UNIVERSITÀ DEGLI STUDI DI PARMA

Dottorato di ricerca in Fisica
Ciclo XXVI

Statistical mechanics models for biological systems: cooperativity in biochemistry and affinity maturation of antibodies

Coordinatore:
Prof. Pier Paolo Lottici

Tutor:
Prof. ssa Raffaella Burioni

Dottorando:
Guido Uguzzoni

Contents

Introduction

The interplay between Physics and Biology has a long history of successes and has been the driven force in the evolution of our quantitative understanding in life sciences. Statistical Mechanics provides useful tools and concepts to deal with collective behavior of many strongly interacting agents. Overlooking the detailed and the specific description of the interactions to focus on the very key features allows to ask different questions concerning the global systemic properties of biological systems. The information processing and statistical inference approach has became more urgent in the last decades due to the large amount of data coming from the exploit of different new experimental techniques. Concepts such as entropy, phase transition and criticality has entered the unavoidable terminology to describe the nature of biological systems at very different level of complexity: from the animal collective behaviour [1],the physiological apparatuses as nervous system [2] and immune system [3] to the biochemical processes in cells [4]. The studies presented in this thesis are placed in this interdisciplinary border context.

The thesis is divided in three main parts. The first is devoted to the more formal aspect of statistical mechanics models of spin systems. We review briefly, in the first chapter, three milestone models of spin systems: the Curie-Weiss, the Sherrington-Kirkpatrick and the Hopfield model. These models constitute the paradigmatic examples of mean-field Statistical Mechanics and will constitute the ground for the studies in biochemical kinetics and immunology presented in the following parts. In the second chapter we report a detailed study of a generalization of the Hopfield model with diluted and correlated patterns [5]. We investigate the topology of the emergent interactions network. We find an exact expression of the coupling distribution that allows to distinguish different regimes varying the dilution parameter. Moreover we study the thermodynamic properties of the model, obtaining explicitly the replica symmetric free-energy coupled with its self-consistence equations. Considering the small overlap expansion of these self consistencies equations we get the critical surface dividing the ergodic phase to the spin-glass one.

The second part of the thesis focus on the investigation of the cooperative behavior in biochemical kinetics through mean field statistical mechanics [6]. Cooperativity is one of the most important properties of molecular interactions in biological systems as it is often invoked to account for collective features in binding phenomena. It constitutes a fundamental tool that nature developed to modulate the chemical response of biological systems to varying stimuli. Statistical mechanics offers a valuable approach as, from its first principles, it aims to figure out collective phenomena, allowing a unified and broader theory for complex chemical kinetics. In this way different cooperative behaviors, described by the related binding curves, can be analysed in an unified framework. We compare the theoretical curves predicted by the model with experimental data found in literature, finding an overall good agreement and extrapolating the values of the effective interactions between the binding sites, which can be put in direct correspondence with the standard coefficient that measure cooperativity (Hill number). Moreover, an extension of

the model allows to take into account heterogeneity that can affect both the couplings between the multiple active sites (allosteric regulation) and the chemical potentials in the binding of the ligands.

The last part is dedicated to a statistical inference analysis on deep sequencing data of an antibodies repertoire with the purpose of studying the process of antibodies affinity maturation [7]. A partial antibodies repertoire from a HIV-1 infected donor presenting broadly neutralizing serum is used to infer a probability distribution in the space of sequences . The idea is to use the model to study the structure of the affinity with an antigen as a function of the antibody sequence. We test this strategy using neutralization power measurements and the deposited crystallographic structure of a deeply matured antibody. The work is still in progress, but preliminary results are encouraging and are presented here.

The original part of the thesis is based on the following works:

- Elena Agliari, Lorenzo Asti, Adriano Barra, Raffaella Burioni and Guido Uguzzoni *Analogue neural networks on correlated random graphs*, J.Phys.A: Math. Theor. 45 365001 (2012)
- Elena Agliari, Adriano Barra, Raffaella Burioni, Aldo Di Biasio and Guido Uguzzoni *Collective behaviours: from biochemical kinetics to electronic circuits*, Scientific Reports 3 3458 (2013)
- Lorenzo Asti, Paolo Marcatili, Andrea Pagnani, and Guido Uguzzoni. *Multivariate Gaussian Modelling for Abs affinity maturation.* In preparation.

Part I

Mean field statistical mechanics models

Chapter 1

Generalities of some classical models of spin systems

In this chapter we give a brief introduction to the equilibrium statistical mechanics of some classical spin-systems models that constituted the ground floor for the original studies described in the next parts of the thesis. In particular we'll introduce the Curie Weiss model, the mean-field implementation of the Ising model, that will be used in the second part for the description of cooperative behavior in biochemistry. Further, the Sherrington-Kirkpatrick model, that is the paradigmatic example of disordered systems, allow to introduce the main concepts that help the comprehension of the following parts of the thesis. In particular, the disorder systems framework are used in the study of the heterogeneous couplings extension of the model in the second part (section 4.4) and for a generalization of the Hopfield model in the second chapter. The Hopfield model will be treated in the last section of this introductory chapter. Our goal here is not an exhaustive presentation of the subjects. Since there is an huge amount of studies and applications on each single topic, we are going to illustrate basic concepts and results necessary to understand the following sections. We refer for a review to standard textbooks [8, 9, 10].

The main aim of statistical mechanics is to compute the macroscopic properties of a systems in thermal equilibrium, starting from the microscopic interaction law between its component.
Introducing a microscopic law can be done picking from among all degrees of freedom of the system a given number of variables that are relevant to the considered problem. Ideally the effective Hamiltonian is obtained from the original one through integration over the configuration (C) at fixed value of the variables chosen (ϕ):

$$\exp[-\beta H_{\beta}^{eff}(\phi)] = \int d[C]\delta(\phi(C) - \phi)\exp[-\beta H(C)] \tag{1.1}$$

What happeneds in most of the cases is that the integration is hardly done and an effective Hamiltonian is postulated , depending on a few parameters that can be fitted from the experiments or computed in a microscopic way. The art of model building is subtle, the model must be simple enough to be investigated and rich enough to reproduce the essential properties of the systems under studies.

Once the model is defined by writing down the Hamiltonian, you assume that the equilibrium properties of the systems are described by the averages taking over a proper probability distribution of the microstates, $\rho(\sigma)$. The equilibrium internal energy can be

identified with the ensemble average of the energy $H(\sigma)$ over this distribution:

$$U = \langle H \rangle = \sum_{\sigma} \rho(\sigma) H(\sigma), \tag{1.2}$$

while the entropy associated to the probability distribution is given by

$$S = -\sum_{\sigma} \rho(\sigma) \log \rho(\sigma). \tag{1.3}$$

and loosely speaking can thought as the amount of information I which would be needed to specify the micro-state of the body to a receiver who knows its probability distribution [80].

It is reasonable to assume that the probability distribution describing the system at equilibrium is the distribution that minimize the information I (maximize S) taking fixed $U = \langle H \rangle_\rho$ and the normalization condition. So to choose the distribution that *a priori* constrain less the model. Following the old principle that can be stated as "the best explanation of the facts is the shortest". Using the method of Lagrange multipliers, the Boltzmann distribution is easily obtained:

$$\rho(\sigma) = \exp\left[\beta H(\sigma)\right] / Z, \tag{1.4}$$

where Z is called partition function and is fixed by the normalization condition,

$$Z = \sum_{\{\sigma\}} \exp\left[\beta H(\sigma)\right], \tag{1.5}$$

and β is the Lagrange multipliers fixed by the condition $U = \langle H \rangle_\rho$. The link with thermodynamics is completed by identifying $\beta = 1/T$ with the inverse of the absolute temperature.
From this, it follows that the Helmoholtz free energy F is:

$$F = U - TS = -T \log Z, \tag{1.6}$$

and the following thermodynamics relations holds:

$$U = F + \frac{\partial}{\partial \beta} \beta F = -\frac{\partial}{\partial \beta} \log Z \tag{1.7}$$

$$S = \beta^2 \frac{\partial F}{\partial \beta}. \tag{1.8}$$

Another variational formulation that plays a important role is the principle of minimum free energy functional $\Phi[\rho]$:

$$\Phi[\rho] = U[\rho] - \frac{S[\rho]}{\beta}. \tag{1.9}$$

The entropic term, thought as a functional of the distribution ρ, is maximized when the latter is uniform and minimized when the distribution gives nonzero probability only to a single state σ (we refer for simplicity to a discrete space of microscopic states), favoring thus a disordered state. The equilibrium state, identified by the Boltzmann distribution, represents the distribution which minimizes the free energy $\Phi = F = U - TS$, as it is the best tradeoff between internal energy minimization and entropy maximization at a given temperature T. The higher the temperature, then, the more important the relative role of the disorder, encoded in S.

From the partition function one can compute all the thermodynamic functions, and it should be able to see, possibly, phase transitions.
For a finite system, Z is a finite sum of positive analytic functions of the temperature (for $T \neq 0$), and consequently also the free energy density $f_N = -N^{-1}T \log Z$ is an analytic function of T. Since phase transitions correspond to singularities of thermodynamic functions, it is necessary to go to the thermodynamic limit to see them.

1.1 Ising model

In many magnetic materials electrons responsible for magnetic behavior are localized near the atoms of a lattice, and there is an exchange interaction that tend to orient the spins. The Ising model describes this situation in a very simple way. To every point on a lattice, a dicotomic variable $\sigma = \pm 1$ is associated that mimic the spin of the electron. The Hamiltonian in presence of an uniform external field h is

$$H(\sigma; h) = -J \sum_{i,k}^{\text{n. n.}} \sigma_i \sigma_k - h \sum_i \sigma_i, \tag{1.10}$$

where the sum over i and k runs over all possible nearest neighbour of pairs in the lattice. If the exchange constant J is positive the system is ferromagnetic and nearby spin tends to stay parallel; if J is negative the system is antiferromagnetic and antiparallel spin are energetically favoured.

Despite its simplicity the Ising model has been the paradigmatic example for the statistical mechanics studies. A lot of fundamental ideas, such as spontaneous symmetry breaking, order parameter, critical exponents have been formulated and discovered in its framework.

The magnetization density associated with a given spin configuration σ is

$$m(\sigma) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i \tag{1.11}$$

Note that the average magnetization can be expressed through a derivative with respect to h:

$$\langle m \rangle = \frac{1}{\beta N} \frac{\partial}{\partial h} \log Z = -\frac{\partial f}{\partial h} \tag{1.12}$$

In the trivial case of Hamiltonian (1.10) with $J = 0$, the partition function factorizes in the sums over the single uncorrelated spins,

$$\sum_{\{\sigma\}} e^{\beta h \sum_i \sigma_i} = \prod_i \left(\sum_{\sigma_i = \pm 1} e^{\beta h \sigma_i} \right) = \left(2 \cosh(\beta h)\right)^N \tag{1.13}$$

describing a paramagnetic system with magnetization

$$\langle m \rangle = \tanh(\beta h). \tag{1.14}$$

In absence of the external field a very important property of the energy (1.10) is the invariance under reversal of all spins ($\sigma_i \to -\sigma_i$, for all i), while the magnetization changes sign under this operation. This symmetry, called gauge symmetry, is *explicitly* broken by the external field term in the Hamiltonian.

The minimization of the free energy $F = U - TS$ requires the best compromise between minimal energy U (all spins aligned) and maximum entropy S (random spins). For low temperatures it is more efficient to minimize energies in order to minimize F , and consequently the spins can become ordered giving rise to a spontaneous magnetization (ferromagnetic phase). As the temperature increases, entropy plays a major role, so disordered spins are favored (paramagnetic phase) and there is no spontaneous magnetization. A phase transition between these two behaviors is then expected.

Note that, to apply the above energy-entropy argument, it is necessary to explicitly break the spin reversal symmetry by applying, for example, an infinitesimal external field, since the spin reversal symmetry leads immediately to the absence of magnetization in zero external field. The thermodynamic limit of the average magnetization per spin $\langle m \rangle$ should be then computed as

$$\langle m \rangle_0 = \lim_{h \to 0^+} \lim_{N \to \infty} \frac{1}{N} \sum_i \langle \sigma_i \rangle. \tag{1.15}$$

The order of the limits is crucial to obtain the appearance of spontaneous magnetization ($\langle m \rangle_0 \neq 0$), since if we reverse the order we always obtain zero. This situation, in which the low temperature thermodynamic state has a lower degree of symmetry than the Hamiltonian, is called *spontaneous symmetry breaking* and the magnetization m is the *order parameter*.

Despite its simplicity in the formulation, the partition function Z with the Hamiltonian (1.10) can be calculated exactly only for a regular lattice in one dimension and in two dimensions with $h = 0$. In the general case approximate results must be considered, i.e. in lattice with more dimensions or where the model is embedded in different topologies and the spin variable can be associate to vertices in a generic graph. The simplest of these is the mean field approximations that consider each spin interacting with an average of the other spins, a mean field, parametrized by the magnetization. This is essentially equivalent to considered the probability distribution factorized over the spins:

$$\rho[\sigma] = \prod_i \rho_i(\sigma_i) \,, \quad \rho_i(\sigma) = \frac{1 + m_i}{2} \delta_{\sigma,1} + \frac{1 - m_i}{2} \delta_{\sigma,-1} \tag{1.16}$$

where δ is the Kronecker delta.

1.1.1 Curie-Weiss model

Instead of discussing the solutions of the mean field approximation, we introduce an Ising model that have the same solution of the mean field approximation. This model have weak long range interaction and is known as the Curie Weiss model. This model will constitute the basis for modelling the cooperative systems in chapter (4).

$$H = -\frac{J}{N} \sum_{i<k} \sigma_i \sigma_k - h \sum_i \sigma_i \tag{1.17}$$

The sum runs over all the pairs i, k of spins and the interactions are properly rescaled to obtain an extensive internal energy as we are going to see.

The reason why this model recovers the mean field approximation is essentially due to the decorrelation of spins that occur in the thermodynamics limit, where the two point correlation became $\langle \sigma_i \sigma_j \rangle \to m_i m_j$.

The partition function Z can be recast introducing the magnetization m in the following way:

$$Z = \sum_{\{\sigma\}} \exp\left[\frac{\beta J}{N}\sum_{i<j}\sigma_i\sigma_j + \beta h\sum_i \sigma_i\right] = \tag{1.18}$$

$$\simeq \sum_{\{\sigma\}}\int dm\delta\left(Nm - \sum_i \sigma_i\right)\exp\left[\frac{\beta J}{2N}m^2 + \beta hm\right] \tag{1.19}$$

where in the last equality we neglect subdominant term in N. Using the integral expression for the Dirac's delta, we rewrite

$$\begin{aligned} Z &\simeq \sum_{\{\sigma\}}\int_{-i\infty}^{i\infty} dt\int dm \exp\left[N\left(tm - \frac{1}{N}t\sum_i\sigma_i + \beta J/2m^2 + \beta hm\right)\right] = \\ &= \int_{-i\infty}^{i\infty} dt\int dm\exp\left[N\left(tm - \log 2\cosh(t) + \beta J/2m^2 + \beta hm\right)\right] = \\ &= \int_{-i\infty}^{i\infty} dt\int dm\exp\left[N\Phi\right] = \end{aligned}$$

where we have sum over the spin configuration $\{\sigma\}$, $\sum_{\{\sigma\}}\prod_i \exp[-t\sigma_i] = 2\cosh(t)^N$. We call $N\Phi$ the exponent in the integrals. Performing the thermodynamic limit, it is possible to use the saddle point method to evaluate the integrals. We take the maximum of the function Φ to consider the leading term in N:

$$\frac{\partial\Phi}{\partial t} = 0 \quad\rightarrow\quad -m = \tanh(t) \tag{1.20}$$

$$\frac{\partial\Phi}{\partial m} = 0 \quad\rightarrow\quad -t = \beta Jm + \beta h \tag{1.21}$$

Substituting the last equation in (1.20) and in the expression of Φ we obtain the self consistency equation or Curie Weiss equation and the free energy density $f(m;\beta,J,h) = -\frac{1}{\beta}\Phi$:

$$m = \tanh\left[\beta\left(Jm + h\right)\right] \tag{1.22}$$

$$f = \frac{J}{2}m^2 - \frac{1}{\beta}\log\cosh\left[\beta\left(Jm+h\right)\right] - \frac{1}{\beta}\log 2 \tag{1.23}$$

Equations such as (1.22) are known in statistical mechanics as *self consistency relations*. Solving it, checking that the solution is stable, in the sense that it minimizes $f(m)$, and that it represents the absolute minimum of the free energy gives the physical value of the order parameter, namely the one that is exponentially (in the system size) the most probable to find by picking at random a configuration of the system with the Boltzmann probability.
Using the thermodynamics relations $s = -\beta^2\frac{\partial f}{\partial\beta}$ and $f = e - \frac{1}{\beta}s$ we obtain the densities of the internal energy e and the entropy s:

$$e = -\frac{J}{2}m^2 - hm \tag{1.24}$$

$$s = -\beta Jm^2 - \beta hm + \log\cosh\left[\beta\left(Jm+h\right)\right] + \log 2 \tag{1.25}$$

We discuss briefly the solutions of the self-consistency equation in the absence of the external field ($h = 0$), expanding the self consistency equation (1.22) for small m gives

$$m = \beta J m - \frac{\beta J}{3} m^3 \,. \tag{1.26}$$

By studying the solution of this equation for different values of the temperatures it is possible to observe that two very different thermodynamical regimes are separated by the presence of a critical temperature T_c.
For $\beta < \beta_c = 1/J$ the only solution is $m = 0$ and it is stable. This correspond to a high temperature ($T > T_c = J$) paramagnetic phase in which the system displays a null magnetization and it is ergodic in the sense that all the configurations with a non-null statistical measure are connected by the (local) operation of a single spin flip. By performing on a finite system a dynamic that is consistent with the Boltzmann measure, independently on the initial condition, after a transient, one should observe the system exploring configurations where the fluctuations of the magnetization are small, namely they vanish with the system size as $\delta m \to \mathcal{O}(1/\sqrt{N}) \to 0$. In the paramagnetic phase the free energy is simply $f(\beta < \beta_c, J) = -\frac{1}{\beta} \log 2$.
For $\beta > \beta_c = 1/J$ the $m = 0$ solution is no longer stable and two symmetric stable solution $m = \pm m^*$ appear, each corresponding to the system being polarized in one of the two possible directions, spontaneously breaking the symmetry of the Hamiltonian. As the two solutions are mapped one into the other by the transformation under which the Hamiltonian is invariant (the global spin flip: $m \to -m$), the two solutions give the same value of the free energy. This phase is called ferromagnetic. In this phase the system is no longer ergodic. In fact the Boltzmann measure displays two "peaks" in the phase space that are not connected by local moves. In this regime a dynamics at finite size will show the system fluctuating around each of the two symmetric magnetization and jumping between them. Increasing the system size N, fluctuations around one solution vanish as $\delta m \to \mathcal{O}(1/\sqrt{N}) \to 0$ and the time of permanence in one of the two state grows exponentially as $\sim e^N$ so that a very large system becomes trapped in one of the two state and only one "peak" of the Boltzmann measure is explored: the system remains confined to a restricted region of the phase space, being this region smaller as the temperature is lower. This fact is know as *ergodicity breaking* and the (two, in this case) different sets of available configurations in which the system gets restricted are know in statistical mechanics as *pure states* or *ergodic components*.

The presence of a positive external field explicitly breaks the symmetry of the Hamiltonian so that, in this case, in the low temperature phase, the two saddle points value of the magnetization , $m_1 > m_2$, are no longer equivalent: the state corresponding to the greater one has a lower free energy, $f(m_1) < f(m_2)$, and so is the one that gives the leading contribution to the free energy, while the contribution of the smaller one vanishes exponentially. The model with external field h display a *fist order phase transitions* for $h = 0$ in which the order parameter displays a jump at the critical point, $\lim_{h\to 0^+} m(\beta, J, h) \neq \lim_{h\to 0^-} m(\beta, J, h)$.

The consistent way to have a result with spontaneous symmetry breaking in case of vanishing external field, $h = 0$ is to compute the free energy in the thermodynamic limit at non zero field and then to send its value to zero:

$$f(\beta, J, h = 0) = \lim_{h\to 0} f(\beta, J, h \neq 0) = \lim_{h\to 0} \lim_{N\to\infty} f(\beta, J, h \neq 0, N) \,; \tag{1.27}$$

in this way, for example, in the low temperature phase $\langle m(\{\sigma\})\rangle \neq 0$.

The ferromagnetic phase transition at $\beta = \beta_c$ is a *second order phase transition*. In fact a discontinuity appear, at the critical point in the susceptibility $\chi = \partial m/\partial h = \partial^2 f/\partial h^2$ that is the second derivative of the free energy with respect to the control parameter of the system, the external magnetic field h.

1.1.2 Antiferromagnetic Ising model

The antiferromagnetic (AFM) Ising model is obtained from the Hamiltonian (1.10) by simply changing $J \to -J$. For many regular lattices (cubic, honey-comb, body-centered cubic, etc.) it is possible to divide the lattice into two equal sublattices, labelled A and B, in such a way that all the nearest-neighbor sites of any A-site are B-sites and vice versa. The interaction graph is said to be bipartite. Here we consider the mean-field version of the antiferromagnetic Ising model, assuming that each A-site interacts with the same coupling J with all the B-sites (and vice versa). Then the Hamiltonian can be written as

$$H(\sigma; h) = \frac{J}{N} \sum_{i \in A,\, j \in B} \sigma_i^A \sigma_j^B - h \sum_{i \in A,B} \sigma_i \tag{1.28}$$

Note that for $h = 0$ this is the Hamiltonian of a ferromagnet with spin variable σ_i^A and $-\sigma_j^B$ and there is a one-to-one correspondence between microstates of equal energy in the zero-field ferromagnet and antiferromagnet. In fact, reversing the spin on one sublattice and taking $J \to -J$, one can connect the microstates for the two cases. The minimum energy for $T = 0$ is achieved in this case with an antiparallel ordering of all the nearest-neighbor spin pairs.

The magnetizations for the two sublattice are defined by

$$m^A(\sigma^A) = \frac{2}{N} \sum_{i \in A} \sigma_i, \qquad m^B(\sigma^B) = \frac{2}{N} \sum_{j \in B} \sigma_j \tag{1.29}$$

and the overall magnetization is given by their arithmetic average

$$m(\sigma) = \frac{1}{2}(m^A(\sigma^A) + m^B(\sigma^B)). \tag{1.30}$$

An important role is played by the staggered magnetization

$$n(\sigma) = \frac{1}{2}(m^A(\sigma^A) - m^B(\sigma^B)), \tag{1.31}$$

which corresponds to the order parameter for the antiferromagnet.

In presence of a not vanishing external field $h \neq 0$, the correspondence between the states of the antiferromagnet and those of the ferromagnet is lost. However, if either the magnetic field h or the thermal disorder are strong enough with respect to the antiferromagnetic interaction, the equilibrium state will be one in which $\langle m^A \rangle = \langle m^B \rangle = \langle m \rangle$, so that $\langle n \rangle = 0$ (paramagnetic state). On the contrary, for small fields and temperature one expects that the staggered magnetization $\langle n \rangle$ is different from zero (antiferromagnetic state). There will be a transition curve in the (T, h) plane, separating these two phases. This curve should be symmetrical about $h = 0$ and meets the temperature axis at $T = T_c$. Moreover, there exists a critical field h_c on the $T = 0$ axis separing the two phase.

The analogue of the free energy density $f(m,h)$ introduced in the previous section through equation (1.23), is given by the following expression

$$\begin{aligned} f(m^A,m^B,h) =& T\left(\frac{1+m^A}{4}\log(1+m^A)+\frac{1-m^A}{4}\log(1-m^A)\right.\\ &\left.+\frac{1+m^B}{4}\log(1+m^B)+\frac{1-m^B}{4}\log(1-m^B)\right)\\ &+\frac{J}{4}m^Am^B-\frac{h}{2}(m^A+m^B) \end{aligned} \tag{1.32}$$

The first two rows correspond to the entropic term: in this case it is the sum of the entropies of the two sublattices, evaluable for instance through the Stirling formula (observing that $N_A = N_B = N/2$). Last row corresponds to the Hamiltonian (1.28) expressed in terms of the two sublattice magnetizations. The equilibrium state corresponds to the minimum condition of this expression with respect to m^A and m^B, which is achieved if the self-consistence equations

$$m_A = \tanh(-\beta J m_B/2+\beta h) \tag{1.33}$$

$$m_B = \tanh(-\beta J m_A/2+\beta h) \tag{1.34}$$

are satisfied. The paramagnetic state, where $m^A = m^B$, is stable for large values of temperature and field. Imposing that the Hessian of f with respect to m^A and m^B is positive in this minimum, one has the stability condition for the paramagnetic phase; the borderline case of vanishing Hessian gives the critical line $h(T)$. We refer to [11] for a detailed analysis.

1.2 The Sherrington-Kirkpatrick model

In the previous section we described the Ising model characterized by equals couplings among the spins. An important generalization that was introduced around forty years ago [12], has been the ground for the foundation of a training branch of the statistical physics that is the theory of the disordered systems and in particular of the spin glass.
Abandoning the homogeneous framework, the extension considers heterogeneous couplings. The natural way to take into account different coupling values is to describe these in a probabilistic way. The spins interact with a couplings J picked randomly from a distribution $P(J)$.

Originally the model was born to describe some magnetic alloys characterized by lack of long range order and a non periodic arrangement of magnetic moments with a very long scale dynamics at low temperatures, but nowadays the spin glass have a wide range of applications, include optimization theory, computer science, biology, economy [13, 9, 14].

Introducing the average over the realizations of the disorder, one realize that there are basically two way to compute it. One procedure is called *annealing* disorder, and considers thermal fluctuation on the same time scale of the realizations one.

$$\overline{f} = \lim_{N\to\infty} 1/N \log \mathbb{E}Z(J) \tag{1.35}$$

The other more interesting method is to considered the equilibrium states of the system at fixed different realizations of the couplings and to average over them. The random parameters are collectively denoted as *quenched* or "frozen" disorder. From a physical point of view, the word "frozen" means that we are modelling a disordered system whose

impurities have a dynamics which is many orders of magnitude slower than the dynamics of the spin degrees of freedom. Therefore, the disorder does not reach thermal equilibrium on the time scales of the spin relaxation and can be considered as fixed.

$$\overline{f} = \lim_{N\to\infty} 1/N\mathbb{E}\log Z(J) \tag{1.36}$$

This terminology comes from metallurgy and the thermal processing of materials: a "quench" corresponds in this jargon to preparing a sample by suddenly bridging it from high to low temperatures, so that atoms do not change their position apart from small vibrations. In an "annealing" process, on the contrary, the cooling down is slower and gradual, and atoms can move and find favorable positions.

The second key ingredient, following from disorder, is frustration, i.e., competition between different terms in the Hamiltonian, so that they can not all be satisfied simultaneously. Frustration in the network gives rise to a complex energy (and free energy) landscape with many metastable (locally minimizing the energy) configurations.

1.2.1 The replica symmetric solution

To perform the average over the disorder of $\log Z$ we introduce the so called *replica trick* which makes use of the limit

$$\mathbb{E}\log Z = \lim_{n\to 0} \frac{\mathbb{E}Z^n - 1}{n} = \lim_{n\to 0} \frac{1}{n}\log[\mathbb{E}\, Z^n] \tag{1.37}$$

When the parameter n is considered as an integer, computing $\mathbb{E}\, Z^n$ corresponds to compute the partition function for n independent copies of the system with the same realization of the disorder and to average their product. The limit is then interpreted as an analytical continuation to real values of n. This procedure can be mathematically problematic. Moreover, as we will see in practical cases, Z^n is usually computed performing the thermodynamic ($N \to \infty$) limit before the $n \to 0$ one. This exchange of the ordering of the limits is usually not justified. Nevertheless in some cases (as for instance for the Sherrington-Kirkpatrick model that we are going to introduce, see [15]) the procedure is proven to give the exact results by rigorous arguments.

The Sherrington-Kirkpatrick (SK) model is the paradigm for disordered spin systems for the compactness of its definition and the richness of its phenomenology. It is a mean field spin model defined through the Hamiltonian

$$H_{SK} = -\frac{1}{\sqrt{N}} \sum_{i<j}^{N} J_{ij}\sigma_i\sigma_j \,, \tag{1.38}$$

where the couplings J_{ij} are independently thrown from a standard normal distribution:

$$dP(J_{ij}) = \frac{1}{\sqrt{2\pi J}} e^{-\frac{(J_{ij}-\frac{J_0}{\sqrt{N}})^2}{2J}} dJ_{ij} \quad \text{and} \quad P(J_{ij}, J_{kl}) = P(J_{ij})P(J_{kl}), \quad \forall (i,j) \neq (k,l)\,, \tag{1.39}$$

This choice is a matter of convenience. In fact any other symmetric probability distribution with finite moments could be chosen for J_{ij} , without modifying the free energy of the system, apart from error terms vanishing in the thermodynamic limit. The case $J_{ij} = \pm 1$ with equal probability $1/2$, for instance, is often considered in the literature.

The normalization factor $1/\sqrt{N}$ guarantees that $H_N(\sigma; J)/N$ and the free energy density are of order unity in the thermodynamic limit. In the Curie-Weiss model the correct

factor is $1/N$, but in this case the random signs of the couplings J_{ij} produce cancellations among the many terms of the Hamiltonian H_N. The correctness of this choice can be easily understood by considering a duplicated system with configurations σ^a and σ^b, but with the same disorder, and computing the quantity

$$\mathbb{E}(H_N(\sigma^a;J)H_N(\sigma^b;J)) = \frac{1}{N}\sum_{i<j}^{1,N}\sum_{k<l}^{1,N}\mathbb{E}(J_{ij}J_{kl})\sigma_i^a\sigma_j^a\sigma_k^b\sigma_l^b$$

$$= \frac{1}{N}\sum_{1\le i<j\le N}\sigma_i^a\sigma_j^a\sigma_i^b\sigma_j^b = \frac{N}{2}\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^a\sigma_i^b\right)^2 - \frac{1}{2}.$$

The term

$$q_{ab} = q(\sigma^a,\sigma^b) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^a\sigma_i^b, \tag{1.40}$$

which occurs in the previous equation, is a fundamental one, as we will see in the following, and it is called *overlap*. In fact, it measures the resemblance between the configurations of the two copies σ^a and σ^b, going from -1, when each spin of a system is opposed to the corresponding one of the other copy, to $+1$, when they are perfectly aligned. It is related with the Hamming distance $d(\sigma^a,\sigma^b)$, which counts the number of non-aligned spins: $d(\sigma^a,\sigma^b) = 1/2(1-q_{ab})$. So, taking two identical copies $\sigma^a = \sigma^b$, we note that

$$\mathbb{E}\left(H_N(\sigma;J)\right)^2 = \frac{N}{2} - \frac{1}{2}, \tag{1.41}$$

showing that the normalization factor is correct.

Curiously, the title of the paper [16] where Sherrington and Kirkpatrick firstly proposed a solution was "Solvable Model of a Spin-Glass", but also if the authors, using the replica-trick, found an explicit form for the free energy, they realized that their solution was only valid above a certain temperature, and it took some years to understand which was the valid solution, with the seminal works of Parisi [17, 18, 19], where he proposed a formula for the free energy per site in the thermodynamic limit and a description of the pure states of the system. The rigorous proof that the Parisi formula is in fact correct was established only some years ago, split across two works by Guerra [20] and Talagrand [21].

Using the replica trick (1.37),

$$\bar{f}(\beta) = \lim_{N\to\infty} -\frac{1}{\beta N}\mathbb{E}\log Z = \lim_{N\to\infty} -\frac{1}{\beta N}\lim_{n\to 0}\log \mathbb{E}Z^n\,. \tag{1.42}$$

Performing the disorder average of the replicated systems involves just Gaussian integrations so that

$$\mathbb{E}Z^n = \sum_{\{\sigma^1\}\dots\{\sigma^n\}}\int_{-\infty}^{\infty}\prod_{ij}dJ_{ij}\exp\left[-\frac{(J_{ij}-\frac{J_0}{\sqrt{N}})^2}{2J} + \sum_a^n\frac{\beta}{2\sqrt{N}}\sum_{i,j}J_{ij}\sigma_i^a\sigma_j^a\right] \tag{1.43}$$

$$= \sum_{\{\sigma^1\}\dots\{\sigma^n\}}\exp\left[\frac{\beta J_0}{2N}\sum_a^n\sum_{i,j}\sigma_i^a\sigma_j^a + \frac{\beta^2J^2}{4N}\sum_{ij}^{N}\sum_{ab}^{n}\sigma_i^a\sigma_i^b\sigma_j^a\sigma_j^b\right]. \tag{1.44}$$

Now, using the definition of the overlap matrix $q_{ab}(\{\sigma\}) = 1/N\sum_i^N\sigma_i^a\sigma_i^b$ and the magnetization $m_a = 1/N\sum_i^N\sigma_i^a$, calling the integration measures $Dm \equiv \prod_a^n dm^a$ and $Dq \equiv$

$\prod_{ab}^{n} dq_{ab}$ and neglecting the sub-dominant order term in N we get:

$$\mathbb{E}Z^n = \sum_{\{\sigma\}} \int Dm \int Dq \delta(Nm^a - \sum_i \sigma_i^a) \delta(Nq_{ab} - \sum_i \sigma_i^a \sigma_i^b) e^{\frac{\beta J_0 N}{2} \sum_a^n m_a^2 + \frac{N\beta^2 J^2}{4} \sum_{ab}^n q_{ab}^2} =$$

$$= \sum_{\{\sigma\}} \int Dm \int Dq \int_{-i\infty}^{i\infty} D\Lambda \int_{-i\infty}^{i\infty} D\gamma \exp\left[N \sum_a \gamma_a m_a - \sum_a \gamma_a \sum_i \sigma_i^a \right] \otimes \tag{1.45}$$

$$\otimes \exp\left[N \sum_{a,b} \lambda_{ab} q_{ab} - \sum_{a,b} \lambda_{ab} \sum_i \sigma_i^b \sigma_i^a + N \frac{\beta J_0}{2} \sum_a^n m_a^2 + N \frac{\beta^2 J^2}{4} \sum_{ab}^n q_{ab}^2 \right] = \tag{1.46}$$

$$= \int Dm \int Dq \int_{-i\infty}^{i\infty} D\Lambda \int_{-i\infty}^{i\infty} D\gamma \exp\left[-N\beta\Phi \right] \tag{1.47}$$

with

$$\Phi(m, q, \gamma, \lambda) = -\frac{1}{\beta} \sum_a \gamma_a m_a - \frac{1}{\beta} \sum_{a,b} \lambda_{ab} q_{ab} - \frac{J_0}{2} \sum_a^n m_a^2 - \frac{\beta J^2}{4} \sum_{ab}^n q_{ab}^2 + \tag{1.48}$$

$$-\frac{1}{\beta N} \log \left(\sum_{\{\sigma^1\} \ldots \{\sigma^n\}} \exp\left[-\sum_a \sum_i \sigma_i^a \gamma_a - \sum_{a,b} \sum_i \sigma_i^a \lambda_{ab} \sigma_i^b \right] \right) = \tag{1.49}$$

that is symmetric under permutations of replica indexes. The physical value of the intensive free energy is given for formula (1.42). Exchanging the orders of the limits in that formula is the only way to compute the integrals over the measures Dq, Dm and $D\gamma$,$D\lambda$. In that case, taking the thermodynamic limit before sending the number of replicas to zeros allows to use the saddle point method. So the extremization of $\Phi(m, q, \gamma, \lambda)$ over m and q allow to eliminate the additional variables λ and γ with

$$\frac{\partial \Phi}{\partial m_a} = 0 \to \gamma_a = -\beta J_0 m_a; \quad \frac{\partial \Phi}{\partial q_{ab}} = 0 \to \lambda_{ab} = -\frac{\beta^2 J^2}{2} q_{ab} \,, \tag{1.50}$$

Substituting we get the expression Φ as a function of the magnetization and the overlap matrix is

$$\Phi(m, q; \beta) = \frac{J_0}{2} \sum_a m_a^2 + \frac{\beta J^2}{4} \sum_{ab} q_{ab}^2 - \frac{1}{\beta} \log \left(\sum_{\{\sigma\}} e^{\beta J_0 \sum_a \sigma^a m_a + \frac{\beta^2 J^2}{2} \sum_{ab} \sigma^a q_{ab} \sigma^b} \right) . \tag{1.51}$$

In last formula it is evident the effect of the replica trick: it has decoupled sites but now replicas are coupled by the overlap matrix. Last expression has to be computed in the saddle point and the limit $n \to 0$ is still to be performed. In order to do this a parametrization of the overlap matrix needs to be made explicit. A good parametrization could permit to decouple replicas.

The simplest *ansatz* for the form of the matrix overlap is the one that respects the symmetry of the $f(q)$, so a matrix which has 1 on the diagonal and a parameter q_0 on all the entries out of the diagonal (and the magnetization of different replica are equal $m_a = m$) i. e.

$$q_{ab} = q_0 + (1 - q_0)\delta_{ab} \,. \tag{1.52}$$

This is known as the *replica symmetric (RS) anstatz*. With this parametrization the free energy can be obtained as the limit $n \to \infty$ (we neglect terms order n^2):

$$\begin{aligned} f(m, q_0; \beta) &= \lim_{n\to 0} \frac{1}{n}\Phi(q_{RS}) = \lim_{n\to 0} \frac{J_0}{2}m^2 + \frac{\beta J^2}{4}(1-q_0^2) + \\ &\quad -\frac{1}{\beta n}\log\left(\sum_{\{\sigma\}} e^{nJ_0\beta m\sigma + \frac{\beta^2 J^2}{2} q_0 \sum_{ab}^{n}\sigma^a\sigma^b + \frac{\beta^2 J^2}{2}(1-q_0)n}\right) = \\ &= \frac{J_0}{2}m^2 - \frac{\beta J^2}{4}(1-q_0)^2 - \lim_{n\to 0}\frac{1}{\beta n}\log\left(\sum_{\{\sigma\}} e^{nJ_0\beta m\sigma + \frac{1}{2}\left(n\beta J\sqrt{q_0}\sigma\right)^2}\right) \end{aligned}$$

Rewriting the last term in the exponential as the result of a Gaussian integration,

$$\begin{aligned} f(m, q_0; \beta) &= \frac{J_0}{2}m^2 - \frac{\beta J^2}{4}(1-q_0)^2 - \lim_{n\to 0}\frac{1}{\beta n}\log\sum_{\{\sigma\}}\int d\mu(z) e^{n(J_0\beta m + z\beta J\sqrt{q_0})\sigma} = \\ &= \frac{J_0}{2}m^2 - \frac{\beta J^2}{4}(1-q_0)^2 - \lim_{n\to 0}\frac{1}{\beta n}\log\int d\mu(z)\left(2\cosh(J_0\beta m + z\beta J\sqrt{q_0})\right)^n \end{aligned}$$

We can now use the replica trick relation (1.37) in the inverse sense, i.e. $\lim_{n\to\infty} 1/n \log \mathbb{E}\Phi^n = \mathbb{E}\log\Phi$, so that we have the final result for the RS free energy of the SK model

$$f(m, q_0; \beta) = \frac{J_0}{2}m^2 - \frac{\beta J^2}{4}(1-q_0)^2 - \frac{1}{\beta}\int d\mu(z)\log\left(\cosh(J_0\beta m + z\beta J\sqrt{q_0})\right) - \frac{1}{\beta}\log 2\,. \tag{1.53}$$

This expression has to be minimized with respect to the parameters m, q and this procedure gives the self consistency relations:

$$m = \int d\mu(z)\tanh(J_0\beta m + z\beta J\sqrt{q_0}) \tag{1.54}$$

$$q_0 = \int d\mu(z)\tanh^2(J_0\beta m + z\beta J\sqrt{q_0})\,. \tag{1.55}$$

By studying equations (1.55) and (1.54) it is possible to verify that for $T > J_0$ only the solution $m = 0$ is stable and for $T > J_0$ and $T > J$ the only solution is $m = q_0 = 0$. For $T < J_0 < J$ and for $T < \max\{J_0, J\}$ a stable solution with $q_0 > 0$ appears, giving the scenario summarized in the following table, and showed in the phase diagram figure (1.1):

	at	from	to
$J_0 > J$	$T = J_0$	$m = q_0 = 0$	$m \neq 0, q > 0$
$J_0 < J$	$T = J$	$m = q_0 = 0$	$m = 0, q_0 > 0$
$T < \max\{J_0, J\}$	$T = J_0(1-q_0)$	$m = 0, q_0 > 0$	$m \neq 0, q_0 > 0$

For the SK models the replica symmetric solutions is not exact and the scientific community was suddenly aware of this as, in their paper [22], Sherrington and Kirkpatrick pointed out that, at low temperatures, their solution gives a negative entropy (that is not possible for discrete variables where the entropy is the logarithm of the number of available configurations at temperature T, that is always greater than one and so its log greater than zero). Studying the fluctuation of $f(q, m)$ around the RS solution $f(q_{RS}, m_{RS})$ is possible to get the *Almeida Thouless* (AT) line that separate the phase where the ergodicity is broken (spin glass phase). The Parisi solution [18] gives the correct structure of the low temperature phase which is much more complex of the RS scenario.

