united nations
educational, scientific
and cultural
organization

international atomic
energy agency

the
abdus salam
international centre for theoretical physics

SMR/1499 - 12

INTERNATIONAL WORKSHOP ON PROTEOMICS: PROTEIN STRUCTURE, FUNCTION AND INTERACTIONS

(5 - 16 May 2003)

"Solvent effects on simple protein models"

presented by:

P. De Los Rios
University of Lausanne
Switzerland

Solvent Effects on Simple Protein Models

Paolo De Los Rios
Giovanni Salvi
Susanne Mölbert

Institut de Physique Théorique
Université de Lausanne

October 1st, 2003

Trieste, 7/5/2003

Can we bridge the gap between explicit and implicit solvent models?

Implicit solvent: simple and fast, but… how reliable?

Anything here?

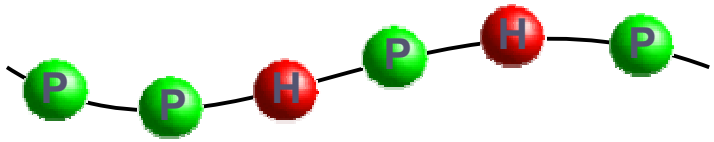
Explicit solvent: detailed and reliable, but… too slow

Outline

- A simple implicit solvent models: the HP model
- The semi-explicit solvent: the Müller-Lee-Graziano model
- A simple semi-explicit solvent model: the HPW model
- Can we *learn* effective pairwise interactions?

Implicit solvent: The HP model (K. Dill, 1989)

Amino-acid alphabet: H (hydrophobic), P (polar)

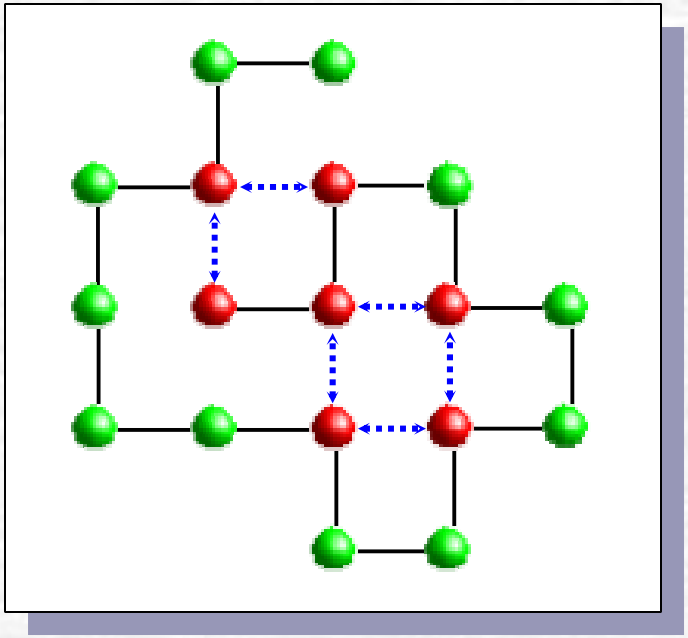


Proteins: self-avoiding walks (SAW) on a 2D lattice

The hydrophobic effect is expressed by an effective interaction between nearest neighbors H amino-acids.

$$\varepsilon_{HH} < \varepsilon_{HP} \; (= \varepsilon_{PH}) \leq \varepsilon_{PP}$$

Mathematical Formulation



Hamiltonian

$$H = \sum_{i<j} \varepsilon_{p_i p_j} \Delta\left(r_i, r_j\right)$$

with:

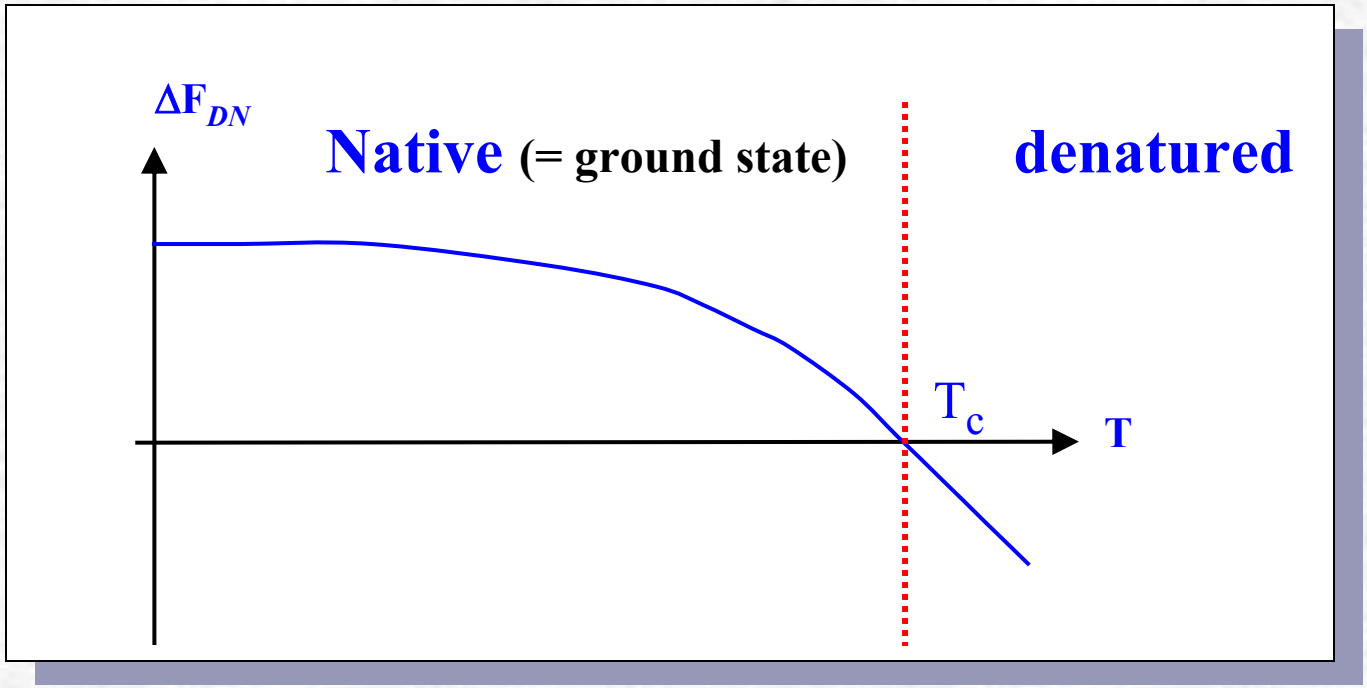
ε_{XY} : **interaction values**

p_i : **H or P**

$$\Delta(r_i,r_j) = \begin{cases} 1 & \text{if } i,j \text{ are } n.n. \\ 0 & \text{otherwise} \end{cases}$$

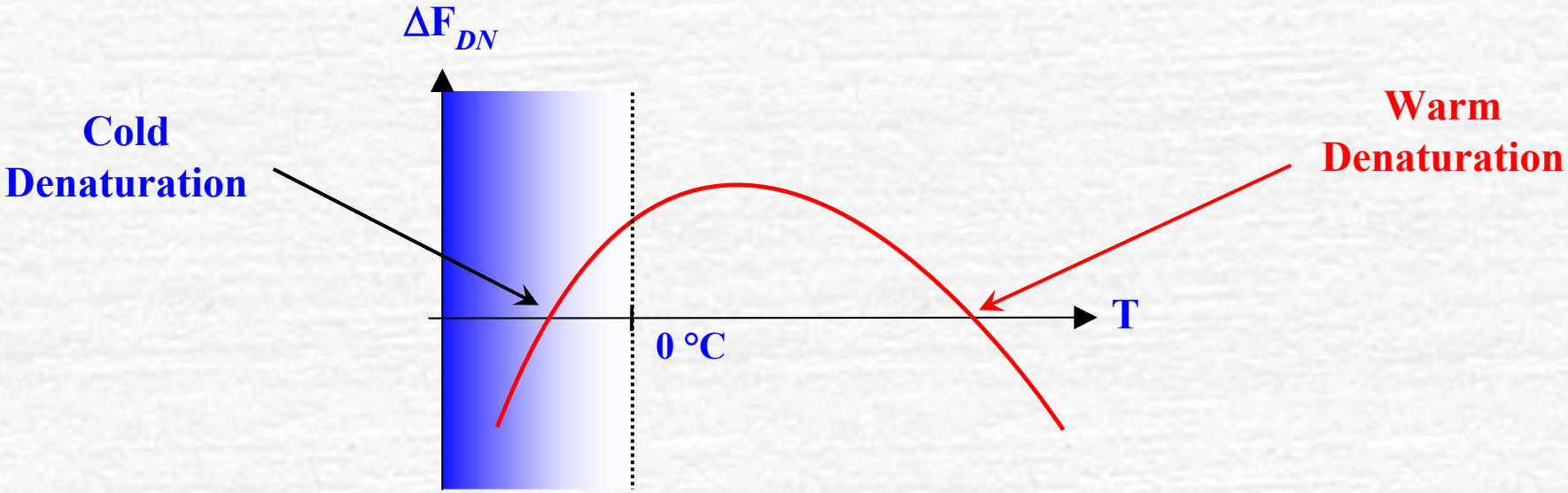
$\varepsilon_{HH}=-1 \quad \varepsilon_{HP} = \varepsilon_{PH} = \varepsilon_{PP} = 0$

