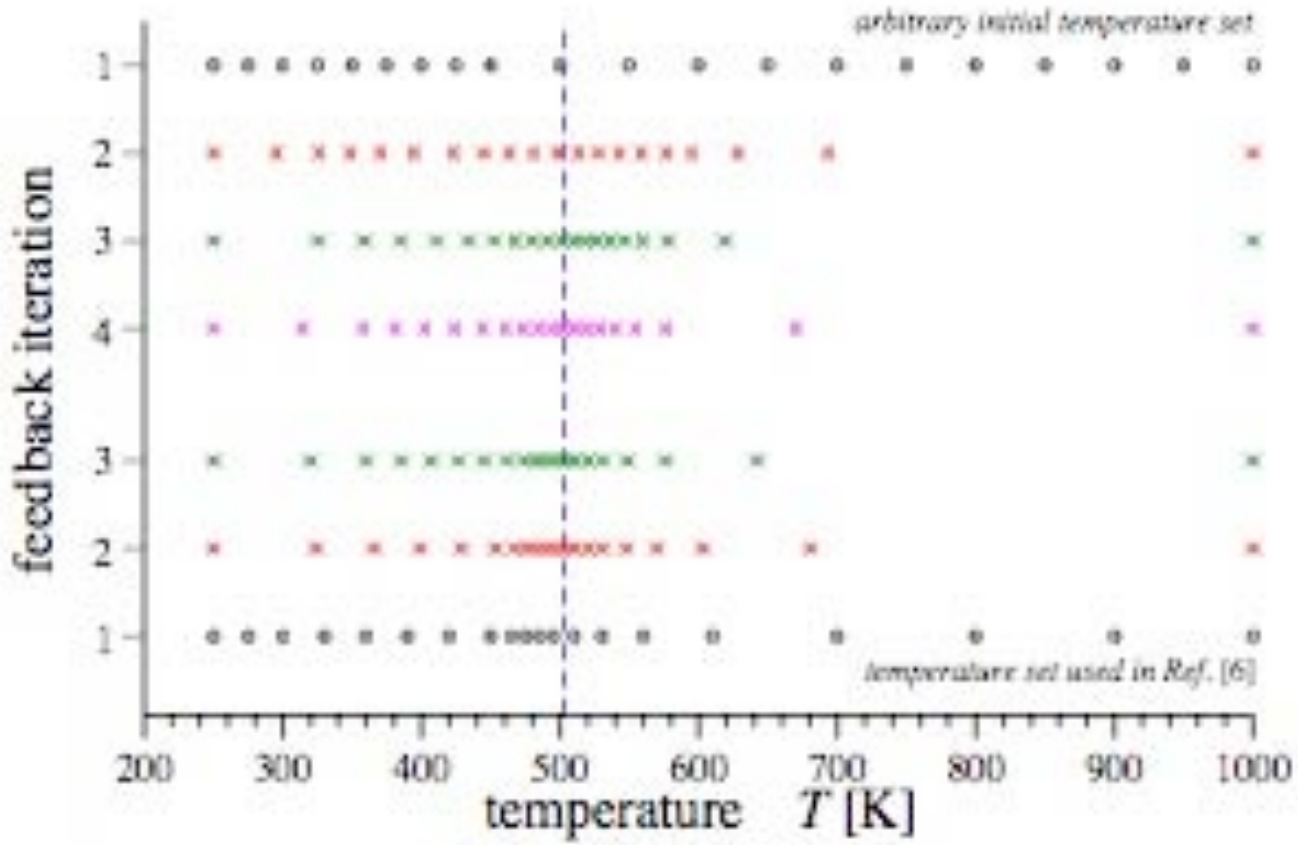


# Acceptance probabilty



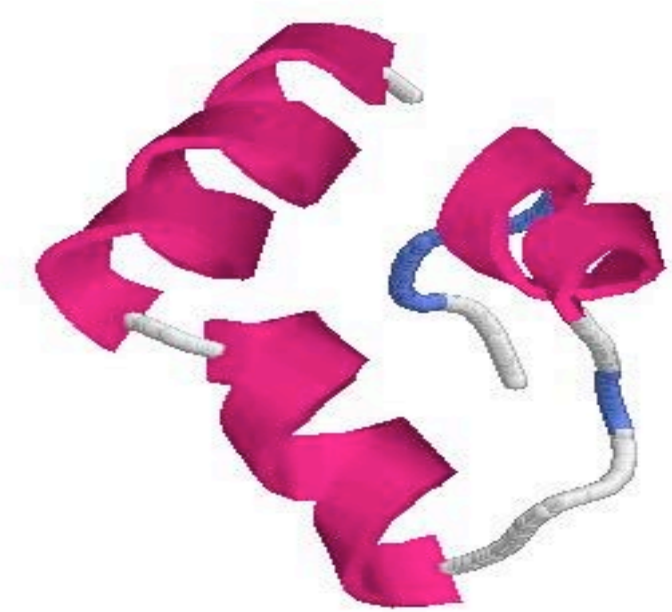
# Iteration of Temperature Distribution

S.Trebst, M. Troyer and U.H.E.Hansmann, J. Chem. Phys. 124 (2006) 174903



# Low-energy configuration of HP-36

(simulation with solvent-accessible surface area term)

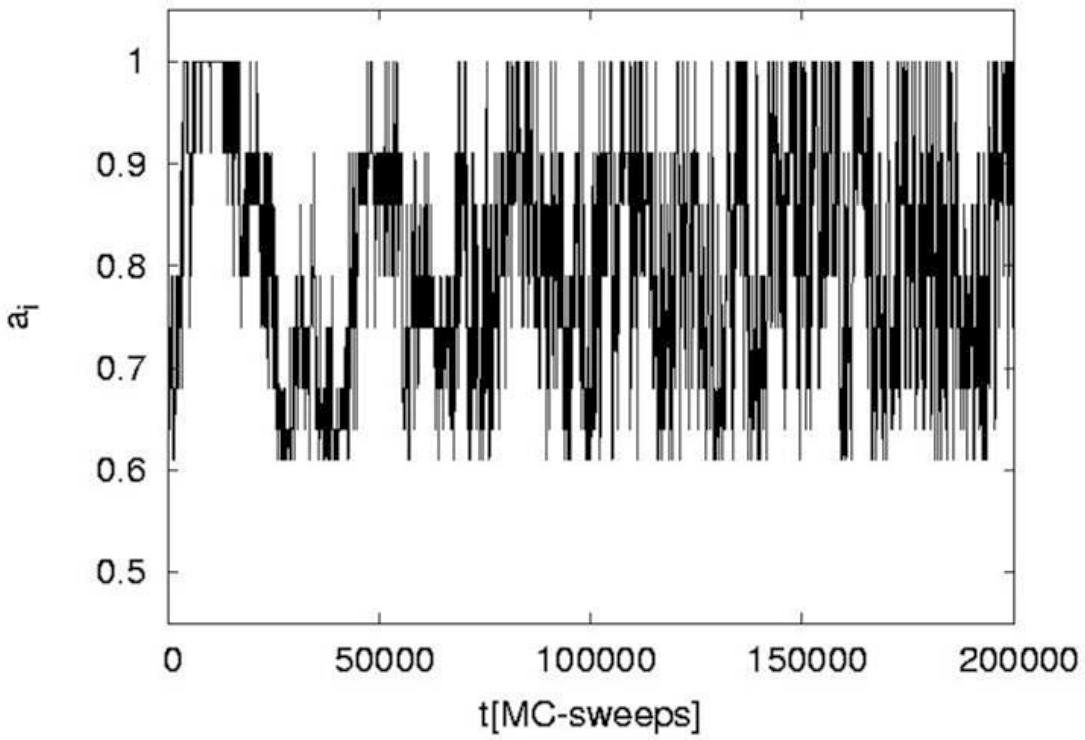


# “Model Hopping” in Protein Simulations

W. Kwak & U.H.E. Hansmann, PRL **95** (2005) 138102

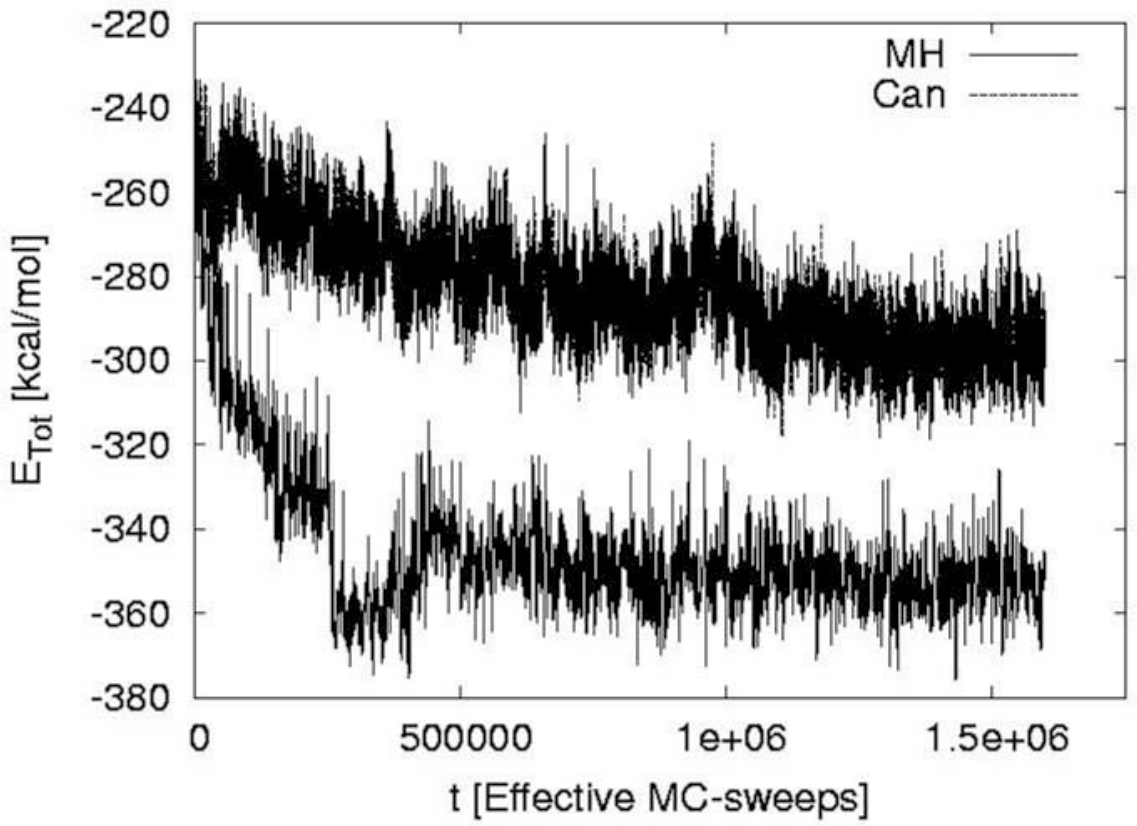
- Energy barriers often due to **vdW**-repulsion
- **Model Hopping** (MH) “tunnels” through barriers by random walk over non-physical models
- $E = E_{Rest} + E_{vdW} \rightarrow E_i = E_{Rest} + a_i E_{vdW}$
- $W(i,j) = \min(1,\exp(\Delta a \; \Delta E_{vdW}))$
- First test: **HP-36** and **protein A** fragment (48 AA)
- Other realization: multiscale modeling (in preparation)

