

Simple Solvable Models of Primary and Secondary Structure Formation in Hetero-Polymers

ACC Coolen

*N Skantzos, J van Mourik,
D Sherrington, H Chakravorty*

Proteins – an Introduction

Secondary Structure Formation

*1 + ∞ dimensional models
mean-field & random field techniques*

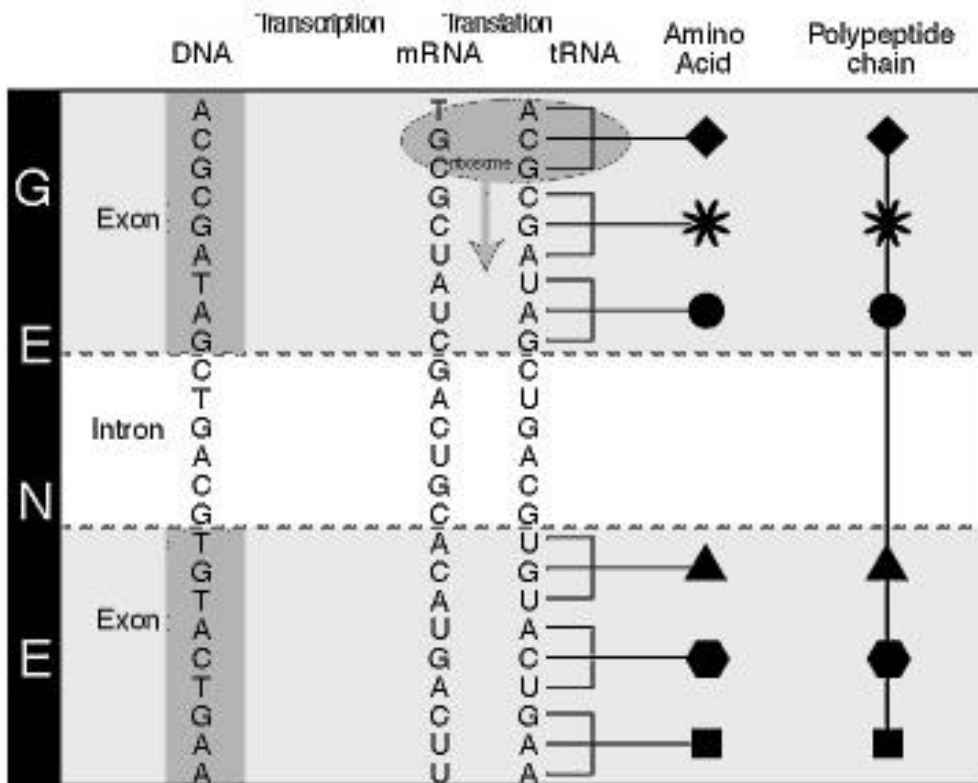
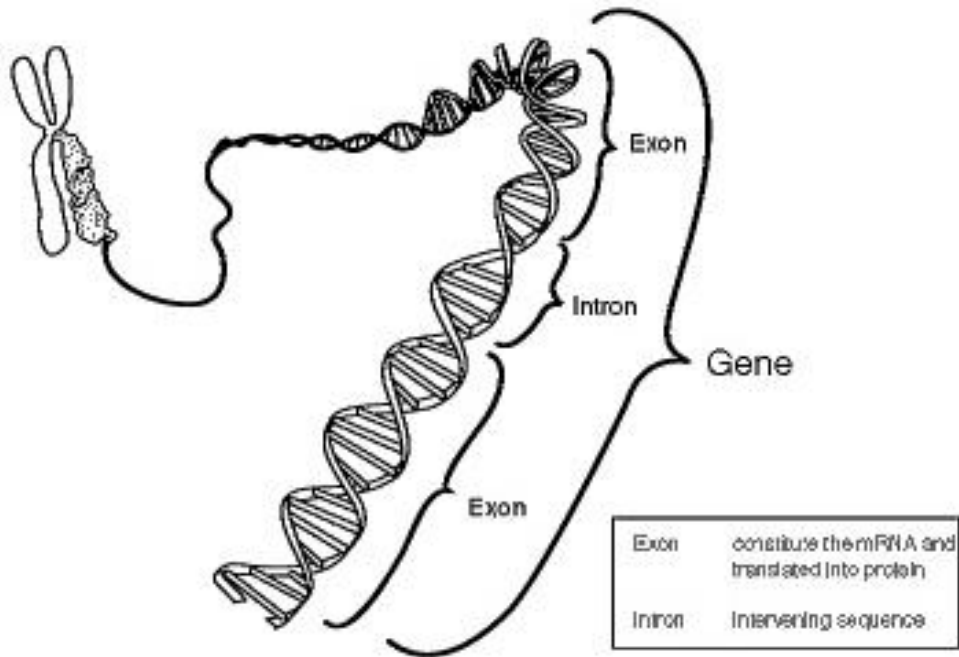
Primary Structure Formation

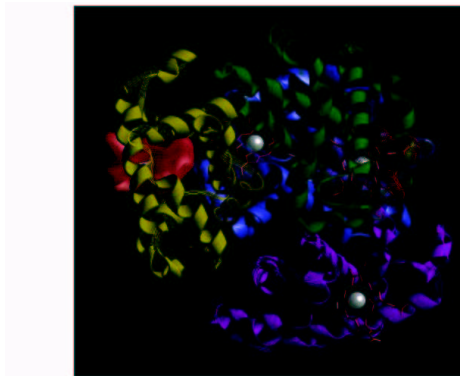
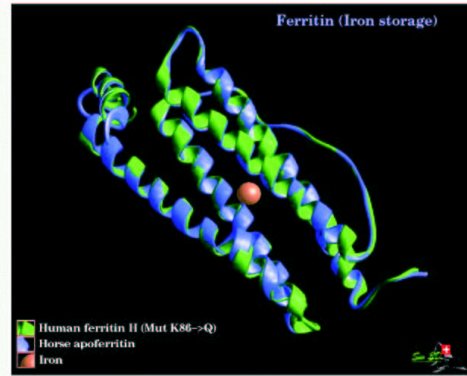
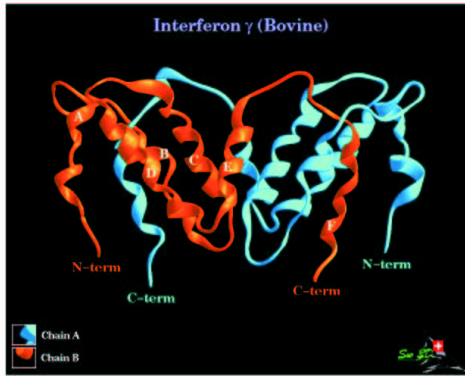
*coupled dynamics of folding and genetic selection
mean-field model, finite- n replica analysis*

