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 antibody interaction space and network, 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, with decreasing frequencies of interactions. B-cell development and immune responses can be 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, with one tail representing negatively selected events, the other tail with a network exponent of phi square representing the antibody network.

Thus, the RADARS model suggests that an absolute sequence space is reduced to a thermodynamically viable structure space by means of a network of interactions. 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. 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

138 of uniform reactivity against such common public self-antigens may lead to the positive
139 selection of CDR3 with similar binding properties, and thereby similar sequences.
140 Sequencing of the complete heavy chain variable domains can be readily used to follow
141 changes in repertoire size and diversity during B-cell development and response to
142 immunization (33).
143 Whatever the depth and methodology, sequence similarity relationships can be used for the
144 construction of family trees, often displayed in circular forms. These trees usually start
145 classification with the V segment, clustering clones with common V use (34). While this
146 approach may be useful for classification, the use of the complete VDJ-H sequence as a first
147 stage classifier, followed by VJ-L use better reflects the natural development scheme of B
148 cells (Fig 1C).
149 Antibody repertoire sequencing now follows guidelines to help the integration of data (35–
150 37), several tools and databases devoted especially for these data have been established
151 (7,9,38–40).
152
153 **3. Antibody interaction network representation in structure space**
154
155 In contrast to this graded qualitative scheme, which may well serve the purpose of tracking
156 peripheral clonal expansions accompanied by affinity maturation, a quantitative scheme
157 should place genetic changes into structure rather than sequence space. Furthermore, because
158 it is not just antibody structure but also the availability of targets and the structure of those
159 targets that determine the development of antibody repertoire and the architecture of the
160 network, we shall talk about interaction space, as explained below.
161 *3.1. Structural resolution of molecular recognition as a measure of interaction strength*
162 While sequence can be defined with various levels of certainty of an amino acid occupying a
163 given position in the sequence, molecular structure can be defined at various levels of
164 resolution. As we are talking about antibody molecules structural resolution is on the atomic
165 scale, crystal structures define atomic coordinates on the Ängstrom scale. The binding site of
166 an antibody can also be characterized by the surface area that comes into close contact with
167 the antigen (41,42). Water molecules are displaced from this area as a function of the
168 goodness of fit. The so-called buried surface area (BSA) is therefore a good predictor of
169 binding energy of protein interactions (43). Another measure of goodness of fit is the
170 decrease of free energy of the antibody molecule upon binding. All these approaches are
171 correlated: higher resolution “description” of a structure by the antibody corresponds to

172 greater BSA and to a higher binding energy. In other words, the resolution of molecular
173 recognition is the goodness of fit in terms of number and strength of non-covalent bonds
174 forming between antibody and target and can be expressed as standard free energy change or
175 as equilibrium constant of binding. The advantage of using thermodynamic description for
176 the characterization of structural resolution is that it conveys the sense of function: higher
177 binding energy means higher affinity of antibody to target, which in turn means more
178 efficient clearance (27). Besides defining resolution of molecular recognition, which is a
179 general descriptor, the identification of a given interaction requires the description of target
180 shape, a distinct molecular structure. The higher the resolution the more information is
181 required for defining shape, translating into a better fit between antibody and target.
182 Going in the opposite direction, by decreasing the resolution of interaction, we can reach a
183 theoretical minimum. At this minimum all interactions are defined by a single identical
184 quantity of interaction strength. Let this quantity be 1, so that an equilibrium constant with
185 the value 1 can represent the lowest possible resolution.
186 As standard free energy change is defined as
187
188 $\Delta G° = -RT\ln K_A$
189 by using RT as the unit of energy, called nat, we obtain
190 $\Delta G°_{nat} = -\ln K_A$
191 standard free energy change expressed in nats.
192 Using $\Delta G°_{nat}$ as a measure of resolution, at the minimal resolution binding to any target is
193 possible and can be interpreted as minimal random contact upon collision of molecules, with
194 the value being 0. This minimal resolution characterizes the center of the system, represented
195 by the center of interaction space in Figure 2. By starting to increase resolution we shall be
196 able to distinguish between different shapes, the higher the resolution the more shapes
197 becoming distinct. Because of the structural complexity of an antibody binding surface area,
198 the distinction between all possible shapes at high resolution would require a
199 multidimensional space. Let us gradually generate a multidimensional interaction space by
200 considering a point of origin, the center of the system, from which a particular direction
201 represents a particular shape. In this representation the extent by which we leave the point of
202 origin corresponds to the resolution at which we can define the direction. Thus, going away
203 from the minimal resolution of 0 we can define shape at gradually higher resolutions,
204 corresponding to larger free energy decrease of the interacting molecule (Fig. 2A,B).

205 Different levels of resolution, that is different levels of binding energies appear in our scheme
206 as shells of a sphere. Theoretically the number of directions originating from a single point is
207 infinite, so the shapes available in this representation are also infinite if we go far enough.
208 The other way round, higher resolution of molecular recognition is required to distinguish
209 two similar antigenic structures. Practically, considering a reversible interaction, the
210 resolution is limited by the binding energy of reversible interactions. Beyond this limit
211 structure of the binding partners becomes fixed, rigid, corresponding to covalent binding.
212 This model of the evolution of interactions of a system we shall call 'RAdial ADjustment of
213 System Resolution' or RADARS in short. The abbreviation intentionally reminds of
214 radiolocation, where emitted electromagnetic waves interact with objects in their way and are
215 reflected to provide a virtual image of the surrounding. The RADARS model implies that
216 elements of the growing system interact with the surroundings, gaining information and
217 adjusting system growth accordingly.
218 *3.2. B-cell development in interaction space*
219 Immunological interpretation of the model requires us to fit B-cell development and antibody
220 network evolution into this interaction space. We shall assume that a common lymphoid
221 progenitor has the potential to generate any and all functional VDJ-VJ sequences and
222 therefore to produce via sequential differentiation and maturations steps antibody against any
223 and all targets. By functional VDJ-VJ sequences we mean all sequences that are physically
224 and biologically viable. This means thermodynamic stability (able to fold into a structure
225 compatible with the Ig domain), ability to pair, forming a VH-VL functional binding unit,
226 and ability to sustain a B-cell via delivering survival, differentiation and proliferation signals.
227 A differentiation step that reduces this total potential introduces restrictions in structural
228 space. This will appear as a step towards increased resolution the cognate target structure.
229 Expression of the surrogate light chain (SLC) marks the first step towards BCR formation.
230 These pro-B cells represent the founders of all B cells (Fig. 2C). While signaling via the SLC
231 may be required, it is uncertain whether binding of SLC is required for further differentiation,
232 therefore we assume that these cells seed the complete antibody interaction space.
233 Rearrangement of the heavy chain introduces a structural restriction: a particular functional
234 heavy chain variable domain (VH) sequence has a limited range of targets. Pre-B cells
235 displaying the pre-BCR composed of VH-SLC pairs will divide until and as long as ligands
236 (antigens) are available. Cells with different VH sequences will populate the structural space
237 and share this space according to the availability of target. Cells with more abundant targets
238 expand more, cells with less frequent targets remain in lower numbers, until optimal BCR

