bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Network organization of antibody interactions in sequence and structure space: the RADARS model

József Prechl

Diagnosticum zrt, Attila u. 126, Budapest, Hungary

Correspondence: jprechl@gmail.com

Keywords: antibody; network; sequence; structure; clonality; B cell; systems biology; quantitative biology; immunodominance; consecutive reactions; energy partition

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Adaptive immunity in vertebrates represents a complex self-organizing network of protein interactions that develops throughout the lifetime of an individual. While deep sequencing of the immune-receptor repertoire may reveal clonal relationships, functional interpretation of such data is hampered by the inherent limitations of converting sequence to structure to function.

In this paper a novel model of antibody interaction space and network, termed radial adjustment of system resolution, RADARS, is proposed. The model is based on the radial growth of interaction affinity of antibodies towards an infinity of directions in structure space, each direction representing particular shapes of antigen epitopes. Levels of interaction affinity appear as free energy shells of the system, where hierarchical B-cell development and differentiation takes place. Equilibrium in this immunological thermodynamic system can be described by a power-law distribution of antibody free energies with an ideal network degree exponent of phi square, representing a scale-free fractal network of antibody interactions. Plasma cells are network hubs, memory B cells are nodes with intermediate degrees and B1 cells represent nodes with minimal degree.

Thus, the RADARS model implies that antibody structure space develops against an infinite antigen structure space via interactions that are individually immunologically controlled, but on a systems level are organized by thermodynamic probability distributions. The network of interactions, which control B-cell development and differentiation, represent pathways of antigen removal on systems level. Understanding such quantitative network properties of the system should help the organization of sequence-derived structural data, offering the possibility to relate sequence to function in a complex, self-organizing biological system.

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1. Introduction

Appearance of complex multicellular life was accompanied by the evolution of a system that maintains cellular and molecular integrity in the host organism (1). The adaptive immune system is a complex system in the physical sense, being composed of a vast number of cells that engage in interactions, self-organize and - most impressively - adapt to the molecular and cellular environment. Its mere size, with cell numbers in the range of 10^{11}, suggests that system could be described by statistical properties. In fact, the host is more than an organism: a supraorganism (2,3) with microbial communities, and immunity maintains a continuity of interactions in a wide range of energy scale, rather than simply discriminating self from non-self (4). Technological advances now allow us to measure and characterize this complexity in ever growing details, at the gene, transcript, protein and cellular levels, driving the field of systems immunology (5). The vast amount of data generated requires not only data storage and analysis capacity, but also theoretical frameworks, models that simplify data organization and systems level interpretation.

Humoral adaptive immunity comprises the cells and mechanisms that lead to the production of antibodies. In human adults B-cells develop in the bone marrow throughout life and build up a system of effector and memory cells, which accumulate with a lifetime of immunological experiences. Continuously emerging naive B cells only differentiate further if selected for immunological actions based on their B-cell antigen receptor (BCR) specificity. The primary role of B cells and antibodies is the regulation of antigen removal and thereby adjustment of antigen concentrations in the host, operating over several orders of magnitude (6). Because this specificity is genetically coded in the individually rearranged and mutated immunoglobulin heavy and light chain sequences, it is possible to capture the antibody repertoire in a given sample of B cells. Deep sequencing or next generation sequencing (NGS) is capable of generating sequence data of antibody repertoires with varying resolution and length (7–11).

It is also possible to profile the antibody repertoire functionally, based on the identification of antibodies binding to huge sets of potential targets (12,13). This approach is biased by the fact that a priori knowledge of targets is not always possible and only those antibodies that bind to the tested antigens are identified. Antigen microarray assays are useful for the focused analysis of antibodies related to allergy, autoimmunity, infection or cancer (14–18). Such functional analyses provide a more meaningful profile in the immunological sense and, if carried out from blood, it is less prone to sampling error than cell-based sequencing approaches.

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The relationship between antibody sequence and structure is on one hand like that of proteins in general: polypeptide chains of a given sequence fold into structures, which are responsible for function. In the enormous sequence space allowed by permutating amino acids only the thermodynamically stable structures materialize as proteins (19). Proteins capable of interacting with molecules in a way that improves chances of survival of the host organism will themselves survive and evolve. Unlike proteins in general, antibodies evolve within the lifetime of the host. While thermodynamic constraints still hold, their "survival", meaning the producing cell clone being selected into long-lived B-cell populations, is determined by interactions with self and foreign molecules, the antigens. Importantly, because there are so many more sequences than structures and because changing a few critical amino acids can result in different structures, mapping sequence space to structure space in far from trivial. The combined length of the complementarity determining regions (CDR) of heavy and light immunoglobulin chains, which forms the antigen binding site, is around 14-50 amino acids (IMGT definition (20)), which allows extensive diversification. By employing screening and selection mechanisms, coupled with cycles of random mutagenesis, targeting primarily these amino acids, the immune system is capable of developing high-affinity antibodies against most targets. Understanding these processes on the systems level preferably requires the prediction of structures from NGS data (21) because of the complex sequence-to-space relationship, as noted above.

The architecture and functioning of complex systems can be assessed by network science, which in the case of antibodies identifies antibody-antigen interaction networks (22). The development of concepts of the immune system as a network were key steps in our current perception of immunity (23,24). Efforts are now under way to describe the immune system as a network (termed network systems immunology) using NGS data and network science (25). Since the system is organized by structure rather than sequence, the conceptualization of an antibody interaction network based on physical properties should help better definition of the system.

In this paper, following a brief introduction to the sequence space of antibodies, a model for the molecular organization of antibody structure space or interaction space is proposed. The model builds on the generalized quantitative model of antibody homeostasis (6,26,27), thus approaches antibody function from the physico-chemical perspective: antibodies are organized into a network based on their binding affinity to cognate target. The model also considers the architecture of B-cell development and hierarchy and provides a power law-based quantitative network description of the humoral immune system at rest.

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2. Antibody clonal representation in sequence space

Sequence space in the biological sense is a theoretical space comprising collections of nucleic acid or protein sequences of interest. We usually talk about protein sequence space and define what protein sets are involved (proteome of a given species, cells, etc.) and whether any restrictions hold (fully random, functional, identified, etc.). An amino acid sequence with a given length 'd' and full randomization with 20 amino acids occupies a sequence space with size 20^d (Fig 1A). An exact sequence with no ambiguity defines an exact position in sequence space; moves in this space are discrete steps along a given dimension. Exponential growth is incredibly fast, leading to the generation of vast amounts of sequence space in high dimensions. It is accepted that only a fraction of all theoretically possible sequences are thermodynamically stable and protein evolution can be interpreted as a search for acceptable and functional structures in sequence and structure space (19). Thinking along these lines, the evolution of antibody binding surface, the paratope, is a search for the thermodynamically stable sequences and the selection from among these the ones meeting immunological criteria for B-cell survival. The set of viable antibody sequences, functional antibody sequence space, lies much below the theoretically possible (28) and close to the already observed and annotated antibody sequence space (29).

Collections of antibody protein sequences obtained by translating DNA or RNA of deep sequencing data ideally span the whole variable domain of heavy (VH) and light chains (VL) and can also pair these two. In such a case the gene segments contributing to the rearrangement of VH and VL can be predicted and visualized in 3D and 2D respectively, as shown in Figure 1. A repertoire can be represented by identifying coordinates of rearrangements identified, and symbol size or color can represent segment frequencies (30). While the use of gene segments for classification allows tremendous reduction in dimensionality, it is not best suited for functional network analysis, where the use of complete rearranged and mutated sequences is preferable (31).

In a much simpler approach, heavy chain CDR3 regions only are used as an estimate of diversity. Though this region is often regarded as being most important for determining binding specificity, identical H-CDR3 sequences have been found to be present in functionally unrelated cells and therefore H-CDR3 seems insufficient for functional classification (32). Selection of the pre-BCR bearing cells depends on signals that may be triggered by ubiquitous ligands present in the bone marrow microenvironment. The presence of uniform reactivity against such common public self-antigens may lead to the positive selection of CDR3 with

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

similar binding properties, and thereby similar sequences. Sequencing of the complete heavy chain variable domains can be readily used to follow changes in repertoire size and diversity during B-cell development and response to immunization (33).

Whatever the depth and methodology, sequence similarity relationships can be used for the construction of family trees, often displayed in circular forms. These trees usually start classification with the V segment, clustering clones with common V use (34). While this approach may be useful for classification, the use of the complete VDJ-H sequence as a first stage classifier, followed by VJ-L use better reflects the natural development scheme of B cells (Fig 1C). Antibody repertoire sequencing now follows guidelines to help the integration of data (35–37), several tools and databases devoted especially for these data have been established (8,10,38–40).

3. Antibody interaction representation in structure space

In contrast to this graded qualitative scheme, which may well serve the purpose of tracking peripheral clonal expansions accompanied by affinity maturation, a quantitative scheme should place genetic changes into structure rather than sequence space. Furthermore, because it is not just antibody structure but also the availability of targets and the structure of those targets that determine the development of antibody repertoire and the architecture of the network, we shall talk about interaction space, as explained below.