New Directions

*full primary + secondary generation
tertiary structure*

I: PROTEINS - AN INTRODUCTION





Primary structure:

amino-acid sequence (sequence of monomers)

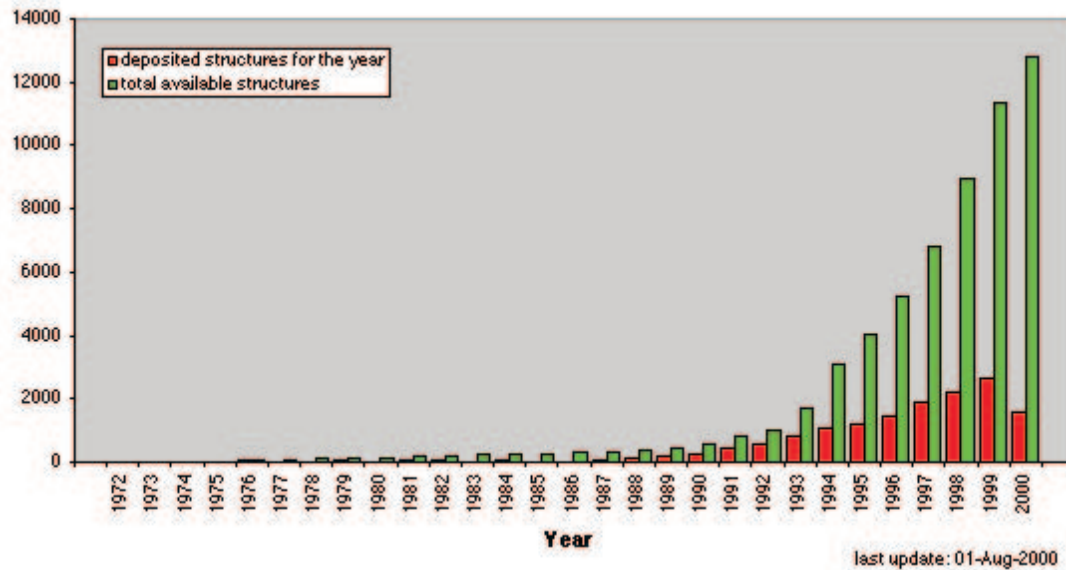
Secondary structure:

local conformation of α -helices, β -sheets, etc

Tertiary structure:

3D arrangement of secondary structure elements

‘Knowledge of a protein’s tertiary structure is a prerequisite for the proper understanding and engineering of its function.’



Primary structures: $\pm 182,000$ (growing at $\pm 12,000/\text{year}$)
 Tertiary structures: $\pm 13,000$ (growing at $\pm 3,000/\text{year}$)

‘This makes the [protein] folding problem, the successful prediction of a protein’s tertiary structure from its amino-acid sequence, central to rapid progress in post-genomic biology’

The protein folding problem

Understand how and why proteins find their unique functional three-dimensional structures, and predict these structures from the amino-acid sequences

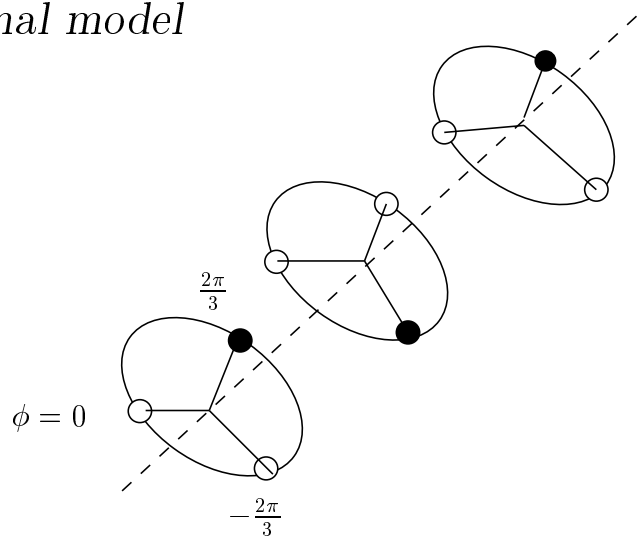
The inverse protein folding problem

Design artificial proteins, which fold into desired structures and hence perform desired biological functions