engagement is achieved (25). As a result, interaction space as represented at this resolution will be filled with different pre-B cell clones according to the availability of the common self-antigens.

The next level of interaction 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 resolution grows, in a hierarchy similar to that proposed by Derényi (45). To maintain cell numbers in all levels of differentiation the ratio of differentiation rates is proposed to be the average of ϕ rate expansion and no expansion,

differentiation rate $\lambda = (\phi+1)/2 = \phi^2/2$

This practically means that at a given level of differentiation for every 2 incoming cells there are $\phi+1$ outgoing cells. As a result, the distribution of cells in interaction space will be determined by the rate of division and differentiation involving 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

272 of fitting target structures. As a result, the emerging population of immature B cells will bear
273 the antigenic signature of the bone marrow environment. This can be interpreted both as
274 deletion of highly self-reactive clones to prevent autoimmunity (46), and as selection for
275 mildly autoreactive clones to help homeostatic antibody functions and setting a reference for
276 recognition of non-self (47).

277 *3.3. Immune response in interaction space*

278 The development of cells reacting with antigens that are only temporarily present in the host
279 presents the risk of investing energy into clones that will become useless once that antigen
280 disappears. Therefore, such clonal expansions beyond the border of self only takes place
281 when 2nd signals inform the host of danger. This is the development of an immune response,
282 aided by various molecular and cellular help signals. Thymus independent and primary
283 thymus dependent responses expand populations of B cells without improving their affinity,
284 thus keeping them on the same level in the interaction space. We shall call these cells effector
285 B cells. Thymus dependent responses aided by helper T cells lead to the formation of
286 germinal centers where affinity maturation takes place. This adjustment of affinity focuses
287 interactions into a particular direction, shape in our model, leading to increased resolution of
288 interaction space only in that direction. In this region of interaction space, affinity and
289 corresponding resolution of target recognition is high, but once target is cleared cells go into
290 dormancy. These are the memory B cells that conserve the genetic information acquired
291 during affinity maturation but minimize their activity: no divisions, no antibody secretion;
292 remaining guards against future encounter with the same or similar target (Fig. 2C). Another
293 set of cells that remains after target has been cleared is plasma cells, which presumably arise
294 from effector cells finding appropriate niche for long-term survival.
295

296 **4. Characterization of the antibody interaction network**

297

298 The cell biological aspects of the RADARS model are summarized in Figure 3A. Expansion
299 and differentiation of cells originating from CLPs create a cellular repertoire in the bone
300 marrow, then further expansions and differentiation steps increase this repertoire and
301 supplement it with antibody secretion. While selection starts in the bone marrow, events in
302 the periphery show more self-organization because interactions with the outside molecular
303 world begins to take control.

Further differentiation into higher resolution of structure space along lnKA (Fig. 3B) is accompanied by competition between clones via antibody secretion overlapping in structure space.

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 (48). 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. The continuous removal of target antigens requires the continuous production of antibodies and maintenance of a reaction quotient, which we will also assume to follow a lognormal distribution. The combination of these two, Q/KA is the exponential of interaction free energy $e^{\Delta rG}$, which then again possesses lognormal distribution.

If B cell differentiation takes place in this interaction energy space, then network nodes will be individual B cell clones, moving along the $\Delta G°$ radius at $\lambda=\phi 2/2$ steps. Reed (49), Mitzenmacher (50) and Newman (51), described that the exponential sampling of lognormal distributions produces power law distribution with two tails, called double-Pareto distribution. We can apply their approach in our model by sampling Q/KA at λ rate as defined by

$$e^{-\lambda RT}$$

where R is the universal gas constant and T is thermodynamic temperature. For the sake of simplicity, we can use 1 nat = RT as thermodynamic unit (Fig.3B). The two tails of the power law relationship are

$$f(x) \approx x^{-1-SQRT(2\lambda)} \quad , x>1$$

$$f(x) \approx x^{-1+SQRT(2\lambda)} \quad , x>1$$

meaning that the slopes of probability density of interaction energies (Fig.3B) are both ϕ-1, ϕ itself taking up negative or positive values.

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 4, interpreting the events as the immune system being outside and developing inwards. Practically this means that we use 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.

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. 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 (52).

The RADARS model suggests that the greater the resolution of structural recognition the more restricted is the number of shapes recognized. However, with the growth 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.4). 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.

Substituting $phi^2/2$ for λ in the double Pareto distribution we obtain the probability distribution function of the antibody interaction network according to the RADARS model:

$$f(k) \approx k^{phi-1} \quad , \text{for } 0 < k \leq 1$$

$$f(k) \approx k^{-phi-1} \quad , \text{for } k > 1$$

where k is $e^{\Delta rG}$, related to the energy of interaction in the system (Fig. 5).