3.1. Structural resolution of molecular recognition as a measure of interaction strength

While sequence can be defined with various levels of certainty of an amino acid occupying a given position in the sequence, molecular structure can be defined at various levels of resolution. As we are talking about antibody molecules, structural resolution is on the atomic scale, crystal structures define atomic coordinates on the Ängstrom scale. The binding site of an antibody can also be characterized by the surface area that comes into close contact with the antigen (41,42). Water molecules are displaced from this area as a function of the goodness of fit. The so-called buried surface area (BSA) is therefore a predictor of binding energy of protein interactions (43,44). Another measure of goodness of fit is the decrease of free energy of the antibody molecule upon binding. All these approaches are correlated: higher resolution "description" of a structure by the antibody corresponds to greater BSA and to a higher binding energy. In other words, the resolution of molecular recognition is the goodness of fit in terms of number and strength of non-covalent bonds forming between antibody and target and can be expressed as standard free energy change $\Delta G°$ or as equilibrium constant K of binding. These two are related by the equation

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$\Delta G° = -RT\ln K$ 1)

alternatively

$K = \exp(-\Delta G°/RT)$ 2)

where R is the product of kB, the Boltzmann constant, and NA, the Avogadro number, and T is thermodynamic temperature in Kelvins. The Boltzmann constant serves to relate the energy of particles to temperature, NA expresses energy per mole particles. Considering that the recognition of a given antigen Ag_i is changing in the system by the adjustment of Ab fit, we can characterize this maturation of affinity by examining the free energy level of the antibody, such as

$\Delta G° = F_b - F_f$ 3)

where F_b is the free energy of the bound (also called native) and F_f is the free (non-native) form of the antibody; $\Delta G°$ is the Gibbs free energy of formation of 1 mole antibody under standard conditions.

The advantage of using thermodynamic description for the characterization of structural resolution is that it conveys the sense of function: higher binding energy means higher affinity of antibody to target, which in turn means more efficient clearance (6). Besides defining resolution of molecular recognition, which is a general descriptor, the identification of a given interaction requires the description of target shape, a distinct molecular structure. The higher the resolution, the more information is required for defining shape, corresponding to higher interaction energy and a better fit between antibody and target. This fit can be characterized by the number of bonds forming between pairs of atoms, which we shall denote by "r", and the fluctuation in energy of coupling ΔE. The statistical distribution of affinity, based on random energy models (45,46) and protein folding models (47), that are also used for modeling BCR energy landscape (48), is a normal distribution of free energy ΔG around its mean $\langle \Delta G \rangle$ (49), according to

$f(\Delta G) \sim \exp(-(\Delta G - \langle \Delta G \rangle)^2/2r\Delta E^2)$ 4)

By increasing resolution we shall be able to distinguish between different shapes, the higher the resolution the more shapes becoming distinct. Because of the structural complexity of an antibody binding surface, the distinction between all possible shapes at high resolution would require a multidimensional space. Let us gradually generate an interaction space by considering a point of origin, the center of the system, from which a particular direction represents a particular shape. In this representation the extent by which we leave the point of origin corresponds to the resolution at which we can define the direction. Thus, going away from the

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

minimal resolution we can define shape at gradually higher resolutions, corresponding to larger free energy decrease of the interacting molecule (Fig. 2A,B). Different levels of resolution, that is different levels of binding energies appear in our scheme as shells of a sphere. Theoretically the number of directions originating from a single point is infinite, so the shapes available in this representation could also be infinite. Practically, considering a reversible interaction, the resolution is limited by the binding energy of reversible interactions and system size, that is the number of B cells available.

This model of the organization of interactions of a system we shall call 'RAdial ADjustment of System Resolution' or RADARS in short. The abbreviation intentionally reminds of radiolocation, where emitted electromagnetic waves interact with objects in their way and are reflected to provide an image of the surroundings. The RADARS model implies that elements of the growing system interact with the surroundings, gaining information and adjusting system growth accordingly.

3.2. B-cell development in interaction space

Immunological interpretation of the model requires us to fit B-cell development and antibody network growth into this interaction space. We shall assume that a common lymphoid progenitor (CLP) has the potential to generate any and all functional VDJ-VJ sequences and therefore to produce via sequential differentiation and maturations steps antibody against any and all targets. By functional VDJ-VJ sequences we mean all sequences that are physically and biologically viable. This means thermodynamic stability (able to fold into a structure compatible with the Ig domain), ability to pair, forming a VH-VL functional binding unit, and ability to sustain a B-cell via delivering survival, differentiation and proliferation signals (26,50). Of note, the value of r is 0 for a CLP and therefore the probability of interactions is 0, which confirms its position at the core of the system.

A differentiation step increases this total potential but immediately introduces restrictions in shape space because of its direction. This will appear as a step towards increased resolution of cognate target structure recognition. Expression of the surrogate light chain (SLC) marks the first step towards BCR formation. These pro-B cells represent the founders of all B cells (Fig. 2C). While signaling via the SLC may be required, it is uncertain whether binding of SLC is required for further differentiation, therefore we assume that these cells seed the complete antibody interaction space. Rearrangement of the heavy chain introduces a structural restriction: a particular functional heavy chain variable domain (VH) sequence has a limited range of targets. Pre-B cells displaying the pre-BCR composed of VH-SLC pairs will divide until and as long as ligands (antigens) are available. Cells with different VH sequences will

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

populate the interaction space and share this space according to the availability and direction of target. Cells with more abundant targets expand more, cells with less frequent targets remain in lower numbers, until optimal BCR engagement is achieved (26). As a result, interaction space as represented at these energies will be filled with different pre-B cell clones according to the availability of the common self-antigens.

The next levels of interaction resolution, introducing further focusing in interaction space, comes with the rearrangement of the light chain. Individual pre-B cells rearrange their light chains independently and randomly. Therefore, all pre-B cells reserve a particular area on the next level of interaction energy. The size of this area again will correspond to the nature of the rearranged VL domain, with those finding more available targets expanding more. The pool of immature B cells fills thus the outer energy level in the bone marrow (Fig. 2C).

Taking a somewhat unique route of differentiation are the B1 cells. These cells seem to generate antibodies that keep B1 cells in a continuous state of low-level activation. This may reflect their ability to respond to soluble, highly abundant antigens (26), or an intrinsic ability of the BCR of sustained signaling, perhaps due to structural properties (51). In any case, B1 cells represent a stable population with the highest affinity towards self and non-self, which is achieved without affinity maturation. Meanwhile B2 cells are continuously generated but die in a few days unless recruited as effector cells for further differentiation (Fig. 2C).

As implied above, selection in the bone marrow is a passive process: randomly generated sequences find their positions in interaction space, expanding or dying according to the availability of interacting target structures. As a result, the emerging population of immature B cells will bear the low-resolution antigenic signature of the bone marrow environment. This can be interpreted both as deletion of highly self-reactive clones to prevent autoimmunity (50), and as selection for mildly autoreactive clones to help homeostatic antibody functions and setting a reference for recognition of non-self (52).

3.3. Immune responses in interaction space

The development of cells reacting with antigens that are only temporarily present in the host presents the risk of investing energy into clones that will become useless once that antigen disappears. Therefore, such clonal expansions beyond the border of self only take place when 2nd signals inform the host of danger. This is the development of an immune response, aided by various molecular and cellular help signals. Thymus independent and primary thymus dependent responses expand populations of B cells without improving their affinity, thus keeping them on the same level in the interaction space. Thymus dependent responses aided by helper T cells lead to the formation of germinal centers where affinity maturation takes

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

place. This adjustment of affinity by repeated cycles of random somatic mutation and selection focuses interactions into a particular direction in antigenic structure space, leading to the growth of system only in that direction. Post-germinal center B cells will have accumulated somatic hypermutations to increase their affinity. This corresponds to increased values of r, the number of non-covalent bonds in the binding surface. In this region of interaction space, most effector cells are eliminated by apoptosis and some cells go into resting state, once target is cleared. These latter are the memory B cells that conserve the genetic information acquired during affinity maturation but minimize their activity: no divisions, no antibody secretion; remaining guards against future encounter with the same or similar target (Fig. 2C) (53–55). Another important type of cell that remains in the system is the long-lived plasma cell, which is a terminally differentiated B cell that has become independent of BCR signals and is fully devoted to antibody secretion (56,57).

