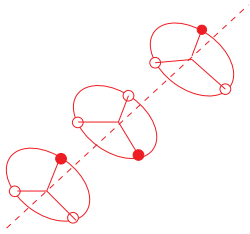
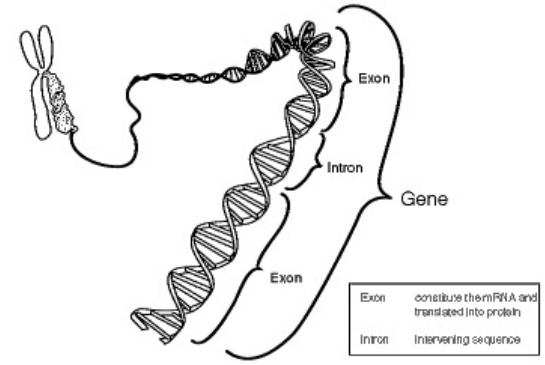
A solvable model of the genesis of amino-acid sequences via coupled dynamics of folding and slow genetic variation

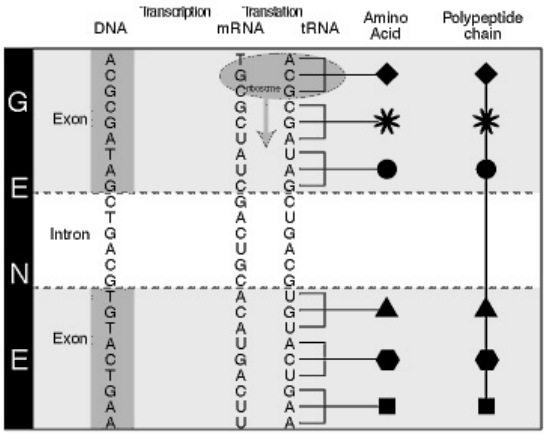
ACC Coolen, S Rabello, CJ Pérez-Vicente and F Fraternali

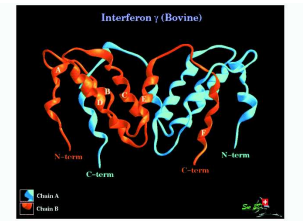


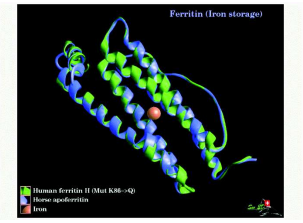
Motivation: proteins have non-random disorder ...
Dynamics of folding and sequence selection
Finite-n replica analysis, replicated transfer matrices
The limit $n \to \infty$, deterministic sequence selection
Numerical results, simulations
Summary and outlook

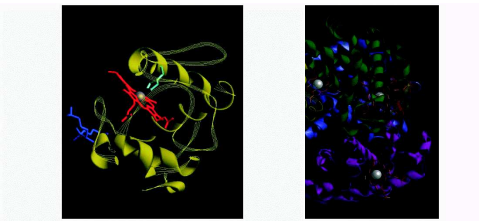
1. MOTIVATION











Proteins are disordered systems, but with non-random disorder ...

Primary structure: monomer sequence (the disorder), DHJKAFACGD ...
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.'*

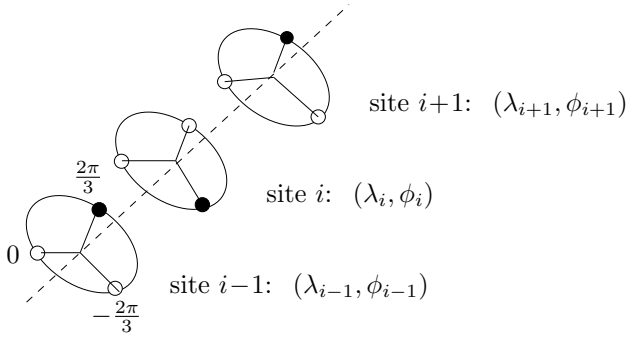
Problem for statistical mechanics
To use disordered systems techniques a la Parisi,
we need a *formula* for the disorder statistics ...

- Random amino-acid sequences do not fold into unique conformations, amino-acid sequences of proteins have been selected during evolution
- Our options for ensembles of sequences:
  (i) find a *formula for a nontrivial ensemble* of random amino-acid sequences?
  (ii) empirically: download all sequences from biomedical data base?

