Phase Transition in Disordered RNA Model

Andrea Pagnani

(Roma2 “Tor Vergata”, Italy)

A simple model of binary bases RNA in which only planary links are allowed, is introduced. Using a very simple recursion rule on the partition function, the thermodynamic of the model is investigated. A spin-glass like order parameter is also introduced. Relevant hints of replica simmetry breaking (RSB) are found in the cold phase of the model.

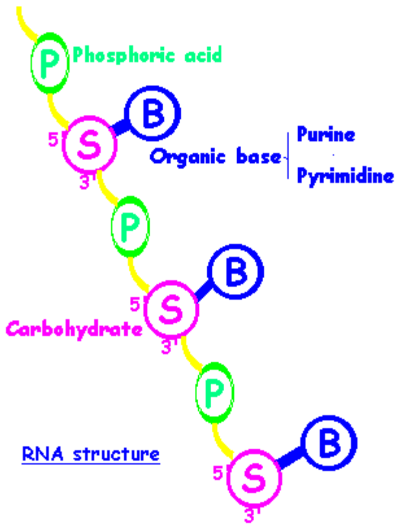
In collaboration with:

Giorgio Parisi, University of Roma 1

&

Federico Ricci-Tersenghi, ICTP Trieste

The Components of RNA



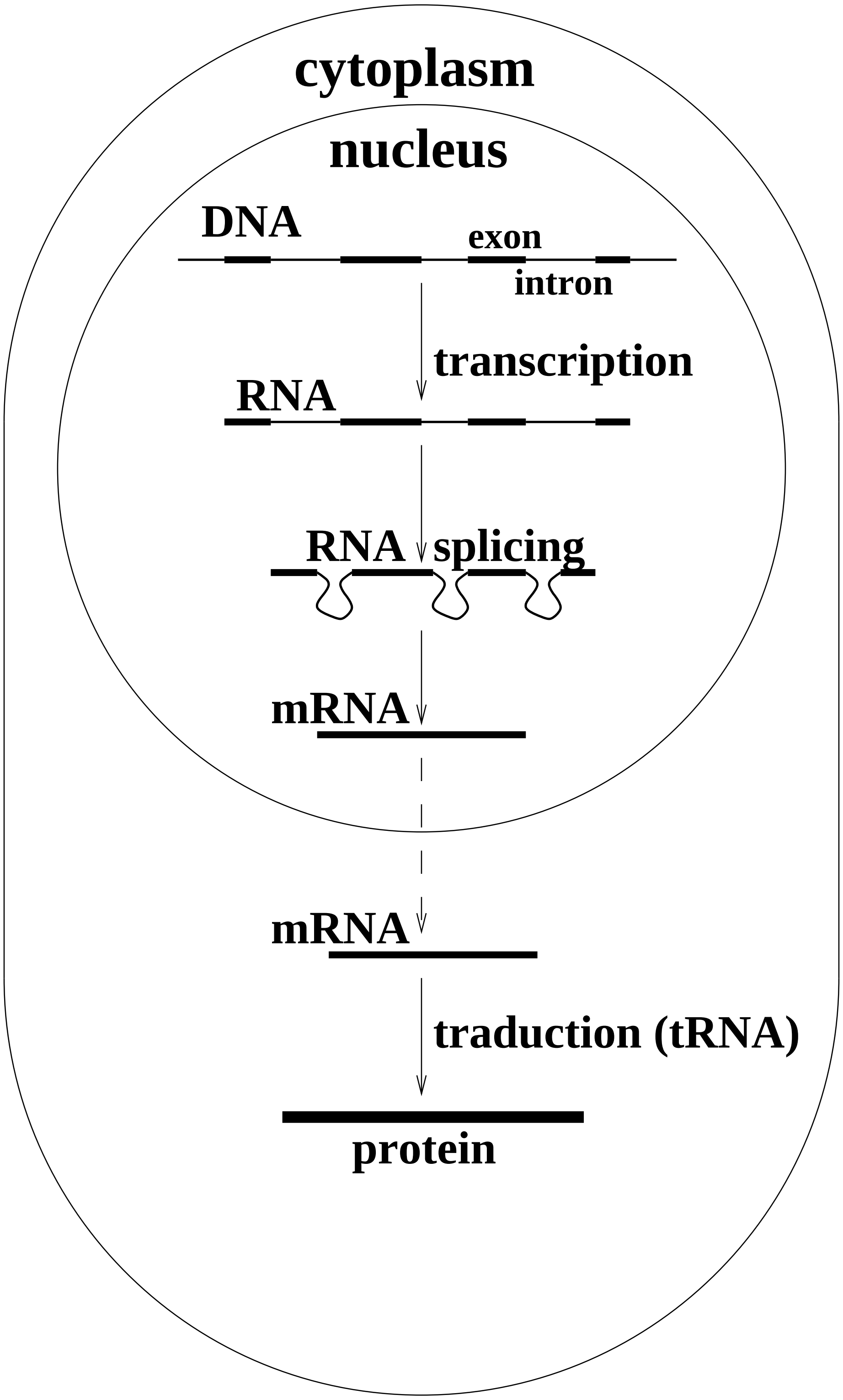
1. Ribose a pentose, a sugar with 5 carbons.
2. Phosphoric Acid
3. Nitrogenous bases: Purines (Andenine and Guanine) and Pyrimidines (Cytosine and Uracil)

The combination of abase plus a pentose constitutes a nucleoside.