4. Characterization of the antibody interaction network

4.1 Distribution of binding energies in the system

Cell and systems biological aspects of the RADARS model are summarized in Figure 3. Expansion and differentiation of cells originating from CLPs create a cellular repertoire in the bone marrow, then further expansions and differentiation steps increase this repertoire and supplement it with antibody secretion. The development of B cells bearing surface antibodies, B-cell receptors, with increasing affinities takes place in an environment loaded with a huge diversity of macromolecules. These antibodies thus develop in a system characterized by constant reversible, non-covalent interactions. These interactions in the system can be described mathematically by the frequency distribution of interaction energy. According to the random energy model, the free energy of biomolecular interactions is normally distributed (49). Testing a given antigen against a universe of randomly generated antibodies would yield a normal distribution of binding energies. However, it is exactly the role of the immune system to adjust the binding energy against particular antigens according to their quality of selfness and dangerousness. The Generalized Quantitative Model of antibody homeostasis proposes that antigen concentrations can be best regulated in the host by setting the equilibrium dissociation constants nearly equal to the desired antigen concentration

$[Ag] \approx K_D$

Substituting into eq 2),

$$[Ag] \approx \exp(\Delta G^\circ/RT) \qquad 5)$$

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Thus the model suggests an exponential relationship between the adjusted free antigen concentration [Ag] and the binding free energy. Since the availability of antigen corresponds to the probability of an interaction, expansion of the antibody system in the proposed interaction space towards antigen shape space will also be exponentially distributed. The combination of Gaussian distributed interactions of random structures with exponentially distributed antibody frequencies yields a lognormal distribution of free energy levels (Fig.4). Consequently, a normal distribution with mean interaction energy $\mu=0$ and variance $\sigma^2=RT^2$ is transformed into a lognormal distribution with mean = $\exp(\mu + \sigma^2/2)$. This approach gives a mean energy of 27.67 kJ/mol, which is equivalent to a K_D of $2x10^{-5}$. This is in a good agreement with the observed lower limits of antibody binding energy and is also the lower limit of BCR sensitivity (58).

In order to define the behavior of a system of molecules we can introduce a system equilibrium constant

$K_{sys} = K_A/\langle K_A\rangle = e^{-(\Delta G^\circ - \langle\Delta G^\circ\rangle)/RT}$ 6)

where < > enclose median equilibrium constant and mean free energy of binding in the system of antibodies, respectively. K_{sys} represents the binding propensity of a given molecule in the system. The reason for introducing this parameter is that whereas in a bimolecular interaction K_A is a sufficient measure for determining the ratio of bound and free molecules in equilibrium, in a system a given molecule can interact with any other component of the system and its relative free energy will determine its behavior in the system. Instead of looking at particular bimolecular interactions we are interested in the flow of antigen arising from the sequential interaction with antibodies with increasing affinity.

To further explore the properties of the system we shall consider the combination of a lognormal distribution and an exponential distribution by exponential sampling, as has been described by Reed and Mitzenmacher (59,60). For the whole system, we need to consider the exponentially distributed number of normal distributions of interaction energy, as explained above, corresponding to exponentially distributed number of lognormal distributions of equilibrium constants. The probability density function of K_{sys} with such distribution is given by

equation 7)

$$f(Ksys) = \int_{r=0}^{\infty} \lambda e^{-\lambda \Delta E^2} \frac{\sqrt{\lambda}}{\sqrt{\pi} 2 \Delta E Ksys} e^{-\frac{\lambda(\Delta G^\circ sys)^2}{4r\Delta E^2}} dr$$

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

using a λ rate of exponential decrease of antigen concentrations, and a variance of $2/\lambda$ for the normal distribution. We can regard the energy fluctuation associated with ligand independent tonic signaling (61) as the unit variance ΔE^2.

This combination of exponential and lognormal distributions generates a double-Pareto distribution of K_{sys}, which for $K_{sys}>1$ is given by

$$f(K_{sys})= \varphi/2 * K_{sys}^{-1-\lambda} \qquad 8)$$

which is a power law distribution of system affinity constants with degree exponent λ-1 (Fig.4 and 5). From this distribution we can now obtain the network description of a system of antibody interactions. A more detailed description of the deduction of power law distribution and the interpretation of interaction space is given in supplementary text 1.

4.2 Hierarchy of binding events and geometry of the system

Following an active expansion phase of the immune response, when a number of short-lived plasma cells is generated and antigen is cleared by means of the antibodies produced, the immune response settles and the system retracts. Memory B cells and long lived plasma cells are formed, which provide walls of protection against future intrusion by that particular antigen (62). While MBC can enter GC reaction upon antigenic re-challenge, LLPC are terminally differentiated. This also means that while MBC are dependent on tonic BCR signaling, LLPC are independent of such signals.

In the RADARS model these walls of protection can be thought of as an outermost surface of high affinity antibodies produced by LLPC, and an inner multiple layer of MBC. The affinity of the antibody towards cognate antigen determines the distance of the surface of protection in that particular direction in interaction space. This translates to a maximal concentration of that antigen, since high affinity antibodies can reduce target concentration with an efficiency determined by affinity. Should antigen concentration rise further, MBC of lower energy layers are triggered to enter germinal centers again, expand and differentiate, producing new antibody secreting cells and a stronger wall of protection (54).

The important question here is the organization of this hierarchy. Both MBC and LLPC may persist for a lifetime (57,63), so the optimal coverage of shape space in balance with affinity requires organization of interaction energy and space. In the case of LLPC we know of no antigen specific deletion mechanism, and competition for niche, which is not antigen specific, might be the only attrition factor (64,65). MBC may undergo further rounds of affinity maturation in GC and can therefore be reorganized (66). The affinity of antibodies produced by LLPC is generally expected to be high (67), though initial low affinity of BCR favors differentiation into LLPC following GC reaction, while initial high affinity has opposite effects

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(68). This observation underlines that hierarchy needs to be established rather than single high affinity cells selected during an immune response.

4.3 A scale-free network of interactions

The systems organization of binding events, as outlined above, can be interpreted as a network of interactions. While antibodies seek to minimize their free energy by finding their best fitting target, antigens are passed on from antibody to antibody. Such shared consecutive binding represents links in the network between nodes of antibodies. Antibodies with the highest K_{sys} value are the most avid binders and as such will take over and handle most antigens, which are channeled to these molecules by antibodies underneath in the binding hierarchy. These antibodies will have the highest number of links and therefore the highest network degree k. The RADARS model in combination with the GQM (27) suggests that long-lived plasma cells act as network hubs. By constantly secreting antibodies, these terminally differentiated cells provide for the binding and removal of antigen of all network nodes, represented by memory B cells, below in the hierarchy. This prevents activation of memory B cells and maintains their resting state. At the bottom of the hierarchy B1 cells, producing natural antibodies, serve as a first line of defense and relay agents for antigen (Fig.6). The higher the network degree and corresponding K_{sys}, the more cells and structure space is covered by a plasma cell.

Assuming that $k=K_{sys}$ we obtain the probability density function of the antibody network

$$p(k) \sim k^{-\gamma} \qquad 9)$$

where γ is the degree exponent and is equal to λ-1 (Fig.5).

Assuming that each subnetwork of LLPC has a similar organization determined by thermodynamic laws, the whole system is expected to have a self-similar fractal topology (69). Fractality in such networks is characterized by the box dimension dB, which is related to the degree exponent (70) by

$$d_B=(\gamma-1)/(\gamma-2) \qquad 10)$$

The network degree distribution exponent in the RADARS model describes the connectivity between shells of free energy levels, on the other hand fractal box dimension describes the rate of appearance of clusters in the network projected onto the surface of the system. Therefore a further level of similarity can be introduced if these two parameters are equal. An ideal network with equal network degree distribution exponent and fractal dimension is satisfied by one condition only, if d_B=golden ratio+1 since $\phi/(\phi-1)=\phi^2=\phi+1$ (golden ratio:$\phi=(1+\sqrt{5})/2$. It is proposed that an optimal resting immune network in thermodynamic equilibrium could be

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

characterized by this ideal network with degree distribution exponent $\gamma=\phi+1=\phi^2$. Then the probability that an antibody can cross-react with k other antibodies will decay as a power law $p(k) \sim k^{-\gamma}$ and the number of boxes N_B needed to tile a renormalized network, e.i. the probability of finding antibody interaction clusters characterized by the shortest path l_B, will decay as a power law $N_B(l_B)/N \sim l_B^{-\gamma}$. In our model l_B is in fact a distance along the projected surface of the system, and as such it represents structural similarity.

The power law relationship of antibody interactions is a hallmark of scale-free networks (71). This scale-free network is an energy transfer system physically and an antigen transfer system immunologically. This is an optimization of antibody differentiation in the sense that the minimal number of high free energy antibodies (network hubs) are used for the removal of the maximal diversity of antigens, covering the maximum of immunologically relevant structure space. The generation of an antibody network, with network hubs represented by plasma cells secreting antibodies, reveals the physical aspect of the system: all interactions of such an antibody contribute to the clearance of many target antigens sharing structural homology. A new node in the network, a new B cell in the structure space, will preferentially attach to an existing subnetwork as a low-affinity clone, in agreement with the preferential attachment model of growth in scale-free networks (72). Preferential attachment may explain immunodominance and antigenic sin, phenomena arising from the preference of the immune system for known epitopes, which correspond to hubs in the network.

4.4 An inverted view of the system

The model presented so far follows a view of the system growing outwards, with cells supplied from within and extending into the world of antigens. However, we can turn this view inside out, as shown in Figure 6, interpreting the events as the immune system being outside and developing inwards. Practically this means that we use the dissociation constant K_D instead of K_A as a measure for the system. Though it is more difficult to visualize a system that grows inwards, this view helps our perception of the developing antibody network and of antigen removal.