2. MODEL DEFINITIONS

slow process:		*fast process:*
genetic selection of sequences $\boldsymbol{\lambda}$	$\longrightarrow$	*folding of residue orientations* $\boldsymbol{\phi}$
Hamiltonian $H_{\rm eff}(\boldsymbol{\lambda})$	$\longleftarrow$	*Hamiltonian* $H_{\rm f}(\boldsymbol{\phi}\vert\boldsymbol{\lambda})$

- No defn of sequence statistics: define genetic *dynamics* of sequences
- Simple Hamiltonians, focus on secondary structure
- Solve coupled dynamics for disparate timescales using finite n replica method
- Exploit 1D nature of proteins: replicated transfer matrices

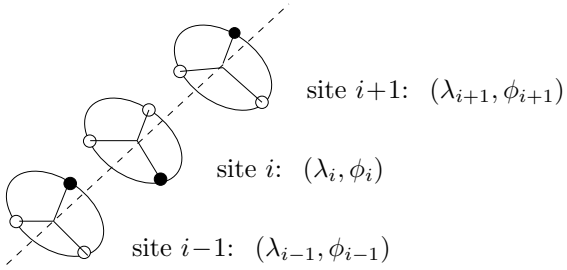


λ_i: the local amino-acid type
ϕ_i: residue angle relative to 'backbone'

primary structure: $(\lambda_1,\ldots,\lambda_N)$
secondary structure: $(\phi_1,\ldots,\phi_N)$

The fast process: folding

Variables: angles $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)$
$\phi_i = \{0, 2\pi/q, \ldots, (q-1)2\pi/q\}$,
$q = 2, 3, \ldots$

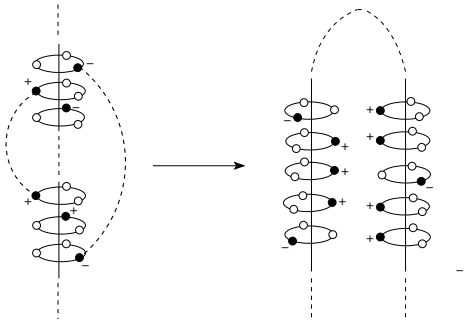


$$H_{\rm f}(\boldsymbol{\phi}|\boldsymbol{\lambda}) = -\frac{J_p}{N}\sum_{ij}\xi(\lambda_i)\xi(\lambda_j)\;\delta_{\phi_i,\phi_j} - J_s\sum_i \cos[(\phi_{i+1}-\phi_i) - (\phi_i-\phi_{i-1}) - a(\lambda_i)]$$

polarity energy *steric energy*

- polarity energy:
  proxy for energy gain by folding in 3D
  $\xi(\lambda)$: polarity of residue λ,
  $\xi > 0$: hydrophobic, $\xi < 0$: hydrophilic

- steric energy: mechanical constraints,
  residues 'stick out', distort homogeneous winding
  $a(\lambda)$: winding shift induced by residue λ



The slow process: genetic selection of sequences

sequence fitness:

(i) sequence must give protein with reproducible conformation
(ii) structure is useful, e.g. can act as catalyst of some reaction

translate into minimization of

$$H_{\text{eff}}(\boldsymbol{\lambda}) \;=\; U(\boldsymbol{\lambda}) + V(\boldsymbol{\lambda}) + F_{\text{f}}(\boldsymbol{\lambda})$$

- $U(\boldsymbol{\lambda})$: biological utility as catalyst
- $F_{\text{f}}(\boldsymbol{\lambda})$: free energy of folding process
  (low free energy = proxy for reproducible conformation)
- $V(\boldsymbol{\lambda})$: energetic cost of not having strictly hydrophilic 'surface residues' and strictly hydrophobic 'core residues'

stochastic minimization of Glauber type, noise level $\tilde{T}$:
genetic selection evolves to equilibrium state,

$$P_\infty(\boldsymbol{\lambda}) \;\propto\; \exp[-\tilde{\beta} H_{\rm eff}(\boldsymbol{\lambda})]$$

combined model solved in equilibrium
by calculating effective free energy

$$f_N \;=\; -\frac{1}{\tilde{\beta} N} \log \sum_{\boldsymbol{\lambda}} e^{-\tilde{\beta} H_{\rm eff}(\boldsymbol{\lambda})} = -\frac{1}{n\beta N} \log \sum_{\boldsymbol{\lambda}} e^{-n\beta[U(\boldsymbol{\lambda})+V(\boldsymbol{\lambda})]} [\mathcal{Z}_{\rm f}(\boldsymbol{\lambda})]^n$$

- temperature ratio $n = \tilde{\beta}/\beta$
- folding partition function $\mathcal{Z}_{\rm f}(\boldsymbol{\lambda}) = \sum_{\boldsymbol{\phi}} \exp[-\beta H_{\rm f}(\boldsymbol{\phi}|\boldsymbol{\lambda})]$
- solvable with replica method (finite n version)
- $n \to 0$ ($\tilde{T} \to \infty$): free energy of system with quenched random sequences
- effective free energy as generator of observables:

$$H_{\rm f}(\boldsymbol{\phi}|\boldsymbol{\lambda}) \to H_{\rm f}(\boldsymbol{\phi}|\boldsymbol{\lambda}) + \chi N G(\boldsymbol{\phi}, \boldsymbol{\lambda}) : \qquad \langle\langle G(\boldsymbol{\phi}, \boldsymbol{\lambda})\rangle_{\rm fast}\rangle_{\rm slow} = \lim_{\chi \to 0} \frac{\partial}{\partial \chi} f_N$$

Connections with earlier studies

- Skantzos, Van Mourik, ACCC *J. Phys. A 2001*
  random sequences (no genetic dynamics), but included hydrogen bonds
- Chakravorty, ACCC, Sherrington *J. Phys. A 2002*
  genetic dynamics, but only (long-range) polarity forces, $J_s = J_g = 0$

