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

Abstract

Adaptive immunity in vertebrates represents a complex network of protein interactions that develops throughout the lifetime of an individual. Based on the key molecule of humoral adaptive immunity, the antibody, development of this system comprises molecular genetic, cell biologic and immunologic mechanisms, and as a network the system is likely governed and can be characterized by physical rules as well. While deep sequencing can provide vast amounts of data related primarily to 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 structural interaction space, termed radial adjustment of system resolution, or RADARS, is proposed. The model is based on the radial growth of resolution of molecular recognition, corresponding to increasing affinity of immune reactivity, and the virtual infinity of directions of growth, corresponding to the ability to respond to almost any molecular structure. Levels of interaction strength appear as shells of the sphere representing the system. B-cell development and immune responses can be readily interpreted in the model and quantitative properties of the antibody network can be inferred from the physical properties of a quasi-spherical system growing multi-radially. The system is described by a type of power-law distribution, the double-Pareto distribution, sampling the lognormally distributed binding equilibrium constants at a rate of phi square. Understanding such quantitative network properties of the system should help the organization of sequence-derived structural data, while the proposed physical network properties are directly related to function. Finally, general strategies for merging antibody sequence space into structural space are also outlined.

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. 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. 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 (6–10).

It is also possible to profile the antibody repertoire functionally, based on the identification of antibodies binding to huge sets of potential targets (11,12). 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 (13–17). 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 (18). 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 (19)). 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 (20) 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 (21). The development of concepts of the immune system as a network were key steps in our current perception of immunity (22,23). Efforts are now under way to describe the immune system as a network (termed network systems immunology) using NGS data and network science (24). 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 (25–27), 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 (18). 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 (7,9,38–40).

3. Antibody interaction network 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 sequencespace. 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 Ångstrom 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 good predictor of binding energy of protein interactions (43). 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 or as equilibrium constant of binding. 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 (27). 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.

Going in the opposite direction, by decreasing the resolution of interaction, we can reach a theoretical minimum. At this minimum all interactions are defined by a single identical quantity of interaction strength. Let this quantity be 1, so that an equilibrium constant with the value 1 can represent the lowest possible resolution. At this minimal resolution binding to any target is possible and can be interpreted as minimal random contact upon collision of molecules. This minimal resolution characterizes the center of the system, represented by the center of interaction space in Figure 2. 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 immense 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 extent of precision by which we can define the direction. Thus, going away from the minimal resolution of 1 we can define shape at gradually higher resolutions, corresponding to larger BSA and binding energies (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 round, 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. Beyond this limit structure of the binding partners becomes fixed, rigid, corresponding to covalent binding.

This model of the evolution 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 a virtual image of the surrounding. 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 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. A differentiation step that reduces this total potential introduces restrictions in structural space. This will appear as an expansion along a radius that point towards the cognate target. 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 of target. Cells with more abundant targets expand more, cells with less frequent targets remain in lower numbers, until optimal BCR engagement is achieved (25). As a result, structural space as represented by this resolution will be filled with different pre-B cell clones according to the availability of the common self-antigens.

The next level of structural resolution, as well as the next level of restriction of binding 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 (25), or an intrinsic ability of the BCR of sustained signaling, perhaps due to structural properties (44). 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 a cell that can divide further and another one that differentiates (Fig. 2C). Assuming equal time and energy investments for division and differentiation this scheme results in a Fibonacci series number of cells after every time unit. In our model dividing cells remain on the same level of resolution, differentiating cells move outwards if interaction affinity grows. The scheme is modified by the selection steps though, dictated by the continuous selection of BCR reactivity against target. As a result, the distribution of cells in interaction space will be determined by the rate of division and differentiation along with random sequence generation pruned by interaction-based selection events.

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 fitting target structures. As a result, the emerging population of immature B cells will bear the 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 (46).

3.3. Immune response 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. 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, shape in our model, leading to increased resolution of interaction space only in that direction. 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 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). Another set of cells that remains after target has been cleared is plasma cells, which presumably arise from effector cells finding appropriate niche for long-term survival.

4. Characterization of the antibody interaction network

What is the use of devising a theoretical space for the description of antibody network development in the host? Antibody interaction space might be small enough to be handled by today's databases and tools. With the growth of interaction databases and structural biological data we might simply merge these sets of data and be able to predict binding specificity and affinity from sequence, so why introduce another model?

The main reason could be the rational organization of all those data. By following the natural flow of events of antibody evolution, by understanding the natural organization of sequences we should be able to simplify and rationalize our databases and tools. The model presented here can be developed into a quantitative network and mathematical tools applied for the analysis of the system.