II: SECONDARY STRUCTURE FORMATION IN RANDOM HETERO-POLYMERS

Random hetero-polymers

a simple $1 + \infty$ dimensional model



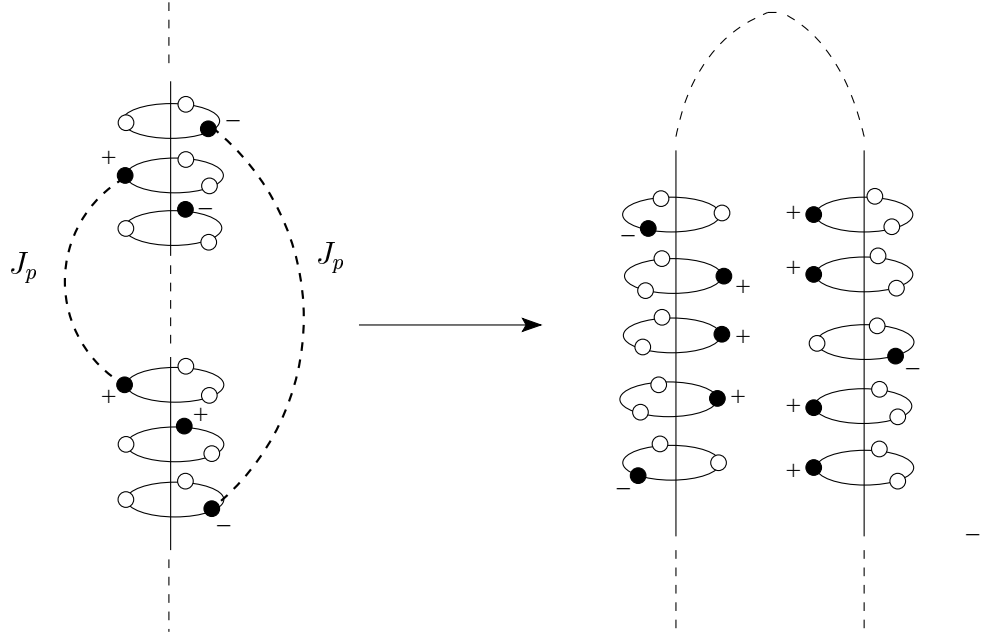
$$\phi_i \in \left\{ \frac{(2k + 1)\pi}{q}; \quad k = 0, \dots, q - 1 \right\}$$

$$H(\phi) = H_s(\phi) + H_p(\phi) + H_{\text{Hb}}(\phi)$$

Polarity energy (hydrophobic vs hydrophilic)

Steric energy (mechanical constraints)

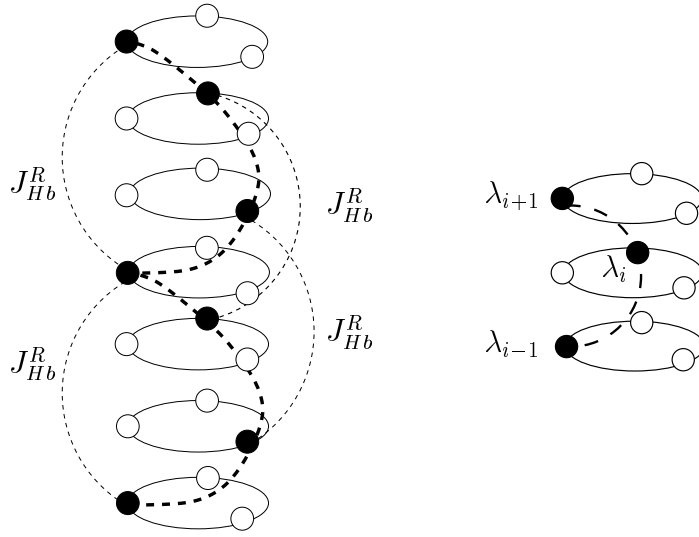
Hydrogen bonding energy



$$H_p(\phi) = -\frac{J_p}{N} \sum_{ij} \xi_i \xi_j \delta_{\phi_i, \phi_j}$$

hydrophobic : $\xi_i = 1$

hydrophilic : $\xi_i = -1$



$$H_{Hb}(\phi) = -\sum_i \left\{ J_{Hb}^L \prod_{k=0}^{q-1} \delta_{\phi_{i+k+1} - \phi_{i+k}, \frac{2\pi}{q}} + J_{Hb}^R \prod_{k=0}^{q-1} \delta_{\phi_{i+k+1} - \phi_{i+k}, \frac{-2\pi}{q}} \right\}$$

$$H_s(\phi) = -J_s \sum_i \cos[(\phi_{i+1} - \phi_i) - (\phi_i - \phi_{i-1}) - a_i]$$

Solution of the Model for $q = 2$

$$\phi_i = \frac{1}{2}\pi\sigma_i, \quad \sigma_i = \pm 1$$

$$H_p(\boldsymbol{\sigma}) = -\frac{J_p}{2N} \sum_{ij} \sigma_i \xi_i \xi_j \sigma_j$$

$$H_{Hb}(\boldsymbol{\sigma}) = -\frac{1}{2} J_{Hb} \sum_i [1 - \sigma_i \sigma_{i+1}] [1 - \sigma_i \sigma_{i-1}] \quad J_{Hb} = \frac{1}{2} (J_{Hb}^L + J_{Hb}^R)$$

$$H_s(\boldsymbol{\sigma}) = -J_s \sum_i \eta_i \sigma_{i+1} \sigma_{i-1} \quad \eta_i = \cos[a_i]$$

Observables:

$$\chi(\boldsymbol{\sigma}) = \frac{1}{4N} \sum_i [1 - \sigma_i \sigma_{i+1}] [1 - \sigma_i \sigma_{i-1}] \quad m(\boldsymbol{\sigma}) = \frac{1}{N} \sum_i \xi_i \sigma_i$$

Solution:

$$f = -\lim_{N \rightarrow \infty} \frac{1}{\beta N} \log \int dm d\hat{m} e^{-\beta N G_N(m, \hat{m})}$$

$$G_N(m, \hat{m}) = -im\hat{m} - \frac{1}{2} J_p m^2 - \frac{1}{\beta N} \log Z_N(-i\beta\hat{m})$$

$$Z_N(x) = \sum_{\sigma_1 \dots \sigma_N} e^{\frac{1}{2}\beta J_{Hb} \sum_i [1 - \sigma_i \sigma_{i+1}] [1 - \sigma_i \sigma_{i-1}] + \beta J_s \sum_i \sigma_{i-1} \eta_i \sigma_{i+1} + x \sum_i \sigma_i \xi_i}$$