In the immune-side out view of the RADARS model distance from the outside boundary of the system represents resolution of molecular recognition, which is now K_D. Directions still correspond to theoretical targets: individual molecular shapes. If all these targets have binding energy distributions as predicted by the universal distribution model, then our interaction space will represent a collection of these statistical distributions.

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The RADARS model suggests that the greater the resolution of structural recognition the more restricted is the number of shapes recognized. However, with the development of high affinity clones the ability to react with related structures also grows, a phenomenon called cross reactivity. The further the system grows in a given direction the more focused is the recognition of cognate target but affinity to related structures inevitably grows as well. This is not mere polyreactivity, however, but rather organized cross-reactivity. With the supply of B-cell precursors outside and the organization of antigen removal inside, we can best interpret effector antibody function as an antigen sink (Fig.6). In this sink multiple sinkholes develop as the immune system matures. The sinkholes themselves correspond to immunodominant epitopes: structures preferred by the system as targets.

5. Merging sequence space to interaction space

Network theory has always been considered as a key to understand and define immunity on a systems level. The network hypothesis of Niels Jerne (23), its modified version leading to the concept of clonal selection (73), mathematical and computational simulations (74,75), various re-interpretations (24), experimental approaches using NGS (25,31) or antigen microarrays (22) all strive to describe this highly complex system as connected elements of a network. There are two new aspects of the RADARS model that may improve our view of this network. First, it introduces physical units, binding energy, as the measure of interactions and as a measure of system architecture. Natural networks are formed as a result of energy dispersal (76,77), therefore network theories should consider energy transduction in the system. Second, it proposes an architecture for the whole network, characterized by the scale-free distribution, and an optimal value for the degree exponent of power-law relationship.

The network architecture of antibody repertoires was recently computed based on high-throughput sequencing data from more than 100.000 unique antibody sequences (78). This study revealed that pre-B cell and naïve B-cell clones form homogenously interconnected assortative networks, in contrast to the disassortative networks of plasma cell clones, which covered smaller but more focused regions of sequence space. This contrasting behavior of antigen-naïve and antigen experienced, post-germinal center B cells corresponds to the antibody-centric view in our model. The low-affinity region with developing B-cells is homogenously interconnected by clonal relationships and shared usage of gene segments (Fig.2 and 3). The high affinity side of the distribution is the narrowing, focusing interaction space of

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

plasma cells, where growth of subnetworks with different specificity into different directions of antigen shape space appears as hub repulsion and fractal topology (69).

Considering that our technological capability is ripe for the high-resolution determination and comprehensive analysis of antibody sequence space, current efforts focus on the conversion of sequence space data into datasets in interaction space. By providing a physical and mathematical description of relationship between antibody clones the RADARS model may help in the final integration of sequence data. The model also suggests that sequence-based network properties of early B-cell developmental stages also need to be determined (25), in addition to the mature and antigen-experienced repertoire (79), and comprehensive and selective analysis of the B1 repertoire is very important for capturing network properties of the system.

The model presented here with a network degree exponent φ^2 depicts an ideal state of the system of antibody interactions. It is the fluctuations and disturbances in the system that we observe as immune response during infections, and distortions are autoimmunity and allergy. Besides suggesting how antibody sequence space fits into structural space and into an interaction network, the model may potentially lead to the ability to model whole immune system and simulate its functioning.

6. Concluding remarks

This theoretical study introduces the concept of antibody interaction space, which arises from structure space, and is based on the distribution of interaction energy. In a physical chemistry perspective, the immune system is a grand canonical ensemble: the number of states with given energy corresponds to conformational diversity of antibodies, the probability of states is dictated by the availability of antigen, the host serves as a heat bath and particle bath, maintaining temperature and being a source of antigen, finally, chemical potential is kept the same by immune mechanisms adjusting antibody concentration according to K_D. A self-organizing system, such as the humoral adaptive immune system, is based on the organization of interactions. Since molecular interactions are determined by structure, organized interaction space should mean an organized structure space. The RADARS model proposes that a universal organization of an immense number of structures in a huge but finite system is possible by adjusting the resolution of structural recognition, which is the adjustment of interaction energy. Radial adjustment of system resolution generates a network of interactions. The network of interactions is scale-free and is characterized by a power law distribution of

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

free energy of interactions. Overall, this organization allows the energy optimized controlled removal of antigens from the host system.

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgments

The paper is dedicated to my first mentor in molecular biology and antibody cloning, István Kurucz. The author is partly supported by VEKOP 2.1.1-15-00098 grant to Diagnosticum. I wish to thank István Csabai for reading and discussing the manuscript and anonymous referees for their constructive criticism. Special thanks to Tamás Pfeil and Ágnes Kovács for enlightening discussions about exponential distributions.

Author contributions

J.P. developed theoretical model, designed and prepared figures, wrote paper

Conflicts of interest

The author declares no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Müller V, de Boer RJ, Bonhoeffer S, Szathmáry E. An evolutionary perspective on the systems of adaptive immunity. Biol Rev Camb Philos Soc. 2017 Jul 26;93(1):505–528.
2. Eberl G. A new vision of immunity: homeostasis of the superorganism. Mucosal Immunol. 2010 Sep;3(5):450–460.
3. Sleator RD. The human superorganism - of microbes and men. Med Hypotheses. 2010 Feb;74(2):214–215.
4. Pradeu T, Carosella ED. On the definition of a criterion of immunogenicity. Proc Natl Acad Sci U S A. 2006 Nov 21;103(47):17858–17861.
5. Davis MM, Tato CM, Furman D. Systems immunology: just getting started. Nat Immunol. 2017 Jun 20;18(7):725–732.
6. Prechl J. A generalized quantitative antibody homeostasis model: maintenance of global antibody equilibrium by effector functions. Clin Transl Immunology. 2017 Nov 17;6(11):e161.
7. Boyd SD, Joshi SA. High-Throughput DNA Sequencing Analysis of Antibody Repertoires. Microbiol Spectr. 2014 Oct;2(5).
8. Vergani S, Korsunsky I, Mazzarello AN, Ferrer G, Chiorazzi N, Bagnara D. Novel Method for High-Throughput Full-Length IGHV-D-J Sequencing of the Immune Repertoire from Bulk B-Cells with Single-Cell Resolution. Front Immunol. 2017 Sep 14;8:1157.
9. Friedensohn S, Lindner JM, Cornacchione V, Iazeolla M, Miho E, Zingg A, et al. Synthetic standards combined with error and bias correction improve the accuracy and quantitative resolution of antibody repertoire sequencing in human naïve and memory B cells. Front Immunol. 2018 Jun 20;9:1401.
10. Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V. Computational Strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune Repertoires. Front Immunol. 2018 Feb 21;9:224.
11. Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and statistical analysis of adaptive immune repertoires. Trends Immunol. 2015 Nov;36(11):738–749.
12. Prechl J, Papp K, Erdei A. Antigen microarrays: descriptive chemistry or functional immunomics? Trends Immunol. 2010 Apr;31(4):133–137.
13. Merbl Y, Zucker-Toledano M, Quintana FJ, Cohen IR. Newborn humans manifest autoantibodies to defined self molecules detected by antigen microarray informatics. J Clin Invest. 2007 Mar;117(3):712–718.
14. Bacarese-Hamilton T, Bistoni F, Crisanti A. Protein microarrays: from serodiagnosis to whole proteome scale analysis of the immune response against pathogenic microorganisms. BioTechniques. 2002 Dec;Suppl:24–29.
15. Wang J, Lin J, Bardina L, Goldis M, Nowak-Wegrzyn A, Shreffler WG, et al. Correlation of IgE/IgG4 milk epitopes and affinity of milk-specific IgE antibodies with different phenotypes of clinical milk allergy. J Allergy Clin Immunol. 2010 Mar;125(3):695–702, 702.e1.
16. Grötzinger C. Peptide microarrays for medical applications in autoimmunity, infection, and cancer. Methods Mol Biol. 2016;1352:213–221.