The 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. Events in the bone marrow represent self-organization because of control mechanisms involved, events in the periphery are more stochastic because interactions with the outside molecular world takes control. To describe the network of antibody interactions we follow convention and generate a diagram with the most central node to the rightmost position on an axis representing connectivity of nodes (Fig. 3B). Thus, CLPs, cells with the potential to generate all further B-cell developmental forms are regarded as the central node of the system. The radius of the network grows leftwards in this

representation, corresponding to decreasing K_D and free energy changes, that is increasing affinity and interaction strength.

Deduction of network properties requires the knowledge of distributions of nodes and their connectivity. We shall make two assumptions here regarding these parameters.

Universal laws governing the distribution of binding energies for biomolecular recognition have been uncovered by Zheng and Wang (47). Based on energy landscape theory they suggest that binding free energy follows a Gaussian distribution, while the equilibrium constant K_A follows a log-normal distribution, accordingly. In our model distance from the center represents resolution of molecular recognition, which was defined as K_A above, and directions 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. Mean binding energy, the energy gap between native configuration and the average unbinding states in the universal binding energy distribution model, corresponds to the average binding energy of antibodies. This value is around -30kJ/mol and is generally regarded as the lower limit of BCR sensitivity (48).

The RADARS model suggests that the greater the resolution of structural recognition the more restricted is the number of shapes recognized. The further the system grows in a given direction the more focused is the recognition. We proposed above that a balanced growth and maintenance of the system is only possible if cellular division and differentiation rates are evenly balanced. Here it is further proposed that the generation of a network with such radially adjusted resolution is possible if rate of node generation is defined by the golden ratio, phi ($\varphi \approx 1.618$). This is a consequence of the Fibonacci branching scheme of cell division and differentiation, shown in Figure 2. By increasing radius length phi times, the surface area of the sphere, representing the next level of differentiation and the next level of interaction strength, increases proportionally to phi square. Since phi square is equal to phi plus one (one of the special properties of phi), it is possible to promote cells from the previous interaction level (r x 1) and add new cells in proportion to the rate of increase (r x phi). A balance between exponential growth of cells (nodes) and of radius is established, at the same time the number of links (clonal relationships) is determined.

The combination of a lognormal process and sampling at an exponential rate is defined by the double-Pareto distribution, as suggested by Reed (49) and further developed by Mitzenmacher (50). Based on these mathematical approaches, probability distribution function of the antibody interaction network according to the RADARS model is given by:

$$F(r)= \frac{phi^2}{2\text{phi}} r^{\text{phi-1}} \quad , \text{for } 0 < r \leq 1$$

$$F(r)= \frac{phi^2}{2\text{phi}} r^{\text{-phi-1}} \quad , \text{for } r > 1$$

where r is the radius of the system, the equilibrium dissociation constant normalized to system size and binding energy distribution. Here we shall call this the system equilibrium distribution constant

$$K_{sys}= e^{-\frac{\Delta G^\circ - \Delta \overline{G}^\circ}{RT}}$$

where ΔG° is the standard free energy change, ΔG° with bar is the mean of standard free energy changes in the system, R is the universal gas constant and T is temperature. K_{sys} describes the molecular behavior of a component of the system based on the relative affinity of the component compared to the mean affinity of the system, and stands for the probability of the component to be free or bound in the system (Fig 3B). Further detailed explanation and analysis of the network and physical characteristics of the RADARS model are given in an accompanying paper (manuscript in preparation).

The network properties implied by the model are supported by both theoretical and experimental approaches. A recent analysis of the architecture of cell differentiation hierarchy in self-renewing tissues proposes that the ratio of differentiation rates is identical for all levels in the hierarchy and that it has a minimum value of 2 (51). This corresponds to the power law behavior of B-cell development network with a degree exponent of -phi-1 = -(phi+1)= - phi^2 ≈ -2.618. Our model establishes a balanced load of cellular division and differentiation, rather than minimizing divisional load, which results in a higher ratio of division rates and a concomitant higher ratio of cells at increasing levels of hierarchy. On the other side of the distribution, representing plasma cell generation and antibody production, division rates show an inverse relationship. Indeed, rather than expanding, cells here differentiate into long lived plasma cells once the effector phase of immune response is over. The network architecture of antibody repertoires was recently computed based on high-throughput sequencing data from more than 100.000 unique antibody sequences (24). 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 two

sides of the double-Pareto distribution in our model. The low-affinity side (Fig. 3B) with developing B-cells (Fig. 3A) is a scale-free network with power-law degree distribution, homogenously interconnected by clonal relationships and shared usage of gene segments (Fig.2). The high affinity side of the distribution is the narrowing, focusing interaction space of plasma cells with an inverse power law behavior (exponent is phi-1= 1/phi ≈ 0.618). The immunological implication of the RADARS model is that the naive repertoire generates a low affinity image of the human antigenome available in the bone marrow environment. Such an image has been called the immunological humunculus, as an analogy to the representation of human body in the brain cortex (52). While this property of the repertoire often leads investigators to conclude that autoreactive antibodies are enriched in the naive and B1 repertoire, the practical use is quite obvious, as has been suggested by Irun Cohen (46). Telling the difference between molecular structural self and non-self can be quite easy is the difference is striking. For identifying small differences, it is useful to have a starting repertoire that represents self and then increase resolution only in regions where encountered dangerous structures show up in structural space. The importance of self-recognition has also been recognized in antibody-mediated physiological functions (27,53).