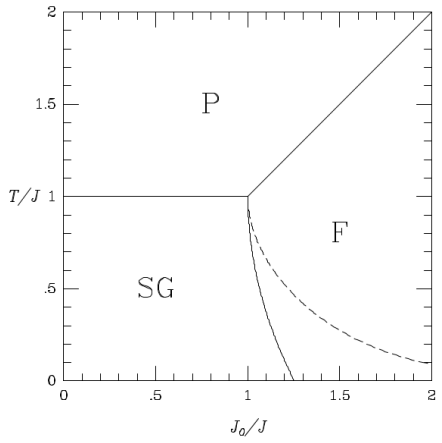


Figure 1.1: Phase diagram of the SK model obtain from the replica-symmetric solution. Solid lines separated the paramagnetic phase (P) where $m = q_0 = 0$, the ferromagnetic phase (F) where $m \neq 0, q > 0$ and the spin glass phase (SG) where $m = 0, q > 0$. Dashed line is the so called Almeida Thouless*(AT) line that separate the region of the parameter space where the RS* ansatz *gives the right solution from the region in which it represents an approximation. Figure from [10].*

As we have shown in Section 1.1.1 for the case of the mean field Ising model, in interacting systems ergodicity can break in the low temperature phase where the phase space can split into ergodic regions where the system remains trapped. Let us indicate states with Greek letters. Pure states, in mean field systems (not at the critical point), at equilibrium, have the *clustering properties* which states that spins are independent inside a generic state α:

$$\langle \sigma_i \sigma_j \rangle_\alpha = \langle \sigma_i \rangle_\alpha \langle \sigma_j \rangle_\alpha \, . \tag{1.56}$$

In finite dimension systems, the property still holds for spins that are far apart in the space structure, i.e. for $|i - j| \to \infty$. Calling $Z_\alpha = \sum_{\{\sigma \in \alpha\}} e^{-\beta H(\sigma)}$ and $w_\alpha = Z_\alpha / Z$, such that $\sum_\alpha w_\alpha = 1$ and $\sum_\alpha Z_\alpha = Z$, thermal average can be split into ergodic components as

$$\langle \Phi(\sigma) \rangle = \sum_{\{\sigma\}} \Phi(\sigma)\, e^{-\beta H(\sigma)} = \sum_\alpha w_\alpha \langle \Phi(\sigma) \rangle_\alpha \tag{1.57}$$

We can use the above identity to write the distribution $P(m)$ and $P(q)$ of the magnetization and the overlaps in equilibrium as

$$\begin{aligned} P(m) &= \frac{1}{Z} \sum_{\{\sigma_a\}} e^{-\beta H(\sigma^a)} \delta(m - m_a) = \\ &= \sum_{\alpha\beta} w_\alpha \delta(m - m_\alpha) \end{aligned} \tag{1.58}$$

$$\begin{aligned} P(q) &= \frac{1}{Z} \sum_{\{\sigma_a \sigma_b\}} e^{-\beta H(\sigma^a) - \beta H(\sigma^b)} \delta(q - q_{ab}) = \\ &= \sum_{\alpha\beta} w_\alpha w_\beta \, \delta(q - q_{\alpha\beta}) \end{aligned} \tag{1.59}$$

From this relation is possible to understand the role played by replicas in spin glasses: they serves as probes to explore the structure of states of the system which, as we will see,

can be very complex. So last quantity is of uttermost importance for spin glasses and in fact the true order parameter of this kind of systems turns out to be the full average overlap distribution $\mathbb{E}P(q)$. A hint that could justify this claim comes from the definitions of the moments of the average overlap distribution that, by the use of the clustering property, can be easily shown to be

$$q^{(k)} = \int dq\, q^k\, \mathbb{E}P(q) = \frac{1}{N^k} \sum_{i_1 \dots i_k} \mathbb{E}\langle \sigma_{i_1} \dots \sigma_{i_k} \rangle^2 \,. \tag{1.60}$$

In particular the first moment is

$$q^{(1)} = \int dq\, q\, \mathbb{E}P(q) = \frac{1}{N} \sum_{i} \mathbb{E}\langle \sigma_i \rangle^2 \equiv q_{EA} \,. \tag{1.61}$$

This quantity, also known as the *Edward-Anderson order parameter*, gives a very intuitive information about the presence of an "ordered" phase in spin glasses. In fact as the couplings are randomly distributed around zero, the naif idea about the broken phase is that spins are frozen around random directions evenly distributed among sites. So in spin glasses the global magnetization $m = 1/N \sum_i \langle \sigma_i \rangle$ vanishes also if spins are frozen in random directions and the local magnetizations $m_i = \sum_i \langle \sigma_i \rangle \neq 0$; while, thanks to the presence of the power two in its definition, q_{EA} is a global quantity that can distinguish between the paramagnetic phase, where $\langle \sigma_i \rangle = 0$, $\forall i$, and the spin glass phase $\langle \sigma_i \rangle \neq 0$ but with different signs depending on the site. We stress anyway that, even if q_{EA} can give a very effective information on the presence of symmetry breaking, the true order parameter of mean field spin glasses is the full distribution $\mathbb{E}P(q)$. In order to compute it, the replica trick is needed. In fact it can be used in general to compute thermal averages of a generic quantity $\Phi(\sigma)$ as

$$\begin{aligned} \langle \Phi(\sigma) \rangle &= \frac{\sum_{\{\sigma\}} e^{\beta H(\sigma)} \Phi(\sigma)}{\sum_{\{\sigma\}} e^{\beta H(\sigma)}} = \lim_{n \to 0} \sum_{\{\sigma\}} e^{\beta H(\sigma)} \Phi(\sigma) \left[e^{\beta H(\sigma)} \right]^{n-1} = \\ &= \lim_{n \to 0} \sum_{\{\sigma\}} \Phi(\sigma^1)\, e^{\beta \sum_a^n H(\sigma^a)} = \\ &= \lim_{n \to 0} \sum_{\{\sigma\}} \frac{1}{n} \sum_{a}^{n} \Phi(\sigma^a)\, e^{\beta \sum_b^n H(\sigma^b)} \,. \end{aligned} \tag{1.62}$$

So the average overlap distribution can be computed as

$$\mathbb{E}P(q) = \mathbb{E} \sum_{\{\sigma\}} \frac{2}{n(n-1)} \sum_{ab} \delta(q - q_{ab})\, e^{-\beta \sum_c^n H(\sigma^c)} \,. \tag{1.63}$$

Using the exchange of limits and the saddle point methods exactly as for the computation of the free energy, the result is

$$\mathbb{E}P(q) = \frac{2}{n(n-1)} \sum_{ab} \delta(q - q_{ab}^{SP}) \,, \tag{1.64}$$

where q_{ab}^{SP} stands for the overlap matrix computed at the saddle point.
The insertion of the RS *ansatz* give as a result

$$\mathbb{E}P(q)_{RS} = \delta(q - q_0) \,, \tag{1.65}$$

where $q_0 = q_{EA}$ and satisfies (1.54).

1.3 The Hopfield model

In this section, we will present the Hopfield model, firstly introduced by Pastur and Figotin and independently considered by Hopfield, in 1982, as a model for associative memory [23]. This model can be seen as generalization of the mean-field Ising model with many ground states. Call the P ground states $\boldsymbol{\xi}^\mu = \{\xi_1^\mu, .., \xi_N^\mu\}$ with $\mu = 1, .., P$, the Hamiltonian of the model is defined as

$$H_{\text{hopfield}}(\sigma; \boldsymbol{\xi}) = -\frac{1}{2N}\sum_\mu^P \left(\sum_i^N \xi_i^\mu \sigma_i\right)^2 = -\frac{1}{2N}\sum_{i,j}^N \left(\sum_\mu^P \xi_i^\mu \xi_j^\mu\right)\sigma_i\sigma_j - \frac{1}{2}P \tag{1.66}$$

The couplings matrix $J_{ij} = 1/N\sum_\mu \xi_i^\mu \xi_j^\mu$ is then a sum of projectors over the P vectors $\boldsymbol{\xi}^\mu$. The configurations of spins that are parallel (or anti-parallel) to the vectors $\boldsymbol{\xi}^\mu$, are local energy minimum for every μ. If some conditions are fulfilled [9], these configurations are also the ground state of the energy landscape, for example if $P \ll N$ and the patterns are more or less orthogonal.

The interpretation as a model for associative memory comes from considering the dynamics. Under some conditions, if the system has an initial configuration that is similar to one of the vectors, ξ_i^1 for example, it will evolve to the configuration $\sigma_i = \text{sign}(\xi_i^1)$, i.e. the initial state is in the basin of attraction of the first vector. In this context, the P vectors $\{\boldsymbol{\xi}^1, ..., \boldsymbol{\xi}^P\}$ are called memories (or patterns) that the (dynamical) model can retrieve starting from similar input configurations as, for instance, in the visual identification of an object. As shown in [9] the system can retrieve up to $\alpha_c N$ stored binary uncorrelated patterns with $\alpha_c = 0.138$. The failing of the retrieval of the patterns is due to states that are mixture of different patterns that are metastable state where the dynamics can be trapped at low temperature and can reduce their free energy for high temperature and for correlation between patterns.

The Hamiltonian (2.22) can be rewritten using the patterns overlap m_μ as:

$$H_{\text{hopfield}}(\sigma; \boldsymbol{\xi}) = -\frac{N}{2}\sum_\mu^P m_\mu^2, \qquad m_\mu = \frac{1}{N}\sum_i \xi_i^\mu \sigma_i \tag{1.67}$$

The simplest choice for the entries of patterns $\xi_i^\mu \in \{-1, 1\}$ is to randomly drawn it according to a Bernoulli distribution, $P(\xi_i^\mu = \pm 1) = 1/2$, in order to have uncorrelated patterns, and the number of patterns P to remains finite in the thermodynamic limits, so that:

$$\Phi(\boldsymbol{\xi}_i) = 2^{-P}\sum_{\xi_i \in \{-1,1\}} \Phi(\boldsymbol{\xi}_i), \qquad \langle \xi_i^\mu \xi_i^\nu \rangle = \delta_{\mu\nu} \tag{1.68}$$

The free energy becomes

$$f(\beta) \;=\; -\frac{1}{N\beta}\mathrm{E}_{\boldsymbol{\xi}} \log \int_{-\infty}^{+\infty} \prod_{\mu=1}^p dm_\mu \sum_{\{\sigma\}} \delta\big(Nm_\mu - \sum_i^N \xi_i^\mu \sigma_i\big) e^{\frac{1}{2}\beta N \boldsymbol{m}^2 + p/2} \tag{1.69}$$

where we introduce the vector $\boldsymbol{m} = \{m^1, .., m^\mu\}$.
Following step by step the method used in Section 1.1.1, introducing the Fourier representation of the δ function and applying the saddle point method to evaluated the integral, in the thermodynamic limit, the physical values taken by the order parameters are the solutions of the P equations:

$$\boldsymbol{m} = \langle \boldsymbol{\xi} \tanh(\beta\, \boldsymbol{m} \cdot \boldsymbol{\xi}) \rangle_{\boldsymbol{\xi}} \tag{1.70}$$

that locally extremize the free energy

$$f(\boldsymbol{m}) = \frac{1}{2}\boldsymbol{m}^2 - \frac{1}{\beta}\langle \log\cosh(\beta \boldsymbol{m}\cdot\boldsymbol{\xi})\rangle_{\boldsymbol{\xi}} + \frac{\alpha}{2} - \frac{1}{\beta}\log 2 \tag{1.71}$$

In order to discriminate between stable and non-stable, to get the pure state, we have to select the solutions of the self consistency equation (1.70) that minimizes the free energy (1.71), from the study of the spectrum of the Hessian matrix.

In this case, we can show that if $\beta < 1$, the only solution to the saddle-point equations is $m^\mu = 0$, which corresponds to a paramagnetic phase, while for $\beta > 1$, non-trivial solutions of (1.70) do exist. The solutions that are global minima of the free-energy are the states where the magnetization over one pattern m^μ is non-zero while the others are zero, which correspond exactly to the thermodynamic states where one retrieves the μ-th pattern. Dealing with the case of $P = \alpha N$, for α finite is considerably harder. It can be however treated through a calculation similar to that we will see in Section 1.2 using the replica trick [24]. In this case, the system has a ferromagnetic phase, where it retrieves one of the patterns, a paramagnetic phase and a spin glass phase. In the solution of Amit et al. [24], as in the Sherrington-Kirkpatrick model, one needs to make a replica-symmetric hypothesis to solve the saddle-point equations of the problem. In the case of the Hopfield model, for all but the very lowest temperatures the replica-symmetric solution yields the correct expression of the free energy. The phase diagram of the model is depicted in Fig. 1.2. In the next chapter we are gone to see how to deal to a different choice of the

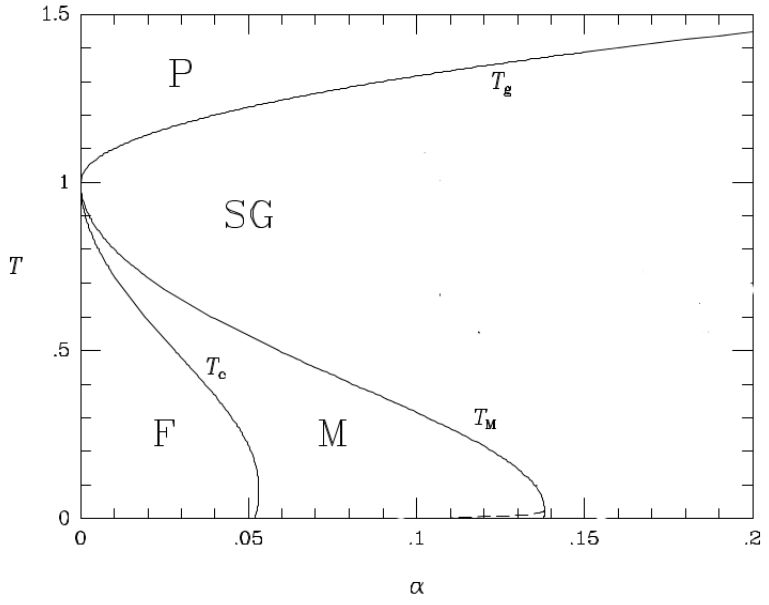


Figure 1.2: Phase diagram of the Hopfield model. P: paramagnetic state; SG: spin glass phase; M: mixed phase; F: retrieval state. Figure from [10].

distribution of patterns in which the entries are biased and diluted giving rise to a not trivial topology of the interactions network.

Chapter 2

Hopfield model with gaussian and diluted patterns

In this chapter we consider a generalization of the Hopfield model, introduced previously in section 1.3, where the entries of patterns are Gaussian and diluted. More precisely, patterns entries are drawn from a normal distribution $\mathcal{N}[0,1]$ with a probability $(1+a)/2$ or set equal to zero with probability $(1-a)/2$, where $a \in [-1,+1]$ is a tunable parameter controlling the degree of dilution of patterns. We focus on the high-storage limit, namely the amount of patterns L is linearly diverging with the system size N, i.e. $L = \alpha N$.

In the first part (section 2.1) we investigate analytically the topological properties of the emergent network. We find that, by properly tuning the dilution in the pattern entries, the network can recover different topological regimes characterized by peculiar scalings of the average coordination number with respect to the system size. Moreover, even if the network is very sparse, it turns out to display a large degree of cliquishness due to the Hebbian rule underlying its couplings. The coupling distribution is also explicitly calculated and shown to be central and with extensive variance, as expected.

In the second part (section 2.2) we investigate the thermodynamic properties of the model. We obtain explicitly the replica symmetric free-energy and the self-consistence equations for the overlaps (order parameters of the theory), which turn out to be a weighted sums of "sub-overlaps" defined on all possible sub-graphs. Finally, a study of criticality is performed through a small-overlap expansion of the self-consistencies and through a whole fluctuation theory developed for their rescaled correlations: both approaches show that the net effect of dilution in pattern entries is to rescale accordingly the critical noise level at which ergodicity breaks down. The study has been published in [5].

The model

Given N Ising spins $\sigma_i = \pm 1$, $i \in (1,...,N)$, we aim to study a mean-field model whose Hamiltonian has the form

$$\tilde{H} = -\frac{1}{D}\sum_{ij}^{N} J_{ij}\sigma_i\sigma_j \,, \tag{2.1}$$

where the couplings are built in a Hebbian fashion [25][23] as

$$J_{ij} = \sum_{\mu=1}^{L} \xi_i^\mu \xi_j^\mu, \tag{2.2}$$

and D is a denominator whose specific form is discussed in Section 3. In fact, in general, as the coordination number may vary sensibly according to the definition of patterns ξ, in order to ensure a proper linear scaling of the Hamiltonian (2.1) with the volume, D has to be a function of the system size N and of the parameters through which patterns ξ are defined.

We consider the high-storage regime [10], such that, in the thermodynamic limit (i.e. $N \to \infty$), the following scaling for the amount of stored memories (patterns) is assumed

$$\lim_{N\to\infty} \frac{L}{N} = \alpha \in \mathbb{R}^+, \tag{2.3}$$

even though we use the symbol α for the ratio between the number of patterns and the system size also at finite N, bearing in mind that the thermodynamic limit has to be performed eventually.

The quenched entries of the memories ξ_i^μ are Gaussian and diluted, namely they are set to zero with probability $(1-a)/2$, while, with probability $(1+a)/2$, they are drawn from a standard Gaussian distribution:

$$P(\xi_i^\mu) = \left(\frac{1-a}{2}\right)\delta(\xi_i^\mu) + \left(\frac{1+a}{2}\right)\mathcal{N}_{[0,1]}(\xi_i^\mu)\,. \tag{2.4}$$

The parameter a can in principle be varied in the range $a \in [-1,1]$, and, in general, small values correspond to highly diluted regimes. As proved in Section 2.1.2, a scaling law for this parameter has to be introduced in order to avoid the topology to become trivial in the thermodynamic limit. Thus, we consider the following scaling

$$a = -1 + \frac{\gamma}{N^\theta}\,, \tag{2.5}$$

where θ determines the topological regime of the network, while γ plays the role of a fine tuning within it. Notice that fixing $\theta = 0$ and $\gamma = 2$ yields to $a = 1$, corresponding to the standard analogical Hopfield model [26]. In general, due to constraint on a, $\gamma \in [0, 2N^\theta]$, and, of course, for $\gamma = 0$ we get $P(\xi_i^\mu) = \delta(\xi_i^\mu)$, that is, there is no network, so we discuss only the case $\gamma > 0$.

2.1 Topological analysis

2.1.1 Coupling distribution

Let us consider the definition of the coupling strength in Equation (2.2): the probability p that the μ-th term $\xi_i^\mu \xi_j^\mu$ is zero corresponds to the probability that at least one between ξ_i^μ and ξ_j^μ is zero, which is

$$p \equiv \left(\frac{1-a}{2}\right)^2 + 2\left(\frac{1-a}{2}\right)\left(\frac{1+a}{2}\right) = \frac{3-a^2-2a}{4}\,, \tag{2.6}$$

while its complement is the probability that a Gaussian number is drawn for both entries, that is $(1-p) = [(1+a)/2]^2$. Thus, the probability that the link connecting i and j has

strength J_{ij} can be written as

$$P(J_{ij}) = p^L\delta(J_{ij}) + \sum_{k=1}^{L} p^{L-k}(1-p)^k \binom{L}{k} P_k\Big(\sum_{\nu}^{k} \xi_i^\nu \xi_j^\nu = J_{ij}\Big) = \tag{2.7}$$

$$= p^L\delta(J_{ij}) + \sum_{k=1}^{L} f(k)\ P_k\Big(\sum_{\nu}^{k} \xi_i^\nu \xi_j^\nu - J_{ij}\Big) =$$

$$= p^L\delta(J_{ij}) + \sum_{k=1}^{L} f(k)\ \int_{-\infty}^{+\infty} \frac{dl}{2\pi} \frac{e^{-ilJ_{ij}}}{(1+l^2)^{k/2}}\,, \tag{2.8}$$

where $P_k\big(\sum_\nu^k \xi_i^\nu \xi_j^\nu = J_{ij}\big)$ is the probability that k pairs of Gaussian entries, pairwise multiplied, sum up to J_{ij}, namely

$$P_k\Big(\sum_{\nu}^{k} \xi_i^\nu \xi_j^\nu = J_{ij}\Big) = \int_{-\infty}^{\infty} \prod_{\nu=0}^{k} d\xi_i^\nu d\xi_j^\nu\ P(\xi_i^\nu)P(\xi_j^\nu)\,\delta\Big(\sum_{\nu=0}^{k} \xi_i^\nu \xi_j^\nu - J_{ij}\Big) =$$

$$= \int_{-\infty}^{+\infty} \frac{dl}{2\pi} \prod_{\nu=0}^{k} d\xi_i^\nu d\xi_j^\nu\ \frac{e^{-\frac{(\xi_i^\nu)^2}{2}}}{\sqrt{2\pi}} \frac{e^{-\frac{(\xi_i^\nu)^2}{2}}}{\sqrt{2\pi}}\ e^{il\left(\xi_i^\nu \xi_j^\nu - J_{ij}\right)} = \int_{-\infty}^{+\infty} \frac{dl}{2\pi} \frac{e^{-ilJ_{ij}}}{(1+l^2)^{k/2}}\,. \tag{2.9}$$

From Equation (2.8) one can easily specify the characteristic function of the coupling distribution

$$F(l) \equiv \int_{-\infty}^{+\infty} e^{ilJ} P(J) dJ = p^L + \sum_{k=1}^{L} \frac{f(k)}{(1+l^2)^{k/2}}\,, \tag{2.10}$$

where we dropped the indices i and j, due to the arbitrariness of the couple of nodes considered. From $F(l)$ it is possible to obtain all the momenta by simple differentiation. For instance, first and second moments read respectively as

$$\mathbb{E}[J_{ij}] = (-i)\frac{\partial F(l)}{\partial l}\bigg|_{l=0} = 0, \tag{2.11}$$

$$\mathbb{E}[J_{ij}^2] = (-i)^2\frac{\partial^2 F(l)}{\partial l^2}\bigg|_{l=0} = L(1-p) = L\left(\frac{1+a}{2}\right) = \frac{\alpha\gamma}{2}N^{1-\theta}\,. \tag{2.12}$$

Now, for fixed a and α, we expect that J, being a sum of Gaussian variables, is also normally distributes (except the point $J = 0$), at least for large N. Indeed, numerical simulations confirm that the distribution $P(J)$ converges in the thermodynamic limit ($L \to \infty$) to a Gaussian distribution with zero mean and variance given by Equation (2.12) (see Figure 2.1), except for the point $J = 0$ which will be discussed in the following section.

2.1.2 Link Probability and topology regimes

Let us consider the bare topology. The quantity of interest is the average link probability P_{link}:

$$P_{\text{link}} = 1 - P(J = 0)\,. \tag{2.13}$$

Looking at Equation (2.7), in principle $P(J = 0)$ has two contributions: one from the delta function and one from the sum over k random numbers, but the latter has a null measure in the limit $L \to \infty$. To show this, we consider the second term in Equation (2.7)

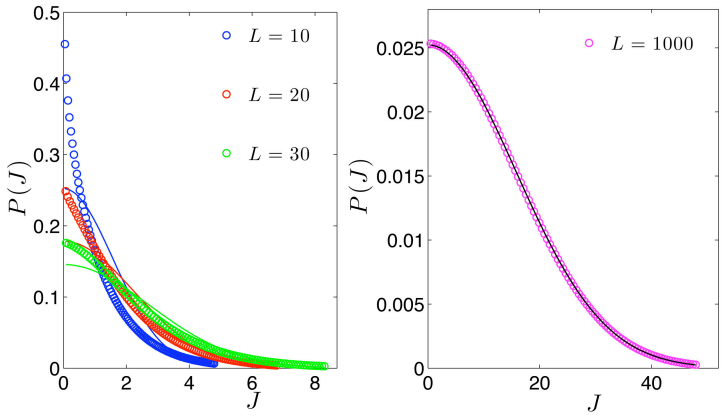


Figure 2.1: Coupling distribution $P(J)$ for $L = 10$ (left panel, blue), $L = 20$ (left panel, red), $L = 30$ (left panel, green) and $L = 1000$ (right panel). Circles represent the coupling distribution calculated according to Equation2.7, while the continuous lines represent normal distributions with momenta given by Equation2.11 and Equation2.12: as the thermodynamic limit is approached, the agreement gets better and better. Notice that, when L grows, the divergence in $J = 0$ becomes weaker. Only the positive values of J are considered due to the symmetry.

and we calculate its measure over the interval $J_{ij} \in [-\epsilon, +\epsilon]$, highlighting for clarity the term $k = 1$:

$$\int_{-\epsilon}^{\epsilon} dJ_{ij} \sum_{k=1}^{L} f(k)\ P_k(\sum_{\nu}^{k} \xi_i^\nu \xi_j^\nu - J_{ij}) =$$

$$= p^{L-1}\ (1-p) \int_{-\epsilon}^{\epsilon} P_1(r)dr\ +\ \sum_{k=2}^{L} \binom{L}{k} p^{L-k}\ (1-p)^k \int_{-\epsilon}^{\epsilon} P_k(r)dr\,. \qquad (2.14)$$

In the first term $P_1(r)$ has a weak divergence in $r = 0$ and its integral scale as $\sim \epsilon \log(\epsilon)$, so that the divergence is suppressed by the prefactor in the limit $L \to \infty$. As for $P_{k>1}(r)$, its integral is non-diverging and can be upper bounded [1] to show that the second term is negligible in the limit $L \to \infty$.

Hence, in the thermodynamic limit, $J_{ij} = 0$ only if, in $\xi_i^\mu \xi_j^\mu = 0$, for any μ, namely

$$P(J_{ij} = 0) = p^L = \left(\frac{3 - a^2 - 2a}{4}\right)^L . \qquad (2.15)$$

Now, looking at (2.13) and (2.15) in the thermodynamic limit, it is clear that, if we consider a as finite and constant, only two trivial topologies can be realized. In fact, if $a = -1$, P_{link} is zero and the system is fully disconnected, while, with $a > -1$, P_{link} tends to one exponentially fast with the system size, and the graph becomes fully connected.

Nevertheless, with the scaling (2.5),

$$P_{\text{link}} = 1 - \left(1 - \frac{\gamma^2}{4N^{2\theta}}\right)^{\alpha N} \simeq 1 - e^{-\frac{\alpha\gamma^2}{4N^{2\theta-1}}}, \qquad (2.16)$$

[1] from the the two inequalities $\int_{-\epsilon}^{\epsilon} P_k(r)dr < 2\epsilon P_k(0)$, $P_k(0) < \mathrm{C}\frac{\log(k)}{k}$ (with C a constant) follows

$$\tilde{P}_2 = \sum_{k=2}^{L} \binom{L}{k} p^{L-k}(1-p)^k \int_{-\epsilon}^{\epsilon} P_k(r)dr < 2\epsilon \mathrm{C} \sum_{k=2}^{L} \binom{L}{k} p^{L-k}\ (1-p)^k \frac{\log(k)}{k}$$

that goes to zero in the $L \to \infty$ limit .

where the last expression holds for large N. Now, by tuning the value of θ, we realize different topological regimes; within each regime the parameter γ acts as a fine tuning. Following a mean-field approach we can distinguish:

- $\underline{\theta = 0}$: $P_{\text{link}} = 1 - e^{-\frac{\alpha\gamma^2}{4}N} \to 1$. Fully Connected graph, with average degree equal to the system size ($\bar{z} = N - 1$).
  The coupling distribution converges to the Gaussian one with $\text{Var}[J] \propto N$, as in the with Sherrington-Kirkpatrick model.

- $\underline{0 < \theta < 1/2}$: $P_{\text{link}} = 1 - e^{-\frac{\alpha\gamma^2}{4}N^k} \to 1$ (where $0 < k < 1$). Fully Connected graph, with average degree equal to the system size ($\bar{z} = N - 1$).
  The coupling distribution converges to the Gaussian one with $\text{Var}[J] \propto N^q$, where $0 < q < 1$.

- $\underline{\theta = 1/2}$: $P_{\text{link}} \simeq \frac{\alpha\gamma^2}{2} = \text{const}$. The link probability is finite and the average coordination number is linearly diverging with the system size, namely $\bar{z} = \alpha\gamma^2(N-1) = \mathcal{O}(N)$.

- $\underline{1/2 < \theta < 1}$: $P_{\text{link}} = 1 - e^{-\frac{\alpha\gamma^2}{4}N^{-k}} \simeq \frac{\alpha\gamma^2}{4N^k} \to 0$ (where $0 < k < 1$). Extreme Diluted Graph, characterized by a sublinearly diverging average coordination number, $\bar{z} = \mathcal{O}(N^{1-k})$.

- $\underline{\theta = 1}$: $P_{\text{link}} = 1 - e^{-\frac{\alpha\gamma^2}{4N}} \simeq \frac{\alpha\gamma^2}{4N} \to 0$. Finite Coordination Regime with $\bar{z} = \alpha\gamma^2/4$. When $\alpha\gamma^2 > 4 (< 4)$, then $z > 1 (< 1)$ and the graph, within a mean-field approach, is expected to be overpercolated (underpercolated).

- $\underline{\theta > 1}$: $P_{\text{link}} = 1 - e^{-\frac{\alpha\gamma^2}{4}N^{-k}} \simeq \frac{\alpha\gamma^2}{4N^k} \to 0$ (where $k > 1$). Fully Disconnected Regime with coordination number vanishing for any choice of α and γ.

Finally, we notice that, recalling the variance for the coupling distribution $\text{Var}[J] = \mathbb{E}[J_{ij}^2] = \frac{\alpha\gamma}{2}N^{1-\theta}$, we have that it is vanishing - and correspondingly $P(J) = \delta(J)$ - for $\theta > 1$, consistently with the picture above. A contour plot of P_{link} as a function of γ and θ is shown in Figure 2.2.

2.1.3 Small-world properties

Small-world networks are characterized by two main properties: a small diameter and a large clustering coefficient, namely, the average shortest path length scales logarithmical (or even slower) with the system size and they contain more cliques than what expected by random chance [27]. The small-world property has been observed in a variety of real networks, including biological and technological ones [28].

First, we checked that, in the overpercolated regime, the structures considered here display a diameter growing logarithmically with N, as typical for random networks [29].

As for the clustering coefficient C, it is basically defined as the likelihood that two neighbors of a node are linked themselves, that is, for the i-th node,

$$c_i = \frac{2E_i}{z_i(z_i - 1)} , \tag{2.17}$$

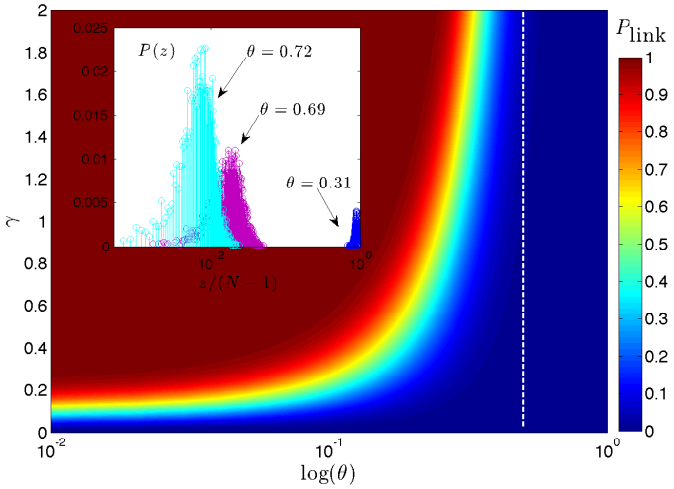


Figure 2.2: The main figure represents the contour plot of P_{link} (see Equation 2.16) as a function of γ and of $\log\theta$. The dashed, vertical line corresponds to $\theta = 0.5$ and demarcates the onset of a disconnected regime. The inset represents the degree distribution $P(z)$ as a function of the normalized number of nearest neighbors; three values of θ are considered as specified. Notice that, as expected, larger values of θ yields to sparser graphs. Both figures refer to systems made up of $N = 6000$ nodes, with $\alpha = 0.05$ and $\gamma = 1$.

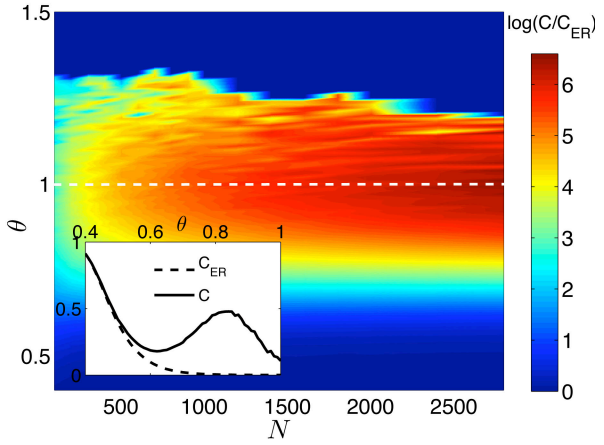


Figure 2.3: Inset: Mean clustering coefficient C (continuous line) for different choices of the parameter θ, while $N = 1600$, $\gamma = 2$ and $\alpha = 0.5$ are kept fixed; the mean has been performed over all nodes making up the graph and over 10^2 realizations. An analogous ER random graph is also considered and the related clustering coefficient C_{ER} (dashed line) is shown for comparison. Notice the qualitative different behaviors of C and C_{ER}. Main figure: contour plot for the logarithm of the ratio C/C_{ER}, as a function of N and θ. Notice that, although for θ close to 1 both graphs are sparse, $C \gg C_{ER}$. On the other hand, for $\theta > 1$, both coefficient converge to zero, in the thermodynamic limit; the non-null values appearing in the figure are due to finite-size effects.

where z_i is the number of nearest neighbors of i and E_i is the number of links connecting any couple of neighbors; when E_i equals its upper bound $z_i(z_i-1)/2$ the neighborhood of i is fully connected. The global clustering coefficient then reads as

$$C = \frac{1}{N}\sum_{i}^{N} c_i. \tag{2.18}$$

A clustering coefficient close to 1 means that the graph displays a high *cliquishness*, while a value close to 0 means that there are few triangles.

It is easy to see that for the Erdős-Rényi graph, where each link is independently drawn with a probability P, the average clustering coefficient is $C_{\rm ER} = P$. Therefore, for our network, we measure C and we compare it with the average link probability $P_{\rm link}$; results obtained for different choices of θ are shown in Figure 2.3.

First, we notice that, for a given system size N, the behavior of C and of $C_{\rm ER}$, with respect to θ, is markedly different (see the inset): the latter decreases monotonically due to the analogous decrease of the link probability, while the former exhibits two extremal points at a relatively large degrees of dilution. In fact, as long as the networks are highly connected, the disappearance of a few links yields, in both cases, a modest drop in the overall cliquishness. On the other hand, when dilution is significant, the intrinsic structure of the "Hebbian graph" matters: as patterns get sparser and sparser, surviving links are those connecting nodes whose related patterns display matching non-null entries. In this way, the neighbors of a node are also likely to be connected [30, 31, 32] and the clustering coefficient grows. Finally, at a very large degree of dilution, the system approaches the fully-disconnected regime and the clustering coefficient decreases.

In order to compare more effectively our graph and an analogous ER graph, we also considered the ratio $C/C_{\rm ER}$ (see the main figure). Interestingly, for θ relatively large, as N gets larger this ratio grows confirming that the few links remaining are very effective in maintaining the cliques. This can be understood as follows: to fix ideas let us take $\theta = 1$, so that the average number of non-null entries in a string is $L\gamma/(2N)$ which equals $\gamma\alpha/2$ in the high-storage regime under investigation. For simplicity, let also assume that $\gamma\alpha/2 \approx 1$, and that this holds with vanishing variance for all nodes. Therefore, if the node i has k neighbors, its (local) clustering coefficient is either 0 (if $k \leq 2$) or 1 (if $k > 2$). Hence, the expected local clustering coefficient can be estimated as the probability for a node to display $k > 2$ nearest neighbors, namely $1-(1-q)^{N-1} - Nq(1-q)^{N-2}$, where $q = \gamma/(2N)$ is the probability that the pattern of an arbitrary node $j \neq i$ has the non-null entry matching with the one of ξ_i. With some algebra we get $c_i \approx 1 - e^{-\gamma/2}(1+\gamma/2)$, which remains finite also in the thermodynamic limit, in agreement with results from simulations. For $\theta > 1$, $q = \gamma N^{1-\theta}/2$ and c_i converges to zero.

2.2 The statistical mechanics analysis

In this section, we study the thermodynamic properties of the system introduced: At first, we show its equivalence to a bipartite spin-glass and figure out the order parameters of the theory, then we define an interpolating free-energy which generalizes the multiple stochastic stability developed in [26]; this technique allows to obtain the replica-symmetric solution in form of a simple sum rule. As a last step, we extremize the free energy finding self-consistencies for the order parameters, whose critical behavior is also addressed.

2.2.1 The equivalent diluted bipartite spin-glass

As we deal with a structure whose average coordination number may range in $[0, N]$, from a statistical mechanics perspective, we aim to define the normalization constant D for the Hamiltonian in Equation (2.1), in such a way that its average (which defines the extensive energy of the system and is denoted symbolically with the brackets) is linearly diverging with the system size, namely $\langle \tilde{H} \rangle \propto N$.

By a direct calculation, it is possible to show that this condition is fulfilled by

$$D = N^{1-\theta}, \tag{2.19}$$

so that, using the explicit definition for the couplings, we can write

$$\tilde{H} = -\frac{1}{N^{1-\theta}} \sum_{i<j}^{N} \sum_{\mu}^{L} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j = -\frac{1}{N^{1-\theta}} \sum_{i<j}^{N} J_{ij} \sigma_i \sigma_j \,, \tag{2.20}$$

For a single realization of the disorder encoded in the memories, the partition function reads off as:

$$\tilde{Z}_{N,L}(\beta; \xi) = \sum_{\{\sigma\}} \exp \left\{ \frac{\beta}{2N^{1-\theta}} \sum_{i,j}^{N} \sum_{\mu}^{L} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j \right\} . \tag{2.21}$$

Note that, as usual in the Hopfield model, the diagonal term gives an extensive contribution to the partition function. In the above expression, we neglected this diagonal term, directly by adding it as a term $\frac{\alpha\beta}{2}(\frac{1+a}{2}) = \frac{\alpha\beta\gamma}{4N^\theta}$ to the final expression of the free energy [33] (see Equation (2.41)).

Now, we can introduce another party made up of L *soft* spins $\{z_\mu\}$, namely i.i.d. variables with an intrinsic standard Gaussian distribution $\mathcal{N}[0, 1]$, that interact only with the original party of dichotomic spins $\{\sigma_i\}$ via the couplings $\{\xi_i^\mu\}$; the related partition function is

$$Z_{N,L}(\beta; \xi) = \sum_{\{\sigma\}} \int \prod_{\mu} d\mu(z_\mu) \exp \left\{ \sqrt{\frac{\beta}{N^{1-\theta}}} \sum_{i}^{N} \sum_{\mu}^{L} \xi_i^\mu \sigma_i z_\mu \right\} , \tag{2.22}$$

with $d\mu(z_\mu)$ standard Gaussian measure for all the z_μ. By applying Gaussian integrations as usual [31], it is easy to see that $\tilde{Z}_{N,L}(\beta; \xi)$ and $Z_{N,L}(\beta; \xi)$ are thermodynamically equivalent. The advantage of the expression (2.22) is that it is linear with respect to the memories ξ_i^μ so that the bare topology is simply that of a bipartite random graph with link probability $p_{\text{link}} = (1 + a)/2$, like in [30].

Taken O as a generic observable, depending on the spin configurations $\{\sigma, z\}$, we define the Boltzmann state $\omega_\beta(O)$ at a given level of (fast) noise β as

$$\omega_\beta(O) = Z_{N,L}(\beta; \xi)^{-1} \sum_{\{\sigma\}} \int \prod_{\mu} d\mu(z_\mu) O(\{\sigma, z\}) e^{\sqrt{\frac{\beta}{N^{1-\theta}}} \sum_i^N \sum_\mu^L \xi_i^\mu \sigma_i z_\mu}, \tag{2.23}$$

and we introduce a product space on several replicas of the system as $\Omega_s = \omega_1 \bigotimes \omega_2 \bigotimes ... \bigotimes \omega_s$ [26].

For a generic function of the memories $F(\xi)$, the quenched average will be defined by the symbol $\mathbb{E}$ and performed in two steps: first, we fix the number l of links between the two parties and we perform the average over the Gaussian distribution of the memories:

$$\mathbb{E}_\xi^{(l)}[F(\xi)] \equiv \int_{-\infty}^{+\infty} \prod_{(i,\mu)=1}^{l} \frac{d\xi_i^\mu}{\sqrt{2\pi}} e^{-\frac{(\xi_i^\mu)^2}{2}} F(\xi_i^\mu) = \int F(\xi)\, d\mu_l(\xi) \equiv f(l)\,; \tag{2.24}$$

then, we perform the average over the binomial distribution for the number of links,

$$\mathbb{E}_l[f(l)] \equiv \sum_{l=0}^{NL} \binom{NL}{l} \left(\frac{1+a}{2}\right)^l \left(\frac{1-a}{2}\right)^{(NL-l)} f(l)\,, \tag{2.25}$$

so that $\mathbb{E} \equiv \mathbb{E}_l \mathbb{E}_\xi^{(l)}$. Indeed, for example, $\mathbb{E}[\xi_i^\mu] = 0$ and $\mathbb{E}[(\xi_i^\mu)^2] = (1+a)/2$.

Moreover, as we will see, for a natural introduction of the order parameters, it is useful to define the number of links, l, as the product of two independents binomial variables

$$l \doteq l_\eta l_\chi\,, \tag{2.26}$$

where the symbol $\doteq$ stands for the equality in distribution and

$$P(l_\eta) = \binom{N}{l_\eta} \sqrt{\frac{1+a}{2}}^{l_\eta} \sqrt{\frac{1-a}{2}}^{N-l_\eta}, \tag{2.27}$$

$$P(l_\chi) = \binom{L}{l_\chi} \sqrt{\frac{1+a}{2}}^{l_\chi} \sqrt{\frac{1-a}{2}}^{L-l_\chi}. \tag{2.28}$$

Of course, a product of two binomial variables is not a binomial variable itself, so at finite size this definition is not consistent; nevertheless, in the thermodynamic limit, the central limit theorem ensures that only the first two momenta of the distributions survive so that the definitions become consistent.

We also use the symbol $\langle\cdot\rangle$ to mean $\langle\cdot\rangle = \mathbb{E}\Omega(\cdot)$ and $\langle\cdot\rangle_G = \mathbb{E}_\xi^{(l)}\Omega(\cdot)$.

The main thermodynamical quantity of interest is the intensive pressure defined as

$$A(\alpha,\beta) = \lim_{N\to\infty} A_N(\alpha,\beta) = -\beta f(\alpha,\beta) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\log Z_{N,L}(\beta,\xi)\,. \tag{2.29}$$

Here $f(\alpha,\beta) = u(\alpha,\beta) - \beta^{-1}s(\alpha,\beta)$ is the free-energy density, $u(\alpha,\beta)$ the internal energy density and $s(\alpha,\beta)$ the entropy density.

Finally, we define two infinite (in the thermodynamic limit) sets of order parameters, the restricted overlaps, as

$$q_{12}^{l_\eta} = \frac{1}{l_\eta}\sum_i^{l_\eta} \sigma_i^1\sigma_i^2,$$

$$p_{12}^{l_\chi} = \frac{1}{l_\chi}\sum_\mu^{l_\chi} z_\mu^1 z_\mu^2\,, \tag{2.30}$$

which define the overlaps (restricted on sub-networks) between two replicas made up by parties with l_η and l_χ nodes, respectively.

2.2.2 Free energy interpolation and general strategy

In what follows, we assume that no real external fields (as magnetic inputs or partial information submission for retrieval) act on the network, but fields insisting on each spin are strictly generated by other spins. Thus, the overall field felt by an element of a given party is the sum (weighted through the couplings), of the states of the spins in the other party. Note that spins are connected in loops using the other party as a mirror, therefore, the equivalent analogical neural network is a recurrent network.

In this section, we show that the free-energy can be calculated in specific cases (e.g. at the replica symmetrical level) by using a novel technique that has been developed in [26] for fully connected spin-glass models and extended in [30] to diluted ferromagnetic models. This technique introduces an external field acting on the system which "imitates" the internal, recurrently generated input, by reproducing its average statistics. While the external, fictitious input does not reproduce the statistics of order two and higher, it represents correctly the averages. These external inputs are denoted as η and χ (one for each spin in each party) and are distributed following the Gaussian distributions with zero mean and whose variances scale according to the underlying topology (as a function of α, θ, γ) and coherently approaches zero when the network topology disappears.

In order to recover the second order statistics, the free-energy is interpolated smoothly between the case in which all fields are external, and all high order statistics is missing, and the case in which all fields are internal, describing the original network: Following the original Guerra's schemes [15, 26, 34, 35], this allows a powerful sum rule. We use an interpolating parameter $t \in [0, 1]$ for this morphing, such that for $t = 0$ the fields are all external and the calculation straightforward, while for $t = 1$ the original model is fully recovered.