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 antibody-centric view in our model. The low-affinity region with developing B-cells is

370 homogenously interconnected by clonal relationships and shared usage of gene segments
371 (Fig.2 and 3). The high affinity side of the distribution is the narrowing, focusing interaction
372 space of plasma cells.
373 The immunological implication of the RADARS model is that the naive repertoire generates
374 a low affinity image of the human antigenome available in the bone marrow environment.
375 Such an image has been called the immunological humunculus, as an analogy to the
376 representation of human body in the brain cortex (53). While this property of the repertoire
377 often leads investigators to conclude that autoreactive antibodies are enriched in the naive
378 and B1 repertoire, the practical use is quite obvious, as has been suggested by Irun Cohen
379 (47). Telling the difference between molecular structural self and non-self can be quite easy if
380 the difference is striking. For identifying small differences, it is useful to have a starting
381 repertoire that represents self and then increase resolution only in regions where encountered
382 dangerous structures show up in structural space. The importance of self-recognition has also
383 been recognized in antibody-mediated physiological functions (27,54).
384 The generation of an antibody network with network hubs represented by plasma cells
385 secreting antibodies reveals the physical aspect of the system: all interactions of such an
386 antibody contribute to the clearance of many target antigens sharing structural homology but
387 at an ever-decreasing efficiency as we move away in structure space. A new node in the
388 network, a new B cell in the structure space, will preferentially attach to an existing
389 subnetwork as a low-affinity clone, in agreement with the preferential attachment model of
390 growth in scale-free networks (55). Importantly, the model confirms that the immune system
391 is organized around key targets, dominant epitopes, which are the main hubs in the network.
392
393 **Merging sequence space to interaction space**
394
395 Network theory has always been considered as a key to understand and define immunity on a
396 systems level. The network hypothesis of Niels Jerne (22), its modified version leading to the
397 concept of clonal selection (56), mathematical and computational simulations (57,58),
398 various re-interpretations (23), experimental approaches using NGS (24,31) or antigen
399 microarrays (21) all strive to describe this highly complex system as connected elements of a
400 network. There are two new aspects of the RADARS model that may improve our view of
401 this network. First, it introduces physical units, binding energy, as the measure of interactions
402 and as a measure of system architecture. Natural networks are formed as a result of energy
403 dispersal (59,60), 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.

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 (61), and comprehensive and selective analysis of the B1 repertoire is very important for capturing network properties of the system.

The model presented here depicts an ideal state of the system of interaction. 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.

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 anonymous 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.

436 References

437 1. Müller V, de Boer RJ, Bonhoeffer S, Szathmáry E. An evolutionary perspective on the
438 systems of adaptive immunity. Biol Rev Camb Philos Soc. 2017 Jul 26;93(1):505–528.
439 2. Eberl G. A new vision of immunity: homeostasis of the superorganism. Mucosal
440 Immunol. 2010 Sep;3(5):450–460.
441 3. Sleator RD. The human superorganism - of microbes and men. Med Hypotheses. 2010
442 Feb;74(2):214–215.
443 4. Pradeu T, Carosella ED. On the definition of a criterion of immunogenicity. Proc Natl
444 Acad Sci U S A. 2006 Nov 21;103(47):17858–17861.
445 5. Davis MM, Tato CM, Furman D. Systems immunology: just getting started. Nat
446 Immunol. 2017 Jun 20;18(7):725–732.
447 6. Boyd SD, Joshi SA. High-Throughput DNA Sequencing Analysis of Antibody
448 Repertoires. Microbiol Spectr. 2014 Oct;2(5).
449 7. Vergani S, Korsunsky I, Mazzarello AN, Ferrer G, Chiorazzi N, Bagnara D. Novel
450 Method for High-Throughput Full-Length IGHV-D-J Sequencing of the Immune
451 Repertoire from Bulk B-Cells with Single-Cell Resolution. Front Immunol. 2017 Sep
452 14;8:1157.
453 8. Friedensohn S, Lindner JM, Cornacchione V, Iazeolla M, Miho E, Zingg A, et al.
454 Synthetic standards combined with error and bias correction improve the accuracy and
455 quantitative resolution of antibody repertoire sequencing in human naïve and memory
456 B cells. Front Immunol. 2018 Jun 20;9:1401.
457 9. Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V. Computational
458 Strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune
459 Repertoires. Front Immunol. 2018 Feb 21;9:224.
460 10. Greiff V, Miho E, Menzel U, Reddy ST. Bioinformatic and statistical analysis of
461 adaptive immune repertoires. Trends Immunol. 2015 Nov;36(11):738–749.
462 11. Prechl J, Papp K, Erdei A. Antigen microarrays: descriptive chemistry or functional
463 immunomics? Trends Immunol. 2010 Apr;31(4):133–137.
464 12. Merbl Y, Zucker-Toledano M, Quintana FJ, Cohen IR. Newborn humans manifest
465 autoantibodies to defined self molecules detected by antigen microarray informatics. J
466 Clin Invest. 2007 Mar;117(3):712–718.
467 13. Bacarese-Hamilton T, Bistoni F, Crisanti A. Protein microarrays: from serodiagnosis to
468 whole proteome scale analysis of the immune response against pathogenic
469 microorganisms. BioTechniques. 2002 Dec;Suppl:24–29.
470 14. Wang J, Lin J, Bardina L, Goldis M, Nowak-Wegrzyn A, Shreffler WG, et al.
471 Correlation of IgE/IgG4 milk epitopes and affinity of milk-specific IgE antibodies with
472 different phenotypes of clinical milk allergy. J Allergy Clin Immunol. 2010
473 Mar;125(3):695–702, 702.e1.
474 15. Grötzinger C. Peptide microarrays for medical applications in autoimmunity, infection,
475 and cancer. Methods Mol Biol. 2016;1352:213–221.
476 16. Sjöberg R, Mattsson C, Andersson E, Hellström C, Uhlen M, Schwenk JM, et al.
477 Exploration of high-density protein microarrays for antibody validation and
478 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.