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 (22), its modified version leading to the concept of clonal selection (54), mathematical and computational simulations (55,56), various re-interpretations (23), experimental approaches using NGS (24,31) or antigen microarrays (21) 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 (57,58), therefore network theories should consider energy transduction in the system. Second, it proposes a value for the power-law relationship of division rates between levels of differentiation, and an architecture for the whole network, characterized by the double-Pareto distribution.

Considering that our technological capacity 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 the RADARS model may help in the final integration of data, as summarized in Table 1.

The model also suggests sequence-based network properties of early B-cell developmental stages also need to be determined (24), in addition to the mature and antigen-experienced repertoire (59), and comprehensive and selective analysis of the B1 repertoire is very important for capturing network properties of the system.

These developments should help understand how antibody sequence space fits into structural space and into an interaction network, potentially leading to the ability to model whole immune systems and simulate their function.

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 grant to Diagnosticum. I wish to thank István Csabai for reading and discussing the manuscript and the referees for their constructive criticism.

Author contributions

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

Conflicts of interest

The author declares no conflict of interest.

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. Boyd SD, Joshi SA. High-Throughput DNA Sequencing Analysis of Antibody Repertoires. Microbiol Spectr. 2014 Oct;2(5).
7. 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.
8. 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.
9. 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.
10. Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and statistical analysis of adaptive immune repertoires. Trends Immunol. 2015 Nov;36(11):738–749.
11. Prechl J, Papp K, Erdei A. Antigen microarrays: descriptive chemistry or functional immunomics? Trends Immunol. 2010 Apr;31(4):133–137.
12. 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.
13. 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.
14. 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.
15. Grötzinger C. Peptide microarrays for medical applications in autoimmunity, infection, and cancer. Methods Mol Biol. 2016;1352:213–221.
16. 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.

17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. Jerne NK. Towards a network theory of the immune system. Ann Immunol (Paris). 1974 Jan;125C(1-2):373–389.
23. Coutinho A. The network theory: 21 years later. Scand J Immunol. 1995 Jul;42(1):3–8.
24. 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;
25. 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.
26. 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.
27. 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.
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 Feb;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 Jul;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. Prechl J. Thermodynamic projection of the antibody interaction network: the fountain energy landscape of binding. BioRxiv. 2017 Apr 5;
45. Melchers F. Checkpoints that control B cell development. J Clin Invest. 2015 Jun;125(6):2203–2210.
46. Cohen IR. Biomarkers, self-antigens and the immunological homunculus. J Autoimmun. 2007 Dec;29(4):246–249.
47. 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.
48. 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.

49. 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.
50. Mitzenmacher M. Dynamic models for file sizes and double pareto distributions. Internet Math. 2004 Jan 1;1(3):305–333.
51. 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.
52. Cohen IR, Young DB. Autoimmunity, microbial immunity and the immunological homunculus. Immunol Today. 1991 Apr;12(4):105–110.
53. Lobo PI. Role of Natural Autoantibodies and Natural IgM Anti-Leucocyte Autoantibodies in Health and Disease. Front Immunol. 2016 Jun 6;7:198.
54. 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.
55. FARMER DJ, PACKARD N, PERELSON A. The immune system, adaptation, and machine learning. Physica D: Nonlinear Phenomena. 1986;22(1-3):187–204.
56. Schulz R, Werner B, Behn U. Self-tolerance in a minimal model of the idiotypic network. Front Immunol. 2014 Mar 10;5:86.
57. Mäkelä T, Annila A. Natural patterns of energy dispersal. Phys Life Rev. 2010 Dec;7(4):477–498.
58. Hartonen T, Annila A. Natural networks as thermodynamic systems. Complexity. 2012 Nov;18(2):53–62.
59. 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.