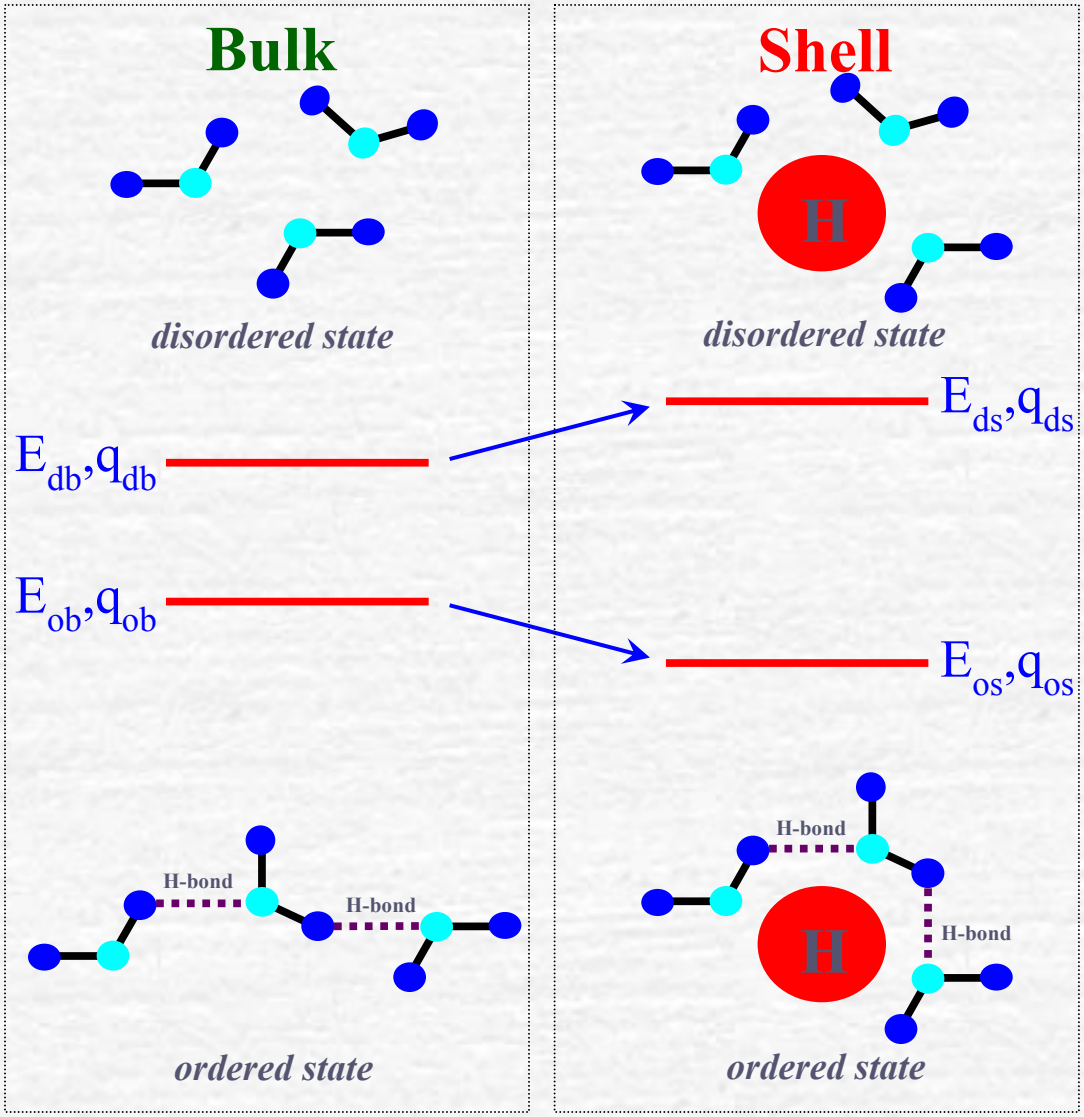
Given a good sequence (= with a non-degenerate native state)



Only a subset of all sequences are good sequences, and have a native state with a hydrophobic core

Folding free energy difference for real proteins





Müller-Lee-Graziano (1996)

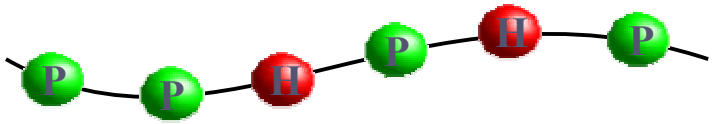
Experiments and simulations suggest

$$E_{ds} > E_{db} > E_{ob} > E_{os}$$
$$q_{ds} > q_{db} > q_{ob} > q_{os}$$

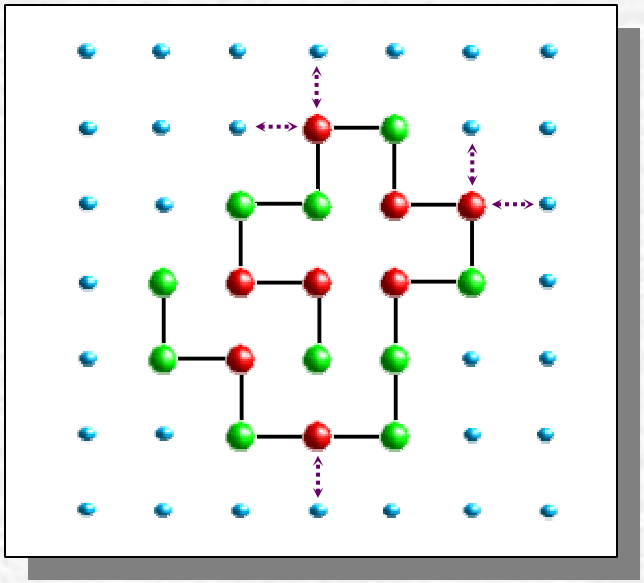
Explicit solvent: The HPW model

P. De Los Rios & G. Caldarelli, Phys. Rev. E **62**, 8449 (2000)

Amino-acids alphabet: H (hydrophobic), P (polar)