Simple choices for remaining parameters

- Sequence utility potential: $U(\boldsymbol{\lambda}) = \sum_i u(\lambda_i)$, $u(\lambda) = \mu\xi(\lambda) + \nu\cos[a(\lambda)]$
- Energetic cost of polarity imbalance: $V(\boldsymbol{\lambda}) = J_g N v(\frac{1}{N}\sum_i \xi(\lambda_i) - k^\star)$
- periodic boundary conditions, N even
- chemical characteristics of amino-acids statistically indep:

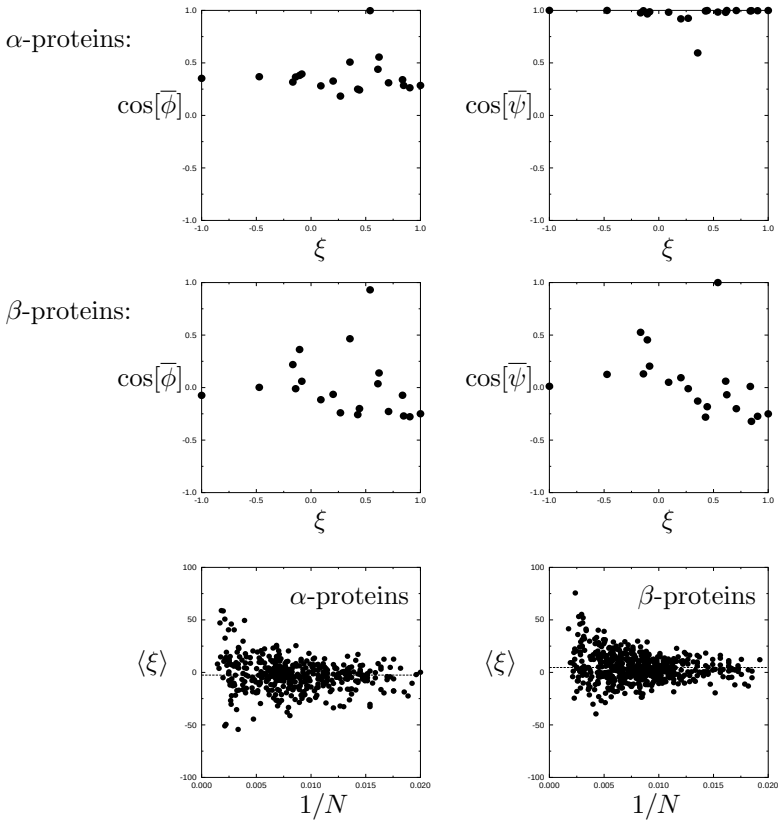
$$w(\xi,\eta) = \frac{1}{\Lambda}\sum_{\lambda=1}^{\Lambda} \delta[\xi - \xi(\lambda)]\delta[\eta - \cos[a(\lambda)]] = w(\xi)w(\eta)$$

Λ : nr of amino acid species, i.e. 20

Assumed amino-acid properties

$a(\lambda)$: winding shift of residue λ
$\xi(\lambda)$: polarity of residue λ

- independence of polarity and steric properties?
- preferred overall polarity $k^\star = N^{-1} \sum_i \xi(\lambda_i)$ (Eisenberg scale)



3. REPLICA ANALYSIS OF THE MODEL

write $\mathcal{Z}_{\rm f}^n(\boldsymbol{\lambda})$ in terms of n replicas of the system,
sum over sequences $\boldsymbol{\lambda}$ before sum over conformations,
$f = \lim_{N\to\infty} f_N = \mathrm{extr}_{\mathbf{z}} \varphi_n(\mathbf{z})$

$$n\varphi_n(\mathbf{z}) = J_p \sum_{\alpha\phi} z_{\alpha\phi}^2 + nJ_g[v(\frac{1}{n}\sum_{\alpha\phi} z_{\alpha\phi} - k^\star) - (\frac{1}{n}\sum_{\alpha\phi} z_{\alpha\phi}) v'(\frac{1}{n}\sum_{\alpha\phi} z_{\alpha\phi} - k^\star)] - \frac{1}{\beta}\log\Lambda$$
$$- \lim_{N\to\infty} \frac{1}{\beta N} \log \sum_{\boldsymbol{\phi}^1 \ldots \boldsymbol{\phi}^n} \prod_i M[\boldsymbol{\phi}_{i-1}, \boldsymbol{\phi}_i, \boldsymbol{\phi}_{i+1}|\mathbf{z}]$$

$$M[\boldsymbol{\phi}_{i-1}, \boldsymbol{\phi}_i, \boldsymbol{\phi}_{i+1}|\mathbf{z}] = \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} e^{\beta\xi(\lambda)\sum_\alpha [2J_p z_{\alpha\phi_i^\alpha} - J_g v'(\frac{1}{n}\sum_{\alpha\phi} z_{\alpha\phi} - k^\star)] + \beta J_s \sum_\alpha \cos[\phi_{i+1}^\alpha + \phi_{i-1}^\alpha - 2\phi_i^\alpha - a(\lambda)] - n\beta u(\lambda)}$$

structure: replicated transfer matrix product
embedded within a mean-field calculation
in principle solvable!
only order pars with one replica index, so RS ok