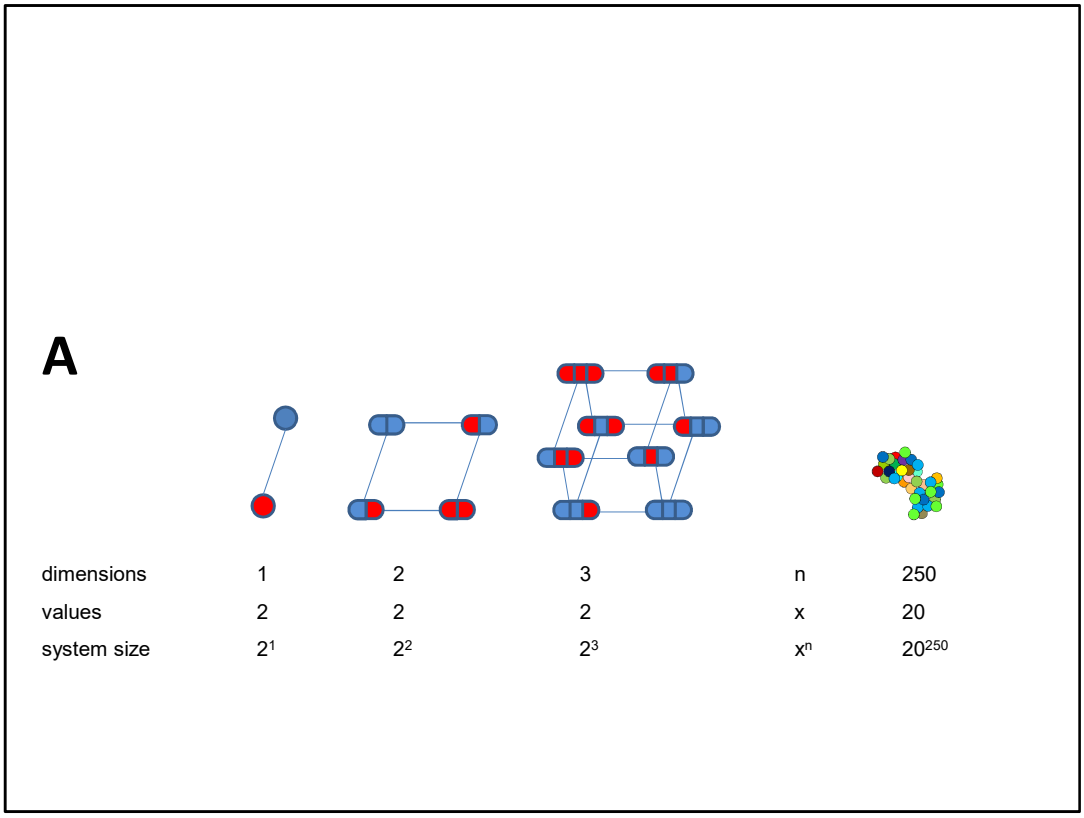


Figure 1. Sequence space and visualization of antibody combinatorial relationships
A) Theoretical diversity of a sequence is determined by its length and the number of values a particular position in the sequence can take. More than 3 dimensions are difficult to visualize in 2D. An antibody Fv region 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 the multidimensional space.

