

RESEARCH ARTICLE

Open Access



Deciphering evolution of immune recognition in antibodies

Harmeet Kaur^{1,2}, Neetu Sain³, Debasisa Mohanty³ and Dinakar M. Salunke^{1,4*}

Abstract

Background: Antibody, the primary effector molecule of the immune system, evolves after initial encounter with the antigen from a precursor form to a mature one to effectively deal with the antigen. Antibodies of a lineage diverge through antigen-directed isolated pathways of maturation to exhibit distinct recognition potential. In the context of evolution in immune recognition, diversity of antigen cannot be ignored. While there are reports on antibody lineage, structural perspective with respect to diverse recognition potential in a lineage has never been studied. Hence, it is crucial to evaluate how maturation leads to topological tailoring within a lineage enabling them to interact with significantly distinct antigens.

Results: A data-driven approach was undertaken for the study. Global experimental mouse and human antibody-antigen complex structures from PDB were compiled into a coherent database of germline-linked antibodies bound with distinct antigens. Structural analysis of all lineages showed variations in CDRs of both H and L chains. Observations of conformational adaptation made from analysis of static structures were further evaluated by characterizing dynamics of interaction in two lineages, mouse V_H1-84 and human V_H5-51 . Sequence and structure analysis of the lineages explained that somatic mutations altered the geometries of individual antibodies with common structural constraints in some CDRs. Additionally, conformational landscape obtained from molecular dynamics simulations revealed that incoming pathogen led to further conformational divergence in the paratope (as observed across datasets) even while maintaining similar overall backbone topology. MM-GB/SA analysis showed binding energies to be in physiological range. Results of the study are coherent with experimental observations.

Conclusions: The findings of this study highlight basic structural principles shaping the molecular evolution of a lineage for significantly diverse antigens. Antibodies of a lineage follow different developmental pathways while preserving the imprint of the germline. From the study, it can be generalized that structural diversification of the paratope is an outcome of natural selection of a conformation from an available ensemble, which is further optimized for antigen interaction. The study establishes that starting from a common lineage, antibodies can mature to recognize a wide range of antigens. This hypothesis can be further tested and validated experimentally.

Keywords: Germline, Mature, Somatic hypermutation, Data science, Antigens, Paratope, Antibody, Simulation, Cluster, Conformation

* Correspondence: dinakar.salunke55@gmail.com

¹Regional Centre for Biotechnology, Biotech Science Cluster, Faridabad, Haryana 121001, India

⁴International Centre for Genetic Engineering and Biotechnology, New Delhi, Delhi 110067, India

Full list of author information is available at the end of the article



Background

The antibody-antigen (Ab-Ag) is a miniature system to understand the process of evolution. Development of a B-cell starting from progenitor lymphoid cell to an immature B-cell marks the event of VDJ recombination leading to the formation of naïve germline antibody (Ab) [1]. Exposure of an antigen (Ag) fosters affinity maturation leading to an iterative process of cell proliferation, extensive mutagenesis on the immunoglobulin gene and stringent Darwinian selection of B-cells producing higher affinity antibodies [2]. Various studies have also reported insertions and deletions during somatic hypermutation [3, 4]. Alterations in the genetic machinery of immunoglobulin are tailored to mediate antigen recognition. Such genetic drifts are strictly dependent on the time of exposure and antigen persistence [5]. Hence, at any given time, the incoming antigen acts as a stimulus that allows somatic mutations to be incorporated in the genome that may be favorable and rarely deleterious. Infrequent deleterious mutations result in formation of self-reactive receptors or B-cell transformation [6]. These outcomes are generally eliminated by regulated cell-cycle checkpoints and negative selection processes in the germinal center. Nonetheless, the processes often undergo challenge that may result in breach of self-tolerance leading to autoimmune disorders. Favorable mutations, on the other hand, undergo affinity-based positive selection and facilitate recognition of bona fide antigen. Therefore, based on the incoming antigen, different pathways of maturation may ensue in the germline B-cells forming different advanced versions or siblings of a lineage. Detailed characterization of dynamics of antigen recognition by mature antibodies should provide unprecedented insight into the immune response. Thus, an Ab-Ag system serves as a perfect miniature model to study the evolution of molecular recognition i.e. how antibodies of an ontology diffuse into isolated molecular environment of the stimulating antigen enabling differentiated fixation of paratope topology. One way of dissecting this is by looking for structural changes in the lineage. While sequence based phylogeny has been implemented to analyze different facets of evolution of immunological recognition, the role of structural adaptation in a lineage has not been systematically explored.