$(1 + \infty)$ -dimensional random field problem

Adaptation of RFIM techniques:

$$Z_{\sigma\sigma'}^{(N)}(x) = \sum_{\sigma_1 \dots \sigma_N} e^{\frac{1}{2}\beta J_{Hb} \sum_i [1-\sigma_i \sigma_{i+1}] [1-\sigma_i \sigma_{i-1}] + J_s \sum_i \sigma_{i-1} \eta_i \sigma_{i+1} + x \sum_i \sigma_i \xi_i} \delta_{\sigma_{N-1}, \sigma} \delta_{\sigma_N, \sigma'}$$

$$k_j^{(1)} \equiv e^{-2x\xi_j} \frac{Z_{++}^{(j)}}{Z_{+-}^{(j)}} \quad k_j^{(2)} \equiv e^{2x\xi_j} \frac{Z_{+-}^{(j)}}{Z_{-+}^{(j)}} \quad k_j^{(3)} \equiv e^{-2x\xi_j} \frac{Z_{-+}^{(j)}}{Z_{--}^{(j)}}$$

stochastic map:

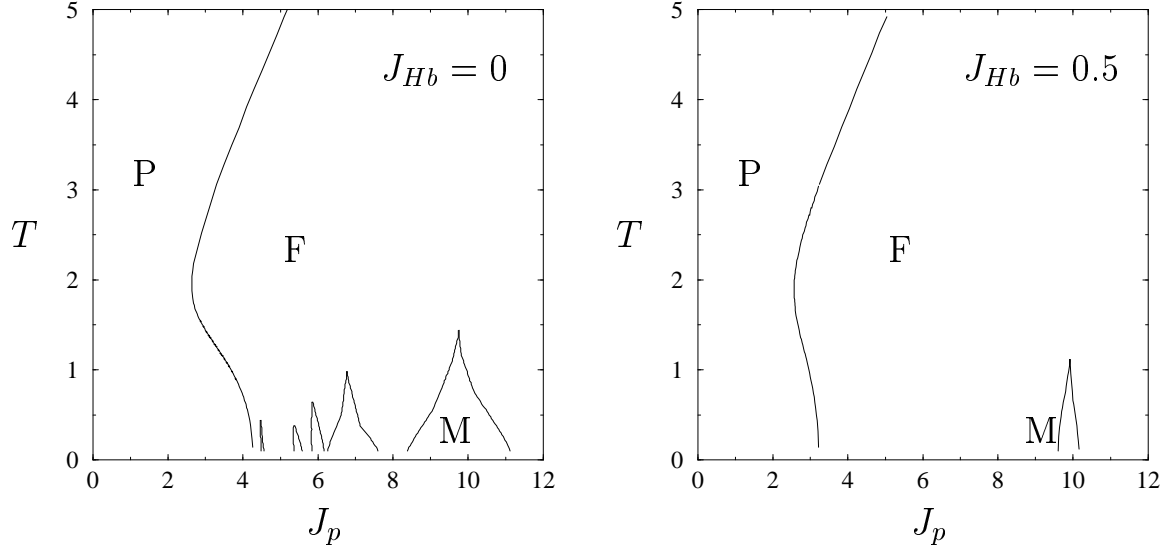
$$k_{j+1}^{(1)} = \frac{e^{\beta J_s \eta_j} k_j^{(1)} k_j^{(2)} + e^{-\beta J_s \eta_j - \beta J_{Hb}}}{e^{-\beta J_s \eta_j} k_j^{(1)} k_j^{(2)} + e^{\beta J_s \eta_j + \beta J_{Hb}}} e^{-\beta J_{Hb}}$$

$$k_{j+1}^{(2)} = \frac{e^{-\beta J_s \eta_j} k_j^{(1)} k_j^{(2)} + e^{\beta J_s \eta_j + \beta J_{Hb}}}{e^{\beta J_s \eta_j + \beta J_{Hb}} k_j^{(2)} k_j^{(3)} + e^{-\beta J_s \eta_j}} k_j^{(3)} e^{2x\xi_j}$$

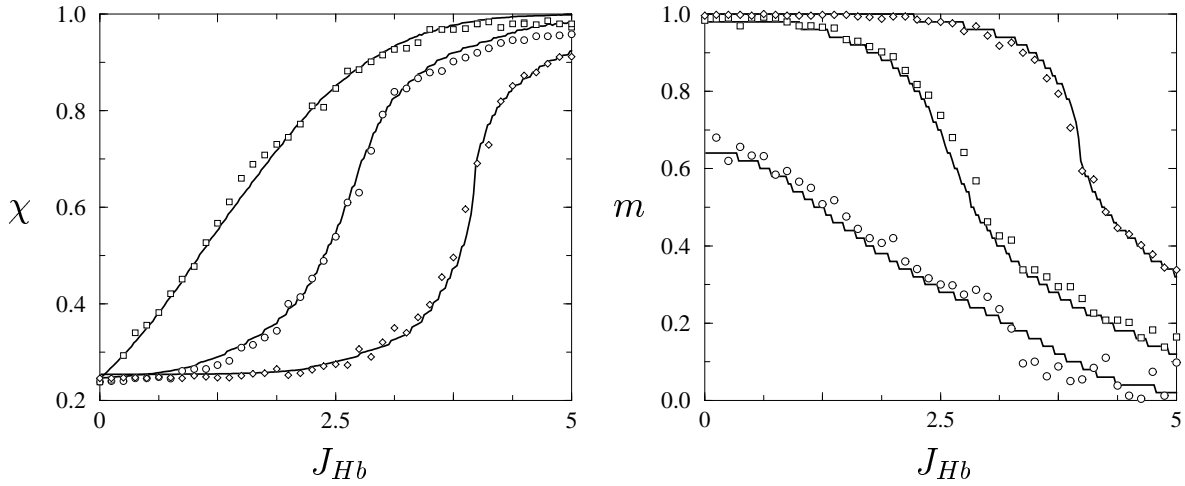
$$k_{j+1}^{(3)} = \frac{e^{\beta J_s \eta_j + \beta J_{Hb}} k_j^{(2)} k_j^{(3)} + e^{-\beta J_s \eta_j}}{e^{-\beta J_s \eta_j - \beta J_{Hb}} k_j^{(2)} k_j^{(3)} + e^{\beta J_s \eta_j}} e^{\beta J_{Hb}}$$

$$P_i(\mathbf{k}|x) = \langle \delta \left[\mathbf{k} - (k_i^{(1)}, k_i^{(2)}, k_i^{(3)}) \right] \rangle_{\boldsymbol{\xi}, \boldsymbol{\eta}}$$

$$f = \text{extr}_m \left\{ \frac{1}{2} J_p m^2 + J_p m p - \frac{1}{\beta} \int d\mathbf{k} P_\infty(\mathbf{k} | \beta J_p m) \int d\eta \tilde{w}[\eta] \log \left[e^{-\beta J_s \eta - \beta J_{Hb}} k^{(2)} k^{(3)} + e^{\beta J_s \eta} \right] \right\}$$