In what follows, for the sake of clearness, we allow ourselves in writing $A = A_N(\alpha, \beta)$ even though α should be introduced only once the thermodynamic limit has been performed. The interpolating quenched pressure $\tilde{A}_N(\alpha, \beta, t)$ at finite N is then defined as

$$\tilde{A}_N(\alpha, \beta; t) = \frac{1}{N}\mathbb{E} \log \sum_{\{\sigma\}} \int \prod_{\mu} d\mu(z_\mu) \exp\left(\sqrt{t}\sqrt{\frac{\beta}{N^{1-\theta}}} \sum_{i,\mu}^{N,L} \xi_i^\mu \sigma_i z_\mu\right) \cdot$$
$$\cdot \exp\left(b\sqrt{1-t}\sum_{i}^{N} \sigma_i \eta_i + c\sqrt{1-t}\sum_{\mu}^{L} z_\mu \chi_\mu + \frac{d}{2}(1-t)\sum_{\mu}^{L} z_\mu^2 \theta_\mu\right). \qquad (2.31)$$

We assume trough the paper that the limit $A(\alpha, \beta) = \lim_{N\to\infty} A_N(\alpha, \beta) = \lim_{N\to\infty} A_N(\alpha, \beta, t = 1)$ exists. The "interpolating fields" distributions are chosen to mimic the local fields behavior, so that η_i, χ_μ and θ_μ have zero value with probability $\sqrt{(1-a)/2}$, while, with probability $\sqrt{(1+a)/2}$, are normally distributed, except for θ_μ which assumes value one [2]. So the number of active fields follows Equation (2.27). While the constant b, c, d have to be chosen properly, as shown in the following.

The strategy for the evaluation of the pressure of the original model, $\tilde{A}(\alpha, \beta, t = 1)$, is to compute the t-streaming of it, namely $\partial_t \tilde{A}(\alpha, \beta, t = 1)$, and use the fundamental theorem of calculus to obtain

$$A_N(\alpha, \beta) = \tilde{A}_N(\alpha, \beta; t = 1) = \tilde{A}_N(\alpha, \beta, t = 0) + \int_0^1 dt' \left(\frac{d}{dt}\tilde{A}_N(\alpha, \beta; t)\right)_{t=t'} \qquad (2.32)$$

When evaluating the streaming $\partial_t \tilde{A}$, we get the sum of four terms, each comes as a consequence of the derivation of a corresponding exponential term appearing into Equation (2.31). The explicit computation is reported in the appendix A.1, the final expression for

[2] Indeed, the presence the field θ_μ has much less physical meaning but simplifies the calculations.

the t streaming is

$$\begin{aligned}\frac{d\tilde{A}_N(\alpha,\beta,t)}{dt} &= \left[\frac{\alpha\beta}{2}N^\theta\left(\frac{1+a}{2}\right)-\frac{\alpha c^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}-\frac{\alpha d}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\right]\langle z^2\rangle + \\ &-\frac{\alpha\beta}{2}N^\theta\left(\frac{1+a}{2}\right)\langle q_{12}^{l_\eta}p_{12}^{l_\chi}\rangle+\frac{b^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\langle q_{12}^{l_\eta}\rangle+\frac{c^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\langle p_{12}^{l_\chi}\rangle + \\ &-\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\frac{b^2}{2}\,. \end{aligned} \tag{2.33}$$

2.2.3 Replica symmetric approximation and fluctuation source

As it is, this streaming encodes the whole full replica-symmetry-breaking complexity [36, 37] of the underlying glassy phase and it is intractable. Our plan is to split this derivative in two terms, one dealing with the averages of the order parameters and one accounting for their fluctuations. To this aim, we introduce the *source of fluctuations*, $S(\alpha,\beta,t)$, as

$$S(\alpha,\beta,t)=-\frac{\alpha\beta}{2}\left(\frac{1+a}{2}\right)N^\theta\langle(q_{12}^{l_\eta}-\overline{q})(p_{12}^{l_\chi}-\overline{p})\rangle, \tag{2.34}$$

with

$$\overline{q}\equiv\mathbb{E}_{l_\eta}\overline{q}^{l_\eta},\quad \overline{p}\equiv\mathbb{E}_{l_\eta}\overline{p}^{l_\eta}\,. \tag{2.35}$$

Notice that the main order parameters $\overline{q}$ and $\overline{p}$ sum every overlap, each with its relative weight, on every possible subnetwork of the whole network according to the approaches [38, 39] and that they recover the standard order parameters of the Hopfield model when dilution is neglected [10, 26].

In order to relate Equation (2.34) to Equation (A.5), let us remember that we still have free parameters that can be chosen as [3]

$$\begin{aligned} b &= \sqrt{\alpha\beta\left(\frac{1+a}{2}\right)^{\frac{1}{2}}N^\theta\overline{p}}=\sqrt{\alpha\beta\overline{p}}\left(\frac{\gamma}{2}\right)^{\frac{1}{4}}N^{\frac{\theta}{4}}, \\ c &= \sqrt{\beta\left(\frac{1+a}{2}\right)^{\frac{1}{2}}N^\theta\overline{q}}=\sqrt{\beta\overline{q}}\left(\frac{\gamma}{2}\right)^{\frac{1}{4}}N^{\frac{\theta}{4}}, \\ d &= \beta N^\theta\left(\frac{1+a}{2}\right)^{\frac{1}{2}}(1-\overline{q})=\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})\,, \end{aligned} \tag{2.36}$$

so to have

$$\frac{d\tilde{A}(\alpha,\beta;t)}{dt}=S(\alpha,\beta,t)+\frac{\alpha\beta}{2}\left(\frac{1+a}{2}\right)N^\theta\,\overline{p}(\overline{q}-1)=S(\alpha,\beta,t)+\frac{\alpha\beta\gamma}{4}\overline{p}(\overline{q}-1)\,. \tag{2.37}$$

In the replica symmetric approximation, the order parameters do not fluctuate with respect to their quenched average in the thermodynamic limit as they get delta-distributed over their replica symmetric averages $\overline{q}$, $\overline{p}$, which have been denoted with a bar. As a consequence, within this approximation, we can neglect the fluctuation source term $S(\alpha,\beta,t)$ and keep only the replica symmetric overlap averages in the expression (2.37) such that its integration is trivially reduced to a multiplication by one.

[3] in particular we choose d to cancel the $\langle z^2\rangle$ terms appearing in the first line of equation (A.5).

In order to get an explicit expression of the sum rule (2.32), we can then proceed to analyze the starting point for the "morphing", namely $\tilde{A}(\alpha,\beta;t=0)$, which can be calculated straightforwardly as it involves only one-body interactions:

$$\begin{aligned}
&\tilde{A}_N(\alpha,\beta,t=0) = \\
&= \frac{1}{N}\mathbb{E}\log\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)\exp\left(\ b\sum_{i}^{N}\sigma_i\eta_i + c\sum_{\mu}^{L}z_\mu\chi_\mu + \frac{d}{2}\sum_{\mu}^{L}\theta_\mu z_\mu^2\right) = \\
&= \sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}\log 2 + \sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}\int d\mu(\eta)\log\cosh(\sqrt{\alpha\beta\overline{p}}\left(\frac{\gamma}{2}\right)^{\frac{1}{4}}N^{\frac{\theta}{4}}\ \eta) + \\
&+\frac{\gamma\alpha\beta\bar{q}}{4}\frac{1}{1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})} - \frac{\alpha}{2}\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{-\frac{\theta}{2}}\log\left(1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})\right),
\end{aligned} \tag{2.38}$$

where we used

$$\begin{aligned}
&\frac{1}{N}\mathbb{E}_{l_\eta}\mathbb{E}_\eta^{l_\eta}\log\sum_{\{\sigma\}}\exp\left(b\sum_{i}^{N}\sigma_i\eta_i\right) = \\
&= \frac{1}{N}\mathbb{E}_{l_\eta}\mathbb{E}_\eta^{l_\eta}\sum_{i}^{N}\log 2\cosh(b\ \eta_i) = \frac{1}{N}\mathbb{E}_{l_\eta}l_\eta\ \mathbb{E}_\eta\log 2\cosh(b\ \eta) = \\
&= \left(\frac{1+a}{2}\right)^{\frac{1}{2}}\log 2 + \left(\frac{1+a}{2}\right)^{\frac{1}{2}}\int d\mu(\eta)\log\cosh(b\ \eta),
\end{aligned} \tag{2.39}$$

and

$$\begin{aligned}
&\frac{1}{N}\mathbb{E}_{l_\chi}\mathbb{E}_\chi^{l_\chi}\log\int\prod_{\mu}d\mu(z_\mu)\exp\left(c\sum_{\mu}^{L}z_\mu\chi_\mu\right)\exp\left(\frac{d}{2}\sum_{\mu}^{L}\theta_\mu z_\mu^2\right) = \\
&= \frac{1}{N}\mathbb{E}_{l_\chi}\mathbb{E}_\chi^{l_\chi}\left(\sum_{\mu}\frac{c^2\chi_\mu^2}{2(1-d\theta_\mu)}\right) + \frac{1}{N}\mathbb{E}_{l_\chi}\mathbb{E}_\chi^{l_\chi}\frac{1}{2}\sum_{\mu}\log\left(\frac{1}{1-d\theta_\mu}\right) = \\
&= \frac{\alpha}{2}\sqrt{\left(\frac{1+a}{2}\right)}\left(\frac{c^2\sigma_\chi^2}{(1-d)}\right) - \frac{\alpha}{2}\sqrt{\left(\frac{1+a}{2}\right)}\log(1-d) = \\
&= \frac{\gamma\alpha\beta\bar{q}}{4}\frac{1}{1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})} - \frac{\alpha}{2}\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{-\frac{\theta}{2}}\log\left(1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})\right)
\end{aligned} \tag{2.40}$$

Now, substituting the expression for $\tilde{A}_N(\alpha,\beta,t=0)$ of Equation(2.38) into (2.32), we obtain the replica-symmetric free energy (strictly speaking the mathematical pressure) of the network as

$$\begin{aligned}
&\tilde{A}_N^{RS}(\alpha,\beta) = \tilde{A}_N(\alpha,\beta,t=0) + \left.\frac{d\tilde{A}_N^{RS}(\alpha,\beta,t}{dt}\right|_{t=0} = \\
&= \sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}\left[\log 2 + \int d\mu(\eta)\log\cosh\left(\sqrt{\alpha\beta\overline{p}}\left(\frac{\gamma}{2}\right)^{\frac{1}{4}}N^{\frac{\theta}{4}}\eta\right)\right] + \\
&+\frac{\alpha\beta\gamma\ \bar{q}}{4}\frac{1}{1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})} - \frac{\alpha}{2}\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{-\frac{\theta}{2}}\log\left(1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})\right) + \\
&+\frac{\alpha\beta\gamma}{4}\overline{p}(\overline{q}-1) + \frac{\alpha\beta\gamma}{4}N^{-\theta}\,.
\end{aligned} \tag{2.41}$$

Despite the last expression is meant to hold in the thermodynamic limit, with a little mathematical abuse, we left the explicit dependence size to discuss some features of the solution: Equation (2.41) may look strange due to the strong presence of various powers of the volume size N, which in principle are potentially unwanted divergencies. We start noticing that, in the limit of zero dilution $\theta = 0$ and homogeneous distribution of fields $\gamma = 2$, the expression for the free-energy recovers the replica symmetric one of the analogical Hopfield model [26] (or digital one without retrieval [9]). Moreover, remembering the various topological regimes outlined in Section 3, we see that when the network changes the topological phase, for instance moving from a fully connected topology to a sparse graph the coordination number may scale with the volume size or be kept constant. These situations are deeply different from a thermodynamical viewpoint because, in order to have no negligible contributions to the free-energy, fields obtained by an extensive number of (finite) terms in the fully connected scenario must be (possibly) turned into fields obtained by a finite number of (infinite) terms in the dilute regime. As the topology changes, the fields must follow accordingly, which is equivalent to a (fast) noise rescaling with the volume size that is another standard approach to diluted network [40, 41, 42].

The physical free-energy is then obtained by extremizing this expression with respect to the order parameters; we only stress here that, as a general property of these neural networks/bipartite spin-glasses, the free-energy now obeys a min-max principle, which will not be deepened here (because it does not change the following procedure and it has been discussed in [26]). As a consequence, the following system defines the values of the overlaps (as functions of $\alpha,\ \beta$) that must be used into Equation (2.41)

$$\begin{aligned} \frac{\partial \tilde{A}}{\partial \overline{q}} &= \frac{\alpha\beta(\frac{\gamma}{2})}{2}\left(\overline{p} - \frac{(\frac{\gamma}{2})^{\frac{1}{2}} N^{\frac{\theta}{2}} \beta \overline{q}}{(1-\beta(\frac{\gamma}{2})^{\frac{1}{2}} N^{\frac{\theta}{2}}(1-\overline{q}))^2}\right) = 0 \\ \frac{\partial \tilde{A}}{\partial \overline{p}} &= \frac{\alpha\beta(\frac{\gamma}{2})}{2}\left[\overline{q} - \int d\mu(\eta) \tanh^2(\sqrt{\alpha\beta\overline{p}}\left(\frac{\gamma}{2}\right)^{\frac{1}{4}} N^{\frac{\theta}{4}}\eta)\right] = 0\,, \end{aligned} \tag{2.42}$$

by which

$$\overline{q} = \int d\mu(\eta) \tanh^2\left(\frac{\sqrt{\alpha\overline{q}}\beta(\frac{\gamma}{2})^{\frac{1}{2}} N^{\frac{\theta}{2}}}{1-\beta(\frac{\gamma}{2})^{\frac{1}{2}} N^{\frac{\theta}{2}}(1-\overline{q})}\,\eta\right)\,. \tag{2.43}$$

All the related models (e.g. Viana-Bray [43], Hopfield [23], Sherrigton-Kirkpatrick [22]) display an ergodicity breaking associated with a second order phase transition and presence of criticality. If we assume the same behavior even for the model investigated here, the self-consistence equation (2.43) can give hints on the critical line (in the parameter space) where ergodicity breaks down. In fact, when leaving the ergodic region (implicitly defined by $\overline{q} = 0,\ \overline{p} = 0$) the order parameters start growing (implicitly defining the critical line as the starting point) and, as continuity is assumed through the second order kind of transition, we can expand the r. h. s. of Equation (2.43) for low $\overline{q}$ and obtain a polynomial expression at both terms. Then, due to the principle of identity of polynomials, we can equate the two sides term by term obtaining

$$\beta_c = \frac{1}{\left(\frac{\gamma}{2}\right)^{\frac{1}{2}} N^{\frac{\theta}{2}}\,(1+\sqrt{\alpha})}\,, \tag{2.44}$$

which is the critical surface of the system.

Mirroring the discussion dealing with the free energy, we note that this result too is clearly a consequence of the choice (2.19) for the normalization factor that gives us an extensive thermodynamics. If we normalize choosing $D = N$, as it is usual in the Hopfield

model [9], we obtain (turning to $T = 1/\beta$ which is most intuitive) $T_c = N^{-\theta/2}(1 + \sqrt{\alpha})\sqrt{2/\gamma}$ (and recover the AGS line for $\theta = 0$ and $\gamma = 2$), such that the overall effect of increasing dilution is to reduce the value of the critical temperature because the couplings on average becomes weaker. In particular, in the finite connectivity regime ($\theta = 1$), the network is built of by N links instead of N^2 which, roughly speaking, implies a rescaling in the temperature proportional to $\sqrt{N}$ (coherently with a spin-glass behavior), as in the ferromagnetic counterpart its rescale is ruled by N instead of $\sqrt{N}$ [40] because the latter is a model defined through the first momentum, while the former by the variance.

Furthermore, we stress that the system displays only one critical surface splitting the ergodic region from the spin glass and there are no further "weak-transitions" for each sub-overlap, coherently with the scenario discussed in [44] for the similar case of the Viana-Bray model [43].

2.3 Fluctuation theory and critical behavior

The plan of this section is studying the regularity of the rescaled (and centered) overlap correlation functions.
The idea is as follows: If the system undergoes a second order phase transition, on the critical surface (2.44), the (extensive) fluctuations of its order parameters should diverge there, hence they should be described by meromorphic functions, whose poles detect the critical surface itself. As a consequence, an explicit knowledge of these functions would confirm (or neglect) the critical picture we obtained through the small overlap expansion of the previous section. However, obtaining them explicitly is not immediate and we sketch in what follows our strategy: At first, we define the (rescaled and centered) fluctuations of the order parameters as

$$\begin{aligned} Q_{ab}^{l_\eta} &= \sqrt{N}\left(q_{ab}^{l_\eta} - \overline{q}^{l_\eta}\right), \\ P_{ab}^{l_\eta} &= \sqrt{L}\left(p_{ab}^{l_\chi} - \overline{p}^{l_\eta}\right), \end{aligned} \tag{2.45}$$

such that, while $q_{ab}^{l_\eta} \in [-1,+1]$, $p_{ab}^{l_\eta} \in [-1,+1]$, $Q_{ab}^{l_\eta} \in \mathbb{R}$, $P_{ab}^{l_\eta} \in \mathbb{R}$, hence, the square of the latter may diverge as expected for second order phase transitions.

Nevertheless, obtaining them explicitly from the original Hamiltonian is prohibitive and we use another procedure, originally outlined in [35]: We evaluate these rescaled overlap fluctuations weighted with the non-interacting Hamiltonian in the Maxwell-Boltzman distribution, hence $\langle Q_{l_\eta 12}^2\rangle_{t=0}$, $\langle Q_{l_\eta 12}P_{l_\chi 12}\rangle_{t=0}$, $\langle P_{l_\chi 12}^2\rangle_{t=0}$, then we derive the streaming of a generic observable O (function in principle, both of the spins of the parties and of the quenched memories), namely $\partial_t\langle O\rangle_t$ such that we know how to propagate $\langle O\rangle_{(t=0)}$ up to $\langle O\rangle_{(t=1)}$ (which is our goal), and finally we use this streaming equation (which turns out to be a dynamical system) on the Cauchy problem defined by $\langle Q_{l_\eta 12}^2\rangle_{t=0}$, $\langle Q_{l_\eta 12}P_{l_\chi 12}\rangle_{t=0}$, $\langle P_{l_\chi 12}^2\rangle_{t=0}$, obtaining the attended result. Once the procedure is completed, the simple analysis of the poles of $\langle Q_{l_\eta 12}^2\rangle_{t=1}$, $\langle Q_{l_\eta 12}P_{l_\chi 12}\rangle_{t=1}$, $\langle P_{l_\chi 12}^2\rangle_{t=1}$ will identify the critical surfaces of the system.

Starting with the study of the structure of the derivative, our aim is to compute the

t-streaming for a generic observable O_s of s replicas. Calling

$$H_s = \sum_{a=1}^{s}\left\{\sqrt{t}\sqrt{\frac{\beta}{N^{1-\theta}}}\sum_{i,\mu}^{l_\eta,l_\chi}\xi_i^\mu\sigma_i^a z_\mu^a + \sqrt{1-t}\,b\sum_i^{l_\eta}\eta_i\sigma_i^a + \right.$$
$$\left. +\sqrt{1-t}\,c\sum_\mu^{l_\chi}\chi_\mu z_\mu^a + (1-t)\frac{d}{2}\theta_\mu\sum_\mu^{l_\chi}(z_\mu^a)^2\right\} \tag{2.46}$$

such that

$$\langle O\rangle_t = \frac{\int\prod_\mu^L d\mu(z_\mu)\sum_\sigma O\exp\left(-\beta H_s\right)}{\int\prod_\mu^L d\mu(z_\mu)\sum_\sigma\exp\left(-\beta H_s\right)},$$

its t-streaming is

$$\frac{d\langle O_s\rangle_t}{dt} = \frac{d}{dt}\mathbb{E}\frac{\sum_{\{\sigma\}}\int\prod_\mu d\mu(z_\mu)O_s e^{H_s}}{\sum_{\{\sigma\}}\int\prod_\mu d\mu(z_\mu)e^{H_s}} = \mathbb{E}\left[\Omega\left(O_s\frac{dH}{dt}\right)\right] - \mathbb{E}\left[\Omega(O_s)\Omega\left(\frac{dH}{dt}\right)\right]. \tag{2.47}$$

In the last equation eight terms contribute. The explicit computation is reported in the appendix A.2. Merging all the terms together we get the streaming:

$$\frac{d}{dt}\langle O_s\rangle_t = \beta\sqrt{\alpha}\frac{\gamma}{2}\left(\sum_{a<b}^{s}\langle OQ_{ab}^{l_\eta}P_{ab}^{l_\chi}\rangle_t - s\sum_a^s\langle OQ_{a,s+1}^{l_\eta}P_{a,s+1}^{l_\chi}\rangle_t + \right.$$
$$\left. + \frac{s(s+1)}{2}\langle OQ_{s+1,s+2}^{l_\eta}P_{s+1,s+2}^{l_\chi}\rangle_t\right). \tag{2.48}$$

In order to control the overlap fluctuations, namely $\langle Q_{l_\eta 12}^2\rangle_{t=1}$, $\langle Q_{l_\eta 12}P_{l_\chi 12}\rangle_{t=1}$, $\langle P_{l_\chi 12}^2\rangle_{t=1}$, ..., noting that the streaming equation pastes two replicas to the ones already involved ($s=2$ so far), we need to study nine correlation functions. It is then useful to introduce them and refer to them by capital letters so to simplify their visualization:

$$\langle Q_{l_\eta 12}^2\rangle_t = A(t),\quad \langle Q_{l_\eta 12}Q_{l_\eta 13}\rangle_t = B(t),\quad \langle Q_{l_\eta 12}Q_{l_\eta 34}\rangle_t = C(t), \tag{2.49}$$
$$\langle Q_{l_\eta 12}P_{l_\chi 12}\rangle_t = D(t),\quad \langle Q_{l_\eta 12}P_{l_\chi 13}\rangle_t = E(t),\quad \langle Q_{l_\eta 12}P_{l_\chi 34}\rangle_t = F(t), \tag{2.50}$$
$$\langle P_{l_\chi 12}^2\rangle_t = G(t),\quad \langle P_{l_\chi 12}P_{l_\chi 13}\rangle_t = H(t),\quad \langle P_{l_\chi 12}P_{l_\chi 34}\rangle_t = I(t). \tag{2.51}$$

Let us now sketch their streaming. First, we introduce the operator "dot" as

$$\dot{O} = \frac{2}{\beta\sqrt{\alpha}\gamma}\frac{dO}{dt},$$

which simplifies calculations and shifts the propagation of the streaming from $t=1$ to $t=\beta\sqrt{\alpha}\gamma/2$. Using this we sketch how to write the streaming of the first two correlations (as it works in the same way for any other):

$$\dot{A} = \langle Q_{l_\eta 12}^2 Q_{l_\eta 12}P_{l_\chi 12}\rangle_t - 4\langle Q_{l_\eta 12}^2 Q_{l_\eta 13}P_{l_\chi 13}\rangle_t + 3\langle Q_{l_\eta 12}^2 Q_{l_\eta 34}P_{l_\chi 34}\rangle_t,$$
$$\dot{B} = \langle Q_{l_\eta 12}Q_{l_\eta 13}\Big(Q_{l_\eta 12}P_{l_\chi 12} + Q_{l_\eta 13}P_{l_\chi 13} + Q_{l_\eta 23}P_{l_\chi 23}\Big)\rangle_t -$$
$$- 3\langle Q_{l_\eta 12}P_{l_\chi 13}\Big(Q_{l_\eta 14}P_{l_\chi 14} + Q_{l_\eta 24}P_{l_\chi 24} + Q_{l_\eta 34}P_{l_\chi 34}\Big)\rangle_t + 6\langle Q_{l_\eta 12}P_{l_\chi 13}Q_{l_\eta 45}P_{l_\chi 45}\rangle_t \tag{2.52}$$

By assuming a Gaussian behavior, as in the strategy outlined in [35], we can write the overall streaming of the correlation functions in the form of the following differential system

$$\begin{aligned}
\dot{A} &= 2AD - 8BE + 6CF, \\
\dot{B} &= 2AE + 2BD - 4BE - 6BF - 6EC + 12CF, \\
\dot{C} &= 2AF + 2CD + 8BE - 16BF - 16CE + 20CF, \\
\dot{D} &= AG - 4BH + 3CI + D^2 - 4E^2 + 3F^2, \\
\dot{E} &= AH + BG - 2BH - 3BI - 3CH + 6CI + 2ED - 2E^2 - 6EF + 6F^2, \\
\dot{F} &= AI + CG + 4BH - 8BI - 8CH + 10CI + 2DF + 4E^2 - 16EF + 10F^2, \\
\dot{G} &= 2GD - 8HE + 6IF, \\
\dot{H} &= 2GE + 2HD - 4HE - 6HF - 6IE + 12IF, \\
\dot{I} &= 2GF + 2DI + 8HE - 16HF - 16IE + 20IF. \qquad (2.53)
\end{aligned}$$

As we are interested in discussing criticality and not the whole glassy phase, it is possible to solve this system starting from the high noise region, once the initial conditions at $t = 0$ are known. As at $t = 0$ everything is factorized, the only needed check is by the correlations inside each party. Starting with the first party, we have to study A, B, C at $t = 0$. As only the diagonal terms give non-negligible contribution, it is immediate to work out this first set of starting points as

$$\begin{aligned}
A(0) &= \langle Q^2_{l_\eta 12}\rangle = N(\langle (q^{l_\eta}_{12})^2\rangle - 2\mathbb{E}_{l_\eta}\overline{q}^{l_\eta}\langle q^{l_\eta}_{12}\rangle + \mathbb{E}_{l_\eta}\overline{q}^{l_\eta} = \\
&= N\big(\mathbb{E}_{l_\eta}\frac{1}{l^2_\eta}\langle\sum_i^{l_\eta}(\sigma^1_i)^2(\sigma^2_i)^2\rangle + \overline{q}^2\big) = \left(\frac{1+a}{2}\right)^{-\frac{1}{2}} - N\overline{q}^2 = \sqrt{\frac{2}{\gamma}}N^{\frac{\theta}{2}} - N\overline{q}^2 \qquad (2.54) \\
B(0) &= \langle Q_{l_\eta 12}Q_{l_\eta 13}\rangle = N(\langle q^{l_\eta}_{12}q^{l_\eta}_{13}\rangle - \overline{q}^2) = \sqrt{\frac{2}{\gamma}}N^{\frac{\theta}{2}}\overline{q} - N\overline{q}^2, \qquad (2.55) \\
C(0) &= \langle Q_{l_\eta 12}Q_{l_\eta 34}\rangle = N(\langle q^{l_\eta}_{12}q^{l_\eta}_{34}\rangle - \overline{q}^2) = N\big(\mathbb{E}_{l_\eta}\frac{1}{l^2_\eta}\langle\sum_i^{l_\eta}\sigma^1_i\sigma^2_i\sigma^3_i\sigma^4_i\rangle - N\overline{q}^2 = \\
&= \sqrt{\frac{2}{\gamma}}N^{\frac{\theta}{2}}\int d\mu(\eta)\tanh^4\left(\frac{\beta\sqrt{\alpha\overline{q}\frac{\gamma}{2}}N^{\theta/2}}{1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\overline{q})}\right) - N\overline{q}^2. \qquad (2.56)
\end{aligned}$$

For the second party we need to evaluate G, H, I at $t = 0$. The only difference with the first party is the lack of dichotomy of its elements such that $z^2_\mu \neq 1$ as for the σ's.

$$\begin{aligned}
G(0) &= \langle P^2_{l_\chi 12}\rangle = N(\langle (p^{l_\chi}_{12})^2\rangle - 2\mathbb{E}_{l_\chi}\overline{p}^{l_\chi}\langle p^{l_\chi}_{12}\rangle + \mathbb{E}_{l_\chi}\overline{p}^2_{l_\chi} = \\
&= N\big(\mathbb{E}_{l_\chi}\frac{1}{l^2_\chi}\sum_\mu^{l_\chi}\langle (z^1_\mu)^2\rangle_G\langle (z^2_\mu)^2\rangle_G - \overline{p}^2\big) = \big(\alpha\frac{\gamma}{2}\big)^{-\frac{1}{2}}N^{\theta/2}\ \omega^2(z^2) - N\overline{p}^2, \qquad (2.57) \\
H(0) &= \langle P_{l_\chi 12}P_{l_\chi 13}\rangle = N(\langle p^{l_\chi}_{12}p^{l_\chi}_{13}\rangle - \overline{p}^2) = \big(\alpha\frac{\gamma}{2}\big)^{-\frac{1}{2}}N^{\theta/2}\ \omega(z)\ \omega(z^2) - N\overline{p}^2 \qquad (2.58) \\
I(0) &= \langle P_{l_\chi 12}P_{l_\chi 34}\rangle = N(\langle p^{l_\chi}_{12}p^{l_\chi}_{34}\rangle - \overline{p}^2) = \big(\alpha\frac{\gamma}{2}\big)^{-\frac{1}{2}}N^{\theta/2}\ \omega^2(z) - N\overline{p}^2. \qquad (2.59)
\end{aligned}$$

$\omega(z^2)$ and $\omega(z)$ are Gaussian integrals and can be explicitly calculated as

$$\omega(z) = \frac{\int d\mu(z) z \exp\left(bz\chi + \frac{d}{2}z^2\right)}{\int d\mu(z) \exp\left(bz\chi + \frac{d}{2}z^2\right)} = \frac{b\langle\chi\rangle}{1-d} = 0, \tag{2.60}$$

$$\omega(z^2) = \frac{\int d\mu(z) z^2 \exp\left(bz\chi + \frac{d}{2}z^2\right)}{\int d\mu(z) \exp\left(bz\chi + \frac{d}{2}z^2\right)} = \frac{1-d+b^2\langle\chi^2\rangle}{(1-d)^2} = \tag{2.61}$$

$$= \frac{1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\overline{q}+\alpha\overline{p})}{\left(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\overline{q})\right)^2}.$$

Finally, we have obviously $D(0) = E(0) = F(0) = 0$, because at $t = 0$ the two parties are independent. As we are interested in finding where ergodicity becomes broken (the critical line), we start propagating t (from 0 to 1) from the annealed region (high noise limit), where $\bar{q} \equiv 0$ and $\bar{p} \equiv 0$. It is immediate to check that, for the only terms that we need to consider, A, D, G (the other being strictly zero on the whole $t \in [0,1]$), the starting points are:

$$A(0) = \sqrt{\frac{2}{\gamma}}N^{\frac{\theta}{2}} = \frac{1}{r}, \tag{2.62}$$

$$D(0) = 0, \tag{2.63}$$

$$G(0) = \frac{N^{\frac{\theta}{2}}}{\sqrt{\frac{\gamma}{2}}(1-\beta\sqrt{\frac{\gamma}{2}}N^{\frac{\theta}{2}})^2} = \frac{1}{rs^2}. \tag{2.64}$$

Where we have defined $r = \sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}$, $s = 1-\beta\sqrt{\frac{\gamma}{2}}N^{\frac{\theta}{2}}$.
The evolution is ruled by

$$\begin{cases} \dot{A} = 2AD \\ \dot{D} = AG + D^2 \\ \dot{G} = 2GD \end{cases} \tag{2.65}$$

Noticing that $\frac{\dot{A}}{\dot{G}} = 0$ by substitution, and that $\frac{A(0)}{G(0)} = s^2$ we obtain immediately :

$$A(t) = G(t)s^2 = G(t)\left(1-\beta\sqrt{\frac{\gamma}{2}}N^{\frac{\theta}{2}}\right)^2. \tag{2.66}$$

The system then reduces to two differential equations; calling $Y = D + Gs$, we have $\dot{Y} = \dot{D} + \dot{G}s = G^2s^2 + D^2 + 2GDs = Y^2$ with solution $Y(t) = \frac{Y(0)}{1-tY(0)}$, and $Y(0) = D(0) + G(0)s = \frac{1}{rs}$ by which we get

$$Y(t = \sqrt{\alpha}\beta\frac{\gamma}{2}) = \frac{1}{rs}\,\frac{1}{1-\sqrt{\alpha}\beta\frac{\gamma}{2}(rs)^{-1}} = \frac{1}{\sqrt{\frac{\gamma}{2}}N^{-\theta/2}\left(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1+\sqrt{\alpha})\right)}, \tag{2.67}$$

i.e. there is a regular behavior up to

$$\beta_c = \frac{1}{\sqrt{\frac{\gamma}{2}}N^{\frac{\theta}{2}}(1+\sqrt{\alpha})}\,, \tag{2.68}$$

which confirms the result obtained in Equation (2.44). Now we can consider separately the evolution equation for G and D:

$$\dot{G} = 2G(t)\Big(Y(t) - sG(t)\Big) = \frac{2}{rs-t}G(t) - 2sG(t), \tag{2.69}$$

where we used $Y(t) = (rs-t)^{-1}$. Dividing both sides by G^2 and calling $Z = G^{-1}$ we get

$$-\dot{Z}(t) = 2Y(t)Z(t) - 2s = \frac{2}{rs-t}Z(t) - 2s. \tag{2.70}$$

Namely, an ordinary first order differential equation for $Z(t)$, that have the solution, for the initial condition $Z(0) = rs^2$:

$$Z(t) = 2s(rs-t) - \frac{1}{r}(rs-t)^2. \tag{2.71}$$

From $Z(t)$ we obtain $G(t)$ that is:

$$G(t) = \frac{1}{(rs-t)(s+\frac{t}{r})}. \tag{2.72}$$

Using Equation (2.66) and remembering that $D(t) = Y(t) - G(t)s$ we obtain the other overlap fluctuations.

$$\langle Q^2_{l_\eta 12}\rangle = \frac{\left(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}\right)^2}{\sqrt{\frac{\gamma}{2}}N^{-\theta/2}\Big(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1+\sqrt{\alpha})\Big)\Big(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\sqrt{\alpha})\Big)}, \tag{2.73}$$

$$\langle Q_{l_\eta 12}P_{l_\chi 12}\rangle = \frac{\sqrt{\alpha}\beta}{\Big(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1+\sqrt{\alpha})\Big)\Big(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\sqrt{\alpha})\Big)}, \tag{2.74}$$

$$\langle P^2_{l_\chi 12}\rangle = \frac{1}{\sqrt{\frac{\gamma}{2}}N^{-\theta/2}\Big(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1+\sqrt{\alpha})\Big)\Big(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\sqrt{\alpha})\Big)}. \tag{2.75}$$

A simple visual inspection at the formula above allows to confirm that the poles are found at

$$\beta\sqrt{\frac{\gamma}{2}}\,N^{\theta/2}\,(1+\sqrt{\alpha}) = 1,$$

confirming the heuristic result previously obtained. We can easily see furthermore that in the fully connected limit ($\gamma = 2$ and $\theta = 0$) we find out the result of [9].

2.4 Conclusions

In this chapter, we introduced and solved, at the replica symmetric level, a disordered mean-field system provides a generalization of the Hopfield model by introducing dilution into its patterns encoding the memories, the latter drawing form a Gaussian distribution.

We focused on the topological properties of the emergent weighted graph of the interactions. We found an exact expression for the coupling distribution, showing that in the thermodynamic limit it converges to a central Gaussian distribution with variance scaling linearly with the system size N. We also calculated the average link probability which, as

expected, depends crucially on the degree of dilution introduced. More precisely, by properly tuning it, the emergent structure displays an average coordination number $\bar{z}$ which can range from $\bar{z} = N$ (fully-connected regime) to $\bar{z} = \mathcal{O}(N)$ (constant link probability), to finite with $\bar{z} > 1$ (overpercolated network) or $\bar{z} < 1$ (underpercolated network).

Then, we moved to the thermodynamical analysis, where, through an interpolation scheme recently developed for fully connected Hebbian kernels [26], we obtained explicitly the replica symmetric free-energy coupled with its self-consistence equations. The overlaps, order parameters of the theory, turn out to be classical weighted sums of sub-overlaps defined on all possible sub-graphs (as for instance discussed in [30][39]). Both a small overlap expansion of these self-consistencies, as well as a whole fluctuation theory developed for their rescaled correlations, confirm a critical behavior on a surface (in the $\alpha, \beta, \gamma, \theta$ iperplane) that reduces to the well known of Amit-Gutfreud-Sompolinsky when the dilution is sent to zero [9]. On the other hand, the net effect of entry dilution in bit-strings (which weakens the coupling strength) is to rescale accordingly the critical noise level at which ergodicity breaks down, as expected.

Without imposing retrieval through Lagrange multipliers (as for analogical patterns it is not a spontaneous phenomenon, see [26]) the system displays only two phases, an ergodic one (where all overlaps are zero) and a spin-glass one (where overlaps are non-zero), split by the second order critical surfaces (over which overlaps start being non-zero) which defines criticality.

Part II

Mean field approach to cooperativity in biochemical reactions

Chapter 3

Cooperative behaviour in biochemistry

In this chapter, we introduce the background material for describing and understanding the cooperative behaviour in biochemistry. This chapter covers the concepts of solution thermodynamics and chemical equilibrium, in particular we will focus on the concept of *cooperativity*, a widespread phenomenon in biochemical reactions involving the binding of ligand molecules to larger biopolymers. As we are going to see, positive cooperativity gives to a system more sensitivity over a narrower range of stimulus, whereas negative cooperativity gives a less sensitive response over a much broader range, and both constitute a fundamental tool that nature developed to modulate the chemical response of biological systems to varying stimuli. We refer to standard textbooks for an extensive account of the topic (see for instance [45] and [46]).

3.1 Thermodynamics of reactions in solutions

A fundamental tool for understanding biochemical reactions is the thermodynamic study of solutions. Generally speaking, a solution is a single-phase system with more than one component, thought as an independently variable chemical substance. Note that, whenever chemical equilibria exist in the solution, there are usually more molecular species than the ones defined as components. For instance, let us consider a solution containing water, hemoglobin (Hb), and dissolved oxygen. Each hemoglobin molecule can bind one to four oxygen molecules, so we could possibly consider the species H_2O, Hb, O_2, HbO_2, $Hb(O_2)_2$, $Hb(O_2)_3$, $Hb(O_2)_4$. Yet, if the binding reactions are in equilibrium at a given temperature T and pressure P, it is sufficient to specify three independent components (solvent plus any two of the others), since equilibrium relationships allow us to determine the others.

The appropriate thermodynamic potential to study process occurring in a cell is the Gibbs free energy G, which is the quantity that is minimized for a system at equilibrium with constant temperature T and pressure P .

It is related to the enthalpy H and the entropy S via the Legendre transform $G = H - TS$. The contribution of the component i of the solution to the total free energy depends on its chemical potential μ_i, also called partial molar Gibbs free energy, defined as the increment of G with respect to the change in the number of moles n_i of the component i:

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{T,P,n_{j\neq i}} . \tag{3.1}$$

This is an intensive quantity, hence depending on T and P, but not on the size of the system, and its name comes from the fact that differences in its values for different components are the driving potentials for chemical reactions.

The total free energy of the solution can be then expressed in terms of the μ_i's as

$$G = \sum_{i=1}^{k} n_i \mu_i. \tag{3.2}$$

If the state of the system undergoes an infinitesimal change the corresponding change in G is

$$dG = \left(\frac{\partial G}{\partial T}\right)_{P,n_i} dT + \left(\frac{\partial G}{\partial P}\right)_{T,n_i} dP + \sum_i \left(\frac{\partial G}{\partial n_i}\right)_{T,P,n_{j\neq i}} dn_i, \tag{3.3}$$

which becomes, using standard thermodynamic relationships for entropy S, volume V, and chemical potential μ_i

$$dG = -SdT + VdP + \sum_i \mu_i dn_i, \tag{3.4}$$

and, if temperature and pressure are kept constant,

$$dG = \sum_i \mu_i dn_i. \tag{3.5}$$

The chemical potential of a substance A in a mixture depends on its concentration, and this dependence is particularly simple in *ideal* solutions, where

$$\mu_A = \mu_A^0 + RT \log \chi_A. \tag{3.6}$$

Here χ_A is the mole fraction of component A (that is the ratio between the number of molecules of solvent A and the total number of solute molecules), R is the gas constant, and T the absolute temperature. For $\chi_A = 1$ we have $\mu_A = \mu_A^0$, hence the chemical potential μ_A^0, which represents the chemical potential in a *standard state*, equals the molar free energy of pure component A.

To be called ideal, a solution should have two properties: first, there should be no difference in interaction energy between solute and solvent molecules, so that the enthalpy change ΔH_m in the solution is zero; second, the entropy change should be the entropy of non-interacting particles mixing, that is,

$$\Delta S_m = -R \sum_{i=k} n_i \log \chi_i. \tag{3.7}$$

In fact, from these two conditions it follows that the free energy for the mixing of a solution is

$$\Delta G_m = \Delta H_m - T\Delta S_m = RT \sum_i n_i \log \chi_i. \tag{3.8}$$

On the other hand, the free energy of n_i moles of a particular pure component is given by $n_i \mu_i^0$, so, using (3.2), we also have

$$\Delta G_m = G(\text{solution}) - \sum_i G_i(\text{pure components}) = \sum_i n_i(\mu_i - \mu_i^0). \tag{3.9}$$

Equating last two equations, and observing that the components n_i are independent variables, we obtain (3.6).

Instead of mole fraction that is an inconvenient quantity to deal with, since experiments usually measure concentrations, thus equation (3.6) for component A is often written in terms of the concentration $[A]$. Thus, we may equally write, for the solute in dilute ideal solutions,

$$\mu_A = \mu_A^0 + RT \log[A]. \tag{3.10}$$

Where $[A]$ and μ_A^0 is intended measured respect to a reference concentration.

To take care of deviations from ideal behavior, is usually substitutes in Eq. (3.10) the concentration $[A]$ with an effective concentration (or *activity*) $a_A = y_A[A]$, writing a formally analogous equation:

$$\mu_A = \mu_A^0 + RT \log a_A. \tag{3.11}$$

The activity is a unit-less quantity and the coefficient y_A has units of inverse concentration. The latter will be in general a function of T, P, and all the solute concentrations. In practice, is usually considered a virial expansions of the chemical potential, consisting in power series expansions in terms of the termodynamic parameters, around the ideal values. The higher order coefficients give a measure of the non-ideal behavior. We expect that, in general, solutes will approach an ideal behavior ($y_A = 1$) if the solution is very dilute.

