Protein folding and disordered systems

Juan J. Ruiz-Lorenzo

Universidad de Extremadura

These lucids will be availables at:

http://chimera.roma1.infn.it/JJRL/juan_ruiz.html

Overview of the talk

1. Some relevant facts in protein folding for a (statistical) physicist.
2. Universality in Physics. Robust Properties of Phase Transitions. The Renormalization Group.
3. Disordered Systems with Frustration.
4. Random Heteropolymers. Some models.
5. Dynamics.
6. Random Energy Model and Potts Glass.
7. Researches in Spain on disordered systems.

Some relevant facts in protein folding

[see for instance Garel, Orland and Pitard]

1. Characterization of the phases:
   - Native State.
   - Denatured State: Coil State and Molten globule.
2. Times Scales:
   - Microscopic. Associated with the vibrational modes of the covalent bonds $\simeq 10^{-15}$ s.
   - Macroscopic. Times for the folding. Typically from 10^{-3} s to 1 s.
3. Structures: Primary, Secondary, Tertiary and Quaternary.
4. Interactions:
   - Bonded. Covalent bonds.
   - Unbonded: Coulomb, Van der Waals and Hydrogen bonds.
   - Solvent. Mainly water.
5. Energy Scales:
   - Bonded Interactions. From 200 kJ/mole to 600 kJ/mole (2 eV/molecule-6 eV/molecule).
   - Unbonded. From 4 kJ/mole to 5 kJ/mole (0.04-0.05 eV/molecule).

6. Typical Size:

   100 aminoacids for small proteins and 500 for long immuno-globulins.

7. The role of solvent (e.g. water).

   55 % of residues in a protein are hydrophobic. There is a 35 % of probability to find a hydrophobic residue on the surface of a protein. $\rightsquigarrow$ Large Frustration.

8. Levinthal Paradox.
   The protein during the folding does not explore all the configuration space only a small part of it $\rightsquigarrow$ Energy funnel.

   Each peptide bond has z different conformations. Hence, the dimension of the conformational space is z^N. Taking for simplicity $z = 2$ and $N = 100$: $2^{10} \simeq 10^{30}$. The minimum time to change the conformation of the peptidic bond is 10^{-15} s, hence an lower bound is $10^{30} \times 10^{-15}$ s $= 10^{15}$ s (which is 0,002 the age of the Universe) to sweep all the states of the conformational space of the protein.

Hence, it is possible to identify:

- Hard degrees of freedom. Linked to covalent bonds and the peptide bond. They are very rigid at room temperature $\left(\text{Energy} \gg k_{\text{B}}T_{\text{room}}\right)$.
- Soft degrees of freedom. Torsion angles along the backbone chain and of the side chains. $\left(\text{Energy} \simeq k_{\text{B}}T_{\text{room}}\right)$.

Dual Requirement for the folding:

- Kinetic accessibility.
- Stability.

Universality in Phase Transitions

Phase Transitions depend only on general properties of the Hamiltonian (e.g. symmetry and dimensionality). In particular critical exponents and amplitude ratios do not depend on the microscopic details of the model.

Only critical amplitudes and the critical temperature depend on the details of the model.

Hence, we have an infinite number of Hamiltonians which yield the same phase transition (i.e. all these Hamiltonians belong to the same Universality class).

All this has been understood in the framework of the Renormalization Group.

For example:

1) The phase transition liquid-vapor in water at the critical point has the same critical exponent that the Ising model (which describes a large class of ferromagnetic and antiferromagnetic materials like FeF_2, $CoCs_2Br_5$,...)

2) Universal behavior in polymer physics:

$$\langle R_g^2 \rangle \simeq A \; N^{2\nu} \left[1 + N^{-\Delta} \right] .$$

R_g is the gyration radius of the polymer, and N is the number of monomers.

$\nu \simeq 0{,}588$ is UNIVERSAL.

Its value is independent of the type of polymer, solvent and temperature (for good solvents and above the theta temperature).

This behavior also holds in Self Avoiding Walks (SAW).

We can study the SAW using a $(\phi^2)^2$ field theory with $O(N)$ symmetry in the limit $N \to 0$.

Disorder and Frustration

DISORDER. It is possible to modelize the impurities in a material introducing the concept of disorder. We do not know the Hamiltonian of the system but its probability distribution.

A magnetic material with impurities has a Hamiltonian $\mathcal{H}(J, S)$ where J describes the impurities and S describes the magnetic moments of the atoms.