# Time Series of Coupling Parameter in a Simulation of HP-36

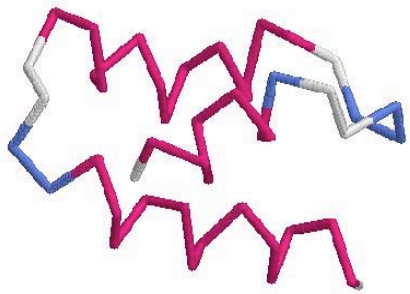


# Comparison of MH with Canonical Simulation

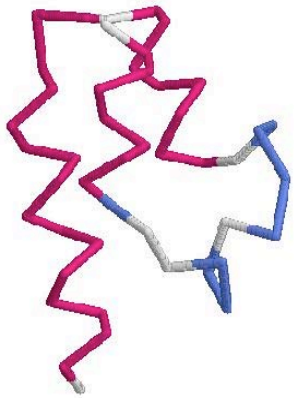
HP-36:



# Protein A configurations (OONS solvent)



PDB - structure



Lowest energy structure
Rmsd: 3.9 d

# Program Package SMMP

- **SMMP** (**S**imple **M**olecular **M**echanics for **P**roteins) is a modern package for simulation of proteins.
- Contains **generalized-ensemble** algorithms and other sophisticated simulation techniques.
- Runs also **parallel** computers
- Written in **FORTRAN**, a **C++** version is in preparation
- The program package is **freeware** and **open source** (http://www.phy.mtu.edu/biophys/smmp.htm)

Reference:

F. Eisenmenger, U.H.E.Hansmann, S.Hayryan & C.K.Hu
[SMMP] - A modern package for simulation of proteins
*Computer Physics Communications* **138** (2001) 192

# Helix vs. Sheet Formation

Y.Peng and U.H.E. Hansmann, *PRE*, **68** (2003) 041911.

- **α-helices** and **β-sheets** are common motifs in proteins
- "Mis-folding" → **aggregation**, often associated with **diseases**
- What **factors** govern formation of secondary structure?
- Our model: **EKAYLRT,** which forms both **α-helices** and **β-sheets**

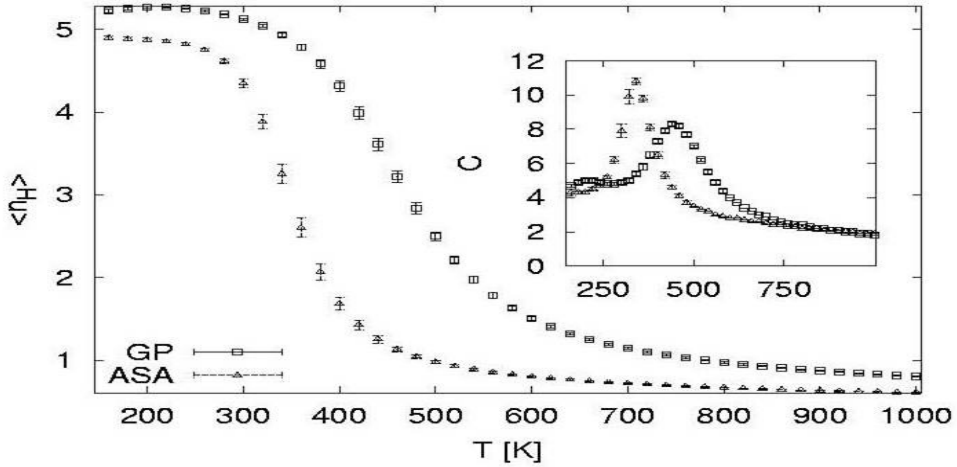
  S. Sudarsanam, *Proteins* **30** (1998) 228
- **Isolated** molecule and **interacting** with a β-strand.
- Simulations in **gas phase** and with an **implicit solvent**

  T.Ooi, *et al.*, *PNAS (USA)* **84** (1987) 3086

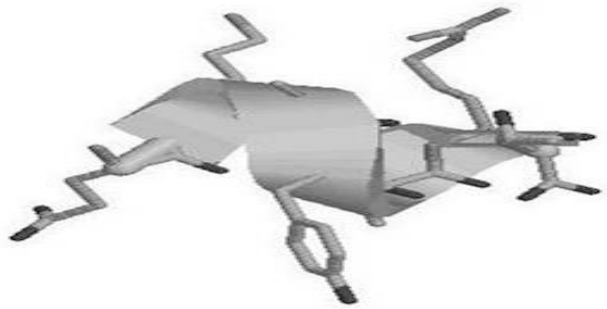
# Isolated EKAYLRT Molecule

Y.Peng and U.H.E. Hansmann, *PRE*, **68** (2003) 041911.

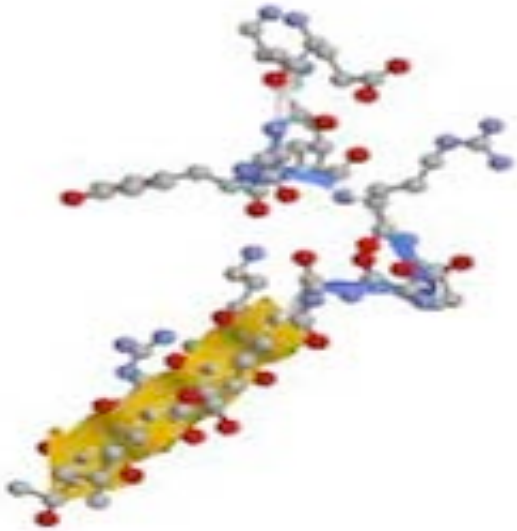
Helicity as function of temperature:




Ground-state structure of EKAYLRT:

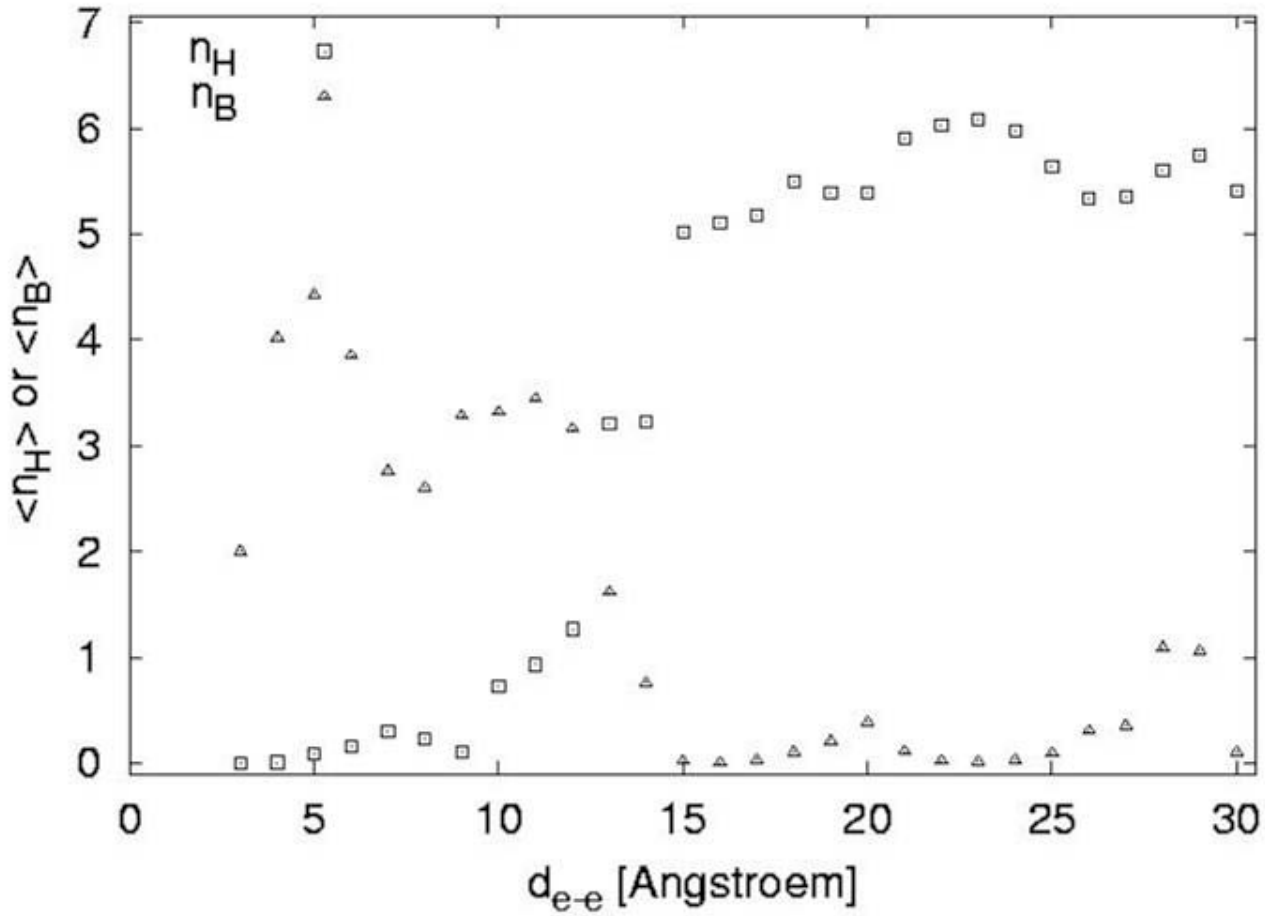


# Simulation of the Peptide Sequence EKAYLRT Interacting with a β-Strand



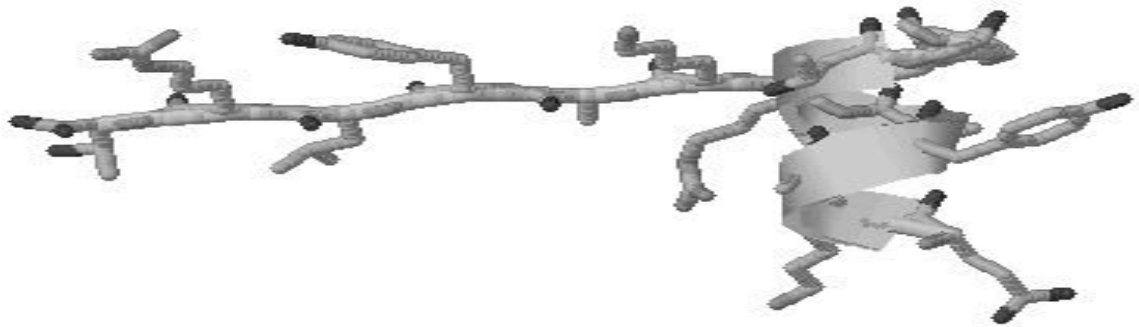
# EKAYLRT interacting with a β-strand

Y.Peng and U.H.E. Hansmann, *PRE*, **68** (2003) 041911.