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

17. Sjöberg R, Mattsson C, Andersson E, Hellström C, Uhlen M, Schwenk JM, et al. Exploration of high-density protein microarrays for antibody validation and autoimmunity profiling. N Biotechnol. 2016 Sep 25;33(5 Pt A):582–592.
18. Bradford TJ, Wang X, Chinnaiyan AM. Cancer immunomics: using autoantibody signatures in the early detection of prostate cancer. Urol Oncol. 2006 Jun;24(3):237–242.
19. Caetano-Anollés G, Wang M, Caetano-Anollés D, Mittenthal JE. The origin, evolution and structure of the protein world. Biochem J. 2009 Feb 1;417(3):621–637.
20. Scaviner D, Barbié V, Ruiz M, Lefranc MP. Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions. Exp Clin Immunogenet. 1999;16(4):234–240.
21. Krawczyk K, Kelm S, Kovaltsuk A, Galson JD, Kelly D, Trück J, et al. Structurally mapping antibody repertoires. Front Immunol. 2018 Jul 23;9:1698.
22. Madi A, Kenett DY, Bransburg-Zabary S, Merbl Y, Quintana FJ, Tauber AI, et al. Network theory analysis of antibody-antigen reactivity data: the immune trees at birth and adulthood. PLoS ONE. 2011 Mar 8;6(3):e17445.
23. Jerne NK. Towards a network theory of the immune system. Ann Immunol (Paris). 1974 Jan;125C(1-2):373–389.
24. Coutinho A. The network theory: 21 years later. Scand J Immunol. 1995 Jul;42(1):3–8.
25. Miho E, Greiff V, Roskar R, Reddy ST. The fundamental principles of antibody repertoire architecture revealed by large-scale network analysis. BioRxiv. 2017 Apr 5;
26. Prechl J. A generalized quantitative antibody homeostasis model: regulation of B-cell development by BCR saturation and novel insights into bone marrow function. Clin Transl Immunology. 2017 Feb 17;6(2):e130.
27. Prechl J. A generalized quantitative antibody homeostasis model: antigen saturation, natural antibodies and a quantitative antibody network. Clin Transl Immunology. 2017 Feb 24;6(2):e131.
28. Elhanati Y, Sethna Z, Marcou Q, Callan CG, Mora T, Walczak AM. Inferring processes underlying B-cell repertoire diversity. Philos Trans R Soc Lond, B, Biol Sci. 2015 Sep 5;370(1676).
29. Kovaltsuk A, Leem J, Kelm S, Snowden J, Deane CM, Krawczyk K. Observed Antibody Space: A Resource for Data Mining Next-Generation Sequencing of Antibody Repertoires. J Immunol. 2018 Sep 14;
30. Rettig TA, Ward C, Bye BA, Pecaut MJ, Chapes SK. Characterization of the naive murine antibody repertoire using unamplified high-throughput sequencing. PLoS ONE. 2018 Jan 10;13(1):e0190982.
31. Bashford-Rogers RJM, Palser AL, Huntly BJ, Rance R, Vassiliou GS, Follows GA, et al. Network properties derived from deep sequencing of human B-cell receptor repertoires delineate B-cell populations. Genome Res. 2013 Nov;23(11):1874–1884.
32. D'Angelo S, Ferrara F, Naranjo L, Erasmus MF, Hraber P, Bradbury ARM. Many Routes to an Antibody Heavy-Chain CDR3: Necessary, Yet Insufficient, for Specific Binding. Front Immunol. 2018 Mar 8;9:395.

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

33. Greiff V, Menzel U, Miho E, Weber C, Riedel R, Cook S, et al. Systems Analysis Reveals High Genetic and Antigen-Driven Predetermination of Antibody Repertoires throughout B Cell Development. Cell Rep. 2017 May 16;19(7):1467–1478.
34. DeFalco J, Harbell M, Manning-Bog A, Baia G, Scholz A, Millare B, et al. Non-progressing cancer patients have persistent B cell responses expressing shared antibody paratopes that target public tumor antigens. Clin Immunol. 2018;187:37–45.
35. Bukhari SAC, O'Connor MJ, Martínez-Romero M, Egyedi AL, Willrett D, Graybeal J, et al. The CAIRR Pipeline for Submitting Standards-Compliant B and T Cell Receptor Repertoire Sequencing Studies to the National Center for Biotechnology Information Repositories. Front Immunol. 2018 Aug 16;9:1877.
36. Breden F, Luning Prak ET, Peters B, Rubelt F, Schramm CA, Busse CE, et al. Reproducibility and reuse of adaptive immune receptor repertoire data. Front Immunol. 2017 Nov 1;8:1418.
37. Rubelt F, Busse CE, Bukhari SAC, Bürckert J-P, Mariotti-Ferrandiz E, Cowell LG, et al. Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data. Nat Immunol. 2017 Nov 16;18(12):1274–1278.
38. Rosenfeld AM, Meng W, Luning Prak ET, Hershberg U. Immunedb, a novel tool for the analysis, storage, and dissemination of immune repertoire sequencing data. Front Immunol. 2018 Sep 21;9:2107.
39. Corrie BD, Marthandan N, Zimonja B, Jaglale J, Zhou Y, Barr E, et al. iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. Immunol Rev. 2018;284(1):24–41.
40. Avram O, Vaisman-Mentesh A, Yehezkel D, Ashkenazy H, Pupko T, Wine Y. ASAP - A Webserver for Immunoglobulin-Sequencing Analysis Pipeline. Front Immunol. 2018 Jul 30;9:1686.
41. Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes in antibody:protein complexes. Mol Immunol. 2013 Jan;53(1-2):24–34.
42. Wedemayer GJ, Patten PA, Wang LH, Schultz PG, Stevens RC. Structural insights into the evolution of an antibody combining site. Science. 1997 Jun 13;276(5319):1665–1669.
43. Chen J, Sawyer N, Regan L. Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. Protein Sci. 2013 Apr;22(4):510–515.
44. Brooijmans N, Sharp KA, Kuntz ID. Stability of macromolecular complexes. Proteins. 2002 Sep 1;48(4):645–653.
45. Derrida B. Random-energy model: An exactly solvable model of disordered systems. Phys Rev B. 1981 Sep 1;24(5):2613–2626.
46. Mézard M, Montanari A. The random energy model. Information, physics, and computation. Oxford University Press; 2009. p. 93–105.
47. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem. 1997;48:545–600.
48. Childs LM, Baskerville EB, Cobey S. Trade-offs in antibody repertoires to complex antigens. Philos Trans R Soc Lond, B, Biol Sci. 2015 Sep 5;370(1676).

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

49. Zheng X, Wang J. The universal statistical distributions of the affinity, equilibrium constants, kinetics and specificity in biomolecular recognition. PLoS Comput Biol. 2015 Apr 17;11(4):e1004212.
50. Melchers F. Checkpoints that control B cell development. J Clin Invest. 2015 Jun;125(6):2203–2210.
51. Prechl J. Thermodynamic projection of the antibody interaction network: the fountain energy landscape of binding. BioRxiv. 2017 Apr 5;
52. Cohen IR. Biomarkers, self-antigens and the immunological homunculus. J Autoimmun. 2007 Dec;29(4):246–249.
53. Tangye SG, Tarlinton DM. Memory B cells: effectors of long-lived immune responses. Eur J Immunol. 2009 Aug;39(8):2065–2075.
54. Shah HB, Smith K, Wren JD, Webb CF, Ballard JD, Bourn RL, et al. Insights From Analysis of Human Antigen-Specific Memory B Cell Repertoires. Front Immunol. 2018;9:3064.
55. Neuberger MS, Ehrenstein MR, Rada C, Sale J, Batista FD, Williams G, et al. Memory in the B-cell compartment: antibody affinity maturation. Philos Trans R Soc Lond, B, Biol Sci. 2000 Mar 29;355(1395):357–360.
56. Good-Jacobson KL, Shlomchik MJ. Plasticity and heterogeneity in the generation of memory B cells and long-lived plasma cells: the influence of germinal center interactions and dynamics. J Immunol. 2010 Sep 15;185(6):3117–3125.
57. Hammarlund E, Thomas A, Amanna IJ, Holden LA, Slayden OD, Park B, et al. Plasma cell survival in the absence of B cell memory. Nat Commun. 2017 Nov 24;8(1):1781.
58. Batista FD, Neuberger MS. Affinity dependence of the B cell response to antigen: a threshold, a ceiling, and the importance of off-rate. Immunity. 1998 Jun;8(6):751–759.
59. Reed WJ, Jorgensen M. The Double Pareto-Lognormal Distribution—A New Parametric Model for Size Distributions. Communications in Statistics - Theory and Methods. 2004 Dec 31;33(8):1733–1753.
60. Mitzenmacher M. Dynamic models for file sizes and double pareto distributions. Internet Math. 2004 Jan 1;1(3):305–333.
61. Monroe JG. Ligand-independent tonic signaling in B-cell receptor function. Curr Opin Immunol. 2004 Jun;16(3):288–295.
62. Akkaya M, Kwak K, Pierce SK. B cell memory: building two walls of protection against pathogens. Nat Rev Immunol. 2019 Dec 13;
63. Jones DD, Wilmore JR, Allman D. Cellular dynamics of memory B cell populations: igm+ and igg+ memory B cells persist indefinitely as quiescent cells. J Immunol. 2015 Nov 15;195(10):4753–4759.
64. Wilmore JR, Allman D. Here, There, and Anywhere? Arguments for and against the Physical Plasma Cell Survival Niche. J Immunol. 2017 Aug 1;199(3):839–845.
65. Lightman SM, Utley A, Lee KP. Survival of Long-Lived Plasma Cells (LLPC): Piecing Together the Puzzle. Front Immunol. 2019 May 3;10:965.
66. Shlomchik MJ. Do memory B cells form secondary germinal centers? yes and no. Cold Spring Harb Perspect Biol. 2018 Jan 2;10(1).

bioRxiv preprint doi: https://doi.org/10.1101/438804; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