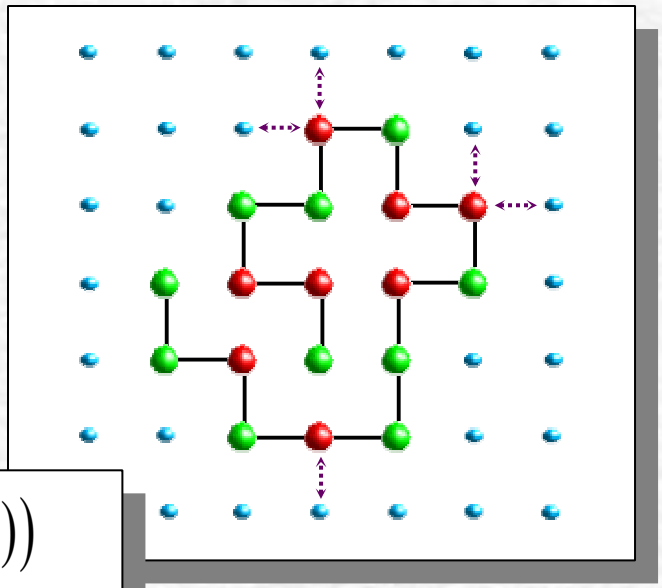
The solvent is (semi)explicitly taken into account. Sites not occupied by a.a. are occupied by *groups of water molecules*.



Proteins:
self-avoiding walks (SAW)
on a 2D lattice

The water behavior is described by the MLG model.

The system: protein + solvent



Energy

$$E=\sum_{\langle i,H\rangle}\left(E_{os}\delta_{i,os}+E_{ds}\left(1-\delta_{i,os}\right)\right)+\sum_{(i,H)}\left(E_{ob}\delta_{i,ob}+E_{db}\left(1-\delta_{i,ob}\right)\right)$$