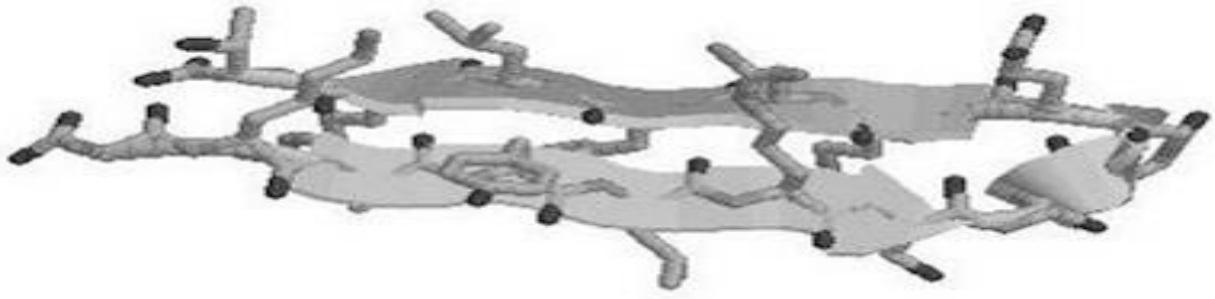


# Low-energy structures of EKAYLRT interacting with a β-strand

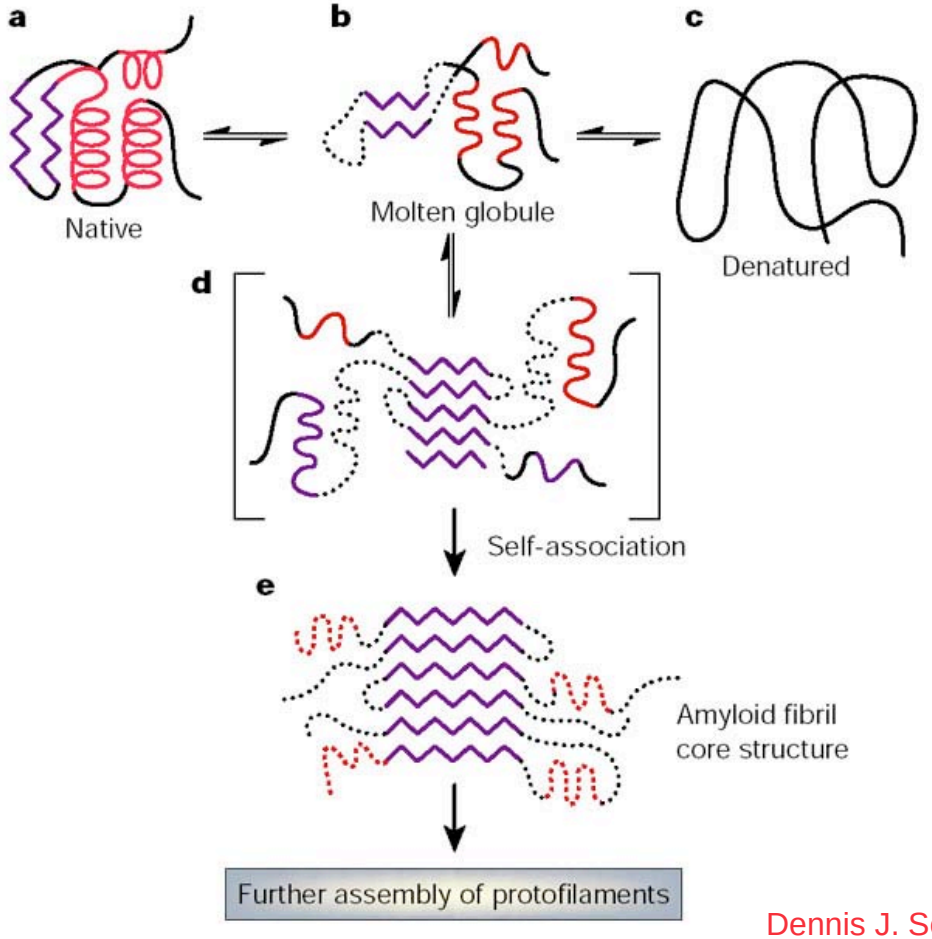
Large end-to-end distance $d_{e\text{-}e}$:



Small end-to-end distance $d_{e\text{-}e}$:



# Aggregation

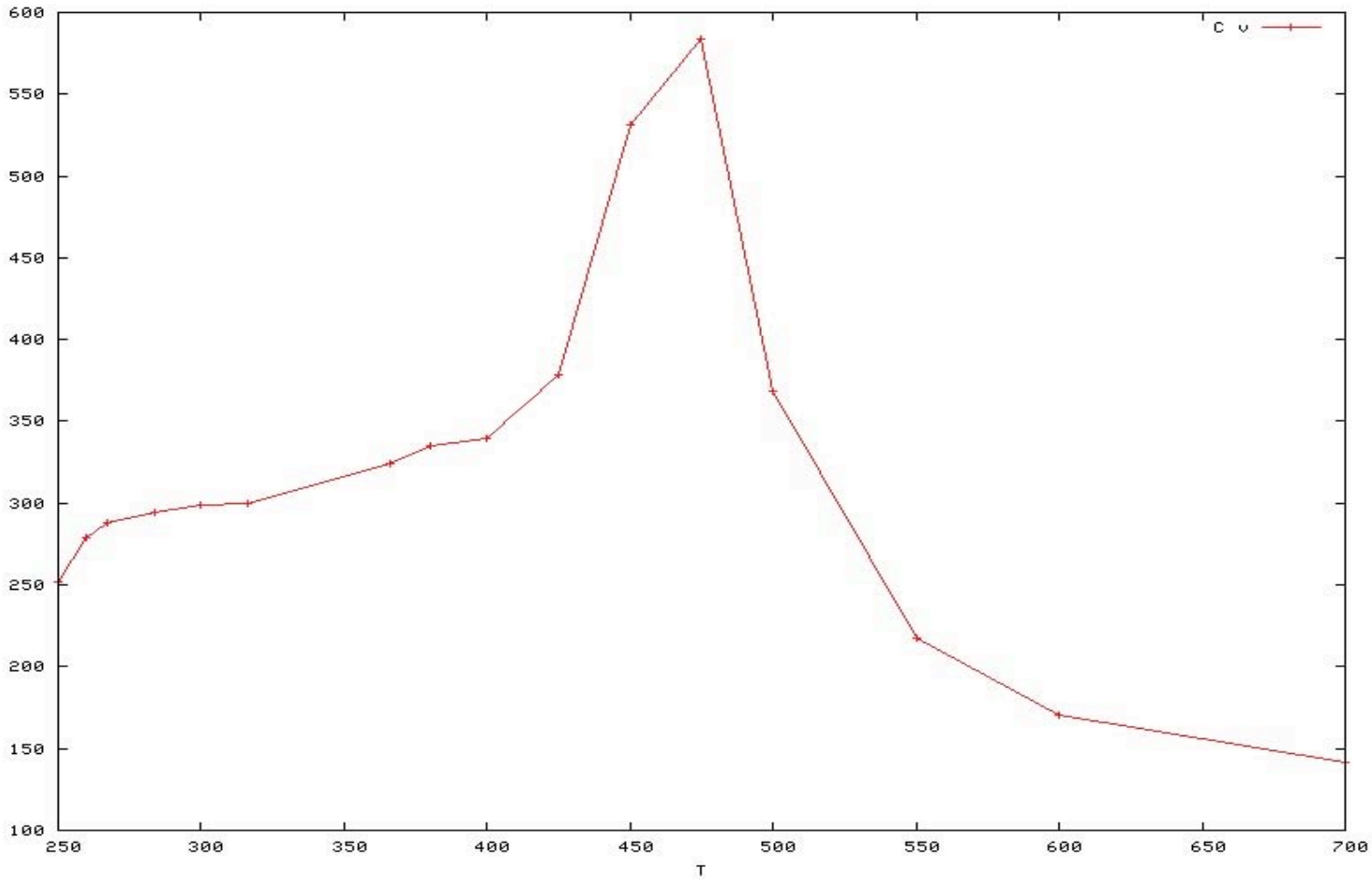



Dennis J. Selkoe Nature 426, 900-904

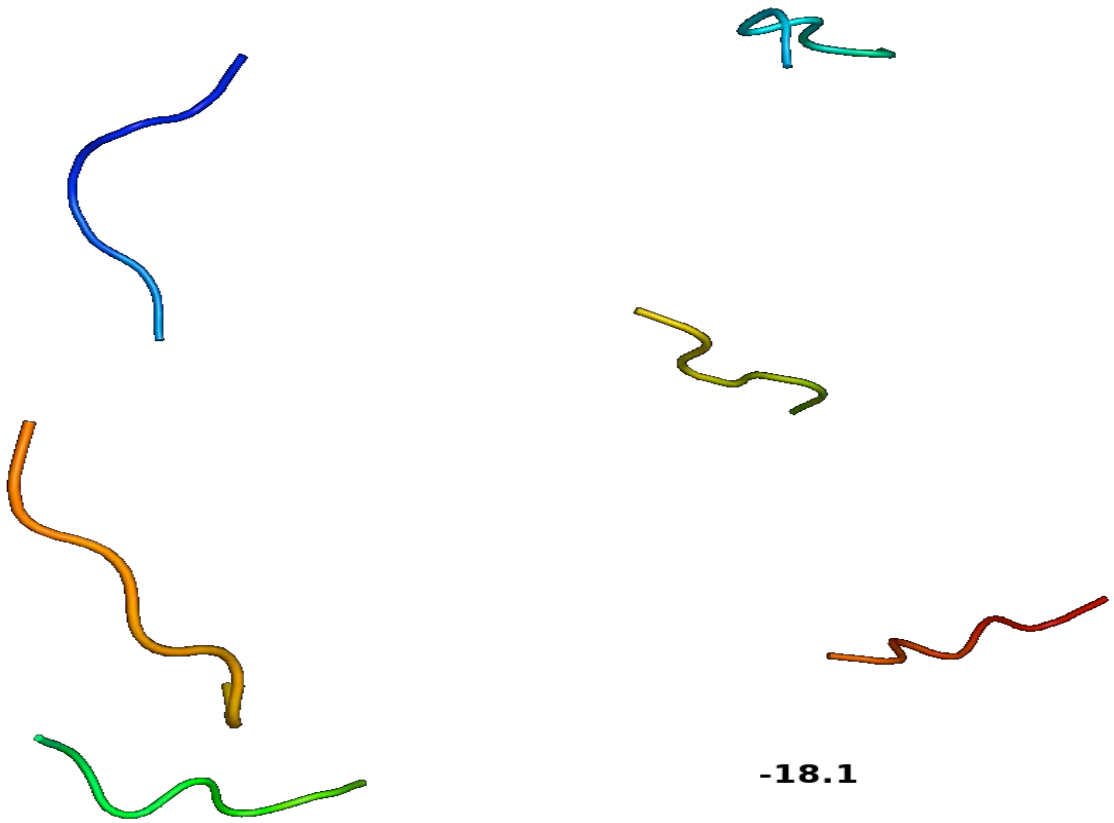
# Aggregation of β -Amaloyd 16–22

- **Fibrils** build out of mis-folded β - Amaloyd peptides are related to outbreak of **Alzheimer disease**.
- Aggregated peptides show **high β - strand** content
- We focus on **segment 16-22** which has high β - strand propensity
- Can we observe fibril formation ***in silico***?