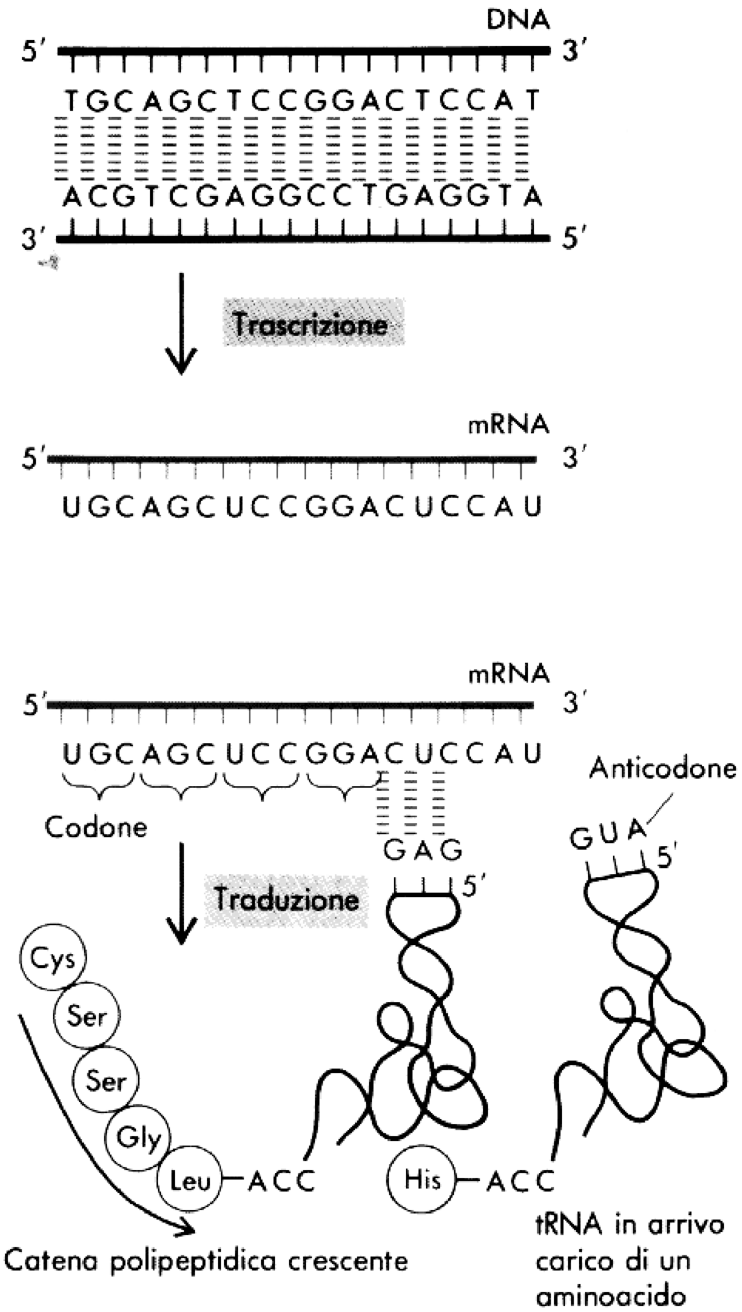
Adenine = Base, Adenosine = adenine + ribose , Adenosine monophosphate = nucleotide

