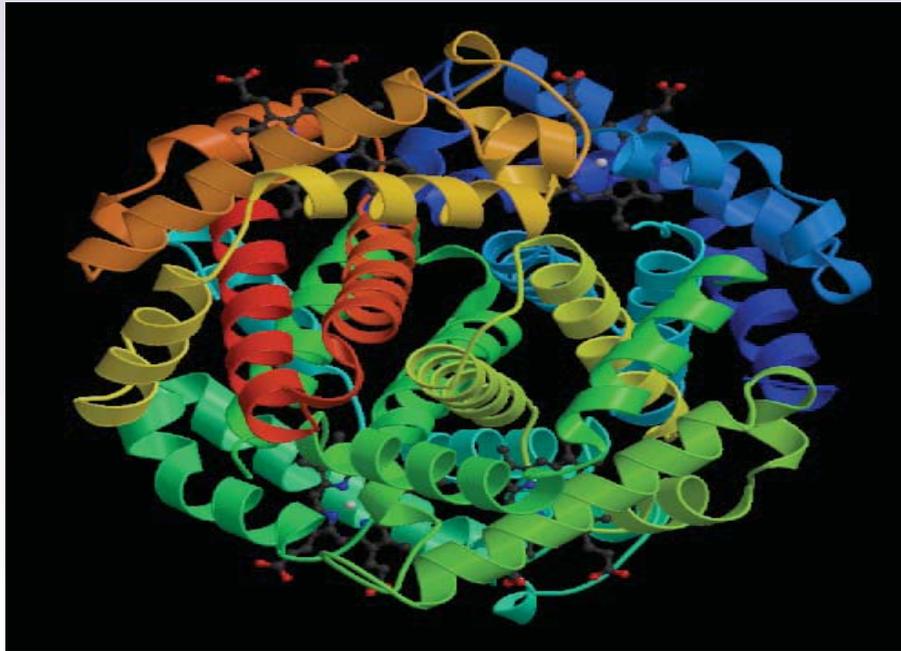


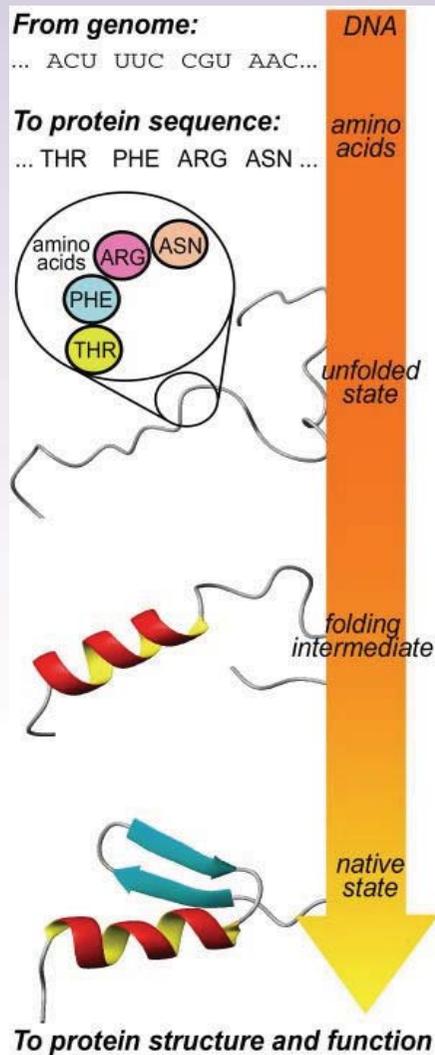
Discrete Nonlinear Schrodinger Equation and Organizing Principle for Protein Folding



Collaborators: M. Chernodub (Tours), U. Danielsson (Uppsala), S. Hu (Tours), Y. Kou (Mount Sinai), A. Khorotkin (Uppsala), M. Lundgren (Uppsala), N. Molkenthin (Tours/Uppsala), S. Nicolis (Tours), X. Peng (Uppsala), F. Sha (Uppsala) ...

