Anergy in self-directed B lymphocytes: A statistical mechanics perspective

Elena Agliari [b], Adriano Barra [b,*], Gino Del Ferraro [a], Francesco Guerra [b], Daniele Tantari [c]

[a] Department of Computational Biology, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden
[b] Dipartimento di Fisica, Sapienza Università di Roma, P.le A. Moro 2, 00185 Roma, Italy
[c] Dipartimento di Matematica, Sapienza Università di Roma, P.le A. Moro 5, 00185 Roma, Italy

HIGHLIGHTS

- The network of interacting B and T clones is described as a bipartite spin-glass.
- The latter has the properties of an attractor network (learning and retrieval).
- The (idiotypic) network of interacting B clones is described as a diluted ferromagnet.
- T-clones dialogue with a subset of B clones only: those highly connected.
- Those clones (idiotypically too connected) that are not signaled are self-directed.

ABSTRACT

Self-directed lymphocytes may evade clonal deletion at ontogenesis but still remain harmless due to a mechanism called clonal anergy. For B-lymphocytes, two major explanations for anergy developed over the last decades: according to Varela theory, anergy stems from a proper orchestration of the whole B-repertoire, such that self-reactive clones, due to intensive feed-back from other clones, display strong inertia when mounting a response. Conversely, according to the model of cognate response, self-reacting cells are not stimulated by helper lymphocytes and the absence of such signaling yields anergy. Through statistical mechanics we show that helpers do not prompt activation of a sub-group of B-cells: remarkably, the latter are just those broadly interacting in the idiotypic network. Hence Varela theory can finally be reabsorbed into the prevailing framework of the cognate response model. Further, we show how the B-repertoire architecture may emerge, where highly connected clones are self-directed as a natural consequence of ontogenetic learning.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Streamlined summary of the biological problem

The adaptive response of the immune system is performed through the coordination of a huge ensemble of cells (e.g. B cells, helper and regulatory T cells, etc.), each with specific features, that interact both directly and via exchanges of chemical messengers such as cytokines and immunoglobulins (antibodies) (Janeway et al., 2005). In particular, a key role is played by B cells, which are lymphocytes characterized by membrane-bound immunoglobulin (BCR) working as receptors able to specifically bind an antigen; upon activation, B cells produce specific soluble immunoglobulin. B cells are divided into clones: roughly speaking, cells belonging to the same clone share the same specificity, that is, they express the same BCR and produce the same antibodies (hyper-somatic mutations apart Janeway et al., 2005). When an antigen enters the host body, some of its fragments are presented to B cells, whereby the clones with the best-matching receptor, after the authorization of helpers through direct contact via CD40-CD40L and via cytokines, undergo clonal expansion and release a huge amount of antibodies in order to kill pathogens and restore order.

This picture, developed by Burnet (1959) in the 50s and verified across the decades, constitutes the "clonal selection theory" and, when focusing on B-lymphocytes only, can be regarded as a one-body theory (Barra and Agliari, 2010a): The growth (drop) of the antigen concentration elicits (inhibits) the specific clones, without any interaction among lymphocytes themselves.

In the 1970s a step forward was taken by Jerne who suggested that, beyond antigenic stimulation, each antibody must also be

detected and acted upon by other antibodies; as a result, the secretion of an atypically large concentration of antibodies by an active B clone (e.g. elicited due to an antigen attack) may even prompt the activation of other B clones that best match those antibodies (Jerne, 1974). This mechanism, experimentally well established (see e.g. Cazenave, 1977; Bernabé et al., 1981), underlies a two-body theory and (possibly) gives rise to an effective network of clones interacting via antibodies, also known as an "idiotypic network".

The B repertoire is enormous ($\sim 10^9$ in humans) and continuously updated due to the random V(D)J recombination occurring during B-cell ontogenesis in the bone marrow (Janeway et al., 2005). The latter process ensures the diversity of the repertoire and therefore the ability of the immune system to recognize many different antigens. On the other hand, it also inevitably produces cells able to detect and attack self-proteins and this possibly constitutes a serious danger. In order to avoid the release of such auto-reactive cells, safety mechanisms are at work during the ontogenesis. Nevertheless some of them succeed in escaping through "receptor editing" (self-reactive cells substitute one of their receptors on their immunoglobulin surface) (Kitamura, 2008) or "clonal anergy" (self-reactive cells that have not been eliminated or edited in the bone marrow become unresponsive, showing reduced expression level of BCR) (Goodnow, 1992, 2005).

In the last decades, two main theories have been proposed to explain clonal anergy, both supported by experimental evidence: The former, introduced by Lundkvist et al. (1989), Stewart et al. (1989), Varela and Coutinho (1991), considers B cells only, while the latter, referred to as the "model of cognate response" (or sometimes "two-signal model") (Goodnow, 1992, 2005; Goodnow et al., 2010), considers both B and helper T cells.

According to Varela theory, each clone μ corresponds to a node in the idiotypic network, with a (weighted) coordination number (i.e. the sum of the binding strengths characterizing its possible interactions with all other clones) W_μ, which represents a measure of the tolerance threshold of the clone. Thus clones corresponding to poorly (highly) connected nodes respond readily (hardly) to the corresponding stimulus. In this way the idiotypic network maintains a regulatory role, where a "core" of highly (weighted) connected clones acts as a safe-bulk against self-reactions. Experimental evidence of this phenomenon has been obtained over the years (Lundkvist et al., 1989; Stewart et al., 1989; Varela and Coutinho, 1991; Lider et al., 1988; Urbain et al., 1977) and even recently (Shoenfeld, 2004). Even so, given the huge size of the B-repertoire and evidences against Varela theory as well, an extensive experimental exploration has always been out of reach, in such a way that the initial promising perspectives offered by the theory were never robustly actualized, and interest in this explanation diminished.

Conversely, according to the modern model of the cognate response, the activation of a B-cell (i.e. antibody production and clonal expansion of its lineage) requires two signals in a given (close) time interval: the first one is delivered by the antigen binding to the BCR, the second one is provided by a helper T lymphocyte, which links to B-cells via direct interaction through *CD40-CD40L* and elicits the B-growth through cytokines. Cytokines constitute a wide class of cell-signaling protein molecules (Theze, 1999) and some of them, e.g. interleukines and interferones, work as immunomodulating agents; the vast majority of these are produced by helper T cells. The effect of a particular cytokine on a given cell depends on the cytokine, its abundance, etc., and it can induce either up regulation or down regulation, leading to enhancement or suppression of the immune response respectively. For example, Interleukin-2 (IL-2) acts in an autocrine manner to stimulate B and T cell proliferation, while Interleukin-10 (IL-10) inhibits responses by blocking activation of accessory cell functions and the synthesis of pro-inflammatory cytokines such as IL-2 (or TNF-α and IL-5) (Janeway et al., 2005). In general, the secretion of a certain cytokine depends on the inflammatory state and on the concentration of ligands on helper receptor (TCR) (Kuchroo et al., 2002), but we will not retain this level of resolution in this paper.

In the absence of the second signal, armed B clones enter a "safe mode" (Kitamura, 2008; Schwartz, 2005), being unable to either proliferate or secrete immunoglobulins. This explanation for anergy has largely prevailed: being based on a local mechanism, its experimental evidence is indisputable. However, it raises the puzzling question of how self-directed B-cells become "invisible" to helpers (Hartley et al., 1991). Furthermore, it does not incorporate previous findings of Varela's picture, whose experimental evidence should however be consistent with this prevailing scheme.

The aim of this paper is to try to answer these questions through techniques stemming from theoretical physics. Interestingly, the scenario we outline robustly evidences the fact that highly connected B cells are transparent to helpers, hence merging the two mechanisms for anergy: Note that this statement does not imply that Varela theory and the model of the cognate response (two-signal model) are equivalent, but, only that Varela theory can be framed within the model of the cognate response and, in this sense, the two theories are only different perspectives of the prevailing one, the cognate response. The major benefit from our investigation is that now all data collected in support of Varela theory, which where somehow split from the rest of modern immunology can be easily framed within the actual scaffold.

1.2. An overview on the statistical mechanics approach

In this work we adopt a statistical-mechanics approach to modeling the immune system. Being theoretically firmly based on the law of large numbers and on the maximum entropy principle (Jaynes, 1957a, 1957b), statistical mechanics aims to figure out collective phenomena, possibly overlooking the details of the interactions to focus on the very key features: trying to summarize its philosophy in a practical example, despite somehow counterintuitive, there is absolutely no need of all the microscopic descriptions regarding a molecule of water to predict that a huge ensemble of these molecules will change its phase (e.g. liquid, vapor) when tuning its temperature across the evaporation threshold. Although this perspective certainly implies a certain degree of simplification, it has been successfully applied to a wide range of fields, e.g., material sciences (Allen and Tildesley, 1987; Frenkel and Smith, 2002), sociology (Daurlauf, 1999; Brock and Daurlauf, 2001), informatics (Mezard and Montanari, 2007), economics (Coolen, 2005; Bouchaud and Potters, 2000), artificial intelligence (Amit, 1992; Coolen et al., 2005), and system biology (Martelli et al., 2009; Kaufman, 1969). Statistical mechanics was also proposed as a candidate instrument for theoretical immunology in the seminal work by Parisi (1990). In fact, many emergent properties of the immune system can be nicely fitted within the systemic perspective offered by statistical mechanics, as, for instance, discussed by Germain: "as one dissects the immune system at finer and finer levels of resolution, there is actually a decreasing predictability in the behavior of any particular unit of function", furthermore, "no individual cell requires two signals (…) rather, the probability that many cells will divide more often is increased by costimulation" (Germain, 2001). Understanding this averaged behavior is just the goal of statistical mechanics.