521 33. Greiff V, Menzel U, Miho E, Weber C, Riedel R, Cook S, et al. Systems Analysis
522 Reveals High Genetic and Antigen-Driven Predetermination of Antibody Repertoires
523 throughout B Cell Development. Cell Rep. 2017 May 16;19(7):1467–1478.
524 34. DeFalco J, Harbell M, Manning-Bog A, Baia G, Scholz A, Millare B, et al. Non-
525 progressing cancer patients have persistent B cell responses expressing shared antibody
526 paratopes that target public tumor antigens. Clin Immunol. 2018 Feb;187:37–45.
527 35. Bukhari SAC, O'Connor MJ, Martínez-Romero M, Egyedi AL, Willrett D, Graybeal J,
528 et al. The CAIRR Pipeline for Submitting Standards-Compliant B and T Cell Receptor
529 Repertoire Sequencing Studies to the National Center for Biotechnology Information
530 Repositories. Front Immunol. 2018 Aug 16;9:1877.
531 36. Breden F, Luning Prak ET, Peters B, Rubelt F, Schramm CA, Busse CE, et al.
532 Reproducibility and reuse of adaptive immune receptor repertoire data. Front Immunol.
533 2017 Nov 1;8:1418.
534 37. Rubelt F, Busse CE, Bukhari SAC, Bürckert J-P, Mariotti-Ferrandiz E, Cowell LG, et
535 al. Adaptive Immune Receptor Repertoire Community recommendations for sharing
536 immune-repertoire sequencing data. Nat Immunol. 2017 Nov 16;18(12):1274–1278.
537 38. Rosenfeld AM, Meng W, Luning Prak ET, Hershberg U. Immunedb, a novel tool for
538 the analysis, storage, and dissemination of immune repertoire sequencing data. Front
539 Immunol. 2018 Sep 21;9:2107.
540 39. Corrie BD, Marthandan N, Zimonja B, Jaglale J, Zhou Y, Barr E, et al. iReceptor: A
541 platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data
542 across federated repositories. Immunol Rev. 2018 Jul;284(1):24–41.
543 40. Avram O, Vaisman-Mentesh A, Yehezkel D, Ashkenazy H, Pupko T, Wine Y. ASAP -
544 A Webserver for Immunoglobulin-Sequencing Analysis Pipeline. Front Immunol. 2018
545 Jul 30;9:1686.
546 41. Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes
547 in antibody:protein complexes. Mol Immunol. 2013 Jan;53(1-2):24–34.
548 42. Wedemayer GJ, Patten PA, Wang LH, Schultz PG, Stevens RC. Structural insights into
549 the evolution of an antibody combining site. Science. 1997 Jun 13;276(5319):1665–
550 1669.
551 43. Chen J, Sawyer N, Regan L. Protein-protein interactions: general trends in the
552 relationship between binding affinity and interfacial buried surface area. Protein Sci.
553 2013 Apr;22(4):510–515.
554 44. Prechl J. Thermodynamic projection of the antibody interaction network: the fountain
555 energy landscape of binding. BioRxiv. 2017 Apr 5;
556 45. Derényi I, Szöllősi GJ. Hierarchical tissue organization as a general mechanism to limit
557 the accumulation of somatic mutations. Nat Commun. 2017 Feb 23;8:14545.
558 46. Melchers F. Checkpoints that control B cell development. J Clin Invest. 2015
559 Jun;125(6):2203–2210.
560 47. Cohen IR. Biomarkers, self-antigens and the immunological homunculus. J
561 Autoimmun. 2007 Dec;29(4):246–249.
562 48. Zheng X, Wang J. The universal statistical distributions of the affinity, equilibrium
563 constants, kinetics and specificity in biomolecular recognition. PLoS Comput Biol.
564 2015 Apr 17;11(4):e1004212.

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. Newman MEJ. Power laws, Pareto distributions and Zipf's law. Contemp Phys. 2005 Sep;46(5):323–351.
52. 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.
53. Cohen IR, Young DB. Autoimmunity, microbial immunity and the immunological homunculus. Immunol Today. 1991 Apr;12(4):105–110.
54. Lobo PI. Role of Natural Autoantibodies and Natural IgM Anti-Leucocyte Autoantibodies in Health and Disease. Front Immunol. 2016 Jun 6;7:198.
55. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004 Feb;5(2):101–113.
56. 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.
57. FARMER DJ, PACKARD N, PERELSON A. The immune system, adaptation, and machine learning. Physica D: Nonlinear Phenomena. 1986;22(1-3):187–204.
58. Schulz R, Werner B, Behn U. Self-tolerance in a minimal model of the idiotypic network. Front Immunol. 2014 Mar 10;5:86.
59. Mäkelä T, Annila A. Natural patterns of energy dispersal. Phys Life Rev. 2010 Dec;7(4):477–498.
60. Hartonen T, Annila A. Natural networks as thermodynamic systems. Complexity. 2012 Nov;18(2):53–62.
61. 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.