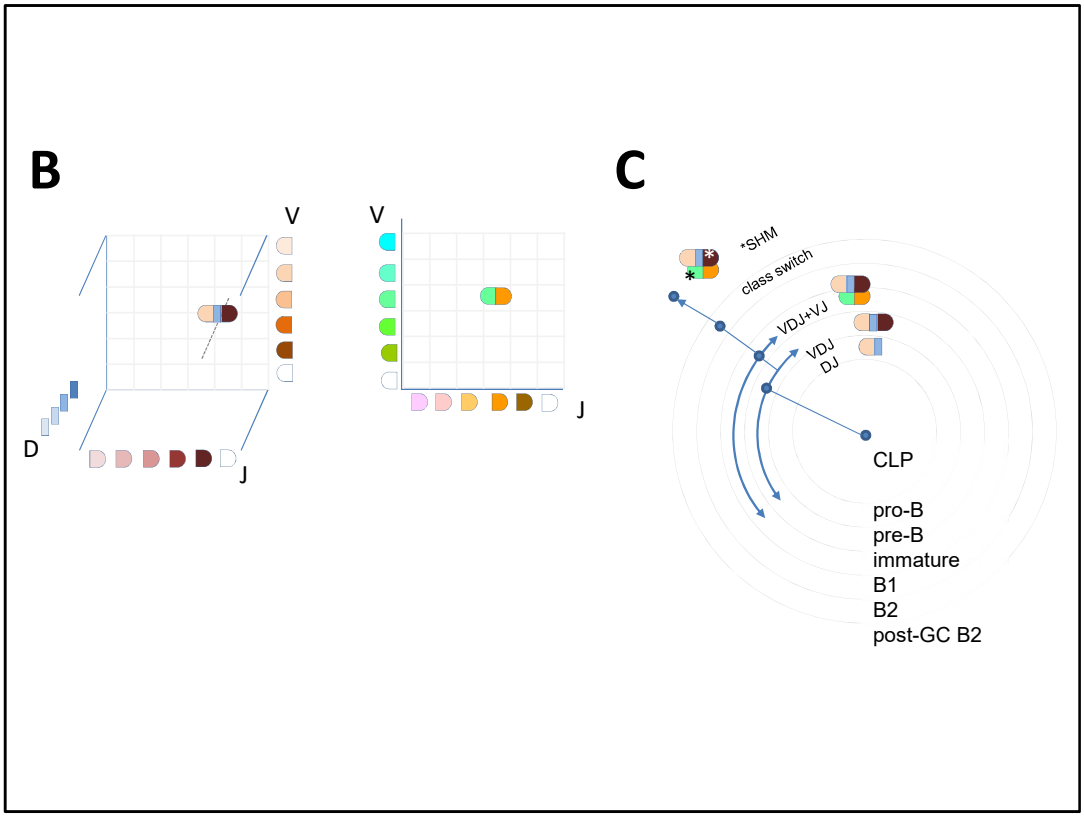


Figure 1
B) Antibody sequences are frequently interpreted as recombined germline sequences. This approach allows the simplified display of repertoires obtained by NGS, preferably with paired heavy and light chain VD identification. Such a display of combinatorial diversity may allow the tracking of specific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

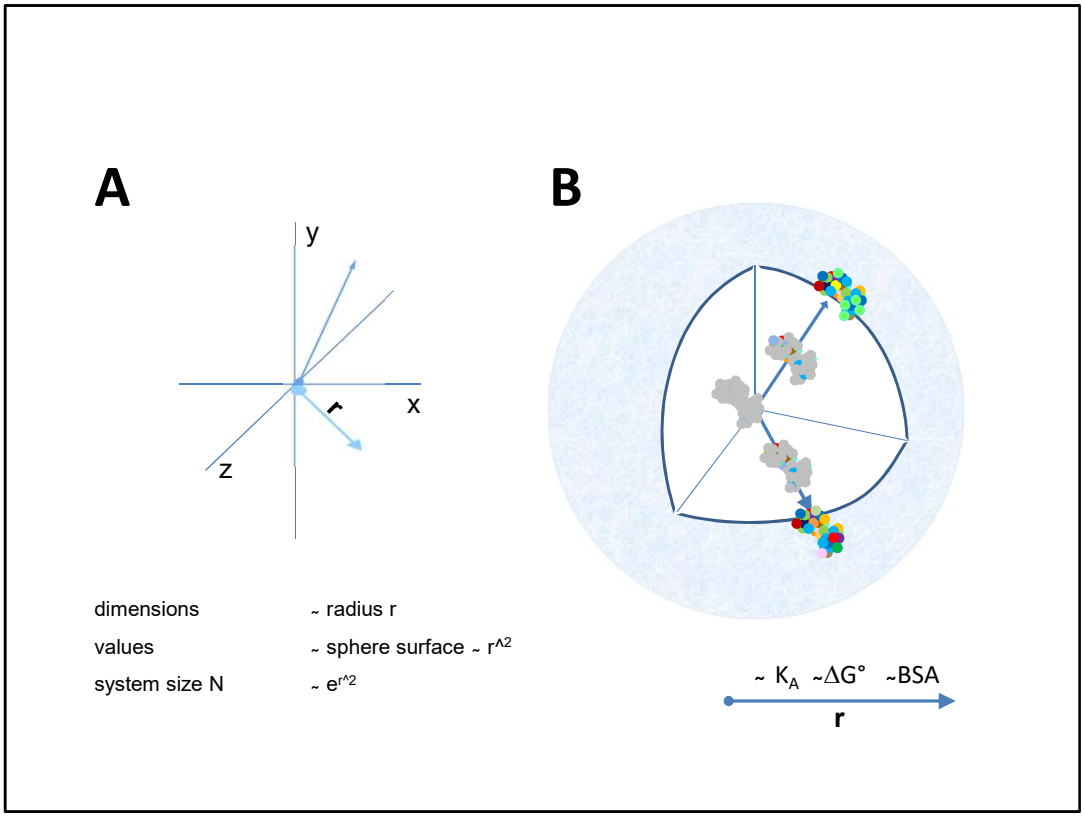


Figure 2. Quantitative interaction space and visualization of antibody network relationships

A) The system of interactions has a center as a reference point in a virtual three dimensional space. Structural diversity appears as directions (two arbitrary arrows) from this center towards the target. Distinct directions can be defined with a precision dependent on the distance from the center, equivalent to the radius (r) of the sphere representing the system. The longer the length of a vector defining a direction, the higher the resolution of the system in that direction. Multidimensionality is theoretically infinite in this representation, practical limits being introduced by resolution.

B) Structural diversity appears as we leave the center, 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.