Protein Folding

Function determined by shape

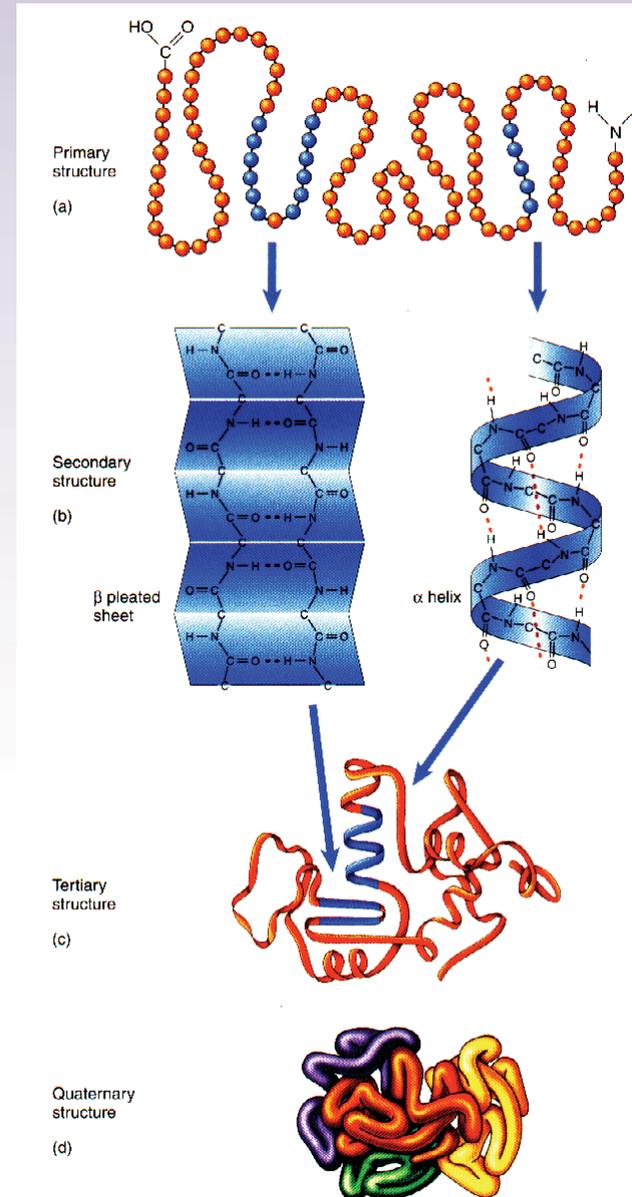


Translation of primary information into secondary, tertiary and quaternary structural information

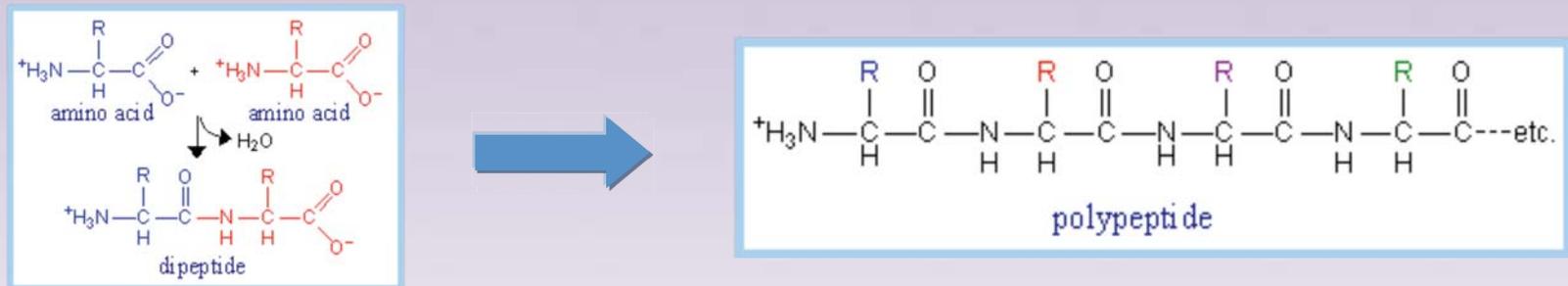
Folding is hydrophobic, hydrophilic or electrically charged

Correct three dimensional folding is essential for function. Misfolding causes diseases

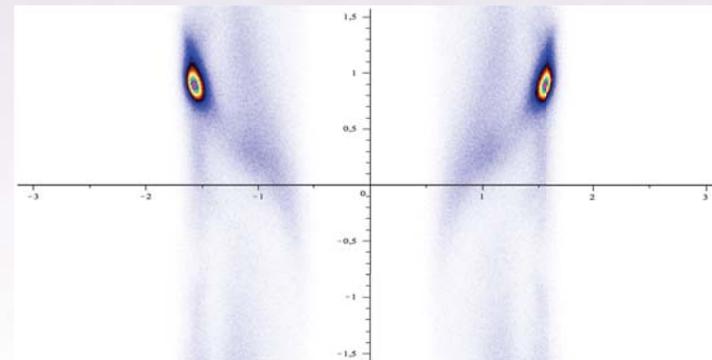
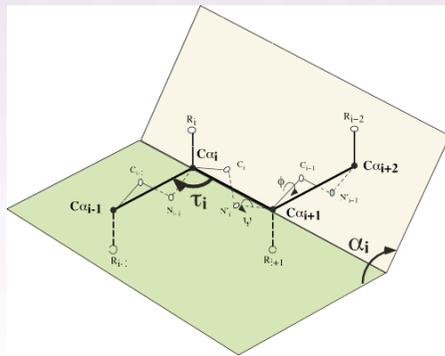
- *Alzheimer*
- *Parkinson*
- *Cancer*
- *Diabetes*



Transcriptome:



Backbone has a rigid structure - angles are (almost) fixed

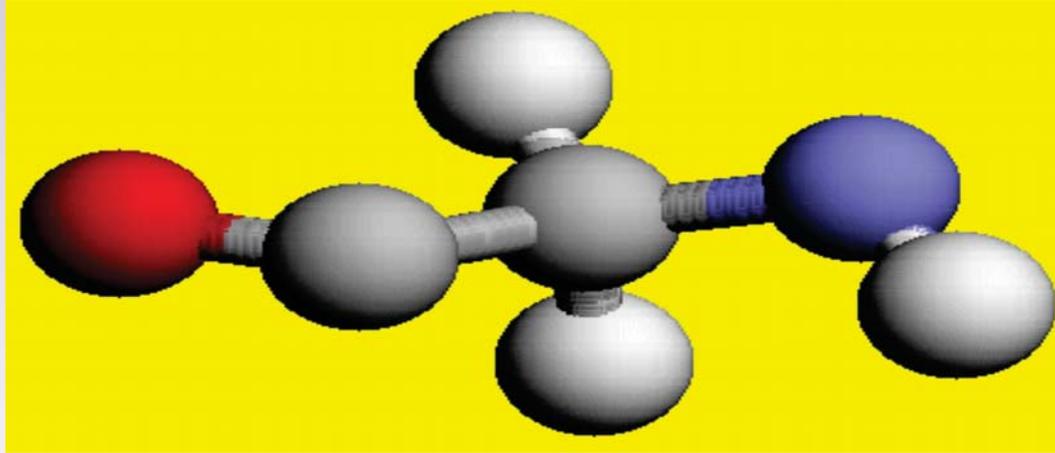


PDB data

But:

- X-ray crystallography, electron microscopy, NMR, ... (close packing, preparation, low resolution ...)
- Proteins subject to conformational fluctuations
- Not conclusive, not the whole story, ...

Classical Molecular Dynamics:



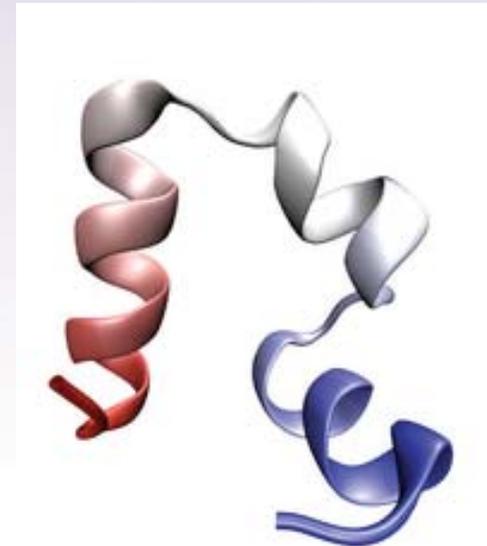
- Each atom moves by Newton's 2nd law $\mathbb{F} = \nabla E$

$$E = \sum_{(i,j)} k_{ij} r_{ij}^2 + \sum_{(i,j)} \frac{q_i q_j}{\epsilon r_{ij}} + \sum_{(i,j)} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{(i,j)} \left(\frac{C_{ij}}{r_{ij}^4} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \dots$$

State-of-Art:

- empirical force fields: AMBER, CHARMM, ...
- massive parallelization: Folding@Home, ...
- special purpose computers

- Trp-cage
- Villin
- WW-domain



RMSD distance $\approx 1.63 \text{ \AA}$

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_i^N (\mathbf{r}_i - \mathbf{r}_i^0)^2}$$

Shortcomings:

$$E = \sum_{(i,j)} k_{ij} r_{ij} + \sum_{(i,j)} \frac{q_i q_j}{\epsilon r_{ij}} + \sum_{(i,j)} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{(i,j)} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \dots$$

- Background fixed - no chemical reactions
- Linearization - harmonic approximation

Levinthal's Paradox:

- Consider a small protein chain of ~ 50 amino acids (typically 100 - 3000)
- Assume each amino acid has 10 conformations (vast underestimation!!)
- Total number of conformations: 10^{50}
- Assume one Monte Carlo step per femtosecond

EXHAUSTIVE SEARCH FOR THE GROUND STATE WILL

TAKE 10^{27} YEARS – BUT PROTEINS FOLD IN MICROSECONDS

NP complete

Example:

$$m\ddot{x} = -\frac{dV}{dx}$$

$$\begin{aligned} V(x) &= \frac{1}{2}k(x) \cdot (x - a)^2 \approx \frac{1}{4}\kappa(x + b)^2 \cdot (x - a)^2 \\ &\approx \frac{1}{4}\kappa b^2(x - a)^2 \cdot \left(1 + \mathcal{O}\left(\frac{x}{b}\right)\right) \end{aligned}$$

But there is an exact solution that can not be found by adding up any amount of small local fluctuations