3.1.1 Equilibrium in chemical reactions

Let us consider now a generic chemical reaction

$$aA + bB + ... \rightarrow pP + qQ + ... \tag{3.12}$$

in which a moles of molecule A, b moles of B, and so forth, react and form p moles of P, q moles of Q, and so forth, at molar concentrations $[A], [B], ..., [P], [Q], ...$ We stress that we are not assuming equilibrium here, but we simply take many moles of the reactants and convert them to the corresponding moles of products, under some given arbitrary conditions (temperature, pressure and concentrations). The driving force of the reaction is the free energy change ΔG, and, if the reaction occurs at constant temperature and pressure, Eq. (3.2) gives

$$\Delta G = p\mu_P + q\mu_Q + ... - a\mu_A - b\mu_B - ... \tag{3.13}$$

whence, using (3.11),

$$\begin{aligned} \Delta G =& (p\mu_P^0 + q\mu_Q^0 + ... - a\mu_A^0 - b\mu_B^0 - ...) \\ &+ RT \log \frac{[P]^p[Q]^q}{[A]^a[B]^b} + RT \log \frac{{y_P}^p {y_Q}^q}{{y_A}^a {y_B}^b}. \end{aligned} \tag{3.14}$$

Here the first term on the r.h.s, involving μ_i^0 values, is called the standard-state free energy change, and it is indicated with ΔG^0. It represents the free energy change that would be observed if a moles of A, and so forth, in the standard state, formed p moles of P, and so forth, also in the standard state. In the second term, the actual concentrations of reactants and products are taken into account, and the third term, involving only the activities coefficients, can be neglected, assuming that all the components behave ideally. In this case we may write Eq. (3.14) as

$$\Delta G = \Delta G_0 + RT \log \frac{[P]^p[Q]^q}{[A]^a[B]^b}, \tag{3.15}$$

and at equilibrium ($\Delta G = 0$) we find

$$\Delta G_0 = -RT \log \left(\frac{[P]^p [Q]^q}{[A]^a [B]^b} \right)_{\text{eq}} = -RT \log K \tag{3.16}$$

where we have defined the equilibrium constant K as

$$K = \left(\frac{[P]^p [Q]^q}{[A]^a [B]^b} \right)_{\text{eq}} \tag{3.17}$$

This constant tells us what are the relative concentrations of reactants and products at equilibrium, and will depend on temperature and pressure. The values of ΔG_0 are known for many reactions, and we can also rewrite last equation as

$$K = e^{-\Delta G_0 / RT}. \tag{3.18}$$

To have an idea near room temperature we have $RT \simeq 2.5$ kJ/mol, and if we consider as essentially irreversible a reaction with $K > 10^4$, the corresponding free energy change should be about $\Delta G_0 < -23$ kJ/mol, which is a condition satisfied for many biological reactions.

For a reaction beginning at arbitrary concentrations $[A], [B], [P], [Q], ...$, the second term, if there is not a refuelling of components, will gradually change in time until the concentrations are equal to the equilibrium values and $\Delta G = 0$. Note that the concentrations of reactants and products in many biochemical reactions correspond to values out of equilibrium, in physiological conditions. For example, at the concentrations maintained in most cells, the hydrolysis reaction

$$ATP + H_2O \rightarrow ADP + \text{phosphate} \tag{3.19}$$

gives to the cell an energy $-\Delta G \simeq 40$ kJ/mol, while the equilibrium value would be $\Delta G_0 \simeq -31$ kJ/mol.

3.2 Small ligand binding on macromolecules

Many biological functions involve the interactions of small molecules, acting as metabolites, regulators and signals, with specific sites on larger macromolecules, involved in cellular processes. Some typical examples include enzymes, which bind substrates and effector molecules accelerating the kinetics of reactions, transport proteins such as hemoglobin, which binds oxygen molecules, as well as the many proteins that act as buffers by binding hydrogen ions. Such binding mechanisms involve, in most cases, the formation of some kind of non-covalent bond between the small molecule or ion, called the *ligand*, and a specific region, the *binding site* (or *docking site*), on or near the surface of the macromolecule. A single macromolecule can possess binding sites of varying degrees of strength and specificity for different ligands. Enzymes, for instance, usually have a very specific *key-lock* mechanism to bind a molecule. Sometimes, one of the consequences of the act of binding a ligand on a site can be a conformational change in the biopolymer, which may influence the binding on other sites (*allosteric mechanism*), and in general binding affinities can be interdependent, allowing different metabolites, for instance, to interact indirectly.

In principle, except the case of covalent bonds, there is always an appreciable concentration of free ligands $[S]$ in equilibrium with bound ligands $[S_b]$ under physiological conditions. The total ligand concentration $[S_T]$ is obviously the sum

$$[S_T] = [S] + [S_b] \tag{3.20}$$

and the total molar concentration on macromolecules will be designated by $[P_T]$. The equilibrium process is dynamic, and at any instant different macromolecules P will host different numbers of ligands. The measure of the number of moles of S bound per mole of P is then an average

$$\bar{\nu} = \frac{[S_b]}{[P_T]}, \tag{3.21}$$

which increases monotonically with $[S]$ and should approach the limiting value n, equal to the total number of binding sites of P for molecule S, as the concentration $[S]$ increases. The *fraction saturation* is the correspondent normalized quantity

$$\theta = \frac{\bar{\nu}}{n} \tag{3.22}$$

and the curve describing its functional dependence on $[S]$ is called *binding curve*, or *binding isotherm*.

Among the many aspects in the investigation of such phenomena, it is often interesting to try to understand what is the maximum number of moles of ligands that can be bound for mole of the macromolecule, or, in other words, the number of binding sites, and the possible influence of the binding of a ligand on the other binding sites. This question is directly related with the phenomenon of cooperativity.

3.2.1 Fraction saturation curve

Needless to say, the simplest case to analyse is that of a macromolecule P with a single binding site for a molecule S, and no other species which can bind on it. The reaction describing the equilibrium process is then

$$[\mathrm{S}] + [\mathrm{P}] \rightleftharpoons [\mathrm{PS}] \tag{3.23}$$

where $[PS]$ denotes the concentration of hosting molecules with an occupied site, and $[P]$ and $[S]$ are the concentration of free ligands and macromolecules, respectively. The equilibrium constant for this reaction is then

$$K = \frac{[PS]}{[S][P]}. \tag{3.24}$$

Since there is only one site per hosting molecule, the measurable parameter $\bar{\nu}$ equals in this case the fraction saturation, and can be written as

$$\theta = \bar{\nu} = \frac{[PS]}{[P] + [PS]}, \tag{3.25}$$

or, using (3.24),

$$\theta = \frac{[S]}{K_{-1} + [S]}. \tag{3.26}$$

This corresponds to a simple hyperbolic dependence, as shown in figure 3.1, where the inverse equilibrium constant K^{-1} gives the concentration $[S]$ at half saturation ($\theta =$

1/2), and so the scale of concentrations involved. Note that $K^{-1} = [P][S]/[PS]$ is the equilibrium constant for the dissociation reaction

$$[\mathrm{PS}] \rightleftharpoons [\mathrm{S}] + [\mathrm{P}]. \tag{3.27}$$

When we consider the general case, that is a macromolecule P which can bind more than a single ligand ($n > 1$), the expression for $\bar{\nu}$ becomes more complicated, as there can be some molecules binding one ligand, some binding two, and so on, up to n. Let us now formalize such behavior by considering a hosting molecule P that can bind N identical molecules S on its structure; calling P_j the complex of a molecule P with $j \in [0, N]$ molecules attached, the reactions leading to the chemical equilibrium are the following

$$S + P_{j-1} \rightleftharpoons P_j,$$

hence the time evolution of the concentration of the unbounded protein P_0 is ruled by

$$\frac{d[P_0]}{dt} = -K_{+1}^{(1)}[P_0][S] + K_{-1}^{(1)}[P_1], \tag{3.28}$$

where $K_{+1}^{(1)}, K_{-1}^{(1)}$ are respectively the forward and backward rate constants for the state $j = 1$, and their ratios define the *association constant* $K^{(1)} = K_{+1}^{(1)}/K_{-1}^{(1)}$ and *dissociation constant* $\tilde{K}^{(1)} = K_{-1}^{(1)}/K_{+1}^{(1)}$. Focusing on the steady state we get, iteratively,

$$K^{(j)} = \frac{[P_j]}{[P_{j-1}][S]}.$$

Unfortunately, measuring $[P_j]$ is not an easy task and one usually introduces, as a convenient experimental observable, the average number $\bar{A}$ of substrates bound to the protein as

$$\bar{A} = \frac{[P_1] + 2[P_2] + ... + n[P_n]}{[P_0] + [P_1] + ... + [P_n]} = \frac{K^{(1)}[S] + 2 \cdot K^{(2)}[S]^2 + ... + N \cdot K^{(N)}[S]^N}{1 + K^{(1)}[S] + K^{(1)}K^{(2)}[S]^2 + ... + K^{(1)} \cdot K^{(n)}[S]^N}, \tag{3.29}$$

which is the well-known Adair equation [47], whose normalized expression defines the *saturation function* $\theta = \bar{A}/N$.

In a *non-cooperative* system, one expects independent and identical binding sites, whose steady states can be written as

$$0 = -NK_+[P_0][S] - K_-[P_1], \quad (j = 1), \tag{3.30}$$

$$0 = -(N-1)K_+[P_1][S] + 2K_-[P_2], \quad (j = 2), \tag{3.31}$$

where K_+ and K_- are the rates for binding and unbinding on any arbitrary site. Being $K \equiv K_+/K_-$ the intrinsic association constant, we get

$$K = \frac{[P_1]}{N[P_0][S]} = \frac{K^{(1)}}{N}, \quad (j = 1), \tag{3.32}$$

$$K = \frac{2[P_2]}{(N-1)[P_1][S]} = \frac{2K^{(2)}}{(N-1)}, \quad (j = 2), \tag{3.33}$$

and in general $K^{(j)} = (N - j + 1)\, K/j$. Plugging this expression into the Adair equation we get

$$\bar{A} = \frac{NK[S]}{1 + K[S]} \Rightarrow \theta = \frac{K[S]}{1 + K[S]}, \tag{3.34}$$

which is the well-known Micaelis-Menten equation [47].

If interaction among binding sites is expected, the kinetics becomes complex. Let us first sketch the limit case where intermediates steps can be neglected, hence

$$[P_0] + N[S] \rightleftharpoons [P_N],$$

then

$$\bar{A} = \frac{N[P_N]}{[P_0] + [P_N]} = \frac{NK[S]^N}{1 + [S]^N}, \tag{3.35}$$

$$\bar{\theta} = \frac{\theta}{N} = \frac{K[S]^N}{1 + [S]^N}. \tag{3.36}$$

More generally, one can allow for a degree of sequentiality and write

$$\theta = \frac{K[S]^{n_H}}{1 + [S]^{n_H}}, \tag{3.37}$$

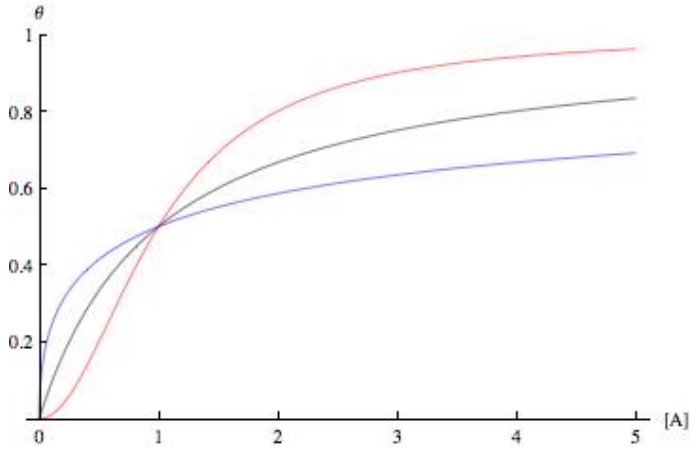
which is the well-known Hill equation [47], where n_H, referred to as Hill coefficient, represents the effective number of substrates which are interacting, such that for $n_H = 1$ the system is said to be *non cooperative* and the Micaelis-Menten law is recovered, while for $n_H > 1$ it is *cooperative*, for $n_H \gg 1$ it is *ultra sensitive*, while for $n_H < 1$ it is *anti cooperative*.
From a practical point of view, from experimental data for $\theta([S])$, one measures n_H as the slope of $\log(\theta/(1-\theta))$ versus $[S]$.

3.2.2 Cooperative binding

Consider a macromolecule with n binding sites: if the binding of a ligand to one site influences the affinity of other sites for the same kind of ligand, the binding is said to be *cooperative*. If the affinity of other sites increases after a ligand has bound, cooperativity is *positive*; the affinity could be also decreased, and in this case cooperativity is said to be *negative*. In the case of a protein with a number of binding sites, the mechanism causing this affinity change is usually a small change in the tertiary structure, after the binding of a molecule. Such effects are classified as a part of the general phenomenon of *allostery*. Here and in all the first part of this thesis, we will focus exclusively on *homeoallostery*, which refers to the influence on the binding affinities due to ligand of the same species, but it should be mentioned that many binding phenomena are *heteroallosteric*, that is, the binding of a ligand on a site can influence the binding of other species of ligands on different sites.

The different behaviors of positively and negatively cooperating sites are reflected in different binding curves ($\bar{\nu}$ or θ vs. $[A]$, see figure 3.1). The Michaelis-Menten curve (3.26) for noncooperative systems corresponds to a rectangular hyperbola, as we have seen. If there is positive cooperativity the curve has usually a sigmoidal shape (this is not strictly true, since when cooperativity is weak the binding isotherm has a form similar to the Michaelis-Menten curve, see section...). On the contrary, if cooperativity is negative, the remaining non-occupied sites become weaker as the others are filled, and a typical binding curve, compared with the non-cooperative and positively cooperative case, is shown in figure 3.1.

Figure 3.1: Binding curves for noncooperative (black), positively (red) and negatively (blue) cooperative binding sites. Note that the case of positive cooperativity corresponds to a sigmoidal curve.



The typical example of positive cooperativity is Hemoglobin, which has four binding sites for oxygen molecules. The quaternary structure of this protein consists of four subunits arranged in a roughly tetrahedral form, each containing a heme group (a iron charged atom held in a heterocyclic ring, known as porphyrin), the binding site for oxygen (see Fig. (3.2(b))). More precisely, the molecule contains two copies of two kinds of similar subunits (α and β), therefore it can be considered a dimer of α-β copies. Roughly speaking, when oxygen binds to the heme complex, it causes the iron atom to move backward into the heme which holds it, and this induces a series of modifications in the structure, such that binding of oxygen to the other three sites becomes easier. At low concentrations of oxygen the binding is very weak, because the molecule is in a weakly binding state, and the first ligands bound tend to bind to different molecules. When the first one or two sites are filled on any molecule, affinity on the remaining sites has increased, so that binding becomes stronger and the curve turns upward, in the typical sigmoidal shape (see Fig. 3.2(a)). This makes the hemoglobin a very efficient carrier for oxygen molecules in the blood, since it easily binds oxygen in the lungs, where the concentration is high, and can release it in the tissues of the body where the concentration is low.

3.2.3 Experimental measurements

In most cases, experimental measurements do not allow to observe directly the single ligands and binding sites, but they rather focus on the measure of the total fraction of ligand molecules that are bound, or the fraction of occupied binding sites (this is not strictly true, since some single-molecule techniques, such as atomic force microscopy, allow sometimes observation of certain dynamic interactions under physiologic conditions). An example is equilibrium dialysis, a technique based on membrane equilibrium, in which one places the macromolecule solution inside a semipermeable membrane bag, suspended in a solution containing the small ligands. In a nutshell, since the macromolecules cannot pass through the membrane, at equilibrium the excess of ligands in concentration inside the membrane should correspond to the bound molecules. Hence, measuring the concentrations of the smaller molecules inside and outside the bag gives the requested value.

Other methods are based on the change of some physical measurable properties (light-absorption spectra, fluorescence, nuclear magnetic resonance, and so forth) of the macro-

molecules when a molecule binds. Surface plasmon resonance (SPR), for instance, allows to detect the fraction of occupied sites by measuring the change in refractive index of a dielectric layer of biopolymers attached to a metallic surface, when they bind molecules from a solution. The problem of these methods is that they assume the same linear change in the physical parameter for all the binding sites, which is sometimes inaccurate. Moreover, if the number of sites n is not known a priori, it can be difficult to extrapolate its value from experimental measurements at high concentration of ligands, unless binding is particularly strong. Using a combination of several techniques, however, can give in most cases very accurate measurements.

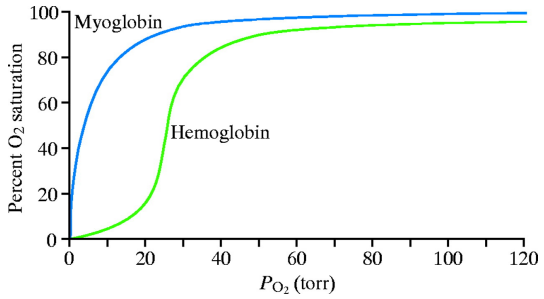
3.3 Model of cooperative binding

A large number of models have been proposed in the past, in order to have a quantitative description of cooperative binding on proteins. The most important, introduced for oxygen binding by hemoglobin, are probably the Monod-Wyman-Changeux (MWC) model [48], and the Koshland-Nemethy-Filmer (KNF) model [49], which in a certain sense lay at two extreme descriptions, among the many proposed.

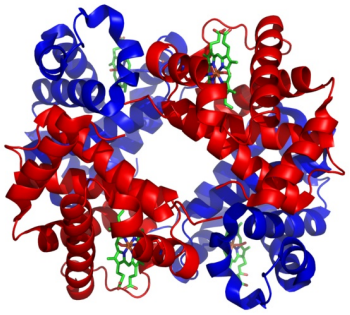
The MWC model is based on the concept of concerted conformational transitions in the subunits of the protein. It is assumed that each subunit contains a binding site: this exists in two states (T and R), with different affinities, but all the subunits undergo the transition in concert. For low concentrations of ligands the T state is favored, while the R form, with a larger affinity for ligands, is more favored at high concentrations. Without going into details, as the concentration of ligands increases from small values, the proportion of unoccupied sites in the R conformation will increase, and they will bind ligands with greater affinity. This model, with all its variants, predicts the behavior observed in homeoallosteric molecules with positive cooperativity, but it cannot describe negative cooperativity.

In the KNC model, on the contrary, it is assumed that the subunits change one at a time from a weak-binding to a strong-binding form, and the interactions between the different pairs of subunit depends on the relative states. This model can describe also negative cooperativity, and can take into account specific topologies for the interactions among binding sites. The model we are going to introduce later has more to do with this approach, as we shall see.

As for the sequential binding of O_2 molecules by hemoglobin, experimental measurements have shown that it has features in common with both the models depicted above, but agrees exactly with neither. In fact, individual α and β subunits can bind O_2 with individual tertiary conformational shifts, as described by KNC model, and a quaternary T$\rightarrow$R change can occur, but only after at least one of the sites on each α-β pair is occupied [50, 51].



(a) Curves describing the binding of oxygen by myoglobin and hemoglobin. Myoglobin has a single binding site for O_2 molecules, and the resulting curve is hyperbolic. Hemoglobin, on the other hand, has four binding sites to which oxygen binds cooperatively, with a sigmoidal shape for the binding isotherm.



(b) Structure of human hemoglobin. The red and blue subunits represent respectively α and β helices, while the four heme groups, containing the atom charged atoms to which oxygen bind, are in green. From Protein Data Bank.

Figure 3.2: Hemoglobin binding curve confronted with myoglobin, and hemoglobin structure.

Chapter 4

Mean-field model for cooperativity

In this chapter, we introduce a mean-field model for cooperative systems, developing the analogy with the Curie-Weiss model described before. We will show how different behaviours, in the binding of small molecules to homeoallosteric biopolymers with multiple docking sites, can be reconstructed and described in this framework. In particular, different cooperative phenomena, such as positive and negative cooperativity, can be interpreted in terms of an effective interaction between sites, and the observables commonly used to describe such phenomena, like the Hill coefficient (see section 3.2.1), are easily related to these interactions. A first study exploiting this framework was previously reported in [52] while the original contribution of the work has been published in [6].

It is important to point out that adopting a mean-field perspective implies the assumptions that the interactions among the system constituents are not limited by any topological or spatial constraint, but are implicitly taken to be long-ranged, as in a system that remains spatially homogeneous. This approach is naturally consistent with the rate-equation picture, typical of chemical kinetics investigations and whose validity is restricted to the case of vanishing correlations [53, 54] and requires a sufficiently high spatial dimension or the presence of an effective mixing mechanism. In general, in the mean-field limit, fluctuations naturally decouple from the volume-averaged quantities and can be treated as negligible noise. The consequence is to abandon a direct spatial representation of binding structures and we introduce a renormalization of the effective couplings. The reward lies in a resulting unique model exhibiting a rich phenomenology (e.g. phase transitions), yet being still feasible for an exact solution. In particular, we obtain an analytical expression for the saturation function which is successfully compared with recent experimental findings, taken from different (biological) contexts to check robustness. Furthermore, from this theory basic chemical kinetics equations (e.g. Michaelis-Menten, Hill and Adair equations) are recovered as special cases.

Consider a macromolecule, like for instance an enzyme, which can bind to its N sites some smaller ligand molecules. We will focus in the following on homeoallosteric binding, that is binding of the same ligand molecules on sites with the same affinities. We will see in the section 4.4 a possible extension to heterogeneous binding sites. The formal description of the global binding state of the macromolecules with the ligands is encripted in a vector of length N:

$$\sigma = \{\sigma_1, \sigma_2, ..., \sigma_N\}.$$

where the elements assume the two values $\sigma_i = \pm 1$, respectively when the site is occupied or empty. In section (3.2.1) we point out that the experimentally interesting macroscopic

quantity is the fraction saturation, that can be expressed as

$$\theta(\sigma) = \frac{1}{N}\sum_i \frac{1+\sigma_i}{2}. \tag{4.1}$$

In terms of $m(\sigma) = \sum_i \sigma_i/N$ this becomes

$$\theta(\sigma) = \frac{1+m(\sigma)}{2}. \tag{4.2}$$

The probability of a configuration $P(\sigma)$ at a given temperature will be influenced, in general, by the total concentration of ligands α and by the possible cooperative effects among sites. The distribution $P(\sigma)$ determines the average fraction of occupied sites

$$\langle\theta(\sigma)\rangle = \sum_\sigma P(\sigma)\theta(\sigma) \tag{4.3}$$

which is the quantity measured by experiments.

4.1 Independent binding sites

In this section, we introduce the formal bridge between the binding of ligands to macromolecules and the Statistical Mechanics. For this purpose, we start with the simplest scenarios where the binding sites are independent, there is no interaction between them and the total probability factorizes in the single probabilities for the occupation of a site: $P(\sigma) = \prod_i p(\sigma_i)$. In this case, cooperativity doesn't arise and we know that the binding curve obeys the Michaelis-Menten law (3.26).
Now, since $\sigma_i^2 = 1$, the most general function of σ_i is a linear function, but, without losing generality, we can also express it as an exponential

$$p(\sigma_i) = \frac{1}{Z}e^{h\sigma_i} \tag{4.4}$$

where Z is a normalization factor, which for the constraint $p(+1)+p(-1) = 1$ is equal to $Z = 2\cosh h$, and h is a parameter that will depend on the total concentration of ligands, here designated by α. The average fraction of occupied sites from (4.3) can be compute

$$\begin{aligned}\langle\theta\rangle &= \sum_\sigma P(\sigma)\frac{1}{N}\sum_i \frac{1+\sigma_i}{2} \\ &= \frac{1}{2} + \frac{1}{2N}\sum_i\left(\sum_{\sigma_i=\pm1} p(\sigma_i)\sigma_i\right) \\ &= \frac{1}{2}\left[1+\tanh(h)\right]\end{aligned} \tag{4.5}$$

This probability also correspond to the distribution of a paramagnetic spin in an external field h, at unitary temperature in which the microscopic energy of a configuration of sites is given by:

$$E(\sigma, h) = -h\sum_i^N \sigma_i \tag{4.6}$$

What is missing here is the link between the concentration of ligands and the probability distribution. From eq. (4.4) we know that the external field is related to the ratio of the

probability of having the site occupied to the probability for an empty site and this can be set proportional to the concentration α

$$\frac{p(+1)}{p(-1)} = e^{2h} = \frac{\alpha}{\alpha_0}, \tag{4.7}$$

with the proportionality constant α_0 that set a reference concentration. When $\alpha > \alpha_0$ this energetic term is positive and tend to favor binding of molecules on sites. On the contrary, if $\alpha < \alpha_0$ the logarithm is negative so that empty sites ($\sigma_i = -1$) are energetically favorable. $(1/2)\log\alpha_0$ corresponds to a sort of standard-state chemical potential.

With this simple assumption the average fraction of occupied sites (4.5) as a function of the concentration α is given by

$$\langle\theta\rangle = \frac{\alpha}{\alpha_0 + \alpha}, \tag{4.8}$$

which describes, as expected, a Michaelis-Menten behavior, with an equilibrium constant $K = \alpha_0^{-1}$, coherently with (3.18).

4.2 Two-sites interactions

The model above can be easily extended to cooperative system, adding a two-sites interactions that gives an Ising-like form for the microscopic energy.

$$E(\{\sigma\}|\mathbf{J},\mathbf{h}) = -\sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j - \sum_{i=1}^{N} h_i\sigma_i. \tag{4.9}$$

The first term represents an interaction energy between couples of sites: the couplings J_{ij} are assumed positive in the case of a cooperative system and the structure of these couplings will depend on the particular macromolecule considered. In any case, one obviously expects that the stronger the couplings, the stronger the cooperative effects.

In general, we could adapt this model to different proteins and enzymes, using a suitable couplings interaction J_{ij}, a paradigmatic example is the case of the haemoglobin considered in [55].

However, if we are interested in a unified picture more than an highly accurate model for specific case (with increasing number of parameters), a natural step is to consider a more simple mean-field approach, with an unique parameter for the interactions among couples of sites. This simple model is able to capture in a quantitative way several behaviors occurring in the binding of small ligands to macromolecules, without considering the detailed microscopic structure.

As we have seen in section 1.1.1, the mean field formulation of the Ising model is the Curie-Weiss model, where the energy of a configuration is

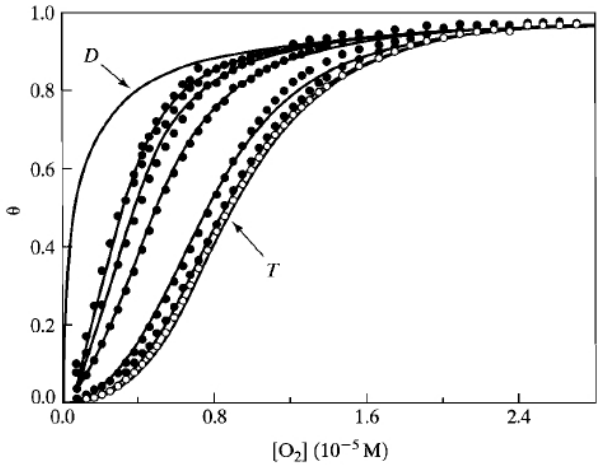
$$E(\sigma) = -\frac{J}{N}\sum_{1\le i<j\le N}\sigma_i\sigma_j - h\sum_i \sigma_i \tag{4.10}$$

The interaction is again between couples of sites, but there is no topology, and every site interacts in the same way with all the others, also if they are localized on different macromolecules. In the previous section, we considered the interactions between sites on a single macromolecule. However, if the concentration of macromolecules is large, we may expect that binding on less filled molecules will be more likely, since there is a large

number of free sites, and the system as a whole can manifest cooperative effects. These effects may be treated at a first stage as mean-field interactions among all the couples of sites, which "feel" the effective presence of molecules bound to the other sites, due to the high concentration of macromolecules. Hence, in the following we may consider N as the total number of binding sites, localized on different macromolecules, for a ligand.

As an example of the effect of high macromolecules concentration, consider the oxygen binding curves for different concentration of hemoglobin, shown in figure 4.1. For low Hb concentration the binding affinity is higher, but the cooperative effects are more strong when the concentration is higher. The energy (4.10) can be considered as an effective

Figure 4.1: Oxygen-binding curves for hemoglobin at different concentrations (decreasing from right to left). For diluted solutions the equilibrium between $\alpha\beta$ dimers and $(\alpha\beta)_2$ tetramers is shifted toward dimers,and this is the prevailing form. The leftmost solid curve is for dissociated $\alpha^1\beta^1$ dimers (D), with $n_H = 1.0$, while the rightmost curve pertains to tetramers (T), with $n_H = 3.3$. The concentration of hemoglobin for intermediate curves ranges form $4\times10^{-8}M$ to $1\times10^{-4}M$. Note the different cooperative effects (and affinities). Figure from [50].



microscopic interaction energy, and the coupling J, as we will see, can be directly related to the Hill number, giving an effective number for the interacting sites. The general interaction parameter we introduce is certainly not sufficient to reproduce correctly the complex mechanisms acting at the elementary level, but it encodes some of the most relevant features observed in binding curves. The main point in our analysis is the effective microscopic pictures that allow for a clear interpretation of a wide range of macroscopic cooperative behaviors and the cooperativity measures as the Hill coefficients in terms of the effective structural properties. Valid descriptions of binding phenomena in terms of this mean-field model are dependent on there being a very large number of solute particles in the sample observed, so that macroscopic fluctuations in properties will be very unlikely. As for biopolymers, we can consider an extreme example: in a solution containing 0.01 mg/ml of a virus of molecular weight 100 million, we still find approximately 1010 particles per milliliter. This is a number that is large enough that we don't need to worry about fluctuations in macroscopic volumes. It is only when we begin considering volumes comparable to that of a single cell that some problems can arise.

Seeking for a general scheme, we replace the fully-connected network of the original Curie-Weiss model by a complete bipartite graph: sites are divided in two groups, referred to as A and B, whose sizes are N_A and N_B ($N = N_A+N_B$), respectively. Each site in A (B)

is linked to all sites in B (A), but no link within the same group is present. The justification to the introduction of this bipartite scheme relies in the unification of the positive and negative cooperativity cases as it will be clear in section 4.2.2. With this structure we mirror dimeric interactions [1], where a ligand belonging to one group interacts in a mean field way with ligands in the other group (cooperatively or competitively depending on the sign of the coupling, see below), and they both interact with the substrate. As a result, given the parameters J and h, the energy associated to the configuration $\{\sigma\}$ turns out be

$$E(\{\sigma\}|J,h) = -\frac{J}{(N_A+N_B)}\sum_{i=1}^{N_A}\sum_{j=1}^{N_B}\sigma_i\sigma_j - h\sum_{i=1}^{N_A+N_B}\sigma_i. \tag{4.11}$$

Some remarks are in order now. This two-groups model can mimic both cooperative and non-cooperative systems but, while for the former case bipartition is somehow redundant as qualitatively analogous results are obtained by adopting a fully-connected structure, for the latter case the underlying competitive interactions intrinsically require a bipartite structure.

The order parameter can be trivially extended as

$$m_A = \frac{1}{N_A}\sum_{i=1}^{N_A}\sigma_i, \quad m_B = \frac{1}{N_B}\sum_{j=1}^{N_B}\sigma_j, \tag{4.12}$$

and, according to statistical mechanics prescriptions, we minimize the free energy coupled to the cost function (4.11) and we get in the thermodynamic limit, the following self-consistencies, has it is shown in the previous chapter

$$\begin{aligned} \langle m_A\rangle &= \tanh\left[J\rho_B\langle m_B\rangle + h\right], \\ \langle m_B\rangle &= \tanh\left[J\rho_A\langle m_A\rangle + h\right]. \end{aligned} \tag{4.13}$$

where $\rho_{A,B} = N_{A,B}/N$. Through equations (4.13), the number of occupied sites can be computed as

$$n_A(\{\sigma\}) = \sum_{i=1}^{N_A}\frac{1}{2}(1+\sigma_i) = N_A\frac{1+\langle m_A\rangle}{2}, \quad n_B(\{\sigma\}) = \sum_{j=1}^{N_B}\frac{1}{2}(1+\sigma_j) = N_B\frac{1+\langle m_B\rangle}{2}, \tag{4.14}$$

from which we get the overall binding isotherm

$$\theta(\alpha) = \frac{\langle n_A(\alpha)\rangle + \langle n_B(\alpha)\rangle}{N}. \tag{4.15}$$

We are now going to study separately the two cases of positive ($J > 0$) and negative ($J < 0$) cooperativity.

4.2.1 Ferromagentic interactions

When the couplings are positive, $J > 0$, interacting units tend to "imitate" each other. In this ferromagnetic context one can prove that the bipartite topology does not induce

[1] Note that for the sake of clearness, we introduced the simplest bipartite structure, which naturally maps dimeric interactions, but one can straightforwardly generalize to the case of an n-mer by an n-partite system and of course values of $\rho_A \neq \rho_B$ can be considered too. We did not perform these extensions because we wanted to recover the broader phenomenology with the smaller amount of parameters, namely J, α only.

any qualitative effects: results are the same (under a proper rescaling) as for the Curie Weiss model; indeed, in this case one can think bipartition as a particular dilution on the previous fully-connected scheme and we know that (pathological cases apart), dilution does not affect the physical scenario.

The solution of the statistical mechanics problem trivially generalizes results from previous sections, assuming an equal number of elements in the two sets ($\rho_A = \rho_B$=1/2), one has obviously $\langle m_A \rangle = \langle m_B \rangle$, and we get the following self-consistence equation for the magnetization,

$$\langle m_A \rangle = \tanh \left[\frac{J}{2} \langle m_A \rangle + \frac{1}{2} \log(\alpha/\alpha_0) \right] \tag{4.16}$$

while the overall binding curve $\frac{1}{2}(1 + \langle m_A \rangle)$ fulfills the following self-consistence equation (here we drop the brackets to have a more readable expression):

$$\theta(\alpha; J, \alpha_0) = \frac{1}{2} + \frac{1}{2} \tanh \left[\frac{J}{2}(2\theta - 1) + \frac{1}{2} \log(\alpha/\alpha_0) \right] \tag{4.17}$$

This expression returns the average fraction of occupied sites corresponding to the equilibrium state for the system. Differently from low-dimensional systems such as the linear Ising-chains, the Curie-Weiss model admits sharp (eventually discontinuous in the thermodynamic limit) transitions from an empty ($\langle m_A \rangle = \langle m_B \rangle = 0$) to a completely filled ($\langle m_A \rangle = \langle m_B \rangle = 1$) configuration as the field h is tuned. This means that the fraction saturation vanishes when the substrate concentration vanishes and it saturates to one when the substrate concentration is large, as expected. It should be clear that $\theta(\alpha)$ is the value corresponding to the average value $\langle \theta \rangle$ with respect to the Gibbs distribution with energy (4.10). The resulting fraction saturation is continuous for $J < J_c$, while for $J = J_c$ we have the scenario, previously depicted, corresponding to a second order phase transition (see figure 4.2). Conversely, when $J > J_c$ and $\alpha = \alpha_0$, transition is first order, that is the fraction saturation θ becomes discontinuous, taking a value smaller than 1/2 when $\alpha \to \alpha_0^-$ and greater than 1/2 when $\alpha \to \alpha_0^+$. This picture holds rigorously just in the thermodynamic limit ($N \to \infty$); for finite systems, beyond $\mathcal{O}(1/N)$ corrections, we recall that the discontinuous functions are mildly smoother, accordingly with the experimental counterparts. For clearance in the notation in the following, we will use α instead of α/α_0 and we will often drop the dependence on α and J (and the brackets) but it should be clear, from the context, that we are talking about this average value.

When couplings vanish, $J \to 0$, no cooperativity is expected (as the model reduces to a one-body theory) and, coherently, we recover the MM binding curve. In fact, eq. (4.17) can be equivalently expressed as

$$\theta(\alpha, J) = \frac{\alpha \exp\left[2J(2\theta(\alpha, J) - 1)\right]}{1 + \alpha \exp\left[2J(2\theta(\alpha, J) - 1)\right]} \tag{4.18}$$

which properly gives, for $J = 0$

$$\theta(\alpha, J)|_{J=0} = \frac{\alpha}{1 + \alpha}. \tag{4.19}$$

From equation (4.19), we see that when $J > 0$ the fraction saturation for a given concentration is smaller than the corresponding value for a non interacting system when $\alpha < 1$, and becomes greater when $\alpha > 1$. In fact, the greater the interaction and the steeper the sigmoidal shape of the curve. The fraction saturation curves resulting from eq. 4.19 are plotted in figure 4.2 versus α, for several values of J. Interestingly, a global change

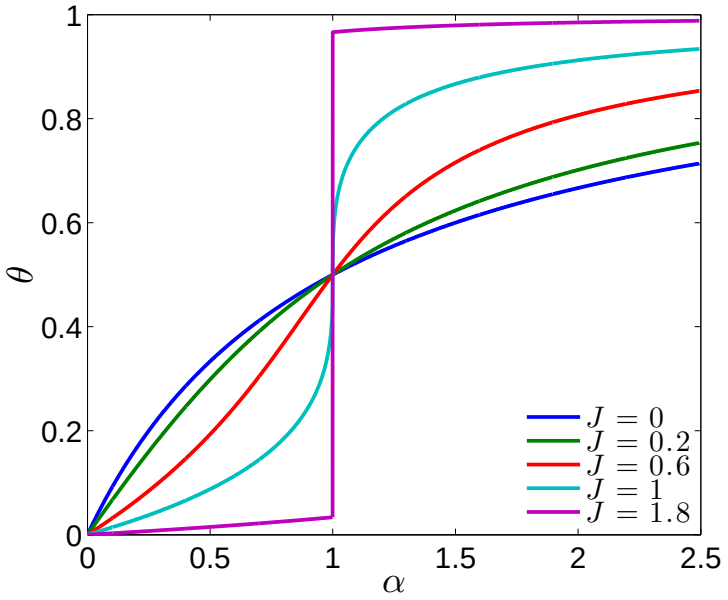


Figure 4.2: Different binding curves obtained by varying the coupling J. For $J = 0$ (blue line) the hyperbolic Michaelis-Menten law represents the curve for a non-interacting system; for $J = 0.2$ (green) the system has a weakly cooperative behavior; for $J = 0.6$ (red) strong cooperativity manifests itself with the typical sigmoidal shape; $J = 1$(light green) is the critical regime: the derivative in the inflection point which gives the Hill coefficient is infinite; $J = 1.8$ (purple) represents the discontinuous case, with an extremely strong cooperativity.

in the system considered, e.g. concerning ph or temperature, may lead to variations in the affinity between binding sites and ligands as well as in the coupling strength between binding sites themselves, giving rise to a curve $\theta(\alpha)$ displaying a different steepness.

Now, the derivative[2] of θ with respect to α, which is strictly related to the Hill coefficient and, consequently, to the cooperativity of the system, can be computed from (4.17):

$$\frac{\partial\theta}{\partial\alpha} = \frac{1}{4\alpha}\frac{1-(2\theta-1)^2}{1-J\left[1-(2\theta-1)^2\right]}. \tag{4.21}$$

This is always positive and finite for $J < 1$, meaning that θ is an increasing function of α, as we expected. In the limit of low concentration of ligands we obtain

$$\left.\frac{\partial\theta}{\partial\alpha}\right|_{\alpha=0} = \exp(-2J) \tag{4.22}$$

so the binding at very low concentration is governed by the two-bodies interaction J: the greater J and the flatter the fraction saturation curve. When $J = 0$, $\partial_\alpha\theta|_{\alpha=0} = 1$ and

[2]Note that in the frame of the Curie-Weiss model this is strictly related to the generalized susceptibility

$$\chi = \frac{\partial m(h)}{\partial h}$$

which measures the response of the system to a change in the field h. In fact, we have

$$\frac{\partial\theta}{\partial\alpha} = \frac{1}{2}\frac{\partial m(h(\alpha))}{\partial\alpha} = \frac{1}{2}\frac{\partial h}{\partial\alpha}\chi(h(\alpha)) = \frac{1}{4\alpha}\chi(h(\alpha)) \tag{4.20}$$

one properly recovers the same trend as that of the MM curve, which has a first order behavior with the same coefficient for small concentrations.

Finally, to recognize the sigmoidal shape typical of cooperative systems, we have to study the second derivative, which can be easily computed and expressed in terms of the first one:

$$\frac{\partial^2\theta}{\partial\alpha^2} = -\frac{1}{\alpha}\frac{\partial\theta}{\partial\alpha}\left[1 + \frac{2\theta - 1}{(1 - J\,[1 - (2\theta - 1)^2])^2}\right]. \tag{4.23}$$

When α ranges in $(1, \infty)$, this is always negative, so that θ is a concave function of α in that range, for any value of J. For $\alpha = 1$ we have $\partial^2_\alpha\theta = -(1/4)/(1 - J)$, so that θ is still concave there. For $\alpha \in (0, 1)$ we can compute numerically the second derivative: it comes out, not surprisingly, that it is not sufficient to have a positive coupling J between binding sites to see a sigmoidal curve. In fact, if this interaction is small, the second derivative is negative for all concentration, and the hyperbolic form will resemble the MM curve.

As explained in section 3.2.1, an usual way to define in a quantitative manner the cooperativity of a system is by the Hill coefficient n_H, obtained from the maximum slope of $\log[\theta/(1 - \theta)]$ vs. $\log\alpha$. If binding on different sites is an independent process, one simply finds $n_H = 1$, while in the extremum case in which sites are either all empty or all occupied $n_H = N$. We call a system cooperative (non cooperative) if $n_H > 1$ ($n_H = 1$), while the cooperativity is said to be negative, meaning that binding is reduced if there are occupied sites, for $n_H < 1$. This number gives then a lower bound for the number of interacting sites, and it is possible to see that it is related to the variance of the mean number of occupied sites, in our model.

The Hill coefficient for our general model depends, as expected, on the interaction J; in particular for $J < 1$ we find

$$n_H \equiv \frac{\partial \log\left[\theta/(1 - \theta)\right]}{\partial \log\alpha} = 4\left.\frac{\partial\theta}{\partial\alpha}\right|_{\alpha=1} = 1/(1 - J). \tag{4.24}$$

Being the derivative of θ for $\alpha = 1$, the Hill coefficient is finite (and greater than one) for $J < 1$ and it diverges for $J \to 1^-$ when the discontinuity appears. An infinite Hill coefficient may seem unrealistic, however it is not an unavoidable feature of our modeling: in fact h scales with the connectivity of the underlying network of interactions and, while the latter diverges in this minimal fully connected representation, diluted mean fields can still work finely. The equation above gives a new interpretation of the Hill number n_H, in terms of microscopic effective interactions among sites. On the other hand, the coupling J can be related in this way to the effective number of interacting sites.

From those considerations on the solutions of the self-consistence equation (4.10) we can individuate three regimes that we are going to described.

J<1/4: Weak cooperativity. The values of the coupling for which the curve is hyperbolic are the ones below the value $J \leq 1/4$. In fact, expanding θ to the first order in α one finds

$$\left.\frac{\partial^2\theta}{\partial\alpha^2}\right|_{\alpha=0} = -2(1 - 4J)\exp(-4J) \tag{4.25}$$

so for this value of the interactions the binding curve θ is everywhere concave, tending, for $J \to 0$, to the hyperbolic MM form (whose second derivative $-2(1 + \alpha)^{-3}$ is always negative). Note that when $J = 0$ the expression (4.25) gives, correctly, the MM value -2. The absence of an inflection point in the region $J \in [0, 1/4]$ allows us to define it as a weak cooperativity region: the shape of the binding isotherm is practically indistinguishable from that of a non-cooperative system.

We can expand the solution of the self-consistency eq. obtaining polynomials at all the desired orders, more typical of the standard route of chemical kinetics. In particular, expanding eq. 4.13 at the first order in J we obtain

$$\theta(\alpha) \approx \frac{(1-J)\alpha + \alpha^2}{1 + 2(1-J)\alpha + \alpha^2}, \tag{4.26}$$

which is nothing but the Adair equation (eq. 2) as far as we set $J = (1 - k_1^{3/2} k_2^{1/2}/2)$ and we rescale $\alpha \to \alpha/\sqrt{k_1 k_2}$.