nucleotide = Ribose + Base + Phosphoric Acid

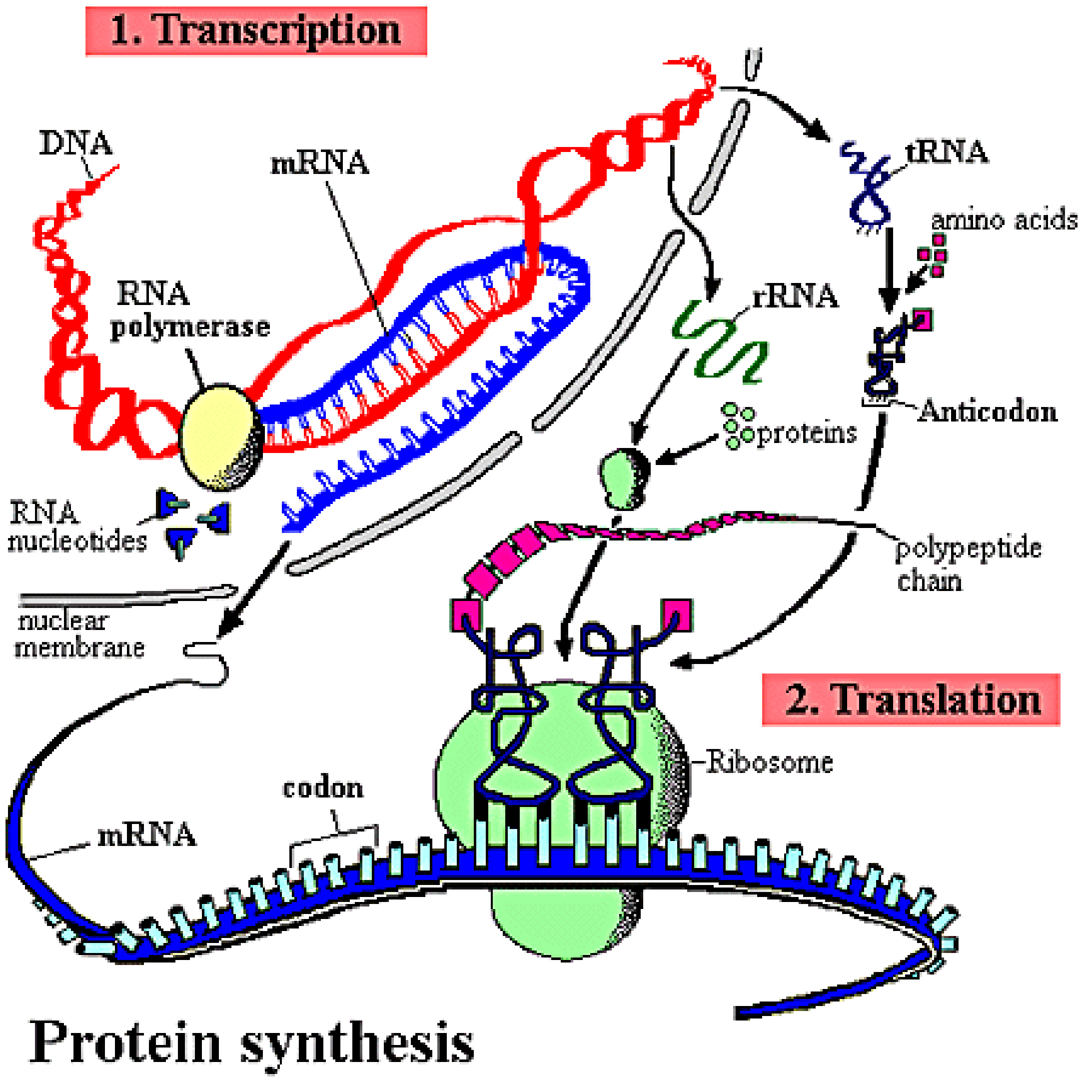
The Functional Cicle of RNA



The Transcription of RNA



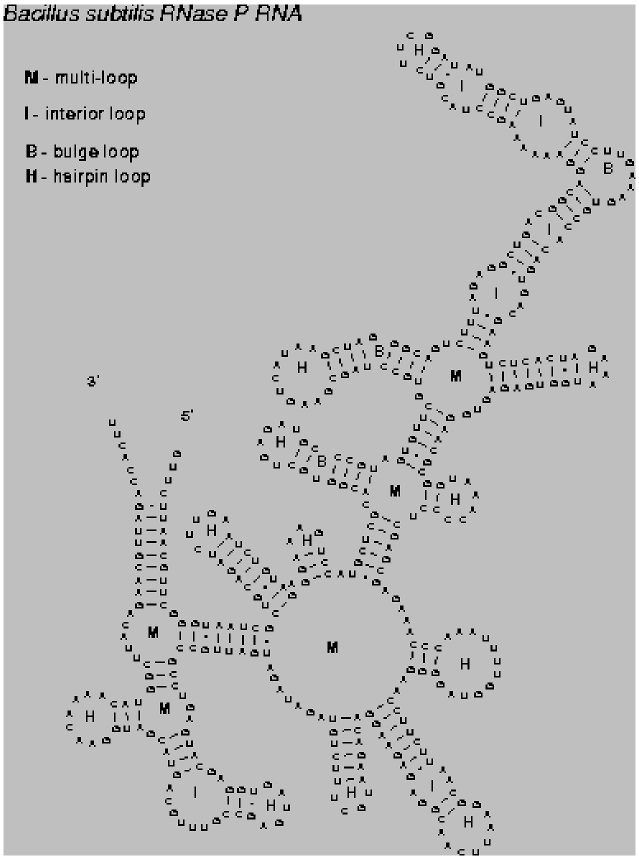
The Ribosomal Cicle

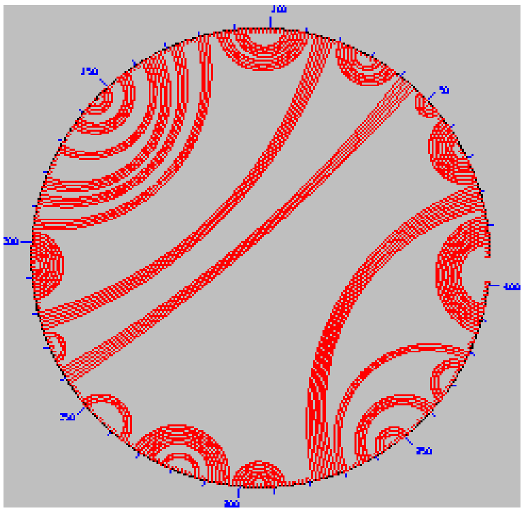


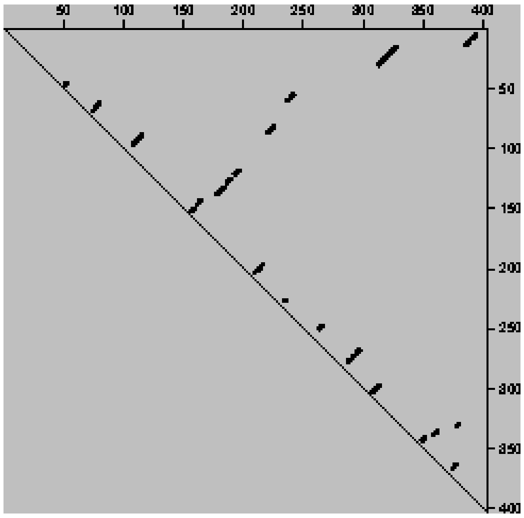
Different kind of RNA

1. **mRNA** $\simeq 1000, 2000$ bases long. It varies a lot and can arrive to 6800 bases. Is the carrier of genetic information on the primary structure of proteins from DNA, along with special features that allow it to attach to ribosomes and function in protein synthesis. Its size depends on the size of the protein for which it codes. It tends to be relatively short-lived, and its lifetime varies from molecular species to molecular species (depending on the biological role of the protein).
2. **tRNA** $\simeq 70, 100$ bases long. It is designed to carry activated amino acids to the site of protein synthesis, the ribosome. It is long-lived (stable).
3. **rRNA** $\simeq 120, 4700$ bases long. It forms the ribosome, the site of protein synthesis, and one rRNA is the catalyst for formation of the peptide bond (Science, June 5, 1992). Various species range in size from 4700 bases to about 120 bases. Eukaryotic and prokaryotic rRNAs are distinctly different. rRNA is also long-lived (stable).