Simplest case $q = 2$: $\phi_i \in \{-\pi/2, \pi/2\}$

$\phi_i = \sigma_i \pi/2$, with $\sigma_i = \pm 1$

$$H_{\rm f}(\boldsymbol{\sigma}|\boldsymbol{\lambda}) = -\frac{J_p}{2N}\sum_{ij}\xi(\lambda_i)\xi(\lambda_j)[1+\sigma_i\sigma_j] - J_s\sum_i \cos[a(\lambda_i)]\sigma_{i+1}\sigma_{i-1}$$

solution:

$$f \;=\; \mathrm{extr}_{m,k}\left\{\frac{1}{2}J_p(m^2{+}k^2) + J_g[v(k{-}k^\star){-}kv'(k{-}k^\star)] - \frac{\log\Lambda}{\beta n} - \frac{1}{\beta n}\log\lambda(m,k)\right\}$$

$\lambda(m,k)$: largest eigenvalue
of $2^n \times 2^n$ transfer matrix

$$M_{\boldsymbol{\sigma}\boldsymbol{\sigma}'}(m,k) \;=\; \langle e^{\beta\eta[J_s\boldsymbol{\sigma}\cdot\boldsymbol{\sigma}'-n\nu]}\rangle_\eta \langle e^{n\beta\xi[J_p(k+\frac{m}{n}\sum_\alpha\sigma_\alpha)-\mu-J_g v'(k-k^\star)]}\rangle_\xi$$

$\langle g(\xi)\rangle = \int d\xi\; w(\xi)g(\xi)$,
$\langle g(\eta)\rangle = \int d\eta\; w(\eta)g(\eta)$

physical meaning of $\{m, k\}$:

$$m = \lim_{N\to\infty} \frac{1}{N} \sum_i \langle \xi(\lambda_i) \langle \sigma_i \rangle_{\text{fast}} \rangle_{\text{slow}} \qquad k = \lim_{N\to\infty} \frac{1}{N} \sum_i \langle \xi(\lambda_i) \rangle_{\text{slow}}$$

saddle-point equations:

$$m = \frac{\sum_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} u^{\text{L}}_{\boldsymbol{\sigma}} \sigma_1 Y_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} u^{\text{R}}_{\boldsymbol{\sigma}'}}{\lambda(m,k) \sum_{\boldsymbol{\sigma}} u^{\text{L}}_{\boldsymbol{\sigma}} u^{\text{R}}_{\boldsymbol{\sigma}}} \qquad k = \frac{\sum_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} u^{\text{L}}_{\boldsymbol{\sigma}} Y_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} u^{\text{R}}_{\boldsymbol{\sigma}'}}{\lambda(m,k) \sum_{\boldsymbol{\sigma}} u^{\text{L}}_{\boldsymbol{\sigma}} u^{\text{R}}_{\boldsymbol{\sigma}}}$$

where

$$Y_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} = \left\langle e^{\beta\eta[J_s \boldsymbol{\sigma}\cdot\boldsymbol{\sigma}' - n\nu]} \right\rangle_\eta \left\langle \xi e^{n\beta\xi[J_p(k+\frac{m}{n}\sum_\alpha \sigma_\alpha) - \mu - J_g v'(k-k^\star)]} \right\rangle_\xi$$

$$\sum_{\boldsymbol{\sigma}'} M_{\boldsymbol{\sigma}\boldsymbol{\sigma}'} u^{\text{R}}_{\boldsymbol{\sigma}'} = \lambda(m,k) u^{\text{R}}_{\boldsymbol{\sigma}}, \qquad \sum_{\boldsymbol{\sigma}'} u^{\text{L}}_{\boldsymbol{\sigma}'} M_{\boldsymbol{\sigma}'\boldsymbol{\sigma}} = \lambda(m,k) u^{\text{L}}_{\boldsymbol{\sigma}}$$

Solution of replicated eigenvalue problem

$$u^{\mathrm{R}}_{\boldsymbol{\sigma}} = \int dx\ \Phi(x) e^{\beta x \sum_\alpha \sigma_\alpha}, \qquad u^{\mathrm{L}}_{\boldsymbol{\sigma}} = \int dy\ \Psi(y) e^{\beta y \sum_\alpha \sigma_\alpha}$$

from replicated spins
to effective fields:

$$\lambda\Phi(x) = \int dx' \Lambda_\Phi(x,x')\Phi(x') \qquad \lambda\Psi(x) = \int dx' \Lambda_\Psi(x,x')\Psi(x')$$

