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

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 antibody 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, or RADARS, is proposed. The model is based on the radial growth of interaction affinity of antibodies towards an infinity of directions representing molecular shapes. Levels of interaction strength appear as energy shells of the system. B-cell development and immune responses are interpreted in the model and quantitative properties of the antibody network are inferred from the physical properties of a quasi-spherical system growing multi-radially. The concept of system equilibrium constant is introduced, which is the median of equilibrium constants in the system and serves to define probability of interactions. This thermodynamic system is described by a power-law distribution of antibody free energies with a network degree exponent of phi square, representing a scale-free 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. As an energy transduction system this network serves to optimize free energy consumption, removing antigens at the required rate at the same time.

Thus, the RADARS model implies that an absolute sequence space is reduced to a thermodynamically viable structure space by means of a network of interactions, which control B-cell development. 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.

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 1011, 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 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 as 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 antigen concentration in the host, operating over several orders of magnitude(6). Because this specificity is genetically coded in the individually rearranged 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.

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 is around 14-50 amino acids (IMGT definition (20)). 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 binders 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 by 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.

2. Antibody clonal network 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 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. As the figure suggests, it is impossible to visualize high-dimensional protein space in 2D. Exponential growth is incredibly fast, leading to the generation of vast amounts of 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 (Figure 1B). 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 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 space 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 ΔG° or as equilibrium constant K of binding. These two are related by the equation

$$\Delta G^\circ = -RT\ln K \quad (1)$$

alternatively

$$K = \exp(-\Delta G^\circ / RT) \quad (2)$$

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

$$\Delta G^\circ = F_b - F_f \quad (3)$$

where Fb is the free energy of the bound (also called native) and Ff is the free (non-native) form of the antibody; ΔG° 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, translating into a better fit between antibody and target.

By starting to increase 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 area, the distinction between all possible shapes at high resolution would require a multidimensional space. Let us gradually generate a multidimensional 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 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 are also infinite if we go

far enough. The other way around, higher resolution of molecular recognition is required to distinguish two similar antigenic structures. Practically, considering a reversible interaction, the resolution is limited by the binding energy of reversible interactions.

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 evolution into this interaction space. We shall assume that a common lymphoid progenitor 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(26,45).

A differentiation step that reduces this total potential introduces restrictions in structural space. 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 populate the structural space and share this space according to the availability and in the 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 resolutions 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 structural resolution. 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 level of resolution 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 (46). 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).

To allow for an even balance between differentiation and division we can employ a Fibonacci branching scheme in our model: every division yields one cell that divides again and another one that differentiates, changing its BCR (Fig. 2C). Assuming equal time and energy investments for division and differentiation this scheme results in a Fibonacci series number of cells with distinct BCR after every time unit. In our model dividing cells remain on the same level of resolution, differentiating cells move outwards if interaction resolution grows, in a hierarchy similar to that proposed by Derényi&Szőllősi (47). To maintain constant cell numbers at all levels of differentiation the ratio of differentiation rates is proposed to be determined by ϕ, the number approached by Fibonacci series. Thus, on a systemic level, for every new incoming cell there will be one differentiated cell entering the next level, along with a differentiated cell already dividing on the next level (Fig. 2C inset). In our model this also means a ϕ^2 increase in surface area to be filled with differentiated cells continuously replenished from lower levels, along with a proportional decrease of directions (shapes) available for a given cell.

As implied above, selection in the bone marrow is a passive process: randomly generated sequences find their positions in the structural space, expanding 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 (45), and as

selection for mildly autoreactive clones to help homeostatic antibody functions and setting a reference for recognition of non-self (48).

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 takes 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. We shall call these cells effector B cells. Thymus dependent responses aided by helper T cells lead to the formation of germinal centers where affinity maturation takes place. This adjustment of affinity focuses interactions into a particular direction, corresponding to molecular shape in our model, leading to increased resolution of interaction space only in that direction. Post-germinal center B cells will have accumulated somatic hypermutations to increase their affinity. In this region of interaction space, affinity and corresponding resolution of target recognition is high, but once target is cleared cells go into dormancy. These 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) (49–51). Another type of cell that remains in the system after target has been cleared is plasma cell, which become long-lived antibody producing cells following terminal differentiation (52,53).

4. Characterization of the antibody interaction network

4.1 Distribution of binding energies in the system

The cell biological aspects of the RADARS model are summarized in Figure 3A. 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 affinity. According to the random energy model the free energy of biomolecular interactions is normally distributed(54).

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 in the host are adjusted by setting the equilibrium dissociation constants as

$[Ag]=K_D$

Substituting into eq 2),

$[Ag]=\exp(\Delta G°/RT)$

There is an exponential relationship between the adjusted free antigen concentration [Ag] and the binding free energy. 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 $2x10^{-5}$ K_D. This is in a good agreement with the observed low affinity antibody binding energy and is also the lower limit of BCR sensitivity (55).