Moreover, concepts such as "decision making", "learning process" or "memory" are widespread both in immunology (Chakraborty and Košmrlj, 2011; Depino, 2010; Floreano and Mattiussi, 2008) and in statistical mechanics (or, more precisely, in a branch of the field,

namely neural networks Amit, 1992; Coolen et al., 2005). Indeed, clones, existing as either active or non-active and being able to collectively interact, could replace the digital processing units (e.g. flip-flops in artificial intelligence Hopfield and Tank, 1987, or neurons in neurobiology Tuckwell, 2005) and cytokines, carrying either eliciting or suppressive chemical signals, could replace connections (e.g. cables and inverters in artificial intelligence, or dendrites and synapses in neurobiology).

As a last remark, we stress that, as it is typical in a statistical mechanics formalization (see e.g. Coolen et al., 2005), we first develop the simplest scenario assuming symmetry for the inter-actions among B and T cells. This assumption has the strong advantage of allowing us to create a clear equilibrium picture that is still able to capture the main features of the system under study, whose off-equilibrium properties (immediately achievable in the opposite, fully asymmetric, limit) should yet retain strong simila-rities with the present picture and will be addressed in future investigations (as has already been done in the neural counterpart Amit, 1992; Coolen, 2005).

Having sketched the underlying philosophy of our work, we highlight our two key results: We first consider the B–T network and show that *T cells are unable to communicate with highly connected B-cells*; Then, we consider the set of B clones and show that a minimal (biased) learning process, during B-cell clonal deletion at ontogenesis, can shape the final repertoire such that *highly connected B clones are typically self-directed*. These two points together allow us to merge the model of the cognate response and the theory of Varela.

The plan of the paper can be summarized by the following logical concatenation of notes:

Part I: *Anergy induced by T cells and the model of the cognate response.*

- Experimental evidence: The (bulk of the) response of B-cells is prompted by two signals: the presence of an antigen and the "consensus" by a helper T lymphocyte (Janeway et al., 2005).
- Theoretical consequence: The ensembles made by B and helper clones interact as a (diluted Agliari et al., 2012a, 2013b) bilayer restricted Boltzmann machine (a two-party spin-glass in the language of disordered statistical mechanics).
- Experimental evidence: Helper cells do not prompt response in B-cells self-directed, thus the latter behave *anergically*.

Part II: *Anergy induced by B cells and the theory of Francisco Varela.*

- Experimental evidence: Antibodies (as any other protein) are not random objects (for instance, randomly generated proteins may not even be able to fold into a stable structure Rabello et al., 2008) (Mora et al., 2010).
- Theoretical consequence: Therefore, once expressed through e. g. bit-strings of information, for example, the related entropy is not maximal (see Appendix Three).
- Theoretical consequence: In the idiotypic network where B-clones are nodes and (weighted) links among them mirror the interactions through the related antibodies, nodes with higher weighted connectivity are more inhibited in reaction and typically self-directed.

Global consequence: The system of helpers[1] is (thermodynami-cally) equivalent to an associative "neural" network, whose equili-brium states correspond to optimal orchestrations of T cells leading to maximal signal on the proper B clones, which are therefore

---
[1] Note that we mean both helpers and suppressors with the term "helpers".

prompted to react. Remarkably, the activation of B-clones with highly weighted connectivity corresponds to negligible basins of attraction in the free energy landscape of T cells, hence they are rarely signaled by helpers. This last point allows to incorporate Varela theory for anergy directly within the two-signal model scenario.

1.3. Preliminary remarks on the structure of the B-network

There are several approaches in estimating the structure, size and shape of the mature B repertoire. For instance, in their pioneering works, Jerne and Burnet used a coarse-grained descrip-tion in terms of epitopes and paratopes (Jerne, 1974; Burnet, 1959), which Perelson extended (and symmetrized) introducing a shape space (Perelson, 1997). De Boer and coworkers dealt directly with peptides of fixed length (Burroughs et al., 2004), while Bialek, Callan and coworkers recently used the genetic alphabet made up of the VDJ genes codifying for the heavy and light chains of the immunoglobulins (Mora et al., 2010; Murugan et al., 2012).

Proceeding along a perspective of general information theory, we associate to each antibody, labeled as μ, a binary string Ψ_μ of length L, which effectively carries information on its structure and on its ability to form complexes with other antibodies or antigens. Since antibodies secreted by cells belonging to the same clone share the same structure, the same string Ψ_μ is used to encode the specificity of the whole related B clone. In this way, the repertoire will be represented by the set $\mathcal{B}$ of properly generated strings and its cardinality $N_B = |\mathcal{B}|$ is the number of clones present in the system. We expect that L must be relatively short with respect to the repertoire size N_B, i.e. $L = \gamma \ln N_B$, $\gamma \in \mathbb{R}^+$ (Barra and Agliari, 2010a). This choice stems from both the probabilistic combinator-ial usage of the VDJ recombination (Mora et al., 2010) (when thinking of bit-string entries as genes) and direct experimental evidence (Dreyer and Bennett, 1965) (when thinking of bit-string entries as epitopes).

Antibodies can bind each other through interactions that are mainly hydrophobic and electrostatic and chemical affinities range over several orders of magnitude (Janeway et al., 2005): This suggests that the more complementary two structures are, the more likely (on an exponential scale) their binding. We therefore define χ as a Hamming distance

$$\chi_{\mu\nu} = \sum_{k=1}^{L} [\Psi_\mu^k(1-\Psi_\nu^k) + \Psi_\nu^k(1-\Psi_\mu^k)], \tag{1}$$

to measure the complementarity between two bit-strings Ψ_μ, Ψ_ν, and introduce a phenomenological coupling (whose details will be deepened in Section 4, see also Barra and Agliari, 2010a; Agliari et al., 2012b)

$$J_{\mu\nu} \propto e^{\alpha\chi_{\mu\nu}}, \tag{2}$$

where α tunes the interaction strength. In this way, a network emerges where nodes are B-clones, and (weighted) links are given by the coupling matrix J (see Fig. 1 (left column, upper panel), and Barra and Agliari, 2010a; Agliari et al., 2012b; Brede and Behn, 2001, 2003; Schmidtchen et al., 2012 for details). This formalizes Jerne's idiotypic network.

For the sake of completeness, we stress that, once a clone has been selected for expansion, it undergoes a process called "hyper-somatic mutation" intended to increase its affinity (complemen-tarity) with the antigen detected (Janeway et al., 2005) and to improve the immune response. In the present paper we do not enter into these details as we want to analyze the development of anergy, which is a mechanism at work before clonal expansions.