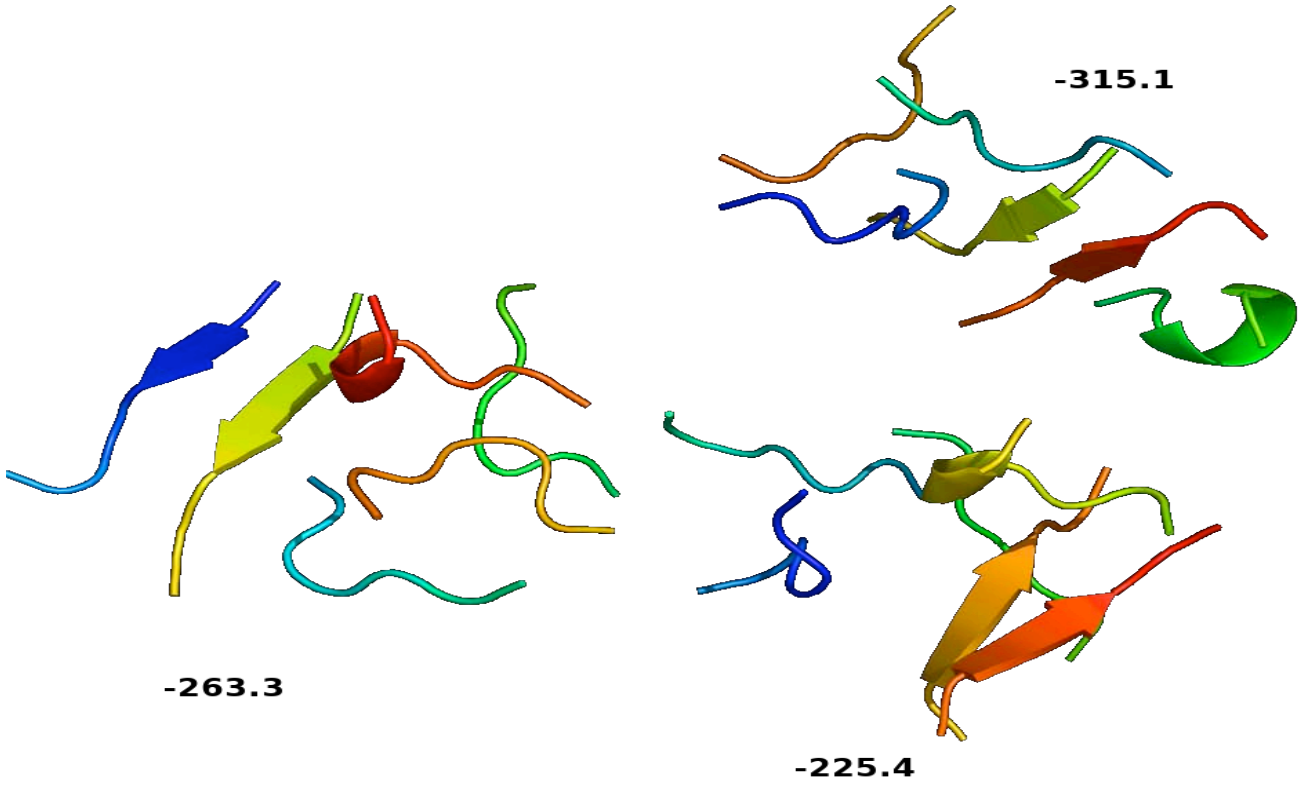
# Spec. Heat



# “Free” chains

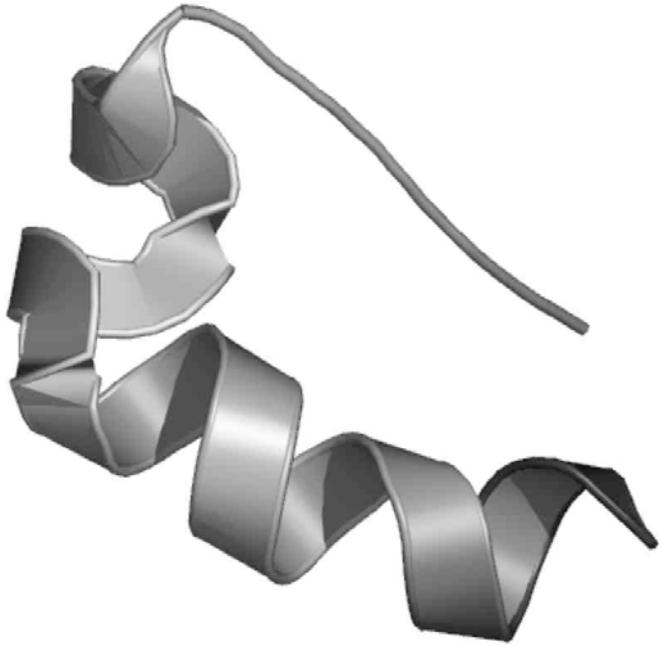


# Low-energy configurations

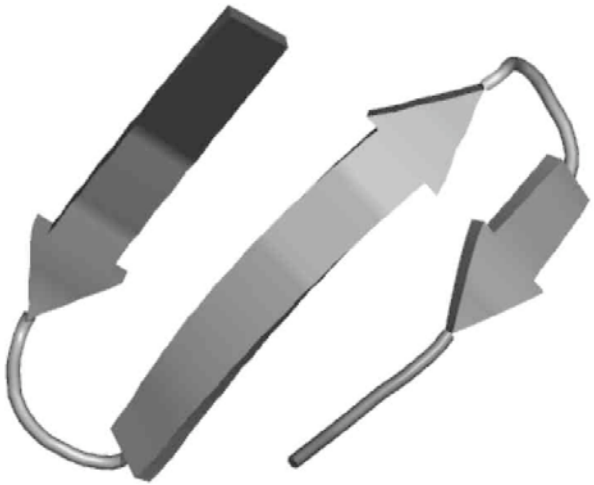


# Three small proteins

S. Mohanty and U.H.E. Hansmann, Biophysical Journal **92** (2006) 3537



1RIJ

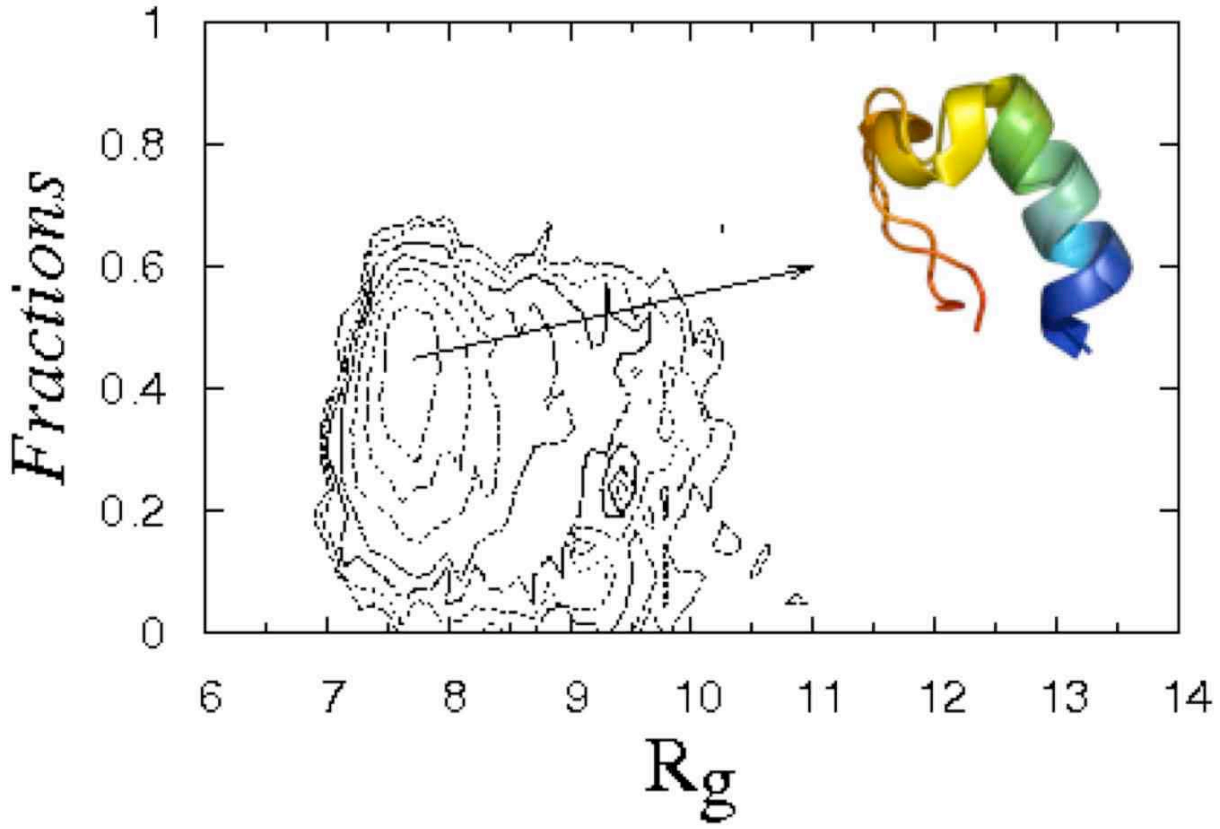