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

$$K_{sys} = K_A/\langle K_A\rangle = e^{-(\Delta G° - \langle \Delta G° \rangle)/RT} \qquad 4)$$

where < > enclose median equilibrium constant and mean free energy of binding, respectively. K_{sys} represents the binding propensity of a given molecule in the system (Fig.3B). 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 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(1,2). The normal distribution of interaction free energy is skewed by each antigen, as explained above, producing lognormal distributions. For the

whole system we need to consider these exponentially distributed number of lognormal distributions, generated by the interaction of the system with antigens of exponentially distributed concentrations. The probability density function of K_{sys} with such distribution is given by

equation 5)

$$f(Ksys) = \int_{k=0}^{\infty} \varphi e^{-\varphi k} \frac{1}{\sqrt{2\pi k} RTKsys} e^{-\frac{(\Delta G° - k<\Delta G°>)^2}{2kRT^2}} dk$$

using a φ rate of differentiation of cells at every level of differentiation, in other words generation of diversity at rate φ. For the obtained double-Pareto distribution of Ksys, for $K_{sys}>1$ $f(K_{sys})= \varphi^2/2\ \varphi\ *K_{sys}^{-1-\varphi}$ which is a power law distribution of system affinity constants with degree exponent $\varphi-1= \varphi^2$. (Fig.4 and 5).

We have now obtained the physical chemical description of a system of antibody interactions.

4.2 A scale-free network of interactions

The systematic 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 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

$p(k) \sim k^{-\gamma}$ 6)

where γ is the degree exponent and is equal to φ^2 (Fig.5).

Both cells and secreted antibodies are continually replaced as they die or are removed along with bound target antigens. This in turn regulates the frequency distribution of antibodies along our radius of $\log K_{sys}$.

The emerging power law relationship of antibody interactions is a hallmark of scale-free networks (57). 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 number 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 (58). 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.

4.3 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 5, 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.

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.7). 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 (59), mathematical and computational simulations (60,61), 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 (62,63), 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 (64). 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 plasma cells.

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 (65), 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. Consecutive binding reactions generate a stationary non-equilibrium system. The network of interactions is scale-free and is characterized by a power law distribution of free energy of interactions. Overall, this organization allows the energy optimized controlled removal of antigens from the host system.

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.

484 References

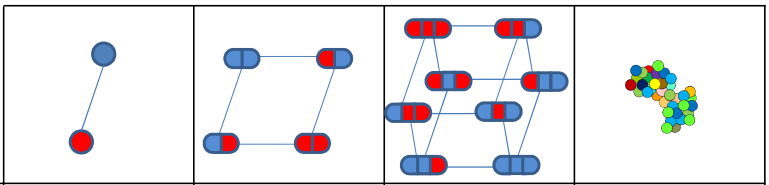
485 1. Müller V, de Boer RJ, Bonhoeffer S, Szathmáry E. An evolutionary perspective on the
486 systems of adaptive immunity. Biol Rev Camb Philos Soc. 2017 Jul 26;93(1):505–528.
487 2. Eberl G. A new vision of immunity: homeostasis of the superorganism. Mucosal
488 Immunol. 2010 Sep;3(5):450–460.
489 3. Sleator RD. The human superorganism - of microbes and men. Med Hypotheses. 2010
490 Feb;74(2):214–215.
491 4. Pradeu T, Carosella ED. On the definition of a criterion of immunogenicity. Proc Natl
492 Acad Sci U S A. 2006 Nov 21;103(47):17858–17861.
493 5. Davis MM, Tato CM, Furman D. Systems immunology: just getting started. Nat
494 Immunol. 2017 Jun 20;18(7):725–732.
495 6. Prechl J. A generalized quantitative antibody homeostasis model: maintenance of
496 global antibody equilibrium by effector functions. Clin Transl Immunology. 2017 Nov
497 17;6(11):e161.
498 7. Boyd SD, Joshi SA. High-Throughput DNA Sequencing Analysis of Antibody
499 Repertoires. Microbiol Spectr. 2014 Oct;2(5).
500 8. Vergani S, Korsunsky I, Mazzarello AN, Ferrer G, Chiorazzi N, Bagnara D. Novel
501 Method for High-Throughput Full-Length IGHV-D-J Sequencing of the Immune
502 Repertoire from Bulk B-Cells with Single-Cell Resolution. Front Immunol. 2017 Sep
503 14;8:1157.
504 9. Friedensohn S, Lindner JM, Cornacchione V, Iazeolla M, Miho E, Zingg A, et al.
505 Synthetic standards combined with error and bias correction improve the accuracy and
506 quantitative resolution of antibody repertoire sequencing in human naïve and memory
507 B cells. Front Immunol. 2018 Jun 20;9:1401.
508 10. Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V. Computational
509 Strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune
510 Repertoires. Front Immunol. 2018 Feb 21;9:224.
511 11. Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and statistical analysis of
512 adaptive immune repertoires. Trends Immunol. 2015 Nov;36(11):738–749.
513 12. Prechl J, Papp K, Erdei A. Antigen microarrays: descriptive chemistry or functional
514 immunomics? Trends Immunol. 2010 Apr;31(4):133–137.
515 13. Merbl Y, Zucker-Toledano M, Quintana FJ, Cohen IR. Newborn humans manifest
516 autoantibodies to defined self molecules detected by antigen microarray informatics. J
517 Clin Invest. 2007 Mar;117(3):712–718.
518 14. Bacarese-Hamilton T, Bistoni F, Crisanti A. Protein microarrays: from serodiagnosis to
519 whole proteome scale analysis of the immune response against pathogenic
520 microorganisms. BioTechniques. 2002 Dec;Suppl:24–29.
521 15. Wang J, Lin J, Bardina L, Goldis M, Nowak-Wegrzyn A, Shreffler WG, et al.
522 Correlation of IgE/IgG4 milk epitopes and affinity of milk-specific IgE antibodies with
523 different phenotypes of clinical milk allergy. J Allergy Clin Immunol. 2010
524 Mar;125(3):695–702, 702.e1.
525 16. Grötzinger C. Peptide microarrays for medical applications in autoimmunity, infection,
526 and cancer. Methods Mol Biol. 2016;1352:213–221.