There are two (limiting) kinds of disorder:

1) Annealing. The impurities are in thermal equilibrium with the spin degrees of freedom (S).

2) Quenched. The impurities are completely frozen. The characteristic time of the impurities is some orders of magnitude greater than that of the spin.

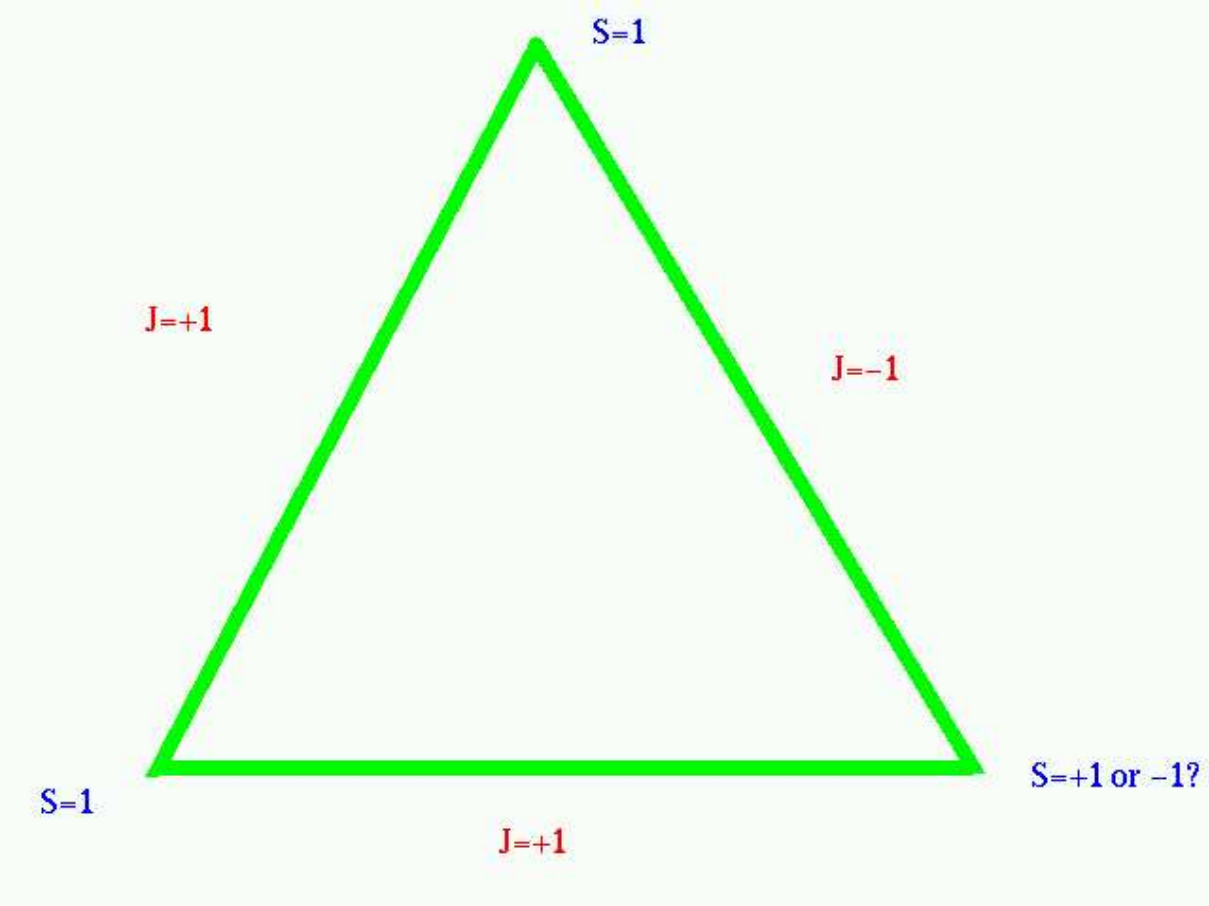
In presence of disorder, besides the Hamiltonian, we should provided the probability distribution of the disorder $p(J)$ in order to have a complete thermodynamical description of the system.

FRUSTRATION. The system cannot satisfy all the constraints (geometric and energetic) at the same time.

The Hamiltonian of an Ising spin glass is :

$$\mathcal{H} = -\sum J_{ij} S_i S_j.$$

Let us consider the following frustrated triangle:



All two configurations have the same energy!!