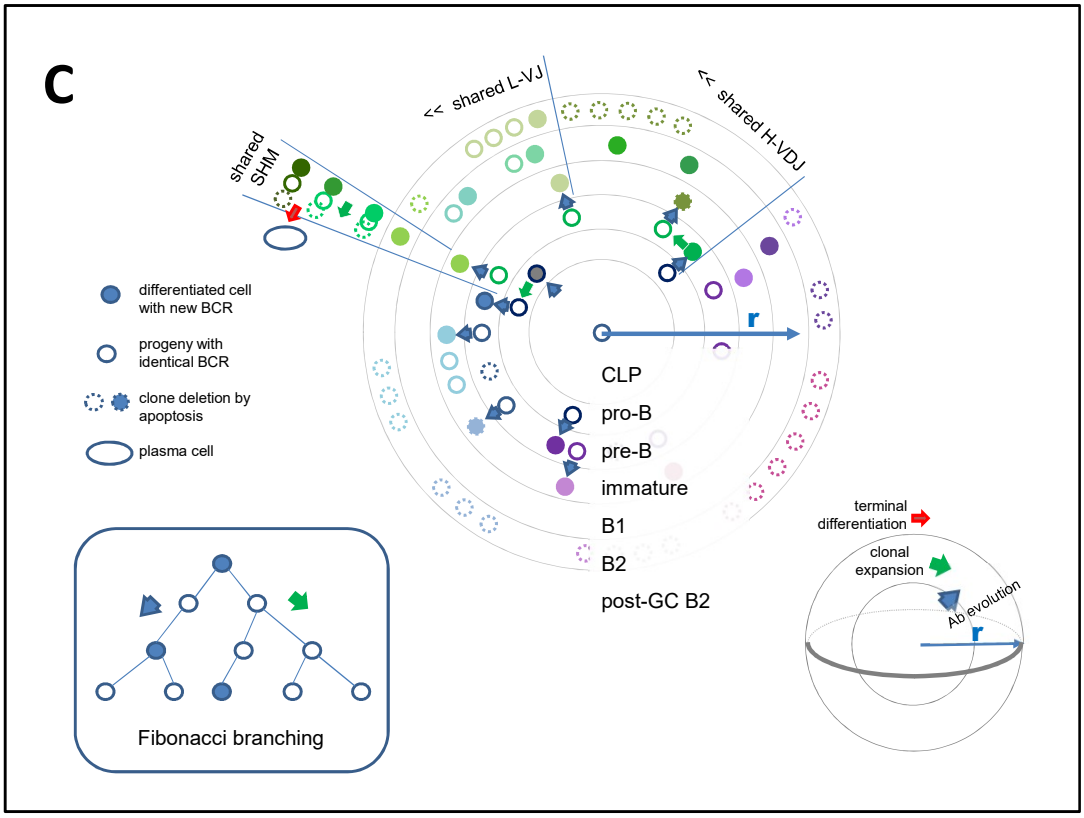


Figure 2.
C) Systemic organization of antibody evolution.
The evolution of the system of antibodies can be interpreted as clones filling the interaction space at various levels of resolution. Along this pathway cells continually increase their specificity and affinity towards their target direction. A common lymphoid progenitor has the potential to develop antibody against any target, equivalent to a radius of 1. Differentiation proceeds outwards from this point of origin, the center of the network. 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. Accordingly, pre-B cells expand based on the ability of rearranged HVD to pair with SLC and signal for clonal expansion. The generated cells rearrange their LVD and exit bone marrow. Primary TD responses and TI responses expand B2 or B1 cells without further evolution of antibody binding. TD responses allow further directed differentiation in germinal centers. 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. Effector cells from clonal expnsions can establish long lived plasma cells if they arrive in the required niche.
Different colors stand for structural differences and relationships.