As several links may stem from the same node, say μ, it is useful to define its weighted degree as $W_\mu = \sum_{\nu=1}^{N_B} J_{\mu\nu}$. When the system is at rest, we can argue that all B clones are inactive, so that
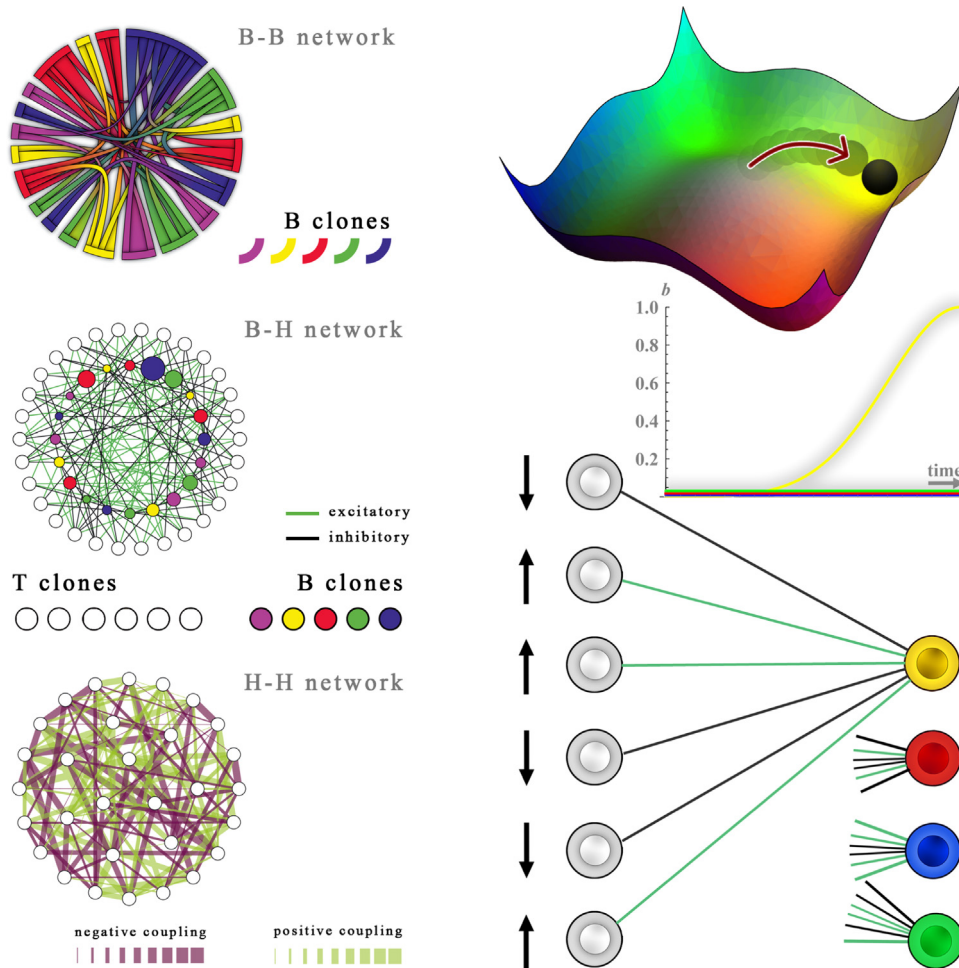
Fig. 1. (left column) Schematic representation of the immune networks considered here, where we fixed $N_{6_T}=30$ and $N_B=20$. The upper graph describes the B–B network: each B-cell μ corresponds to a different arch, whose length is proportional to the related weighted degree W_μ, and the interaction between cells μ and ν corresponds to the link connecting the related arches, whose thickness is proportional to $J_{\mu\nu}$. The middle plot describes the bipartite B–T network: the external set of white circles corresponds to the set of T cells, while the internal set of colored circles corresponds to the set of B cells and their size is proportional to the related weighted degree, according to the plot in the left panel. The interaction ξ between T cells and B cells can be either excitatory (bright link) or inhibitory (dark link). The lower graph describes the T–T network: the white circles correspond to the set of T cells and connections between them are drawn according to the formula $\sum_\mu (\xi_i^\mu \xi_j^\mu)/W_\mu$, as explained in the text; the color and thickness of the link carry information about the sign and the magnitude of the related coupling, respectively. (right column) Schematic representation of the consequence of retrieval capabilities by the helper network in the bipartite network made up of both helpers and B-clones (different colors in the B-branch account for different clonal specificity): in the upper panel a free-energy landscape of the helper network with four minima (each corresponding to retrieval of instruction for a particular B-clone) is shown. The black ball represents the state of the system, which is driven into the yellow hole (e.g. due to antigenic stimulation). In the bipartite network (lower panel) this configuration corresponds to all helpers (white nodes) parallel to the sign of the cytokines linking them to the yellow B clone, here green (black) links represent positive (negative) cytokine interactions. This layout confers maximal strength on the retrieved clone, which undergoes clonal expansion. The latter is represented in the middle plot where the size of the four different B clones is shown versus time. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

if clone μ is stimulated, W_μ can be interpreted as the "inertia" of clone μ in reaction, due to all other cells (Barra and Agliari, 2010b): This mechanism naturally also accounts for the "low dose phenomenon" (Janeway et al., 2005; Barra and Agliari, 2010a, 2010b).

Finally, it is worth considering how W is distributed as this provides information about the occurrence of inertial nodes in the system. Exploiting the fact that couplings $J_{\mu\nu}$ are log-normally distributed (Agliari et al., 2012b), one can approximate the distribution $P(W)$ as

$$P(W) \sim \frac{1}{W\sqrt{2\pi}\sigma} e^{-(\log W - \mu)^2/2\sigma^2}, \qquad (3)$$

in such a way that mean and variance read as $E(W)=e^{\mu+\sigma^2/2}$, $V(W)=(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$, respectively (a detailed discussion on the parameters σ and μ can be found in Section 3 and in Appendix Five).

We underline that the log-normal distribution evidenced here agrees with experimental findings (Carneiro et al., 1996a, 1996b).

Furthermore, its envelope remains log-normal even if the network is under-percolated (Agliari et al., 2012b). Thus, in order to achieve a broad weighted connectivity, the effective presence of a large, connected B-network is not a requisite, but, basically, the mere existence of small-size components, commonly seen in experiments (Cazenave, 1977; Bernabé et al., 1981), is needed.

2. Anergy induced by T cells and the cognate response model

2.1. Stochastic dynamics for the evolution of clonal size

We denote with $b_\mu \in \mathbb{R}$ the "degree of activation" of the B clone μ with respect to a reference value b_0, such that if the clone is in its equilibrium state (i.e. at rest) $b_\mu = b_0$, while if the clone has expanded (is suppressed) $b_\mu > b_0$ ($b_\mu < b_0$). Again, we adopt the simplest assumption and fix a unique reference state $b_0=0$ for all

the clones; the case of tunable b_0 was treated in Agliari et al. (2011).

Concerning T cells, both helper and regulatory sub-classes share information with the B branch in a complex ensemble of paths (Janeway et al., 2005) that here we summarize saying "via cytokines". Hence, we group them into a unique ensemble of size N_T, and denote the state of each clone by h_i ($i = 1, ..., N_T$); hereafter we call them simply "helpers". We take $h_i = \pm 1$ such that $h_i = +1$ stands for an active state (secretion of cytokines) and vice versa for -1. Actually the choice of binary variables is neither a biological requisite nor a mathematical constraint, but it allows us to keep the treatment as simple as possible, while still preserving the qualitative features of the model that we want to highlight.

Now we introduce the relative dynamics for the N_B B clones and the N_T T clones. As the state of the former is denoted by a real variable, the corresponding dynamics will be of Langevin-type. As the state of the latter is denoted by a binary variable, the corresponding dynamics will be of Glauber-type. These are not restrictions imposed by the theory, but rather the simplest (and more CPU-time efficient) ways to show the spontaneous emergence of the phenomenon of anergy, which is the goal of the present investigation.

We define $\epsilon \equiv N_B/N_T$ and, to take advantage of the central limit theorem (CLT), we focus on the infinite volume (thermodynamic limit, TDL), such that, as $N_B \to \infty$ and $N_T \to \infty$, ϵ is kept constant. This mirrors the fact that the global amount of helpers and of B-clones is comparable.

Recalling that B clones receive two main signals, i.e. from other B clones and from T clones, we can introduce the Langevin dynamics for their evolution as

$$\tau \frac{db_\mu(A_k)}{dt} = \sum_{\nu=1}^{N_B} J_{\mu\nu}(b_\nu - b_\mu) + \frac{1}{\sqrt{N_T}} \sum_{i=1}^{N_T} \xi_i^\mu h_i + J_{k\mu} A_k + \sqrt{\frac{\tau}{\beta}} \eta_\mu, \tag{4}$$

where τ rules the characteristic timescale of B cells and β tunes the amplitude of a Brownian noise η. The coupling between the μ-th B clone and the i-th T clone is realized by the ensemble of cytokine ξ_i^μ (see Fig. 1, left column, middle panel) and A_k is a generic k antigenic peptide that interacts with B-clones through the coupling $J_{k\mu}$, which is mediated by antigen presenting cells (APC) (Janeway et al., 2005).

As long as all the interactions are symmetric, the Langevin dynamics admits a Hamiltonian description as

$$\tau \frac{db_\mu(A_k)}{dt} = -\frac{d}{db_\mu} \mathcal{H}_{N_T,N_B}^B(b, h|J, \xi) + \sqrt{\frac{\tau}{\beta}} \eta_\mu,$$

where, by integration over b_μ,

$$\mathcal{H}_{N_T,N_B}^B = \sum_{\mu=1}^{N_B} \sum_{\nu=1}^{N_B} \frac{J_{\mu\nu}}{4}(b_\mu - b_\nu)^2 - \frac{1}{\sqrt{N_T}} \sum_{i=1}^{N_T} \sum_{\mu=1}^{N_B} \xi_i^\mu h_i b_\mu - \sum_{\mu=1}^{N_B} J_{\mu k} b_\mu A_k. \tag{5}$$

Each contribution appearing in the r.h.s. of the previous equation is now deepened:

- The first term comes from B-clone interactions via immuno-globulin, which is translated into a diluted "ferromagnetic" coupling $J_{\mu\nu} \geq 0$ (as B clones tend to "imitate" each another) modulating an elastic membrane-like interaction. According to this term, two different B-clones at different concentrations will tend to reach a common equilibrium at an intermediate level of activation: the presence of a clone stimulates the activation of the other, but at the same time their Abs tend to neutralize each other until reaching a dynamical equilibrium configuration where the two effects are balanced.
- The second term represents the coupling between B and T clones, mediated by cytokines. The cytokine ξ_i^μ is meant to

connect cells of the i-th helper clone and those of the μ-th B one. The message conceived can be either excitatory ($\xi_i^\mu = +1$, e.g. an eliciting IL-2) or inhibitory ($\xi_i^\mu = -1$, e.g. a suppressive IL-35) and here is assumed to be a quenched variable, such that the one with inhibitory effects can be associated to a regulatory cell, conversely, the one with stimulating effect can be associated to an helper cell. Note that the choice ± 1 for ξ_i^μ is only a convenient requisite for encoding two opposite effects, while, clearly, their world is by far richer (Kuchroo et al., 2002), and, in principle, also mathematically accessible.
- The third term mimics the interaction of the generic b_μ clone with the antigenic peptide A_k (shown by an APC), where $J_{\mu k}$ encodes their coupling strength and can be defined according to Eq. (2).