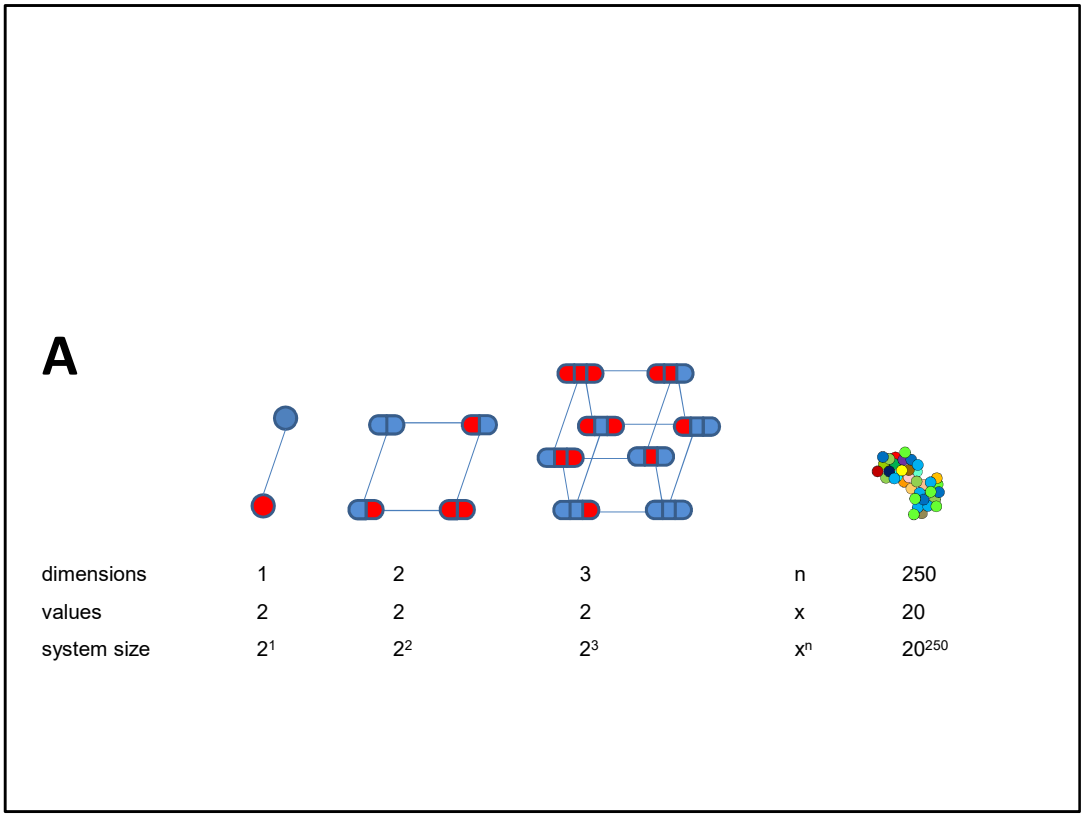


Figure 1. Sequence space and visualization of antibody combinatorial relationships
A) Theoretical diversity of a sequence is determined by its length and the number of values a particular position in the sequence can take. More than 3 dimensions are difficult to visualize in 2D. An antibody Fv region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

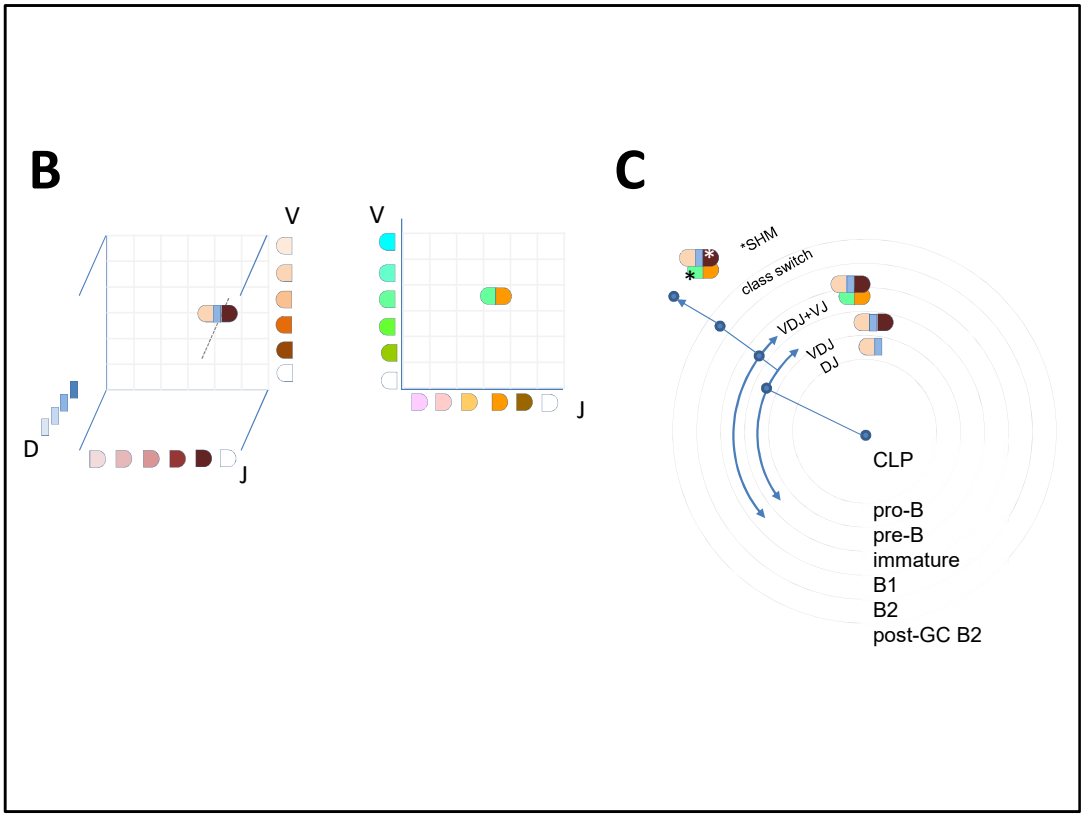
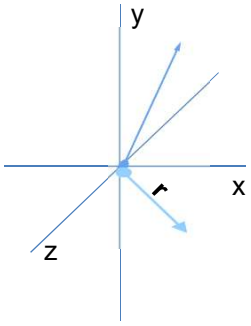


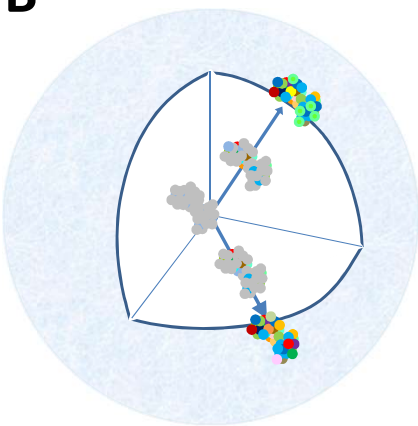
Figure 1
B) Antibody sequences are frequently interpreted as recombined germline sequences. This approach allows the simplified display of repertoires obtained by NGS, preferably with paired heavy and light chain VD identification. Such a display of combinatorial diversity may allow the tracking of specific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