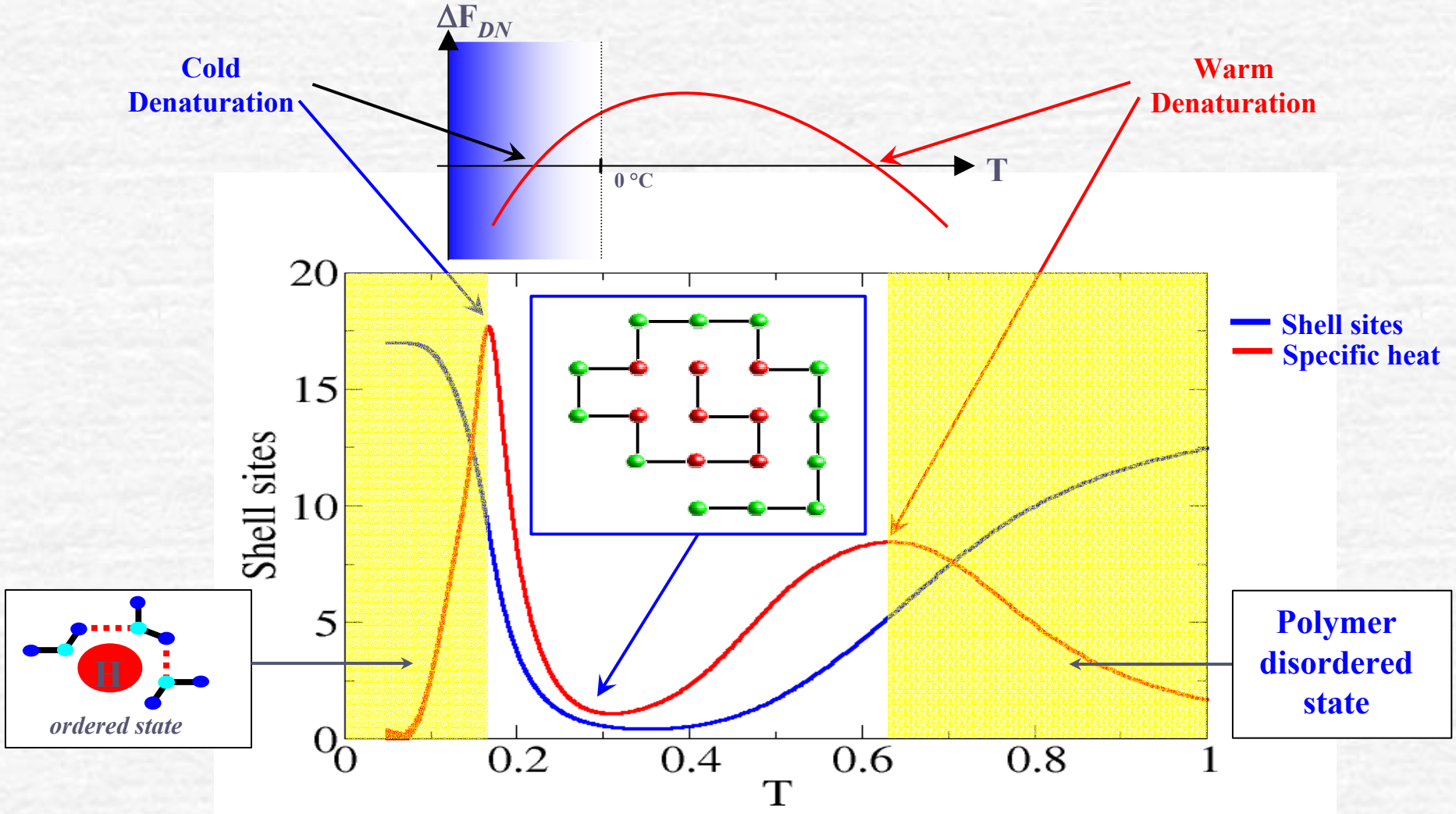
Conformation Partition function

$$Z(S,\Gamma)=\left(q_{ob}e^{-\beta E_{ob}}+q_{db}e^{-\beta E_{db}}\right)^{n_b(\Gamma)}\left(q_{os}e^{-\beta E_{os}}+q_{ds}e^{-\beta E_{ds}}\right)^{n_s(\Gamma)}$$

Γ = conformation

S = sequence

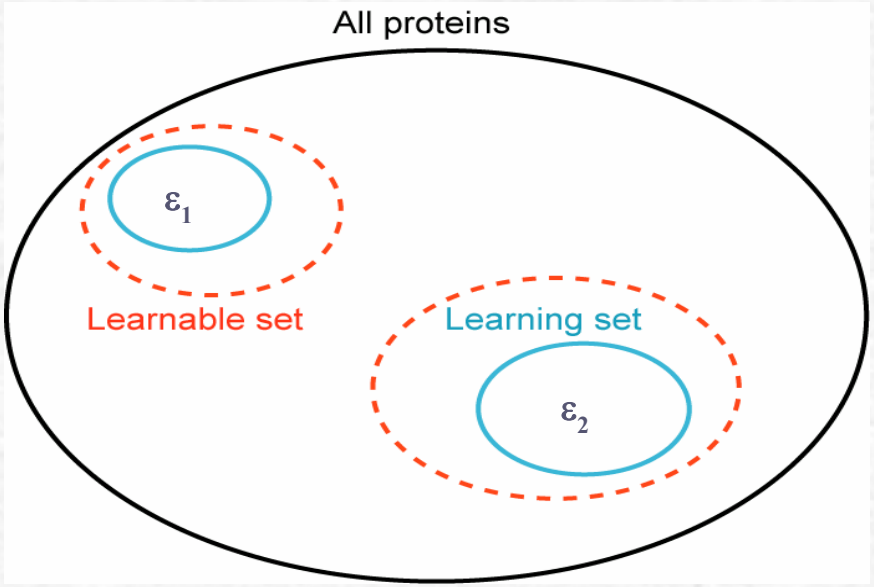
Native state and denaturations



Learning effective potentials

Inference of potentials from known native structures (from the Protein Data Bank)

Problem: no potentials able to stabilize all native states



We need to create our own "virtual" PDB for simple models

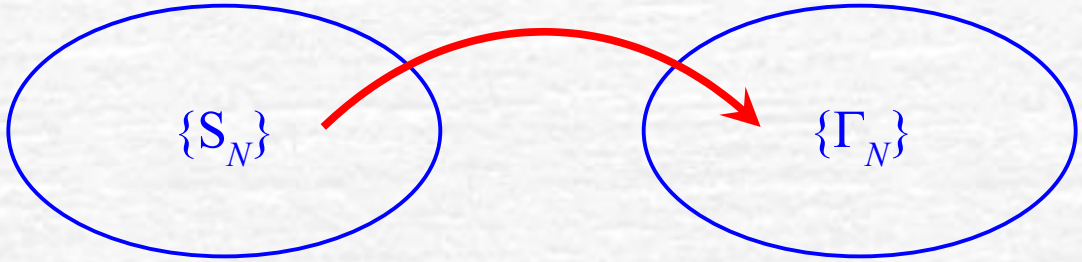
↓

Protein design

Cost function

Partial free energy

$$F(S,\Gamma) = -k_B T \cdot \ln\left(Z(S,\Gamma)/Z(S)\right)$$



↓ **Exact Enumeration**

Good sequences: those with a unique conformation with the lowest cost function value

e.g. L = 20 : # $\{S_{20}\} = 2^{20} \cong 10^6$

$\{S'_{20}\} < 40.000$

Effective a.a. interaction Hamiltonian

$$H = \sum_{i<j} \varepsilon_{p,p,\Delta}(r_i, r_j)$$

$$H(S,\Gamma) - H(S,\Gamma_n) > 0 \quad \forall\, \Gamma \neq \Gamma_n, \forall\, S$$

$$H = \sum_i \varepsilon_i C_i(S,\Gamma) \quad \longrightarrow \quad \sum_i \varepsilon_i C_i(S,\Gamma) - \sum_i \varepsilon_i C_i(S,\Gamma_n) = \vec{\varepsilon} \cdot \vec{C}^{\Delta}(S,\Gamma,\Gamma_n) > 0$$

with: $C^{\Delta}_i(S,\Gamma,\Gamma_n) = C_i(S,\Gamma) - C_i(S,\Gamma_n)$

$\vec{\varepsilon} = (\varepsilon_{HH}, \varepsilon_{HP}, \ldots) =$?