$1/4 < J < 1$: Strong cooperativity From Eq. 4.23 it comes out that when $1/4 < J < 1$, there is a unique inflection point α^* (whose value increases with J), which separates the region where θ is convex (small concentration), to the one where it is concave. For $J = 1/4$ this point corresponds to $\alpha^* = 0$, while it is shifted towards unitary concentrations ($\alpha = 1$) when J is close to 1. As a sigmoidal curve has necessarily an inflection point, we may talk about strong cooperativity in this interval, in contrast to the weak cooperativity previously introduced. These very simple definitions have the advantage of being directly related to an effective microscopic interaction J, so that the experimental behavior of a system could allow one to reconstruct this interaction strength and interpret the binding curve in terms of the mean-field model.

$J > 1$: Ultra-sensitivity This region corresponding in the original Ising model to the "ferromagnetic" phase, the binding curve is still increasing with α, and the expressions (4.22-4.25) remain valid for $\alpha \neq 1$. In this point the curve is discontinuous and the jump is given by $\theta_+(J) - \theta_-(J)$, where

$$\theta_\pm(J) = \lim_{\alpha \to 1^\pm} \theta(\alpha, J).$$

These two limits depend on J: they are both equal to $1/2$ for $J = 1$, when the curve is still continuous, and their difference increases smoothly with the square root of $J - 1$ when $J > 1$ (see Figure 4.2). This means that, starting from vanishing concentration, the system has less sites occupied, for a given α, than the corresponding non interacting one, until the concentration reaches the reference value. Here, it is sufficient to increase infinitesimally the number of free molecules to obtain a large filling (depending on J). After that value, the number of occupied sites is always greater than the corresponding value for MM. Note that, in principle, if the concentration varies slowly one could observe metastability, with a curve which continues growing continuously up to values of $\alpha > 1$. The entire out of equilibrium features of the model are ruled out in this treatment as we deal with equilibrium statistical mechanics, however -as a second step- the bridge could be extended in that direction. If $J \to \infty$ this discontinuity increases, while its derivative in zero vanishes, so that in the large volume limit we obtain a step function. This kind of discontinuous behavior can be observed, for example, in the binding isotherms of small surfactants onto a polymer gel [56]

When $J \to 1$ a second order phase transition appears. This indicates that the correlation between binding sites becomes stronger and the typical trend of thermodynamical observables is a power law. These scalings, in particular those related to the reaction rate, can be used to suggest some new measures for almost discontinuous reaction curves [52].

4.2.2 Anti-ferromagnetic interactions

An interesting aspect of the model under investigation is the property to offer an unify scheme for the description of positive and negative cooperativity. The negative cooperativity is achieved by simply consider the parameter J negative. In this last case the bi-partite form of the energy in (4.10) is essential and constitute the justification for its introduction.

$$E(\{\sigma\}|J,h) = -\frac{J}{(N_A+N_B)}\sum_{i=1}^{N_A}\sum_{j=1}^{N_B}\sigma_i\sigma_j - h\sum_{i=1}^{N_A+N_B}\sigma_i. \tag{4.27}$$

When $J < 0$ we get the antiferromagnetic mean-field model, in which the long range ordering at low temperatures is quite different from that in the ferromagnet, and the correct order parameter for the study of the system is the staggered magnetization $m_A - m_B$. For large values of temperature and field the stable state is paramagnetic ($m_A = m_B$), but lowering these parameters the system goes through a second-order phase transition in an antiferromagnetic state, with a staggered magnetization different from zero (see also Sec 1.1.2). In this state there is an asymmetry between the two subsets, as one of the two is more magnetized. The magnetization, as we have seen, are obtain from the solution of the following equations:

$$\langle m_A\rangle = \tanh\left(-J\rho_B\langle m_B\rangle + h\right), \tag{4.28}$$

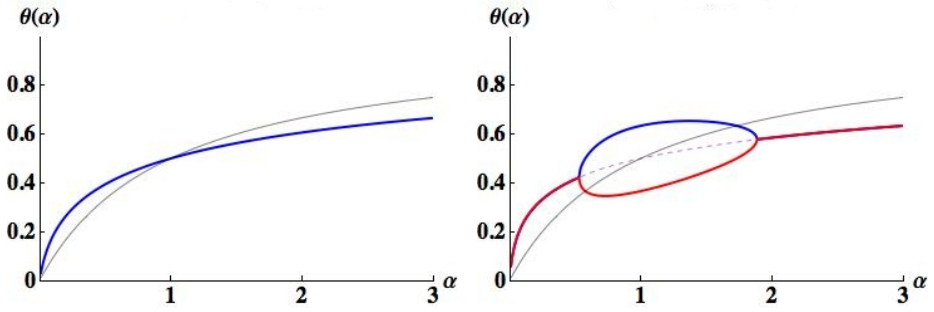
and equivalently for $\langle m_B\rangle$. We recast $J \to -J$ for outline the negative value of the parameter. We consider the case of the two parties with the same size, $N_A = N_B$, $\rho_A = \rho_B = 1/2$ The corresponding self-consistence equations for the two average partial fraction saturations are then:

$$\langle\theta_A\rangle = \frac{1}{2}\tanh\left(-\frac{J}{2}(2\langle\theta_B\rangle - 1) + \frac{1}{2}\log\alpha/\alpha_0\right), \tag{4.29}$$

$$\langle\theta_B\rangle = \frac{1}{2}\tanh\left(-\frac{J}{2}(2\langle\theta_A\rangle - 1) + \frac{1}{2}\log\alpha/\alpha_0\right). \tag{4.30}$$

The binding energy $\log\alpha$, as usual, acts on the same way on the two subsets, as it tends to keep both kind of sites empty when α is small, and filled when α is large. Just like for the antiferromagnetic counterpart, it is possible to check (see also Sec. 1.1.2) that there are two possible behaviors for the system, depending on the interaction strength J, and on the concentration of ligands α: if J is below a critical value J_c, the two partial fractions are always equal, for any concentration of external ligands (Fig. 4.3(a)). However, when the interaction is larger than this value, the two partial fractions are different, in a region of chemical potential $\log(\alpha)$ around zero, as shown in Fig. 4.3(b). Due to the strong interaction and the small chemical potential, it is more convenient for the system to fill sites on one of the subsystems and keep less molecules of ligands on the other subsystem. This region where the two fractions are different corresponds, in the magnetic models, to the anti-ferromagnetic phase, where the staggered magnetization, measuring the long range order of the system, assumes a non-zero value. In this case, starting from low concentrations, there is a critical value $\alpha_c < \alpha_0$ (and, consequently, of the chemical potential $\log\alpha$), above which $\langle\theta_A\rangle - \langle\theta_B\rangle$ start increasing with the concentration, reaches a maximum in $\alpha = \alpha_0$ and then start decreasing, until, from $\alpha = \alpha_c^{-1}$ on, it is again always equal to zero. This value of α_c, and the corresponding interval width $(\alpha_c, \alpha_c^{-1})$ where $\langle\theta_A\rangle \neq \langle\theta_B\rangle$, depends on the interaction strength J; when the average interaction equals J_c, we are in the limiting condition $\alpha_c = \alpha_0$ and, increasing the coupling, this interval of concentrations becomes wider.

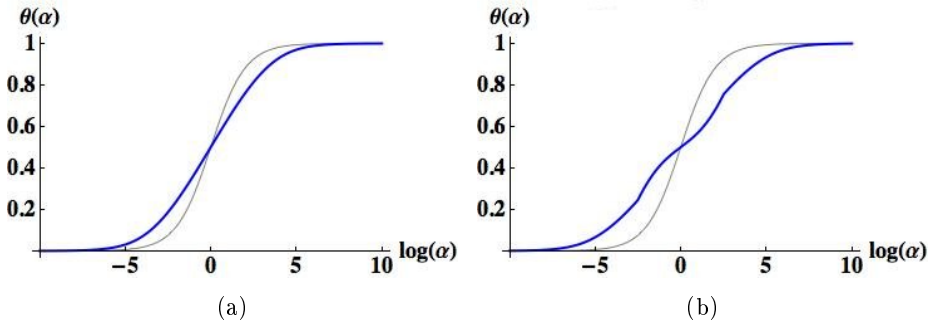
If measurements do not discriminate the subsystem to which a site belongs, we have seen that the total fraction saturation is simply given by $\langle\theta\rangle = (\langle\theta_A\rangle + \langle\theta_B\rangle)/2$. Thus, if the interaction is such that $\langle\theta\rangle_A \neq \langle\theta\rangle_B$ for a given range of concentrations, the total fraction of occupied sites will assume an intermediate value between the two. In all cases,



(a) (b)

Figure 4.3: Theoretical binding curves (fraction saturations) predicted by the model. (a) Theoretical curve for $J = 1.5$ (blue line) confronted with the non-cooperative case (in black). In this case the fraction saturations of the two parties coincide for any value of the concentration of ligands, and the total fraction saturation coincide with them. (b) Fraction saturations of the two subsystems (red and blue lines) for $J = 2.5$; note that in this case there is a difference in the two values for a given range of concentrations, and in this range the total fraction saturation is the average of the two (dashed line). The black line represents the non-cooperative case. In both cases $\alpha_0 = 1$.

the binding curves stay above the value of the corresponding non-cooperative curve for a concentration below the scale α_0 (which we have assumed to be equal to one for simplicity), and below the non-cooperative curve for larger concentrations (Fig. 4.4). This is typical of negatively cooperative systems, which can thus respond to concentrations on a broader range with respect to non-cooperative macromolecules, which can be a useful mechanism in biological systems. On the contrary, we have seen that the effect of positive cooperativity is to trigger a response in the system with small variations of concentrations [57]. Note that we can express the extent of negative cooperativity in terms of a Hill coefficient by generalizing equation (4.24) for negative J, as in this case the resulting number satisfies $n_H = 1/(1-J) < 1$, which is a common feature of negatively cooperative systems.



(a) (b)

Figure 4.4: Theoretical binding curves (blue lines) for the total fraction saturation versus the concentration in log scale, confronted with the non-cooperative case (in black). Binding curve for a negatively cooperative system with (a)$J = 2$ (b) $J = 3.5$. Note that here, unlike the other case, there is a range of concentrations around $\alpha = 1$ for which the steepness of the curve, coming from lower concentrations, decreases and then start increasing again for a while.In both cases $\alpha_0 = 1$.

Binding isotherms are often plotted taking a logarithmic scale for the concentration of ligands, and in fig. 4.4 we show some theoretical binding curves with the total fraction saturation plotted as a function of the logarithm of ligands concentration. Figure 4.4(a) shows the binding curve predicted for a system with interaction strength $J = 2$. Here the partial fraction saturations coincide over all concentrations, with negative cooperativity manifesting itself in the reduced steepness of the curve with respect to the non-cooperative counterpart (black line). In fig. 4.4(b) we show an analogous plot with $J = 3.5$ (in blue). In this case the trend of the curve is more complicated, as in the central region, where the interaction among sites manifests mostly, privileging the occupation of a subsystem with respect to the other ($\langle\theta_A\rangle \neq \langle\theta_B\rangle$), the steepness of the curve decreases coming from low concentrations and then starts increasing for a while for $\alpha > 1$, until it start decreasing again for larger concentrations. Interestingly, this effect can only be seen on a logarithmic scale of concentrations (compare with fig. 4.3(b)). For low concentrations, the sites of the two subsystems are equally filled with ligands, but then, if the negative interaction strength among sites is sufficiently strong, at a certain concentration one of the two subsystems begins binding more ligands than the other. The strong effect of mutual negative cooperativity (with respect to the smaller effect of the chemical potential $\log\alpha$ in that range) decreases the steepness of the global fraction saturation in a range of concentrations around unity, but, when the concentration start increasing over that values, its effect becomes more important, and the steepness of the binding curve starts increasing again, until the partial fraction saturations are re-equilibrated at high concentrations, where it becomes decreasing towards the total saturation of sites.

This particular behaviour is usually find in experimental data (see for instance Fig. 4.11), but while in literature the standard way to handle this experimental data is usually to fit with an interpolation between two Hill functions, in the mean-field model this behavior comes out naturally as a consequence of the antiferromagnetic phase.

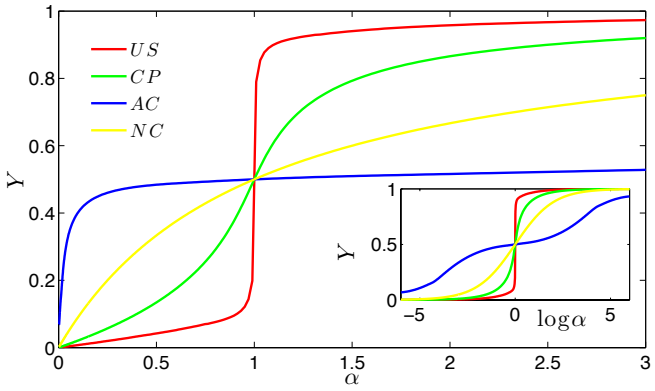


Figure 4.5: Theoretical predictions of typical binding isotherms obtained from SM models. Different colors refer to different behaviors, as explained by the legend. In particular, as the interaction strength J is varied, qualitative different outlines emerge: discontinuous or ultra-sensitive (US) for $J = 6$, cooperative for $J = 2$ (CP), anti-cooperative for $J = 0.5$ (AC), non-cooperative for $J = 1$ (NC).

4.3 Fit with experimental data

We have shown in the previous section the versatility of the model in reproducing different qualitative behaviour of cooperative systems. In this section, we use the model to quan-

titative modelling experimental data. Before passing to describe the experiments and the biological systems we briefly describe the fitting methods used. The fitting parameters of the model are the coupling strength J, that is another measure of the degree of cooperativity and the reference value of the concentration of the free ligand that rescale the external field α_0 that can be obtain form the location of the curve in the midpoint usually known from the experiment.

4.3.1 Fitting procedure

To fit the model, we used a least squared method, where the fitted parameters are chosen minimizing the difference between the observed values and the fitted values provided by the model. The function to minimize is therefore the sum of the squared residuals, namely:

$$S = \frac{1}{M} \sum_{i}^{M} (\theta_i - d_i)^2 \tag{4.31}$$

where d_i is the M experimental data relative to the log-concentration h_i, θ_i is the solution of the self-consistency equations (4.13) calculated in h_i.
The algorithm used to minimize the function is the gradient descent of S, since we have tested that the function is sufficiently smooth. We implement the algorithm with the software Mathematica. More in detail the algorithm consists in three major steps:

- solving the self-consistency equations (4.13) for the different values of h_i, for this purpose we used the built-in function *FindRoot*;
- calculating the sum the squared residuals S, and minimizing with respect to the renormalized data d_i' obtained from the original ones by $d_i' = \text{basal} + \text{maximum} \quad d_i$ changing the zero value *basal* and maximal value *maximum*, using the built-in function *Minimize*. This normalization is necessary for different experimental definitions of the saturation function.
- updating the parameters, the interaction strength J and the mid-values concentration $h_0 = \frac{1}{2} \log \alpha_0$.
- repeat the previous steps until convergence and calculate the errors on the parameters with standard methods for non-linear fitting as shown in Appendix (B.1).

We describe briefly the experimental cases of cooperative systems we have considered. We select in literature three experiments on positive cooperative systems and four experiments on negative cooperative systems in different context (from neurobiology and RNA riboswitch to colloidal nanoparticles). We show in figures (4.6)-(4.11) the comparison of the experimental data and the theoretical prediction of the fitted model.

4.3.2 Positive cooperative cases

1. **CaM kinase II**
   The Calcium-calmodulin-dependent protein kinase II (CaMKII) is one of the most important transducers of $Ca2+$ signals in a variety of cell types and is highly conserved across animal species; in post-synaptic regions, CaMKII is involved in many signaling cascades and is thought to be an important mediator of learning and memory. CaMKII displays 12 kinase domains organized into large symmetrical

holoenzymes. When inactive, CaMKII forms tightly packed auto-inhibited assemblies, which, upon activation, convert into clusters of loosely tethered and independent kinase domains. The activation is prompted by Ca^{2+}/CaM which, binding to CaMKII, removes an auto-inhibitory regulatory segment and this releases the catalytic activity of the enzyme and makes accessible a regulatory residue, namely $Thr-286$. Then phosphorylation of $Thr-286$ by another kinase domain within the oligomeric holoenzyme can take place. Phosphorylation of $Thr-286$ keeps CaMKII active in the absence of Ca^{2+}/CaM by preventing the rebinding of the regulatory segment to the kinase domain.

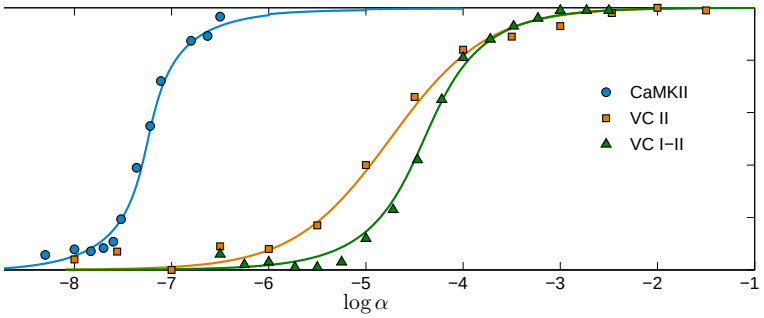
In the experiments performed by Chao and coworkers [58], the binding of calcium-saturated calmodulin to the CaMKII holoenzyme shows clear evidence for positive cooperativity with a Hill coefficient n_H 3.0. This suggests that the binding of one molecule of Ca^{2+}-calmodulin to a kinase subunit, which is expected to activate it, also potentiates adjacent kinase subunits for binding to Ca^{2+}-calmodulin. In the experiment reported in [58] the velocity of substrate phosphorylation at varying calmodulin concentrations is measures and since the velocity of substrate phosphorylation is proportional to the fractional saturation, once normalized those data, we fitted them with Eq. 30 and we show the result in Fig. (4.6). The fitting parameters are discussed in the caption.

2. **mRNA riboswitch**
In molecular biology, a riboswitch is a part of an mRNA molecule that can directly bind a small target molecule, and whose binding of the target affects the gene's activity. Thus, an mRNA that contains a riboswitch is directly involved in regulating its own activity, in response to the concentrations of its target molecule. The discovery that modern organisms use RNA to bind small molecules, and discriminate against closely related analogs Most known riboswitches occur in bacteria, but functional riboswitches of one type (the TPP riboswitch) have been discovered in plants and certain fungi.

   Mandal and coworkers [59], identified a riboswitch class in bacteria that is selectively triggered by glycine. These riboswitches integrate two ligand-binding domains that function cooperatively to more closely approximate a two-state genetic switch (flip-flop). The authors measured the ligand occupancy of two different RNA constructs, i.e. VCI-II and VC-II, versus the concentration of glycine, showing that a change from $\sim 10\%$ to $\sim 90\%$ ligand-bound VCI RNA occurs over a ~ 100-fold increase in glucine concentration, while for VC I-II the same change in ligand occupancy occurs over only a ~ 10-fold increase in glycine concentration. The related Hill coefficients turn out to be $n_H = 0.97$ and $n_H = 1.64$, respectively. In the Fig. (4.6) we show the comparison with the fitted model, see caption for the details.

3. **CaM kinase II, (ultra-sensitivity case)**
In the experiment performed by Bradshaw and coworkes [60], it is evidenced that there can be several issues by which CaMKII responds cooperatively to Ca^{2+}. In particular, one factor is the cooperative binding Ca^{2+}/CaM to CaMKII and another factor is the requirement that Ca^{2+}/CaM bind two different CaMKII subunits for $Thr-286$ auto-phosphorylation to occur. When both this factors are present, great sensitivity ($n_H \sim 5$) is measured from titration curves. This enhanced sensitivity is referred to as ultra-sensitivity and make CaMKII exhibits an intriguing switch-like activation that, as the authors notice, likely is important for changes in synaptic

strength. At the synapse, CaMKII auto-phosphorylation is regulated by the action of phosphatase, particularly protein phosphatase 1 ($PP1$). The switch-like response of a $CaMKII-PP1$ system suggests that CaMKII and $PP1$ may function together as a simple molecular device that specifically translates only strong Ca^{2+} signals into all-or-none potentiation of individual hippocampal synapses. Experimental data and inferred model are shown in Fig. (4.7).



set dati	J	n_H	n_H^{lit}
VC II	0 ± 0.6	1.0 ± 0.7	0.97
VC I-II	0.3 ± 0.1	1.4 ± 0.2	1.64
CaMKII	0.67 ± 0.03	3.0 ± 0.3	3

Figure 4.6: The plots show data from recent experiments (symbols) concerning non cooperative and cooperative binding and fits through our model (Eq. (4.10)). Here θ stands for the fraction saturation and the horizontal axis is logarithmic to the scale 10, with concentrations of ligands α expressed in mol. (●) represent fraction saturation measurements of CaM binding on CaMKII, for which an original Hill fit revealed a coefficient $n_H = 3.0$ (hence positive cooperativity) [58]. Red continuous line is the best fit with our model, predicting $n_H = 2.94 \pm 0.06$, in complete agreement with the literature. The green and blue experimental data (squares and triangles) represent the binding of glycine molecules by the VC I-II RNA ((▲), cooperative) and the VCII RNA ((■), non-cooperative) [59]. The line for the VC II is a Michaelis-Menten curve with $\alpha_0 = 7.5 \times 10^{-5}$ mol ($J \equiv 0$ in our theory) while, the line for the VC I-II is the best fit within our cooperative model, giving $n_H = 1.66 \pm 0.03$. This value is in agreement with the one found in [59] with a Hill fit, which is $n_H = 1.64$. Note that the MM curve has a sigmoidal form, due to the log scale. All the fitted values of the interaction parameter, the inferred Hill number, using (4.24), and the comparison with the Hill number obtain with a fit with the Hill function n_H^{lit} are reported in the table below the figure.

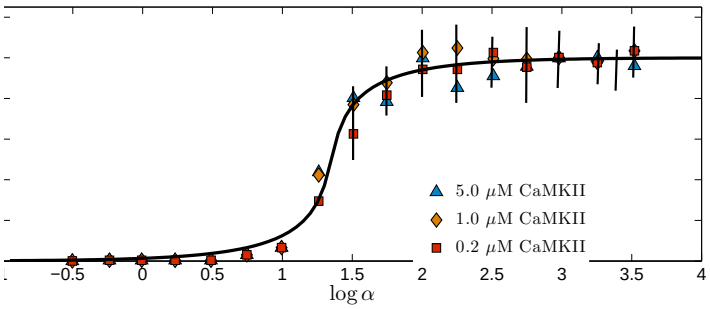


Figure 4.7: Data taken from [60] are still regarding Ca^{2+} -calmodium-dependent protein kinase II, which exhibits a switch-like activation (with a Hill number n_H 5) when coupled with PP1 (protein phosphatase I) at various concentration ($5\mu M$ (▲) ,$1\mu M$ (♦), and $0.2\mu M$ (■), as reported in the legend). Continuous line is our best fit over the whole data, whose prediction is $n_H = 4.89 \pm 0.18$, in total agreement with the literature.

4.3.3 Negative cooperative cases

1. **Neuroreceptors mGlu and GABA**
   The metabotropic glutamate receptors ($mGlu$) and the $GABA_B$ receptors are G-protein-coupled receptors, which play essential role in the central nervous system by regulating fast excitatory and inhibitory transmissions. These receptors are homodimers and each subunit is composed of a ligand binding (VFT) domain and of a heptahelic domain (HD) responsible for G-protein activation. Structural studies identified three states for the dimeric VFT domain: open-open $(1,1)$, closed-open $(0,1)$ and closed-closed $(1,1)$, which, if connected to the HD, give no, partial, and full activity, respectively. In the experiment reported in [61], the authors used the purified soluble VFT dimer of $mGlu1$ as a model, and glutamate as ligand; the Hill analysis of the titration curve showed negative cooperativity.

   Coherently, the experimental results in [61], as also theoretically confirmed in [62], evidence an anti-cooperative behavior. As shown in fig. (4.8), the binding isotherms nicely fit the experimental data and the fit parameters are in agreement with the discussion [62].

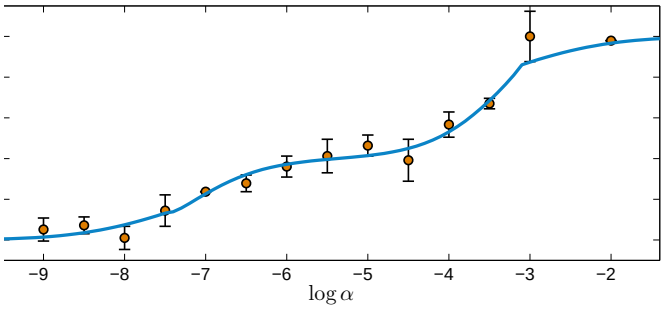


Figure 4.8: Anti-cooperative behavior, experimental data are taken from the figure (5) in [62] (●) representing the fractional response f defined as the fraction of receptors concentration in the active form. The fractional response f is related to the observables n_A and n_B in the model by the relation $f = (f_{max} - f_{min})(pn_A + (1-p)n_B) + f_{min}$, where $p = 0.3$, $f_{min} = 0.41$ and $f_{max} = 0.89$ are fixed parameters taken from [62]. The fitting values of the model parameters are $J = 4.10 \pm 0.05$, $\alpha_0 = 1.0910^5 \pm 10^3$.

2. **Metabotropic glutamate receptor, mGluR**
   Glutamate is a major neurotransmitter in the excitatory synapses of the central nervous system; in particular, the metabotropic glutamate receptor (mGluR) is a G-protein-coupled receptor (GPCR), which induces various cellular responses to glutamate stimulation. Because these cellular responses modulate the degree of synaptic neurotransmission, mGluRs are believed to be involved in higher order neuronal activities such as memory, learning, and so on Metabotropic glutamate receptor (mGluR) has a large N-terminal extracellular ligand binding domain that forms a homo-dimer. Using the intrinsic tryptophan fluorescence change as a probe for ligand binding events, Suzuki and coworkers [63] the authors examined whether allosteric properties exist in the dimeric ligand binding domain of the receptor. A Hill analysis of the saturation binding curves revealed the strong negative cooperativity of glutamate binding between each subunit in the dimeric ligand binding domain.

   It is worth stressing here a possible explanation for the role of negative-cooperativity. In fact, bio-chemists stress that negative cooperativity can extend the ligand concentration range over which the protein can work [64]. Thus one possible role for the negative cooperativity is to extend the glutamate concentration range to which the receptor can respond. This mechanism will be useful in the situations where continuous stimulation takes place at the synapse. Even in such a situation, the receptor will be able to respond, because the ligand binding sites of the receptors on the cell surface would not be completely saturated by glutamate. This mechanism will be useful in the situations where continuous stimulation takes place at the synapse. Even in such a situation, the receptor will be able to respond, because the ligand binding sites of the receptors on the cell surface would not be completely saturated by glutamate. The other advantage for the negative cooperativity is a greater sensitivity for low ligand concentration.

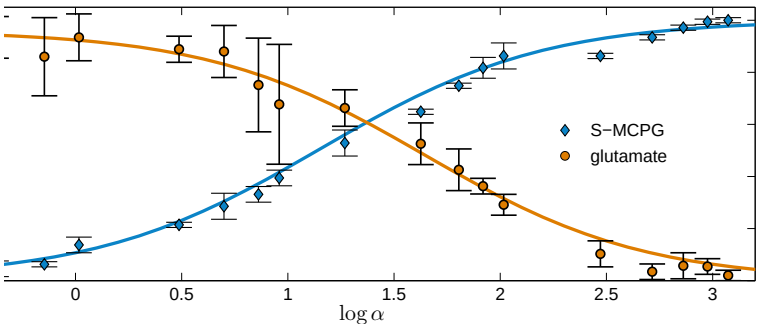


Figure 4.9: Example of negative cooperativity. Data are taken by Suzuki et al. [63] (fig.7 C in their paper), showing titration curves for glutamate (●) and quisqualate (S)-MCPG (♦) binding to the I120ALBD.

3. **Colloidal nano-particles**
   Nano-particles, when exposed to biological fluid, become coated with proteins and other biomoleculer. In [65] the author have explored this phenomenon by considering as a model protein the human serum albumin (HAS), which is a major soluble constituent of human blood plasma, and the their absorption is tested against small polymer-coated FePt and CdSe/ZnS nano-particles negatively charged, whose matching is tested through fluorescence correlation spectroscopy.

   In the experiment, they extracted the hydrodynamic radius R_H and used it to es-

timate the increasing nano-particle size due to protein binding to its surface. The radius R_H is therefore proportional to our saturation function θ and plotted versus ligand concentration α in the figure (4.10).

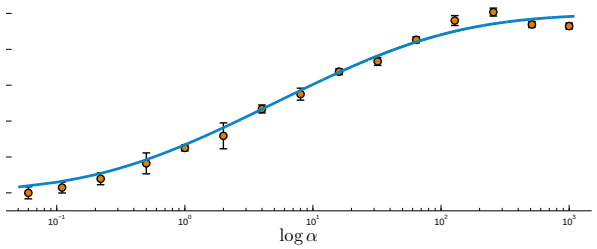


Figure 4.10: Experimental data (●) taken from Figure 3 in [65] for the anti-cooperative binding of human serum albumin onto small (10 - 20 nm) polymer-coated FePt and CdSe/ZnS nanoparticles. The blue line is the best fit within our model (for $a = 1$), with $J = 1.29 \pm 0.05$ and $\alpha_0 = 5.0 \pm 0.3$ mol.

4. **Vitreoscilla hemoglobin-like molecule, VHb**
   The bacterium *Vitreoscilla* sp. synthesizes a soluble hemoglobin-like molecule (VHb) in response to hypoxic environments. Vitreoscilla hemoglobin is involved in oxygen metabolism of this bacterium. Ligand binding isotherms for cyanide, azide, thiocyanate and imidazole association to ferric Vitreoscilla Hb are shown displaying anti-cooperative behavior as monoligation renders thermodynamically unfavorable the ligand association to the second heme site. Ferric Vitreoscilla HB displays an anticooperative ligand binding behavior in solution. This very unusual feature can only be accounted for by assuming ligand-linked conformational changes in the monoligated species. There is an interest in Vitreoscilla biotechnology as a means for improving the efficiency of cell and plant growth processes on industrial scale. Moreover, Vitreoscilla is capable of inducing the synthesis of a homodimeric hemoglobin. The thermodynamics and kinetics for cyanide, azide, thiocyanate and imidazole binding to recombinant ferric vitreoscilla sp homodimeric hemoglobin (Vitreoscilla Hb) have been determined ot pH 6.4 and 20.0 C. Ferric Vitreoscilla HB displays an anticooperative ligand binding behavior in solution. This very unusual feature can only be accounted for by assuming ligand-linked conformational changes in the monoligated species. Error bars non available.

   In Figure (4.11)we show the ligand binding isotherms for ferric VHb in solution: data were taken from [66] and fit is accomplished via eq. (4.30) finding a good agreement. This values of coupling is compatible with a feature of anti-coperativity in agreement with [66].

4.4 Heterogeneous interactions

One assumption that we have done since the beginning of our analysis is that the affinity with ligands of the different binding site and the interaction between them are the same. This homogeneity in the chemical potentials of the binding sites and in the interactions between them is related in our model to the description by a single parameter J for the interactions and α_0 for the chemical potentials.

The model can be extended to take into account a relaxation of these assumptions. The heterogeneity can be introduced in both the coupling parameters and the external

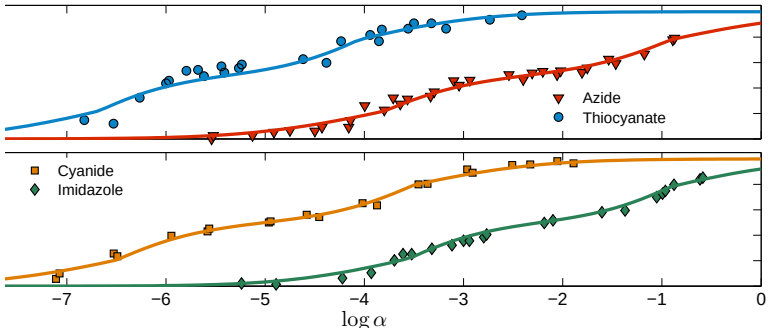


Figure 4.11: Anti-cooperative behavior, experimental data taken from Figure 1 in [66] for the binding of different ligands to the recombinant homodimeric ferric Vitreoscilla Hemoglobin in solution at pH 7.0 and temperature 293 Kelvin(θ stands for the fraction saturation). Comparisons among the fitted solutions of our theory (solid line) and the experimental data (symbols) are shown for the ligand binding isotherms for azide(▼), thiocyanate (●) , cyanide (■) and imidazole (◆). The fitting values of the parameters are (with $a = 1$): (a) cyanide $J = 3.05 \pm 0.05$, $\alpha_0 = 3.6\ 10^4 \pm 10^2$, (b) azide $J = 2.9 \pm 0.05$, $\alpha_0 = 1.0\ 10^2 \pm 1$, (c)thiocyanate $J = 2.9 \pm 0.05$, $\alpha_0 = 5.9\ 10^5 \pm 10^3$ and (d) imidazole $J = 2.9 \pm 0.05$, $\alpha_0 = 5.8\ 10^2 \pm 1$.

fields. The two choices account for different chemical property. In the first case, the bound of the ligand to an active site can module differently the affinities of the other sites (through conformational change of the protein for example). In the homogeneous case, the effect of the bound to a site is the same on all the other, as for the MWC model, while in this heterogeneous generalization can been different as for the KNF model (see section (4.2.2). Otherwise, the heterogeneity on the external fields, introduce different affinities in the binding of the ligands by the multiple active sites of the macromolecule. This can have an effect in the global cooperativity of the systems. Indeed recent experimental findings [67, 68] have highlighted, chemical heterogeneity play a crucial role in determining cooperative effects. Modern single-molecule methods show that heterogeneity exists in many instances: the affinities for a ligand, for example, can also vary in a ensemble of macromolecules which have the same binding cooperativity [67]. One of the consequences of this property is that summing up the individual binding curves, to obtain the overall curve for the ensemble, can lead to a measure of cooperativity (Hill number, for instance) which is less than the cooperativity parameter of single macromolecules (see Fig. 4.13). In the next part of the section, we take into account an extension of the model introducing the heterogeneity among the couplings. We introduce for this aim a formal description of each active site $i = 1, .., N$ by a string ξ_i^μ of length P, $\mu = 1, .., P$.

Each element of the strings encode a physical-chemical features of the particular binding sites (for instance: polar or non-polar, containing or not a given subgroup, etc.). We introduce a functional, which associates to any couple of strings (ξ_i, ξ_j) a proper measure of their coupling strength $J_{ij} = \mathrm{J}(\xi_i, \xi_j)$. We assume the strings to be binary, i.e. for any entry $\mu = 1, ..., P$, $\xi_i^\mu \in \{0, +1\}$. To simulate heterogeneous couplings, the entries in the strings are supposed to be random, and the probability distribution is chosen, seeking for simplicity, as

$$P(\xi_i^\mu = 0) = \frac{1-a}{2}, \qquad P(\xi_i^\mu = 1) = \frac{1+a}{2}, \tag{4.32}$$

where the parameter $a \in [-1, +1]$ tunes the similarity between strings: for $a = \pm 1$ all the strings coincide and inhomogeneity in couplings is lost, and the coupling assumes its maximum value; for $a = 0$ strings are purely random and the inhomogeneity is maximum. For the choice of the couplings, we focus on a similarity-based interaction which enhances

the interaction between similar strings, such that

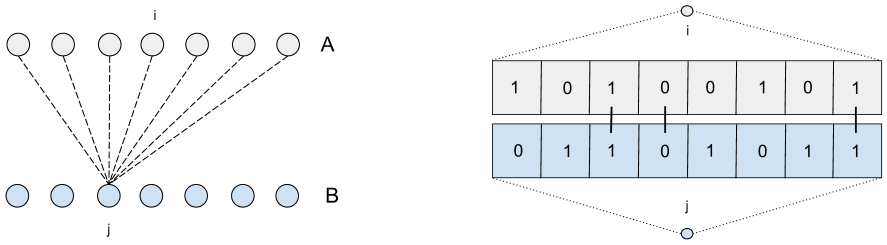
$$J_{ij} = J_0 \frac{1}{P} \sum_{\mu=1}^{P} (\xi_i^\mu \xi_j^\mu + \bar{\xi}_i^\mu \bar{\xi}_j^\mu), \tag{4.33}$$

where we defined $\bar{\xi} = 1 - \xi$ and J_0 is a tuning parameter that regulates the importance of the couplings (Fig. 4.12).

The choice of the interactions resemble very close the hebbian rule in the hopfield model (see section 1.3), indeed is the same recasting the strings $\phi = 2\xi - 1$ apart for a constant term:

$$J_{ij} = J_0 \frac{1}{P} \sum_{\mu=1}^{P} \frac{1}{2} (\phi_i^\mu \phi_j^\mu + 1), \tag{4.34}$$

Figure 4.12: Couplings between elements in different parties.



(a) Bipartite couplings: every element of A interacts with elements in B, and viceversa.

(b) The coupling is determined by the number of overlapping bits between two strings.

The string entries ξ_i^μ are all independent and "quenched[3]" and we denote with $\mathbb{E}$ the expectation with respect to their random values. The resulting average interaction strength is given by $\mathbb{E}J_{ij} = J_0(1 + a^2)/2$ with variance $\sigma_J^2 = J_0^2(1 - a^4)/4P$, so that for a large number of bits P, the coupling distribution approaches to a delta peaked at $\bar{J}$[4]. As a consequence of the coupling rule of Eq. 4.34, the larger the similarity shared by the two strings ξ_i and ξ_j, the stronger the interaction between sites i and j. In particular, for the positively cooperative system, i.e. $J_0 > 0$, the higher the magnitude $\langle J_{ij} \rangle$ and the higher the probability of finding sites i and j both occupied; viceversa, for the negatively cooperative system, i.e. $J_0 < 0$, the higher the magnitude $\langle J_{ij} \rangle$ and the higher the probability of finding one site occupied and the other empty. Finally, the particular choice $J_0 = 0$ recovers again Michaelis-Menten behavior. Clearly, the introduction of the strings associated to the sites is not the only possible approach to bring in heterogeneity among couplings. The parameter we are interested in is, in the end, the heterogeneity tuning parameter a.

As we have seen before, it is possible to associate an effective energy to a system whose

[3]See also part II for a detailed analysis of the meaning of this assumption.

[4]We stress that, whatever the possible scaling between P and N, e.g. $P \sim N^\gamma$, $\gamma > 0$, due to CLT convergence, the distribution of the couplings, hence of the inhomogeneity among ligands, becomes a Gaussian, according with [67].

sites interact with such couplings, and with a concentration of ligands α:

$$H(\sigma,\alpha;\xi) = -\frac{J_0}{N}\sum_{i\in A}\sum_{j\in B}\left(\frac{1}{P}\sum_{\mu=1}^{P}(\xi_i^\mu\xi_j^\mu + \bar{\xi}_i^\mu\bar{\xi}_j^\mu)\right)\sigma_i\sigma_j - \frac{1}{2}\log(\alpha)\sum_{i}^{N}\sigma_i. \tag{4.35}$$

We remind here that the concentration (and consequently the chemical potential) can be rescaled by taking $\alpha \to \alpha/\alpha_0$, with suitable α_0. The probability in this case depends also on the random strings, and is given by

$$P(\sigma;\xi) = \frac{1}{Z}\sum_{\sigma}\exp\left(-H(\sigma,\alpha;\xi)\right), \quad Z = \sum_{\sigma}P(\sigma;\xi). \tag{4.36}$$

One can associate to this a free energy $F = -\frac{1}{N}\log Z$ and compute, for instance, the fraction saturation of a given system in presence of the "disorder" due to the random bits, by properly deriving F with respect to $\log(\alpha)$. This would be a value depending on the particular choice of the strings, and one should then take its average $\mathbb{E}$ with respect to ξ, which will be denoted by the brackets $\langle . \rangle$. Note that in this case there is a double process of averaging.

We show here only the solution of the statistical problem, for the details of the computation see the appendix B.2. The equilibrium state of the system is a simply generalization of the homogeneous case, we have two coupled self-consistence equations for the magnetization in the two parties:

$$\langle m_A\rangle = \tanh\left(J_0\rho_B\frac{1+a^2}{2}\langle m_B\rangle + h\right) \tag{4.37}$$

$$\langle m_B\rangle = \tanh\left(J_0\rho_A\frac{1+a^2}{2}\langle m_A\rangle + h\right) \tag{4.38}$$

$$\tag{4.39}$$

while as usual the saturation function is obtained by $\theta = (1 + \rho_A\langle m_A\rangle + \rho_B\langle m_B\rangle)/2$.

This expression, which returns the average fraction of occupied sites corresponding to the equilibrium state for the system, is analogous to Eq. (4.17), discussed in section 4.2.1. The only difference is in the coupling term in the argument of the hyperbolic tangent, which is replaced here with the average coupling divided by two, since in this case each site interacts with only $N/2$ sites. The resulting fraction saturation is continuous for $J < J_c = J_0(1+a^2)/4$, while for $J = J_c$ we have the scenario, previously depicted, corresponding to a second order phase transition. Conversely, when $J > J_c$ and $\alpha = \alpha_c = 1$, transition is first order, that is the fraction saturation θ becomes discontinuous. The parameter a, introduced for the distribution of bits in the strings that regulate the couplings, acts as a heterogeneity parameter for the interactions: as we have seen, inhomogeneity in the strings is maximum for $a = 0$ and the couplings assume their minimum value in this case, vice versa for increasing $|a|$ the interactions increase, and $|a| = 1$ corresponds to the less homogeneous situation, with maximum coupling. Moreover, the distribution of couplings assumes a Gaussian form for a large number of bits codified in the strings, and the width of this distribution (as well as the average value) depends on the heterogeneity parameter a: the more its absolute value is near to one, the less is the width (but the larger is the average interaction strength), while for small $|a|$ the width of the distribution, and so heterogeneity, is maximum (with the minimum value of the average interaction).

Indeed the effect of the heterogeneity encoded in the parameter a is to rescale the interaction. At fixed interaction parameter J_0, it changes the global cooperativity of the

system. For $|a|$ approaching zero the inhomogeneity increases and the systems is less cooperative. As shown in the positive cooperative case by the rescaling of the Hill number :

$$n_H^{het} = \frac{1}{1 - J\left(\frac{1+a^2}{2}\right)} \tag{4.40}$$

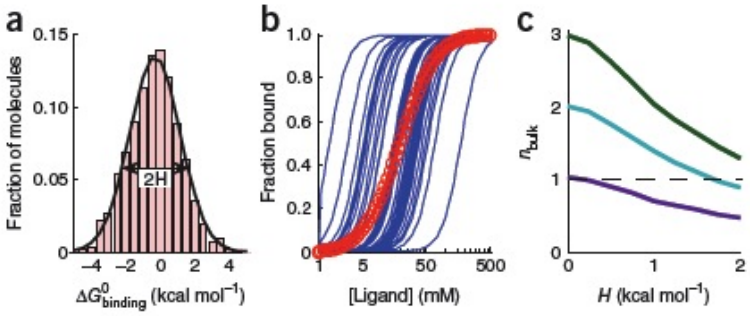


Figure 4.13: (a) Simulated distribution of ligand-binding chemical potentials ΔG^0, with a heterogeneity parameter H. (b) Theoretical Hill binding isotherms (blue lines) with Hill coefficient $n = 3$. The resulting bulk cooperativity parameter for the best Hill fit to the bulk binding curve (red line) is $n_{bulk} = 1.6$ for this example. (c) Reduction of the bulk Hill coefficient as a function of the heterogeneity parameter. From [67]

4.5 Conclusion

In this second part of the thesis we described a mean field statistical mechanics approach to model cooperativity in biochemistry.