Quasi-Planarity of Structures







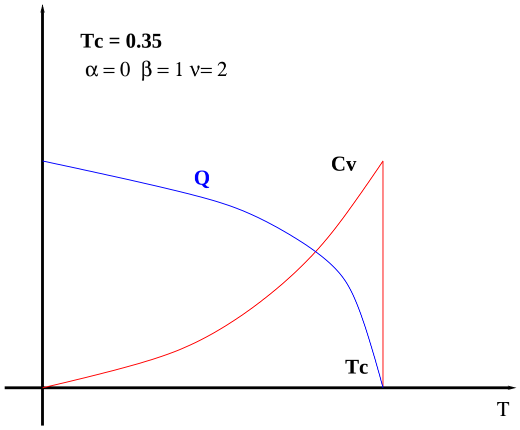
Realistic Models:

- Vienna package (Zuker, Fonatana, Hofacker, ...)

http://www.tbi.univie.ac.at/~ivo/RNA/

Simplified Models

- Bundschuh & Hwa cond-mat/9903089



Analitic solution of a model with unique native state and short distance link allowed.

- Higgs:

1. J. Phys. I (France) **3** , 43 (1993).
2. Phys, Rev. Lett. **76** , 704 (1996).
3. J. Phys A **31** , 3153 (1998).

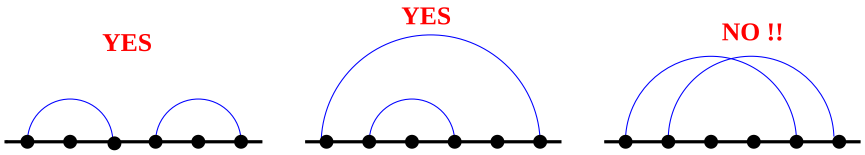
1. The thermodynamical weight of a structure deends on a few parameters (Kcal/mole): G-C= -4.8 A-U=-1.2 G-U=-0.3 .
2. 3. Only WC pairs allowed: (Kcal/mole): G-C=-1 A-U=-1.

The model

- Random sequence of 2 symbols: A,B.
- 2 kinds of base parings:

1. STRONG: A-B B-A $\rightarrow e_S$
2. WEAK: A-A B-B $\rightarrow e_W$

- Planarity condition:

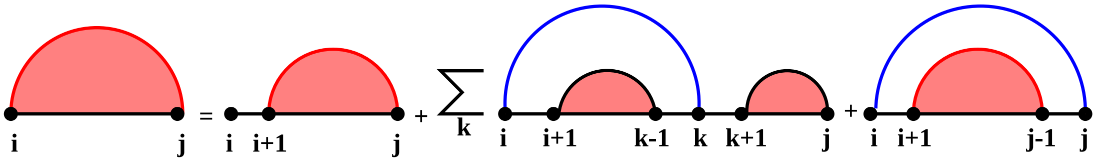


- NO links at distance 1.
- $H = \sum_{i,j=1}^{L} e_{i,j} l_{i,j}$
- Quenched Desorder:$e_{ij} = \{e_S, e_W\}$
- Topological Constraints $\rightarrow$ (Frustration):

$$l_{ij} = \begin{cases} 1 & \text{if } \mathrm{i}-\mathrm{j} \text{ are linked} \\ 0 & \text{otherwise} \end{cases}$$