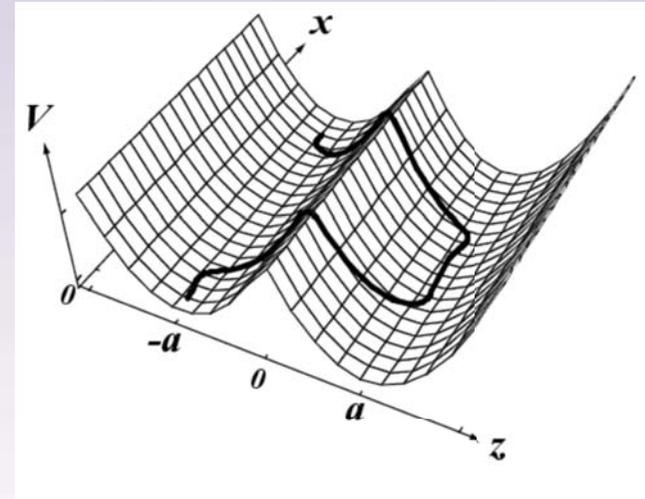
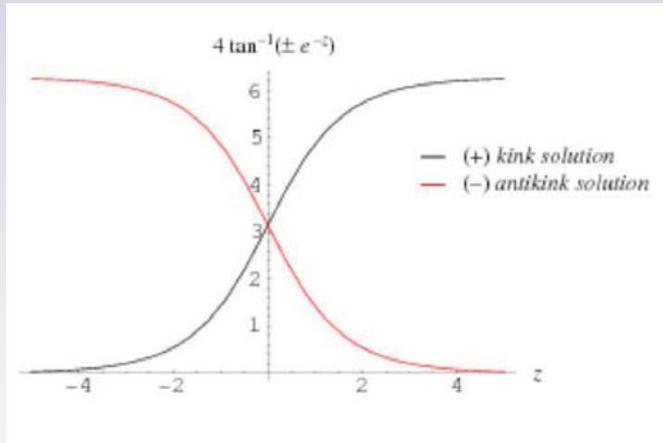
SOLITON:

$$x(t) = -\frac{b \cdot e^{c\sqrt{\frac{\kappa}{2m}}(t-t_0)} - a \cdot e^{-c\sqrt{\frac{\kappa}{2m}}(t-t_0)}}{e^{c\sqrt{\frac{\kappa}{2m}}(t-t_0)} + e^{-c\sqrt{\frac{\kappa}{2m}}(t-t_0)}}$$

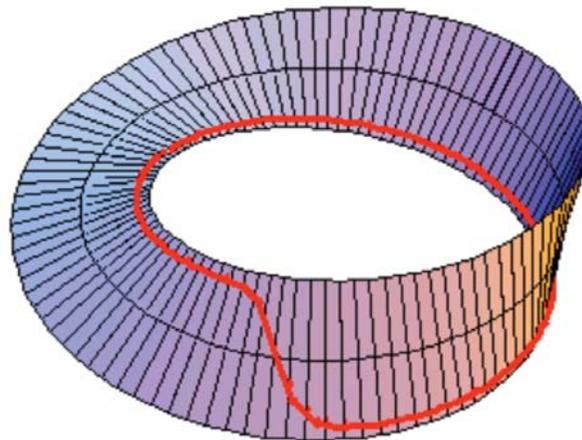
$$\left(= -b \cdot \tanh c\sqrt{\frac{\kappa}{2m}}(t-t_0) \right) \quad \text{for } a = b$$

$$c = -\frac{1}{2}(b + a)$$

Topological soliton interpolates between two different ground states of potential energy



Can not be removed by local perturbations



**AND NOW FOR SOMETHING
COMPLETELY
DIFFERENT.**

Extrinsic Geometry of Strings

- String embedded in \mathbb{R}^3 , coordinate $\mathbf{X}(s)$ with $s \in [0, l]$

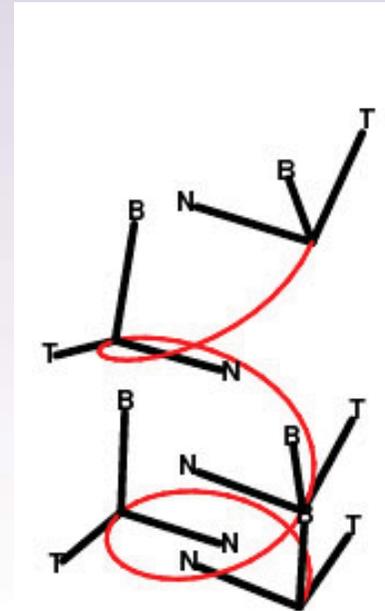
$$\mathbf{t} = \frac{d\mathbf{X}}{ds}$$

Frenet equations:

$$\frac{d}{ds} \begin{pmatrix} \mathbf{t} \\ \mathbf{n} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix} \begin{pmatrix} \mathbf{t} \\ \mathbf{n} \\ \mathbf{b} \end{pmatrix}$$