Cooperative behavior is a widespread phenomenon in biochemical reactions involving the binding of ligand molecules to larger biopolymers with multiple docking sites. The reason of such ubiquity is that it constitute a fundamental tool to regulate and adapt the chemical response of biological systems.

Statistical mechanics offers a valuable approach as, from its basic principles, it aims to figure out collective phenomena, possibly overlooking the details of the interactions to focus on the very key features. Indeed, a statistical mechanics description of reaction kinetics has already been paved through theoretical models based on linear Ising chains [55], spin lattices with nearest neighbors interactions [69], transfer matrix theory [55, 69] and structural probabilistic approaches [70].

We expand such statistical mechanics picture toward a mean-field perspective [52] by assuming that the interactions among the system constituents are not limited by any topological or spatial constraint, but are implicitly taken to be long-ranged, as in a system that remains spatially homogeneous. Neglecting the details of the interaction mechanism and taking into account only the essential ingredients for the description of the collective phenomena, an effective strength among sites and the chemical potential, we achieved two goals.

First unifying the collective behaviors previously described by different phenomenological law into a clear, unified, theoretical scheme. All cooperative behaviors are recover by varying the parameters as different phases of the model. Showing different qualitatively regimes in the cooperative and in the anti-cooperative case. Second, we get an analytical

expression for the saturation function which is successfully compared with recent experimental findings, taken from a plethora of different contexts to check robustness.

Furthermore, recent experimental findings, due to modern single-molecule methods, highlight the importance of heterogeneity in determining cooperative effects. We showed how to exploit this model strategy to extend the study to the effect of heterogeneity in the couplings and in the chemical potential of the single molecules.

Moreover, this approach permits to code collective behavior of biosystems into a cybernetical framework. In fact, cybernetics, meant as the science dedicated to the understanding of self-organization and emergent communication among the constituents of a system, can be naturally described via (mean-field) statistical mechanics [9]. Thus, the latter provides a shared formalism which allows to automatically translate chemical kinetics into cybernetics and vice versa.
In particular, as showed detailed in [6], saturation curves in chemical kinetics mirror transfer functions of different fundamental electronic devices. A conceptual and practical mapping between kinetics of ultra-sensitive, cooperative and anti-cooperative reactions, with the behavior of analog-to-digital converters, saturable amplifiers and flip-flops respectively, highlight how statistical mechanics can act as a common language between electronics and biochemistry.

Part III

Inference statistical approach to the study of the *in vivo* affinity maturation of antibodies

Chapter 5

Elements of the biology of antibodies

This chapter is devoted to a short overview on the basic biology of the antibodies. In particular, we focus on the process of affinity maturation, that is the ensemble of processes that occurs in immune system to produce high responsive antibodies. The aim is to give to the reader unaware of this topics the necessary information to understand the following section. For an extensive account of the topic, we refer to standard textbooks, see for instance [71] and [72]. Finally, it is important to point out that many aspects of the affinity maturation are not clear and undergoes to an intense research and debate [73].

5.1 Adaptive immune system

The immune system is a system of integrated biological structures and processes within an organism that protects against disease. To function properly, an immune system must detect a wide variety of agents, from viruses to parasitic worms, and distinguish them from the organism's own healthy tissue. This task, that is one of the most intriguing and elaborate, is called the self/non-self discrimination. Pathogens can rapidly evolve and adapt, and thereby avoid detection and neutralization by the immune system. One strategy, for instance, is to mimic organism self molecular features to escape the radar of the immune system. The unavoidable function of detecting agents potentially dangerous for the organism needs a careful regulation that involved a huge number of cells and process.

Adaptive immune system is composed by all the process and cells involved in a response that requires the recognition of specific "non-self" antigens (small fragments of the pathogen) during a process called antigen presentation. Antigen specificity allows for the generation of responses that are tailored to specific pathogens or pathogen-infected cells. The ability to mount these tailored responses is maintained in the body by "memory cells". In case a pathogen infect the body more than once, these specific memory cells are used to quickly eliminate it.

The major functions of the acquired immune system include:

- the recognition of specific "non-self" antigens in the presence of "self", during the process of antigen presentation.
- the generation of responses that are tailored to maximally eliminate specific pathogens or pathogen-infected cells.
- the development of immunological memory, in order to quickly eliminate a pathogen if subsequent infections occur.

The cells of the adaptive immune system are special types of white blood cells (leukocyte), called lymphocytes. B cells and T cells are the major types of lymphocytes and are derived from hematopoietic stem cells in the bone marrow.

Both B cells and T cells carry receptor molecules that recognize specific targets.
T cells recognize a "non-self" target, such as a pathogen, only after antigens have been processed and presented in combination with a "self" receptor called a major histocompatibility complex (MHC) molecule. There are two major subtypes of T cells: the killer T cell and the helper T cell. Killer T cells only recognize antigens coupled to Class I MHC molecules, while helper T cells only recognize antigens coupled to Class II MHC molecules. These two mechanisms of antigen presentation reflect the different roles of the two types of T cell.
In contrast, the B cell (BCs) antigen-specific receptor is an antibody molecule on the B cell surface, and recognizes whole pathogens without any need for antigen processing. Each lineage of B cell expresses a different antibody, so the complete set of B cell antigen receptors represent all the antibodies that the body can manufacture.

A B cell identifies pathogens when antibodies on its surface bind to a specific foreign antigen. This antigen/antibody complex is taken up by the B cell and processed by proteolysis into peptides. The B cell then displays these antigenic peptides on its surface MHC class II molecules. This combination of MHC and antigen attracts a matching helper T cell, which releases lymphokines and activates the B cell. As the activated B cell then begins to divide, its offspring (plasma cells) secrete millions of copies of the antibody that recognizes this antigen. These antibodies circulate in blood plasma and lymph, bind to pathogens expressing the antigen and mark them for destruction by complement activation or for uptake and destruction by phagocytes. Antibodies can also neutralize challenges directly, by binding to bacterial toxins or by interfering with the receptors that viruses and bacteria use to infect cells.
This response is called the humoral immune response, whereas T cells are involved in cell-mediated immune response.

5.2 Antibody

As we have seen, the antibodies are key elements in the immune response. An antibody (Ab) is a protein produced by B-cells, also known as an immunoglobulin (Ig), on surface as membrane receptor or on solution.

Antibodies contribute to immunity in two main ways:

- neutralizing antigen, they prevent pathogens from entering or damaging cells by binding to them and interfering in the infection process ;

- targeting for effector cells and processes, they stimulate removal of pathogens by macrophages and other cells by coating the pathogen and they trigger destruction of pathogens by stimulating other immune responses such as the complement pathway.

All these functions depends on the specificity and the high affinity of the binding with the antigen. This two features are accomplished by the presence of a big variability in the repertoire of antibodies and by selective pressure to bind the antigen.

The structure of the immunoglobulins achieve two main tasks: to have a modular organization to permits the genetic recombination mechanism, that is the main source of the variability in the repertoire, and to maintain the stability gives this huge variations

of parts of the sequence. Flexibility to structural change on one side, maintaining folding stability for different sequence arrangements on the other side.

The immunoglobulin is a large Y-shaped glycoproteins, composed by four polypeptide chains: two identical light chains (L) of length around 210-220 amino acid (can be different from different antibodies) and two identical heavy chains (H) of length around 450-550 a.a.. Every light chain is bind to a heavy chain by a disolphour bond (that link strongly two polypeptides) and others non-covalent bonds. Similar combination of disolphour bond and non-covalent interactions link the two couples (H-L) of light and heavy chains together (see figure 5.1).

The first around 110 a.a. of the amino terminal region are very different for different antibodies and constitutes the variable region, region V (for light chain V_L and V_H for heavy chain). The specificity in the antigenic recognition are due to differences in this part. In the variable region (in human) there are three regions called CDR (complementary determine region) in which are accumulated the differences between different monoclonal antibodies and in which usually lie the binding regions with the antigen, surrounding by three framework region with a rate of variability lower.

The other part of the chain are called constant region, or region C, the constant part of the heavy chain are divide in five different classes called isotypes (called $\mu, \gamma, \alpha, \delta$ and ϵ), while the constant region of the light chain are of two types κ and λ.

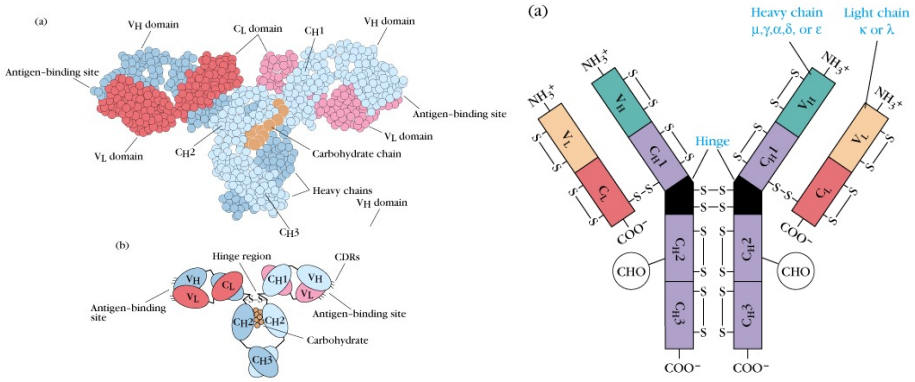


Figure 5.1: On the right, a schematic representation of the structure of an antibody. Are visible the two light chains (L) and the heavy chains (H) connected by the disolphour bonds, in both are highlighted the separation between the variable region (V) and the constant one (C) On the left, is shown the quaternary structure, the division in the subunits and the the typical position of the CDRs and the antigen binding sites. Images taken from [72].

5.2.1 Genetic of antibody

In the DNA of a germline there are different genetic segments codifying for different portions of light and heavy chain. This genetic segment of light and heavy chains can't transcript and traduced in polypeptide unless they are rearranged in functional genes. During the maturation of the lymphocyte one genetic segment per family is casually selected and reorganized in a dynamic genetic process that can generate a huge number of different combinations. The mature B lymphocyte presents only one single functional sequence for the variable region of heavy chain and one for the light chain. The chromosomic DNA of the mature B cell is not any more the DNA in the germline configuration.

The light and heavy chains are codified by different multi-gene families localize in different chromosomes. Each family contains different coding sequences (gene segments)

separate by non-coding regions. The families of the light chains are V,J and C while for the heavy are V,D,J and C. The V,D and J are codifying for the variable region (V,J for the light case) while the D for the constant. In human, there are 51 gene segments in the V family, 27 in D and 6 in J, while for the $\kappa(\lambda)$ light chain there are 40(30) in V and 5(4) in J.

The process of production of one coding sequence for the antibody from the rearrangement of the segment of the DNA of the germline is called V(D)J recombination. It is a very complex molecular process, highly regulate and involve a collection of enzymes some are lymphocyte specific, and some are expressed in many cell types. What happened is that one gene segment for each family is casually selected and cut, after they are joined together sequentially to form a coding sequence for the chain.

A diverse human antibody repertoire is a key element of the acquired immune response and is critical to the effective prevention and clearance of microbial infections. As we have seen, vast diversity in the antibody repertoire is generated initially through a process of combinatorial rearrangement in which gene segments are assembled into a complete immunoglobulin sequence. This initial variability is increased through the use of antigen-driven somatic hypermutation. These affinity maturation processes result in the creation of distinct memory populations that contain only antigen-experienced B cells. The main sources of variability in the repertoire of a B cells during the lymphocyte maturation are:

- V(D)J recombination
  - junctional flexibility
  - adds of "P" and "N" nucleotides
- association between light and heavy chains

The combinatorial genetic process and the association between light and heavy chains alone give rise to a variability of order of 10^{11} of possible immunoglubilins.

Not all the sequences create by the V(D)J recombination correctly codify for antibody, indeed during the process could occur errors that insert stop codon, giving rise to a not-productive rearrangement. In addition for productive antibodies could be unstable or recognize for self-protein and has too be eliminate. These processes that involved also some intriguing features of immune system like the self-nonself distinction, occur in thymus during the maturation process of the lymphocyte and give rise to a filtering or negative selection of the antibodies repertoire.

5.2.2 Affinity maturation process

While VDJ recombination attributed the initial generation of Ig diversity to the combinatorial rearrangement of gene segments, clonal selection account for the subsequent expansion of B cell clones whose surface Ig "recognized" specific antigens led to the production of serum antibodies.

In fact, after immunization the affinity of the antibodies in serum increased dramatically with time, in a phenomenon known as affinity maturation. This is the consequence of iterative rounds of Darwinian-like selection of high-affinity mutants generated by somatic hypermutation (SHM).

During this process B cell receptor locus undergoes an extremely high rate of somatic mutation that is at least $10^5 - 10^6$ fold greater than the normal rate of mutation across the genome. Variation is mainly in the form of single base substitutions, with insertions and deletions being less common. These mutations occur mostly at "hotspots" in the

DNA, known as hypervariable regions. These regions correspond to the complementarity determining regions (CDR).

The combination of SHM and affinity-based selection thus provides the fine-tuning of low-affinity germline VDJ rearrangements, greatly expanding the range of antigenic determinants to which Igs can bind with high affinity.

Affinity maturation takes place in structures known as germinal centers (GCs). These are developed in the secondary lymphoid organs such as spleen and lymph nodes contain primary follicles . When a B cell recognizes an antigen, migrate from the primary focus of infection into the primary follicles and begin monoclonal expansion give rise to a germinal center.

During proliferation, the B cell receptor locus undergoes an extremely high rate of somatic mutation that is at least $10^5 - 10^6$ fold greater than the normal rate of mutation across the genome. Variation is mainly in the form of single base substitutions, with insertions and deletions being less common. These mutations occur mostly at "hotspots" in the DNA, known as hypervariable regions. These regions correspond to the complementarity determining regions (CDR); the sites involved in antigen recognition on the immunoglobulin.

B cells that have undergone SHM must compete for limiting growth resources, including the availability of antigen. The follicular dendritic cells (FDCs) of the germinal centers present antigen to the B cells. B cells in this stage are in a state of activated apoptosis and compete for survival signals from FDCs that present the antigen, meaning the cells will die unless "rescued" by these survival signals. This rescue process is dependent on the affinity of the antibody to the antigen. That is, if a B cell has mutated to have an antibody with more affinity to an antigen, it will be more likely to survive.

The functional B-cells then have to interact with helper T cells to get final differentiation signals. The interaction with T cells is believed to prevent the generation of autoreactive antibodies.

Over several rounds of proliferation, mutation and selection, the resultant secreted antibodies produced will have effectively increased affinities for antigen.

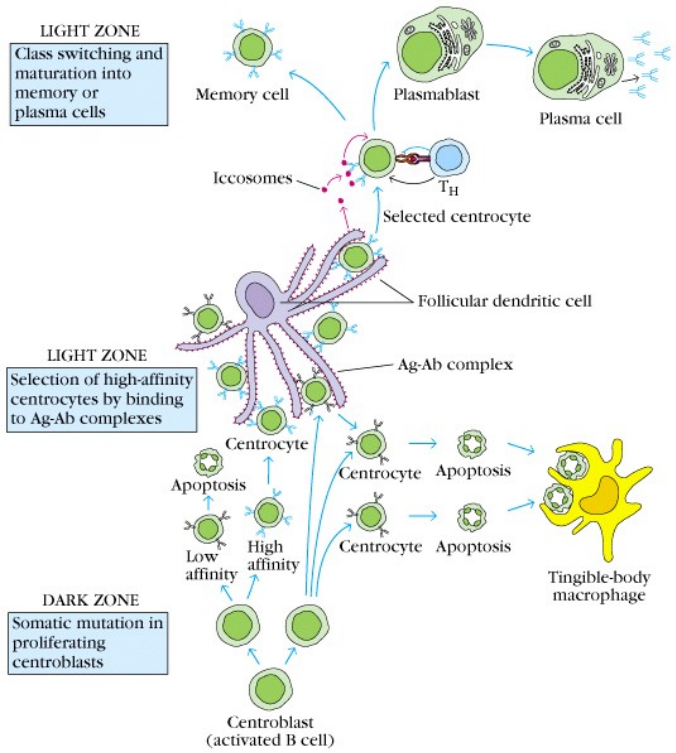


Figure 5.2: Overview of the events occurring in the germinal center. GC is divided in two zones that have different functional roles. A dark zone where the activated B cell proliferates and undergoes somatic hypermutation and a light zone where FDCs present the antigen and the B cell compete for the binding and the rescue signal from apoptosis. Selected BCs return in the dark zone for other turns of mutation and selection or differentiate in memory cell or plasma cell and diffuse in the organism. Image from [72].

Chapter 6

Multivariate Gaussian modeling for the repertoires of antibodies

6.1 Multivariate Gaussian Modeling for protein families

A central problem in structural biology is predicting the tridimensional structure of a protein from the only knowledge of its amino acid sequence. A general recipe for the approach of the problem is still far to be found [74]. Nevertheless, several strategies have been followed in order to restrict the problem. For example in [75], a phenomenological model for the folding of the primary structure is solved through replica method. Here we focus on the research field that deals with the statistical study of correlated substitutions within *multiple sequence alignments* (MSAs) of sequences classified in protein families.

A protein family is defined as a group of evolutionary related proteins that share a common ancestor. Proteins in a family are said to be *homologous* and usually have the same functional role in different species. So, despite the differences in sequence, that can emerge from evolutionary noise, homologous proteins have, in general, rather convergent tridimensional structures.

The main idea beyond correlation analysis is that correlation patterns can be related to structural ones. In particular, contacts between pairs of residues in the native structure could be related to the presence of strongly correlated substitutions in two columns in the MSA of the protein family. Quoting [76], "the basic hypothesis connecting correlated substitution patterns and residue-residue contacts is very simple: If two residues of a protein or a pair of interacting proteins form a contact, a destabilizing amino acid substitution at one position is expected to be compensated by a substitution of the other position over the evolutionary timescale, in order for the residue pair to maintain attractive interaction".

The first attempts to use simple covariance analysis to predict residue-residue contacts ([77], [78]) partially succeeded in identifying some of them but presented a high false positive rate. As often happens in correlation analysis, the main reason beyond this mediocre performance is related to the difficulties in disentangling correlations caused by direct interactions from those produced by indirect ones.

Strategies to overcome the above cited problem have been developed. In particular, a method called *Direct Coupling Analysis* (DCA) has been proposed in [79] and [76]. The main idea beyond DCA is that of inferring a probabilistic graphical model from the alignments so to analyze the inferred interactions instead of the empirical correlations, disentangling direct and indirect contributions to covariances. In the above cited works, the choice of the probabilistic model to infer is based on *maximum entropy principle* [80], according to whom the least constrained model that reproduce single and pair sites

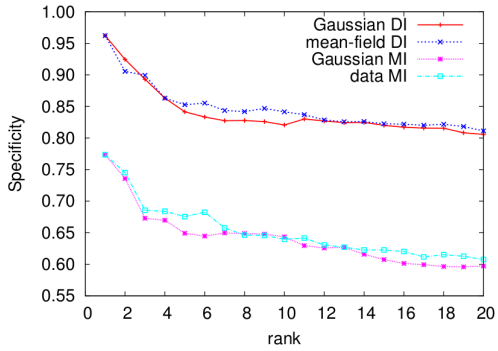


Figure 6.1: DI-ranking-based sensitivity-specificity curves (blue and red) and MI-ranking-based curves (light blue and magenta), averaged over 53 different protein families. The specificity as a function of the rank is defined as the fraction of true positive over the pairs with rank smaller than the indicated one. Gaussian and mean-field Potts method are found, up to small fluctuations, to perform equally well, with a significant and systematic improvement over MI-based methods. Figure from [81].

amino acids frequencies in the alignment is the Inverse Potts model. This model allows to define a scalar quantity called *Direct Information* (DI) that is characteristic of every pair of positions in the alignment and that is correlated to the magnitude of the inferred interaction between them. High DI couples are in fact a good predictor of native contacts.

In [79] a message passing approach is used in order to solve the inference problem paying the price of slow computational times; in [76] the inference is performed within a mean field approximation that would be exact only in the limit of very weak interactions. Both works represented a breakthrough for the performances of contact predictions based on residues coevolution.

Starting from the ideas of the above cited works, some of the authors propose an even simpler probabilistic approach to the problem. That consists of identifying the relevant (real) variables with the fraction of times a given amino acid is present in a given position of an alignment of homologous proteins and to hypothesize that they fluctuate according to a simple multivariate Gaussian distribution. This (strong) assumption permits to completely solve the inference problem by exact analytical computations. This method, that will be soon published in [81] under the name of *Multivariate Gaussian Modeling* (MGM), allows to achieve competitive performances for contact prediction. The fidelity of the method is highlighted in Figure 6.1 where the average *True Positives rate* (TP-rate, or specificity) of the Direct Information computed with both MGM and mean field Inverse Potts model is compared with that of the Mutual Information (MI) that represents a naïve pairs covariance analysis.

In the next Section, we expose the mathematical details of MGM [82].

6.1.1 The mathematical method

In a MSA of P, sequences (whose length after the alignment is indicated with N) are formed by the 20 letters coding for the different amino acids, and may contain alignment gaps ("$-$"), such that the total alphabet size is $q = 21$.

A MSA can be mapped in a binary data set composed by a $P \times N \times (q-1)$ object $(x_i^a)^\mu$ where the subscript $i \in \{1, \dots, N\}$ runs over different amino acid residues, the superscript $a \in \{1, \dots, q-1\}$, runs over different symbols and the superscript $\mu \in \{1, \dots, P\}$ runs

over different sequences. For example, if for protein μ, site i displays the $a-th$ letter, then $\vec{x}_i^{\mu} = (0,\ldots,0,1,0,\ldots,0)$ will be the unitary vector having only the $a-th$ non null component equal to 1. As in a MSA in every position a symbol (letter or *dash*) is present, while encoding it in a vector $\vec{x}_i^{\mu}$ of length $q = 21$, so one symbol is can be eliminated and the vector of length $q-1$ $\vec{x}_i^{\mu} = (0,\ldots,0)$ will indicate a *dash* in the position i of the $\mu-th$ aligned sequence. The symbol that is eliminated while passing from the MSA to the x representation is arbitrary and we choose conventionally it to be the *dash*. We will refer to this freedom in the parametrization of the data as *gauge invariance*.

Furthermore, we promote the variable x to be real numbers ($x \in \mathbb{R}$) and we assume that the sequences $(x_i^a)^{\mu}$ are drawn from a Multivariate Gaussian distribution:

$$P(\{x\}|\{\Sigma,\mu\}) = \frac{1}{\mathcal{N}} \exp\left[-\frac{1}{2}\sum_{ij,ab}(x_i^a-\mu_i^a)(\Sigma_{ij}^{ab})^{-1}(x_j^b-\mu_j^b)\right] =$$

$$= P(\{x\}|\{J,H\}) = \frac{1}{Z}\exp\left[\frac{1}{2}\sum_{ij,ab} x_i^a J_{ij}^{ab} x_j^b + \sum_{i,a} H_i^a x_i^a\right]$$

where we expressed the Gaussian distribution in the parameters $\{J,H\}$ related to the mean vector μ and the covariance matrix $\hat{\Sigma}$ through the relations: $\hat{J} = -\hat{\Sigma}^{-1}$ and $H_i = -\sum_{j,b}\mu_j^b J_{ij}^{ab}$.

In this framework we are interested in implementing a Maximum Likelihood (ML) approach to infer the sets of the model parameters. This standard method selects the set of values of the model parameters that maximizes the likelihood function. The likelihood of a set of parameter values, $\{J,H\}$, given the data $\boldsymbol{x}$, is equal to the probability of those observed outcomes given those parameter values, that is $\mathcal{L}(J,H|x) = P(x|J,H)$. Intuitively, this maximizes the "agreement" of the selected model with the observed data, maximizing the probability of the observed data under the resulting distribution. In practice it is often more convenient to work with the logarithm of the likelihood function, the log-likelihood. See [80] for an extensive discussion on the topic.
Considering this, the log-likelihood of our model parameters $\{J,H\}$ given the data can be written as:

$$\begin{aligned}
\mathcal{L}(\{J,H\}|\text{DATA}) &= \sum_{\mu=1}^{P}\left[-\frac{1}{2}\sum_{ij}\vec{x}_i^{\mu}\cdot\hat{J}_{ij}\cdot\vec{x}_j^{\mu} + \sum_i \vec{H}_i\cdot\vec{x}_i^{\mu} - \log Z(J,H)\right] \\
&= \sum_{\mu=1}^{P}\left[-\frac{1}{2}\sum_{i,j,a,b}(x_i^a)^{\mu} J_{ij}^{ab}\left(x_j^b\right)^{\mu} + \sum_{i,a} H_i^a (x_i^a)^{\mu} - \log Z(J,H)\right] \\
&= P\left[-\frac{1}{2}\sum_{i,j,a,b} J_{ij}^{ab}\langle x_i^a x_j^b\rangle_{\mathrm{D}} + \sum_{i,a} H_i^a\langle x_i^a\rangle_{\mathrm{D}} + \right. \\
&\quad \left. -\langle\log Z(J,H)\rangle_{\mathrm{D}}\right] \qquad (6.1) \\
Z(J,H) &= \int\prod_{i,a=1}^{N} dx_i^a \exp\left[-\frac{1}{2}\sum_{i,j,a,b} J_{ij}^{ab} x_i^a x_j^b + \sum_{i,a} H_i^a x_i^a\right] \\
&= \exp\left[-\frac{1}{2}\log\det J + \frac{1}{2}\sum_{i,j,a,b} H_i^a \left(J^{-1}\right)_{ij}^{ab} H_j^b + const\right], \qquad (6.2)
\end{aligned}$$

where we show both vectorial and component-wise notation and $\langle O \rangle_D := \frac{1}{P}\sum_{\mu=1}^{P} O^\mu$ is the empirical average of a generic observable O over the different experiments. Upon combining together Equations 6.1 and 6.2, we finally obtain up to irrelevant constants:

$$\begin{aligned}\mathcal{L}(\{J,H\}|\mathrm{DATA}) &= P\Big[\sum_{i,j,a,b} -\frac{J_{ij}^{ab}}{2}\langle x_i^a x_j^b\rangle_{\mathrm{D}} + \sum_{i,a} H_i^a \langle x_i^a\rangle_{\mathrm{D}} \\ &\quad - \sum_{i,j,a,b} \frac{H_i^a \left(J^{-1}\right)_{ij}^{ab} H_j^b}{2} + \frac{1}{2}\log\det J\Big] \end{aligned} \tag{6.3}$$

Maximum likelihood estimation of the parameters

Let us define the usual covariance matrix as

$$C_{ij}^{ab} = \langle x_i^a x_j^b\rangle_D - \langle x_i^a\rangle_D \langle x_j^b\rangle_D \,. \tag{6.4}$$

We note that, while $\langle x_i^a\rangle_D$ represents the fraction of times the $a-th$ symbol is present in position i in the MSA, while the second one gives informations about the correlated variation of residues in the same $(i=j)$ or different $(i\neq j)$ positions.

We are now ready to maximize component-wise the log-likelihood $\mathcal{L}$ with respect to the model parameters J and H:

$$0 = -\frac{\partial\mathcal{L}}{\partial J_{ij}^{ab}} = \langle x_i^a x_j^b\rangle_D - \sum_{k,l,c,d} H_k^c \left(J^{-1}\right)_{ki}^{ca}\left(J^{-1}\right)_{jl}^{bd} H_l^d - J_{ij}^{-1}\,, \tag{6.5}$$

$$0 = \frac{\partial\mathcal{L}}{\partial H_i^a} = \langle x_i^a\rangle_{\mathrm{D}} - \sum_{j,b}\left(J^{-1}\right)_{ij}^{ab} H_j^b \,. \tag{6.6}$$

After inserting Equations 6.5 into Eq. 6.6, we get the final relation that gives the maximum likelihood (ML) estimation of the parameter J and H as a function of the data: Defining the usual empirical covariance matrix as

$$C_{ij}^{ab} = \langle x_i^a x_j^b\rangle_D - \langle x_i^a\rangle_D \langle x_j^b\rangle_D \,, \tag{6.7}$$

we get

$$J_{ij}^{ab} = (C^{-1})_{ij}^{ab}\,, \qquad H_i^a = \sum_{j,b} J_{ij}^{ab}\langle x_j^b\rangle \tag{6.8}$$

Direct Information

Once the parameters of the model have been inferred, one scalar quantity for every couple of positions i,j in the alignment can be determined in order to quantify the interaction strength of the sites. With this aim, the Direct Information (DI) has been developed in order to be invariant under gauge transformations.

In order to do that, in the same spirit of what we have done in the previous sections, one could try to infer the ML model in a simplified setting where no interaction is present between variables. In this latter case, the analogous of Equation 6.1 would be

$$\begin{aligned}\mathcal{L}(\{K,L\}|\mathrm{DATA}) &= P\left[-\frac{1}{2}\sum_i \langle \vec{x}_i\cdot\hat{K}_i\cdot\vec{x}_i\rangle_D + \sum_i \vec{L}_i\cdot\langle\vec{x}_i\rangle_D - \langle\log Z(K,L)\rangle_D\right] = \\ &= P\left[-\frac{1}{2}\sum_{i,a,b} K_i^{ab}\langle x_i^a x_j^b\rangle + \sum_{i,a} L_i^a\langle x_i^a\rangle_D - \langle\log Z(K,L)\rangle_D\right], \end{aligned} \tag{6.9}$$

where the model parameter K is the analogous of J and L of H, and from the functional structure of the Likelihood function. The statistical independence among the different proteins is evident. After some simple algebra, one gets that the ML estimation for $\{K, L\}$ is given by

$$\left(K^{-1}\right)_i^{ab} = \langle x_i^a x_i^b \rangle_D - \langle x_i^a \rangle_D \langle x_i^b \rangle_D \,, \tag{6.10}$$

$$L_i^a = \sum_b K_i^{ab} \langle x_i^b \rangle_D \,, \tag{6.11}$$

which are the analogous of Equations 6.5 and 6.6 respectively. Given the ML estimators K, L, we can define the probability of a protein state x given K, L as

$$P^{\text{ind}}(x|K,L) \propto \prod_i \exp\left(-\frac{1}{2}\vec{x}_i \cdot \hat{K}_i \cdot \vec{x}_i + \hat{L}_i \cdot \vec{x}_i\right). \tag{6.12}$$

In a nutshell, the idea is to compare an effective two-interacting-sites model whose interaction matrix is the J computed in Equation 6.5 with P^{ind}. Let us define the probability distribution of the effective two-interacting-sites model as

$$\begin{aligned} &P_{ij}^{\text{dir}}(\vec{x}_i, \vec{x}_j | \hat{J}_{ij}, \hat{W}_i, \hat{W}_j, \vec{G}_i, \vec{G}_j) \propto \\ &\propto \exp\left(-\vec{x}_i \cdot \hat{J}_{ij} \cdot \vec{x}_j - \frac{1}{2}(\vec{x}_i \cdot \hat{W}_i \cdot \vec{x}_i + \vec{x}_j \cdot \hat{W}_j \cdot \vec{x}_j) + (\vec{G}_i \cdot \vec{x}_i + \vec{G}_j \cdot \vec{x}_j)\right) \\ &\propto \exp\left(-\sum_{ab} J_{ij}^{ab} x_i^a x_j^b - \frac{1}{2}\sum_{ab}(W_i^{ab} x_i^a x_i^b + W_j^{ab} x_j^a x_j^b) + \sum_a (G_i^a x_i^a + G_j^a x_j^a)\right) \,, \end{aligned} \tag{6.13}$$

where the parameters $\hat{W}_i, \vec{G}_i$ are chosen such that the single sites marginal of P^{dir} match the single sites probabilities P^{ind}. We do this in two steps by imposing: (i) the equality of the connected second moment, (ii) the equality of the first moment. Let us note that the second connected component in a Gaussian measure is just the inverse of the covariance matrix:

$$\Sigma_{dir} := \langle x_i^a x_j^b \rangle_{dir} - \langle x_i^a \rangle_{dir} \langle x_j^b \rangle_{dir} = \begin{pmatrix} \hat{W}_i & \hat{J}_{ij} \\ \hat{J}'_{ij} & \hat{W}_j \end{pmatrix}^{-1} \tag{6.14}$$

$$\Sigma_{ind} = \langle x_i^a x_j^b \rangle_{ind} - \langle x_i^a \rangle_{ind} \langle x_j^b \rangle_{ind} = \begin{pmatrix} \hat{K}_i & 0 \\ 0 & \hat{K}_j \end{pmatrix}^{-1} = \begin{pmatrix} \hat{K}_i^{-1} & 0 \\ 0 & \hat{K}_j^{-1} \end{pmatrix} \tag{6.15}$$

note that in this representation Σ_{dir} and Σ_{ind} are $2(q-1) \times 2(q-1)$ matrices. We can now impose the equality of the diagonal part of the two matrices:

$$\hat{K}_j^{-1} = \left(\hat{W}_j - \hat{J}'_{ij} \cdot \hat{W}_i^{-1} \cdot \hat{J}_{ij}\right)^{-1} \,, \tag{6.16}$$

$$\hat{K}_i^{-1} = \left(\hat{W}_i - \hat{J}_{ij} \cdot \hat{W}_j^{-1} \cdot \hat{J}'_{ij}\right)^{-1} \,. \tag{6.17}$$

Such equations can be decoupled w.r.t. the variables $\hat{W}_i$ e $\hat{W}_j$:

$$\hat{W}_j = \hat{K}_j + \hat{J}'_{ij}\left(\hat{K}_i + \hat{J}_{ij}\hat{W}_j^{-1}\hat{J}'_{ij}\right)^{-1}\hat{J}_{ij} \,, \tag{6.18}$$

$$\hat{W}_i = \hat{K}_i + \hat{J}_{ij}\left(\hat{K}_j + \hat{J}'_{ij}\hat{W}_i^{-1}\hat{J}_{ij}\right)^{-1}\hat{J}'_{ij} \,; \tag{6.19}$$

and, after some manipulation, one obtains two matricial equations:

$$\begin{aligned} A^2 - A - M = 0 & & (6.20)\\ A &= \hat{K}_i^{-1}\hat{W}_i \\ M &= \hat{K}_i^{-1}\hat{J}_{ij}\hat{K}_j^{-1}\hat{J}'_{ij}\,, \\ B^2 - B - N = 0 & & (6.21)\\ B &= \hat{K}_j^{-1}\hat{W}_j \\ N &= \hat{K}_j^{-1}\hat{J}'_{ij}\hat{K}_i^{-1}\hat{J}_{ij}\,. \end{aligned}$$

Solutions to Eq. 6.20 and Eq. 6.21 can be obtained by diagonalizing matrices M and N and solving $2(q-1)$ independent equations. We are now ready to express the Direct Information as the Kullback-Leibler (KL) divergence of the two Gaussian measures P^{dir} and P^{ind}:

$$\begin{aligned} DI(J,K)_{ij} &\equiv KL(P^{dir}||P^{ind}) = \\ &= \frac{1}{2}\Big\{\mathrm{tr}(\Sigma_{ind}^{-1}\Sigma_{dir}) + (\langle X\rangle_{ind} - \langle X\rangle_{dir})^T\Sigma_{ind}^{-1}(\langle X\rangle_{ind} - \langle X\rangle_{dir}) + \\ &\quad - \log\left(\frac{\det\Sigma_{dir}}{\det\Sigma_{ind}}\right) - 2(q-1)\Big\} \\ &= -\frac{1}{2}\log\left(\frac{\det\Sigma_{dir}}{\det\Sigma_{ind}}\right) \qquad (6.22) \end{aligned}$$

since in our case $\langle X\rangle_{ind} = \langle X\rangle_{dir}$ and $\mathrm{tr}(\Sigma_{ind}^{-1}\Sigma_{dir}) = 2(q-1)$.

Data regularization (pseudocounts)

For the maximum likelihood estimations 6.5 and 6.6 of the parameters to be possible, the covariance matrix needs to be full rank in order to be inverted. As fluctuation in MSAs are generally limited, usually the experimental covariance matrix is rank deficient. To overcome this problem a regularization procedure has to be implemented. The simplest method for that is adding to the sample a number λ of fictitious sequences in which symbols in every site are fairly drawn from a flat distribution. This reduces to manipulate the data as:

$$\langle x_i^a\rangle_D \longrightarrow (1-\pi)\langle x_i^a\rangle_D + \pi\frac{1}{q}\,, \qquad (6.23)$$

$$\langle x_i^a x_j^b\rangle_D \longrightarrow (1-\pi)\langle x_i^a x_j^b\rangle_D + \pi\frac{1}{q^2} \qquad (6.24)$$

where the parameter

$$\pi \equiv \frac{\lambda}{P+\lambda} \qquad (6.25)$$

that is referred to as *pseudocount* parameter, naturally interpolates between the empirical ($\pi = 0$) and completely random ($\pi = 1$) data.

In [81] it is clarified how the use of a pseudocount based regularization in MGM is equivalent to the choice of a normal inverse Wishart prior (the conjugate prior of the multivariate Gaussian distribution) over the parameters of the MGM.

6.1.2 Multivariate Gaussian Modeling for antibodies diversity: the general idea

The possibility of studying the features of the affinity (or of the neutralization power) of antibodies directed towards a certain antigen as a function of the sequence of the variable region is a fundamental issue in structural immunology.

Unfortunately, the size of the available sets of antibodies, for which both the sequence and the neutralization power toward an antigen is known, is, at the moment, as large as some tenth of antibodies. The typical size of an antibody's variable region is of order $N \sim 10^2$ amino acids; as the contributions of the amino acids to the neutralization power are not independent, the least structured function to describe it would contain at least $(Nq)^2 \sim 10^6$ parameters. So up to now, there is no possibility to estimate the parameters of such a function with the size of the available data sets.

Nevertheless, the recent developing of sequencing techniques (*Deep Sequencing*, *Next Generation Sequencing*), that are able to produce in parallel up to $\sim 10^6$ sequence reads, have opened the possibility of extensive experimental studies of the Abs repertoires in different living organisms. This kind of research is known under the the name of Rep-Seq (*Repertoire Sequencing*) experiments (see [83] for a review on the argument). These techniques have permitted for example to study the complete Igs repertoire of simple organisms such as the zebrafish, whose immune system has only ~ 300000 Abs producing B cells, hence determining its complete *antibodyome* (see [84] and [85] for the maximum entropy analysis of the data). In human, the latter amount to $\sim 10^{9-10}$ cells, so that only limited samples of the entire repertoires are available up to now (see for example [86] for Rep-Seq experiment of Igs in human or [87] for a maximum entropy analysis of TCR repertoire in human).

As explained in more detail in the previous chapter, the features of the populations of B cells in host bodies are determined by the processes of genetic recombination, negative selection that eliminate self-directed antibodies, clonal expansion in response to different antigens that are or were present in the host body, random mutations (somatic hypermutation) and positive selection of antibodies that have an high affinity to antigens (affinity maturation). All these mechanisms interact in a complex manner to determine the B cell population present in a body.

In some cases, it should be possible to manipulate Rep-Seq data sets in order to disentangle the processes described above with the aim of obtaining a sample whose evolution is mainly driven by affinity maturation and clonal expansion toward a specific antigen so that the resulting data set is highly correlated with the affinity (or neutralization power) of the antibodies. Once that this step is fulfilled, a probability distribution that is considered to have generated the sample can be inferred. If this probability distribution is indeed observed to be correlated with the affinity (or neutralization power), then it could be used as a proxy to study the features of the neutralization power as a function of the sequences and in principle to propose sequences of high neutralization power.

Moreover, as pointed out in the previous chapter, the affinity maturation is an evolutionary process in which different B cell clones compete for the antigen in the germinal centers. The study of the statistics of the population of sequences could also unveil interesting features of the fitness function in the space of sequences related to this process and in general of the evolutionary dynamics of the B cells. That could be of interest for a statistical population genetics analysis of the affinity maturation process, for example in the spirit of [88].

Within the above described scenario MGM is a useful tool as it interprets the x representation of the MSA as a (discrete) sampling from a Gaussian distribution whose pa-

rameters can be inferred following the procedure outlined above. This distribution over the real x variable is of the form

$$P(x|J,H) = \frac{1}{Z(J,H)} e^{-E(x|J,H)} \tag{6.26}$$

where the *energy* of the model is defined as

$$E(x|J,H) = -\frac{1}{2} \sum_{ij} \vec{x}_i^{\mu} \cdot \hat{J}_{ij} \cdot \vec{x}_j^{\mu} + \sum_i \vec{H}_i \cdot \vec{x}_i^{\mu} \tag{6.27}$$

which is the log-likelihood apart from the normalization factor $Z(J,H)$ defined in Equation (6.2).

6.2 Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing: a review of the experimental work

The general idea that guides this work is to use the probability distribution inferred over a Rep-Seq dataset as a proxy for the affinity function. This approach has been tested using the experimental data published in [89], where the Rep-Seq experiment is performed together with affinity measurements.

In this Section, the experimental work is reviewed, while the preliminary results of our analysis are exposed in Section 6.3.

HIV-1 displays an enormous genetic diversity and in this resides a great part of the infection's strength. Despite this fact, from 10% to 25% of the patients develop cross-reactive neutralizing antibodies after several years of infection. These individuals are said to have a broadly neutralizing serum.

In a previous work [90], the authors of [89] isolated VRC01, VRC02 and VRC03, three similar broadly neutralizing antibodies (bnAbs) from a patient (donor 45) presenting a broadly neutralizing serum. This *VRC01-like Abs* have been seen to be bind gp120, a membrane glycoprotein used by the virus to attach the CD4 receptor on T lymphocytes, HIV target cells. To avoid neutralization by gp120 directed Abs, during evolution, the virus has developed a complex structure for gp120: Highly variable domains hide the site of attachment gp120-CD4 which is the only part of the protein that is under evolutionary pressure and that, for this reason, has relativley low freedom to mutate. There are structural experimental evidences that VRC01-like bind exactly this particular site on gp120 ([90] and [89]).

VRC01 neutralizes 90% of virus isolates with an average neutralization power $IC_{50} \sim 0.3\,\mu$g/ml while its predicted unmutated germline ancsestor has low affinity for the antigen (dissociation constant in the millimolar range). All VRC01-like Abs display an high level of mutation ($\sim$ 30% that has to be compared with the 5-15% of mutations in average Abs) from the inferred germilene. This underlines that, for this kind of antibodies, the improvement of the neutralization power due to affinity maturation is substantial. This fact leads us to the idea that this could be a good system to study the affinity maturation process.

In [89], the authors isolated other VRC01-like antibodies from another donor (donor 74). An example of them is VRC-PG04. Couples of such bnAbs with the same unmutated germiline ancestor, from which they are mutated at about 30% and that come from

different patients (for example VRC-PG04 from donor 74 and VRC-01 from donor 45), are observed to be very different in sequence ($\sim 50\%$); this underlines the fact that the (relatively) optimal sequence to bind the gp120-CD4 attachment site is not unique and that the evolutionary histories of populations of antibodies under analogous evolutionary pressures may be very different.