Random Heteropolymers and Proteins

[See for instance Iori, Marinari & Parisi]

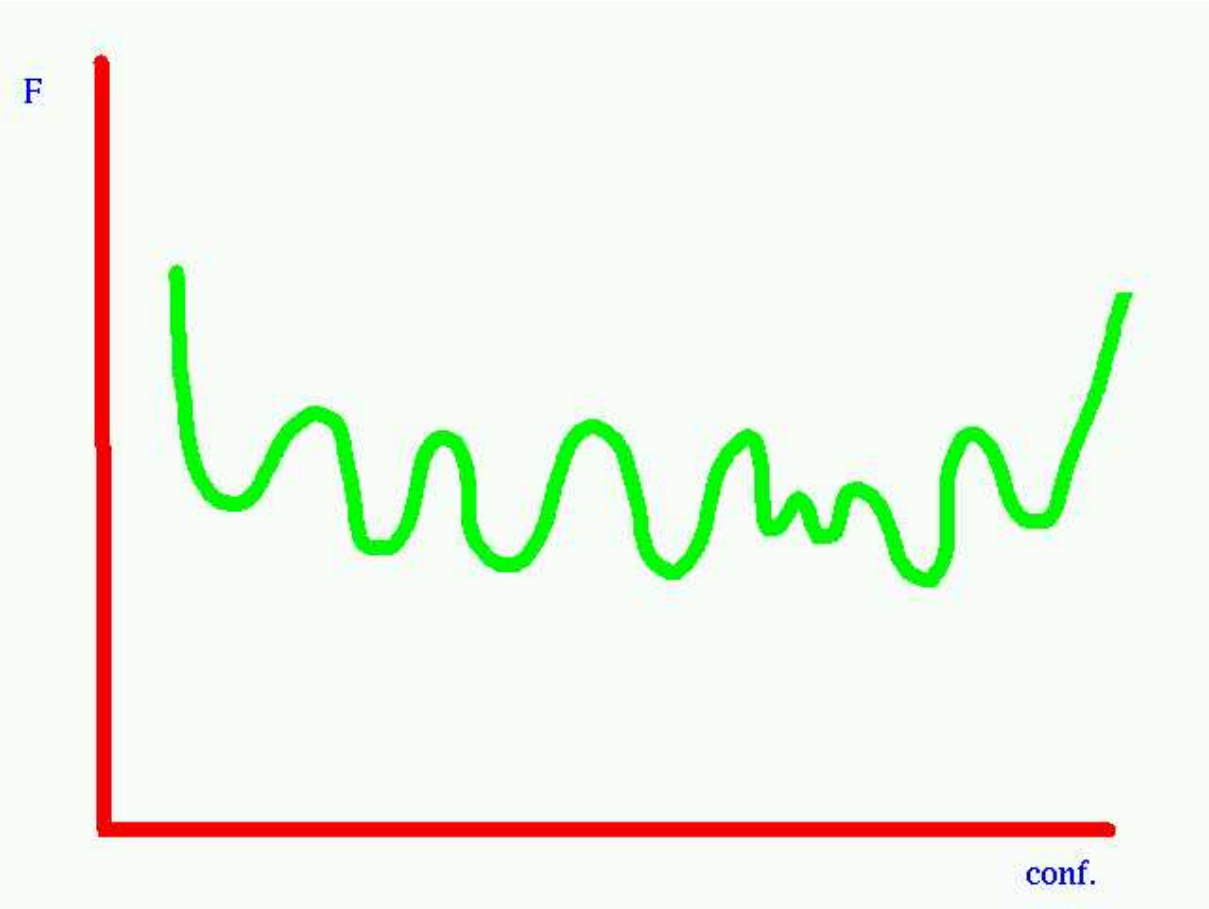
1. The proteins are not random heteropolymers.
2. The proteins are the products of natural evolution and they are NOT random sequences.

BUT

3. Which properties do the proteins share with random heteroplymers?
4. Which properties have been selected by natural evolution?

Free Energy Landscape for a random heteropolymer

A typical free energy landscape of a disordered system with frustration:



Notice the large number of absolute minima and the large number of metastable states (relative minima).

The energy barriers between two metastable states are very high.

For example, in an Ising spin glass (Parisi):

1) Infinite Number of Pure States.