The perceptron

Trial vector ε_n

- Calculate *all* Scalar Products $C^{\Delta}_i \cdot \varepsilon_n$
- Worst Scalar Product $C^{\Delta}_w \cdot \varepsilon_n$
- $\varepsilon_{n+1} = \varepsilon_n + \eta\, C^{\Delta}_w$
- Iterate the process

$\vec{\varepsilon}_n$ converges to $\vec{\varepsilon}$

Decoys from Protein Design

Sequence 1	$\Gamma_n(S_1)$	15 $\Gamma(S_1)$	15 $C^{\Delta}(S_1)$
Sequence N	$\Gamma_n(S_N)$	15 $\Gamma(S_N)$	15 $C^{\Delta}(S_N)$

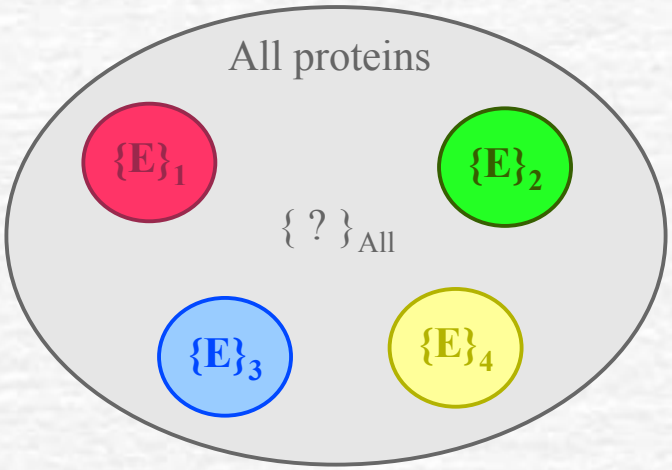
HP Model results

- The perceptron converges to the initial interactions values.

$$\varepsilon_{ij}^{perc} \cong \varepsilon_{ij} \quad i, j = H, P$$

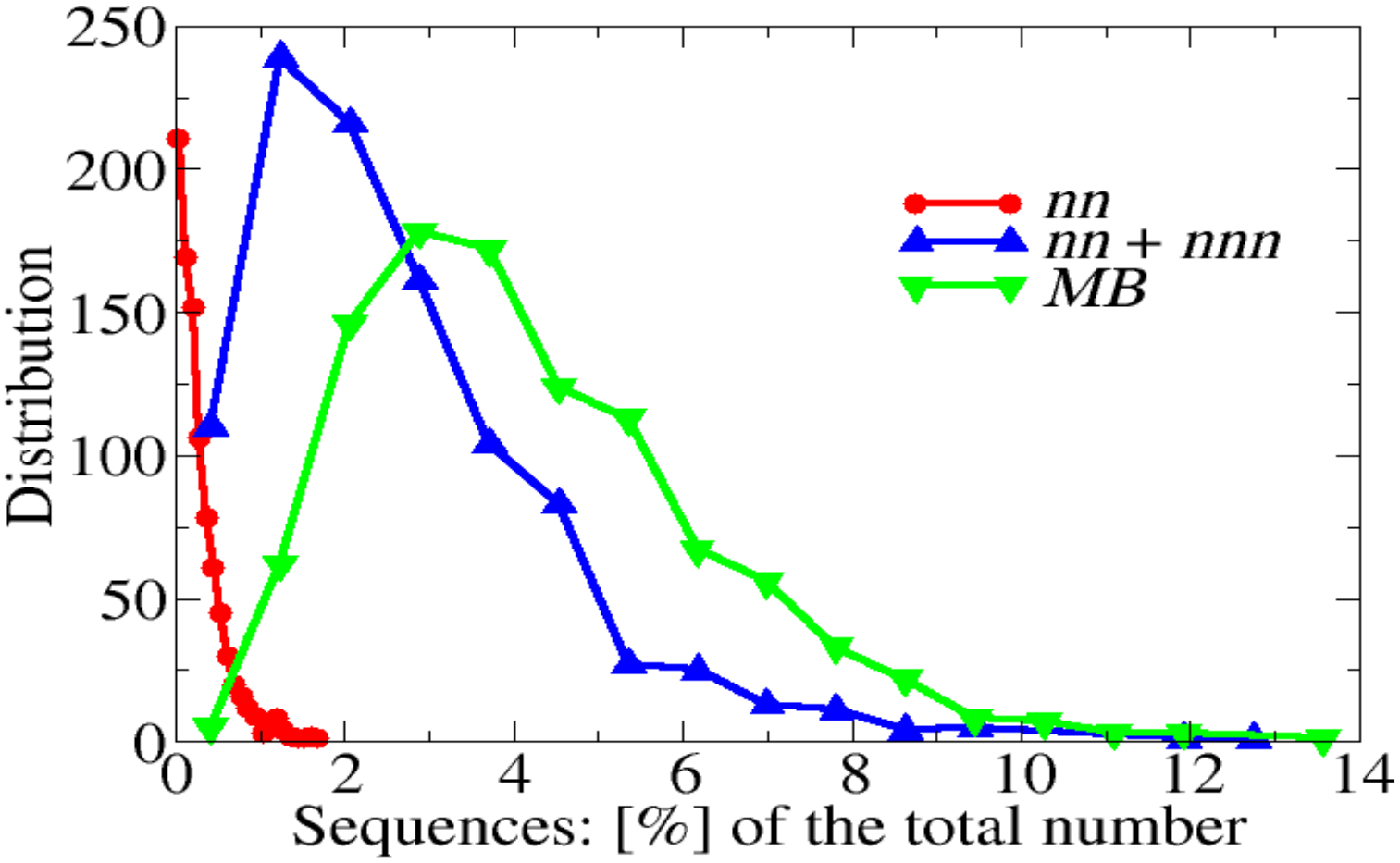
- Only a small set is needed to reach the original values.