P: paramagnetic (i.e. ‘swollen’) state, $m = 0$
 F: Mattis (i.e. ‘compact’) state, two solutions $\pm m \neq 0$
 M: multiple local minima of f (frustration-induced)
 P \rightarrow F: second-order. F \rightarrow M: first-order (dynamical)

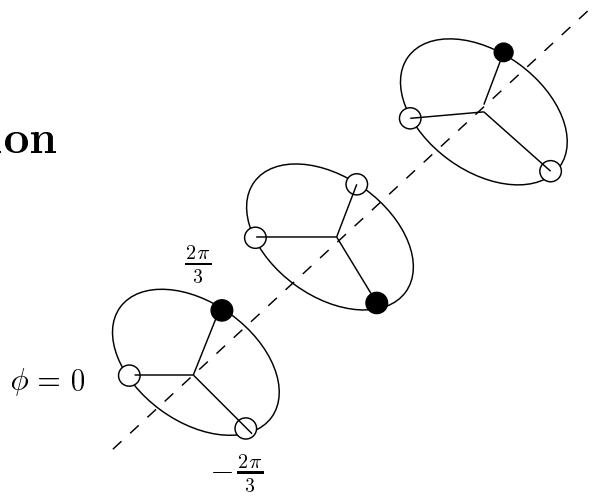


Lines: theory. Markers: simulations
 (measured after 120,000 it/monomer, $N = 1000$)
 $J_p \in \{4, 8, 12\}$ (left: top to bottom; right: bottom to top)
 $T = 2, J_s = 2$

III: PRIMARY STRUCTURE FORMATION IN HETERO-POLYMERS

Primary structure: monomer sequence

**Coupled dynamics of
(slow) sequence selection
and (fast) folding**



$$\phi_i \in \{(2k + 1)\pi/q; \quad k = 0, \dots, q - 1\}$$

$$H_p(\boldsymbol{\phi}) = -\frac{J_p}{N} \sum_{ij} \eta_i \eta_j \delta_{\phi_i, \phi_j} \quad \begin{array}{ll} \text{hydrophobic :} & \eta_i > 0 \\ \text{hydrophilic :} & \eta_i < 0 \end{array}$$

Folding: polarity energy only

Add: genetic dynamics of monomer species

Monomer species: polarity η_i

'fast' degrees of freedom: $\boldsymbol{\phi} = (\phi_1, \dots, \phi_N)$

'slow' degrees of freedom: $\boldsymbol{\eta} = (\eta_1, \dots, \eta_N)$

**Sequence
selection:**

$$\frac{d}{dt}\eta_i = -\frac{\partial}{\partial\eta_i} \{H(\boldsymbol{\phi}, \boldsymbol{\eta}) + V(\boldsymbol{\eta})\} + \xi_i$$

$$\langle \xi_i(t)\xi_j(t') \rangle = 2\tilde{T}\delta_{ij}\delta(t-t')$$

quality of folding : $H(\boldsymbol{\phi}, \boldsymbol{\eta}) = -\frac{J}{N} \sum_{ij} \eta_i \eta_j \delta_{\phi_i, \phi_j}$

functionality : $V(\boldsymbol{\eta})$

*Genetic dynamics
adiabatically slow:*

$$\frac{\partial H(\boldsymbol{\phi}, \boldsymbol{\eta})}{\partial\eta_i} \rightarrow \left\langle \frac{\partial H(\boldsymbol{\phi}, \boldsymbol{\eta})}{\partial\eta_i} \right\rangle_{\boldsymbol{\phi}} = \frac{\partial}{\partial\eta_i} \left\{ -\frac{1}{\beta} \log Z[\boldsymbol{\eta}] \right\}$$

‘genetic Hamiltonian’:

$$H(\boldsymbol{\eta}) = V(\boldsymbol{\eta}) - \frac{1}{\beta} \log Z[\boldsymbol{\eta}] \quad Z[\boldsymbol{\eta}] = \text{Tr}_{\boldsymbol{\phi}} e^{-\beta H(\boldsymbol{\phi}, \boldsymbol{\eta})}$$

free energy per monomer:

$$f_N = -\frac{1}{\tilde{\beta}N} \log \text{Tr}_{\boldsymbol{\eta}} \left[\sum_{\boldsymbol{\phi}} e^{-\beta H(\boldsymbol{\phi}, \boldsymbol{\eta})} \right]^n e^{-\tilde{\beta}V(\boldsymbol{\eta})} \quad n = \frac{\tilde{\beta}}{\beta}$$

Finite n replica problem !

- *functionality potential*: $V(\boldsymbol{\eta}) = \sum_{i=1}^N \mu_i \eta_i$
- *Ising version*: $\eta_i \in \{-1, 1\}$ for all i

$$\tilde{\beta} f = \text{extr}_{\{z_\phi^\alpha\}} \left\{ \frac{1}{4\beta J} \sum_{\phi\alpha} (z_\phi^\alpha)^2 - \langle \log \left[e^{-\tilde{\beta}\mu} \prod_{\alpha} (\sum_{\phi} e^{z_\phi^\alpha}) + e^{\tilde{\beta}\mu} \prod_{\alpha} (\sum_{\phi} e^{-z_\phi^\alpha}) \right] \rangle_{\mu} \right\}$$

Replica Symmetry:
(stable)

$$f_{\text{RS}} = \min_{\{L_\phi\}} \left\{ \frac{J}{4} \sum_{\phi} L_\phi^2 - \frac{1}{\beta n} \langle \log \left[(\sum_{\phi} e^{\beta[JL_\phi - \mu]})^n + (\sum_{\phi} e^{\beta[\mu - JL_\phi]})^n \right] \rangle_{\mu} \right\}$$