Studies on antibodies inheriting the same set of germline genes suggest adoption of fundamentally different binding modes [7]. In the present study, we have examined conformational features associated with recognition of a wide range of antigens by sibling antibodies. Additionally, it is hypothesized that any similarity despite divergence in the structural landscape obtained from simulation will explain their common origin.

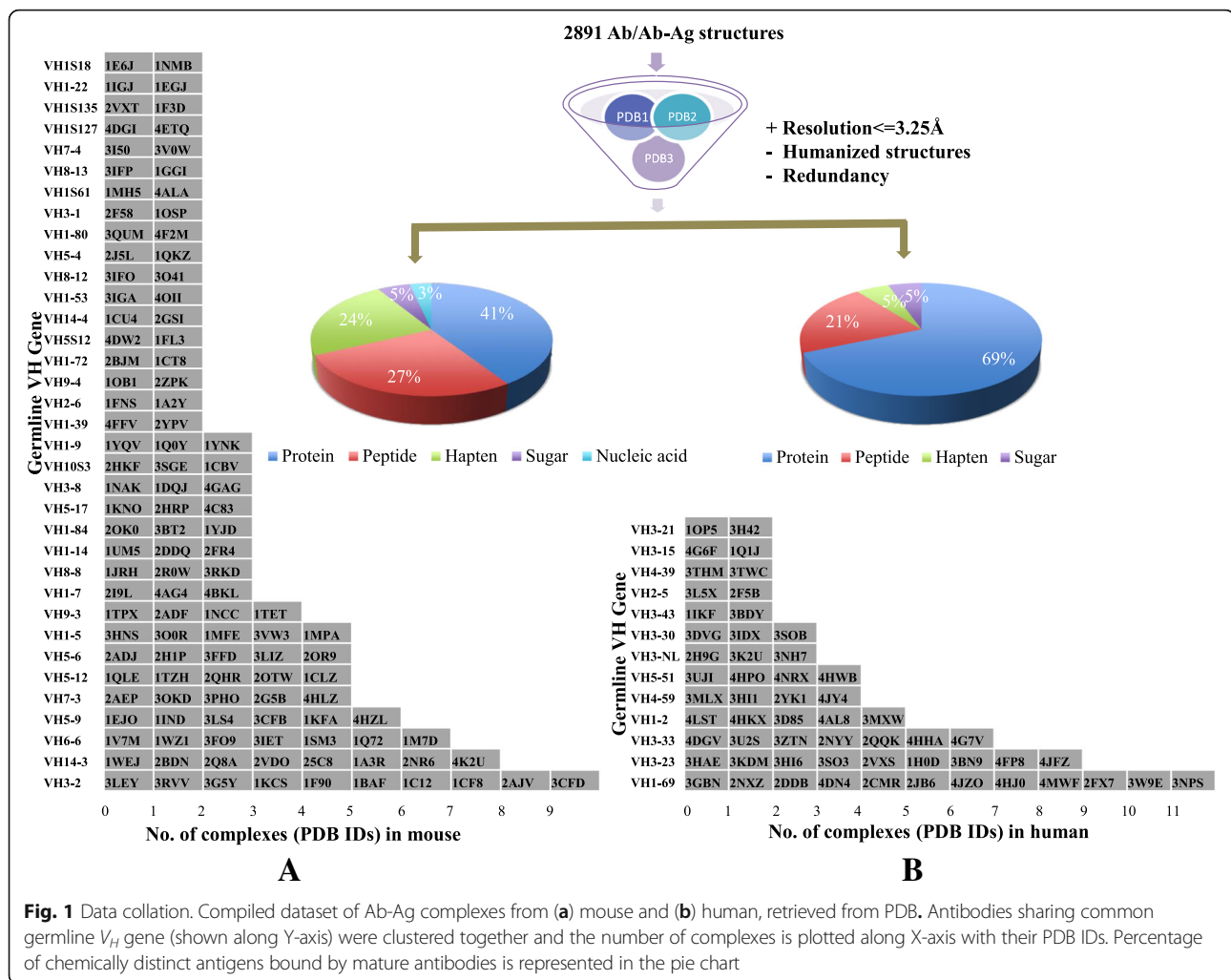
Towards this end, structural data of antibodies of different lineages bound with significantly distinct antigens

were collated and analyzed. Topological alteration of the CDRs, to accommodate distinct antigens was evident within individual lineage. To validate observation from analysis of static structures, antibodies of two germline lineages i.e. carrying the same V_H gene set and bound to significantly different antigens were analyzed separately using explicit solvent molecular dynamics (MD) simulations. While sequence analysis suggested that somatic mutations have altered the geometry of the respective antibody, dynamics showcased that further structural divergence specifically in the paratope was brought about for preferential antigen binding. Despite somatic diversification, similar overall architecture of the descendants and the conformers could be envisaged as retention of germline imprint.