Interestingly, in the Hamiltonian (5), the first term recovers Jerne's idiotypic network theory, the second one captures the two-signal model and the third one recovers Burnet's clonal selection theory. Within this statistical mechanics framework the three approaches are not conflicting, but, rather, interplaying.

As for the helpers, we introduce the Glauber dynamics as

$$P(h_i \to -h_i) = \frac{1}{1 + \exp[2\beta\Delta\mathcal{H}_{N_T,N_B}^H(h_i)]}, \tag{6}$$

where

$$\mathcal{H}_{N_T,N_B}^H = -\frac{1}{\sqrt{N_T}} \sum_{i,\mu}^{N_T,N_B} \xi_i^\mu b_\mu h_i + \sum_i^{N_T} \tilde{J}_{ij} A_l h_i. \tag{7}$$

Two important (and related) differences, and one consequence, should be carefully remarked when comparing the two dynamics.

- First, focusing on BCR and TCR couplings, the B-clone dynamics include an interaction term among B-cells (which accounts for the idiotypic network mimicking BCR–BCR and BCR–Ab interactions), while an analogous term for T cells is neglected in the helper dynamics. This reflects a structural difference between TCR and BCR: beyond contact interactions, BCR's can also be coupled also via antibodies (hence they can interact even beyond a peer-to-peer mechanism, in a "long-range" fashion) while TCR's cannot (Janeway et al., 2005).
- Second, TCR is expected to be much more specific than BCR because the clonal deletion experienced by T-cells during ontogeny in the thymus is much more selective than the corresponding deletion undergone by B-cells in the bone marrow (Košmrlj et al., 2009; Murugan et al., 2012; Efroni et al., 2003, 2007), hence the capacity of interaction between TCRs is further reduced.
- Although beyond the focus of the present paper, we stress that these differences are also reflected in the interactions with the antigens: B cells and T cells are coupled with antigens through different clone-specific parameters, i.e. J and $\tilde{J}$, respectively. Hence the cross-correlation $\langle J\tilde{J} \rangle$ tunes the degree of overlap between their receptors (and acts as a measure of the efficiency of the double check realized by B and T cells through the cognate-response): Clearly, for $\langle J\tilde{J} \rangle \to 1$, T cells do not provide any safety mechanism, while if the correlation is null, hence $\langle J\tilde{J} \rangle \to 0$, no antigen would be able to arm both B and T cells and the immune response would not be mounted: Between these two extrema the diversity of BCR and TCR constitutes the hallmark of the response of adaptive immune systems.

We now proceed with the exploration of the structure of the phase-space displayed by the model and, to this end, we drop the antigenic terms, that is, the "external fields" (hence we set $A_k = 0$ for all k). This is a standard procedure in statistical mechanics as it allows to obtain a complete picture of the unperturbed landscape. In particular, we will see that the system displays several equilibrium configurations,

each corresponding to the activation of a particular subset of clones tuning a particular response. Then, fields (i.e. antigens) will be resumed to show that the presence of an antigen prompts the system to select the configuration where activated clones are exactly those able to attack that particular antigen (and not self-ones).

Moreover, close to equilibrium, the anti-antibodies can be neglected ($b_\mu b_\nu \sim 0$), thus, remembering that $\sum_\nu^{N_B} J_{\mu\nu} = W_\mu$, the global (field-free) Hamiltonian of the process reduces to

$$\hat{\mathcal{H}}_{N_T,N_B} = \mathcal{H}_{N_T,N_B}^B \bigcup \mathcal{H}_{N_T,N_B}^H$$
$$= \frac{1}{2} \sum_{\nu=1}^{N_B} W_\nu b_\nu^2 - \frac{1}{\sqrt{N_T}} \sum_{i=1}^{N_T} \sum_{\nu=1}^{N_B} \xi_i^\nu h_i b_\nu. \tag{8}$$

This means that, while for any given configuration of $\{h_i\}$ the Langevin dynamic for the B clones has an equilibrium distribution $P(b|h) \propto e^{-\beta \mathcal{H}_{N_T,N_B}^B}$ and, in the same way, for any given configuration of $\{b_\mu\}$, $P(h|b) \propto e^{-\beta \mathcal{H}_{N_T,N_B}^H}$ is the equilibrium distribution for the Glauber dynamic of T clones, the global coupled process admits an equilibrium joint probability distribution

$$P(b,h|\beta; \xi, J) \propto e^{-\beta \hat{\mathcal{H}}_{N_T,N_B}}, \tag{9}$$

whose properties will be addressed in the next section through statistical mechanics.

2.2. The equivalence with associative networks

Once the effective Hamiltonian is defined through Eq. (8), the classical statistical mechanics package can be introduced. This implies the partition function

$$Z_{N_T,N_B}(\beta|\xi,W) = \sum_{\{h\}} \int \prod_{\mu=1}^{N_B} db_\mu e^{-(\beta/2) \sum_{\mu=1}^{B} W_\mu b_\mu^2 + (\beta/\sqrt{N_T}) \sum_{\mu,i}^{N_B,N_T} \xi_i^\mu h_i b_\mu} \tag{10}$$

and the quenched free-energy

$$A(\beta, \epsilon | P(W)) = \lim_{N_T,N_B \to \infty} \frac{1}{N_T} \mathbb{E} \ln Z_{N_T,N_B}(\beta|\xi,W), \tag{11}$$

where $\mathbb{E}$ averages over both the ξ and the W distributions.

Notice that the idiotypic contribution in the stochastic process (4) implicitly generates a Gaussian distribution for the activity of the B-clones

$$P(b_\mu|W) \propto \exp(-\beta W_\mu b_\mu^2 / 2). \tag{12}$$

This is consistent with commonly observed data and ensures convergence of the integrals in the partition function (10). Interestingly, W_μ^{-1} plays as variance for $P(b_\mu)$.

A crucial point is that the integrals over $\{b_\mu\}$ in the partition function (10) can be calculated explicitly to give

$$Z_{N_T,N_B}(\beta|\xi,W) = \sum_{\{h\}} \exp\left(\frac{\beta}{2N_T} \sum_{i,j}^{N_T,N_T} \sum_\mu^{N_B} \frac{\xi_i^\mu \xi_j^\mu}{W_\mu} h_i h_j\right), \tag{13}$$

by which the complete Hamiltonian (where a generic antigen A_l is present for the sake of completeness) ruling helper's dynamics can be extracted and reads as

$$\mathcal{H}_{N_T,N_B}^H = -\frac{1}{2N_T} \sum_{ij}^{N_T,N_T} \left(\sum_\mu^{N_B} \frac{\xi_i^\mu \xi_j^\mu}{W_\mu}\right) h_i h_j - \sum_i^{N_T} \tilde{J}_{il} h_i A_l,$$

which is consistent with Eq. (7), upon marginalization over $\{b_\mu\}$. The previous expression deserves attention because it corresponds to a (log-normally weighted) Hopfield model for neural networks (Barra et al., 2012; Amit, 1992) (see Fig. 1, left column, lower panel). Its Hebbian kernel suggests that the network of helpers, driven by the antigen A_l, is able to orchestrate strategies (thought of as patterns of cytokines) if the ratio $\epsilon = N_B/N_T$ does not exceed a threshold (Agliari et al., 2011). This is in agreement with the breakdown of immuno-

surveillance occurring whenever the amount of helpers is too small (e.g. in long-term HIV infections) or the amount of B is too high (e.g. in strong EBV infections): These capabilities of the system are minimally addressed in the present paper and again we refer to Agliari et al. (2011) for further insights and to Agliari et al. (2013c, 2012a) for the investigation of its parallel processing performances (namely the ability of managing several clones simultaneously Vertosick and Kelly, 1989; Agliari et al., 2013a, 2013b). At a first glance, it may appear *strange* that a quiescent T-cell, say $\tilde{h}_j$ (hence $\tilde{h}_j = -1$), supplying the clone b_ν (candidate for clonal expansion) for an inhibiting cytokine $\tilde{\xi}_j^\nu$ (hence $\tilde{\xi}_j^\nu = -1$ too) results in stimulation (as the field experienced by b_ν due to this T-clone is $\tilde{\xi}_j^\nu \tilde{h}_j = +1$); however, as schematically shown in the bottom right picture of Fig. 1, the B-clone receives the maximum expansion signal when all the T-cells (helpers) secreting eliciting cytokines are firing as well as when all the suppressors (hence T-cells secreting inhibitory signals) are quiescent (for otherwise conflicting signals would weaken B expansion).

2.3. High connectivity leads to anergy

To prove that "high connectivity (in the idiotypic network) leads to anergy" we need to show two things: The first is that highly connected B-clones are not signaled by helpers, and this will be achieved in this section. The second step consists in showing that highly connected B-clones are usually self-directed, and this will be studied in the next section.

As anticipated, the network made up of helper cells can work as a neural network able to retrieve "patterns of information". There are overall N_B patterns of information encoded by cytokine arrangement $\{\xi\}$ and the retrieval of the pattern μ means that the state of any arbitrary i-th T clone agrees with the cytokine ξ_i^μ, namely $h_i \xi_i^\mu = +1$. This ultimately means that clone μ is maximally stimulated. A schematic representation of retrieval performed by T cells and of its consequence on the repertoire of B cells is depicted in the right column of Fig. 1.