$$\begin{aligned}
\Lambda_\Phi(x,x') &= \langle\!\langle \delta[x - \xi J_p m - A(x', \eta J_s)] e^{n\beta[B(x',\eta J_s) + \xi(J_p k - \mu - J_g v'(k-k^\star)) - \nu\eta]} \rangle\!\rangle_{\xi,\eta} \\
\Lambda_\Psi(x,x') &= \langle\!\langle \delta[x - A(x' + \xi J_p m, \eta J_s)] e^{n\beta[B(x'+\xi J_p m,\eta J_s) + \xi(J_p k - \mu - J_g v'(k-k^\star)) - \nu\eta]} \rangle\!\rangle_{\xi,\eta}
\end{aligned}$$

with

$$\begin{aligned}
A(x,y) &= \beta^{-1} \tanh^{-1}[\tanh(\beta x)\tanh(\beta y)] \\
B(x,y) &= \frac{1}{2}\beta^{-1} \log[4\cosh[\beta(x+y)]\cosh[\beta(x-y)]]
\end{aligned}$$

everything follows from Φ,Ψ ...

simplify, play around ...

$$m = \int d\xi dh\ W(h,\xi)\ \xi \tanh(\beta h) \qquad k = \int d\xi dh\ W(h,\xi)\ \xi$$

$$W(h,\xi) = \frac{p(\xi)\cosh^n[\beta h]\int dx\ \Psi(x)\Psi(h-x-J_p m\xi)}{\int d\xi' dh' p(\xi')\cosh^n[\beta h']\int dx\ \Psi(x)\Psi(h'-x-J_p m\xi')}$$

in which

$$p(\xi) = \frac{w(\xi)e^{n\beta\xi(J_p k-\mu-J_g v'(k-k^\star))}}{\int d\xi'\ w(\xi')e^{n\beta\xi'(J_p k-\mu-J_g v'(k-k^\star))}}$$

$$\Psi(x) = \frac{\int dx'\Phi(x')\int d\eta\ w(\eta)\delta[x-A(x',\eta J_s)]e^{n\beta[B(x',\eta J_s)-\nu\eta]}}{\int dx'\Phi(x')\int d\eta\ w(\eta)e^{n\beta[B(x',\eta J_s)-\nu\eta]}} \qquad \Phi(x) = \int d\xi\ p(\xi)\Psi(x-J_p m\xi)$$

formulas for f and for

$$\pi(\xi,\eta) = \lim_{N\to\infty}\frac{1}{N}\sum_i \langle\langle \delta[\xi-\xi(\lambda_i)]\delta[\eta-\cos[a(\lambda_i)]]\rangle\rangle$$

e.g. $\pi(\xi,\eta) = \pi(\xi)\pi(\eta)$, $\pi(\xi) = \int dh\ W(h,\xi)$

Simple solutions and special cases

- state without secondary structure (always a soln): $m = 0$

$$\Psi(x) = \Phi(x) = \delta(x), \qquad W(h,\xi) = p(\xi)\delta(h), \qquad k = \frac{\int d\xi\ \xi\ w(\xi) e^{n\beta\xi[J_p k - \mu - J_g v'(k-k^\star)]}}{\int d\xi\ w(\xi) e^{n\beta\xi[J_p k - \mu - J_g v'(k-k^\star)]}}$$

- infinite temperature: $\beta = 0$

$$\Psi(x) = \delta(x), \qquad W(\xi,h) = w(\xi)\delta(h), \qquad m = 0, \qquad k = \int d\xi\ \xi\ w(\xi)$$

$$\lim_{\beta\to 0} \beta f = -n^{-1}\log\Lambda - \log 2$$

- Random sequences: $n \to 0$

$$\Psi(x) = \int dy\ \Psi(y) \langle\!\langle \delta[x - A(y + J_p m\xi, \eta J_s)] \rangle\!\rangle_{\xi,\eta}, \qquad \Phi(x) = \langle \Psi(x - J_p m\xi) \rangle_\xi$$

$$m = \int dx dx'\ \Phi(x')\Psi(x) \langle\langle \xi \tanh[\beta(x + \xi J_p m + A(x', \eta J_s))] \rangle\rangle_{\xi,\eta}$$

recovers Skantzos et al 2001
(random bond chain methods, ratios of constrained partition functions)

4. DETERMINISTIC SEQUENCE SELECTION

choose $v(u) = \frac{1}{2}u^2$,
define natural polarity balance $\qquad k_0 = \dfrac{k^\star - \mu/J_g}{1 - J_p/J_g}$

take $n \to \infty$ in system below:

$$
\begin{aligned}
\Psi(x) &= \frac{\int dx' \int d\xi\ p(\xi)\Psi(x') \int d\eta\ w(\eta)\delta[x - A(x' + J_p m\xi, \eta J_s)] e^{n\beta[B(x' + J_p m\xi, \eta J_s) - \nu\eta]}}{\int dx' \int d\xi\ p(\xi)\Psi(x') \int d\eta\ w(\eta) e^{n\beta[B(x' + J_p m\xi, \eta J_s) - \nu\eta]}} \\
m &= \frac{\int d\xi\ p(\xi)\xi \int dx dy\ \Psi(x)\Psi(y) \tanh[\beta(J_p m\xi + x + y)] \cosh^n[\beta(J_p m\xi + x + y)]}{\int d\xi\ p(\xi) \int dx dy\ \Psi(x)\Psi(y) \cosh^n[\beta(J_p m\xi + x + y)]} \\
k &= \frac{\int d\xi\ p(\xi)\xi \int dx dy\ \Psi(x)\Psi(y) \cosh^n[\beta(J_p m\xi + x + y)]}{\int d\xi\ p(\xi) \int dx dy\ \Psi(x)\Psi(y) \cosh^n[\beta(J_p m\xi + x + y)]} \\
p(\xi) &= \frac{w(\xi) e^{n\beta\xi(J_p - J_g)(k - k_0)}}{\int d\xi'\ w(\xi') e^{n\beta\xi'(J_p - J_g)(k - k_0)}}
\end{aligned}
$$

Form of $\Psi(x)$ for $n \to \infty$

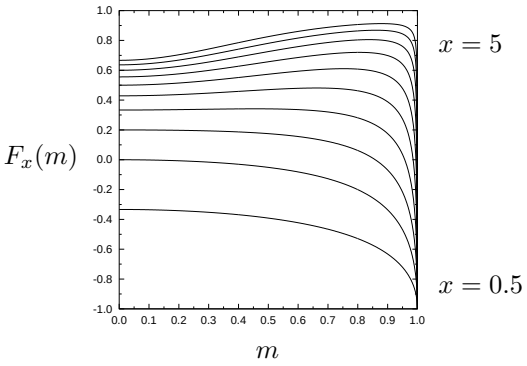
- $\exists \Omega \subseteq [-J_s, J_s]$: $\Psi(x) = 0$ for $x \notin \Omega$, $\Psi(x) = e^{n\beta\psi(x)}$ for $x \in \Omega$
- $\max_{x\in\Omega} \psi(x) = 0$
- need to find Ω and $\lim_{n\to\infty} \psi(x)$

several pages later ...

$$J_g > J_p: \quad k = k_0, \ \text{heteropolar}, \qquad m = 0 \quad \text{or} \quad F_{\beta J_p}(m) = -\tanh(\beta J_s)$$
$$J_g < J_p: \quad k = \pm 1, \ \text{homopolar}, \qquad m = 0 \quad \text{or} \quad F_{\beta J_p}(m) = \operatorname{sgn}(\nu)\tan(\beta J_s)$$

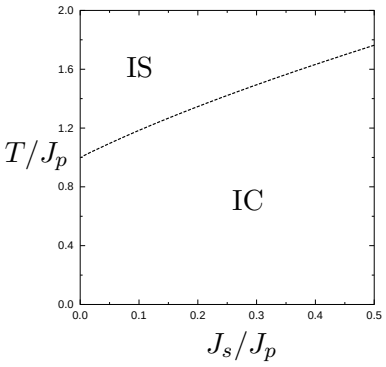
with

$$F_x(m) = \frac{\tanh[\frac{1}{2}xm - \frac{1}{2}\tanh^{-1}(m)]}{\tanh[\frac{1}{2}xm + \frac{1}{2}\tanh^{-1}(m)]}$$

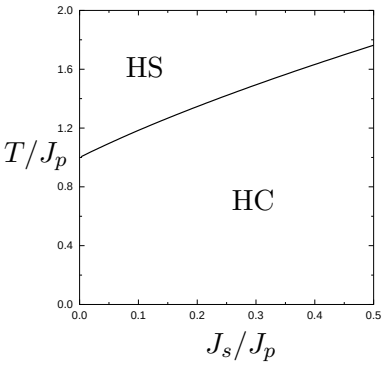


Phase diagrams

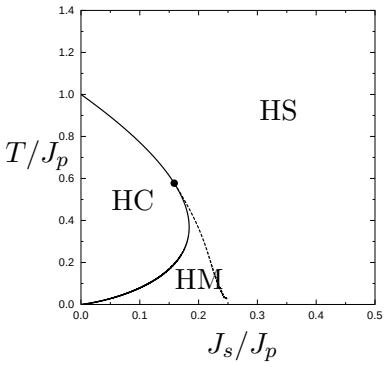
$J_g > J_p$



$J_g < J_p,\ \nu < 0$



$J_g < J_p,\ \nu > 0$



inhom polarity, swollen (IS): $\pi(\xi)$ continuous, $m = 0$
inhom polarity, collapsed (IC): $\pi(\xi) = \frac{1}{2}(1+k_0)\delta(\xi-1)+\frac{1}{2}(1-k_0)\delta(\xi+1)$, $m \neq 0$
hom polarity, swollen (HS): $\pi(\xi) = \delta(\xi \pm 1)$, $m = 0$
hom polarity, collapsed (HC): $\pi(\xi) = \delta(\xi \pm 1)$, $m \neq 0$
hom polarity, mixed (HM): $\pi(\xi) = \delta(\xi \pm 1)$, coexistence of $m = 0$ and $m \neq 0$

$\nu > 0$: favours helices, $\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow \ldots$
$\nu < 0$: favours β-sheets, $\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow \ldots$