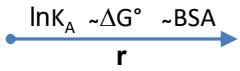
A



dimensions	~ e^{r}
values	~ sphere surface ~ $r^{\wedge 2}$
system size N	~ $e^{r^{\wedge 2}}$

B





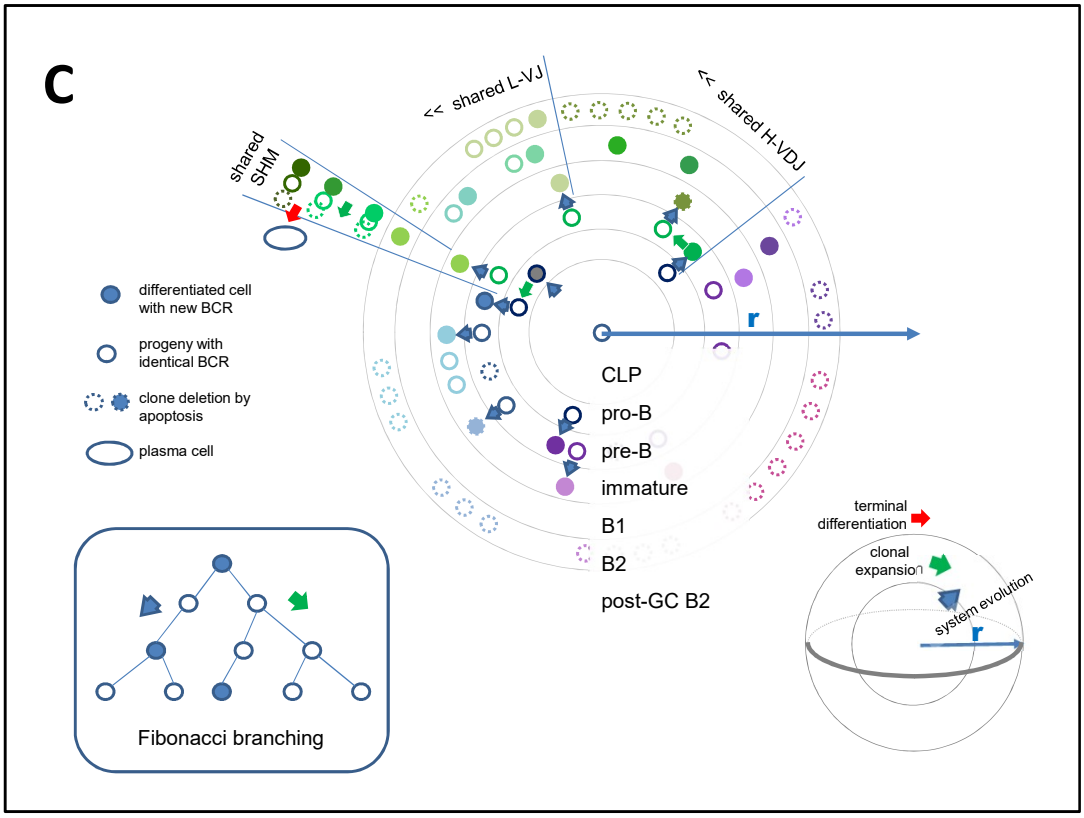
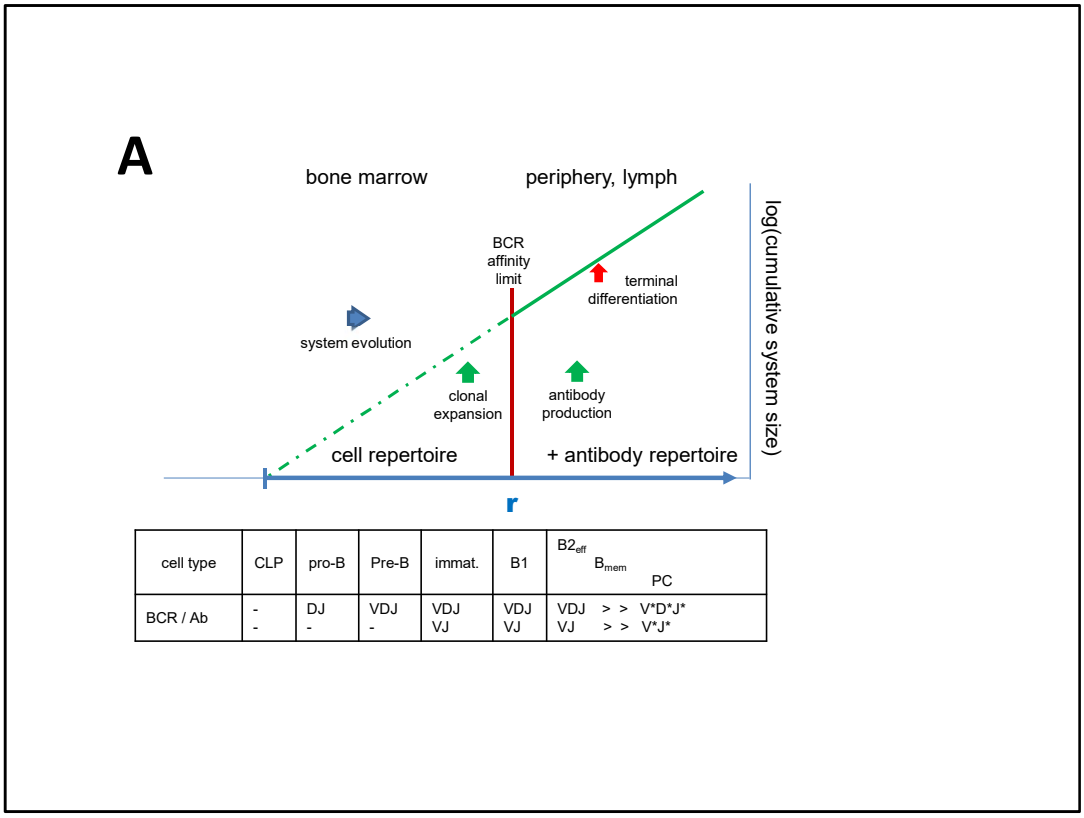


Figure 2.
C) Systemic organization of antibody evolution.
The evolution of the system of antibodies can be interpreted as clones filling the interaction space at various levels of resolution. Along this pathway cells continually increase their specificity and affinity towards their target direction. A common lymphoid progenitor has the potential to develop antibody against any target, equivalent to a radius of ln(1)=0. 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.