The authors performed a crystallographic study on VRC-PG04[1] in complex with the gp120. Comparing this with analogous studies on others VRC01-like Abs, they argue that the most important region for the neutralization are the CDR2 on the heavy chain and the CDR3 on the light chain of the Ab.

All these facts underline that VRC01-like antibodies are an interesting system to study both on the fundamental and practical level. Indeed the authors decided to perform Rep-Seq experiments on donor 45 and donor 74 blood samples with the aim of sampling a part of the antibodies repertoires on these donors depending on the choice of the primers.

The authors choose 454-pyrosequencing as deep sequencing technique as it allow to sequence reads up to a length of 700 bp and so to sequence the whole variable region of both light and heavy chain. The side effect of this choice is that 454-pyrosequencing has a relatively high error rate (1/1000 bp on average) and errors are concentrated on homopolymers.

Moreover, light and heavy chains are translated into different mRNAs molecules; as the sequencing technique capture the mRNA in the sample and mRNAs belonging to different cells are mixed during the procedure, it is only possible to reconstruct separately the light and heavy chains repertoire and there is no way to match the light and heavy chains belonging to the same antibody (B-cell clone).

Data of sequencing experiments on light and heavy chains for donor 45 and two experiments on heavy chains of donor 74 have been deposited to NCBI database[2].

The authors select a set of antibodies highly mutated and close in sequence to VRC-PG04. Then they measured the neutralization power IC50 [3] of 45 successfully produced (out of 70 tried for production) chimeric Abs, in which VRC-PG04 light chain was coupled with heavy chains selected from the highly mutated ones in the sequenced set. The result of the neutralization measurements of 20 HIV-1 isolates, belonging to the clades A,B and C, is that heavy chains that are more similar to VRC-PG04 are in general prone to be (broadly) neutralizing, confirming that several VRC01-like antibodies are present in the sample and that they share some features with the known ones. C50 is not a direct indicator of affinity although the two can be related at least for competitive agonists and antagonists by the Cheng-Prusoff equation.[4] For enzymatic reactions, this equation is:

To summarize, the experimental data available from [89], that we are going to use in the following analysis are:

- Deep sequencing of the heavy chains repertoire (of donor 74). In the sequencing procedure a set of primers are chosen in order to select the germline of the bnAbs

[1] Structure factors and coordinates for antibodies VRC03 and VRC-PG04 in complex with HIV-1 gp120 have been deposited with the Protein Data Bank under accession codes 3SE8 and 3SE9, respectively.

[2] Reference to sequencing data can be found in the Acknowledgment of [89] or at http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?study=SRP006992.

[3] IC50 is not a direct indicator of affinity although the two can be related at least for competitive agonists and antagonists by the Cheng-Prusoff equation. For enzymatic reactions, this equation is:

$$K_i = \frac{IC50}{1 + [S]/K_M}$$

with K_i the binding affinity, $[S]$ the concentration of the substrate (the antigen in this case) and K_m is the Michaelis constant.

VRC-PG04.

- Crystallography of the complex VRC-PG04 antibody and the gp120 protein.
- Measures of neutralization power of 45 chimeric antibodies (with VRC-PG04 light chain) highly mutated against 20 strains of the virus of different clades.

6.3 The Multivariate Gaussian Modeling analysis: preliminary results

In the following, we describe the analysis performed on sequencing data of the repertoire of heavy chains variable region of donor 74 from [89] . This donor has been observed to produce the broadly neutralizing antibody VRC-PG04. The work is still in progress and the result presented here are meant to be preliminary.

The germline of origin of the VRC-PG04 antibody has gene IGHV1-2*02 coding for heavy chain V part and IGHJ2*01 for the region J. As the sequence presents as much as $\sim 30\%$ of mutation from the inferred germline, the D gene of origin is too short to be determined.

For our scopes, from the deposited raw data we have to perform a bioinformatic analysis that will be described in details in [7]. Here we just underline that, starting form the one strand nucleotide sequences, our bioinformatic analysis selects the productive ones and returns a set of amino acid sequences each provided with its multiplicity, i.e. the number of times a nucleotide sequence coding for the same amino acid sequence is present in the set. We point out that, differently from [89], we only retain productive sequences, i.e. sequences for which the V and J genes are in frame and that do not present stop codons. This analysis generates a set 383267 productive amino acid sequences (which reduce to 191661 unique sequences) for the variable part of the heavy chain, provided with the inferred V and J gene of origin. According to the authors of [89], the sequencing primers have been chosen in such a way that, in principle, for all the sequenced reads, the inferred V gene of origin is in the family IGHV1.

Selecting the sequences whose V gene of origin is (one of the alleles of) IGHV1-2 gives a set of 72649 sequences (37839 unique), while the set of sequences that have IGHV1-2 and IGHJ2 as germline genes of origin consist of 6820 sequences (3258 unique).

It has to be kept in mind that sequences and relative proportions of their abundances are strongly affected by sequencing errors and PCR biases. Moreover, despite the fact that the selection of the productive sequences reduces the number of sequences presenting errors, this procedure affects the relative proportions of sequences in the population. So we claim that systematic experimental errors are present and out of our control.

We want to select the antibodies that effectively undergoes affinity maturation against the HIV virus. For this task two consideration are helping us. Almost all the sequences that display a divergence from the germline higher than 25% belong to the set that has IGHV1-2 and IGHJ2 as allelic origin. The broadly neutralizing antibodies individuated by Wu et al. in the blood sample of donor 74 are remarkably highly mutated from the same inferred germline, IGHV1-2 and IGHJ2 genes (that are the same of the VRC-PG04 antibody).

In Figure 6.2 we show the identity/divergence analysis for all the sequence with this germline (see caption for details). In this way is clearly displayed the presence of a cluster of highly mutated sequences more similar to the broadly neutralizing antibody VRC-PG04.

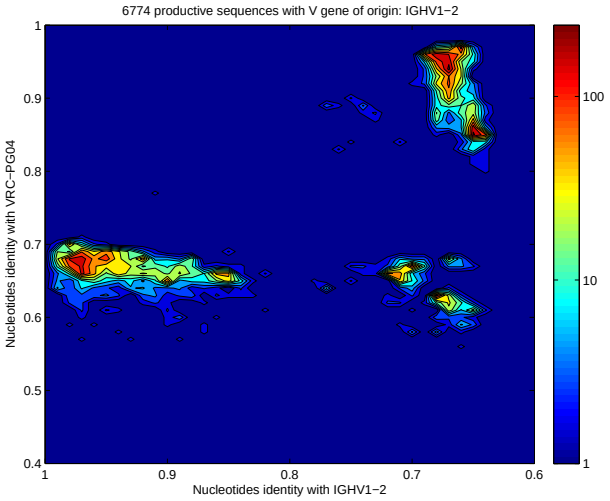


*Figure 6.2: Identity/divergence analysis of Rep-Seq experiment for heavy chains in donor 74 performed in [89]. On the horizontal axis the divergence from the IGHV1-2*02 germline gene of origin is reported while the identity to VRCPG-04 sequence is indicated on the vertical axis. Both quantities are computed at the nucleotides level.*

This cluster is well separated from the cluster of typically mutated sequences, more similar to the V and J germline genes. Remarkably, the same does not happen for Abs with a different IGHV inferred germline gene.

The effective presence of this clustering structure with this two clusters was indeed verified though clustering algorithms as described in the following of this Section.

Clustering algorithms

The set of sequences with inferred germline in the families of IGHV1-2 and IGHJ2 has been submitted to a clustering algorithm that has been proposed in [91]. With this algorithm, it is possible to determine the number of well defined clusters: In fact, varying the parameter λ in the algorithm, the number of clusters in the output develops a plateau exactly when sequences are optimally separated. Figure 6.3 gives the indication that there are two well separated clusters.

Using this last information, we performed a clustering analysis on the same set of sequences (with IGHV1-2 and IGHJ2 allelic origin) by submitting different alignments to the Matlab algorithm `kmeans`, fixing at 2 the input parameter corresponding to the number of clusters.

The consensus sequence of the two clusters were compared with the germline sequence (i.e. to the concatenated sequences of IGHV1-2*02 and IGHJ2*01 genes) and with the VRC-PG04 sequence. As shown in Figure 6.4, the consensus sequence of the normally mutated cluster is similar to the germline genes while for the highly mutated cluster the consensus sequence is similar to that of VRC-PG04. The size of the two obtained clusters are respectively of 3471 sequences (1874 unique) and 3349 sequences (1338 unique). In the following, we will indicate with *clusterVJ* and *clusperPG04* respectively these two sets of sequences.

The different sets in which the sequences have been divided are resumed in Table 6.1.

set description	size	size (unique)
Productive IGHV1 origin	383267	191661
Productive IGHV1-2 origin	72649	37839
Productive IGHV1-2 and IGHJ2 origin	6820	3258
Productive IGHV1-2 and IGHJ2 origin - *clusterVJ*	3471	1874
Productive IGHV1-2 and IGHJ2 origin - *clusterPG04*	3349	1338

Table 6.1: Different sets in which sequences have been divided.

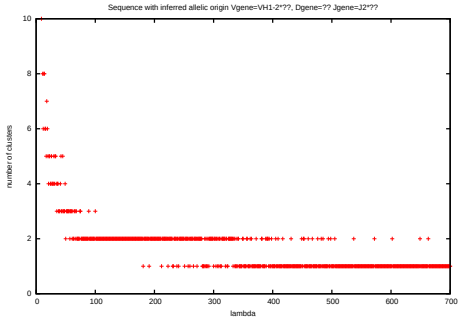


Figure 6.3: Number of clusters as a function of the parameter λ, see [91] for details. The plateau at 2 indicates that an optimal clustering is obtained with two clusters.

Figure 6.4: Alignment of the consensus sequences of the two clusters with the sequences of VRC-PG04 (upper case) and the germline (lower case) .

6.3.1 Comparison between the inferred probability distribution and neutralization power measurements

As already pointed out, in the work of Wu et al. 70 sequenced heavy chains, mostly originating from IGHV1-2, were used to construct chimeric antibodies with the light chain of VRC-PG04. 45 of those have been successfully produced and tested for neutralization power against 20 HIV-1 virus belonging to clades A (6 viruses), B (8 viruses) and C (6 viruses). Although the neutralization power is a complex function of both light and heavy chain sequences (see [89], Figure 4A), these measurements can be used as an approximation of the contribution of the heavy chain to the neutralization power.

The above cited neutralization power measurements (the data relative to which are in Table S19 and S20 in the Supporting Information of [89]) can be compared with the statistical properties of the set. To do this, the highly mutated *clusterPG04* has been selected and a MGM has been inferred on this set considering each sequence with its relative multiplicity. For the 45 Abs that were tested for neutralization power, the IC_{50} [4] has been compared with the sequence energy gives by the inferred model, using equation (6.27).

Inference performed with different values of the pseudocount parameter π display a significant Pearson correlation coefficient between the inferred energy and the neutralization power of single viruses and their average or minimum (see figure 6.5 and figure 6.6 - continuous lines). The viruses in clade A show a better correlation due probably to the infection of the considered patient by a recombinant A/D HIV virus, for the same reason the minimum display a greater correlation then the average [5].

This result is very promising and it indicates that the MGM energy inferred on a proper Rep-Seq dataset could, in general, provides informations about the affinity function landscape in sequences space.

To have a deeper understanding of these correlations, we check that the models inferred from different datasets, in particular the ones depicted in table 6.1, doesn't show the same correlations (see figure 6.7).

Moreover, we repeated the above procedure with a factorized MGM (equation (6.12)) on the same *clusterPG04* set. In order to see if the single site frequency of the residues are enough to provide correlations with the measures of the affinity with the antigen or it's needed the two point correlations (contained in the complete MGM). The result is that the correlation of the energy inferred with the factorized model is less (and less significantly) correlated with the neutralization power then the energy learned with the complete correlated Gaussian model (see Figure 6.6, dashed lines).

The message that should be learned out of the above result is that, as expected, not only single mutations but at least correlated pairs of mutation in the variable region sequence are needed in order to achieve affinity maturation. Special directions in the space of sequences (arising from combinations of the single aligned amino acids) that are more relevant for the affinity can be considered by learning a correlated MGM over the considered set.

Starting from this result we'll investigate, in next sections, the reasons for which the MGM seems to correlate the inferred probability distribution with the neutralization power. Which structural features of the Igs that are important for the neutralization power are captured by MGM? In particular, is it possible to predict the internal contacts through DCA, using the Rep-Seq datasets as it has been done for protein families?

[4] In [89] is not reported the values of IC_{50} greater than 50 and for the following analysis we have considered to fix these values equal to 50.

[5] The minimum of the IC50 values reside in viruses of the clade A.

Furthermore, are the inferred model containing information on the interaction of these antibodies with the antigen gp120? Can we infer the residue in contact with the antigen? Next Sections try to answer these questions.

6.3.2 Recovering the internal contacts

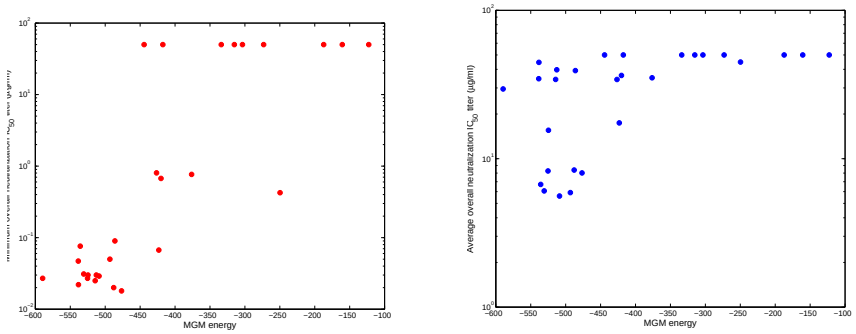
Direct couplings analysis have been developed in order to recover internal native contacts in proteins by analysing multiple sequences alignments of homologous protein families, a natural question is whether it is possible to predict internal contacts of Igs variable domains through coevolution analysis in an individual's repertoire, so by performing DCA treating sets of sequenced Igs as a protein families.

We compare the distance map between internal residues extract from the crystallography study of VRC-PG04 antibody with the DCA predictions of internal contacts.

As shown in figures 6.8 and 6.10, sequences of *clusterPG04* weighted with the related multiplicities are not a good set for the prediction of the internal contacts.

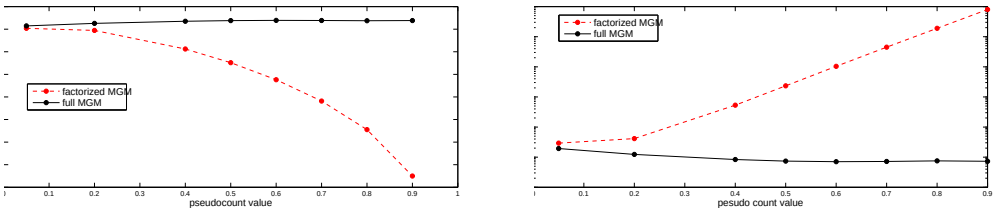
Further attempt can be done on other sets of sequences. As a result, as shown in figures 6.9 and **??**, the best result has been obtained with the largest set, i.e. all sequenced reads without any regards to the inferred germline, other, of course, than the constraint due to the selection of the primers that, as explain select Igs with inferred V gene in the family IGHV1. [6] In any case, also with this larger set, the method doesn't provide a satisfactory performance in recovering the internal contacts of IGs variable domains. This is probably due to the relatively too low degree of variability that is present in the set that does not leave space for covariation of residues in contact.

[6] where is used a reweighting procedure that eliminates the effect of the sequences multiplicity, method described in [76].



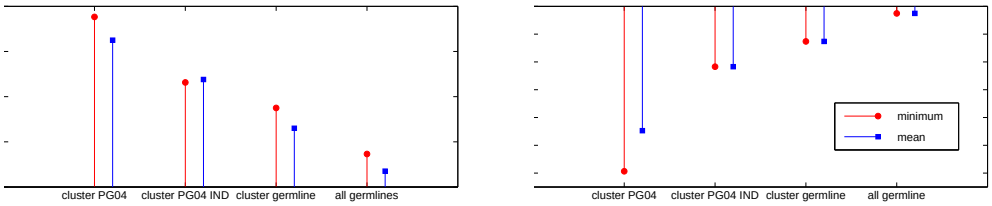
(a) minimum (b) average

Figure 6.5: An example of the comparison between the energy of the MGM inferred on clusterPG04 *and the minimum (Panel (a)) and the average (Panel (b))of the neutralization IC_{50} titer. In both showed examples, the MGM is learned with pseudocount $\pi = 0.5$. The Pearson correlation coefficient is 0.78 for the minimum (p-value 5E-6) and 0.69 (p-value 1.2E-5) for the average.*



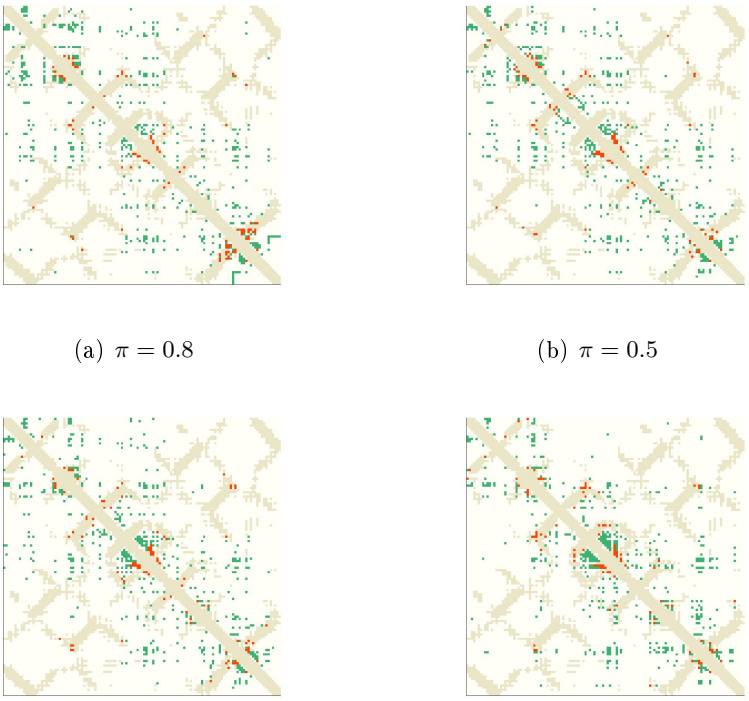
(a) (b)

Figure 6.6: Panel (a): Scaling with the pseudocount of the Pearson correlation coefficient between the energy of the MGM learned on clusterPG04 *and the average IC_{50} neutralization titer against the 20 tested HIV-1 viruses. Black continuous lines correspond to a full MGM while red and dashed lines to a factorized ($J \equiv 0$) MGM.*
Panel (b): Scaling of the p-value of the Pearson correlation coefficient displayed in panel (a). The same legend hold.
The complete MGM seems to significantly explain the neutralization power measurements while the factorized model gives correlation coefficients lesser, that drops for increasing values of pseudocount.



(a) (b)

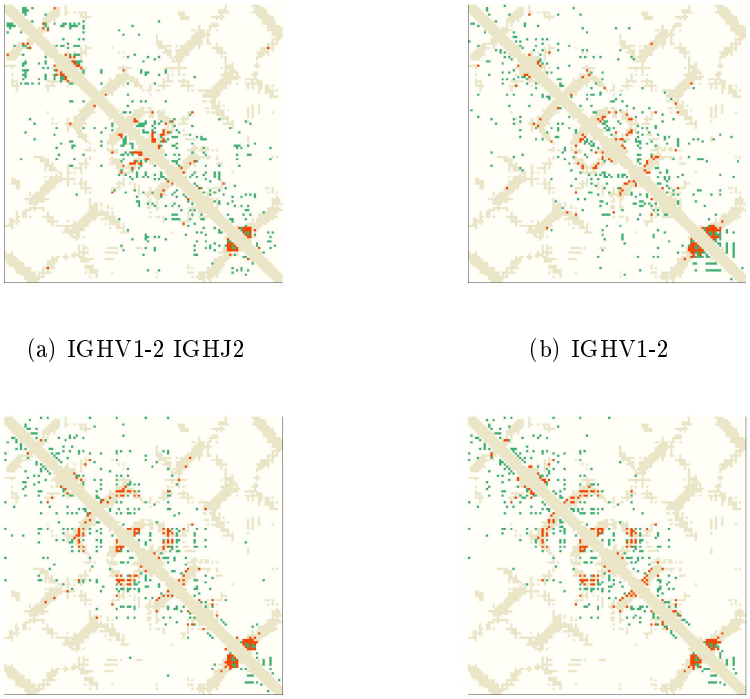
Figure 6.7: Left panel: Pearson correlation coefficient between the energy of the inferred model and the measures of neutralization power, for different data sets:
1) cluster PG04*; 2)*cluster VJ *3)All seqs with inferred IGHV1-2 germline gene;4) Seqs with any IGHV and IGHJ inferred germiline genes.*
In red are shown the average measure and in blue the minimum.
Right panel: Logarithm of the p-value for the same cases.



(a) $\pi = 0.8$ (b) $\pi = 0.5$

(c) $\pi = 0.2$ (d) $\pi = 0.01$

Figure 6.8: Direct Information map computed on the clusterPG04.
The internal contact map of VRC-PG04 heavy chain is shown in grey. Two residues are considered to be in contact if at least a couple of atoms is at distance lower than 8 Å. The first 300 couples with higher DI are displayed in red when they superpose to the internal contacts (true positive internal contact predictions) and in green when they do not (false positive internal contact predictions). Residues with a distance along the a.a. sequence greater than 4 are not considered because they provide trivial contacts [76]. The different maps correspond to different values of the pseudocount parameter, pc: a) $\pi = 0.8$; b) $\pi = 0.5$; c) $\pi = 0.2$; d) $\pi = 0.01$.
In every plot the DI is computed by learning the Gaussian model on the sequences belonging to clusterPG04.



(a) IGHV1-2 IGHJ2 (b) IGHV1-2

(c) all germline (d) all germline reweighted

Figure 6.9: Direct information map computed on different set of sequences. The explanation of the maps and the color legend is the same as in Figure 6.8. Different maps correspond to different sets of sequences. a) Seqs with inferred IGHV1-2 and IGHJ2 germline genes; b) Seqs with inferred IGHV1-2 and any IGHJ inferred germline genes c) Seqs with any IGHV and IGHJ inferred germiline genes; d) Seqs with any IGHV and IGHJ inferred germiline genes that are the result of a reweighting procedure with $\theta = 0.01$ *(see [76]). Every map has been computed with constant pseudocount parameter* π*=0.5.*

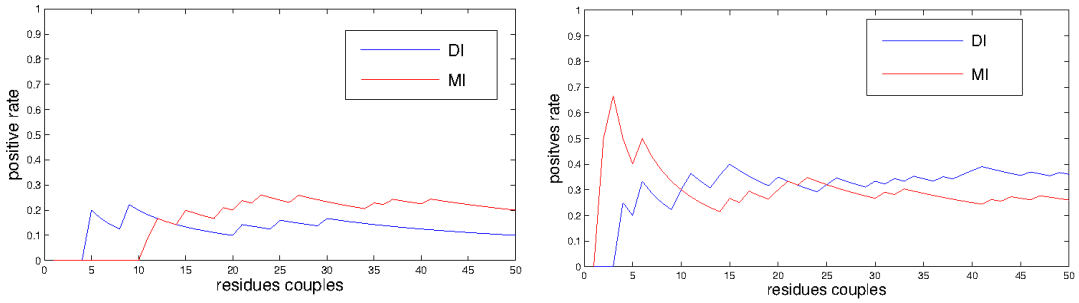


Figure 6.10: Sensitivity-specificity curves: DI-ranking-based in blue *and MI-ranking-based curves in* red. *Specificity is defined as in Section 6.1. Only couples that are at distance larger than 4 amino acid are considered.*

6.3.3 The antigen - heavy chain interaction

Since in the crystallographic structure the broadly neutralizing antibody VRC-PG04 is resolved in complex with the gp120 we can ask if it is possible to predict some feature of the Ab-antigen binding by analysing the population of sequences present in the sample.

The simplest analysis that can be performed is on the variability of the residues in the columns of the alignments. A measure of it are the entropies of the empirical distribution of residues in each column in the alignment. Then, we focus on the residues that the resolved structure say to be in contact with gp120. The site entropies computed on *clusterPG04* have to be compared with the backround ones that encodes the average variability of antibodies maturated from the same germline genes. As a first try, we approximated this background with the *clusterVJ*, see figure 6.11. As an alternative procedure, a library of the deposited antibodies sequences with inferred IGHV1-2 allelic origin has been constructed and assumed to represent the reference background of the IGHV part of the sequences. Both procedures give the convergent information that the amino acids in contact with the antigen have very different behavior with respect to the variability. Indeed some of the contact residues happens to be less variable in the mutated cluster than in the background. This is the case for Arg71, an amino acid which is observed to form a crucial interaction (salt bridge) with an Asp in the binding region of gp120. On the contrary, other contact residues (most of which are in the CDR2 region) are observed to be more variable in the mutated cluster than in the background.

The heterogeneity of the nature of the contacts between the antibody and the antigen does not seems to permit to individuate the contact residues by simply analysing the single site entropies of the populations of antibodies.

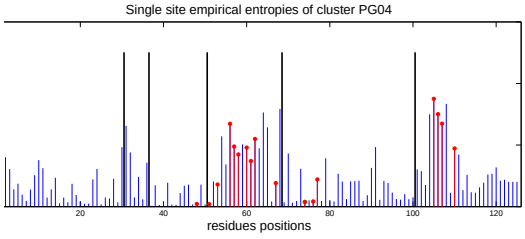
The structure of the correlations between different columns in the MSA is another interesting feature of the populations that can be related to the details of the binding between the antibody and the antigen.

In fact, sequences belonging to *clusterPG04* are peculiar for the emergence of a strong interaction pattern among a set of sites mostly belonging to the CDR2 region. The high DI signal in the CDR2 is more evident for high pseudocounts (see Figure 6.12).

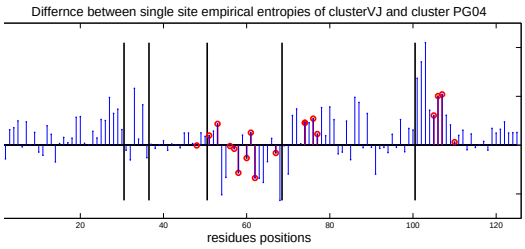
An explanation for this observation could relay in the fact that, in the tridimensional structure, the CDR2 loop appears to be shifted with respect to its typical position in order to permit the exposition of $\mathrm{Arg71}_{\mathrm{VRC\text{-}PG04}}$ that can so from the critical interaction $\mathrm{Arg71}_{\mathrm{VRC\text{-}PG04}}$ - $\mathrm{Asp368}_{\mathrm{gp120}}$. So while the presence of Arg in position 71 is mandatory to establish the interaction, as it can form a salt bridge with Asp, the residues in CDR2 do not feel a strong constraint and are more free to mutate, provided that they evolve in a correlated fashion preserving the exotic position of the loop.

This argument seems to be confirmed by comparing the DI in the two clusters. The high DI pattern of the CDR2 is characteristic of the highly mutated *clusterPG04* and is not present in *clusterVJ* that is more similar to the germline genes, as shown in figure 6.13.

The above explanation of the DI structure is still under investigation and needs to be confirmed by bioinformatic analysis. Anyway this is probably one of the structural characteristics of the population that are recovered by the MGM method in order to generate the results exposed in Section 6.3.1.

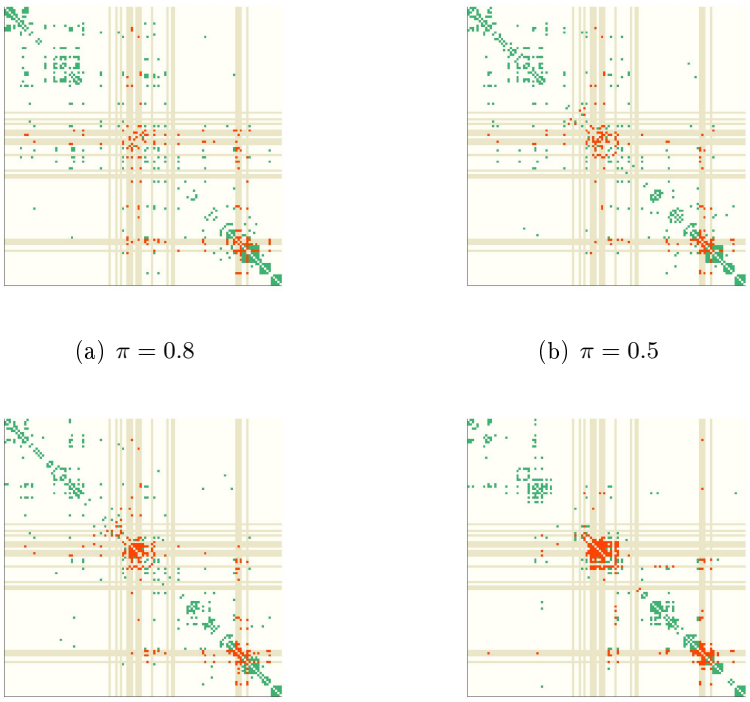


(a)



(b)

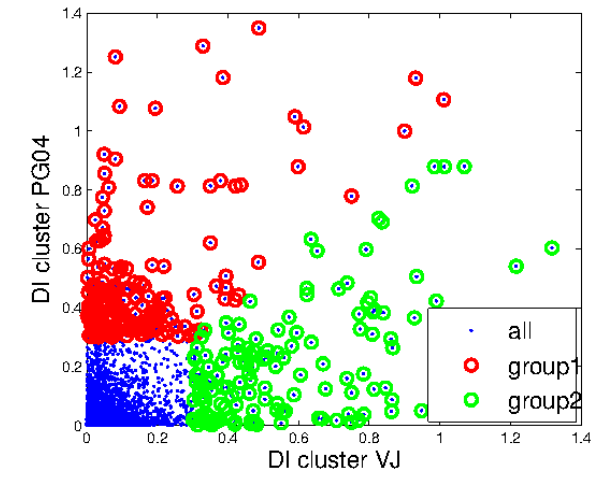
Figure 6.11: Upper panel: single site empirical entropies for clusterPG04*, where are highlighted the residues of VRC-PG04 that are in contact with the antigen.*
Lower panel : difference of the empirical single site entropies computed on clusterPG04 *and the background, the germline cluster* clusterVJ*.*
Vertical lines separate frameworks and complementary determining regions, following: FWR1, CDR1, FWR2, CDR2, FWR3, CDR3.

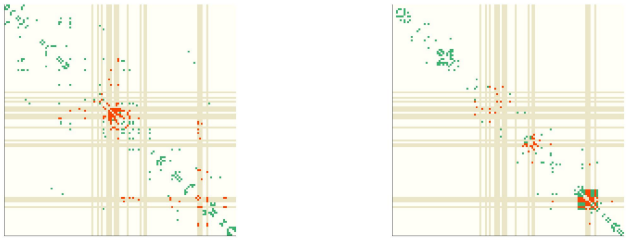


(a) $\pi = 0.8$ (b) $\pi = 0.5$

(c) $\pi = 0.2$ (d) $\pi = 0.01$

Figure 6.12: Direct Information map computed on the clusterPG04.
Third column: Grey lines highlight columns and raws referring to residues of the VRC-PG04 heavy chain than are in contact with the antigen (gp120). Dots represents the first 300 couples with higher DI and they are colored in red if at least one of the amino acid of the couple is in contact with the antigen and in green otherwise.
Different maps correspond to different values of the pseudocount parameter, pc. a) $\pi = 0.8$*; b)* $\pi = 0.5$*; c)* $\pi = 0.2$*; d)* $\pi = 0.01$.
In every plot the DI is computed by learning the Gaussian model on the sequences belonging to clusterPG04.





(b) group 1 (c) group 2

Figure 6.13: Comparison of the DI computed on clusterPG04 *and* clusterVJ *(see main text for details).*
Upper panel: In the scatter plot the DI computed on clusterPG04 *(vertical axis) is compared with the DI computed on* clusterVJ *(orizontal axis). Circled couples correspond to pair of residues for which the DI>2.5 when computed on both clusters. Pairs for which the DI is higher when computed on cluster PG04 are circled in green (group 1), while those for which the opposite happens are circled in green (group 2).*
Left panel: Pairs belonging to group 1 are compared with the contacts with the antigen as explained in Figure 6.8.
Right panel: Same as in left panel but for group 2.

6.4 Conclusions

In this part of the thesis, we have reported the preliminary results on an analysis performed on deep sequencing data of a repertoire of antibodies through Multivariate Gaussian modeling. The used data set is very promising for the study of the affinity maturation process as it presents immunoglobulins with many mutations with respect to the germline genes due to the strong evolutionary pressure provided by the presence of HIV-1 gp120 as an antigen. This cluster of antibodies (*clusterPG04*) contains highly mutated antibodies that underwent the affinity maturation process.

The experimental work provides, in addition to the sequencing data, the measures of neutralization power of a set of chimeric antibodies construct from the high mutated cluster and the crystallography study of the complex antigen with VRC-PG04 Ab, which is a broadly neutralizing Ab similar in sequence to this cluster. These data permit to design some strategies to test if the inferred model can be used as a proxy to study the affinity function on the sequences space and to extract information on structure of the antibodies and its interaction with the antigen.

Up to now, the central result of the analysis is that, having learned a MGM on the heavy chains of this group of highly mutated Abs, the inferred energy of the measured Abs correlates with the neutralization titers. The result needs to be deepened but, if confirmed, it would allow to use Rep-Seq data to study the structure of the neutralization power as a function of the sequence.

As shown in [89], gp120 directed affinity maturation can walk through different directions in the space of sequences. In fact, bnAbs with the same inferred germline and convergent tridimensional structure but found in different donors display highly divergent sequences ($\sim 50\%$). This fact suggests that the neutralization power in the space of sequence can be imagined as a multivalley landscape and that a single valley is explored by different patients. The perspective of an integration of multiple Rep-Seq experiments performed on different donors seems an interesting step forward for the understanding of the neutralization power optimization.

On the structural level, Gaussian DCA analysis on Rep-Seq data does not seem to produce good results for the prediction of internal native contacts. This suggests that the evolutionary process doesn't have the possibility to explore enough the sequence space (due probably to temporal constraint) to shows variations on the backbone of the structures, condition needed for the success of the DCA approach.

Nevertheless, the inferred model seems to contain information on the Abs-antigen interaction. Despite the fact that a naïve analysis on the variability of the different positions of the residues in the MSA don't seem to represent a satisfactory tool for predicting antigen-heavy chain contacts, functional high DI patterns appear and can be related to more complicated structural features that can be involved in the optimization of the antigen binding. Considering this, instead of finding a unique mark for the detection of residues in contacts, ongoing research attempts to develop different tools to account the heterogeneous nature of the acquired mutations (key strong bonds, modifications to increase structural stability, etc.).

Appendices

Appendix A

part I

A.1 Detailed computation of the t streaming of thermodynamic pressure $\tilde{A}$

In this section we report the detailed computation necessary to get the expression of the t streaming of the thermodynamic pressure $\tilde{A}(\alpha, \beta, t)$ report in Sec. 2.2.2 in Eq. (A.5). When evaluating the streaming $\partial_t \tilde{A}$, we get the sum of four terms $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$; each comes as a consequence of the derivation of a corresponding exponential term appearing into Equation (2.31). In order to proceed we need to compute them explicitly:

$$\begin{aligned}
\mathcal{A} &= \frac{1}{N}\mathbb{E}\frac{\sqrt{\beta}}{2\sqrt{t}N^{(1-\theta)/2}}\sum_{i,\mu}^{l_\eta,l_\chi}\xi_i^\mu\omega(\sigma_i,z_\mu) = \frac{1}{N}\mathbb{E}\frac{\sqrt{\beta}}{2\sqrt{t}N^{(1-\theta)/2}}\sum_{i,\mu}^{l_\eta,l_\chi}\partial_{\xi_i^\mu}\omega(\sigma_i,z_\mu) \\
&= \frac{1}{N}\mathbb{E}\frac{\beta}{2N^{(1-\theta)}}\sum_{i,\mu}^{l_\eta,l_\chi}\left[\omega(\sigma_i^2,z_\mu^2)-\omega^2(\sigma_i^2,z_\mu^2)\right] = \\
&= \frac{1}{N}\frac{\beta}{2N^{(1-\theta)}}\sum_{l_\eta,l_\chi}P(l_\eta,l_\chi)l_\eta l_\chi\left[\langle z_\mu^2\rangle_G - \langle q_{12}^{l_\eta}p_{12}^{l_\chi}\rangle_G\right] = \\
&= \frac{1}{N}\frac{\beta}{2N^{(1-\theta)}}NL\left(\frac{1+a}{2}\right)\left[\langle z_\mu^2\rangle - \langle q_{12}^{l_\eta}p_{12}^{l_\chi}\rangle\right] = \frac{\alpha\beta}{2}\frac{\gamma}{2}\left[\langle z_\mu^2\rangle - \langle q_{12}^{l_\eta}p_{12}^{l_\chi}\rangle\right] ,
\end{aligned} \tag{A.1}$$

where in the first passage we used integration by parts and, in the fourth, the factorization properties of the quenched averages [30, 92, 41, 93, 94] (which should be understood in the thermodynamic limit).