Results

Germline-linked mature antibodies reveal structural heterogeneity

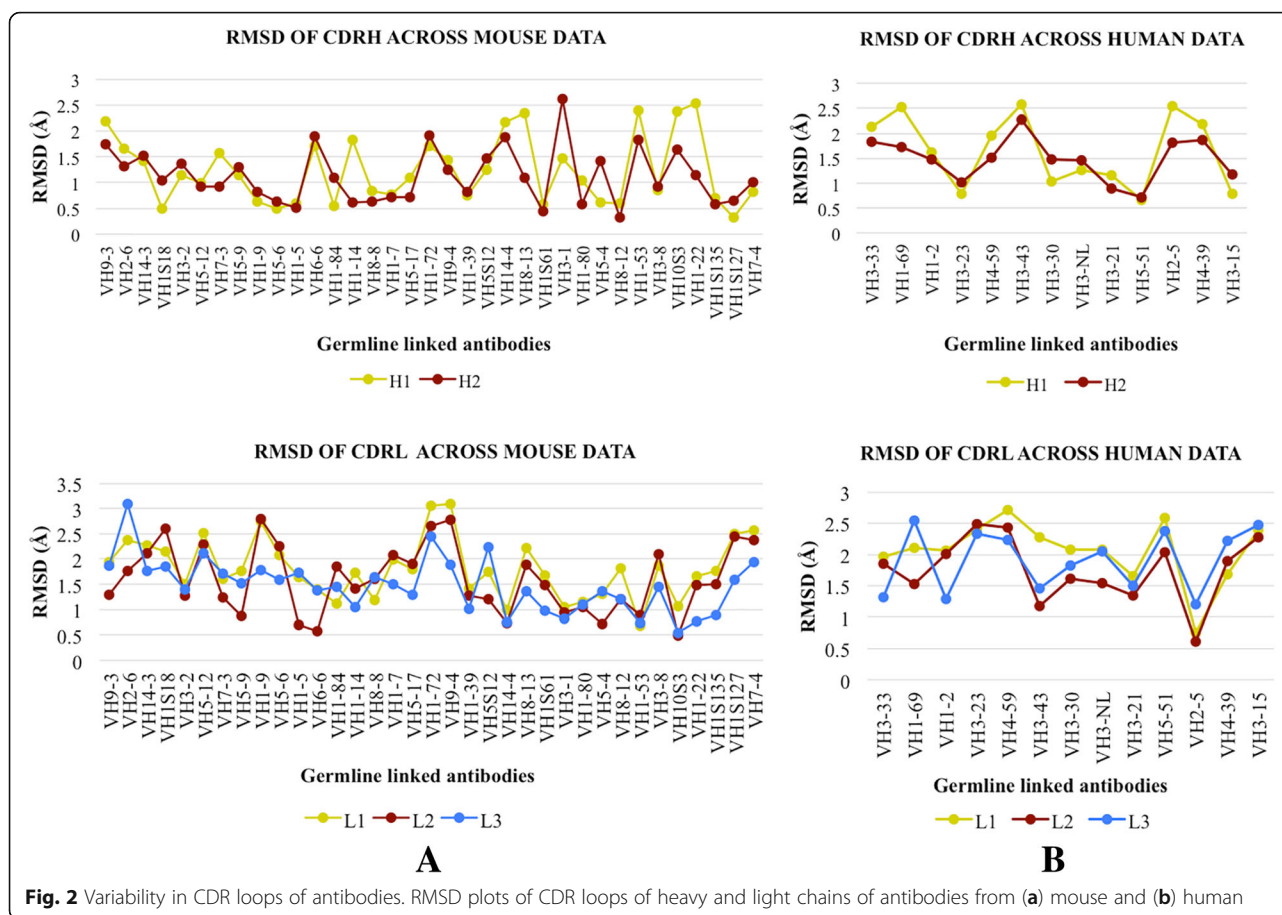
Two thousand eight hundred ninety-one hits were obtained from RCSB PDB that contained both standalone antibody molecules and Ab-Ag complexes [8]. Coordinate files of complexes were retrieved and were segregated based on the source of antibody into two groups, mouse and human, to obtain homogenous sets of data. The hits were filtered based on the following criteria: Only X-ray crystallographic structures with resolution of 3.25 Å or higher were retained. Humanized or chimeric structures were excluded, only a single representative of an antibody was taken to avoid redundancy. Criteria for the nature of antigen or antibody isotype was not set. Antibodies that are deemed to share the same lineage are those that are produced from a single B-cell clone. It has been documented that the H-chain is the primary player in antigen recognition and binding [9, 10]. The L-chain, although, acts in concert with H-chain, is largely known to be an accessory component. Therefore, at the molecular level, the data were subsequently clustered based on a recent study that involved antibody evolution from a common ancestral V_H (variable region of heavy chain) gene [11]. The coding genes of retrieved antibodies were identified based on highest score and lowest e-value. If the antibodies aligned with similar scores to several distinct germline genes, the gene that was also common in other antibodies and enabled their clustering was selected. Those that shared no common IGHV origin were rejected. The final compiled dataset comprised of 115 mature antibodies from mouse that were grouped into 35 clusters based on common germline V_H genes (Fig. 1a and Additional file 1: Table S1). Likewise, 57 mature antibodies from human were clustered into 13 germline V_H gene groups (Fig. 1b and Additional file 2: Table S2).