Here, with respect to standard Hopfield networks, Hebbian couplings are softened by the weighted connectivity W_μ and this has some profound effects. In fact, the patterns of information which can be better retrieved (i.e. the clones which can be more intensively signaled) are those corresponding to a larger signal, that is, a smaller W. Thus, B-clones with high weighted connectivity (the safe-bulk) cannot be effectively targeted and, in the TDL, those B-clones exhibiting $W \to \infty$ are completely "transparent" to helper signaling, hence allowing finally to include data collected in support of Varela theory into the two-signal model.

Deepening this point is now mainly technical. We introduce the N_B pattern-overlaps $\langle m_\mu \rangle$, which measure the extent of pattern retrieval, i.e. signaling on clone μ, and are defined as $\langle m_\mu \rangle = \mathbb{E} N_T^{-1} \Omega(\sum_i^{N_T} \xi_i^\mu h_i)$, where Ω is the standard Boltzmann state (Ellis, 1985) associated to the free energy (11). The Hamiltonian corresponding to Eq. (13) can therefore be recast as

$$\mathcal{H}_{N_T,N_B}(h|\xi,W) = -\frac{1}{2N_T} \sum_{i,j}^{N_T,N_T} \left(\sum_{\mu=1}^{N_B} \frac{\xi_i^\mu \xi_j^\mu}{W_\mu}\right) h_i h_j$$
$$= -\frac{N_T}{2} \sum_{\mu=1}^{N_B} \frac{m_\mu^2}{W_\mu}. \tag{14}$$

Now, free energy minimization implies that the system spontaneously tries to reach a retrieval state where $\langle m_\mu \rangle \to 1$ for some μ. Of course, this is more likely for clones μ with smaller W_μ, while highly connected ones are expected not to be signaled (pathological cases apart, i.e. no noise $\beta \to \infty$, or giant clonal expansions $b_0 \to \infty$).

Note that $\langle m_\mu \rangle = 1$ (gauge-invariance apart) means that all the helpers belonging to the clone i are parallel to their corresponding cytokine. Hence if ξ_i^μ is an eliciting messenger, the corresponding helper h_i will be firing, conversely for $\xi_j^\mu = 1$ the corresponding helper h_j will be quiescent, so to confer to the b_μ clone, overall, the maximal expansion signal (the random environment becomes a deterministic field).

In order to figure out the concrete existence of this retrieval, we solved the model through standard replica trick (Mezard et al., 1987), at the replica symmetric level (see Appendix One), and we numerically found solutions for the obtained self-consistence equations, which read off as

$$\langle m_1(\epsilon, \beta) \rangle = \langle\langle \xi^1 \tanh\left[\beta(m_1\xi^1/W_1 + \sqrt{\epsilon r}z)\right]\rangle_z\rangle_{\xi,W},$$

$$\langle q(\epsilon, \beta) \rangle = \langle\langle \tanh^2\left[\beta(m_1\xi^1/W_1 + \sqrt{\epsilon r}z)\right]\rangle_z\rangle_{\xi,W},$$

$$\langle r(\epsilon, \beta) \rangle = \lim_{N_T \to \infty} \frac{1}{\epsilon N_T} \sum_{\mu > 1}^{N_B} \frac{q}{[W_\mu - \beta(1-q)]^2}. \tag{15}$$

In this set of equations, we used the label 1 to denote a test B-clone $\mu = 1$, which can be either a self-node (i.e. with a high value of W_1, infinite in the TDL) or a non-self-clone (i.e. with a small value of W_1, zero in the TDL). While the first equation defines the capability of retrieval by the immune network as explained above, q is the Edward–Anderson spin-glass order parameter (Mezard et al., 1987) and r accounts for the slow noise in the network due both to the number of stored strategies and to the weighted connectivity (these equations generalize the Hopfield equations Amit, 1992 which clearly are recovered by setting $W_\mu = 1$ for all $\mu = 1, .., N_B$).

As shown in the Appendix Two, the equations above can be solved in complete generality. Here, for simplicity, we choose $\mu = 1$ as the test-case (hence B_1 is the B-clone that has to be expanded) and describe the outcome obtained by replacing all W_μ with $\mu \neq 1$ with their average behavior, namely $\langle W \rangle = \int dW P(W)W$; this assumption makes the evaluation of the order parameter r much easier, while still preserving the qualitative outcome.

We now focus on the two limiting cases: $W_1 \ll \langle W \rangle$, which accounts for a non-self-node, and $W_1 \gg \langle W \rangle$, which mirrors the self-counterpart.

In the former case, the slow noise is small (vanishing as $\langle W \rangle \to \infty$). Consequently, the non-self-nodes live in a free environment and the corresponding equations for their retrieval collapse to the non-saturated Hopfield model (Amit, 1992). Hence, retrieval should always be possible (ergodic limit apart), therefore, in this case, helpers can effectively signal clone 1.

Conversely, in the latter case, for a self-node, it is straightforward to check that the noise rescaling due to W implies a critical noise level for the retrieval $\beta^{-1} \sim W_1^{-1} \sim 0$ (as W_1 is ideally diverging in the thermodynamic limit, see Fig. 4 and Section 5). An alternative view is in terms of the basins of attraction for the free energy of the B–H network: Mirroring the toy phase diagram built of four minima (i.e. instructions for B-clones) only of Fig. 1 (top right), in Fig. 2 we show how progressively, while increasing the connectivity of two (out of the four) clones, their minima in the corresponding landscape disappear, hence the related B-clones can no longer be signaled. As a result, under normal conditions, the retrieval of patterns enhancing self-node clonal expansions is never performed by helpers: This behavior mimics anergy as a natural emergent property of these networks.

3. Anergy induced by B cells and Varela theory

So far we have shown that helper cells are unable to exchange signals with highly connected B-clones, however, the reason why the latter should be self-directed is, so far, still puzzling. We now build a basic model for the ontogenetic process of B cells, which solely assumes that self-proteins are not random objects, and we show that survival clones expressing large self-avidity are those highly connected.

3.1. Ontogeny and the emergence of a biased repertoire

During ontogenesis in the bone marrow, B-cell survival requires sufficiently strong binding to at least one self-molecule (positive selection), but those cells which bind too strongly are also deleted (negative selection). Such conditions ensure that surviving B cells are neither aberrant nor potentially harmful to the host (Košmrlj et al., 2009, 2008).

To simulate this process, we model the ensemble of self-molecules as a set $\mathcal{S}$ of strings Φ_μ, of length L, whose entries are extracted independently via a proper distribution. The overall number of self-molecules is $|\mathcal{S}| = N_S$, that is, $\mu = 1, ..., N_S$.

As stated in the Introduction, although a certain degree of randomness seems to be present even in biological systems, proteins are clearly non-completely random objects (Rabello et al., 2008): Indeed, the estimated size of the set of self-proteins is much smaller than expected from randomly generated sets (Mora et al., 2010). Within an information theory context, this means that the entropy of such repertoire is not maximal, that is, within the set $\mathcal{S}$ some self-proteins are more likely than others (see Appendix Three).

In order to account for this feature, we generate $\mathcal{S}$ extracting each string entry i according to the simplest biased-distribution

$$P_{\text{self}}(\Phi_i^\mu | \overline{a}) = \delta(\Phi_i^\mu + 1)\frac{1+\overline{a}}{2} + \delta(\Phi_i^\mu - 1)\frac{1-\overline{a}}{2}, \tag{16}$$

where $\delta(x)$ is the Dirac delta (returning 1 if $x = 0$ and 0 otherwise) and $\overline{a} \in [-1, 1]$ is a parameter tuning the degree of bias, i.e. the likelihood of repetitions among string-bits. Of course, when $\overline{a} = 0$ the complete random scenario is recovered. We stress that here, looking for minimal requisites, we neglect correlations among string entries (Mora et al., 2010), in favor of a simple mean-field approach where entries are identically and independently generated.

As underlined above, a newborn B cell, represented by an arbitrary string Ψ, undergoes a screening process and the condition for survival can be restated as

$$\chi_P < \max_{\Phi \in \mathcal{S}}\{\chi(\Psi, \Phi)\} < \chi_N, \tag{17}$$

χ_P and χ_N being the thresholds corresponding to positive and negative selection, respectively.

As explained in Appendix Four, the value of the parameters χ_P and χ_N can be fixed according to indirect measurements, such as the survival probability of new-born B cells. It is widely accepted that human bone marrow produces daily $\sim 10^7$ B cells, but only $\sim 10^6$ are allowed to circulate in the body (Chakraborty and Košmrlj, 2011; Košmrlj et al., 2008). The remaining 90% undergo apoptosis since they are targeted as self-reactive ones (Wardemann et al., 2003; Allman et al., 1993; Rolink et al., 1998); therefore the expected survival probability for a new-born B cell is $P_{\text{surv}} = 0.1$ (see Fig. 3, left panel).

Thus, having properly fixed χ_P and χ_N, we extract randomly and independently a string Ψ and we check whether Eq. (17) is fulfilled; if so, the string is selected to make up the repertoire $\mathcal{B}$. We proceed sequentially in this way until the prescribed size N_B is attained (see Appendix Four for more details).