beta3s

bba5

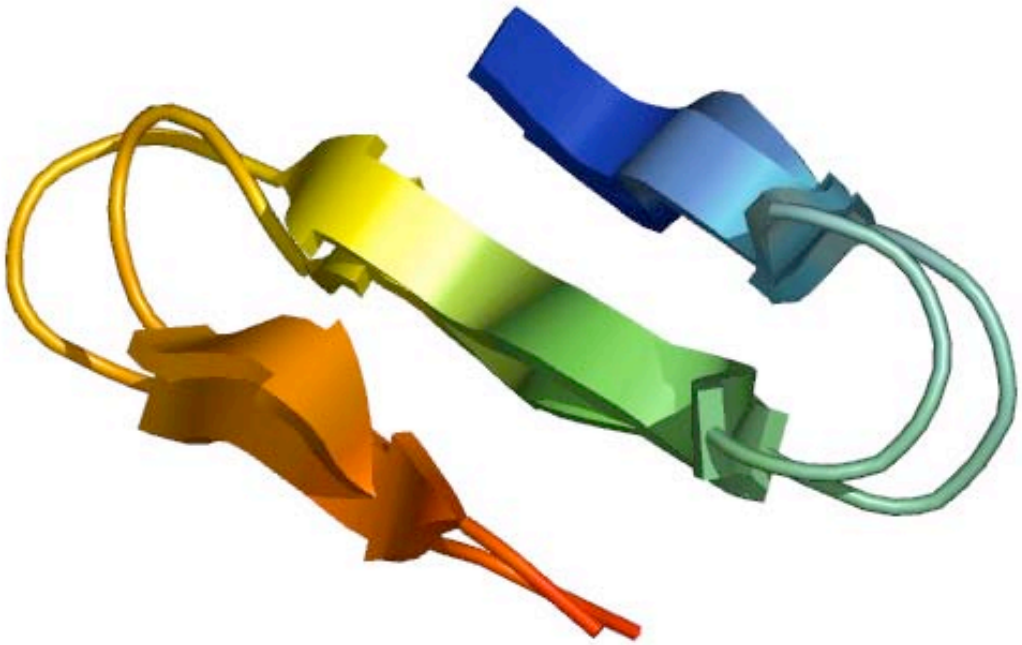
# Energy Landscape of 1RIJ

S. Mohanty and U.H.E. Hansmann, Biophysical Journal **92** (2006) 3537



# Lowest energy configuration

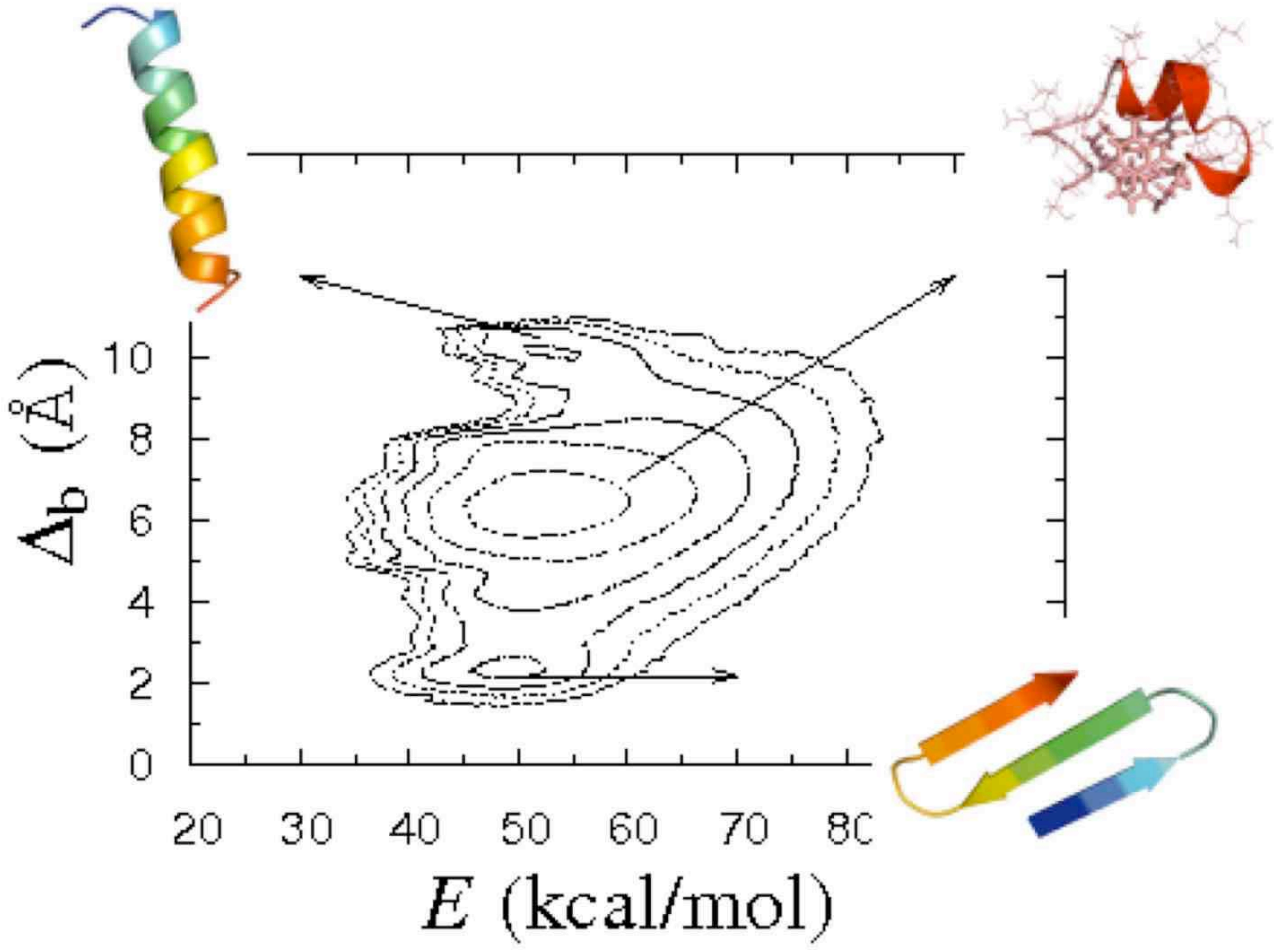
S. Mohanty and U.H.E. Hansmann, Biophysical Journal **92** (2006) 3537