In the collated database, the chemistry of antigens bound by antibodies evolved from a common IGHV was analyzed. The data revealed a high level of diversity and individual uniqueness of the antigens (Fig. 1; Additional file 1: Table S1 and Additional file 2: Table S2). Antigens of individual sets could not be aligned because of distinctness in their chemical nature and topology. Even for protein or peptide antigens, sequence alignment revealed no immuno-dominant epitopes as seen for anti-peptide antibodies of V_{H1S127} , V_{H1-53} , V_{H1-39} , V_{H1-80} etc. lineages from mouse and V_{H3-30} , V_{H1-2} , V_{H3-33} , V_{H5-51} etc. lineages from human (Additional file 1: Table S1 and Additional file 2: Table S2). Chemically, the spectrum of epitopes was wide for mouse V_{H3-2} , V_{H7-3} , V_{H1-84} , V_{H1-5} and V_{H5-17} lineages and for human V_{H1-69} , V_{H3-23} , V_{H4-59} and V_{H4-39} lineages ranging from proteins, peptides, nucleic acid to sugar and haptens.

Despite significant diversity in antigens, it was intriguing that they were bound by sibling antibodies. Structure superposition of all antibodies of respective lineages was

done to check topological changes undergone in CDRs to accommodate respective antigens. Conformational variability of the paratope was apparent for all groups of mouse and human (Additional file 3: Figures S1 and S2). Variability was observed in all CDR loops of both heavy and light chains. However, the contribution of individual CDR loop was variable for different datasets. To further quantify this observation, average RMSDs of CDR loops were calculated from structural alignment. In case of mouse, 58 and 52% data show RMSD > 1 Å for CDRs H1 and H2 respectively while 98, 80 and 80.5% data show RMSD > 1 Å for CDRs L1, L2 and L3 respectively (Fig. 2a). In case of human, 76 and 84% data had RMSD > 1 Å for CDRs H1, H2 while 92, 92 and 100% data for CDRs L1, L2 and L3 (Fig. 2b). Due to recombination events at V-D and D-J junctions [12], the sequence and length of H3 is significantly altered within a lineage across dataset, hence its RMSD could not be computed. Thus, from analysis of structural alignment it is clear that all CDR loops exhibit seemingly comparable variability; the RMSD values being slightly lower for H1 and H2 in mouse antibodies. The



structural alignment provides a basic understanding of the conformational re-arrangement that facilitates a favorable micro-environment for accommodating the antigen. It does not necessarily translate into actual contacts made by CDRs with the antigen.