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

Figure 3. Predicted network properties of the humoral immune system.
A) Clonal network development 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.

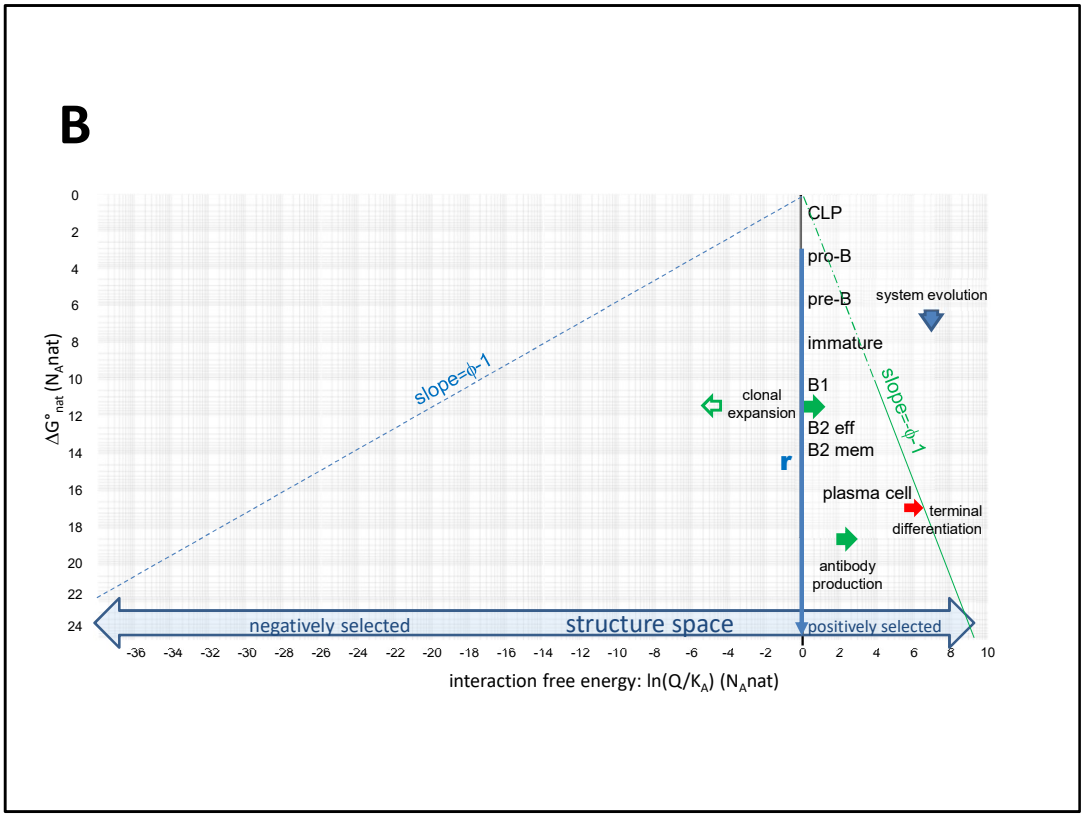


Figure 3.
B) Double-Pareto distribution of antibody interaction free energies in the system. Differentiation of B cells out of the bone marrow is driven by antigen. Here antibodies with increasing free energy, as represented by DG°nat, bind to available targets with increasing tendency to decrease their free energy and achieve a conformation as close to the native bound form as possible. As a consequence antibodies also bind to targets in nearby structure space, a phenomenon called cross-reactivity. Cross-reactivity generates a network of interactions governed by the expansion of the immune system into this structure space. The negative side represents negative selection events, which correspond to thermodynamically non-viable structures in the system. The system is thus defined by Gibbs free energy of reactions and standard free energy of interacting antibody.
*somatic hypermutations; CLP, common lymphoid progenitor; immat., immature; B2eff, effector B2 cell; Bmem, memory B cell; PC, plasma cell