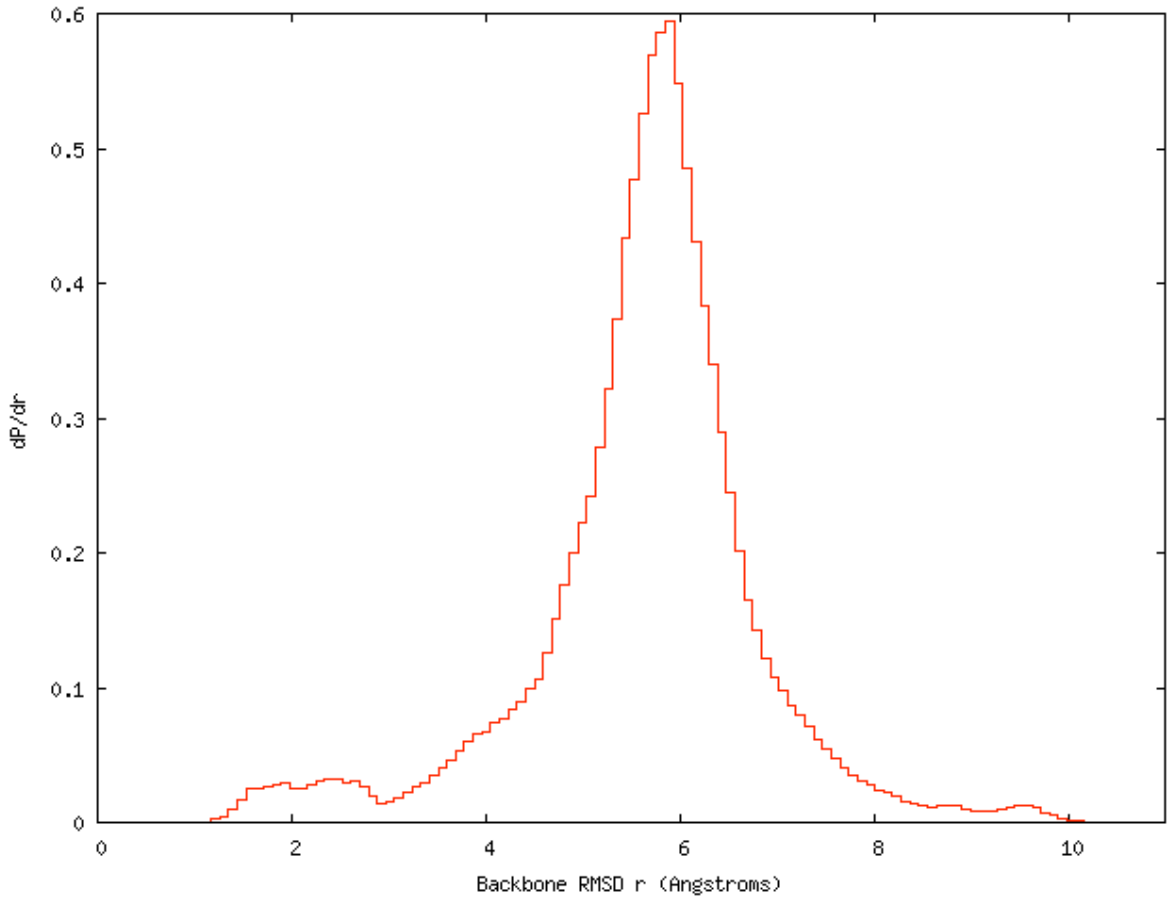
Rmsd: 2 Å

# Energy landscape of beta3s

S. Mohanty and U.H.E. Hansmann, Biophysical Journal **92** (2006) 3537

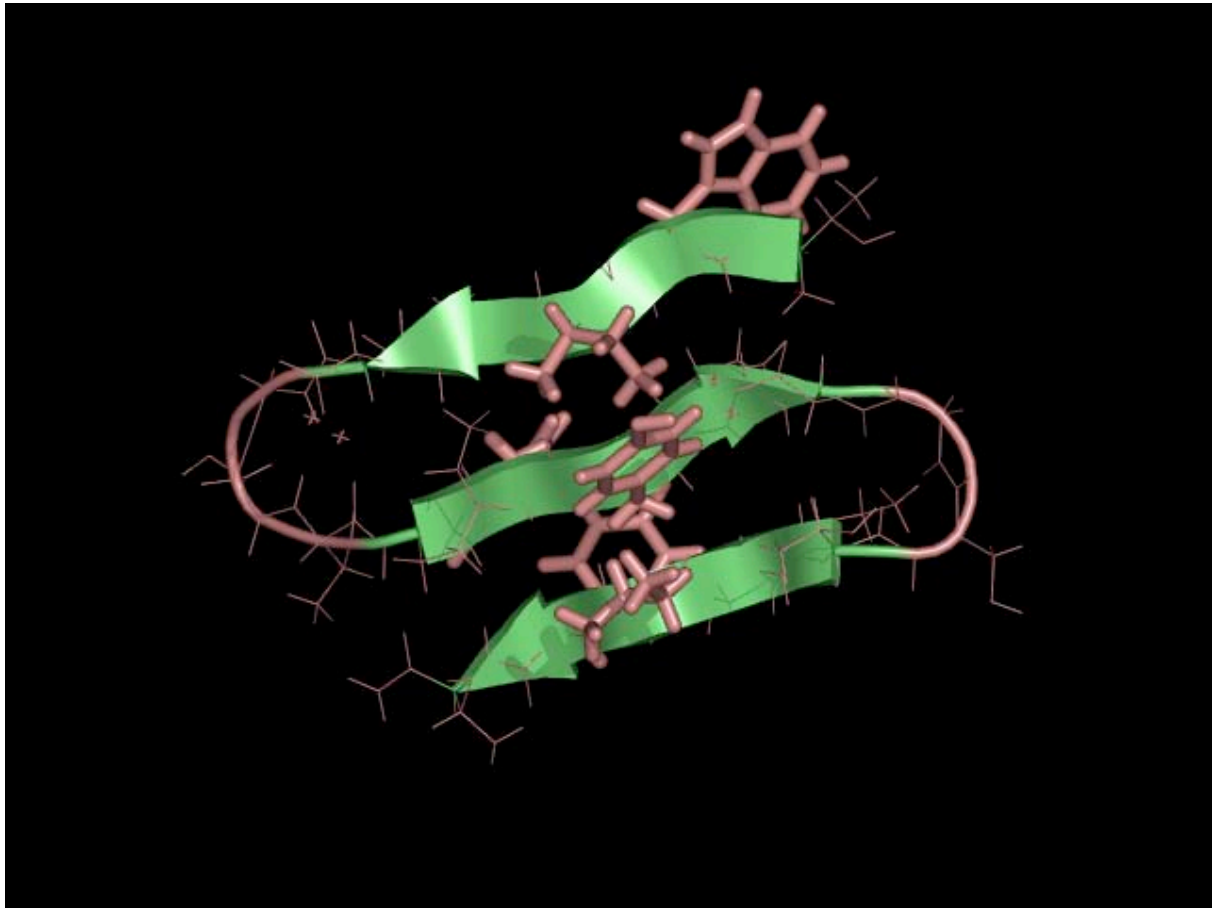


# Propensity of configurations



# Folding of Beta3s

S. Mohanty and U.H.E. Hansmann, Biophysical Journal **92** (2006) 3537



# Folding of Beta3s

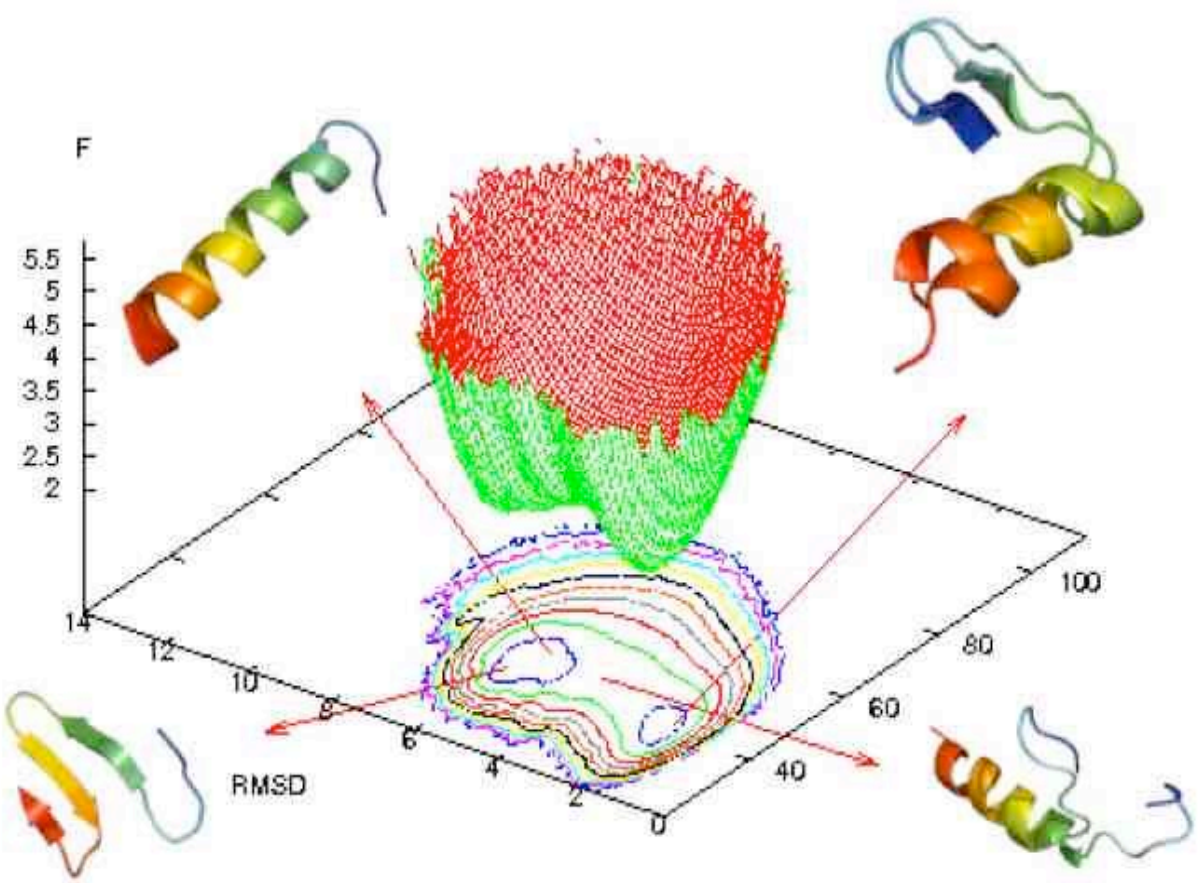
- Collapse **precedes** secondary-structure formation
- **Zipper**-like formation of hairpins
- No particular order in that hairpins are formed
- Once formed it **catalyzes** formation of second hairpin

# Folding of BBA5

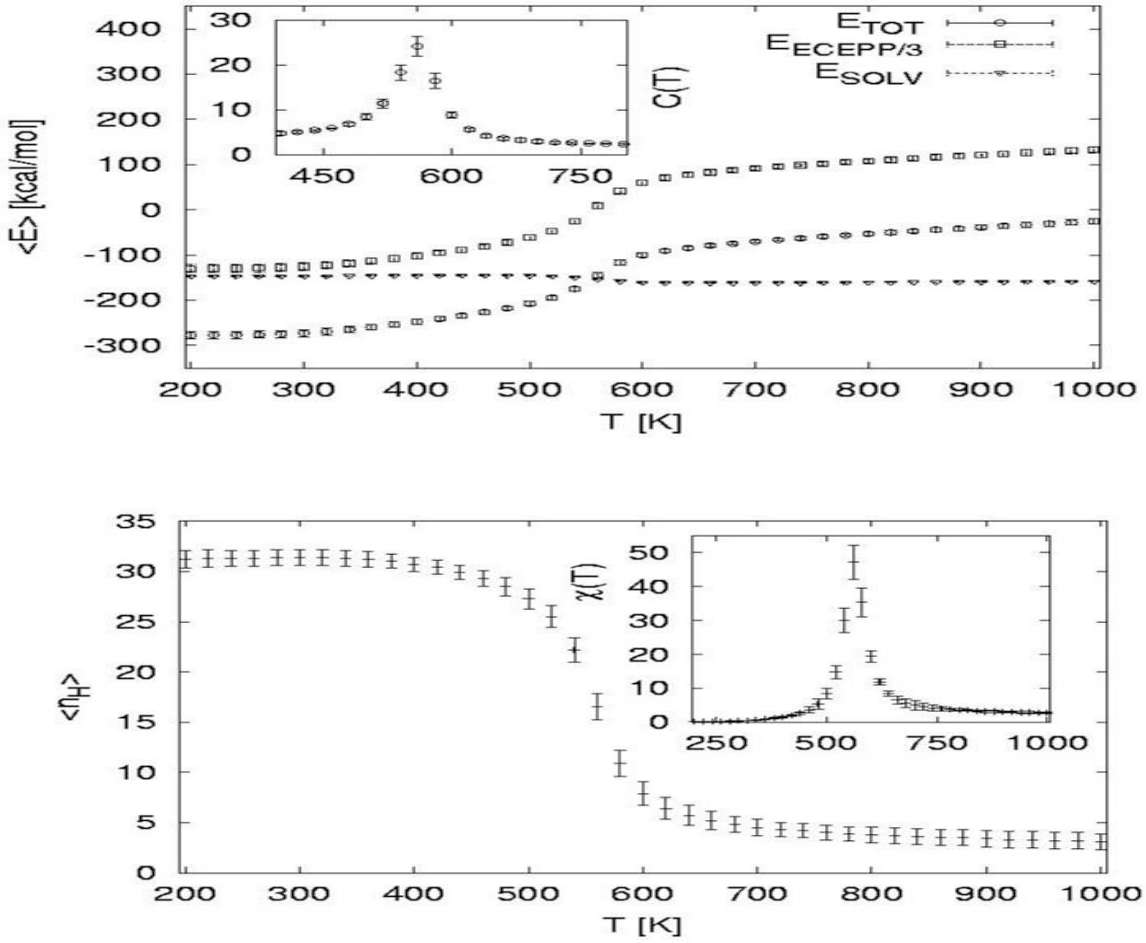
- Both **helix** and **sheet**
- Secondary structure elements form **independently**
- Both  formed **before** folding into final shape