In order to further assess contributions of CDR loops, interaction profile of the complexes in the light of H-bond formation was investigated. Engagement of H-chain was higher than L-chain in 71% mouse and 78% human data, with contributions of all three loops (Additional file 3: Figures S3 and S4). 14% (mouse) and 24.5% (human) data show no involvement of L-chain. In 12% mouse (PDB 1TPX of V_{H9-3} , 1E6J of V_{H1S18} , 1BAF of V_{H3-2} etc.) and 1% human (PDB 4DGV of V_{H3-33} , 2JB6 of V_{H1-69} , 4JFZ of V_{H3-23} etc.) data, H3 does not form any H-bond, implying significant contributions of CDRs H1 and H2. But, H3, in conjunction with other loops, is crucial for antigen recognition and binding as seen across rest of the data [10]. In general, H3 plays a predominant role in defining the topography of the binding site [13]. Shorter ones can create a cavity to accommodate peptides while long H3 loops can generate a definite finger-like topography [14, 15]. Thus, structural re-arrangement in all CDRs of the collated

dataset and especially in H-chain that bear common genetic elements was surprising.

Thus, sibling antibodies exhibit conformational heterogeneity of the paratope to accommodate distinct antigens and interaction is mediated by different CDR loops with contribution of varying degrees across dataset. Even though H3 is a crucial player in antigen recognition, it may not necessarily be involved in direct contact with the antigen (as observed from H-bond analysis), signifying the involvement of other loops. The appropriately screened database thus served as an ideal resource to understand divergent evolution. However, mechanistic details would be better revealed by investigating dynamics of a lineage as it would shed light on time-dependent structural changes.

Towards this direction, two lineages were chosen from the compiled datasets, V_{H5-51} from human and V_{H1-84} from mouse (Table 1, for details, see Methods). Important requisites for selection of systems were that antigens bound by sibling antibodies were different and forcefield parameters for antigens were available for subsequent simulation studies or could be generated using available tools. The sequences and structures of these antigens were analyzed to ensure they were different.

Table 1 Ab-Ag complexes as test systems for study

| Organism | PDB ID | Antibody | Antigen | V_H -gene | Score (Bits) | e-value |
|----------|--------|-----------|---------------------------------------------------------------|-------------|--------------|---------|
| Mouse | 2OK0 | ED10 | DNA | 1-84*01 | 171 | 2e-44 |
| | 3BT2 | Anti-uPAR | Urokinase plasminogen activator surface receptor, Vitronectin | 1-84*01 | 162 | 1e-41 |
| | 1YJD | 5.11A1 | T-cell-specific surface glycoprotein CD28 | 1-84*01 | 140 | 5e-35 |
| Human | 4NRX | m66 | Gp41 MPER peptide | 5-51*03 | 219 | 5e-59 |
| | 4HWB | 10G5H6 | Ectodomain D3 of IL-13 | 5-51*03 | 237 | 3e-64 |
| | 3UJI | 2558 | Gp160 V3-domain | 5-51*03 | 204 | 2e-54 |
| | 4HPO | CH58 | Gp120 V2-domain | 5-51*03 | 243 | 5e-66 |

Mapping somatic mutations in mature antibodies

Sequence analysis of the variable region of heavy chain in V_H1-84 antibodies suggests that Fab ED10 (as a complex in PDB ID: 2OK0), Fab 5.11A1 (as a complex in PDB ID: 1YJD) and anti-uPAR antibody (as a complex in PDB ID: 3BT2) share 81.25, 83.67 and 85.71% identity with their germline V_H1-84 gene each carrying 17, 16 and 13 V_H mutations respectively in the CDR and framework regions (Fig. 3a and b). Presence of allelic variants of germline gene also contributes to the difference in sequence within a lineage. High-throughput sequencing of human antibody repertoire has reported various polymorphic forms of IGHV and IGHD loci [16, 17], but such a report is rare in mouse [18]. At a few positions, somatic mutation between the antibodies is conserved (V2, D65) while most map to different regions (Fig. 3a, Table 2). A few mutations facilitate interaction with the antigen in Fab 5.11A1 (S31) and anti-uPAR antibody (D55, E58) while most lie in the vicinity of interacting residues (Table 2).

Sequence analysis of V_H5-51 antibodies suggests that Fab m66 (as a complex in PDB ID: 4NRX) has 90.72% identity with its germline V_H5-51 gene harboring as many as 8 V_H mutations in the CDR and framework regions while Fab 10G5H6 (as a complex in PDB ID: 4HWB) is 97.96% identical carrying 2 V_H mutations (Fig. 4a and b). None of the somatic mutations is involved in interaction with antigen. However, a few of the mutations lie in the vicinity of the interacting residues (Table 2). The number, as well as nature of mutations in both antibodies, is different and they map to different regions (Table 2).