κ : curvature

τ : torsion (non-planarity)



- Orthogonal plane to curve: (\mathbf{n}, \mathbf{b})

Solving Frenet equations gives unique curve $\mathbf{X}(s)$ up to Galilean motions

U(1) gauge structure

Frame rotation: $\mathbf{e}^\pm = \mathbf{n} + i\mathbf{b} \rightarrow e^{\pm i\theta} \mathbf{e}^\pm$

$$\Rightarrow \kappa_\pm \rightarrow e^{\pm i\theta} \kappa_\pm \quad \& \quad \tau \rightarrow \tau - \partial_s \theta$$

- Abelian Higgs multiplet $(\kappa_\pm, \tau) \sim (\phi, A_1)$
- Landau-Ginsburg internal energy

$$F = \int_0^L ds [|(\partial_s + iA_1)\phi|^2 + \lambda(|\phi|^2 - a^2)^2] + c \int_0^L ds A_1$$

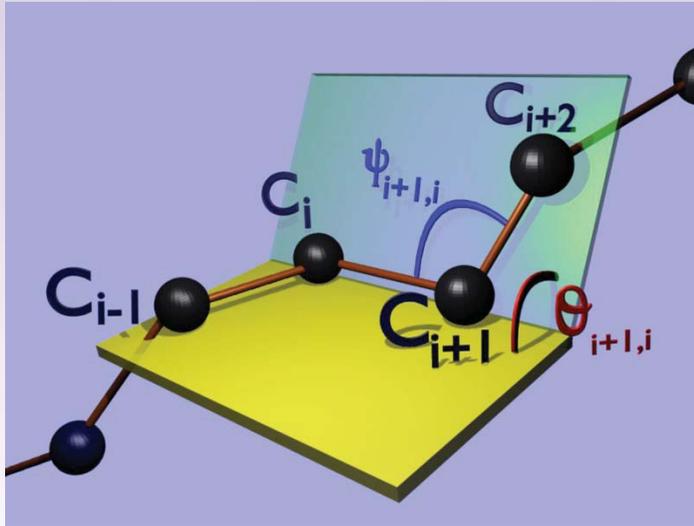
Chern-Simons

$$\left\{ \begin{array}{l} \phi \rightarrow \rho \cdot e^{i\theta} \\ A_i \rightarrow J_i = \frac{i}{2|\phi|^2} \{ \phi^* (\partial_i + iA_i)\phi + c.c. \} \end{array} \right.$$

$$\Rightarrow F = \int_0^L ds [(\partial_s \rho)^2 + (\rho^2 + m^2) J^2 + \lambda(\rho^2 - a^2)^2] + c \int_0^L ds J$$

Proca mass

Discrete Frenet Equation:



$$\mathbf{t}_i = \frac{\mathbf{R}_{i+1} - \mathbf{R}_i}{|\mathbf{R}_{i+1} - \mathbf{R}_i|}$$

$$\mathbf{b}_i = \frac{\mathbf{t}_{i-1} \times \mathbf{t}_i}{|\mathbf{t}_{i-1} \times \mathbf{t}_i|}$$

$$\mathbf{n}_i = \mathbf{b}_i \times \mathbf{t}_i$$

$$\begin{pmatrix} \mathbf{n}_{i+1} \\ \mathbf{b}_{i+1} \\ \mathbf{t}_{i+1} \end{pmatrix} = \exp\{-\tau_i \cdot T^2\} \cdot \exp\{-\alpha_i \cdot T^3\} \begin{pmatrix} \mathbf{n}_i \\ \mathbf{b}_i \\ \mathbf{t}_i \end{pmatrix}$$

Landau free energy for folding proteins:

$$E = - \sum_{i=1}^{N-1} 2 \tau_{i+1} \tau_i + \sum_{i=1}^N \{ 2\tau_i^2 + c \cdot (\tau_i^2 - m^2)^2 \} \\ + \sum_{i=1}^N \{ b \tau_i^2 \alpha_i^2 + d \alpha_i + e \alpha_i^2 + q \tau_i^2 \alpha_i \}$$

Equation for torsion angle:

$$\frac{\partial E}{\partial \alpha_i} = 2b\tau_i^2 \alpha_i + 2e\alpha_i + d + q\tau_i^2 = 0$$

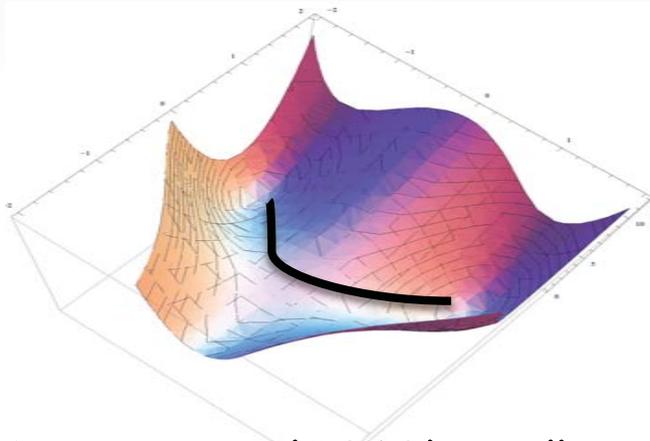
$$\Rightarrow \alpha_i[\tau_i] = - \frac{1}{2} \frac{d + q\tau_i^2}{e + b\tau_i^2}$$