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.

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. Melchers F. Checkpoints that control B cell development. J Clin Invest. 2015 Jun;125(6):2203–2210.
46. Prechl J. Thermodynamic projection of the antibody interaction network: the fountain energy landscape of binding. BioRxiv. 2017 Apr 5;
47. Derényi I, Szöllősi GJ. Hierarchical tissue organization as a general mechanism to limit the accumulation of somatic mutations. Nat Commun. 2017 Feb 23;8:14545.
48. Cohen IR. Biomarkers, self-antigens and the immunological homunculus. J Autoimmun. 2007 Dec;29(4):246–249.

49. Tangye SG, Tarlinton DM. Memory B cells: effectors of long-lived immune responses. Eur J Immunol. 2009 Aug;39(8):2065–2075.
50. 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.
51. 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.
52. 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.
53. 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.
54. 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.
55. 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.
56. Dorogovt͡sev SN (Sergeĭ N., Mendes JFF (José FF. Evolution of networks: From biological nets to the Internet and WWW. Paperback. Oxford: Oxford University Press; 2003.
57. Albert R, Barabási A-L. Statistical mechanics of complex networks. Rev Mod Phys. 2002 Jan 30;74(1):47–97.
58. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004 Feb;5(2):101–113.
59. 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.
60. FARMER DJ, PACKARD N, PERELSON A. The immune system, adaptation, and machine learning. Physica D: Nonlinear Phenomena. 1986;22(1-3):187–204.
61. Schulz R, Werner B, Behn U. Self-tolerance in a minimal model of the idiotypic network. Front Immunol. 2014 Mar 10;5:86.
62. Mäkelä T, Annila A. Natural patterns of energy dispersal. Phys Life Rev. 2010 Dec;7(4):477–498.
63. Hartonen T, Annila A. Natural networks as thermodynamic systems. Complexity. 2012 Nov;18(2):53–62.
64. 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.
65. 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.

A



dimensions d	1	2	3	250
values v	2	2	2	20
system size $N = d^v$	2^1	2^2	2^3	20^{250}

Figure 1. Sequence space and visualization of antibody sequence relationships
A) Theoretical diversity of a sequence is determined by its length and the number of values a particular position in the sequence can take. More than 3 dimensions are difficult to visualize in 2D. An antibody Fv region of 250 amino acids has an astronomical sequence diversity if full randomization is allowed. If the sequences are exact the positions in sequence space are discrete, but there are no structurally meaningful directions or distances in this multidimensional sequence space with N different sequences.

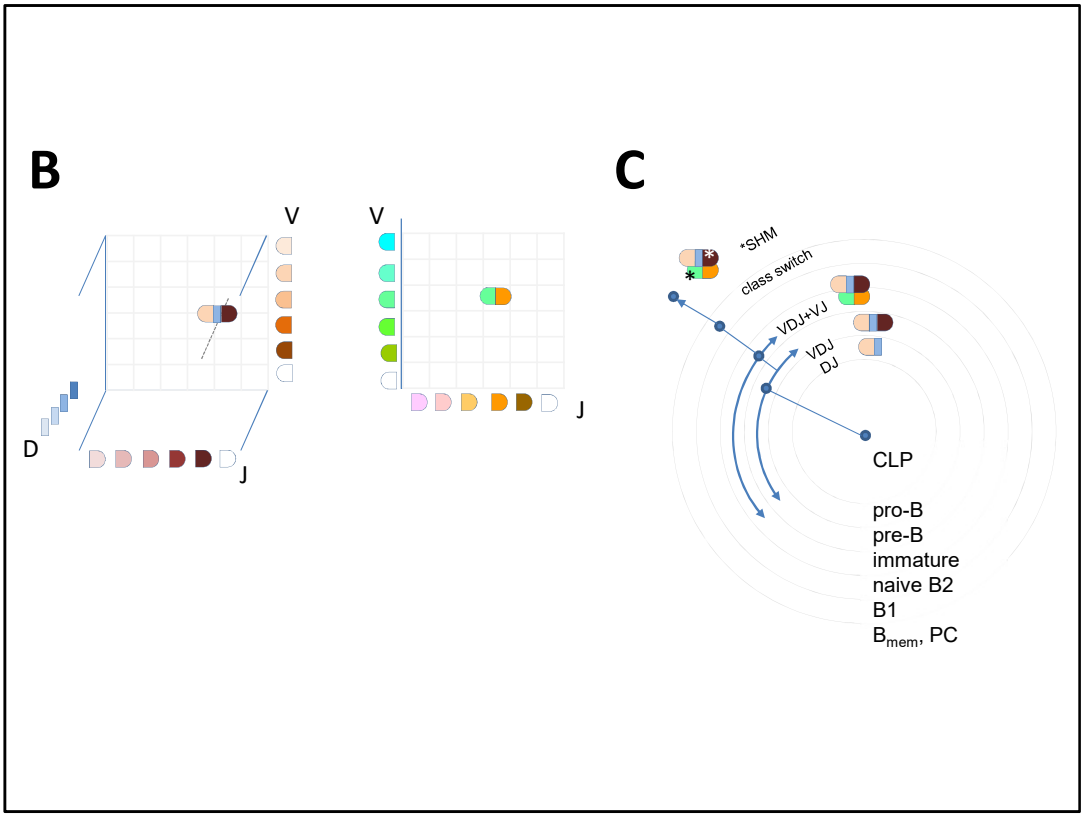


Figure 1
B) Antibody sequences are frequently interpreted as recombined germline sequences. This approach allows the simplified display of repertoires obtained by NGS, preferably with paired heavy and light chain VD identification. Such a display of combinatorial diversity may allow the tracking of specific 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.