Complex landscape in configurational space

How to calculate the partition function

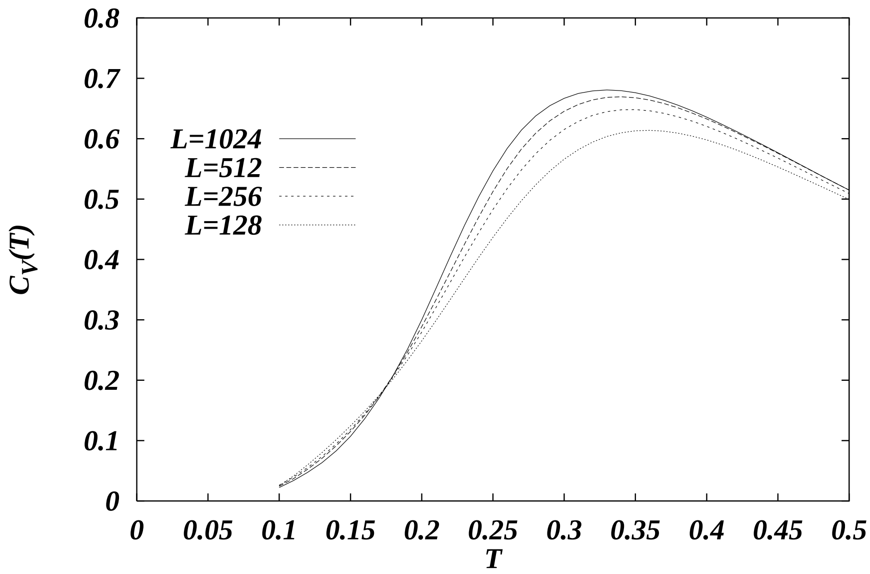


$$
\begin{aligned}
Z_{i,j} &= Z_{i+1,j} + \sum_{k=i+2}^{j-1} Z_{i+1,k-1} e^{-\beta e_{i,k}} Z_{k+1,j} \\
&+ e^{-\beta e_{i,j}} Z_{i+1,j-1}
\end{aligned}
$$

- This algorithm is very effective: The calculation time of $Z_{1,L}$ scales as L^3.
- In order to reach low values of T we need to rescale the variables according to the following rule:

$$
{Z'}_{i,j} = Z_{i,j} e^{\beta e_0 |i-j|} \text{ with } e_0 = \frac{E(T=0)}{L}
$$

Thermodynamics



It is very easy to implement a similar recursion for calculating:

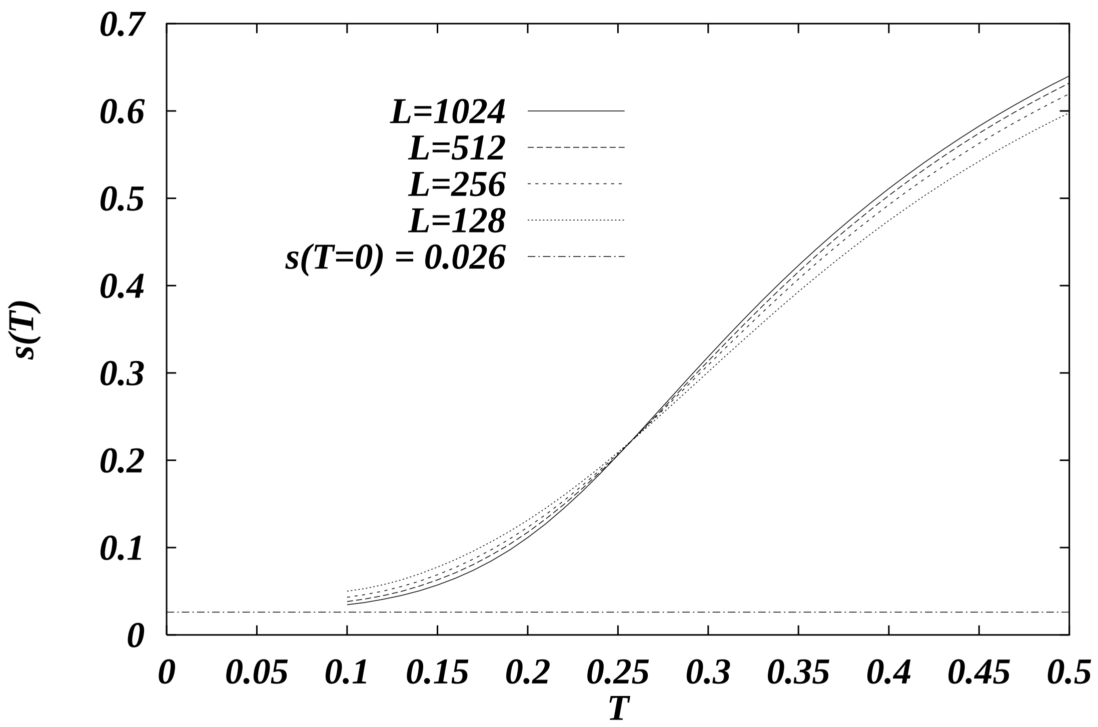
- the internal equilibrium energy $U = \langle E[\mathcal{S}] \rangle$.
- the second moment $U^{(2)} = \langle E^2[\mathcal{S}] \rangle$.

We can than calculate:

$$C_V(T) = \frac{1}{L\ T^2} \left[\frac{U^{(2)}_{1,L}}{Z_{1,L}} - \left(\frac{U_{1,L}}{Z_{1,L}} \right)^2 \right]$$