67. Phan TG, Paus D, Chan TD, Turner ML, Nutt SL, Basten A, et al. High affinity germinal center B cells are actively selected into the plasma cell compartment. J Exp Med. 2006 Oct 30;203(11):2419–2424.
68. O'Connor BP, Vogel LA, Zhang W, Loo W, Shnider D, Lind EF, et al. Imprinting the fate of antigen-reactive B cells through the affinity of the B cell receptor. J Immunol. 2006 Dec 1;177(11):7723–7732.
69. Song C, Havlin S, Makse HA. Origins of fractality in the growth of complex networks. Nat Phys. 2006 Apr;2(4):275–281.
70. Kim JS, Goh KI, Kahng B, Kim D. Fractality and self-similarity in scale-free networks. New J Phys. 2007 Jun 28;9(6):177–177.
71. Albert R, Barabási A-L. Statistical mechanics of complex networks. Rev Mod Phys. 2002 Jan 30;74(1):47–97.
72. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004 Feb;5(2):101–113.
73. Burnet FM. A modification of Jerne's theory of antibody production using the concept of clonal selection. CA Cancer J Clin. 1976 Apr;26(2):119–121.
74. FARMER DJ, PACKARD N, PERELSON A. The immune system, adaptation, and machine learning. Physica D: Nonlinear Phenomena. 1986;22(1-3):187–204.
75. Schulz R, Werner B, Behn U. Self-tolerance in a minimal model of the idiotypic network. Front Immunol. 2014 Mar 10;5:86.
76. Mäkelä T, Annila A. Natural patterns of energy dispersal. Phys Life Rev. 2010 Dec;7(4):477–498.
77. Hartonen T, Annila A. Natural networks as thermodynamic systems. Complexity. 2012 Nov;18(2):53–62.
78. Miho E, Roškar R, Greiff V, Reddy ST. Large-scale network analysis reveals the sequence space architecture of antibody repertoires. Nat Commun. 2019 Mar 21;10(1):1321.
79. Kono N, Sun L, Toh H, Shimizu T, Xue H, Numata O, et al. Deciphering antigen-responding antibody repertoires by using next-generation sequencing and confirming them through antibody-gene synthesis. Biochem Biophys Res Commun. 2017 May 27;487(2):300–306.

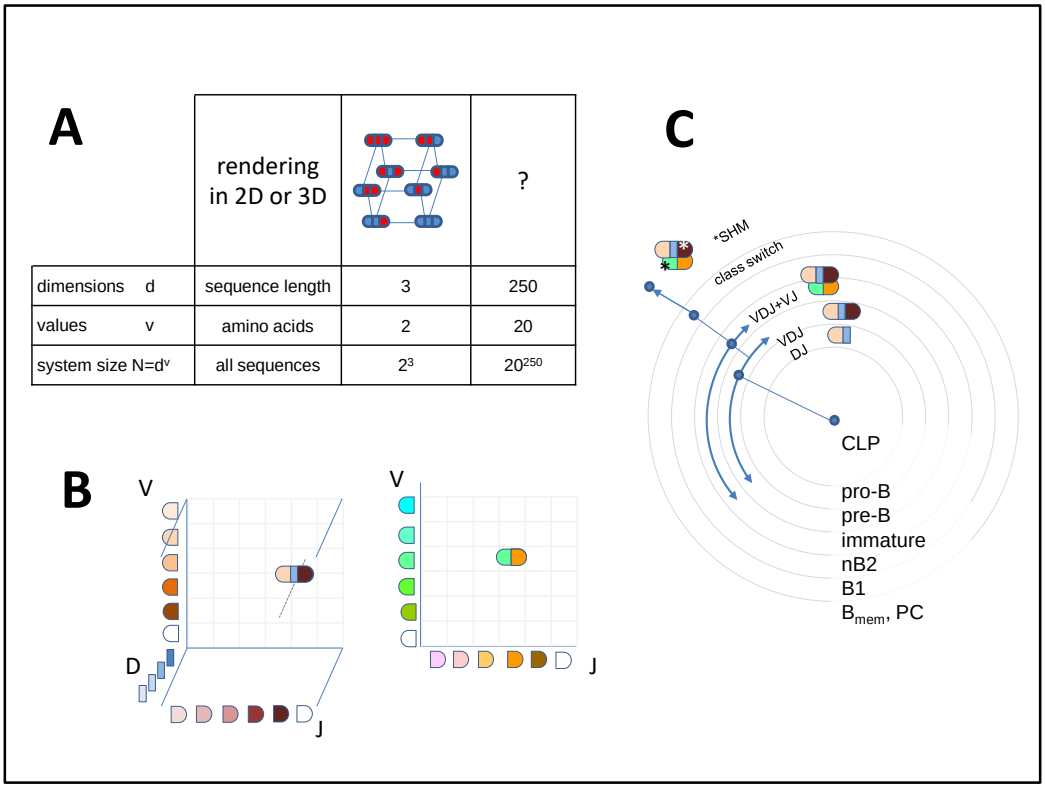
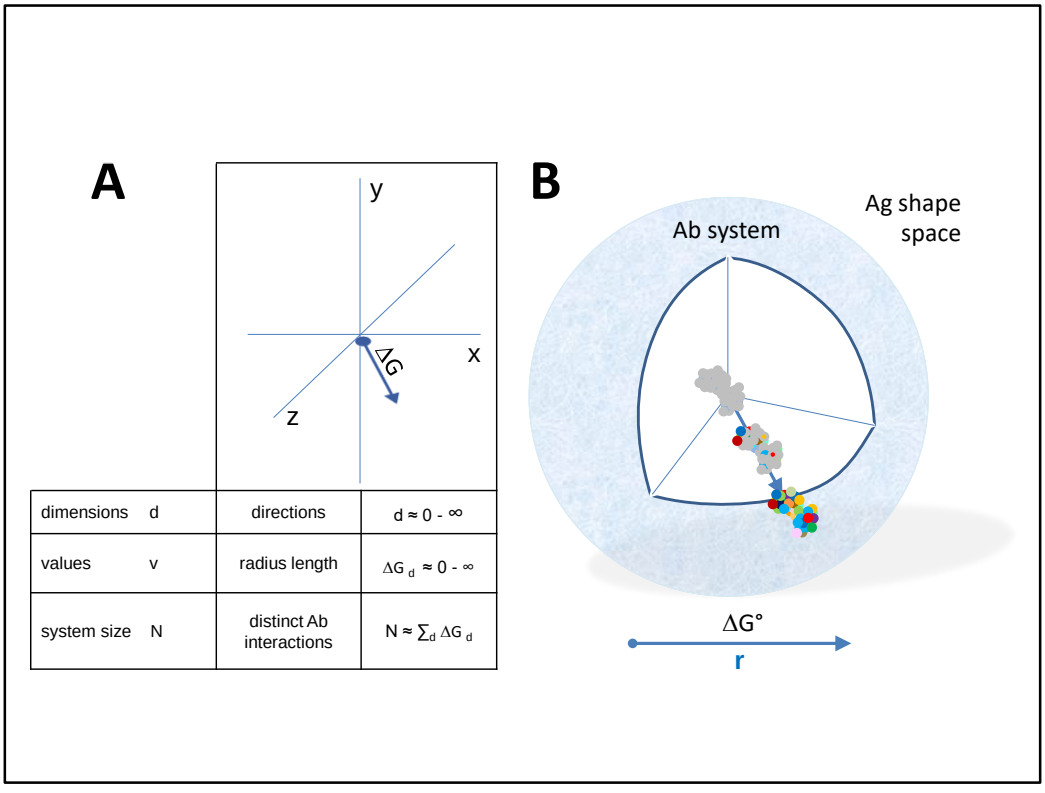


Figure 1. Sequence space and visualization of antibody sequence relationships
A) Theoretical diversity (system size) of a sequence is determined by its length (dimension) and the number of values a particular position in the sequence can take. An antibody Fv region of 250 amino acids has an astronomical sequence diversity if full randomization is allowed. Exact sequences take up discrete positions in sequence space, but there are no structurally meaningful directions or distances in this multidimensional sequence space with N different sequences.
B) Antibody sequences are frequently interpreted as recombined germline sequences. This approach allows the simplified display of repertoires obtained by NGS, preferably with paired heavy and light chain VD identification. Such a display of combinatorial diversity may allow the tracking of specific clonal expansions and further diversification by SHM but reveals little about the overall functional network of interactions.
C) The potential development scheme of a given antibody clone is shown with antibody sequence development along with B-cell differentiation steps. Arching arrows represent combinatorial diversification by V-D-J rearrangement and light chain paring.



dimensions	d	directions	$d \approx 0 - \infty$
values	v	radius length	$\Delta G_d \approx 0 - \infty$
system size	N	distinct Ab interactions	$N \approx \sum_d \Delta G_d$

Figure 2. Quantitative interaction space and B-cell differentiation
A) The system of antibodies has a center as a reference point in a conceptual three dimensional metric space of interactions. Structural diversity, that is different shapes, appear as directions (exemplary arrow) from this center towards the target structure in Ag shape space. Distinct directions can be defined with a precision dependent on the distance from the center, equivalent to the radius of the system extending in that particular direction. This distance is measured as the free energy change of antibody-antigen interaction. Multidimensionality is theoretically infinite in this representation, practical limits being introduced by the maximum interaction energy (DG°) in the system. The total size of the system N is the sum of all interactions in all directions.
B) Diversity appears as we leave the center of the system, spherical shells representing various levels of resolution of molecular recognition. Colors represent distinct amino acids of antibody binding site engaging in non-covalent bonding with the target molecule. Higher energy interactions engage more residues.