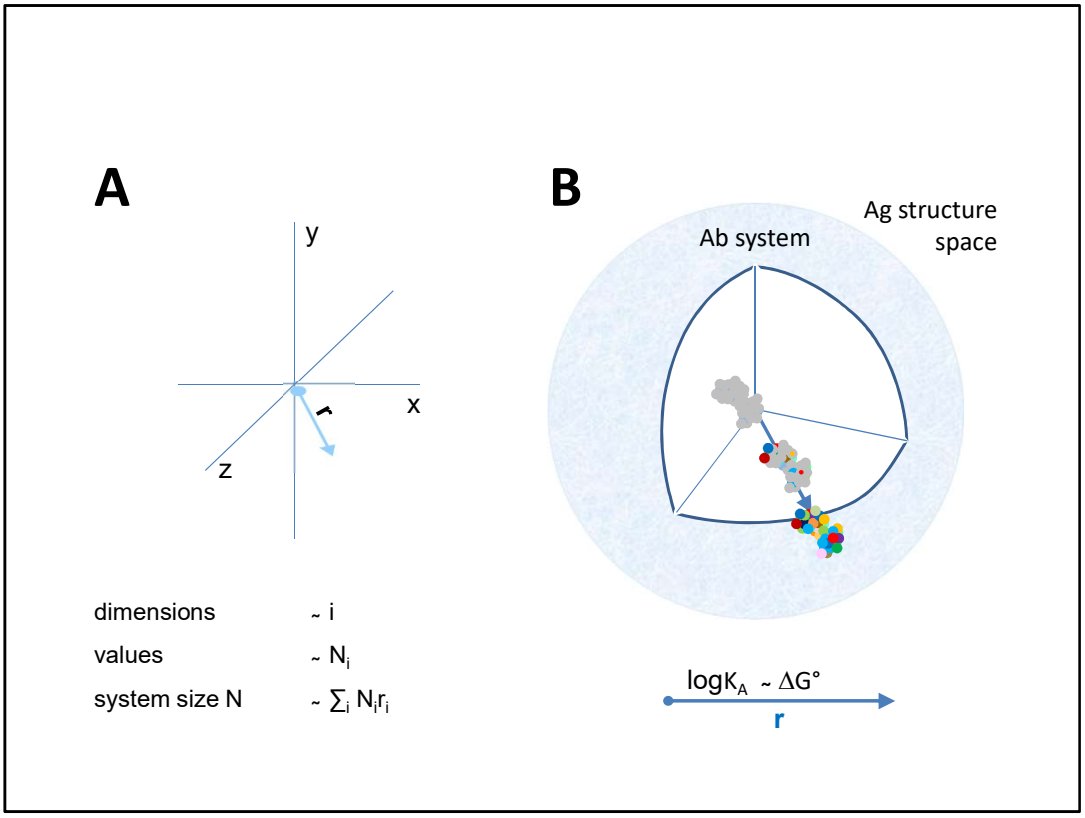


Figure 2. Quantitative interaction space and B-cell differentiation
A) The system of interactions has a center as a reference point in a conceptual three dimensional space. Structural diversity, that is different shapes, appear as directions (exemplary arrow) from this center towards the target structure in Ag structure space. Distinct directions can be defined with a precision dependent on the distance from the center, equivalent to the radius (r) of the system extending in that particular direction. Multidimensionality is theoretically infinite in this representation, practical limits being introduced by the maximum interaction energy (DG°) in the system. For i dimensions with radius ri, each with Ni number of distinct structures, the total size of the system N is the sum of structures of all dimensions.
B) Structural diversity appears as we leave the center, spherical shells representing various levels of resolution of molecular recognition, measured as $\log K_A$. Colors represent distinct amino acids of antibody binding site engaging in non-covalent bonding with the target molecule. As the Ab molecule approaches its native, bound state the number of engaged residues increases.