Before proceeding it is worth noticing that the final repertoire $\mathcal{B}$ is expected to exhibit a certain degree of inhomogeneity possibly described by a proper probability distribution including self-addressed cells as well. However, due to the mean-field
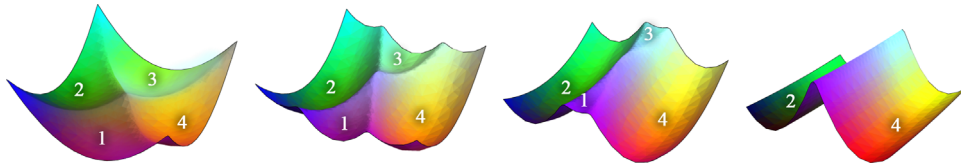
Fig. 2. Schematic representation of the (free-energy) basins of attractions for a toy system starting (at left) with four minima (hence four retrievable patterns). Each minimum contains information addressed to the corresponding B-clone so that four B-clones B_1, B_2, B_3, B_4 can be instructed in the initial configuration. From left to right we fix $W_2 = W_4 = 1$ always, while we increase progressively $W_1 = W_3 = 1, 5, 10, 100$ (and we show the resulting basins of attraction from left to right). Note that at the value of the weighted connectivity $W_1 = W_3 = 100$, the corresponding minima completely disappear. Hence instructions to the corresponding B-clones (which are broadly interacting as their W is much higher than $W_2 = W_4 = 1$) cannot be supplied by helpers.
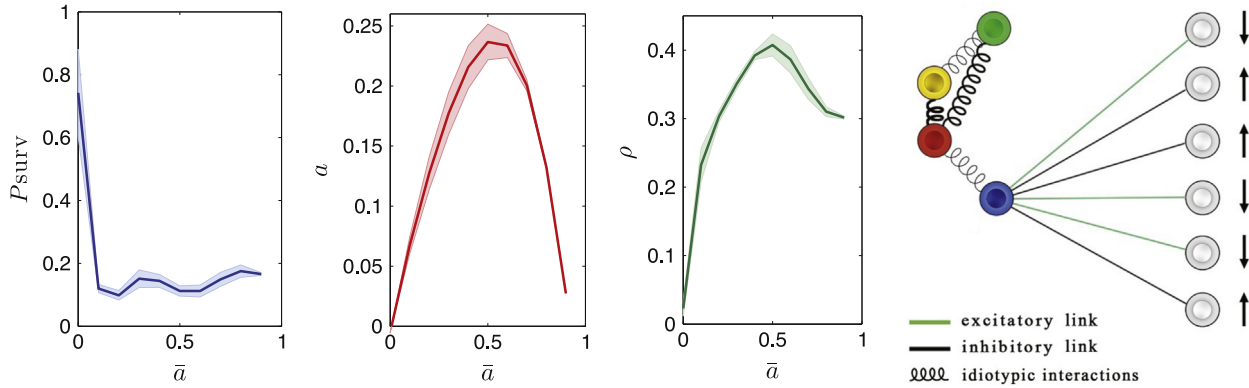

Fig. 3. From left to right: the three panels show plots from simulations where we generated random strings Ψ and we compared them with those in $\mathcal{S}$ which are generated according to the distribution of Eq. (16). Strings Ψ fulfilling the condition (17) are retained and their survival probability P_{surv} is measured and plotted versus $\bar{a}$ (left panel). The final repertoire $\mathcal{B}$ also proves to be biased with degree a depending on $\bar{a}$ (central panel). Moreover, we measured the Spearman correlation coefficient ρ, averaged over $\mathcal{B}$, between W_μ and $\max_{\Phi \in \mathcal{S}} \{\chi(\Psi_\mu, \Phi)\}$ (right panel). Note that a positive value denotes the existence of correlation and gives strong numerical evidence for Varela's theory. Data represented in these plots refer to a system where we fixed the size of the B-repertoire $N_B = 10^5$ and $\gamma = 2$, $c = 0.5$, $\Delta = 0.4$, $\chi_P = 0.6L$ (see Appendix Four of the Supplementary Information file for more details). Data were averaged over 10^3 realizations. The rightmost figure represents a portion of the B–T network: colored and white circles represent B and T cells, respectively; green and black lines mean interactions via cytokines, while springs of different size mean idiotypic coupling of different magnitude. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

assumption underlying Eq. (16), we expect that no fine structure for $\mathcal{B}$ is captured. A possible way to include this feature is to incorporate an ultrametric structure (Mezard et al., 1987) to obtain a tree-like organization for the repertoire. This would implicitly define a hierarchy of antibodies similar to the "cognitive paradigm" (which actually shares several similarities with our approach), discussed for instance by Cohen (1992a, 2007), which suggests the existence of "dominant self-antigens" able to properly channel the immune response.

Here, following our assumptions we analyze the final repertoire and, as anticipated, we find that the occurrence of string entries is not completely random, but is compatible with a biased distribution such as

$$P_{rep}(\Psi_i^\mu | a) = \delta(\Psi_i^\mu + 1)\frac{1 + a}{2} + \delta(\Psi_i^\mu - 1)\frac{1 - a}{2}, \qquad (18)$$

where a proves to be correlated with $\bar{a}$.

More precisely, positive values of $\bar{a}$ (i.e. corresponding to strings with more entries equal to 0) yield a biased mature repertoire with $a > 0$ (see Fig. 4, central panel) as this ensures that a self-string and a string from the repertoire $\mathcal{B}$ typically display a low degree of complementarity and therefore no auto-reactions are expected. Consequently, in the set $\mathcal{B}$ generated in this way, nodes with larger W_μ are those carrying strings more dissimilar with respect to the typical one, namely those displaying a relative large number of entries equal to 1. Such strings are also likely to exhibit large affinity with the self-repertoire.

Otherwise stated, due to negative selection, the $\mathcal{B}$ repertoire displays a bias concordant with the one of the self-repertoire $\mathcal{S}$, for otherwise two arbitrary strings from $\mathcal{B}$ and $\mathcal{S}$, respectively, could interact strongly. Now, as a consequence of the bias and the fact

that interactions are based on complementarity, the set $\mathcal{B}$ presents "hubs" because nodes associated to strings which deviates from the typical biased pattern are expected to interact with the whole bulk as well as with the self-repertoire. This picture is in (qualitative) agreement with recent findings (Madi et al., 2011).

To corroborate this fact we measured the correlation ρ between the weighted degree W_μ of a node and the affinity $\max_{\Phi \in \mathcal{S}} \{\chi(\Psi_\mu, \Phi)\}$ with the self-repertoire finding a positive correlation (see Fig. 3, right panel). We also checked the response of the B-repertoire when antigens are presented, finding that, when a string $\Phi_\nu \in \mathcal{S}$ is taken as antigen, the best-matching node, displaying large W, needs an (exponentially) stronger signal on BCR in order to react.

Such results mirror the theory of Varela (Stewart et al., 1989; Varela and Coutinho, 1991), according to which "self-directed" nodes display a high (weighted) connectivity, which, in turn, induces inhibition. Moreover, the distinction between self-addressed and non-self-addressed is not dichotomic but rather continuous, consistently with Cohen (1992b).

Finally, it is worth underlining that, by taking a biased distribution for string entries (i.e., $a \neq 0$), the distribution $P(W)$ for weights occurring in the idiotypic network still retains its logarithmic shape, namely

$$P(W) = \frac{1}{W\sqrt{2\pi}\sigma} e^{-(\log W - \mu)^2 / 2\sigma^2}, \qquad (19)$$