Generalized Discrete Nonlinear Schroedinger Equation:

$$\tau_{i+1} - 2\tau_i + \tau_{i-1} = U'[\tau_i]\tau_i \equiv \frac{dU[\tau]}{d\tau_i^2} \tau_i$$

$$U[\tau] = - \left(\frac{bd - eq}{2b} \right)^2 \cdot \frac{1}{e + b\tau^2} - \left(\frac{q^2 + 8bcm^2}{4b} \right) \cdot \tau^2 + c \cdot \tau^4$$

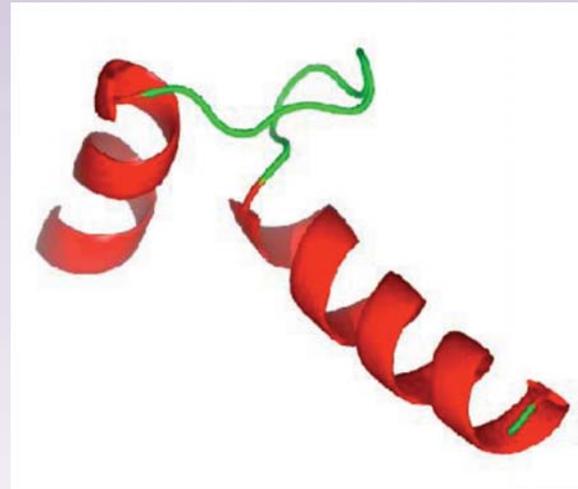
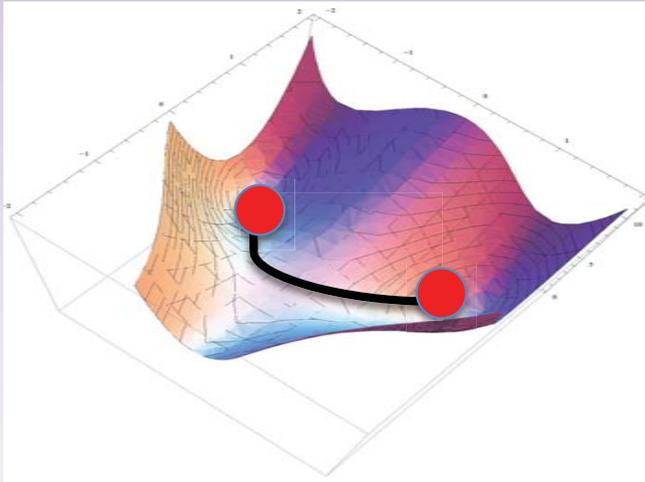
Hamiltonian:
$$H = -2 \sum_{i=1}^{N-1} \tau_{i+1}\tau_i + \sum_{i=1}^N \{2\tau_i^2 + U[\tau_i]\}$$



Topological soliton
that connects the
two ground states?

Herrmann (2010): A (heteroclinic) soliton solution exists

Topological soliton = helix-loop-helix motif



SOLVE THE EQUATION:

$$\tau_i^{(n+1)} = \tau_i^{(n)} - \epsilon \left\{ \tau_i^{(n)} U'[\tau_i^{(n)}] - (\tau_{i+1}^{(n)} - 2\tau_i^{(n)} + \tau_{i-1}^{(n)}) \right\}$$

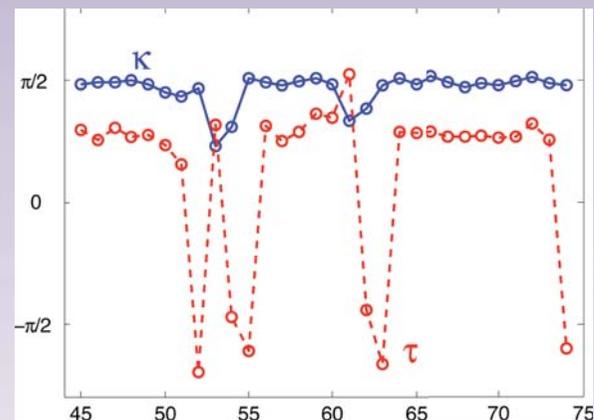
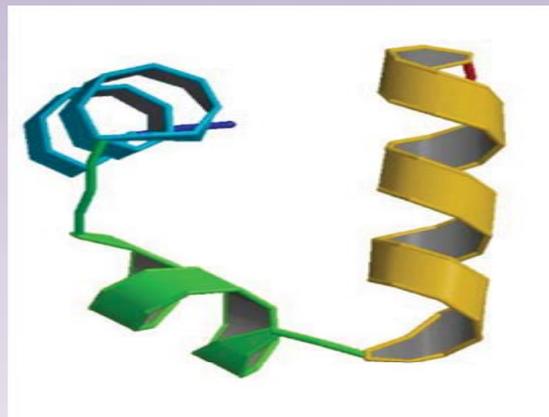
Fixed point is a solution

$$\alpha_i[\tau] = \frac{b}{2} \frac{1 + c \cdot \tau_i^2}{1 + d \cdot \tau_i^2}$$