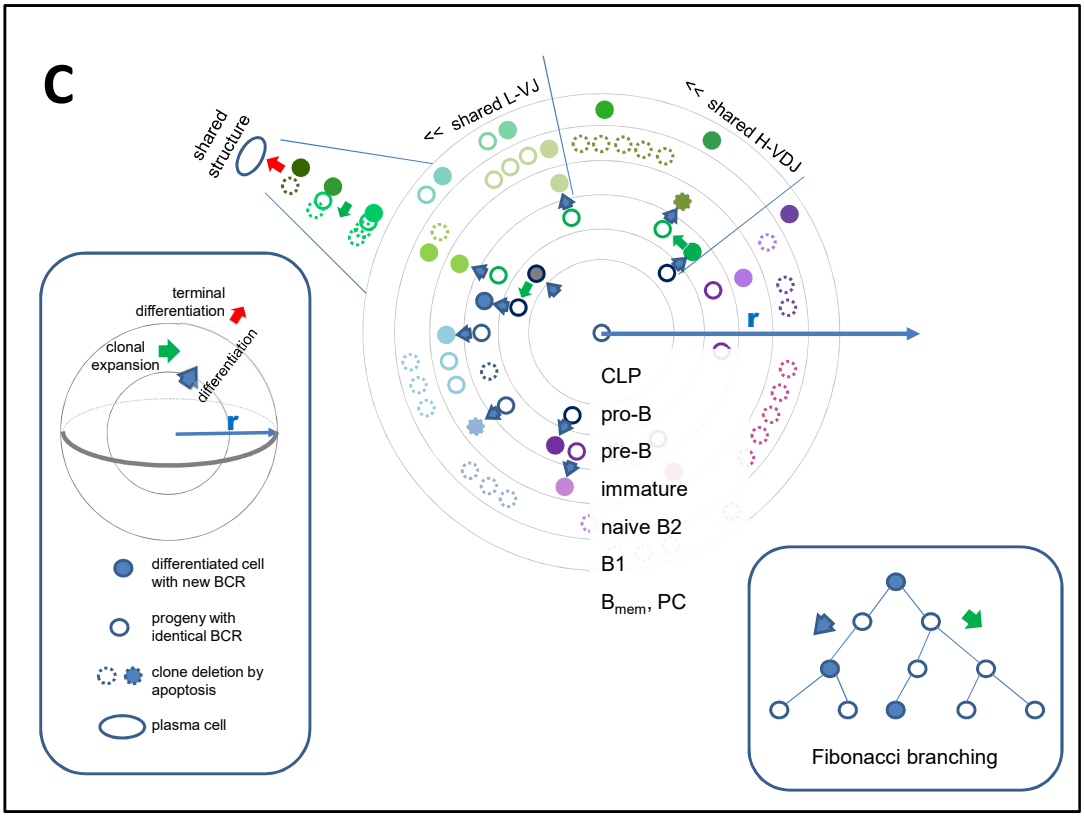
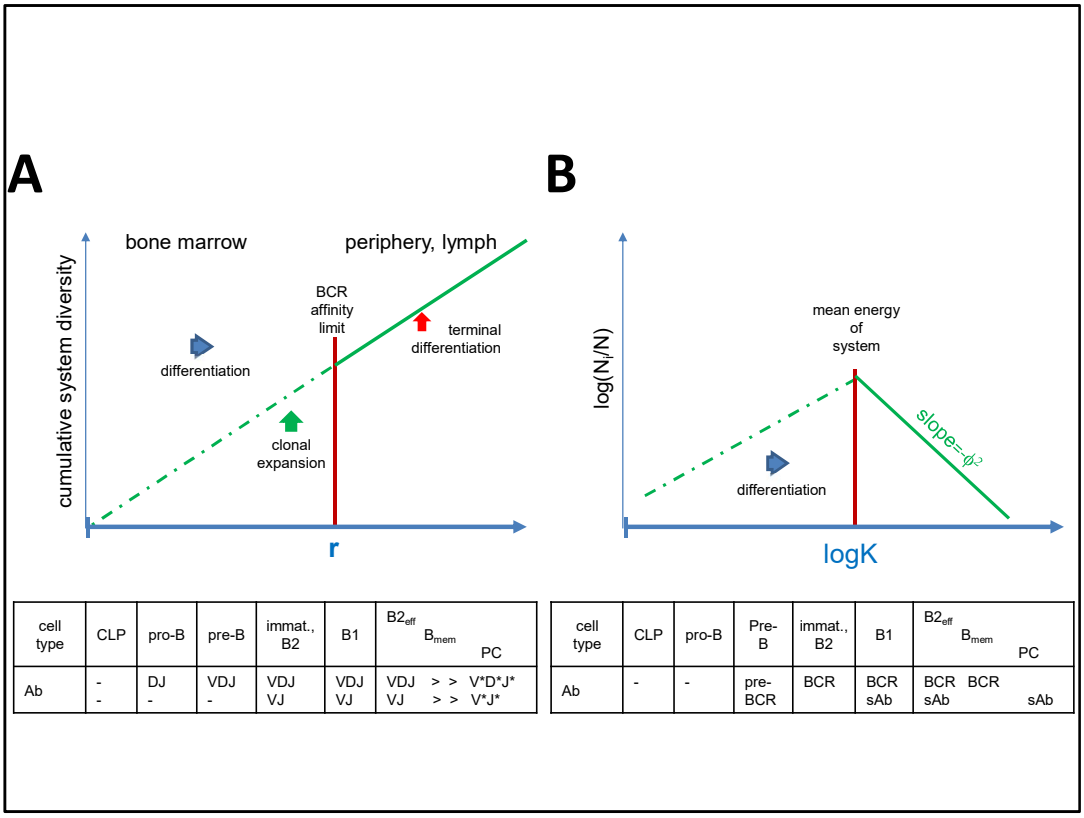


Figure 2.
C) Systemic organization of antibody evolution.
The evolution of the system of antibodies can be interpreted as clones filling the interaction space at various levels of resolution. Along this pathway cells continually increase their specificity and affinity towards their target direction. 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 asymmetric divisions fill the level with a clone, expansion being dictated by antigen driven selection of clones with ideal receptor engagement. TD responses allow further directed differentiation via somatic hypermutations in germinal centers, yielding post-germinal center B cells, that is memory B cells and plasma cells. 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. Fibonacci branching of dividing and differentiating cells defines ratios of cells on various levels. Different colors stand for structural differences and relationships.



cell type	CLP	pro-B	pre-B	immat., B2	B1	$B2_{eff}$ B_{mem} PC
Ab	- -	DJ -	VDJ -	VDJ VJ	VDJ VJ	VDJ > > V*D*J* VJ > > V*J*

cell type	CLP	pro-B	Pre-B	immat., B2	B1	$B2_{eff}$ B_{mem} PC
Ab	-	-	pre-BCR	BCR	BCR sAb	BCR BCR sAb sAb