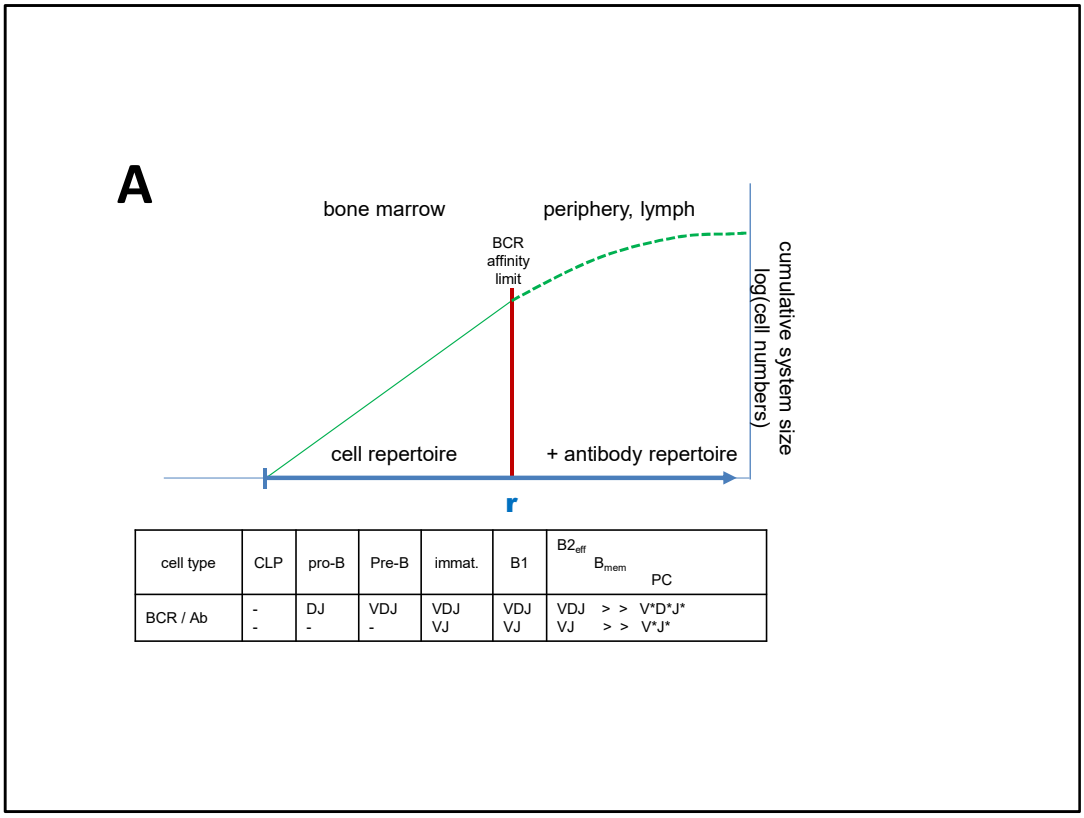


Figure 3. Predicted network properties of the humoral immune system.
A) Biological network structure of the humoral immune system, shown as a cumulative distribution function of evolving antibody clones. Self-organizing system in the bone marrow meets environment in the periphery to maintain immunological self and generate immunological antibody profile. ‚r' represents the radius of the system.
*somatic hypermutations; CLP, common lymphoid progenitor; immat., immature; B2eff, effector B2 cell; Bmem, memory B cell; PC, plasma cell