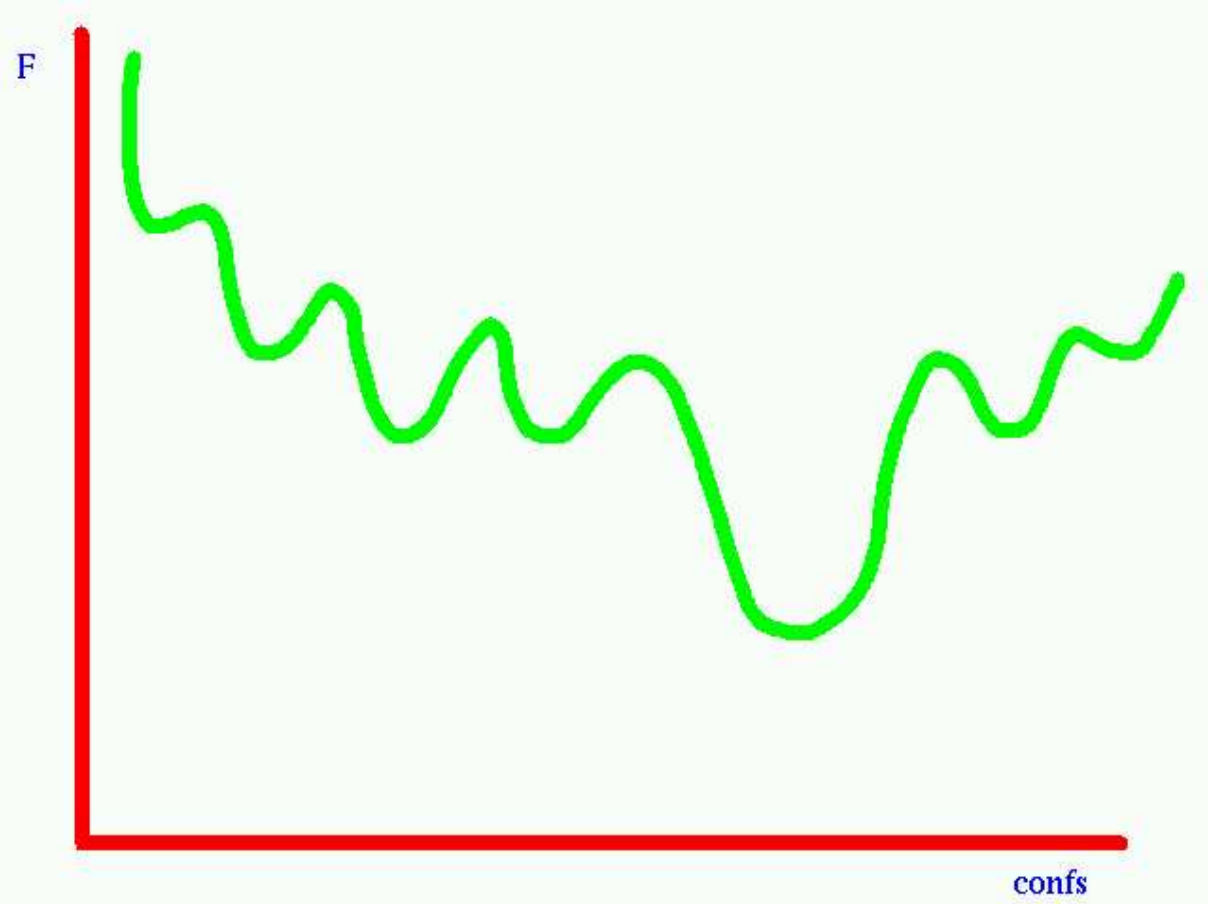
2) All these pure states are organized in an ultrametric fashion.

Continuous Breaking of the Replica Symmetry!!

Note: The statistical distribution of European dicotyledons and monocotyledons follow the same laws as the pure states (species) in Ising Spin Glass. (Epsein&Ruelle)

Free Energy Landscape for a Protein

[see Tang]



Random Heteropolymers

Can Random Heteropolymers mimic the folding of Real Proteins?

Favor.

- The folding mechanism is very robust and so, we will expect no dependence on the microscopic details of the model.
- We will obtain average properties of the folding mechanism (the robust properties of the folding).
- It is easier to solve a random heteropolymer than a given chain of amino acids.
- The disorder should be quenched to account the fixed character of the chemical sequence (the chain of aminoacids).
- Some disordered systems show similar phenomenology (in the statics as well in the dynamics) that proteins: Potts glass, REM, ...
- The analogy between real glasses (with no disorder at all) and spin glasses has been very successfully.
- The same identification works very well in Neural Networks with the Hopfield Model (a spin glass).
- The free energy landscape of disordered systems is very rough and similar (qualitatively) of that of proteins.
- ...