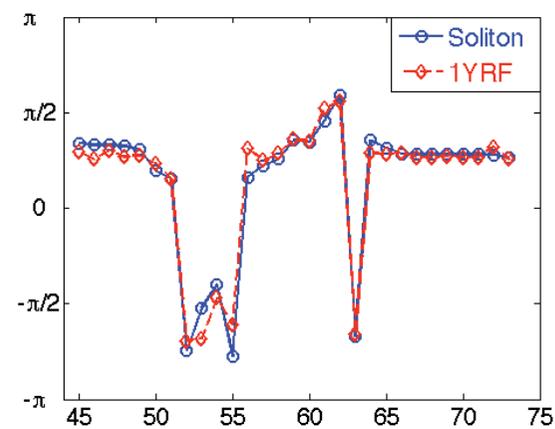
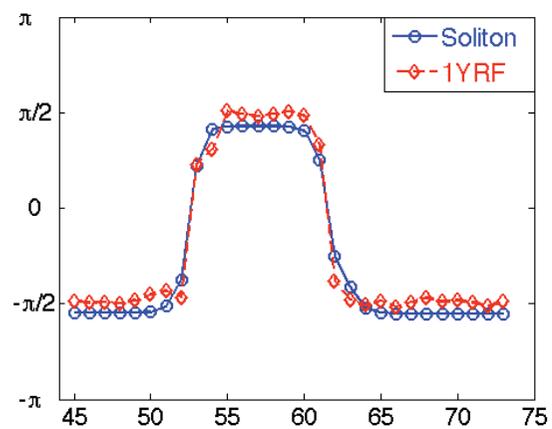
ONE function SIX parameters ALL with global interpretation

ENORMOUS reductions from Levinthal's estimate

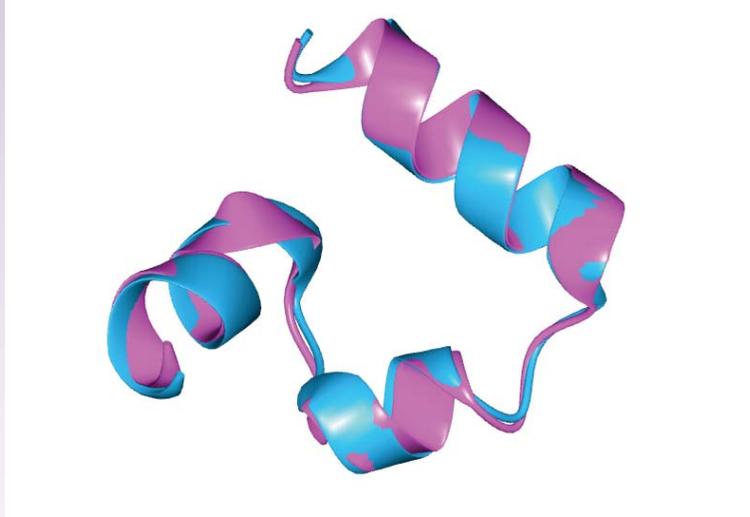
Example:
1YRF



\mathbb{Z}_2 gauge transformation



Soliton solution with $\text{RMSD} = 0.73 \text{ \AA}$



Universality: There are over 7.000 proteins with same solitons with $\text{RMSD} < 1.0 \text{ Angstrom}$