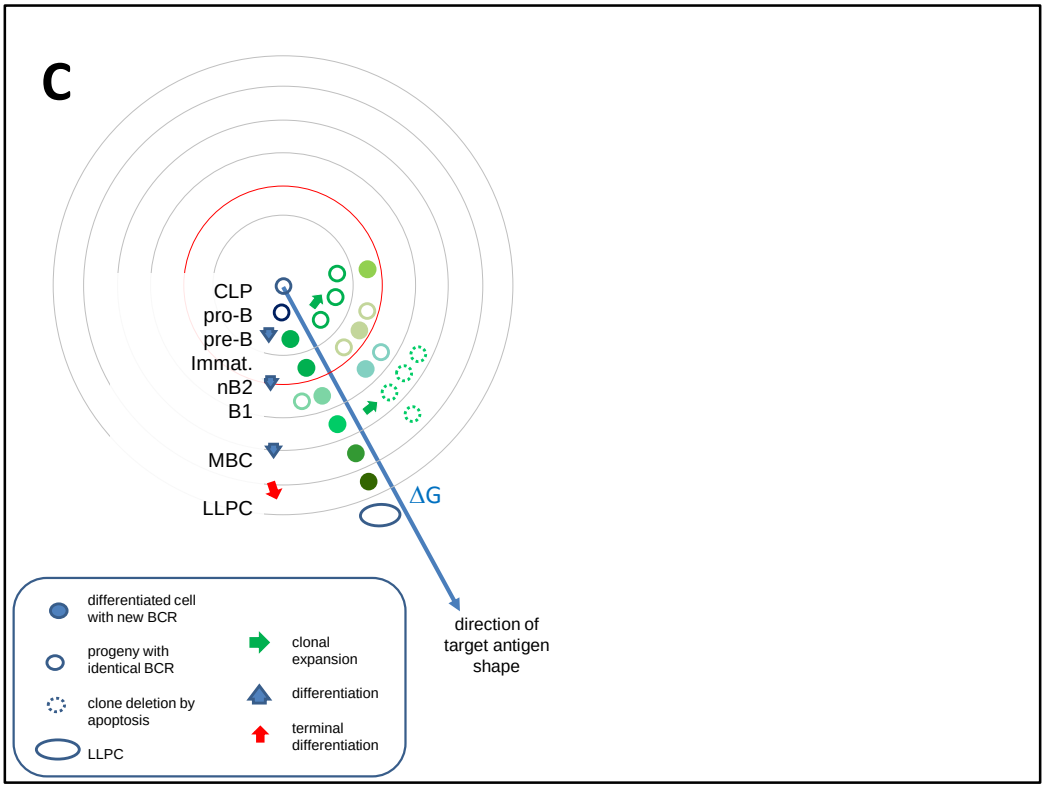
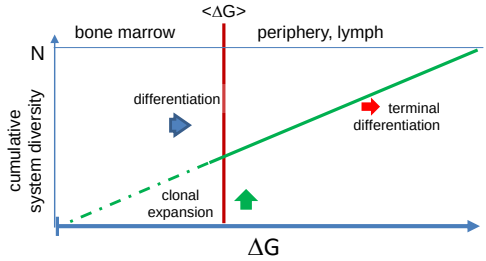


Figure 2.
C) Antigen driven evolution of interaction space.
The evolution of the system of antibodies can be interpreted as clones filling the interaction space at various levels of resolution. Along this pathway cells continually increase their BCR affinity towards the target. A common lymphoid progenitor has the potential to develop antibody against any target. Differentiation proceeds outwards from this point of origin. At every level of differentiation clones expand by asymmetric divisions, expansion being dictated by antigen driven selection of clones with ideal receptor engagement. Naive B2 cells are continuously generated and only survive if recruited for specific immune responses. B1 cells, on the contrary, survive in an activated state producing antibodies and dividing rapidly upon activation. Beyond the mean energy of the system, represented by red circle, TD responses allow further directed differentiation via somatic hypermutations in germinal centers. Clonal expansions are however followed by retraction, effector cells eliminated by apoptosis, only memory B cells and plasma cells surviving. Different colors stand for structural differences and relationships.
LLPC, long-lived plasma cell; MBC, memory B cell; CLP, common lymphoid progenitor



cell type / antibody	CLP	pro-B	pre-B	immat., nB2	B1	MBC	PC
Heavy chain Light chain	- -	DJ -	VDJ -	VDJ VJ	VDJ VJ	V*D*J* V*J*	V*D*J* V*J*
BCR Secreted Ab	- -	- -	Pre -	+ -	+ +	+ -	- +

Figure 3. Generation of diversity in the humoral immune system.
Network development of the humoral immune system, shown as cumulative distribution of evolving antibody clones. Diversity generation in the bone marrow utilizes random recombination of gene segments (V,D,J) selecting for B cells with high diversity but modest affinity. Environmental antigens in the periphery drive the generation and selection of high affinity antibodies, employing somatic hypermutation, leading to further diversification in the system. The energy of these interactions is adjusted for each antigen based on its immunological properties, thereby secreted antibodies can adjust the concentration of the target antigen in the host.
*somatic hypermutations; CLP, common lymphoid progenitor; immat., immature; B2eff, effector B2 cell; Bmem, memory B cell; PC, plasma cell; BCR, B-cell receptor; sAb, secreted antibody

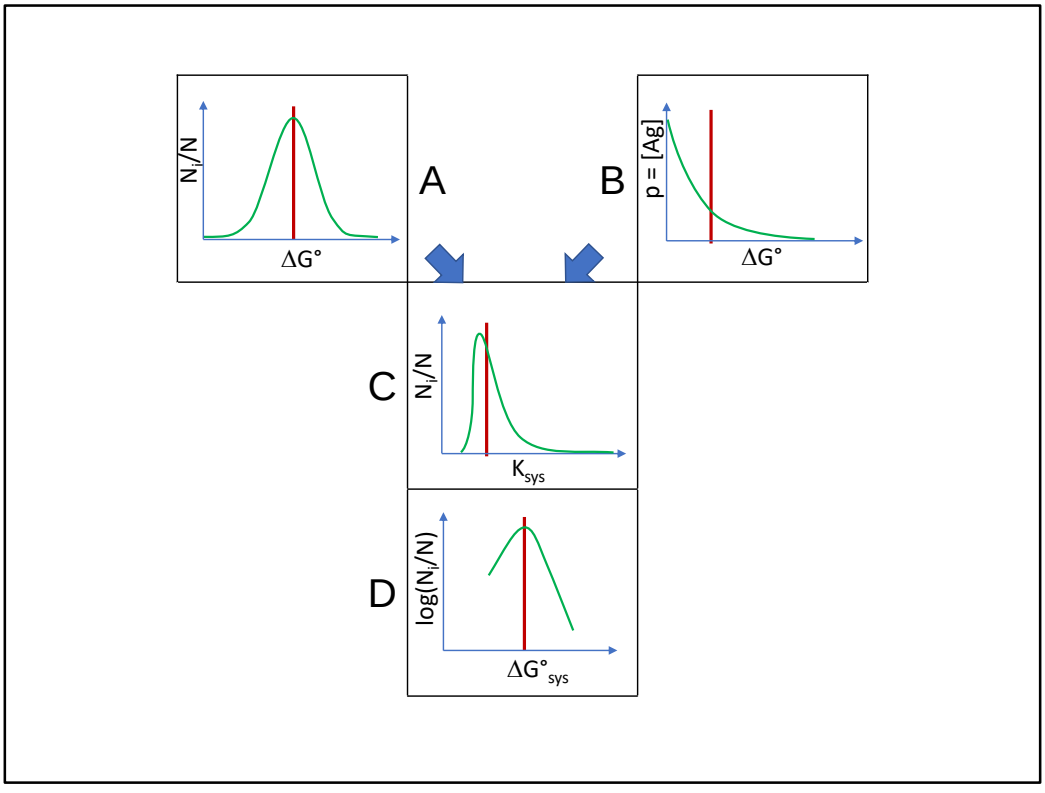


Figure 4. Combination of distributions.
Normally distributed interaction energies (A) are selected by exponentially distributed probability of interactions with antigen (B), resulting in double-Pareto distribution of system equilibrium interaction constants (C), which corresponds to the distribution of conformational entropy in interaction energy levels (D) when logarithm is taken.
Red lines represent the expected values of the distributions.