Against.

- Very short chains in order to define thermodynamics (in the diluted limit). No problem as the number of chains in the solvent is large.

  Possible solution: Define finite volume pure states.

- Sometimes it is interesting to study the folding of a definite sequence of aminoacids.

- Typically proteins do not exhibit glassy dynamics, whereas random heteropolymers show it.$\rightsquigarrow$ Refinement of the random heteropolymer models.

- Little feedback between biologists, chemists and physicists!!

- ...

Some Models

Garel-Orland-Leibler Model (GLO).

$$\mathcal{H} = \frac{1}{2}\sum_{i\neq j} v_{ij} + \frac{1}{6}\sum_{i\neq j\neq k} \omega_3 \delta_{ij}\delta_{ik} + \frac{1}{24}\sum_{i\neq j\neq k\neq l} \omega_4 \delta_{ij}\delta_{jk}\delta_{kl}$$

$$v_{ij} = v_0 + \beta(\lambda_i + \lambda_j)\delta_{ij} \ ,$$

v_0 being a suitable short range interaction and λ_i are independent Gaussian variables (Quenched disorder):

$$P(\lambda_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\lambda_i - \lambda_0)^2}{2\sigma^2}\right] \ .$$

If $\lambda_0 > 0$ the majority of the residues are hydrophilic.

The model displays an interesting phase diagram when $\lambda_0 < 0$.

Iori-Marinari-Parisi Model (IMP).

$$\mathcal{H} = \sum_{1\leq i<j\leq N} \left(\delta_{i+1,j}\ r_{ij}^2 - \frac{A}{r_{ij}^6} + \frac{R}{r_{ij}^{12}} + \frac{\epsilon\eta_{ij}}{r_{ij}^6}\right) \ .$$

$$\langle \eta_{ij} \rangle = 0 \ , \ \langle \eta_{i,j}\eta_{k,l} \rangle = \delta_{(i,j),(k,l)} \ .$$

η_{ij} quenched disorder.

If $\epsilon = 0$, the Hamiltonian describes a homopolymer and the model shows the usual coil-gobule transition.

If $\epsilon > 0$ the model shows two different phase transition: 1) coil-globule and 2) globule-folded.

Lattice Models

The monomers ( ) live on a lattice (e.g. cubic lattice):

$$\mathcal{H} = \sum_{i<j} \delta(|r_i - r_j| - a) B_{ij} \ .$$

B_{ij} is the contact interaction between the monomers i and j. a is taken to be the distance between the α-carbons in the polypeptide chain: $a \simeq 3{,}8$ Å.

One can choose for B_{ij}:

1. In order to mimic the interactions between aminoacids in Nature.
2. Random (e.g. Gaussian).

Simplifications:

- Go model.
- HP model. Only two kinds of aminoacids: Hydrophobic (H) and Polar (P).

Again, the model shows two different phase transition:

1. Collapse Transition: T_θ.
2. Folding Transition: T_F.

Relation between the foldability of a protein and the ratio:

$$\sigma_T = \frac{T_\theta - T_F}{T_\theta}$$

The time to fold grows with σ_T.

Dynamics

[See Garel, Orland & Pitard]

Phenomenological approach

The starting point is the Master equation:

$$\frac{dP_\alpha}{dt} = \sum_\gamma \left[W_{\alpha\gamma} P_\gamma(t) - W_{\gamma\alpha} P_\alpha(t) \right] \ .$$

$P_\alpha(t)$ is the probability that the state α being occupied at the time t. $W_{\alpha\gamma}$ is the transition probability from the state γ to the state α. To reach the equilibrium it is sufficient to have the detailed balance relation:

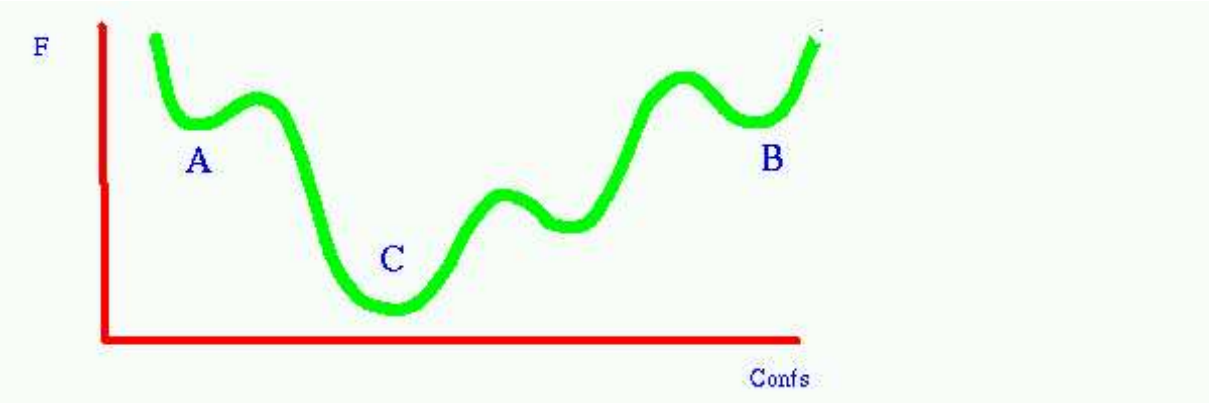
$$\frac{W_{\alpha\gamma}}{W_{\gamma\alpha}} = e^{-\beta(E_\alpha - E_\gamma)}$$

Different choices of $W_{\alpha\gamma}$ yield different phenomenological models.

For instance, an Arrhenius law:

$$W_{AB} = W_0 e^{-\beta(E_A - E_C)}$$

This transition probability depends only on the final state of the transition.



This model provides stretched exponential behavior.

REM and Potts Glasses

We can generalize the Ising spin glass (two body interactions) to a p-body interactions, the p-spin model:

$$\mathcal{H}_J(S) = - \sum_{i_1 \leq i_2 \leq \cdots \leq i_p \leq N} J_{i_1 i_2 \cdots i_p} S_{i_1} S_{i_2} \cdots S_{i_p} \ .$$

The quenched variables $J_{i_1 i_2 \cdots i_p}$ are Gaussian distributed with zero mean and variance J^2.

We can compute the probability distribution of the energy

$$P(E) = \overline{\langle \delta(E - \mathcal{H}_J(S)) \rangle_J}$$

In the limit $N \to \infty$ and then $p \to \infty$, $P(E)$ follows the Gaussian distribution.

Besides, if we introduce a second copy of the system, we can compute the probability that the first copy have energy E_1 and the second one E_2:

$$P(E_1, E_2) = P(E_1) P(E_2).$$

In general, if we have M copies of the system (all with the same disorder), we obtain that all the energy are uncorrelated:

$$P(E_1, E_2, \cdots, E_M) = \prod_{i=1}^{M} P(E_i).$$

We have obtain that in the limit $p \to \infty$ the p-spin model is equivalent to the Random Energy Model (REM) which consists in 2^N levels with independent random energies (distributed following a Gaussian distribution).

The REM has been relevant in this study of protein folding in the last decades. REM was proposed as a caricature of the protein dynamics on phenomenological grounds [Bryngelson & Wolynes].

The energies in the GLO model (with $\omega_3 = 0$) also follows the Gaussian law of the REM.

The REM shows a phase transition:

1) If $T > T_c$, all the 2^N levels are distributed with $E \geq -E_0$ [E_0 is the minimum energy of the system].

2) If $T < T_c$ all the 2^N levels have energy $E = -E_0$. The entropy in this state is zero.

The p-spin has been solved using replicas [Gross & Mezard] finding that the solution only need one step of replica symmetry breaking (RSB).

The Potts glass with $q > 2$ states, with Hamiltonian:

$$\mathcal{H}_J = -\sum_{i,j} J_{ij}\ \delta(S_i, S_j)$$

where $S_i = 0, \cdots, q-1$.

was solved using replicas [Gross & Sompolinsky] finding again an one-step RSB solution.

Properties of the one-step solution:

1) There are an infinite number of pure states (phases).

2) These infinite number of states are maximally different (the overlap between two different states is zero).

Researches in Spain on disordered systems

1. Spin Glasses in finite dimensions.
2. Neural Networks.
3. Diluted systems.
4. Random Field models.
5. Relation between real glasses and spin glasses.
6. Dynamics with quenched disorder.
7. Growth of surfaces on disordered substratums.
8. Disordered models in condensed matter