5. NON-DETERMINISTIC SEQUENCE SELECTION

Transitions for finite n, increased genetic noise

Generally hard ...
except *continuous* transitions away from $m = 0$

$$m \to \Delta m, \qquad k \to k + \Delta k, \qquad \Psi(x) \to \delta(x) + \Delta\Psi(x)$$

gives

$$\begin{aligned}
\Delta\Psi(x) &= \frac{\int dx' [\Delta\Psi(x') - J_p k \Delta m \delta'(x')] \int d\eta \; w(\eta) \delta[x - A(x', \eta J_s)] e^{n\beta[B(x', \eta J_s) - \nu\eta]}}{\int d\eta \; w(\eta) e^{n\beta[B(0, \eta J_s) - \nu\eta]}} \\
\Delta m &= 2k \int dh \; \tanh(\beta h) \cosh^n(\beta h) \Delta\Psi(h) + \beta J_p \Delta m \int d\xi \; p(\xi) \xi^2 + \mathcal{O}(\Delta^2)
\end{aligned}$$

soln:

$$\Delta\Psi_A(x) = \frac{\lambda J_p k}{\lambda - 1} \; \delta'(x) \Delta m \qquad \lambda = \frac{\int d\eta \; w(\eta) \tanh(\beta \eta J_s) e^{n\beta[B(0, \eta J_s) - \nu\eta]}}{\int d\eta \; w(\eta) e^{n\beta[B(0, \eta J_s) - \nu\eta]}}$$

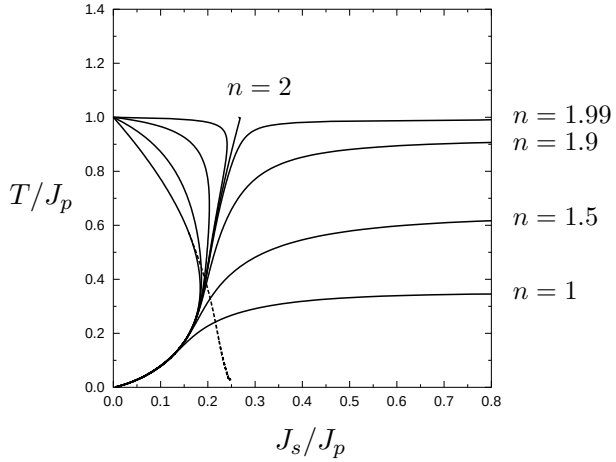
continuous $m \neq 0$ bifurcations:

$$\Delta m \neq 0: \qquad 1 = \beta J_p[\int d\xi\ \xi^2 p(\xi) - \frac{2\lambda k^2}{\lambda - 1}]$$

$$p(\xi) = \frac{w(\xi) e^{n\beta\xi(J_p - J_g)(k-k_0)}}{\int d\xi' w(\xi') e^{n\beta\xi'(J_p - J_g)(k-k_0)}} \qquad \lambda = \frac{\int d\eta\ w(\eta) \tanh(\beta\eta J_s) e^{n\beta[B(0,\eta J_s) - \nu\eta]}}{\int d\eta\ w(\eta) e^{n\beta[B(0,\eta J_s) - \nu\eta]}}$$

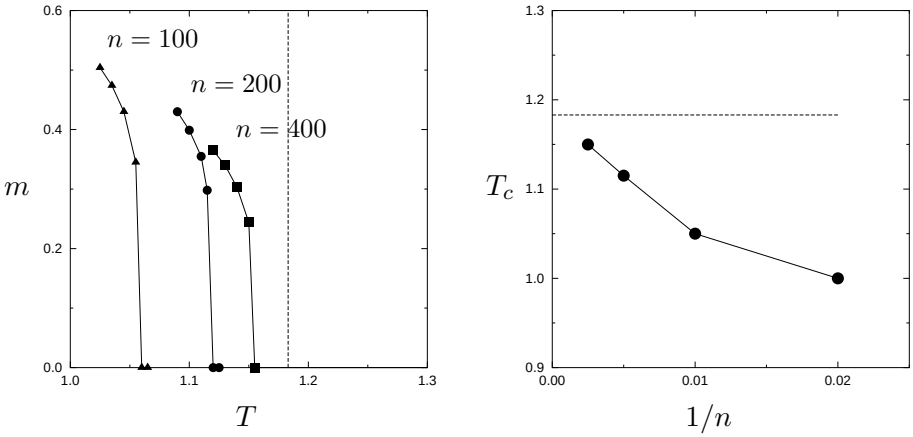
e.g. $J_g <= \frac{1}{2} J_p$, $\nu = \frac{1}{2}$:

onset of discont transition at $n = 2$!
(as in other coupled dynamics models)