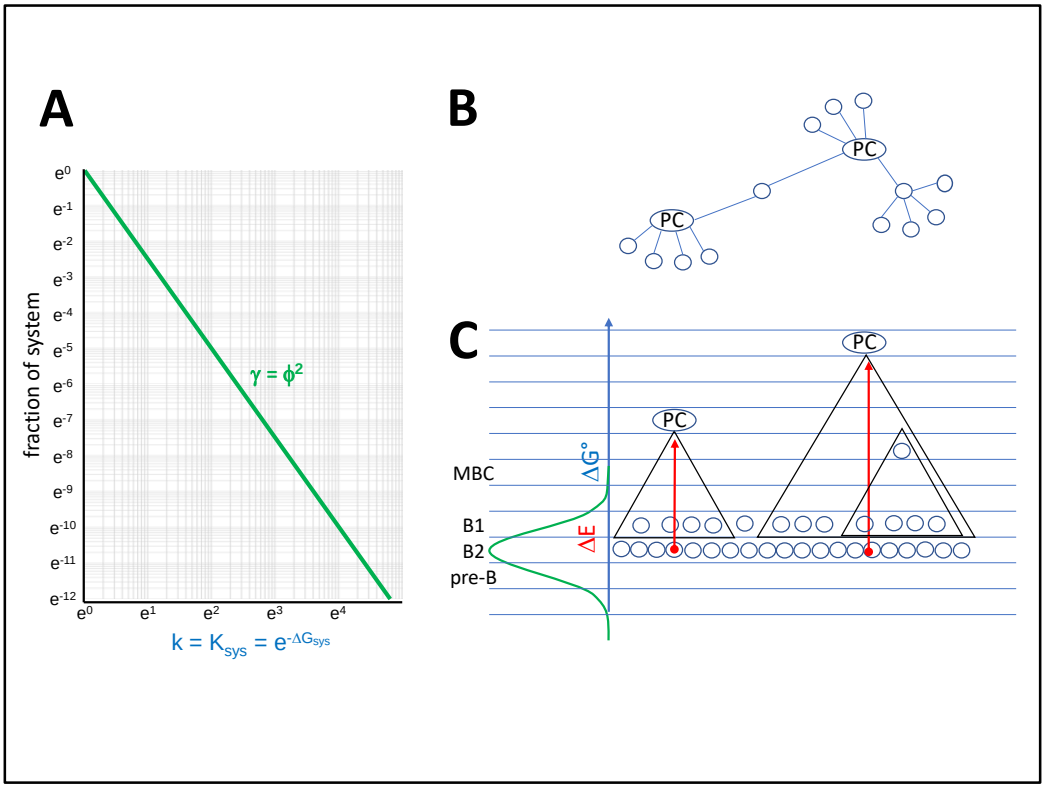


Figure 5. Properties of the antibody interaction network.
A In the antibody network a node is a B cell at a particular differentiation stage and a given BCR sequence. Links are pathways of antibody-antigen interactions, along which antigen is transferred from a lower energy node to a higher energy node. Degree distribution of the antibody network corresponds to the probability distribution of relative free energy of antibody interactions, where the system equilibrium constant is the degree of node and Ni/N is relative frequency of nodes with that degree in the interaction network.
B and C. Proposed schematic network organization of antibody interactions.
In subnetworks plasma cells act as hubs, being connected to all memory B cells and B1 cells with shared structures at lower free energy levels. By secreting antibodies, plasma cells control activation of all connected cells. Triangles represent subnetworks, blue lines indicate energy levels, red arrows stand for energy fluctuation.

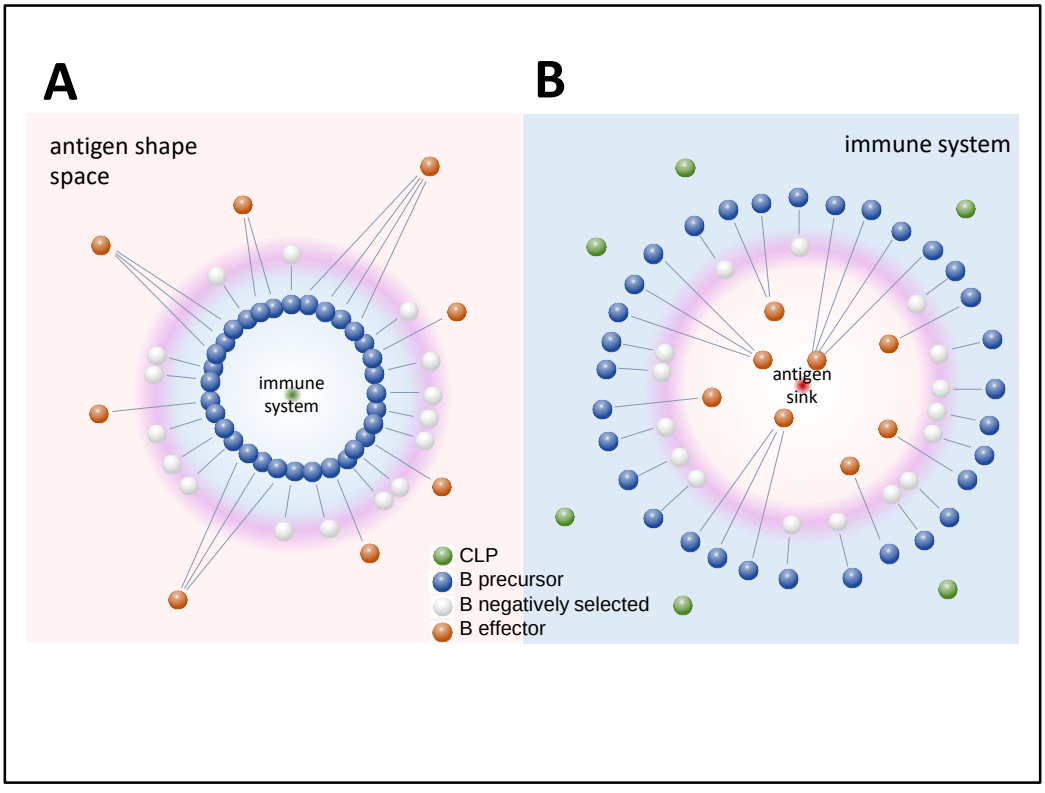
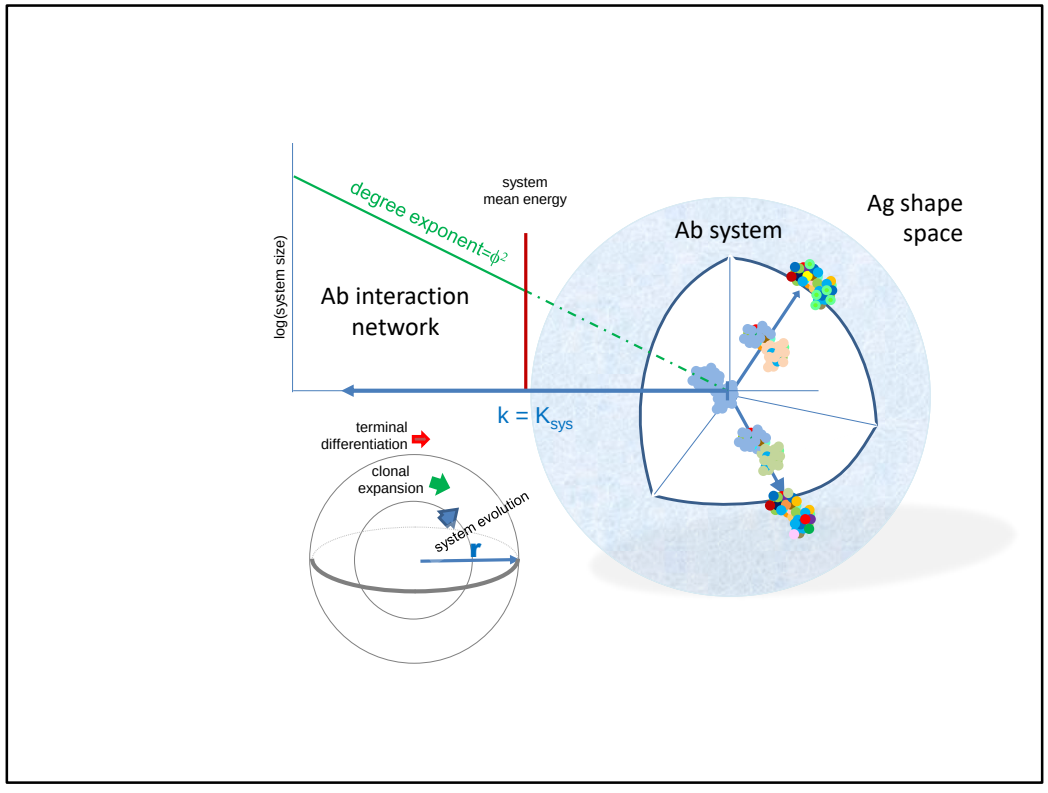


Figure 6. Antibody-centric and antigen-centric views of the organization of adaptive humoral immunity.
The antibody-centric view (A) corresponds to the clonal development and expansion of B cells, generating and maintaining a spherical system with a radius of interaction energy. The system carves out a niche in the molecular structural world of antigens. We can turn the system inside out to obtain an antigen-centric view (B), where the system grows „inwards" with a radius corresponding to the inverse of interaction energy. In this representation the development of antibodies with higher affinity pull antigen down an antigen sink with increasing efficiency, collecting related antigens into coalescing flows.



Graphical abstract