Analysis of multiple sequence alignment of the mature and respective germline forms suggests that mutations are random within a lineage and mutational frequency is high in the CDRs as compared to framework regions. Variations in the framework region contribute to the orientation of V_H-V_L pairing and support the antigen-binding site [19, 20]. Despite sequence variation, the canonical structures of CDRs L2 and H1 of V_H1-84 antibodies were conserved as they belonged to class 1 category. CDRL2 of V_H5-51 antibodies assumed a

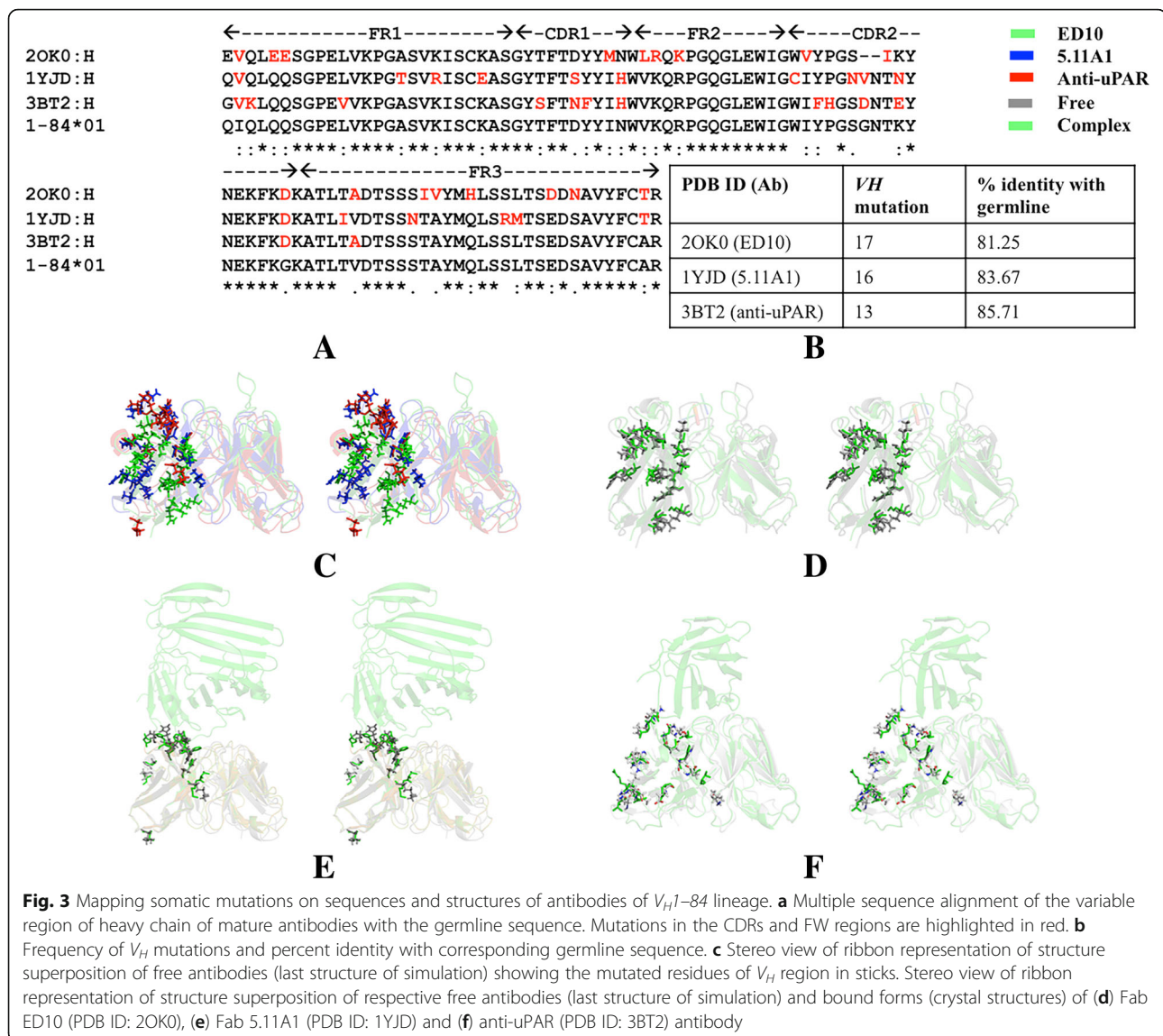
common canonical class 1 category. This indicates that despite somatic mutations led sequence variation, some CDR loops of the antibodies had conserved structural framework. Further examination of the conformational ensemble using molecular dynamics simulation will advance our understanding of the structural principles that govern maturation associated binding to unrelated antigens, while bearing some degree of structural connectivity.