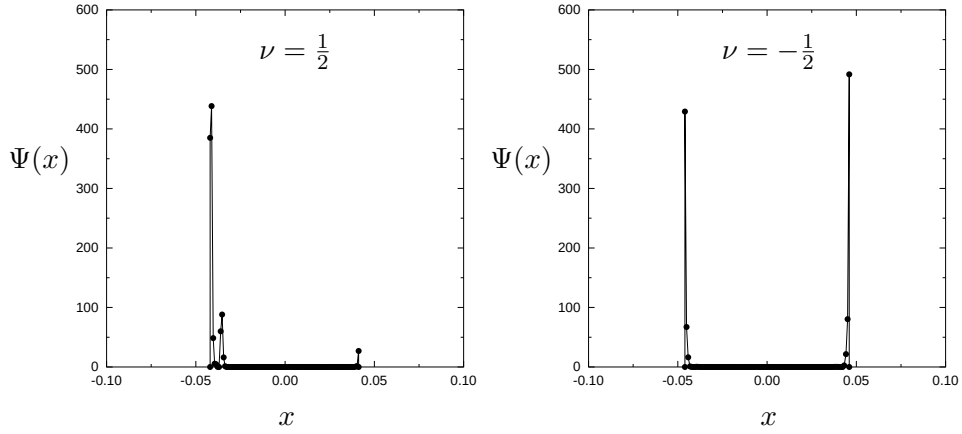
Numerical solution via population dynamics

n appears in exponents,
which limits numerical analysis to $n \leq 400$



$(J_s, J_p, J_g) = (0.1, 1, 2)$, $k_0 = 0.7$, $\mu = 0.2$, $\nu = 0.5$
$n \to \infty$: continuous IS→IC transition at $T_c = 1.183$
large but finite n: discontinuous

prediction: $\lim_{n\to\infty} \Psi(x) = \frac{1}{2}\delta(x + x^\star) + \frac{1}{2}\delta(x - x^\star)$
finite n corrections: $\mathcal{O}(n^{-1/2})$ for $\nu > 0$, $\mathcal{O}(n^{-1})$ for $\nu < 0$

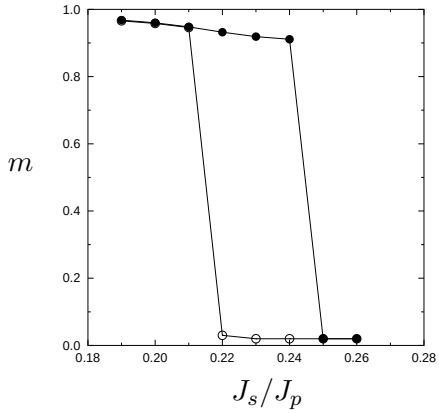


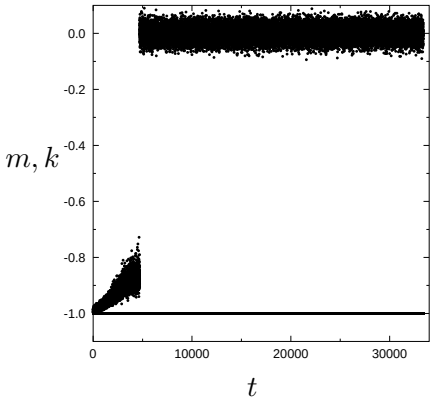
$(J_s, J_p, J_g) = (0.1, 1, 2)$, $n = 200$, $k_0 = 0.2$, $\mu = 0.7$, and $T = 1.07$

$\nu > 0$: favours helices, $\uparrow\downarrow\uparrow\downarrow\uparrow\downarrow\uparrow \ldots$
$\nu < 0$: favours β-sheets, $\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow \ldots$

6. NUMERICAL SIMULATIONS

requires two nested equilibrations of disordered systems,
inner 'loop' of the code: disordered Ising chain ... HARD!
N too small: no transitions, N too large: no equilibration





$J_s/J_p = 0.25$

$N = 1000$, at $T = 0.3$ and $n = 200$
$\nu = J_g = \frac{1}{2}$, $J_p = 1$, and and $k^\star = 0$. For $n \to \infty$: coexistence & remanence

7. SUMMARY AND OUTLOOK

nice:

- solvable models describing protein structure formation
  circumvent the obstacle of non-random amino-acid sequences
- nested equilibration of slow/fast processes: finite n replica method
  short-range frozen random forces: diagonalization of replicated transfer matrix
- exact results for phase transitions,
  especially for deterministic sequence selection, $n \to \infty$

not so nice:

- many simplifications:
  one angle per residue (should be two), simple phenomenological Hamiltonian
  no hydrogen bonds, only primary & secondary structure
- potential for evolution to homo-polar polymers,
  artifact of Hamiltonian? probably ...
- statements on *ensemble* of hetero-polymers,
  not solution of protein folding problem (not even approximate)

Future directions

If driven by passion for theory ...

- introduce contact maps to replace present long-range forces,
  structure similar to 'small-world' topologies,
  more sophisticated order parameters of finitely connected graphs, RSB, etc
- real-valued residue orientations, i.e. $q \to \infty$
  diagonalization of replicated transfer kernels

If driven by passion for biology ...

- increase level of biological detail:
  two residue angles, with real rather than discrete values,
  more realistic Hamiltonians:
    work our steric effects for real amino-acids
    include hydrogen bonds
    more realistic modeling of tertiary structure influence, via contact maps

there is overlap!