Saddle-point eqns:

$$L_\psi = (1+p) \frac{e^{\beta JL_\psi}}{\sum_{\phi} e^{\beta JL_\phi}} - (1-p) \frac{e^{-\beta JL_\psi}}{\sum_{\phi} e^{-\beta JL_\phi}}$$

$$p = \left\langle \frac{(\sum_{\phi} e^{\beta[JL_\phi - \mu]})^n - (\sum_{\phi} e^{\beta[\mu - JL_\phi]})^n}{(\sum_{\phi} e^{\beta[JL_\phi - \mu]})^n + (\sum_{\phi} e^{\beta[\mu - JL_\phi]})^n} \right\rangle_{\mu}$$

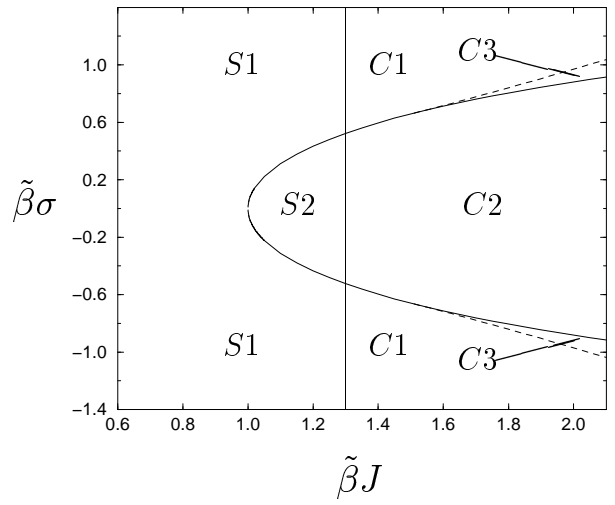
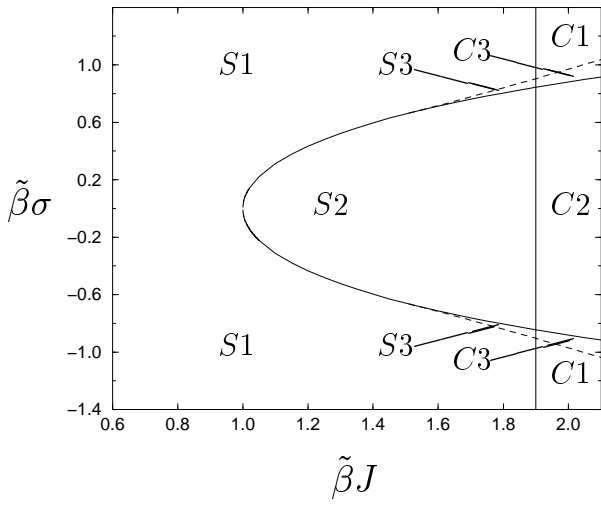
Physical meaning:

$$L_\phi = (1+p)P_+(\phi) + (1-p)P_-(\phi)$$

$$P_{\pm}(\phi) = \lim_{N \rightarrow \infty} \frac{\sum_i \langle \delta_{\phi, \phi_i} \delta_{\pm 1, \eta_i} \rangle}{\frac{1}{2}(1 \pm p)N} \quad p = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \langle \eta_i \rangle$$

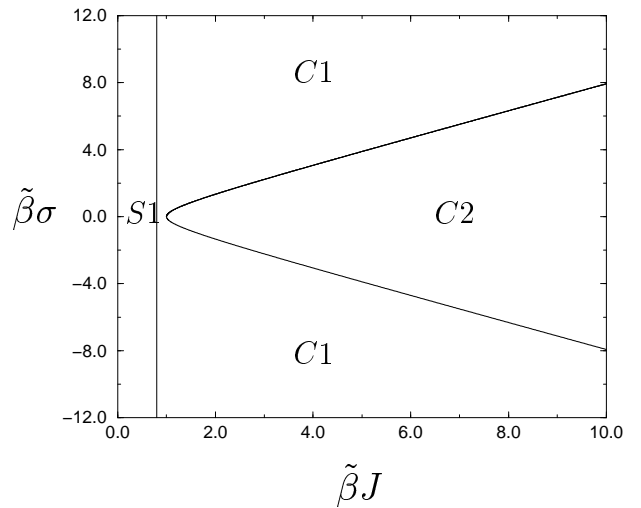
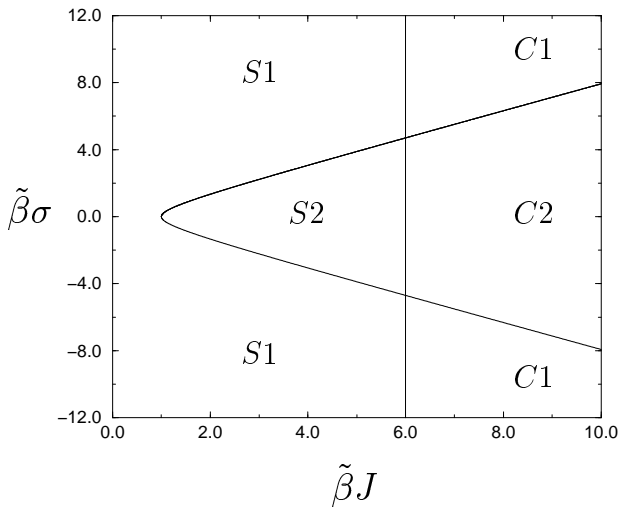
(Dynamical) Phase diagrams for $q = 2$

S_ℓ/C_ℓ : swollen/compact phase, ℓ locally stable values for p
dashed/solid: discontinuous/continuous transitions



$$P(\mu) = \frac{1}{2}\delta[\mu - \sigma] + \frac{1}{2}\delta[\mu + \sigma]$$

Left: $n = 5$. Right: $n = \frac{5}{4}$.