HPW Model results

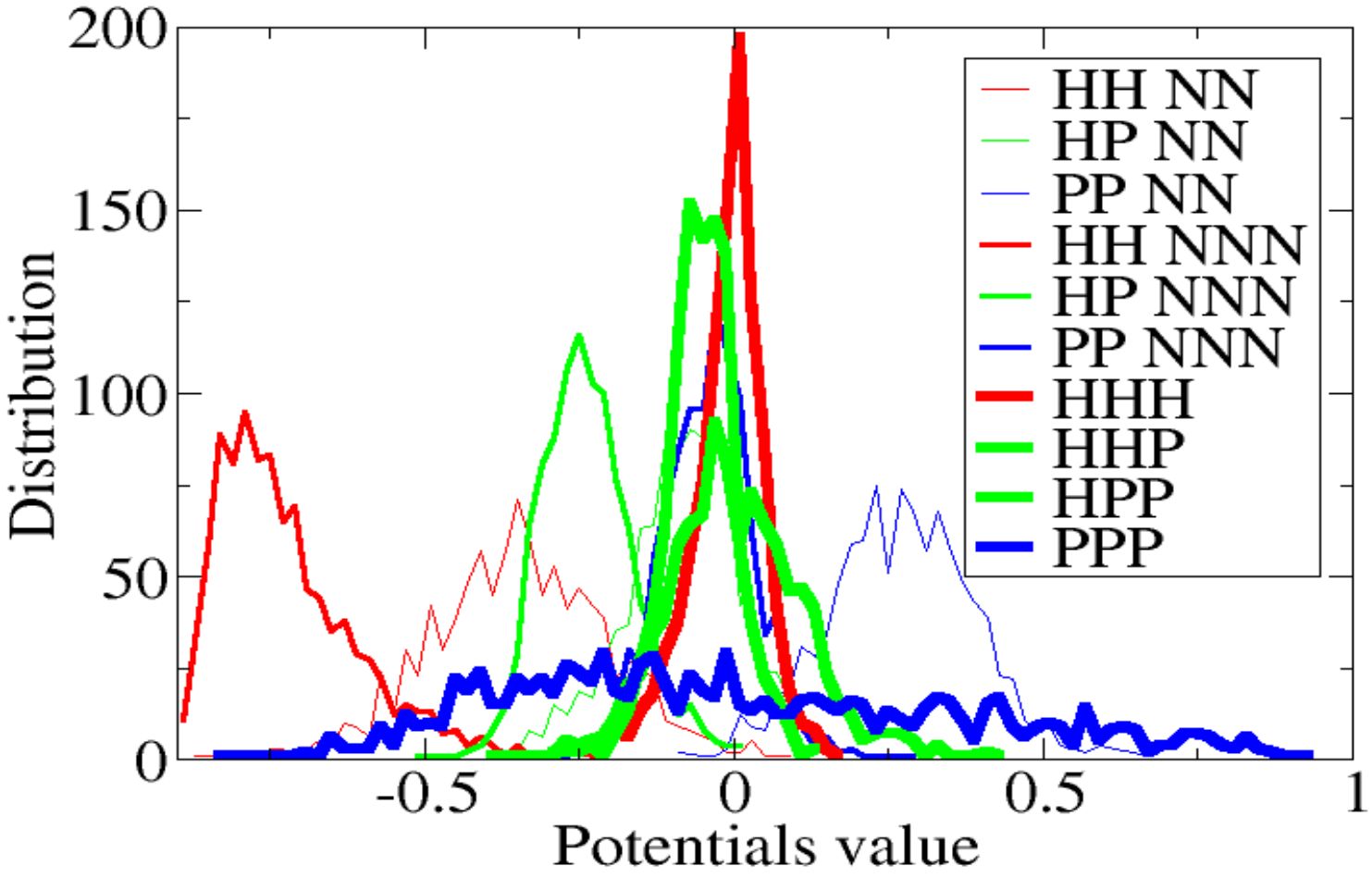


Same problem as for real proteins !