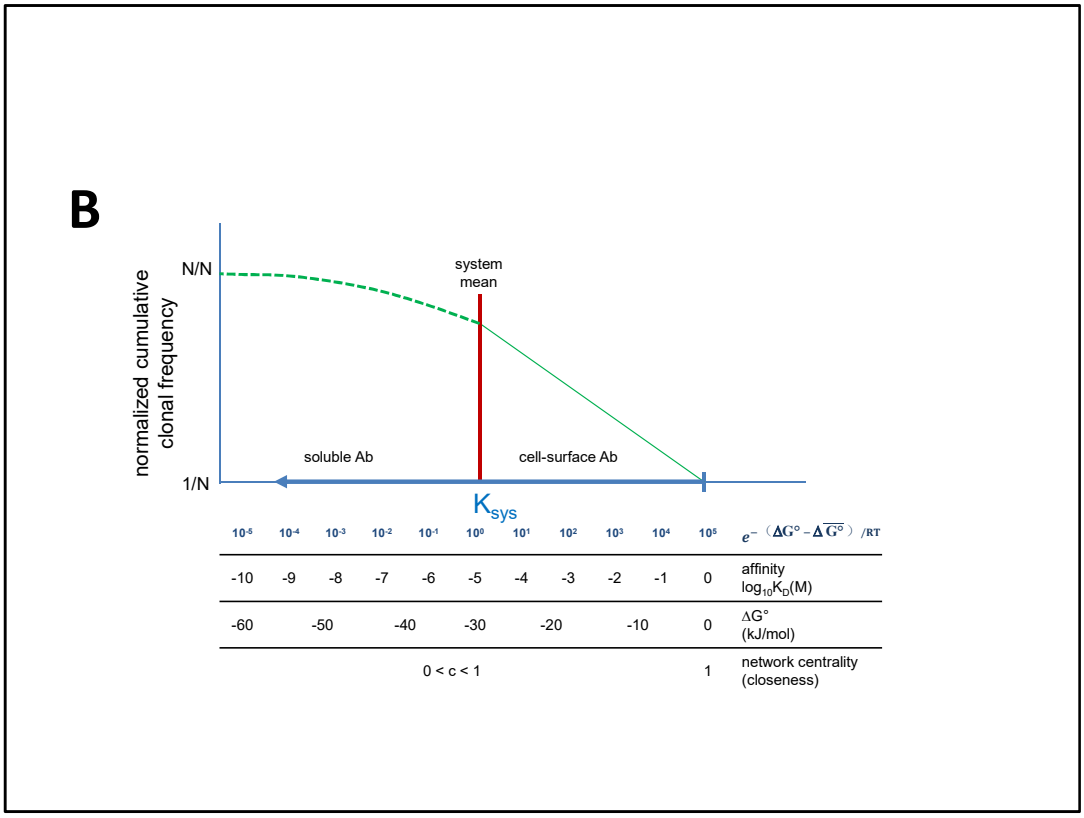


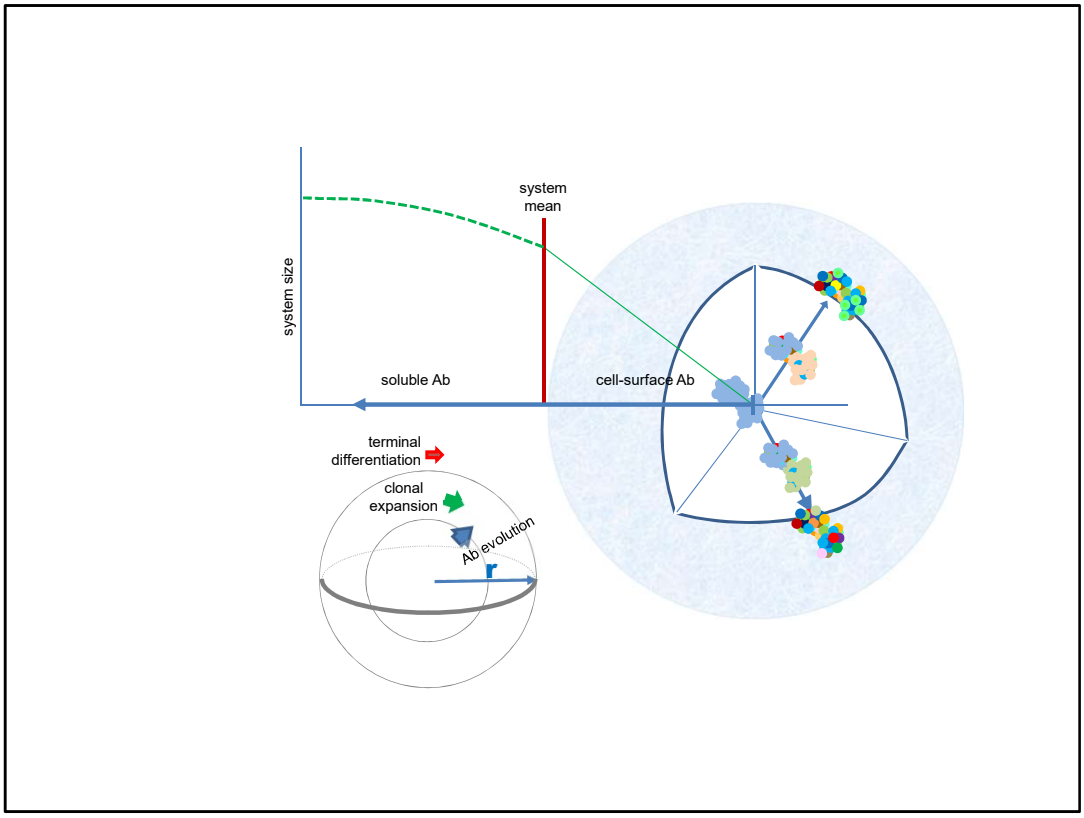
Figure 3.
B) Complementary cumulative distribution of antibody clones as a function of equilibrium distribution, affinity, free energy of binding and network centrality. Based on the radial growth interaction network model the network structure of the humoral immune system can be inferred. Network nodes are functionally identical B cells, identically extending into the direction of cognate target structure but may comprise distinct sequences, as long as binding is identical. Network centrality was defined as normalized network closeness of a complete graph: the reciprocal of the average length of the shortest paths between the node and all other nodes in the graph. Common lymphoid progenitors with germline configuration of antibody-coding genes are at the center of the network.

data	germline origin	sequence	antibody structure	interaction network	antigen structure
aquisition and analysis	bioinformatics	NGS	monoclonal measured polyclonal predicted	multiplex HTS binding assays + deep analysis of interaction databases	molecular structure determination
diversity	combinatorial	sampling restricted	database restricted	assay restricted	database restricted
conversions	prediction	measurement	prediction	prediction	measurement	prediction
data type	qualitative	quantitative discrete	quantitative discrete	quantitative continuous	quantitative discrete
system type	sequence	sequence	structure static	structure flexible	structure static

Network theory (RADARS)

simulation of system dynamics

Table 1. Characterization of data suitable for systems analysis of humoral immunity. The table describes the origin, properties and convertibility of information regarding antibody interactions.



Graphical abstract