$$P(\mu) = [2\pi\sigma^2]^{-1/2} e^{-\frac{1}{2}\mu^2/\sigma^2}$$

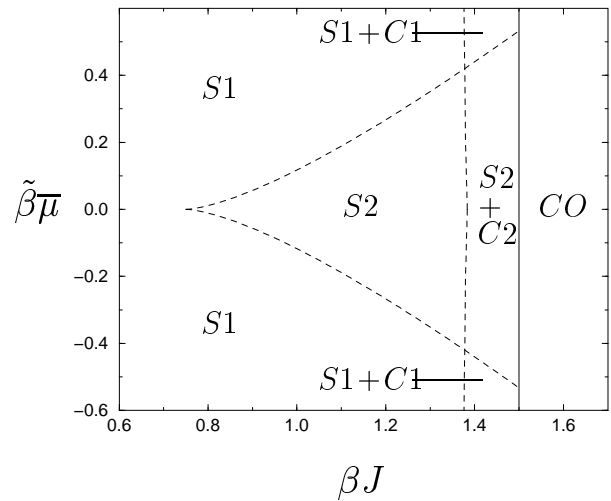
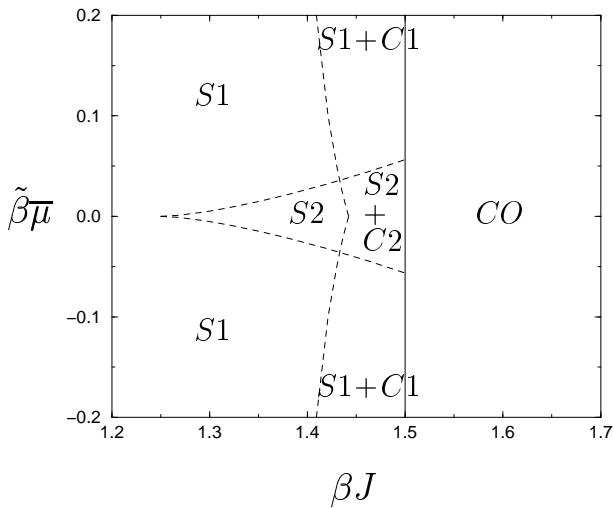
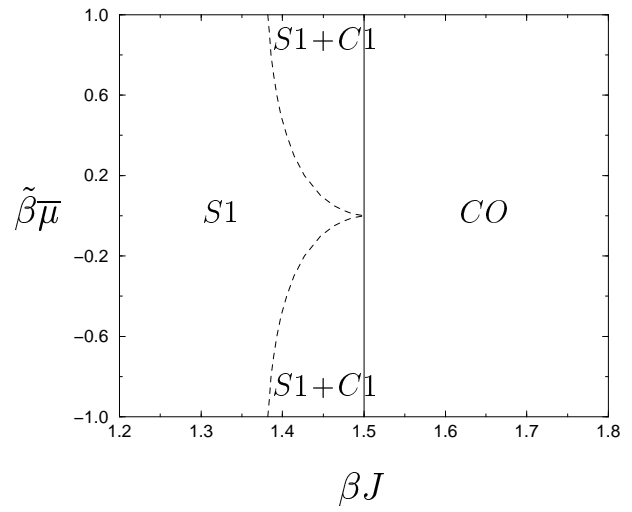
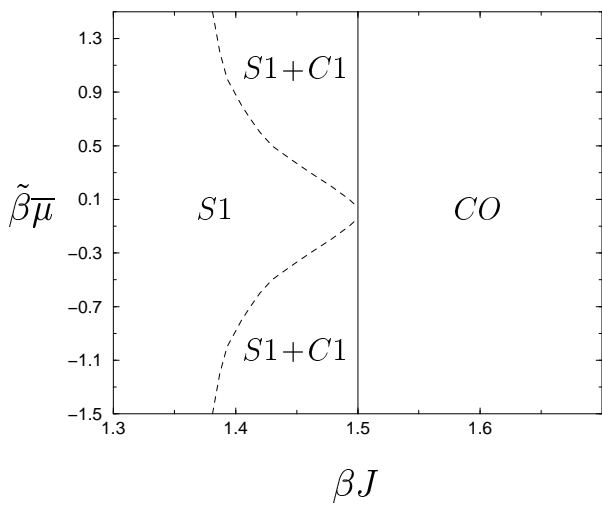
Left: $n = 6$. Right: $n = \frac{3}{4}$.

(Dynamical) Phase diagrams for $q = 3$

S_ℓ/C_ℓ : swollen/compact phase, ℓ locally stable values for p

CO : compact states only

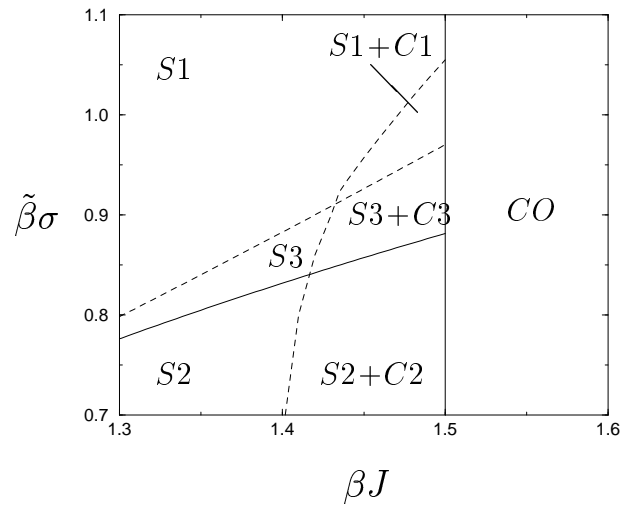
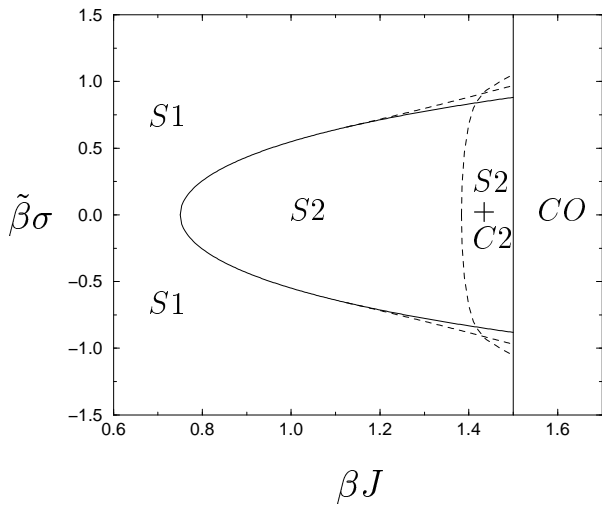
dashed/solid: discontinuous/continuous transitions