# Energy landscape of BBA5

S. Mohanty and U.H.E. Hansmann, Biophysical Journal **92** (2006) 3537

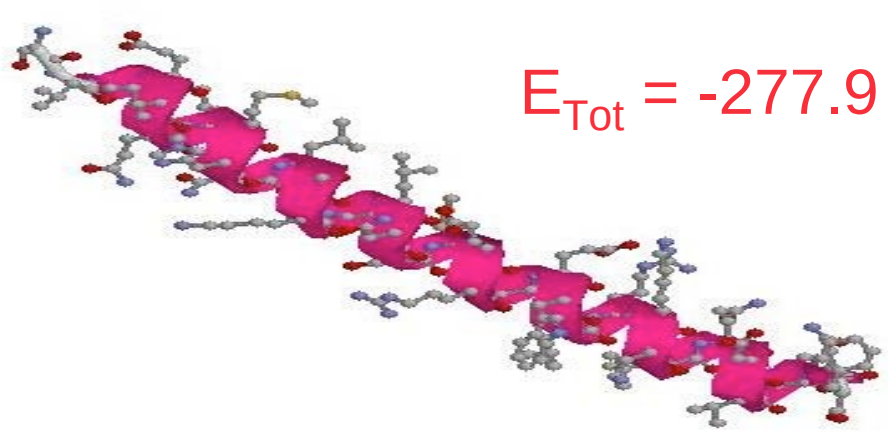


# Thermodynamics of PTH(1-34)



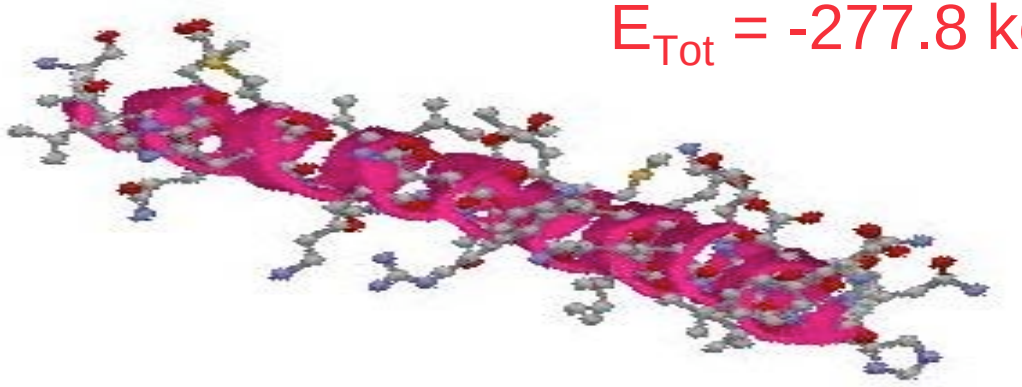

Transition Temperature: $T_\theta = 560 \pm 20$ K

# Crystal Structure (1ET1) of PTH(1-34)



$E_{Tot}$ = -277.9 kcal/mol

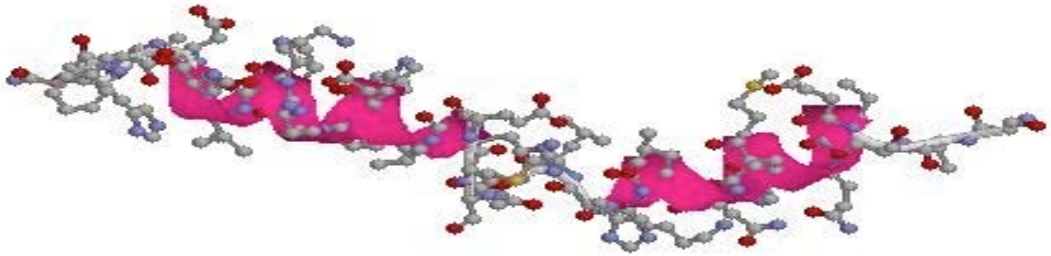
## Lowest-energy structure



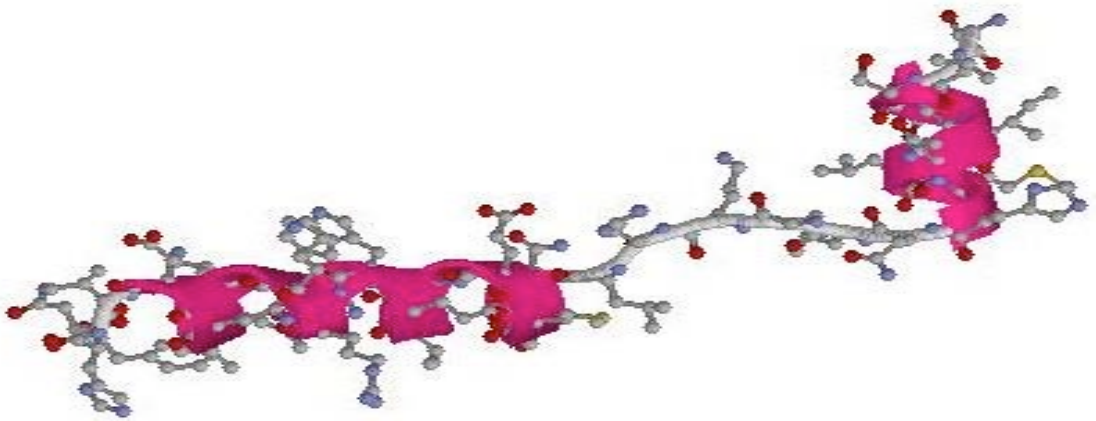
$E_{Tot}$ = -277.8 kcal/mol

RMSD: 0.9 Å

# NMR Structure (1HPY) of PTH(1-34)



Structure found at T=540 K

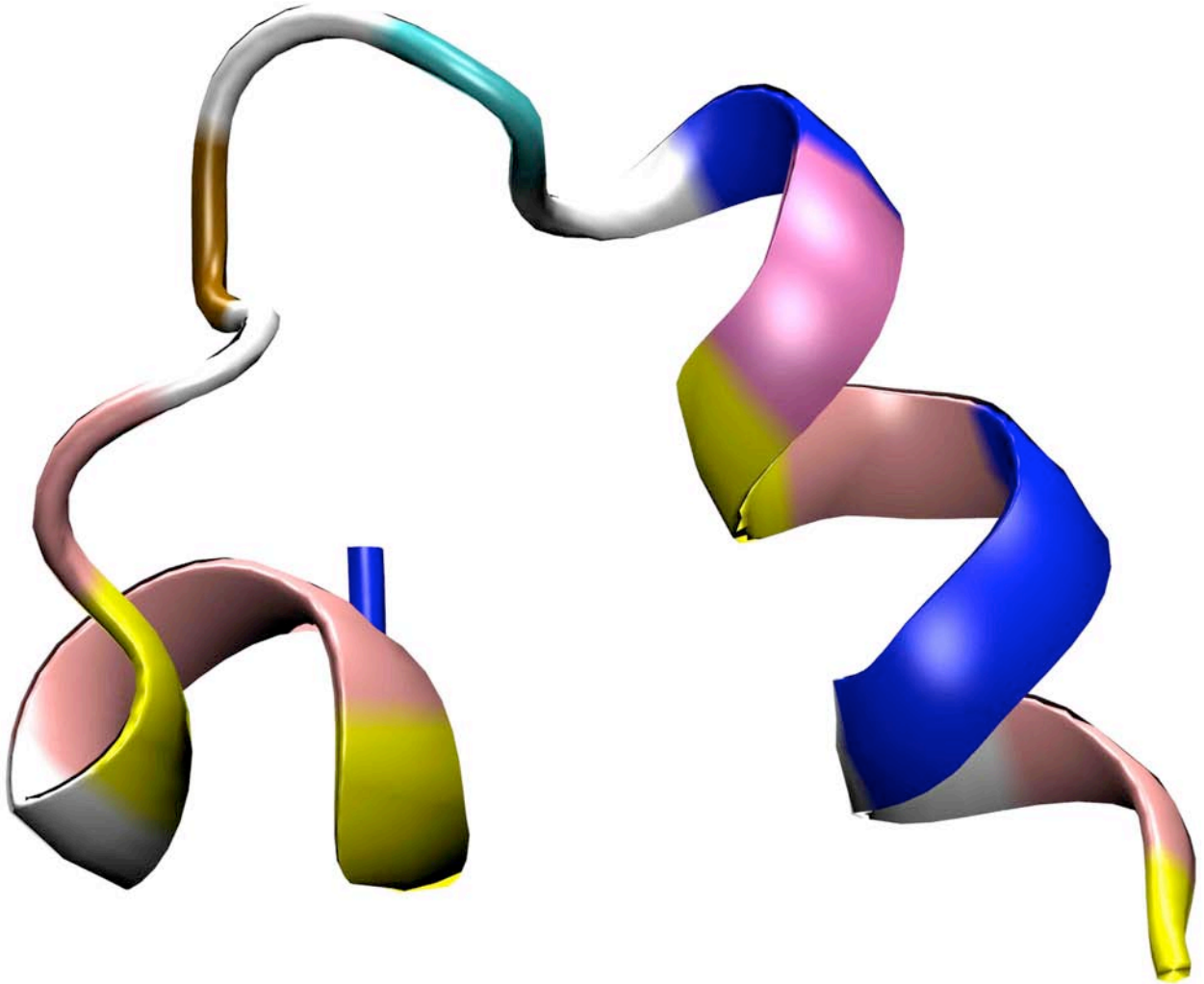


# PTMD of a Signal Peptide

S. Höfinger and U.H.E. Hansmann, Proteins, in press

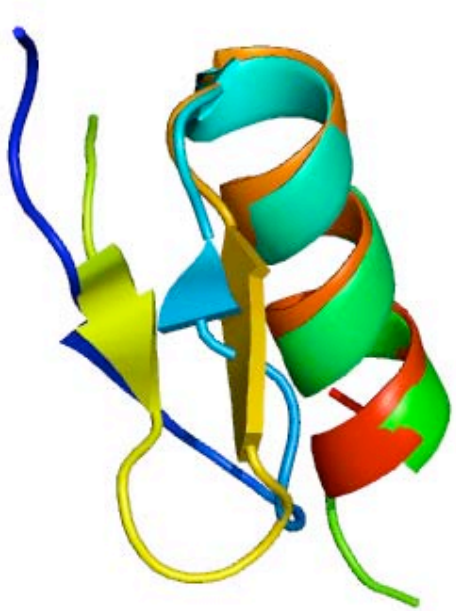
- Signal peptides have strong **hydrophobic** character
- Folding only in **membran** environment?
- But biological considerations suggest folding in **aqueous** environment
- **22-residue** signal peptide of rat liver aldehyde dehydrogenase
- NMR analysis shows a **helix-loop-helix** motif
- Parallel tempering molecular dynamics in **explicit water**