$[\cdot] \equiv$ QUENCHED AVERAGES

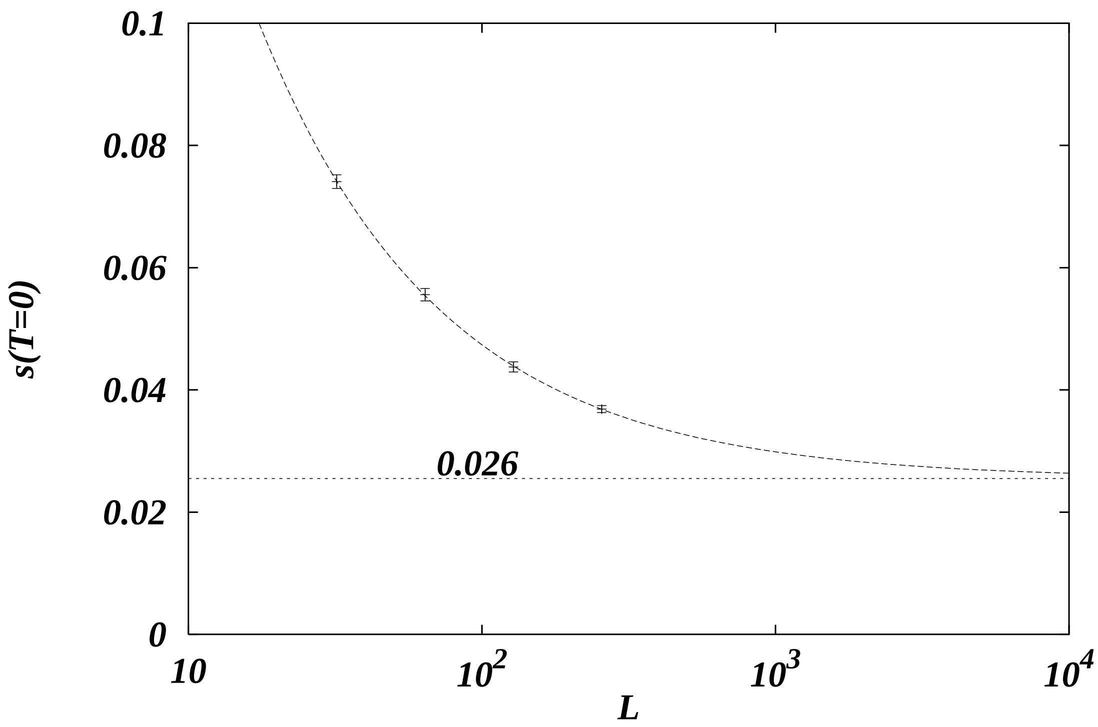
ENTROPY



- Intersection around $T \simeq 0.27 \rightarrow$ shrinking of accessible phase space.
- Curves seems to extrapolate to a finite value $T \rightarrow 0$.

EXPONENTIAL DEGENRACY OF GROUND STATES

$T = 0$ ENTROPY



In order to correctly extrapolate the $T = 0$ entropy we implemented a quite sofisticated backtracking algorithm.

- The number of GSS (*i.e.* the degeneracy) strongly depends on the sequence. We have found sequences with degeneracies ranging from 1 to $\mathcal{O}(10^7)$.
- how these GSS are organized ?

2 DIFFERENT SCENARIO

1. GSS are all very similar (ordered or "ferromagnetic" behavior).
2. GSS are sparse over the whole configurationel space.

We need a notion of distance in order to discern which is the actual scenario.

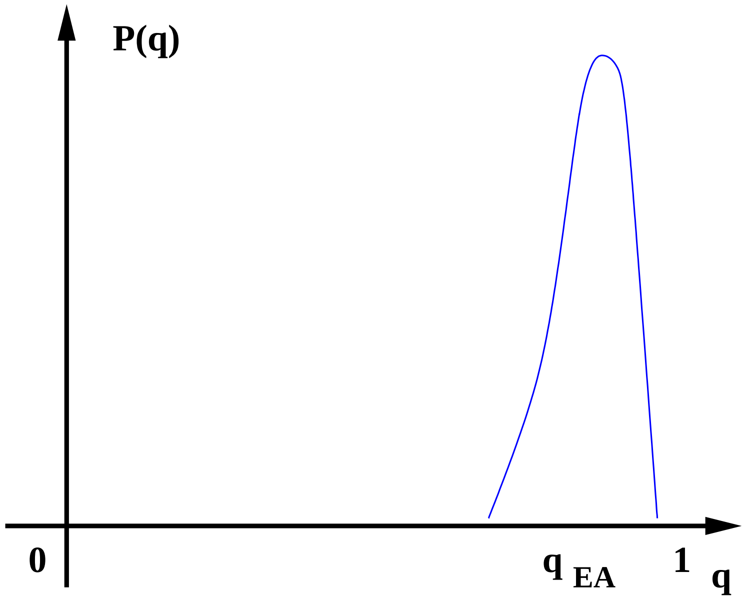
The Overlap ...

- In spin models $q = \frac{1}{N} \sum_i s_i {s'}_i \in [0, 1]$
- In our model : $q \equiv \frac{1}{L} \sum_{i,j} l_{i,j} {l'}_{i,j}$