Figure 3. Predicted interaction properties of the humoral immune system.
A) Clonal network development of the humoral immune system, shown as cumulative distribution of evolving antibody clones. The self-organizing system in the bone marrow utilizes recombination of gene segments (V,D,J) and meets environment in the periphery to maintain immunological self and generate immunological antibody profile, here also employing selected somatic mutations (*). ‚r' represents the radius of the system. B) Distribution of equilibrium constant of antibody interactions in the system. *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

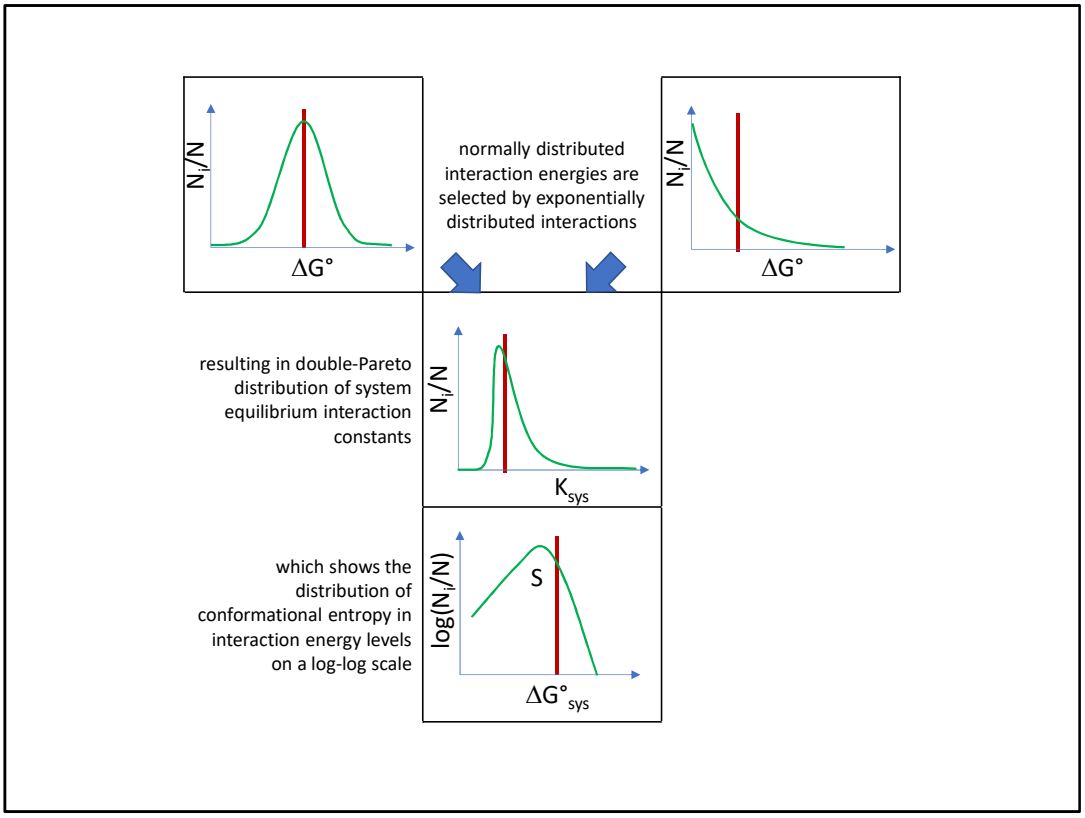


Figure 4. Combination of distributions.
Red lines represent the expected values of the distributions.