with

$$\mu = \log \left[\frac{N_B \langle J \rangle_a^2}{\sqrt{\langle J \rangle_a^2 + (\langle J^2 \rangle_a - \langle J \rangle_a^2)/N_B}} \right], \qquad (20)$$
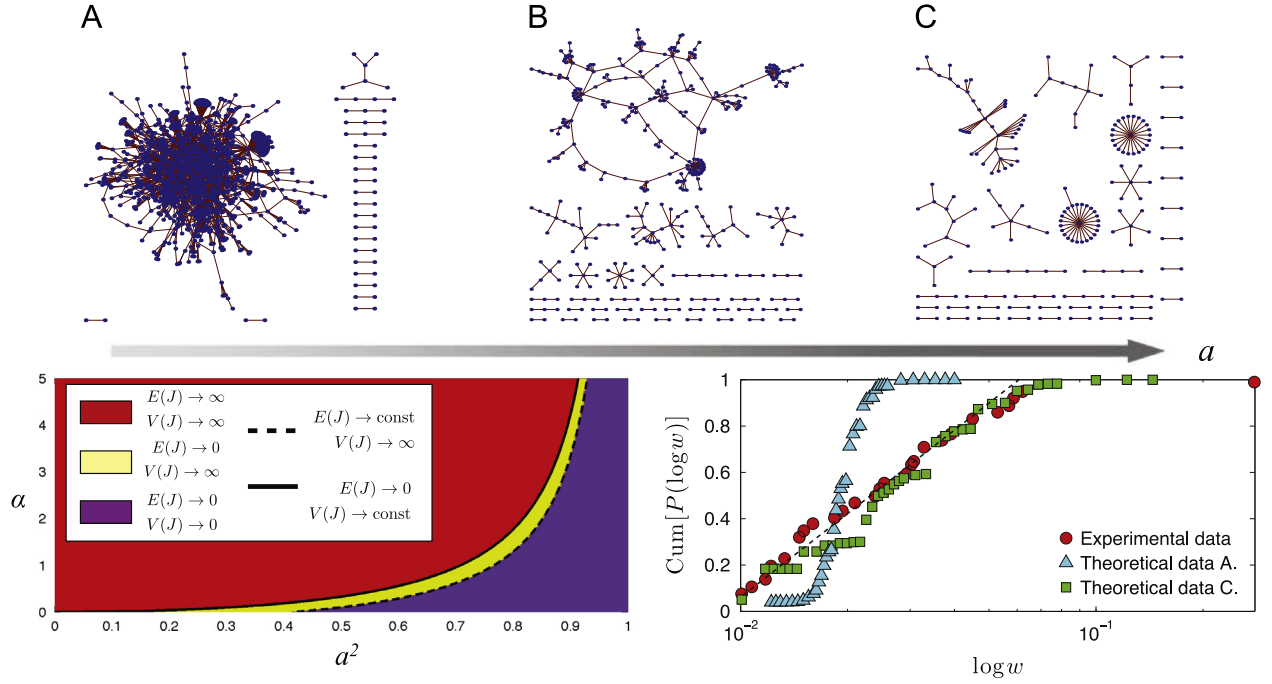
Fig. 4. Upper panels: three examples of idiotypic networks with decreasing connectivity (from A to C), referring to the particular choice $\alpha = 0.7$, $N = 10^4$ and to different values of a (see also Agliari et al., 2012b). Lower panels: in the left plot a phase diagram depicting the distribution of the interaction strengths of the idiotypic network is shown. More precisely, $\gamma = 2$ being fixed, we highlight different regions of the (a^2, α) plane, each corresponding to a different behavior of the average coupling $E(J) = \langle J \rangle_a$ and of the variance $V(J) = \langle J^2 \rangle_a - \langle J \rangle_a^2$, as explained by legend. Different behaviors of $E(J)$ and $V(J)$ can be recast into different topological regimes as visualized by the graphs depicted in the lower panels, representing particular realizations of the idiotypic network. In the right plot the cumulative of $P(W)$ is shown for the upper panel network corresponding to A (triangles) and C (squares), while circles represent real data on mice performed via ELISA technology in Carneiro et al. (1996a, 1996b). Note that while an overpercolated network (case A △) – whose existence has been criticized over the years – is in contrast with data from mice (●), an under-percolate network (case C □) made up of motifs commonly seen in experiments (Cazenave, 1977) displays remarkable agreement.

$$\sigma^2 = \log\left[\frac{\langle J^2 \rangle_a - N_B \langle J \rangle_a^2}{N_B \langle J \rangle_a^2} + 2\right], \tag{21}$$

where $\langle J \rangle_a$ and $\langle J^2 \rangle_a$ are, respectively, the mean value and the mean squared value of coupling $J_{\mu\nu}$ defined in Eq. (2). A detailed derivation of these values can be found in the Appendix Five. Here we simply notice that, by properly tuning a and α, one can recover, in the thermodynamic limit, different regimes characterized by different behaviors (finite, vanishing or diverging) for the average $E(J) \equiv \langle J \rangle_a$ and the variance $V(J) \equiv \langle J^2 \rangle_a - \langle J \rangle_a^2$, respectively, as reported in Fig. 4.

4. Discussion

In this paper we have modeled a subset of the adaptive response of the immune system by means of statistical mechanics. In particular, we have focused on the emergent properties of the interacting lymphocytes starting from minimal assumptions on their local exchanges and, as a fine test, we have searched for the emergence of subtle possible features such as the anergy shown by self-directed B-cells.

First, we reviewed and framed into a statistical mechanics description, the two main strands for its explanation, i.e. the cognate response model and the idiotypic network. For this task we described the mutual interaction between B cells and (helper and suppressor) T cells as a bi-partite spin glass, and we showed its thermodynamical equivalence to an associative network made up of T cells (helpers and suppressors) only. Then, the latter is shown to properly orchestrate the response of B cells as long as their connection within the bulk of the idiotypic network is rather small. In the second part we adopted an information theory perspective to infer that highly connected B clones are typically

self-directed as a natural consequence of learning during ontogeny.

By merging these results we get that helpers are always able to signal non-self-B lymphocytes, in such a way that the latter can become activated, proliferate and produce antibodies to fight against non-self-antigens. On the other hand, self-lymphocytes, due to their large connectivity within the idiotypic network, do not sense the signal sent by helpers.

Therefore, a robust and unified framework where the two approaches act synergically is achieved and in particular Varela theory is finally absorbed in the two-signal model. Interestingly, this picture ultimately stems from a biased learning process at ontogenesis and offers, as a sideline, even a theoretical backbone to the theory of Varela. We stress that, while certainly Jerne's interactions among B cells act as a key ingredient (and the existence of anti-antibodies or small reticular motifs has been largely documented), an over-percolated B network is not actually required as the distribution of the weighted clonal connectivity remains broad even for extremely diluted regimes (far beyond the percolation threshold). This point may deserve further attention because, from a mathematical and physical perspective, at the time both Jerne and Varela suggested a possible role for an underlying network of chemical exchanges, theories were developed regarding only overpercolated networks (which do not work for B-cell interactions), while only recently a satisfactory picture of underpercolated networks is finally available: Note that outcomes from both the over-percolated and under-percolated networks have been shown in Fig. 4 and only the underpercolated is in agreement with experimental support.

Furthermore, we stress that, within our approach, while the theory of Varela is reabsorbed into the model of the cognate response, the converse is not true as clearly other cells (beyond highly connected ones in the B-repertoire), through other paths,

may lack helper signaling for other motivations and still become anergic. Hence the cognate-response model is not necessarily reabsorbed into the theory of Varela. For instance the entire hierarchical organization of antigens underlying the theory of immunological homunculus cannot be resolved within our replica-symmetric framework (and we plan to develop a full replica symmetry breaking scenario, where these other features may spontaneously appear, in future works).

Still, the model developed is able to reproduce several other aspects of real immune networks beyond showing anergy as a spontaneous phenomenon, such as the breakdown of immuno-surveillance by unbalancing the leukocitary formula (the ratio N_B/N_T), the low-dose tolerance phenomenon, the link between lymphocytosis and autoimmunity (for instance as is well documented in the case of A.L.P.S. Agliari et al., 2011), the capability of the system to simultaneously cope with several antigen (Agliari et al., 2013c, 2012a) acting as an autonomous parallel processor and, last but not least, it shows also a remarkable scaling agreement with experimental data on network connectivity.

Despite these achievements, several assumptions underlying this minimal model could be relaxed or improved in future developments, ranging from the symmetry of the interactions to the fully connected topology of the B–T interactions (whose investigation has already started in Agliari et al., 2013a, 2013b).

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this paper can be found in the online version at http://dx.doi.org/10.1016/j.jtbi.2014.05.006.

References

Agliari, E., et al., 2013a. Immune networks: multitasking capabilities at medium load. J. Phys. A 46 (33), 335101.

Agliari, E., et al., 2013b. Immune networks: multitasking capabilities near saturation. J. Phys. A 46 (41), 415003.

Agliari, E., Barra, A., Guerra, F., Moauro, F., 2011. A thermodynamical perspective of immune capabilities. J. Theor. Biol. 287, 48–63.

Agliari, E., Barra, A., Galluzzi, A., Guerra, F., Moauro, F., 2012a. Multitasking associative networks. Phys. Rev. Lett. 109, 268101.

Agliari, E., Asti, L., Barra, A., Ferrucci, L., 2012b. Organization and evolution of synthetic idiotypic networks. Phys. Rev. E 85, 051909.

Agliari, E., Barra, A., Bartolucci, S., Galluzzi, A., Guerra, F., Moauro, F., 2013c. Parallel processing in immune networks. Phys. Rev. E 87, 042701.

Allen, M.P., Tildesley, D.J., 1987. Computer Simulation of Liquids. Oxford Press, Oxford, UK.

Allman, D.M., Ferguson, S.E., Lentz, V.M., Cancro, M.P., 1993. Peripheral B cell maturation. J. Immunol. 151, 4431.

Amit, D.J., 1992. Modeling Brain Function: The World of Attractor Neural Network. Cambridge University Press, Cambridge, UK.

Barra, A., Agliari, E., 2010a. A statistical mechanics approach to autopoietic immune networks. J. Stat. Mech. 07004.

Barra, A., Agliari, E., 2010b. Stochastic dynamics for idiotypic immune networks. Physica A 389, 5903–5911.

Barra, A., Bernacchia, A., Contucci, P., Santucci, E., 2012. On the equivalence of Hopfield networks and Boltzmann machines. Neural Netw. 34, 1–9.

Bernabé, R.R., Coutinho, A., Cazenave, P.A., Forni, L., 1981. Suppression of a "recurrent" idiotype results in profound alterations of the whole B-cell compartment. Proc. Natl. Acad. Sci. 78 (10), 6416–6420.

Bouchaud, J.P., Potters, M., 2000. Theory of Financial Risk and Derivative Pricing. Cambridge University Press, Cambridge, UK.