# Lowest-energy configuration of the SP



# 28 residue-protein FSD

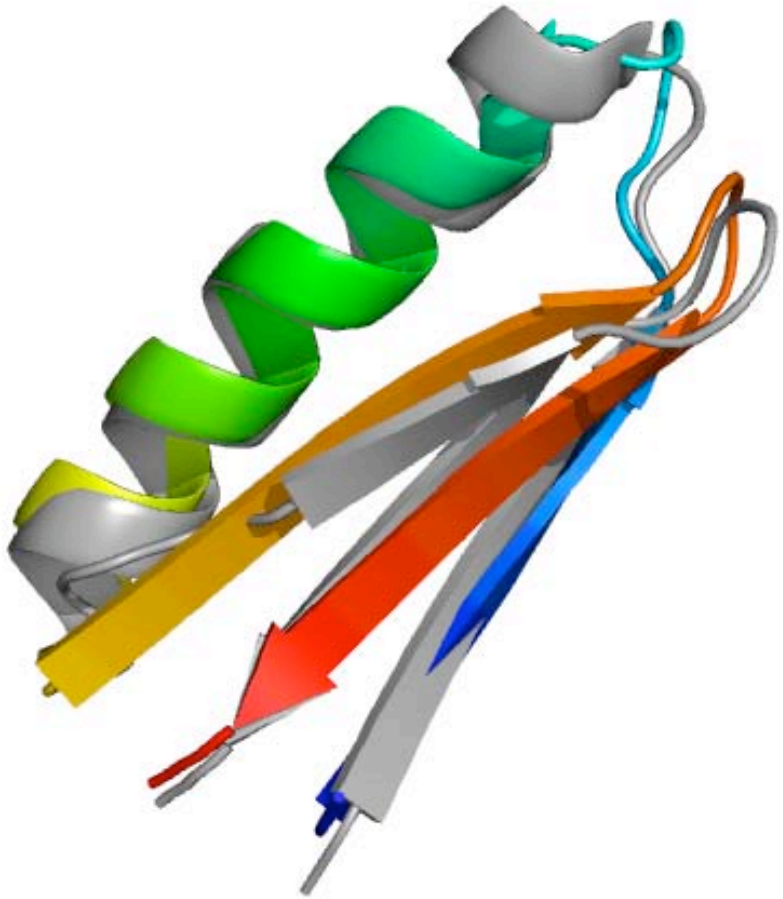
S. Mohanty and U.H.E. Hansmann, in preparation



Rmsd: 2.8 Å

# 49 residue-protein Cfr

S. Mohanty and U.H.E. Hansmann, in preparation



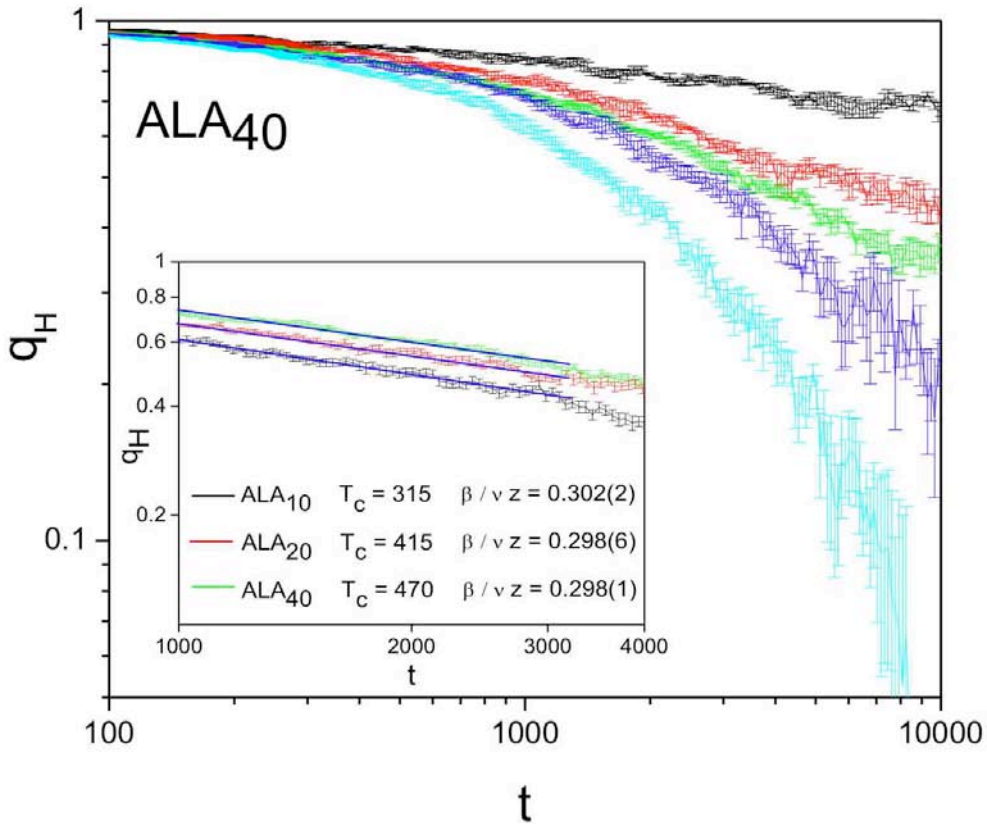
Rmsd: 1.9 Å

# Conclusion

- We have now **efficient** simulation techniques
- **Experimental results** can be reproduced for small proteins
- What happens for **larger molecules** (protein L)?
- **Limitations** of protein models?

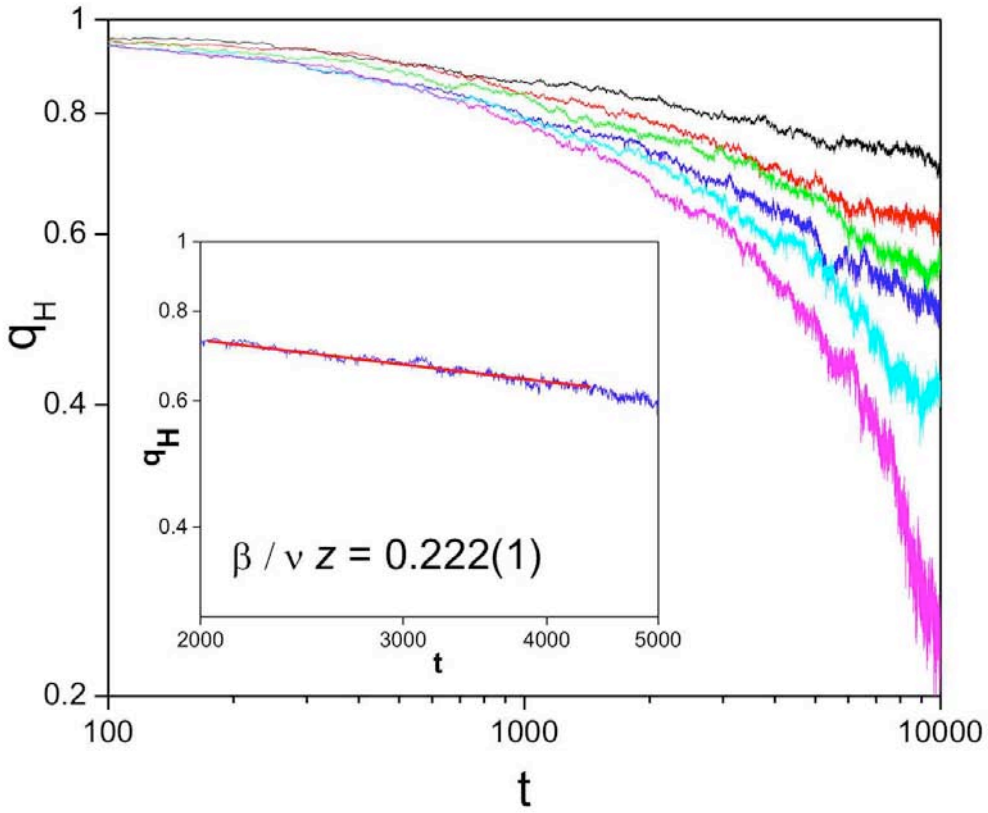
# Short-time dynamics of Poly-alanine

E. Arashiro, J.R. Drugowich de Felicio & U.H.E. Hansmann, PRE **73** (2006) 040902



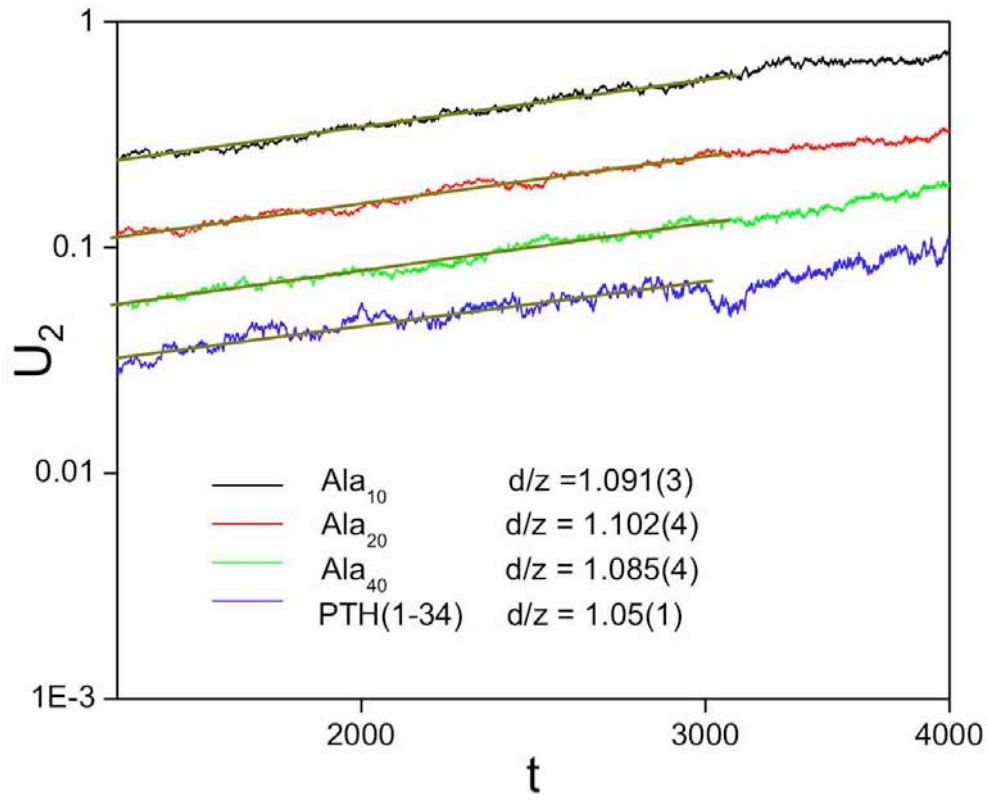
# Short-time dynamics of PTH(1-34)

E. Arashiro, J.R. Drugowich de Felicio & U.H.E. Hansmann, PRE **73** (2006) 040902



# Short-time dynamics

E. Arashiro, J.R. Drugowich de Felicio & U.H.E. Hansmann, PRE **73** (2006) 040902



# Critical Exponents for Helix-Coil Transition

E. Arashiro, J.R.D. de Felicio & U.H.E. Hansmann, J. Chem. Phys. 126 (2007) 045107

- Poly-Alanine (L=10 $\Rightarrow$ 40):
  - d/z = 1.1(1), $\beta$ = 0.38(1), d$\nu$ = 1.41(1)

- PTH(1-34):
  - d/z = 1.1(3), $\beta$ = 0.38(1), d$\nu$ = 1.42(1)