The same procedure can be used in the computation of the other terms, so to get:

$$\begin{aligned}
\mathcal{B} &= -\frac{1}{N}\mathbb{E}\frac{b}{2\sqrt{1-t}}\sum_i^{l_\eta}\eta_i\omega(\sigma_i) = -\frac{1}{N}\mathbb{E}\frac{b}{2\sqrt{1-t}}\sum_i^{l_\eta}\partial_{\eta_i}\omega(\sigma_i) \\
&= -\frac{1}{N}\mathbb{E}\frac{b^2}{2}\left(l_\eta - \sum_i^{l_\eta}\omega^2(\sigma_i)\right) = -\frac{b^2}{2N}\sum_{l_\eta}P(l_\eta)\,l_\eta\left(1-\langle q_{12}^{l_\eta}\rangle_G\right) \\
&= -\frac{b^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\left(1-\langle q_{12}^{l_\eta}\rangle\right) - \frac{b^2}{2}\sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}\left(1-\langle q_{12}^{l_\eta}\rangle\right) ;
\end{aligned} \tag{A.2}$$

$$\begin{aligned}\mathcal{C} &= -\frac{1}{N}\mathbb{E}\frac{c}{2\sqrt{1-t}}\sum_{\mu}^{l_\chi}\chi_\mu\omega(z_\mu) = -\frac{1}{N}\mathbb{E}\frac{c}{2\sqrt{1-t}}\sum_{\mu}^{l_\chi}\partial_{\chi_\mu}\omega(z_\mu) = \\ &= -\frac{1}{N}\mathbb{E}\frac{c^2}{2}\left(\sum_{\mu}^{l_\chi}\omega(z_\mu^2) - \sum_{\mu}^{l_\chi}\omega^2(z_\mu)\right) = -\frac{c^2}{2N}\sum_{l_\chi}P(l_\chi)\,l_\chi\left(\sum_{\mu}^{l_\chi}\langle z_\mu^2\rangle_G - \langle p_{12}^{l_\chi}\rangle_G\right) \\ &= -\frac{\alpha c^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\left(\langle z^2\rangle - \langle p_{12}^{l_\chi}\rangle\right) = -\frac{\alpha c^2}{2}\sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}\left(\langle z^2\rangle - \langle p_{12}^{l_\chi}\rangle\right)\ ; \end{aligned} \tag{A.3}$$

$$\mathcal{D} = -\frac{1}{N}\mathbb{E}\frac{d}{2}\sum_{\mu}^{l_\chi}\omega(z_\mu^2) = -\frac{\alpha d}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\langle z^2\rangle\,. \tag{A.4}$$

Now, the t-streaming of the pressure reads off as

$$\begin{aligned}\frac{d\tilde{A}_N(\alpha,\beta,t)}{dt} &= \left[\frac{\alpha\beta}{2}N^\theta\left(\frac{1+a}{2}\right) - \frac{\alpha c^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}} - \frac{\alpha d}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\right]\langle z^2\rangle + \\ &\quad -\frac{\alpha\beta}{2}N^\theta\left(\frac{1+a}{2}\right)\langle q_{12}^{l_\eta}p_{12}^{l_\chi}\rangle + \frac{b^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\langle q_{12}^{l_\eta}\rangle + \frac{c^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\langle p_{12}^{l_\chi}\rangle + \\ &\quad -\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\frac{b^2}{2}\,. \end{aligned} \tag{A.5}$$

A.2 Detailed computation of the t streaming for a generic observable O_s

In this section we report the detailed computation of the results, showed in Eq. (2.48) (section 2.3) of the t streaming of a generic observable O_s of s replicas.

$$\langle O\rangle_t = \frac{\int\prod_\mu^L d\mu(z_\mu)\sum_\sigma O\exp\left(-\beta H_s\right)}{\int\prod_\mu^L d\mu(z_\mu)\sum_\sigma\exp\left(-\beta H_s\right)}, \tag{A.6}$$

with

$$\begin{aligned}H_s &= \sum_{a=1}^{s}\left\{\sqrt{t}\sqrt{\frac{\beta}{N^{1-\theta}}}\sum_{i,\mu}^{l_\eta,l_\chi}\xi_i^\mu\sigma_i^a z_\mu^a + \sqrt{1-t}\;b\sum_i^{l_\eta}\eta_i\sigma_i^a + \right. \\ &\quad \left. +\sqrt{1-t}\;c\sum_\mu^{l_\chi}\chi_\mu z_\mu^a + (1-t)\frac{d}{2}\theta_\mu\sum_\mu^{l_\chi}(z_\mu^a)^2\right\} \end{aligned} \tag{A.7}$$

The t-streaming of $\langle O\rangle_t$ is then

$$\frac{d\langle O_s\rangle_t}{dt} = \frac{d}{dt}\mathbb{E}\frac{\sum_{\{\sigma\}}\int\prod_\mu d\mu(z_\mu)O_s e^{H_s}}{\sum_{\{\sigma\}}\int\prod_\mu d\mu(z_\mu)e^{H_s}} = \mathbb{E}\left[\Omega\left(O_s\frac{dH}{dt}\right)\right] - \mathbb{E}\left[\Omega(O_s)\Omega\left(\frac{dH}{dt}\right)\right]\,. \tag{A.8}$$

In the last equation eight terms contribute. Let us call them A_1, B_1, C_1, D_1, A_2, B_2, C_2, D_2 and compute them explicitly:

$$\begin{aligned}
A_1 &= \frac{1}{2}\sqrt{\frac{\beta}{tN^{1-\theta}}}\mathbb{E}\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)O_s\sum_a\sum_{i,\mu}^{l_\eta,l_\chi}\xi_i^\mu\sigma_i^a z_\mu^a\frac{e^{H_s}}{Z_s} = \\
&= \frac{1}{2}\sqrt{\frac{\beta}{tN^{1-\theta}}}\mathbb{E}\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)O_s\sum_a\sum_{i,\mu}^{l_\eta,l_\chi}\sigma_i^a z_\mu^a\partial_{\xi_i^\mu}\frac{e^{H_s}}{Z_s} = \\
&= \frac{\beta}{2N^{1-\theta}}\mathbb{E}\left[\sum_{a,b}^{s}\sum_{i,\mu}^{l_\eta,l_\chi}\Omega(O_s\sigma_i^a z_\mu^a\sigma_i^b z_\mu^b)-\sum_{a,b}^{s}\sum_{i,\mu}^{l_\eta,l_\chi}\Omega(O_s\sigma_i^a z_\mu^a)\Omega(\sigma_i^b z_\mu^b)\right] = \\
&= \frac{\beta}{2N^{1-\theta}}\mathbb{E}\left[\sum_a l_\eta\sum_\mu^{l_\chi}\Omega(O_s(z^a)^2)+\sum_{a\neq b}^{s}l_\eta l_\chi\Omega(O_s q_{ab}^{l_\eta}p_{ab}^{l_\chi})-s\,l_\eta l_\chi\sum_a\Omega(O_s q_{s+1,a}^{l_\eta}p_{s+1,a}^{l_\chi})\right] = \\
&= \frac{\beta LN^\theta}{2}\left(\frac{1+a}{2}\right)\left[\sum_a\langle O_s(z^a)^2\rangle+\sum_{a\neq b}^{s}\langle O_s q_{ab}^{l_\eta}p_{ab}^{l_\chi}\rangle-s\sum_a\langle O_s q_{s+1,a}^{l_\eta}p_{s+1,a}^{l_\chi}\rangle\right]. \qquad \text{(A.9)}
\end{aligned}$$

$$\begin{aligned}
A_2 &= -\frac{1}{2}\sqrt{\frac{\beta}{tN^{1-\theta}}}\mathbb{E}\Omega O_s\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)\sum_a\sum_{i,\mu}^{l_\eta,l_\chi}\xi_i^\mu\sigma_i z_\mu\frac{e^{H_s}}{Z_s} = \\
&= -\frac{1}{2}\sqrt{\frac{\beta}{tN^{1-\theta}}}\mathbb{E}\Omega O_s\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)\sum_a\sum_{i,\mu}^{l_\eta,l_\chi}\sigma_i z_\mu\partial_{\xi_i^\mu}\frac{e^{H_s}}{Z_s} = \\
&= -\frac{\beta}{2N^{1-\theta}}\mathbb{E}\Omega O_s\left[\sum_{a,b}^{s}\sum_{i,\mu}^{l_\eta,l_\chi}\Omega(\sigma_i^a z_\mu^a\sigma_i^b z_\mu^b)-\sum_{a,b}^{s}\sum_{i,\mu}^{l_\eta,l_\chi}\Omega(\sigma_i^a z_\mu^a)\Omega(\sigma_i^b z_\mu^b)\right] = \\
&= -\frac{\beta}{2N^{1-\theta}}\mathbb{E}\left[\sum_{a<b}\sum_\mu^{l_\chi}\Omega(O_s\sigma_i^{s+1}z_\mu^{s+1}\sigma_i^{s+2}z_\mu^{s+2})+\sum_a^s\Omega(O_s)\Omega((z_\mu^a)^2)+\right. \\
&\quad \left.+\quad s^2\Omega(O_s\sigma_i^{s+1}z_\mu^{s+1}\sigma_i^{s+2}z_\mu^{s+2})\right] = \\
&= -\frac{\beta}{2N^{1-\theta}}\mathbb{E}\left[\frac{s(s-1)}{2}\,l_\eta l_\chi\Omega(O_s q_{s+1,s+2}^{l_\eta}p_{s+1,s+2}^{l_\chi})+sl_\eta l_\chi\Omega(O_s(z_\mu^{s+1})^2)+\right. \\
&\quad \left.-s^2\Omega(O_s q_{s+1,s+2}^{l_\eta}p_{s+1,s+2}^{l_\chi})\right] = \\
&= -\frac{\beta LN^\theta}{2}\left(\frac{1+a}{2}\right)\left[s\langle O_s q_{s+1,s+2}^{l_\eta}p_{s+1,s+2}^{l_\chi}\rangle-s\langle O_s(z^{s+1})^2\rangle\right]. \qquad \text{(A.10)}
\end{aligned}$$

And with analogous calculations

$$B_1 = -\frac{b^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}N\left[\sum_{a<b}\langle O_s q_{ab}^{l_\eta}\rangle-s\langle O_s q_{s+1a}^{l_\eta}\rangle+s\langle O_s\rangle\right], \qquad \text{(A.11)}$$

$$B_2 = \frac{b^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}N\left[-s\langle O_s q_{s+1,s+2}^{l_\eta}\rangle+s\langle O_s\rangle\right], \qquad \text{(A.12)}$$

$$C_1 = -\frac{c^2}{2}\alpha\left(\frac{1+a}{2}\right)^{\frac{1}{2}} N\left[\sum_a \langle O_s((z^a)^2)\rangle + \sum_{a<b}\langle O_s p^{l_\chi}_{a,b}\rangle - s\sum_a \langle O_s p^{l_\chi}_{s+1,a}\rangle\right], \quad (A.13)$$

$$C_2 = -\frac{c^2}{2}\alpha\left(\frac{1+a}{2}\right)^{\frac{1}{2}} N\left[s\langle O_s(z^{s+1})^2\rangle - \frac{s(s+1)}{2}\langle O_s p^{l_\chi}_{s+1,s+2}\rangle\right], \quad (A.14)$$

$$D_1 = -\frac{d\alpha N}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\sum_a \langle O_s(z^a)^2\rangle, \quad (A.15)$$

$$D_2 = \frac{d\alpha N}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}} s\langle O_s(z^{s+1})^2\rangle. \quad (A.16)$$

Therefore, merging all these terms together, the streaming is

$$\frac{d}{dt}\langle O_s\rangle_t = \beta\sqrt{\alpha}\frac{\gamma}{2}\left(\sum_{a<b}^{s}\langle OQ^{l_\eta}_{ab}P^{l_\chi}_{ab}\rangle_t - s\sum_a^s \langle OQ^{l_\eta}_{a,s+1}P^{l_\chi}_{a,s+1}\rangle_t + \right.$$
$$\left. + \frac{s(s+1)}{2}\langle OQ^{l_\eta}_{s+1,s+2}P^{l_\chi}_{s+1,s+2}\rangle_t\right). \quad (A.17)$$

Appendix B

part II

B.1 Errors on the fitting parameters

We show in this section how to calculate the errors on the fitting parameters using standard methods for non-linear fitting procedure. For simplicity we consider the positive cooperative case, without loosing generality. To obtain the errors on the fitting parameters we have to compute the Hessian matrix $\hat{H}$

$$\frac{\partial^2 S}{\partial J^2}, \qquad \frac{\partial^2 S}{\partial h_0^2}, \qquad \frac{\partial^2 S}{\partial J \partial h_0} \tag{B.1}$$

Consider the first element of the matrix.

$$\frac{\partial S}{\partial J} = \sum_k \frac{\partial S}{\partial \theta_k}\frac{\partial \theta_k}{\partial J}; \qquad \frac{\partial^2 S}{\partial J^2} = \sum_k \frac{\partial^2 S}{\partial \theta_k \partial J}\frac{\partial \theta_k}{\partial J} + \frac{\partial S}{\partial \theta_k}\frac{\partial^2 \theta_k}{\partial J^2} = \sum_{k,j} \frac{\partial^2 S}{\partial \theta_k \partial \theta_j}\frac{\partial \theta_k}{\partial J}\frac{\partial \theta_j}{\partial J}$$

where in the last passage we have neglected the term $\frac{\partial^2 \theta_k}{\partial J^2}$. The first term is trivial to compute $\frac{\partial^2 S}{\partial \theta_k \partial \theta_j} = 2\frac{N}{N-2}\delta_{k,j} \simeq 2\delta_{k,j}$. Differentiating the self-consistency equation we get the second term $\frac{\partial \theta_k}{\partial J}$:

$$\frac{\partial m}{\partial J} = (1-m^2)\left(J\frac{\partial m}{\partial J} + m\right) \Rightarrow \frac{\partial m}{\partial J} = \frac{(1-m^2)m}{1-J(1-m^2)}$$

Recall that $\theta = 1/2 + m/2$, we get:

$$\frac{\partial^2 S}{\partial J^2} = \sum_k \left(\frac{(1-m^2)}{1-J(1-m^2)}\right)^2$$

The other elements of the Hessian matrix $\partial^2 S/\partial h_0^2$, $\partial^2 S/\partial J \partial h_0$, are computed in the same way, considering that

$$\frac{\partial m}{\partial h_0} = \frac{(1-m^2)}{1-(J(1-m^2)}$$

Finally the errors on the parameters are obtained by the product of the squared residuals and the inverse of the Hessian matrix $\mathbf{e} = S\text{Diag}\{(\hat{H})^{-1}\}$.

The Hill number is obtained by the derivatives of the function in the half-value of θ in α^*. Deriving the hill function, namely $\theta_H = 1/(1+(\alpha/\alpha_0)^{-n_H})$, we gets:

$$\frac{\partial \theta_H}{\partial \alpha} \mid_{\alpha^*} = \frac{n_h}{4\alpha/\alpha_0}\frac{(\alpha/\alpha_0)^{-n_H}}{(1+(\alpha/\alpha_0)^{-n_H))^2} \mid_{\alpha^*} = \frac{n_h}{4} \tag{B.2}$$

where we use that $\theta(\alpha^*) = 1/2$, $\alpha^*/\alpha_0^* = 1$. Recalling that $h = (1/2)\log(\alpha)$, the derivatives in the half-curve value in our model is:

$$\frac{\partial\theta}{\partial\alpha} = \frac{1}{4\alpha/\alpha_0}\frac{1-(2\theta-1)^2}{1-(1-(2\theta-1)^2)J} = \frac{1}{4(1-J)} \tag{B.3}$$

Recall that $n_H = 1/(1-J)$, the error on the Hill number is $\Delta n_H = \Delta J/(1-J)^2$. Note that approaching the critical point $J \to 1^-$ the error on the Hill number diverges.

B.2 Calculations of the self-consistencies in the heterogeneous case

In this appendix we outline in details the calculations required to obtain explicit expressions for $\langle m_A\rangle$ and $\langle m_B\rangle$, in the heterogeneous extension of the model, described in section 4.4.

As explained in the main text, we consider a bipartite system made of two subsets A and B whose sizes are N_A and N_B elements, respectively, with $N_A + N_B = N$. The spin of subset interact we the spin of the other through this coupling :

$$J_{ij} = -\frac{J_0}{P}\sum_{\mu=1}^{P}(\xi_i^\mu\xi_j^\mu + \bar{\xi}_i^\mu\bar{\xi}_j^\mu), \tag{B.4}$$

we denote with ξ_i^μ the μ^{th} bit of the i^{th} string, which can assume the values 0, 1 according to the probability distribution

$$P(\xi_i^\mu) = \frac{1+a}{2}\delta_{\xi_i^\mu,1} + \frac{1-a}{2}\delta_{\xi_i^\mu,0}, \tag{B.5}$$

where a is a parameter ranging in the interval $[-1,1]$ and we denote with $\bar{\xi}_i^\mu = 1-\xi_i^\mu$ the complement of a bit. The more the two strings ξ_i and ξ_j are similar, the more the spins σ_i and σ_j tend to "disagree", viceversa, the smaller the string overlap, the weaker the tendency of the spins to interact; this corresponds to a random anti-ferromagnetic interaction.

The Hamiltonian which defines the model is then the sum of the energies due to the interaction between spins and with the external field:

$$H(\sigma;\xi,h) = \frac{J_0}{NP}\sum_{i\in A,\, j\in B}\sum_{\mu=1}^{P}(\xi_i^\mu\xi_j^\mu + \bar{\xi}_i^\mu\bar{\xi}_j^\mu)\sigma_i\sigma_j - h\sum_{i=1}^{N}\sigma_i. \tag{B.6}$$

We now introduce the pattern overlap $q_\mu^A(\sigma^A)$ and the magnetization $m_A(\sigma^A)$ for the subsystem A as

$$q_\mu^A(\sigma^A) = \frac{1}{N_A}\sum_{i\in A}\xi_i^\mu\sigma_i, \quad m_A(\sigma^A) = \frac{1}{N_A}\sum_{i\in A}\sigma_i, \tag{B.7}$$

which respectively measure the resemblance between a microscopic state σ^A and one particular pattern $\xi_{i=1\ldots N}^\mu$ and the net number of active elements. Note that the quantity $q_\mu^A(\sigma^A)$ can be also interpreted as the magnetization of a subgraph of the subsystem A specified by the non-null bits of the string ξ_i^μ. We similarly define the quantities $q_\mu^B(\sigma^B)$

and $m_B(\sigma^B)$ for the subsystem B. Then we may rewrite the Hamiltonian as

$$H(\xi;\sigma,h) = J_0\frac{N_A N_B}{N}\left[\frac{2}{P}\sum_{\mu=1}^{P} q_\mu^A(\sigma^A)q_\mu^B(\sigma^B) + m_A(\sigma^A)m_B(\sigma^B) - \frac{1}{P}\sum_{\mu=1}^{P} q_\mu^A(\sigma^A)m_B(\sigma^B) - \frac{1}{P}\sum_{\mu=1}^{P} q_\mu^B(\sigma^B)m_A(\sigma^A)\right] - h\left(N_A m_A(\sigma^A) + N_B m_B(\sigma^B)\right). \tag{B.8}$$

The first term, in square brackets, is clearly large when the magnetizations (and pattern overlaps) of the two systems have the same sign, while the interaction with the external field tends to align all the spins with the sign of h. Due to this competition, we expect that the system goes trough different phases by varying the ratio J/h: when the external field is small, the leading contribution to the free energy comes from the coupling term J, so in this case the phase of system is anti-ferromagnetic, with two different values for the magnetizations of the two subsystems. On the other hand, when the external field is sufficiently strong or J small enough, we expect the two magnetizations being equal so that the phase is paramagnetic (in a field), and there is a phase transition separating the two.

In the zero noise case, the entropic term is vanishing and we may easily evaluate the critical field $h_c(J)$ for which the transition occurs. Assuming without loss of generality $N_B < N_A$ and $h > 0$, in the anti-ferromagnetic phase (AF) we have all the spins of the subsystem A aligned with the field and all the spins belonging to subsystem B oriented in the opposite direction, while in the paramagnetic phase (P) both the subsystems are aligned with the field:

$$H_{AF} = -\frac{J_0}{NP}\sum_{i\in A}\sum_{i\in B}\sum_{\mu}\left(\xi_i^\mu\xi_j^\mu + \bar{\xi}_i^\mu\bar{\xi}_j^\mu\right) - h(N_A - N_B) \tag{B.9}$$

$$H_P = +\frac{J_0}{NP}\sum_{i\in A}\sum_{i\in B}\sum_{\mu}\left(\xi_i^\mu\xi_j^\mu + \bar{\xi}_i^\mu\bar{\xi}_j^\mu\right) - h(N_A + N_B). \tag{B.10}$$

Equating the two and averaging over the disorder, one obtains the critical field at zero noise level

$$h_c(J_0,a) = J_0\frac{N_A}{N}\frac{1+a^2}{2}. \tag{B.11}$$

To describe the system as a whole, we introduce the global pattern overlap q_μ and the global magnetization m:

$$q_\mu(\sigma) = \frac{N_A}{N}q_\mu^A(\sigma^A) + \frac{N_B}{N}q_\mu^B(\sigma^B), \quad m(\sigma) = \frac{N_A}{N}m_A(\sigma^A) + \frac{N_B}{N}m_B(\sigma^B). \tag{B.12}$$

Moreover, we need some quantities which are able to describe an antiferromagnetic phase, so we introduce the staggered overlap p_μ and staggered magnetization n:

$$p_\mu(\sigma) = \frac{N_A}{N}q_\mu^A(\sigma^A) - \frac{N_B}{N}q_\mu^B(\sigma^B), \quad n(\sigma) = \frac{N_A}{N}m_A(\sigma^A) - \frac{N_B}{N}m_B(\sigma^B). \tag{B.13}$$

In terms of these quantities, the Hamiltonian is

$$H(\sigma;\xi,h) = NJ_0\left[\frac{1}{2P}\sum_{\mu=1}^{P}\left(q_\mu^2(\sigma)-p_\mu^2(\sigma)\right)+\frac{1}{4}(m^2(\sigma)-n^2(\sigma))\right.$$
$$\left.-\frac{1}{2P}\sum_{\mu=1}^{P}q_\mu(\sigma)m(\sigma)-\frac{1}{2P}\sum_{\mu=1}^{P}p_\mu(\sigma)n(\sigma)\right]-Nhm(\sigma). \quad \text{(B.14)}$$

Saddle point equations for the pattern overlaps

The quenched free energy $F(J_0,h)$ and the partition function $Z(\xi;J_0,h)$ are introduced as follows:

$$F(J_0,h) = -\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\log Z(\xi;J_0,h) = -\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\log\sum_{\sigma}e^{-H(\sigma;\xi,h)}. \quad \text{(B.15)}$$

Since the Hamiltonian depends on σ only through the quantities (B.12,B.13) we can choose these as order parameters and define the so-called constrained partition function $Z(\mathbf{q},\mathbf{p},m,n)$ as

$$Z = \int d\mathbf{q}\int d\mathbf{p}\int dm\int dn\; Z(\mathbf{q},\mathbf{p},m,n) \quad \text{(B.16)}$$
$$= \int d\mathbf{q}\int d\mathbf{p}\int dm\int dn\; D(\mathbf{q},\mathbf{p},m,n)\exp\left[-N\mathcal{H}(\vec{q},\vec{p},m,n)\right] \quad \text{(B.17)}$$

where

$$\mathcal{H}(\vec{q},\vec{p},m,n) = J_0\left[\frac{1}{2P}\sum_{\mu=1}^{P}\left(q_\mu^2-p_\mu^2\right)+\frac{1}{4}(m^2-n^2)-\frac{1}{2P}\sum_{\mu=1}^{P}q_\mu m-\frac{1}{2P}\sum_{\mu=1}^{P}p_\mu n\right]-hm. \quad \text{(B.18)}$$

and $D(\vec{q},\vec{p},m,n)$ is the density of states:

$$D(\vec{q},\vec{p},m,n) = \sum_{\sigma}\delta(m-m(\sigma))\delta(n-n(\sigma))\prod_{\mu,\nu=1}^{P}\delta(q_\mu-q_\mu(\sigma))\delta(p_\nu-p_\nu(\sigma)). \quad \text{(B.19)}$$

Using the integral representation for the δ and summing over σ this can be written as

$$D(\vec{q},\vec{p},m,n) = \left(\frac{N}{2\pi}\right)^{2(P+1)}\int_{-\infty}^{+\infty}d\vec{x}\int_{-\infty}^{+\infty}d\vec{y}\int_{-\infty}^{+\infty}dz\int_{-\infty}^{+\infty}dw\exp[Ns(\vec{q},\vec{p},m,n,\vec{x},\vec{y},z,w)] \quad \text{(B.20)}$$

where

$$s(\vec{q},\vec{p},m,n,\vec{x},\vec{y},z,w) = i\Big[\vec{x}\cdot\vec{q}+\vec{y}\cdot\vec{p}+zm+wn+\frac{1}{N}\sum_{i=1}^{N}\log 2\cos\left(\sum_{\mu=1}^{P}\xi_i^\mu\left(x_\mu+\epsilon_i y_\mu+z+\epsilon_i w\right)\right)\Big].$$

Here ϵ_i is a function of the site which takes the value $+1$ if $i\in A$ and -1 for $i\in B$. The constrained entropy $s(\vec{q},\vec{p},m,n)$, i.e., the log-density of states, can be evaluated by saddle point integration of the (B.20), taking the extremum respect to the $2(P+1)$ parameters $\vec{x},\vec{y},z,w$. Then, finding the minimum of the constrained free energy respect to the order parameters $\vec{q},\vec{p},m,n$

$$f(\vec{q},\vec{p},m,n) = \mathcal{H}(\vec{q},\vec{p},m,n)-s(\vec{q},\vec{p},m,n), \quad \text{(B.21)}$$

one obtains the real free energy per spin of the system, in the thermodynamic limit, and the self-consistence equations which rule the behavior of the system for a given realization of the disorder

$$q_\mu = \frac{1}{N}\sum_i \xi_i^\mu \tanh\theta(\vec{q},\vec{p},m,n) \tag{B.22}$$

$$p_\mu = \frac{1}{N}\sum_i \epsilon_i\xi_i^\mu \tanh\theta(\vec{q},\vec{p},m,n) \tag{B.23}$$

$$m = \frac{1}{N}\sum_i \tanh\theta(\vec{q},\vec{p},m,n) \tag{B.24}$$

$$n = \frac{1}{N}\sum_i \epsilon_i \tanh\theta(\vec{q},\vec{p},m,n) \tag{B.25}$$

with

$$\theta(\vec{q},\vec{p},m,n) = -J_0\frac{m-\epsilon_i n}{2}\left(1-\frac{1}{P}\sum_{\nu=1}^{P}\xi_i^\nu\right)+\frac{J_0}{P}\sum_{\nu=1}^{P}\frac{q_\nu-\epsilon_i p_\nu}{2}\left(1-2\xi_i^\nu\right)+h. \tag{B.26}$$

Then it is easy to obtain from these the self-consistence equations for the subsystem parameters,

$$q_\mu^A = \frac{1}{N_A}\sum_{i\in A}\xi_i^\mu \tanh\left(-J_0\frac{N_B}{N}m_B\frac{\sum_\nu\bar{\xi}_i^\nu}{P}+J_0\frac{N_B}{N}\frac{\sum_\nu q_\nu^B\left(1-2\xi_i^\nu\right)}{P}+h\right) \tag{B.27}$$

$$m_A = \frac{1}{N_A}\sum_{i\in A}\tanh\left(-J_0\frac{N_B}{N}m_B\frac{\sum_\nu\bar{\xi}_i^\nu}{P}+J_0\frac{N_B}{N}\frac{\sum_\nu q_\nu^B\left(1-2\xi_i^\nu\right)}{P}+h\right) \tag{B.28}$$

with the corresponding equations for $\vec{q}^B$ and m_B which are easily obtained by exchanging A and B. Looking at equation (B.28), we notice that the magnetization is a sum of terms which may be interpreted as the (thermal) average of the local magnetizations $\omega(\sigma_i)$, so that

$$\omega(\sigma_i) = \tanh\left(-J_0\frac{N_B}{N}m_B\frac{\sum_\nu\bar{\xi}_i^\nu}{P}+J_0\frac{N_B}{N}\frac{\sum_\nu q_\nu^B\left(1-2\xi_i^\nu\right)}{P}+h\right) \tag{B.29}$$

for $i\in A$. The effective field acting on the spin, corresponding to the arguments of the hyperbolic tangent, contain two terms, besides the external field. The first contribution is given by the opposite magnetization of the other subsytem, weighted by a proper factor which is proportional to the fraction of zero bits in the string associated with the spin i, and the second term contains the correlations between the string ξ_i and the strings of the interacting spins, encoded in q_ν.

In the particular cases in which $a=1$ and $a=-1$ the interaction is the same (maximum) for all the couples of spins, and in fact there is no randomness. The only difference is that in the first case the pattern overlaps correspond all to the magnetization, while in the second they all vanish. We may then recover the mean-field equations for a bipartite antiferromagnetic system:

$$m_A = \tanh\left(-J_0\frac{N_B}{N}m_B+h\right) \tag{B.30}$$

$$m_B = \tanh\left(-J_0\frac{N_A}{N}m_A+h\right). \tag{B.31}$$

To obtain the average pattern overlap $\langle q_\mu^A \rangle$ and magnetization $\langle m_A \rangle$ for a generic a we first observe that, being all the bits equivalent, the average pattern overlaps are all equivalent and we may drop the index μ, so that $\langle q_\mu^A \rangle = \langle q^A \rangle$. Then, looking at the arguments of equations (B.27, B.28) - if we approximate q_μ^B and m_B with the averages over the disorder of these quantities - it is clear that they only depend on the bits through the number of non-zero bits k in each string, which has a binomial distribution:

$$\begin{aligned} \langle q^A \rangle &= \frac{1+a}{2} \sum_{k=0}^{P-1} \binom{P-1}{k} \left(\frac{1+a}{2}\right)^k \left(\frac{1-a}{2}\right)^{P-1-k} \\ &\quad \tanh\left(-J_0 \frac{N_B}{N} \frac{P-1-k}{P} \langle m_B \rangle + J_0 \frac{N_B}{N} \frac{P-2(k+1)}{P} \langle q^B \rangle + h\right) \end{aligned} \tag{B.32}$$

$$\begin{aligned} \langle m_A \rangle &= \sum_{k=0}^{P-1} \binom{P}{k} \left(\frac{1+a}{2}\right)^k \left(\frac{1-a}{2}\right)^{P-k} \\ &\quad \tanh\left(-J_0 \frac{N_B}{N} \frac{P-k}{P} \langle m_B \rangle + J_0 \frac{N_B}{N} \frac{P-2k}{P} \langle q^B \rangle + h\right). \end{aligned} \tag{B.33}$$

If P (and $P(1+a)/2$) is large enough, k behaves as a Gaussian random variable, we can get explicitely

$$\langle q^A \rangle = \frac{1+a}{2} \langle m_A \rangle \tag{B.34}$$

$$\langle m_A \rangle = \int_{-\infty}^{+\infty} \frac{dx}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_0)^2}{2\sigma^2}} \tanh\left(-J_0 \frac{N_B}{N} \left(\frac{1-a}{2} + az\right) \langle m_B \rangle + h\right) \tag{B.35}$$

with

$$z_0 = \frac{1+a}{2}, \qquad \sigma^2 = \frac{1-a^2}{2P}. \tag{B.36}$$

Approximating z with its mean value one eventually obtains

$$\langle m_A \rangle = \tanh\left(-J_0 \frac{N_B}{N} \frac{1+a^2}{2} \langle m_B \rangle + h\right), \tag{B.37}$$

and equivalently $\langle m_B \rangle$. These two can then be averaged to get $\theta(\alpha)$ which is the binding isotherm.

Bibliography

[1] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, *et al.*, "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study," *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1232–1237, 2008.

[2] W. Bialek, F. Rieke, R. d. R. Van Steveninck, and D. Warland, "Reading a neural code," *Science*, vol. 252, no. 5014, pp. 1854–1857, 1991.

[3] A. Barra and E. Agliari, "A statistical mechanics approach to autopoietic immune networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2010, no. 07, p. P07004, 2010.

[4] A. De Martino, D. De Martino, R. Mulet, and G. Uguzzoni, "Reaction networks as systems for resource allocation: A variational principle for their non-equilibrium steady states," *PloS one*, vol. 7, no. 7, p. e39849, 2012.

[5] E. Agliari, L. Asti, A. Barra, R. Burioni, and G. Uguzzoni, "Analogue neural networks on correlated random graphs," *Journal of Physics A: Mathematical and Theoretical*, vol. 45, no. 36, p. 365001, 2012.

[6] E. Agliari, A. Barra, R. Burioni, A. Di Biasio, and G. Uguzzoni, "Collective behaviours: from biochemical kinetics to electronic circuits," *Scientific reports*, vol. 3, 2013.

[7] L. Asti, P. Marcatili, A. Pagnani, and G. Uguzzoni, "Multivariate Gaussian Modeling for antibodies affinity maturation," *in preparation*.

[8] G. Parisi, *Statistical field theory*, vol. 28. Perseus Books New York, 1998.

[9] D. Amit, *Modeling brain function: The world of attractor neural networks*. Cambridge Univ. Pr., 1992.

[10] A. Coolen, R. Kühn, and P. Sollich, *Theory of neural information processing systems*. Oxford University Press, USA, 2005.

[11] D. A. Lavis and G. M. Bell, *Statistical Mechanics of Lattice Systems: Volume 2: Exact, Series and Renormalization Group Methods*, vol. 2. Springer, 1999.

[12] F. S. Edwards and P. W. Anderson, "Theory of spin glasses," *J. Phys. F*, vol. 5, no. 5, p. 965, 1975.

[13] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin glass theory and beyond*, vol. 9. World scientific Singapore, 1987.

[14] M. Mézard, G. Parisi, and R. Zecchina, "Analytic and algorithmic solution of random satisfiability problems," *Science*, vol. 297, no. 5582, pp. 812–815, 2002.

[15] A. Barra, E. Mingione, and F. Guerra, "Interpolating the sherrington-kirkpatrick replica trick," *Philosophical Magazine*, vol. 92, pp. 78–97, 2012.

[16] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin glass," *Phys. Rev. Lett.*, vol. 35, no. 26, pp. 1792–1795, 1975.

[17] G. Parisi, "Toward a mean field theory for spin glasses," *Phys. Lett. A*, vol. 73, no. 3, pp. 203–205, 1979.

[18] G. Parisi, "Infinite number of order parameters for spin-glasses," *Phys. Rev. Lett.*, vol. 43, no. 23, pp. 1754–1756, 1979.

[19] G. Parisi, "A sequence of approximated solutions to the s-k model for spin glasses," *J. Phys. A*, vol. 13, no. 4, pp. 115–121, 1980.

[20] F. Guerra, "Broken replica symmetry bounds in the mean field spin glass model," *Commun. Math. Phys.*, vol. 233, no. 1, pp. 1–12, 2003.

[21] M. Talagrand, "The parisi formula," *Annals of Mathematics*, vol. 163, no. 1, pp. 221–263, 2006.

[22] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin-glass," *Physical review letters*, vol. 35, no. 26, pp. 1792–1796, 1975.

[23] J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the national academy of sciences*, vol. 79, no. 8, p. 2554, 1982.

[24] D. J. Amit, H. Gutfreund, and H. Sompolinsky, "Spin-glass models of neural networks," *Physical Review A*, vol. 32, no. 2, p. 1007, 1985.

[25] D. Hebb, *The organization of behavior: A neuropsychological theory*. Lawrence Erlbaum, 2002.

[26] A. Barra, G. Genovese, and F. Guerra, "The replica symmetric approximation of the analogical neural network," *Journal of Statistical Physics*, vol. 140, no. 4, pp. 784–796, 2010.

[27] D. Watts and S. Strogatz *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[28] S. Dorogovtsev and J. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, USA, 2003.

[29] S. Dorogovtsev, A. Goltsev, and J. Mendes, "Critical phenomena in complex networks," *Reviews of Modern Physics*, vol. 80, no. 4, p. 1275, 2008.

[30] E. Agliari and A. Barra, "A hebbian approach to complex-network generation," *EPL (Europhysics Letters)*, vol. 94, p. 10002, 2011.

[31] A. Barra and E. Agliari, "Equilibrium statistical mechanics on correlated random graphs," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, p. P02027, 2011.

[32] E. Agliari, L. Asti, A. Barra, and L. Ferrucci, "Correlation effects in synthetic idiotypic networks," *Physical Review E*, vol. 85, p. 051909, 2011.

[33] A. Barra and F. Guerra, "About the ergodicity in hopfield analogical neural network," *Journal of Mathematical Physics*, vol. 49, p. 125217, 2008.

[34] F. Guerra and F. Toninelli, "The thermodynamic limit in mean field spin glass models," *Communications in Mathematical Physics*, vol. 230, no. 1, pp. 71–79, 2002.

[35] F. Guerra, "Sum rules for the free energy in the mean field spin glass model," *Fields Institute Communications*, vol. 30, p. 161, 2001.

[36] M. Mezard, G. Parisi, and M. Virasoro, *Spin glass theory and beyond*, vol. 9. World scientific Singapore, 1987.

[37] H. Nishimori, *Statistical physics of spin glasses and information processing: an introduction*, vol. 111. Oxford University Press, USA, 2001.

[38] B. Wemmenhove and A. Coolen, "Finite connectivity attractor neural networks," *Journal of Physics A: Mathematical and General*, vol. 36, p. 9617, 2003.

[39] I. Castillo, B. Wemmenhove, J. Hatchett, A. Coolen, N. Skantzos, and T. Nikoletopoulos, "Analytic solution of attractor neural networks on scale-free graphs," *Journal of Physics A: Mathematical and General*, vol. 37, p. 8789, 2004.

[40] E. Agliari, A. Barra, and F. Camboni, "Criticality in diluted ferromagnets," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, p. P10003, 2008.

[41] A. Barra and L. De Sanctis, "Stability properties and probability distributions of multi-overlaps in dilute spin glasses," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, p. P08025, 2007.

[42] F. Guerra and F. Toninelli, "The high temperature region of the viana–bray diluted spin glass model," *Journal of statistical physics*, vol. 115, no. 1, pp. 531–555, 2004.

[43] L. Viana and A. Bray, "Phase diagrams for dilute spin glasses," *Journal of Physics C: Solid State Physics*, vol. 18, p. 3037, 1985.

[44] A. Barra, "Driven transitions at the onset of ergodicity breaking in gauge-invariant complex networks," *International Journal of Modern Physics B*, vol. 24, no. 30, p. 5995, 2010.

[45] K. E. Van Holde, W. Johnson, and P. Shing Ho, *Principles of Physical Biochemistry*. Pearson Prentice Hall, 2006.

[46] T. L. Hill, *Cooperativity theory in biochemistry*. Springer series in molecular biology, 1985.

[47] J. House, *Principles of chemical kinetics*. Elsevier Press, 2007.

[48] J. Monod, J. Wyman, and J. P. Changeux, "On the nature of allosteric transition: A plausible model," *J. Mol. Bio.*, vol. 12, pp. 88–118, 1965.

[49] D. E. Koshland, G. Nemethy, and D. Filmer., "Comparison of experimental binding data and theoretical models in proteins containing subunits," *Biochemistry*, vol. 5, pp. 365–385, 1966.

[50] G. K. Ackers, M. L. Doyle, D. Myers, and M. A. Daugherty, "Molecular code for cooperativity in hemoglobin," *Science*, vol. 255, pp. 54–63, 1992.

[51] G. K. Ackers, J. M. Holt, E. S. Burgie, and C. S. Yarian, "Analyzing intermediate state cooperativity in hemoglobin," *Methods in Enzymology*, vol. 379, pp. 3–28, 2004.

[52] A. Di Biasio, E. Agliari, A. Barra, and R. Burioni, "Mean-field cooperativity in chemical kinetics," *Theor. Chem. Acc.*, vol. 131, no. 3, pp. 1–14, 2012.

[53] D. ben Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems*. Cambridge University Press, 2000.

[54] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library, 2007.

[55] C. J. Thompson, *Mathematical statistical mechanics*. Princeton University Press, 1979.

[56] Y. Murase, T. Onda, K. Tsujii, and T. Tanaka, "Discontinuous binding of surfactants to a polymer gel resulting from a volume phase transition," *Macromolecules*, vol. 32, pp. 8589–8594, 1999.

[57] D. E. Koshland Jr, "The structural basis of negative cooperativity: receptors and enzymes," *Curr. Opin. Struct. Bio.*, vol. 6, pp. 757–761, 1996.

[58] L. H. Chao, P. Pellicena, S. Deindl, L. A. Barclay, H. Schulman, and J. Kuriyan, "Intersubunit capture of regulatory segments is a component of cooperative camkii activation," *Nature Struct. and Mol. Bio.*, vol. 17, pp. 264–272, March 2010.

[59] M. Mandal, M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo, and R. R. Breaker, "A glycine-dependent riboswitch that uses cooperative binding to control gene expression," *Science*, vol. 306, pp. 275–279, October 2004.

[60] J. M. Bradshaw, Y. Kubota, T. Meyer, and H. Schulman, "An ultrasensitive ca2+/calmodulin-dependent protein kinase ii-protein phosphatase 1 switch facilitates specificity in postsynaptic calcium signaling," *Proc. Natl. Acad. Sci.*, vol. 100, no. 18, pp. 10512–10517, 2003.

[61] J. Kniazeff, A. S. Bessis, D. Maurel, H. Ansanay, L. Prézeau, and J. P. Pin, "Closed state of both binding domains of homodimeric mglu receptors is required for full activity," *Nature struct. & mol. bio.*, vol. 11, no. 8, pp. 706–713, 2004.

[62] X. Rovira, D. Roche, J. Serra, J. Kniazeff, J. P. Pin, and J. Giraldo, "Modeling the binding and function of metabotropic glutamate receptors," *J. Pharm. and Exper. Therapeutics*, vol. 325, no. 2, pp. 443–456, 2008.

[63] Y. Suzuki, E. Moriyoshi, D. Tsuchiya, and H. Jingami, "Negative cooperativity of glutamate binding in the dimeric metabotropic glutamate receptor subtype 1*," *J. Biol. Chem.*, vol. 279, pp. 35526–35534, August 2004.

[64] A. Levitzki and D. E. Koshland Jr, "Negative cooperativity in regulatory enzymes," *Proc. Natl. Acad. Sci.*, vol. 62, no. 4, pp. 1121–1128, 1969.

[65] C. Röcker, M. Pötzl, F. Zhang, W. J. Parak, and G. U. Nienhaus, "A quantitative fluorescence study of protein monolayer formation on colloidal nanoparticles," *Nature nanotech.*, vol. 4, no. 9, pp. 577–580, 2009.

[66] M. Bolognesi, A. Boffi, M. Coletta, A. Mozzarelli, A. Pesce, C. Tarricone, and P. Ascenzi, "Anticooperative ligand binding properties of recombinant ferric *Vitreoscilla* homodimeric hemoglobin: A thermodynamic, kinetic and x-ray crystallographic study," *J. Mol. Bio.*, vol. 291, no. 3, pp. 637–650, 1999.

[67] S. S. Solomatin, M. Greenfeld, and D. Herschlag, "Implications of molecular heterogeneity for the cooperativity of biological macromolecules," *Nature Struct. & Mol. Bio.*, vol. 18, pp. 732–734, 2011.

[68] J. L. Macdonald and L. J. Pike, "Heterogeneity in EFG-binding affinities arises from negative cooperativity in an aggregating system," *Proc. Natl. Acad. Sci.*, vol. 105, pp. 112–117, 2008.

[69] T. Chay and C. Ho, "Statistical mechanics applied to cooperative ligand binding to proteins," *Proc. Natl. Acad. Sci.*, vol. 70, pp. 3914–3918, 1973.

[70] J. Wyman and P. Phillipson, "A probabilistic approach to cooperativity of ligand binding by a polyvalent molecule," *Proc. Natl. Acad. Sci.*, vol. 71, pp. 3431–3434, 1974.

[71] A. Abbas, A. Lichtman, and S. Pillai, "Cellular and molecular immunology," 2012.

[72] T. Kindt, B. Osborne, and R. Goldsby, *Kuby Immunology, 6th Editio.* New York: WH Freeman and Company, 2007.

[73] G. D. Victora and M. C. Nussenzweig, "Germinal centers," *Annual review of immunology*, vol. 30, pp. 429–457, 2012.

[74] A. Tramontano, *The ten most wanted solutions in protein bioinformatics.* CRC Press, 2005.

[75] S. Rabello, A. Coolen, C. Perez-Vicente, and F. Fraternali, "A solvable model of the genesis of amino-acid sequences via coupled dynamics of folding and slow-genetic variation," *Journal of Physics A: Mathematical and Theoretical*, vol. 41, no. 28, p. 285004, 2008.

[76] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. Marks, C. Sander, R. Zecchina, J. Onuchic, T. Hwa, and M. Weigt, "Direct-coupling analysis of residue coevolution captures native contacts across many protein families," *Proceedings of the National Academy of Sciences*, vol. 108, no. 49, pp. E1293–E1301, 2011.

[77] R. White, H. Szurmant, H. J.A., and H. T., "Features of protein–protein interactions in two-component signaling deduced from genomic libraries," *Methods Enzymols*, vol. 422, pp. 75–101, 2007.

[78] J. Skerker, B. Perchuk, A. Siryaporn, E. Lubin, O. Ashenberg, M. Goulian, and M. T. Laub, "Rewiring the specificity of two-component signal transduction systems," *Methods Enzymols*, vol. 133, pp. 1043–1054, 2008.

[79] M. Weigt, R. White, H. Szurmant, J. Hoch, and T. Hwa, "Identification of direct residue contacts in protein–protein interaction by message passing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 1, pp. 67–72, 2009.

[80] D. MacKay, *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[81] C. Baldassi, M. Zamparo, C. Feinauer, A. Procaccini, R. Zecchina, and A. Pagnani, "Multivariate Gaussian Modeling of protein families: Predicting residue contacts and protein-interaction partners, in preparation," 2013.

[82] A. Pagnani, "Private communication," 2013.

[83] J. Benichou, R. Ben-Hamo, Y. Louzoun, and S. Efroni, "Rep-Seq: uncovering the immunological repertoire through next-generation sequencing," *Immunology*, vol. 135, no. 3, pp. 183–191, 2012.

[84] J. Weinstein, N. Jiang, R. White, D. Fisher, and S. Quake, "High-throughput sequencing of the zebrafish antibody repertoire," *Science*, vol. 324, no. 5928, pp. 807–810, 2009.

[85] T. Mora, A. Walczak, W. Bialek, and C. Callan, "Maximum entropy models for antibody diversity," *Proceedings of the National Academy of Sciences*, vol. 107, no. 12, pp. 5405–5410, 2010.

[86] K. Larimore, M. McCormick, H. Robins, and P. Greenberg, "Shaping of human germline igh repertoires revealed by deep sequencing," *The Journal of Immunology*, vol. 189, no. 6, pp. 3221–3230, 2012.

[87] A. Murugan, T. Mora, A. Walczak, and C. Callan, "Statistical inference of the generation probability of T-cell receptors from sequence repertoires," *Proceedings of the National Academy of Sciences*, vol. 109, no. 40, pp. 16161–16166, 2012.

[88] R. Neher and B. Shraiman, "Competition between recombination and epistasis can cause a transition from allele to genotype selection," *Proceedings of the National Academy of Sciences*, vol. 106, no. 16, pp. 6866–6871, 2009.

[89] X. Wu, T. Zhou, J. Zhu, B. Zhang, I. Georgiev, C. Wang, X. Chen, N. Longo, M. Louder, K. McKee, *et al.*, "Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing," *Science*, vol. 333, no. 6049, pp. 1593–1602, 2011.

[90] T. Zhou, I. Georgiev, X. Wu, Z.-Y. Yang, K. Dai, A. Finzi, Y. Do Kwon, J. Scheid, W. Shi, L. Xu, *et al.*, "Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01," *Science*, vol. 329, no. 5993, pp. 811–817, 2010.

[91] M. Bailly-Bechet, S. Bradde, A. Braunstein, A. Flaxman, L. Foini, and R. Zecchina, "Clustering with shallow trees," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 12, p. P12010, 2009.

[92] M. Aizenman and P. Contucci, "On the stability of the quenched state in mean-field spin-glass models," *Journal of statistical physics*, vol. 92, no. 5, pp. 765–783, 1998.

[93] P. Contucci, "Stochastic stability: a review and some perspectives," *Journal of Statistical Physics*, vol. 138, no. 1, pp. 543–550, 2010.

[94] S. Ghirlanda and F. Guerra, "General properties of overlap probability distributions in disordered spin systems. towards parisi ultrametricity," *Journal of Physics A: Mathematical and General*, vol. 31, p. 9149, 1998.