Brede, M., Behn, U., 2001. Architecture of idiotypic networks: percolation and scaling behavior. Phys. Rev. E 64, 011908.

Brede, M., Behn, U., 2003. Patterns in randomly evolving networks: idiotypic networks. Phys. Rev. E 67, 031920.

Brock, W.A., Durlauf, S.N., 2001. Discrete choice with social interactions. Rev. Econ. Stud. 68, 235.

Burnet, F.M., 1959. The Clonal Selection Theory of Acquired Immunity. Cambridge University Press, London.

Burroughs, N.J., De Boer, R.J., Kesmir, C., 2004. Discriminating self from nonself with short peptides from large proteomes. Immunogenetics 56, 311–320.

Carneiro, J., Coutinho, A., Stewart, J., 1996a. A model of the immune network with B–T cell co-operation. The simulation of ontogenesis. J. Theor. Biol. 182 (4), 531.

Carneiro, J., Coutinho, A., Stewart, J., 1996b. A model of the immune network with B–T cell co-operation. Prototypical structures and dynamics. J. Theor. Biol. 182 (4), 531.

Cazenave, P.A., 1977. Idiotypic-anti-idiotypic regulation of antibody synthesis in rabbits. Proc. Natl. Acad. Sci. 74 (11), 5122–5125.

Chakraborty, A.K., Košmrlj, A., 2011. Statistical mechanics concepts in immunology. Annu. Rev. Phys. Chem. 61, 283–303.

Cohen, I.R., 1992a. The cognitive paradigm and the immunological homunculus. Immunol. Today 13 (12), 490.

Cohen, I.R., 1992b. The cognitive principle challenges clonal selection. Immunol. Today 13, 441–444.

Cohen, I.R., 2007. Real and artificial immune systems: computing the state of the body. Nat. Rev. Immunol. 7, 569–574.

Coolen, A.C.C., 2005. The Mathematical Theory of Minority Games – Statistical Mechanics of Interacting Agents. Oxford Press, Oxford, UK.

Coolen, A.C.C., Kuehn, R., Sollich, P., 2005. Theory of Neural Information Processing Systems. Oxford Press, Oxford, UK.

Durlauf, S.N., 1999. How can statistical mechanics contribute to social science? Proc. Natl. Acad. Sci. 96 (19), 10582.

Depino, A.M., 2010. Pro-Inflammatory Cytokines in Learning and Memory. Nova Science Publishers, Hauppauge, NY, USA.

Dreyer, W.J., Bennett, J.C., 1965. The molecular bases of antibody formation. Proc. Natl. Acad. Sci. 54, 864.

Efroni, S., Harel, D., Cohen, I.R., 2003. Toward rigorous comprehension of biological complexity: modeling, execution and visualization of thymic T-cell maturation. Genome Res. 13, 2485–2497.

Efroni, S., Harel, D., Cohen, I.R., 2007. Emergent analysis of the thymocyte development and lineage determination. PLoS Comput. Biol. 3 (1), e13.

Ellis, R.S., 1985. Entropy, Large Deviations and Statistical Mechanics. Springer-Verlag, New York, USA.

Floreano, D., Mattiussi, C., 2008. Bio-Inspired Artificial Intelligence: Theories, Methods, and Technologies. The MIT Press, Cambridge, MA, USA.

Frenkel, D., Smith, B., 2002. Understanding Molecular Simulation. Academic Press, London, UK.

Germain, R.N., 2001. The art of probable: system control in the adaptive immune system. Science 293, 240–245.

Goodnow, C.C., 1992. Transgenic mice and analysis of B-cell tolerance. Annu. Rev. Immunol. 10, 489–518.

Goodnow, C.C., 2005. Cellular and genetic mechanisms of self tolerance and autoimmunity. Nature 435, 590.

Goodnow, C.C., Vinuesa, C.G., Randall, K.L., Mackay, F., Brink, R., 2010. Control systems and decision making for antibody production. Nat. Immun. 8 (11), 681.

Hartley, S.B., Crosbie, J., Brink, R., Kantor, A.B., Basten, A., Goodnow, C.C., 1991. Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens. Nat. Lett. 353 (24), 765.

Hopfield, J.J., Tank, D.W., 1987. Computing with neural circuits: a model. Science 233 (4764), 625.

Janeway, C., Travers, P., Walport, M., Shlomchik, M., 2005. Immunobiology. Garland Science Publisher, New York.

Jaynes, E.T., 1957a. Information theory and statistical mechanics. Phys. Rev. E 106, 620.

Jaynes, E.T., 1957b. Information theory and statistical mechanics. II. Phys. Rev. E 108, 171.

Jerne, N.K., 1974. Towards a network theory of the immune system. Ann. Immunol. 125C, 373–389.

Kaufman, S.A., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. 22, 437.

Kitamura, D., 2008. Self–Nonself Recognition Through B-cell Antigen Receptor, How the Immune System Recognizes Self and Non Self. Springer Press, Noda, Japan.

Košmrlj, A., Jha, A.K., Huseby, E.S., Kardar, M., Chakraborty, A.K., 2008. How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. Proc. Natl. Acad. Sci. 105 (43), 16671–16676.

Košmrlj, A., Chakraborty, A.K., Kardar, M., Shakhnovich, E.I., 2009. Thymic selection of T-cell receptors as an extreme value problem. Phys. Rev. Lett. 103, 068103.

Kuchroo, V.K., Sarvetnick, N., Hafler, D.A., Nicholson, L.B. (Eds.), 2002. Cytokine and Autoimmune Diseases. Humana Press, Amsterdam, Netherlands.

Lider, O., Reshef, T., Beraud, E., Ben-Nun, A., Cohen, I.R., 1988. Anti-idiotypic network induced by T cell vaccination against experimental autoimmune encephalomielitis. Science 239, 181–183.

Lundkvist, I., Coutinho, A., Varela, F., Holmberg, D., 1989. Evidence for a functional idiotypic network among natural antibodies in normal mice. Proc. Natl. Acad. Sci. 86, 5074–5078.

Madi, A., et al., 2011. Network theory analysis of antibody-antigen reactivity data: the immune trees at birth and adulthood. PLoS-One 8:6 (3), e17445.

Martelli, C., De Martino, A., Marinari, E., Marsili, M., Perez-Castillo, I., 2009. Identifying essential genes in Escherichia coli from a metabolic optimization principle. Proc. Natl. Acad. Sci. 106 (8), 2607.

Mezard, M., Montanari, A., 2007. Information Physics and Computation. Cambridge Press, Oxford, UK.

Mezard, M., Parisi, G., Virasoro, M.A., 1987. Spin glass theory and beyond. In: Lecture Notes in Physics, vol. 9. World Scientific, Singapore.

Mora, T., Walczak, A.M., Bialek, W., Callan, C.G., 2010. Maximum entropy models for antibody diversity. Proc. Natl. Acad. Sci. 107 (12), 5405–5410.

Murugan, A., Mora, T., Walczakc, A.M., Callan, C.G., 2012. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. Proc. Natl. Acad. Sci. 17 (early edition September).

Parisi, G., 1990. A simple model for the immune network. Proc. Natl. Acad. Sci. 87, 429–433.

Perelson, A., 1997. Immunology for physicists. Rev. Mod. Phys. 69, 1219.

Rabello, S., Coolen, A.C.C., Pérez-Vicente, C.J., Fraternali, F., 2008. A solvable model of the genesis of amino-acid sequences via coupled dynamics of folding and slow genetic variation. J. Phys. A 41, 285004.

Rolink, A.G., Andersson, J., Melchers, G., 1998. Characterization of immature B cells by a novel monoclonal antibody, by turnover and by mitogen reactivity. Eur. J. Immunol. 28, 3738.

Schmidtchen, H., Thune, M., Behn, U., 2012. Randomly evolving idiotypic networks: structural properties and architecture. Phys. Rev. E 86, 011930.

Schwartz, R.H., 2005. Natural regulatory T cells and self-tolerance. Nat. Immun. 4 (6), 327.

Shoenfeld, Y., 2004. The idiotypic network in autoimmunity: antibodies that bind antibodies that bind antibodies. Nat. Med. 10, 17–18.

Stewart, J., Varela, F., Coutinho, A., 1989. The relationship between connectivity and tolerance as revealed by computer simulation of the immune network: some lessons for an understanding of autoimmunity. J. Autoimmun. 2, 15–23.

Theze, J., 1999. The Cytokine Network and Immune Functions. Oxford University Press, Oxford, UK.

Tuckwell, H.C., 2005. Introduction to Theoretical Neurobiology. Cambridge University Press, New York, USA.

Urbain, J., Wikler, M., Franssen, J.D., Collignon, C., 1977. Idiotypic regulation of the immune system by the induction of antibodies against anti-idiotypic antibodies. Proc. Natl. Acad. Sci. 74, 5126–5130.

Varela, F., Coutinho, A., 1991. Second generation immune networks. Immunol. Today 2, 159–166.

Vertosick, F.T., Kelly, R.H., 1989. Immune network theory: a role for parallel distributed processing. Immunology 66, 1–7.

Wardemann, H., Yurasov, S., Schaefer, A., et al., 2003. Predominant autoantibody production by early human B cell precursors. Science 301, 1374–1377.