$$P(\mu) = \delta[\mu - \bar{\mu}]$$

Top: $n = \frac{1}{2}$ (L), $n = 1$ (R). Bottom: $n = \frac{6}{5}$ (L), $n = 2$ (R).

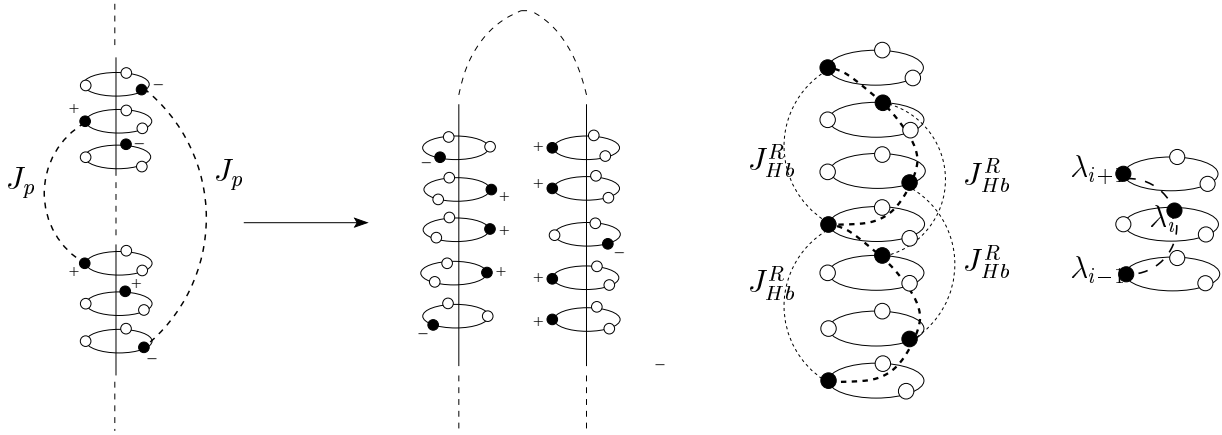
S_ℓ/C_ℓ : swollen/compact phase, ℓ locally stable values for p
 CO : compact states only
 dashed/solid: discontinuous/continuous transitions



$$P(\mu) = \frac{1}{2}\delta[\mu - \sigma] + \frac{1}{2}\delta[\mu + \sigma], \quad n = 2$$

IV: NEW DIRECTIONS

Full primary & secondary generation



$$H_p = -\frac{J_p}{N} \sum_{ij} \xi_i \xi_j \delta_{\phi_i, \phi_j}$$

$$H_{Hb} = -J_{Hb}^L \sum_i \prod_{k=0}^{q-1} \delta_{\phi_{i+k+1} - \phi_{i+k}, \frac{2\pi}{q}} - J_{Hb}^R \sum_i \prod_{k=0}^{q-1} \delta_{\phi_{i+k+1} - \phi_{i+k}, \frac{-2\pi}{q}}$$

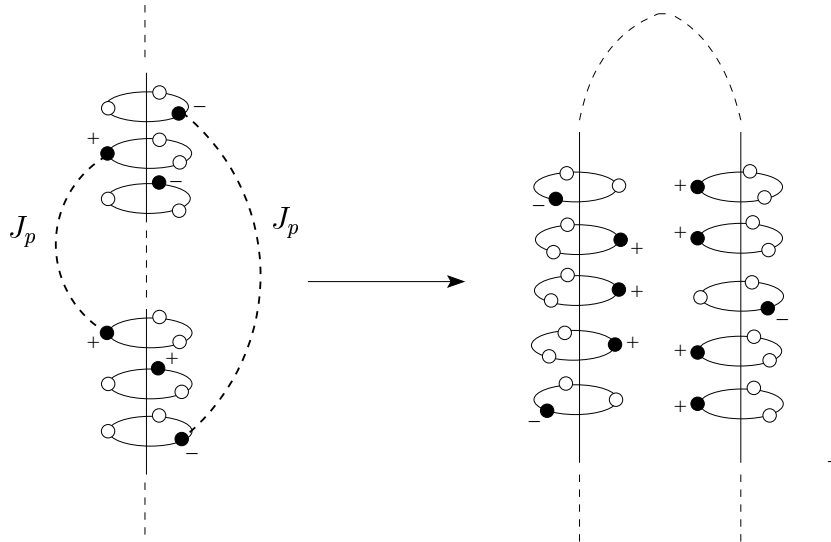
$$H_s = -J_s \sum_i \cos[(\phi_{i+1} - \phi_i) - (\phi_i - \phi_{i-1}) - a_i]$$

But now

- fast variables: amino-acid orientations $\{\phi_i\}$
- slow variables: amino-acid species $\{\xi_i, a_i\}$

theory with n -replicated transfer matrices
($n > 0$)

Tertiary structure generation



- given a distribution of orientations and polarities $\{\phi_i, \eta_i\}$ along the chain:
predict the 3D structure statistics
- express the energy gain from the 3D structure statistics
- find a self-consistent replacement for

$$H_p = -\frac{J_p}{N} \sum_{ij} \xi_i \xi_j \delta_{\phi_i, \phi_j}$$

- use the result for tertiary structure prediction

References

NS Skantzos, J Van Mourik, ACCC (2001)

J. Phys. A34 4437

H Chakravorty, ACCC, D Sherrington (2002)

J. Phys. A35 8647