Conformational selection leads to structural divergence during maturation

Affinity maturation associated changes in V_H1-84 and V_H5-51 lineages were comprehended by carrying out 0.5 μ s all-atom molecular dynamics simulation of bound and free mature counterparts. Conformational ensembles were analyzed by subjecting trajectories of bound and respective free forms together to k-mean clustering protocol. Frames with similar conformations could be clustered using $C\alpha$ radius of 1.5 Å from the centroid. Apart from overall structural variations and connectivity of structural region, changes in the core of the paratope, in particular, were traced.

Bound antibodies sampled fewer conformers as compared to their free forms. The number of sampled conformers for the free V_H1-84 antibodies was 8 (Fab 5.11A1) and 6 (anti-uPAR, Fab ED10). Bound forms had a highly restricted window of 2 (bound 5.11A1, bound ED10) and 3 (bound uPAR) clusters (Fig. 5). For V_H5-51 lineage, free states sampled 5 (Fab 10G5H6) and 4 (Fab m66) conformers respectively while bound m66 and 10G5H6 Fabs sampled 1 and 2 conformers respectively (Fig. 6). A single molecule was present in one of the clusters of bound 10G5H6, hence it was insignificant to be considered. Thus the alteration in the micro-environment brought about by antigenic stimulus narrows the specificity window of antibody.

Further analysis of the sampled conformers showed that in all cases the bound form of the antibodies selected one from the spectrum of conformers sampled by the free form (same colour code of conformers in Figs. 5 and 6 between bound and free forms represent



common cluster). The number of frames in a cluster was expressed as a percentage of the population adopting a conformation (Lower panels in Figs. 5 and 6). Conformation with highest percent population is the dominant conformation of the molecule. In case of ED10 the population of the dominant conformation for DNA bound antibody was 99.6%, while the population of the same conformer was only 11.6% in its free form. For bound anti-uPAR it was 92.4% as against 0.4% in its free form (Lower panel in Figs. 5). Similarly for *V_H*5–51 antibodies, dominant conformation for bound 10G5H6 was 80.2% as opposed to 29.2% in its free form and the population for bound m66 was 100% as against 1.4% in its free form (Lower panel in Fig. 6). Thus the free form of antibody samples all possible conformers of which one is naturally selected and best optimized for preferential antigen binding. In case of 5.11A1, the free form did

not sample the dominant conformation of the bound form. This indicates that the antibody presumably undergoes induced-fit and assumes a different conformation of the paratope to accommodate antigen CD28 (Lower panel in Fig. 5a).

Representative frames from conformational landscape of bound as well as free forms were analyzed by structure superposition to assess changes in the paratope. Each of the bound mature variants adopted a distinct conformation that did not overlap conformational window of the other bound antibodies. The landscape remained highly restricted; corresponding to the native crystal structures (upper panel in Fig. 7). This was true for the free forms as well barring the long loops of CDRL1 in ED10 which is also reflected as jumps in the RMSD plot (Fig. 5b) and CDRH3 in m66 that showed relatively wide conformational space (lower panel in Fig. 7). As observed in the compiled data,

Table 2 Somatic mutation and interacting residues in antibodies of V_H5-51 and V_H1-84 lineages

| Germline lineage | PDB ID | V_H mutation | Interacting residues on H-chain |
|------------------|--------|----------------------------------------------------------------------------------|-----------------------------------------|
| V_H5-51 | 4NRX | Q1, V24, A30, E32, K58, I68, N76, I89 | S31, W33, D54, D56, Y100C, R100F, T100G |
| | 4HWB | V51, Y58 | S32, D56, S57, R60, N102, W34 |
| V_H1-84 | 2OK0 | V2, E5, E6, M33, L36, R37, K39, V51, I56, D65, A71, I77, V78, H81, D85, N87, T92 | N35, G98 |
| | 1YJD | V2, T16, R19, E23, S31, H35, C50, N55, V56, N59, D65, I71, R85, M86, T97 | S31, Y101, G102, D104 |
| | 3BT2 | V2, K3, V11, S28, N31, F32, H35, F52, H52A, D55, E58, D65, A70, | Y33, W50, D55, N56, T57, E58 |