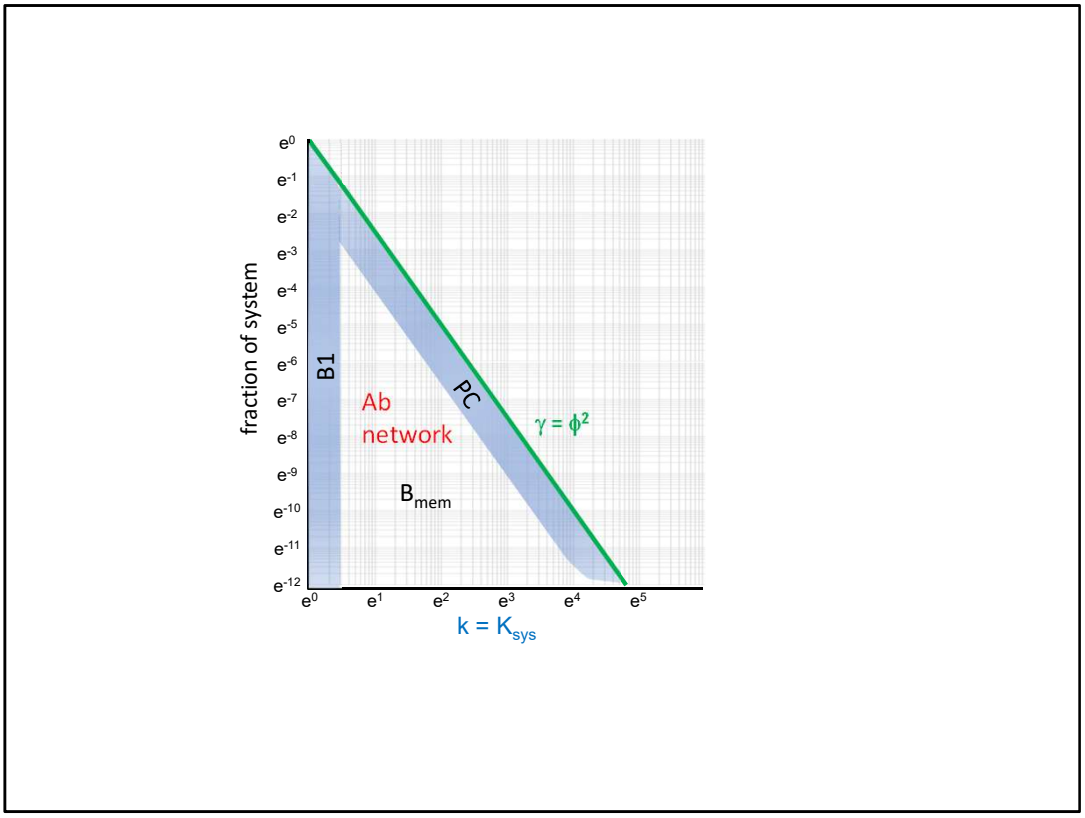


Figure 5. Properties of the antibody interaction network.
Degree distribution of the antibody network corresponds to the probability distribution of relative free energy of antibody interactions, where the system equilibrium constant is the degree of a B-cell clone node and Ni/N is relative frequency of nodes with that degree in the interaction network. Note that this figure corresponds to the lower right quadrant of Figure 3B, except that not the axis values but the scale is logarithmic. B1 cells have the lowest degree, plasma cells have the highest energy at a particular frequency, memory B cells form the rest of the nodes.

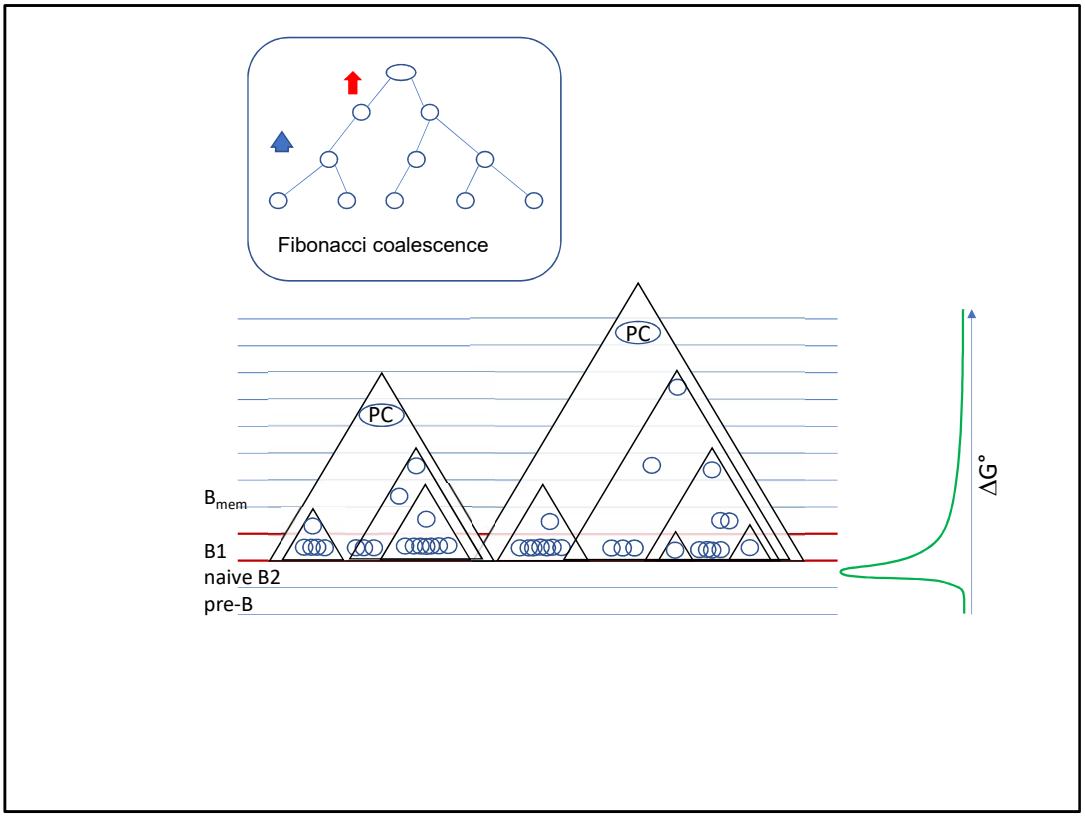


Figure 6. Network organization of antibody responses
Assuming that the rate of coalescence of structural diversity is identical to the diversification, on a systemic scale it will be phi. In subnetworks plasma cells act as hubs, being connected to all memory B cells and B1 cells with shared structures at lower free energy levels. By secreting antibodies, plasma cells control activation of all connected cells. Triangles represent subnetworks, blue lines indicate energy levels.