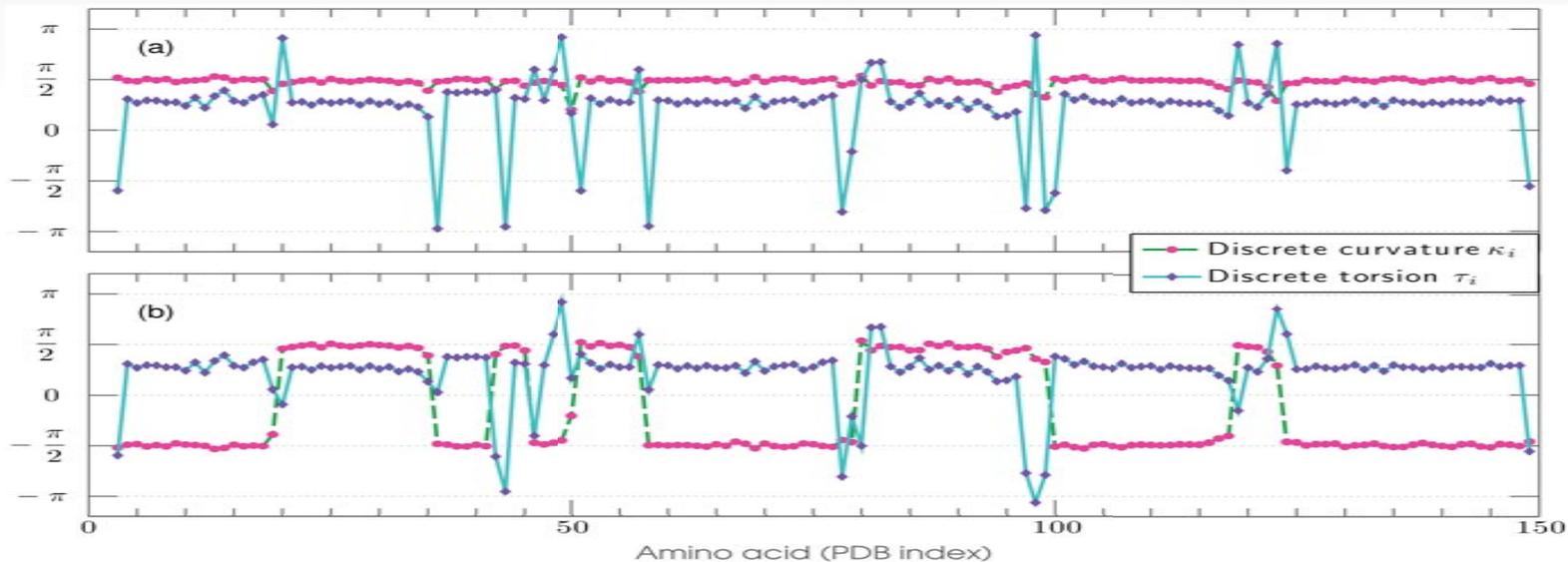
Exact profiles:

$$x(t) = -\frac{b \cdot e^{c\sqrt{\frac{\kappa}{2m}}(t-t_0)} - a \cdot e^{-c\sqrt{\frac{\kappa}{2m}}(t-t_0)}}{e^{c\sqrt{\frac{\kappa}{2m}}(t-t_0)} + e^{-c\sqrt{\frac{\kappa}{2m}}(t-t_0)}}$$

$$\tau_i = (-1)^{r+1} \frac{m_{r1} \cdot e^{c_r(i-s_r)} - m_{r2} \cdot e^{-c_r(i-s_r)}}{2 \cosh[c_r(i-s_r)]}$$

$$\alpha_i = -\frac{b_r}{2} \frac{1 + d_r \tau_i^2}{1 + e_r \tau_i^2}$$

Myoglobin:

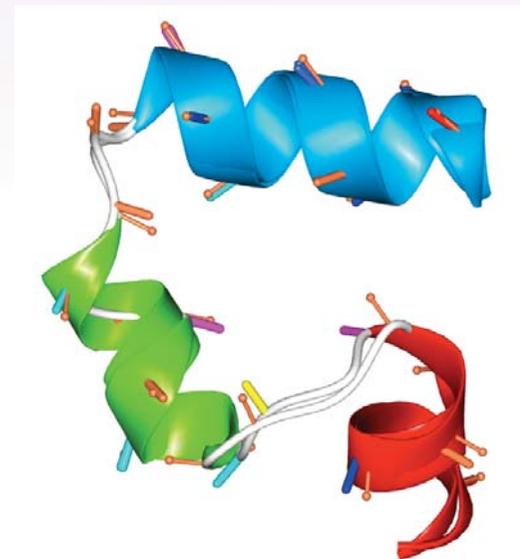
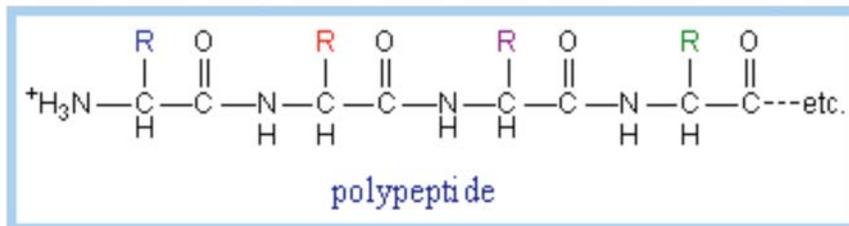


RMSD = 1.2 Å

Universality?



Side-Chains:



Final Remarks:

- *Model describes folded proteins in terms of topological solitons*
- *Folded proteins can even be described in terms of a universal elementary function*
- *A given soliton structure appears in many different proteins, from eukaryotes to prokaryotes - universality!*

Future challenges include

- ✓ *classification of solitons*
- ✓ *inclusion of sidechains*
- ✓ *protein dynamics*
- ✓ *identifying solitons from genome*
- ✓ *...*