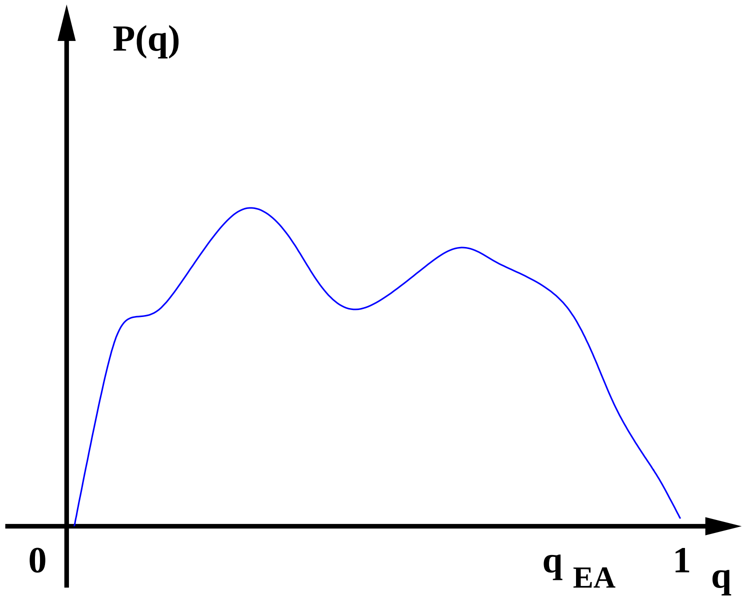
... and its distribution

1. $T = 0 \quad P(q) = \left[\sum_{\alpha,\beta} \delta(q - q^{\alpha\beta})_J \right]$
2. $T \neq 0 \quad P(q) = \left[\sum_{\alpha,\beta} w_\alpha w_\beta \delta(q - q^{\alpha\beta})_J \right]$
   where w_α is the Boltzmann weight factor.

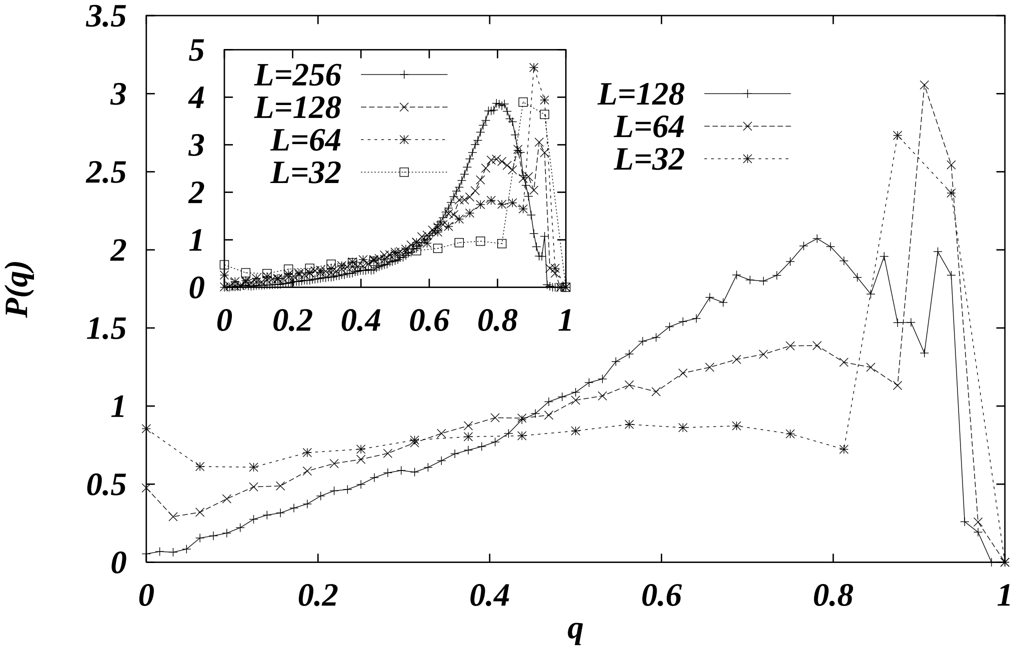
Few similar GSS



Lot of different GSS



Phase Space Structure at $T = 0$



The figure suggest :

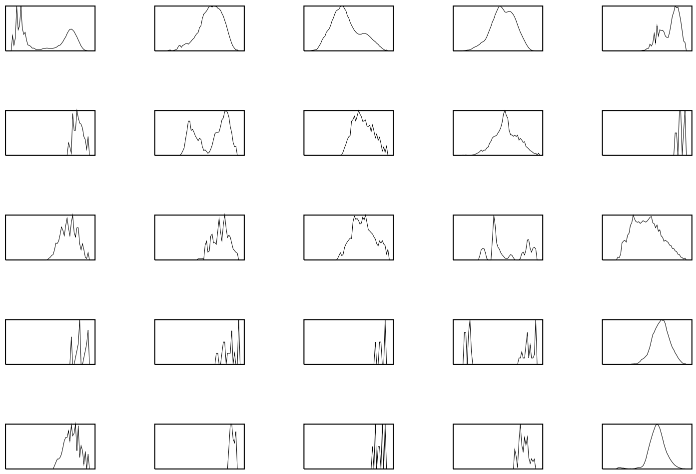
- a coexistence of lots of different GSS
- a structured peak around $q \equiv q_{EA} = 0.8$.
- a more sparse region for $q < q_{EA}$

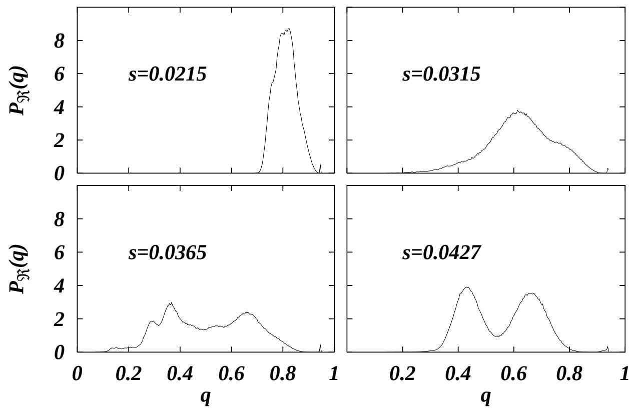
Will all those features survive in the thermodynamical limit ?