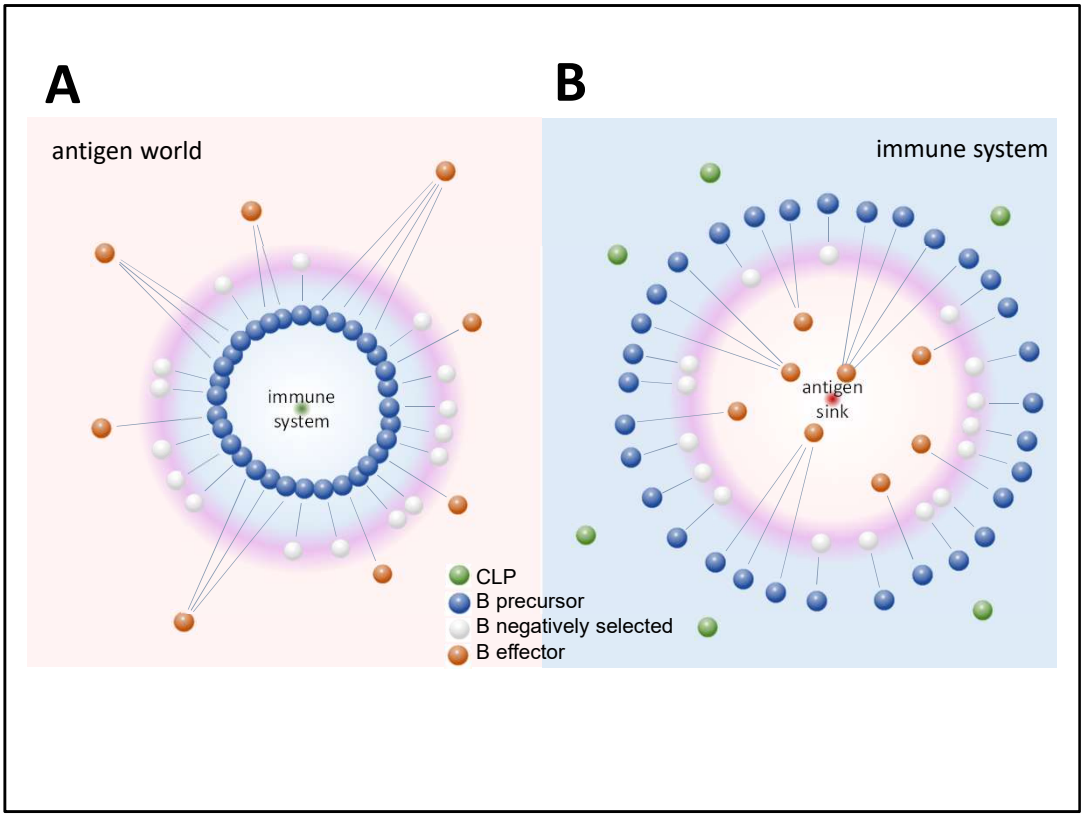
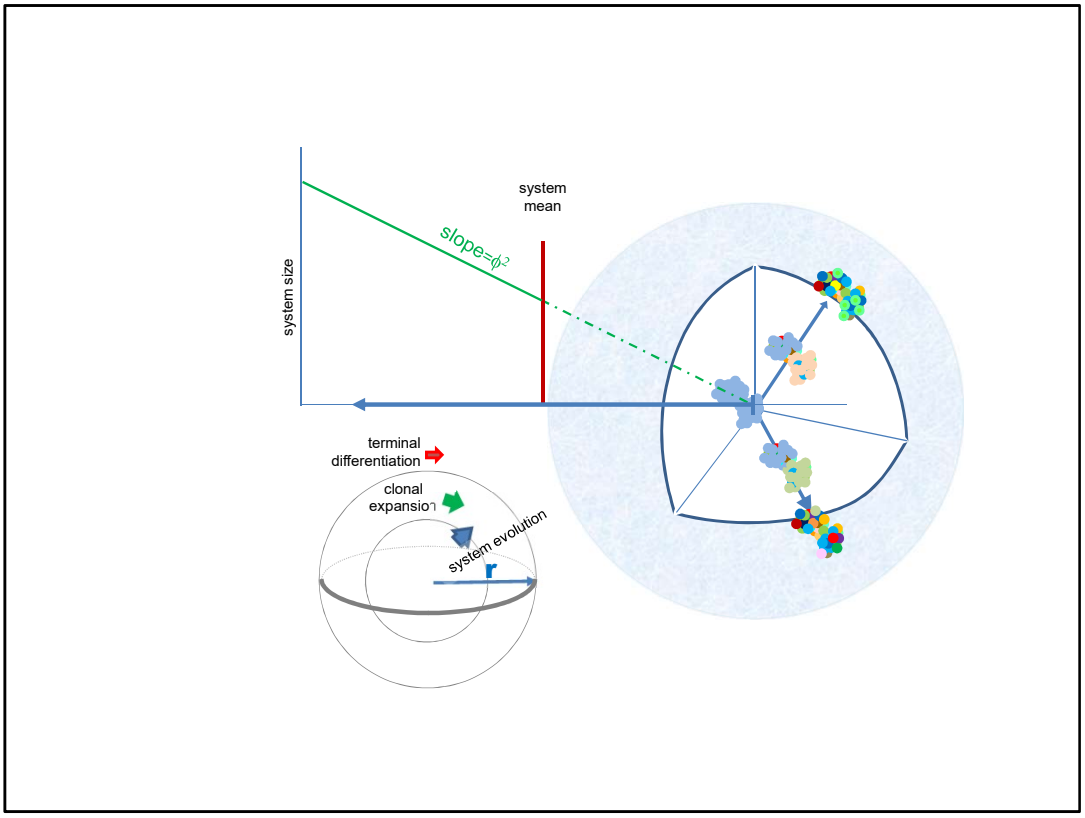


Figure 7. Antibody-centric and antigen-centric views of the organization of adaptive humoral immunity.
The antibody-centric view (A) corresponds to the clonal development and expansion of B cells, generating and maintaining a spherical system with a radius of $\ln K_{sys}$. The system carves out a niche 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 $\ln(1/K_{sys})$. In this representation the development of antibodies with higher affinity pull antigen down an antigen sink with increasing efficiency, capturing antigen of a wider structural diversity.



Graphical abstract