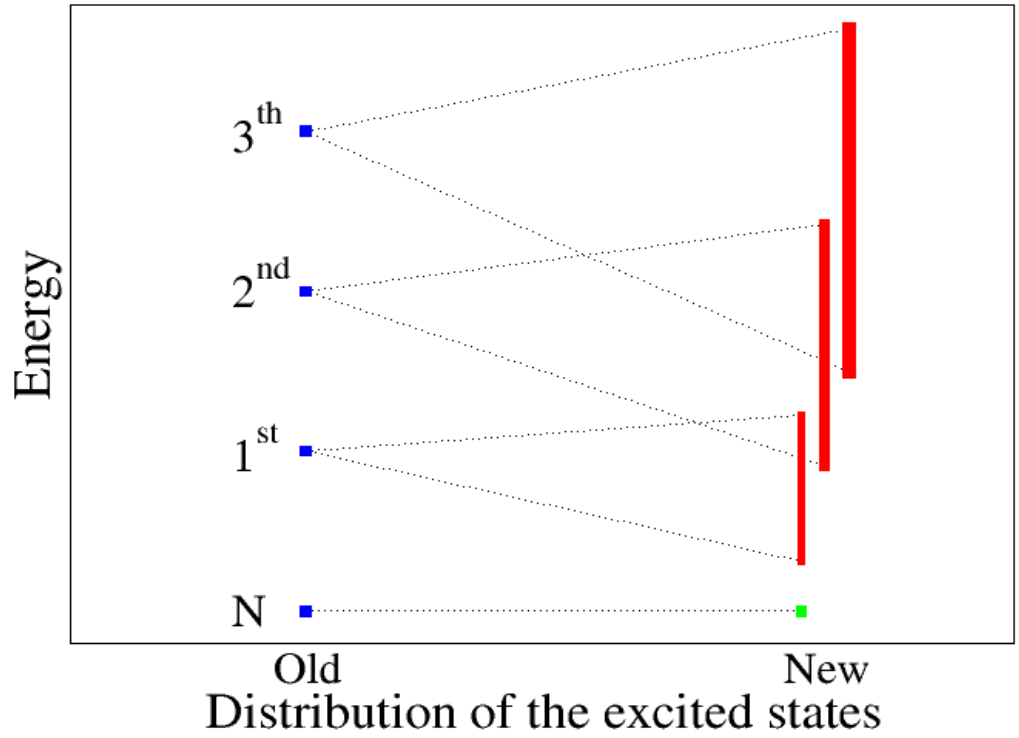
Learnable sequences



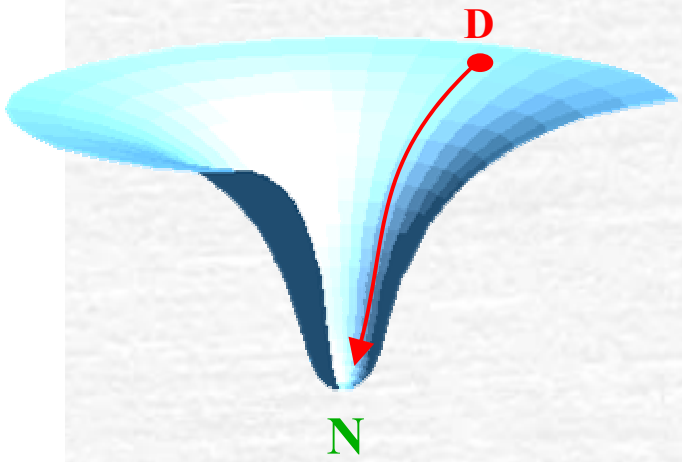
Interaction values



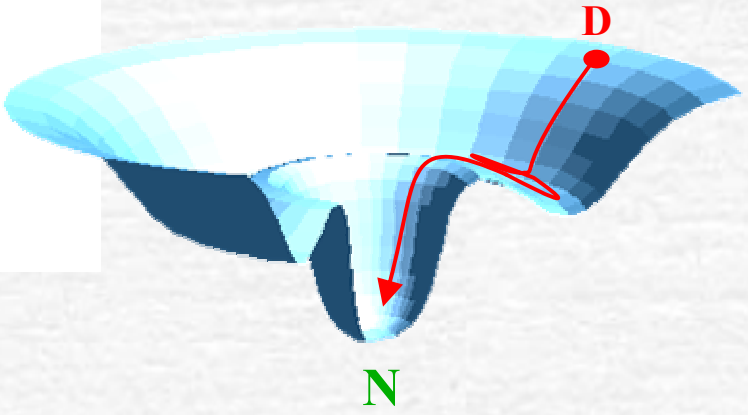
Energy landscapes



Explicit solvent



Effective interactions



Conclusions

- **The thermodynamics of proteins can be reproduced over a broad range of temperatures by means of a semi-explicit solvent (cold and warm denaturation within the same framework)**

- **The solvent effects can not be reproduced by effective residues interactions.**

- **The energy landscape changes : the dynamical folding process may change too. Some care should be taken when studying dynamics.**

Bibliography

- P. De Los Rios & G. Caldarelli, Phys. Rev. E 62, 8449 (2000)
- P. De Los Rios & G. Caldarelli, Phys. Rev. E 63, 031802 (2001).
- G. Caldarelli & P. De Los Rios, J. Biol. Phys. 27, 229 (2001).
- G. Salvi, S. Mölbert & P. De Los Rios, Phys. Rev. E 66, 061911 (2002).
- S. Mölbert & P. De Los Rios, cond-mat/0305013
  (accepted on Macromolecules)

Perceptron Main algorithm