A possible answer is:

Who cares: real RNA is finite (mesoscopic)

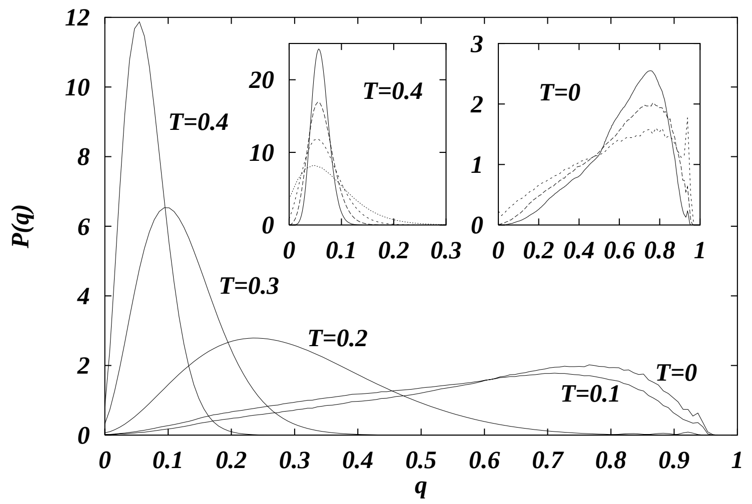
Sample dependent $P_J(q)$.





The whole zoology presented suggest a RSB breaking scenario that maybe collapse in the thermodinamical limit.

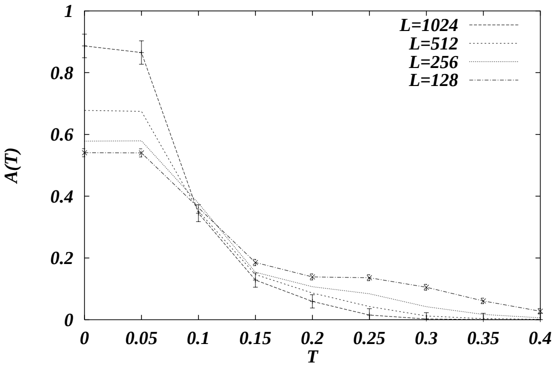
Phase Space at $T \neq 0$

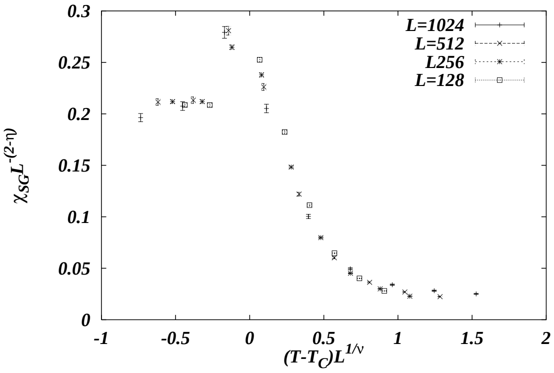


We have calculated the $P(q)$ at different temperatures ($T \in [0, 0.4]$). While at $T = 0$ we were able to sum over the whole sets og GSS, at a finite temperature we performed a Monte Carlo sampling.

- $P(q)$ changes abruptly from a narrow peak in the low-q region to a broader one which extends over almost the whole support.
- for $T > 0.15$ size dipendence show a delta (the width of the distribution goes to zero as $\Delta q \propto L^{-1/2}$).
- for $T < 0.15$ we see the persistence of the broad shape feature.

Confirming the RSB





$$
\begin{aligned}
\chi^J_{\text{SG}} &= L[\langle q^2\rangle_J] - \langle q\rangle_J^2 \\
A &= \frac{[(\chi^J_{\text{SG}})^2] - [\chi^J_{\text{SG}}]^2}{[\chi^J_{\text{SG}}]^2}
\end{aligned}
$$

- The crossing point of different curves in signals the existence of a low-temperature spin glass phase, where the $P_{\mathcal{J}}(q)$ become non-self-averaging.
- The critical temperature of this transition seems to be located between $T = 0.1$ and $T = 0.15$.
- We have determined the best estimates for T_c and for the critical exponent η requiring the best collapse for the $\chi_{\text{SG}} \equiv [\chi^{\mathcal{J}}_{\text{SG}}]$ data, scaled assuming the usual finite-size formula $\chi_{\text{SG}} = L^{2-\eta} f(L^{-1/\nu}(T - T_c))$

Conclusions and Perspectives

We have for the first time indicated a quantitative description of the phase transition of a disordered model of RNA.

- At our knowledge this is the first model with a complex behavior and a replica symmetry breaking transition, which is, on the other hand, not computationally complex (the computation time grows as L^3).
- We have also found that all GSS share in common around a 25% of common links structures . We define this set the back-bone.
- The presence, in the low temperature phase, of persistent links (backbone) suggests that the transition is driven by a collective behavior of mainly the bases belonging to the backbone (the so-called "hot sites" in the protein folding language).
- Because of the disorder, the backbone changes from sequence to sequence, it is in Spin Glass jargon, a non-self-averaging quantity. It is responsible for the fluctuating part of the $P_J(q)$.
- Despite of their name, it would be intersting to investigate wether from a dynamical point of view the "hot links" are really cold, and if can be found a true time scale separation among cold and hot links.