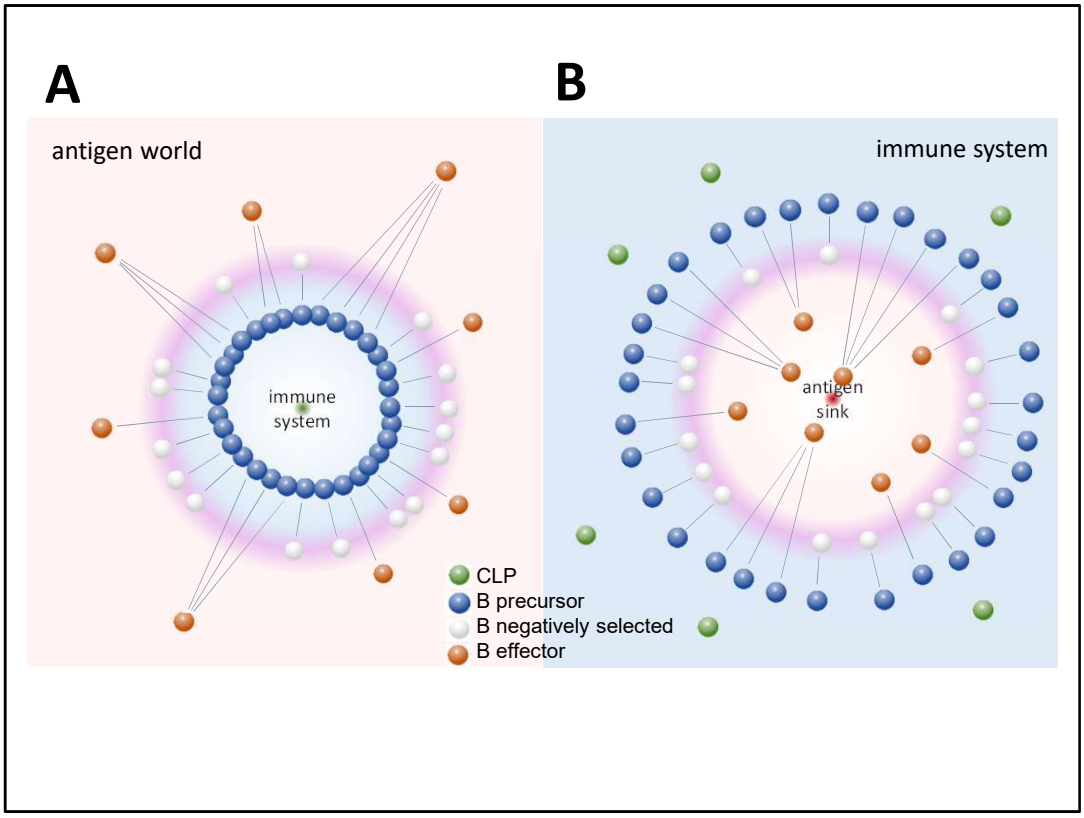


Figure 4. Antibody-centric and antigen-centric views of the organization of adaptive humoral immunity.
The antibody-centric view (A) corresponds to the clonal development and expansion of B cells, with a radius of lnKA=ΔG°nat. We can turn the system inside out to obtain an antigen-centric view (B), where the system grows „inwards" with a radius corresponding to lnKD. In this representation the development of antibodies with higher affinity pull antigen down an antigen sink with increasing efficiency.

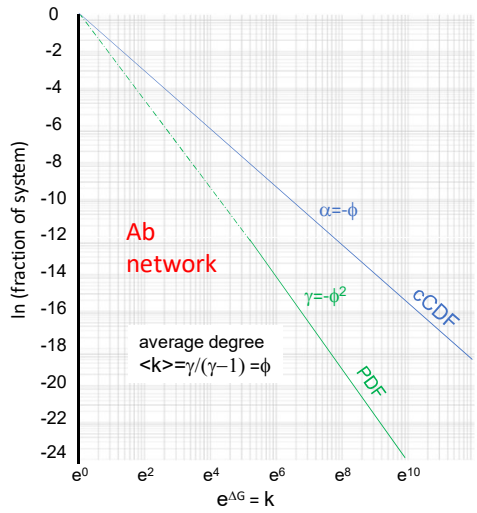
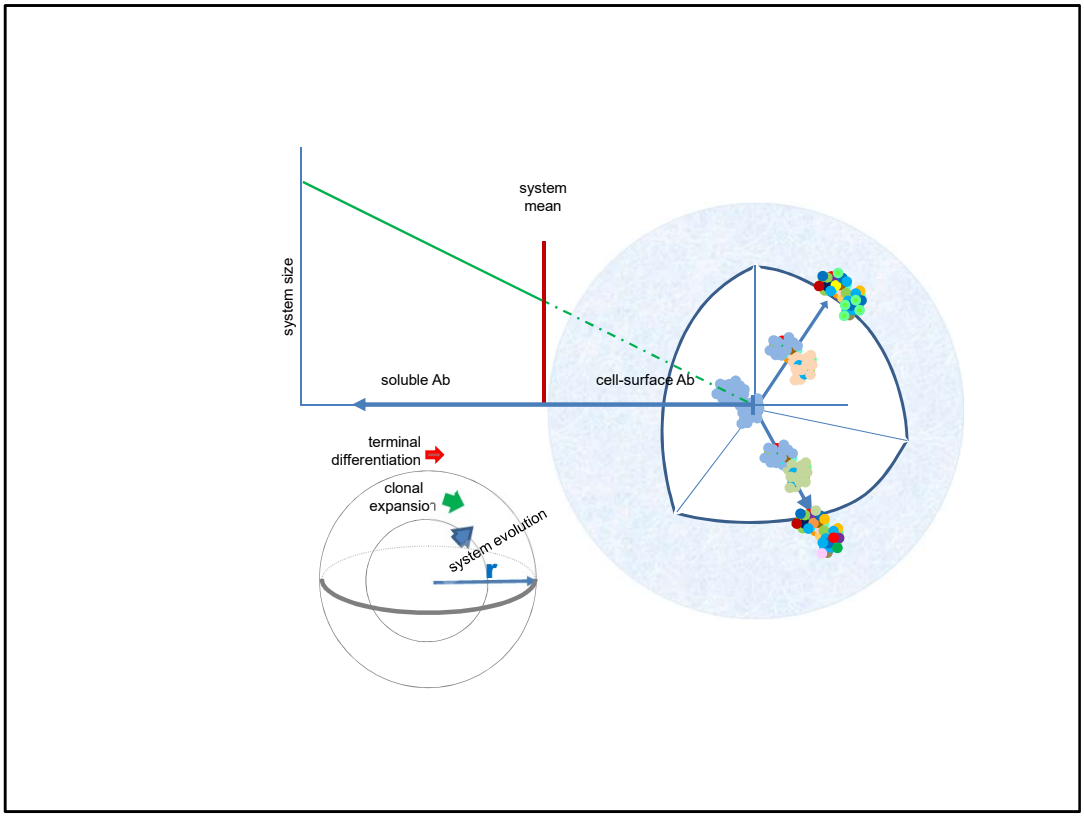


Figure 5. Properties of the antibody interaction network.
Degree distribution of the antibody network corresponds to the free energy of antibodiy interactions and can be expressed as e^{DG}. Because of the special properties of phi, the degree exponent, the average degree, and the slope of the complementary cumulative distribution can all be xepressed using phi. Note that because of the inverse relationship between KA and KD, which determines frequency, the same scale is used as in figure 3 but with negative values.



Graphical abstract