variability of CDRs of both chains was apparent within a lineage across the dataset. So, we evaluated the dynamics of the loops in the complexes of both lineages. While variability was seen in all loops, CDRH2 was relatively less impacted by flexibility of these loops. The mobility of CDRH3 was complemented by the relative shifting of VL and marginally by CDRH1 in V_H1-84 lineage. Apparently, since the conformational spectrum was narrow and the conformers of individual antibody showed good alignment in both lineages, not much information could be drawn. While the paratope in each case seemed to have diversified, pairwise comparison of all conformers and crystal structures suggested very similar topology of the backbone between sibling antibodies. Additional file 3: Figures S5 and S6 show heatmaps generated from RMSD matrix obtained from a one-to-one comparison of all conformers and crystal structures (that included the entire variable regions composed of CDRs and framework regions). Blue regions indicate conformational integrity between the antibodies presumably reflecting structural imprint of their common lineage. These findings suggest that despite an overall backbone structural connectivity, conformational selection of the paratope is imperative to favour preferential antigen binding within the available space. Our results are in conformity with the canonical theory of affinity maturation that suggests an optimized binding pocket while following isolated molecular pathways of maturation [13, 21].

Structural dynamics in light of V_H mutation

In order to check if the structural divergence between the mature antibodies is due to somatic mutation or the degree of variability increases due to binding of incoming antigen, the last structure of free antibody obtained from simulation was compared with that of the crystal structure of bound antibody by structure superposition. Superposition of the free states of all antibodies

obtained from simulation suggested structural divergence (Figs. 3c and 4c). Pairwise structural alignment of the free and respective bound states showed significant variation in the paratope region and particularly in mutated residues (Figs. 3d-f and 4d-e). Thus it can be said that while mutations tailor the overall geometry of individual antibodies of a lineage, it is the interaction with respective antigen that leads to additional topological alteration in their paratope.

Analyses of bonding pattern and binding energy of the complexes

The trajectories were further examined to check interface bonding pattern. For distance cut off of 3.5 Å, H-bonds between antibody and antigen were calculated using CPPTRAJ module of AMBER14. H-bonds with or above 30% occupancy across the trajectories are presented in Additional file 3: Tables S3 and S4. In V_H1-84 lineage, anti-uPAR complex had 15 bonds (ASN_190 and TRP_325, GLY_217 and TYR_308, GLN_189 and TYR_483 being the most stable ones with 93, 93 and 88% occupancy respectively), Fab 5.11A1 complex had 4 (stable ones being GLU_316 and TYR_205, GLY_206 and TYR_280 each with 49% occupancy) and Fab ED10 complex had 6 H-bonds (with the highest occupancy being 66% between different atoms of DT5_1 with ASN_147 and TYR_145) (Additional file 3: Table S3). Fab m66 and Fab 10G5H6 complexes of V_H5-51 lineage had 10 (stable being between TYR_107 and SER_245 with 74% occupancy and between LEU_240 and TRP_33 with 73% occupancy) and 15 H-bonds (most stable was between TYR_318 and LYS_93 with 94% occupancy) respectively (Additional file 3: Table S4). Different sets of H-bonds in the trajectories of the mature variants was a manifestation of mutations during maturation, suggesting independent developmental pathways. Binding surfaces for individual antigens followed the standard principle [22]. A flattened surface was assumed by anti-protein antibodies viz. anti-uPAR, Fab 5.11A1 and Fab 10G5H6 antibodies as opposed to anti-peptide antibody (Fab m66). Due to the small size of DNA fragment, Fab ED10 showed a small groovy binding site buried deeper in the V_H-V_L interface. Together these findings suggest how the physicochemical factors govern reorganization to foster shape complementarity.

Change in overall number of H-bonds during maturation has been linked with the type of antigen bound, but, its direct association with binding energetics has not been established [23–26]. As a measure of affinity, the free energy of binding (ΔG) was calculated from the trajectories using molecular mechanics generalized Born surface area (MM-GB/SA) approach. While entire 0.5 μ s trajectories were used for computing binding free energies, average ΔG for each complex was evaluated from the last 100 ns of the trajectories. The average ΔG was -38.75 kcal/mol for Fab

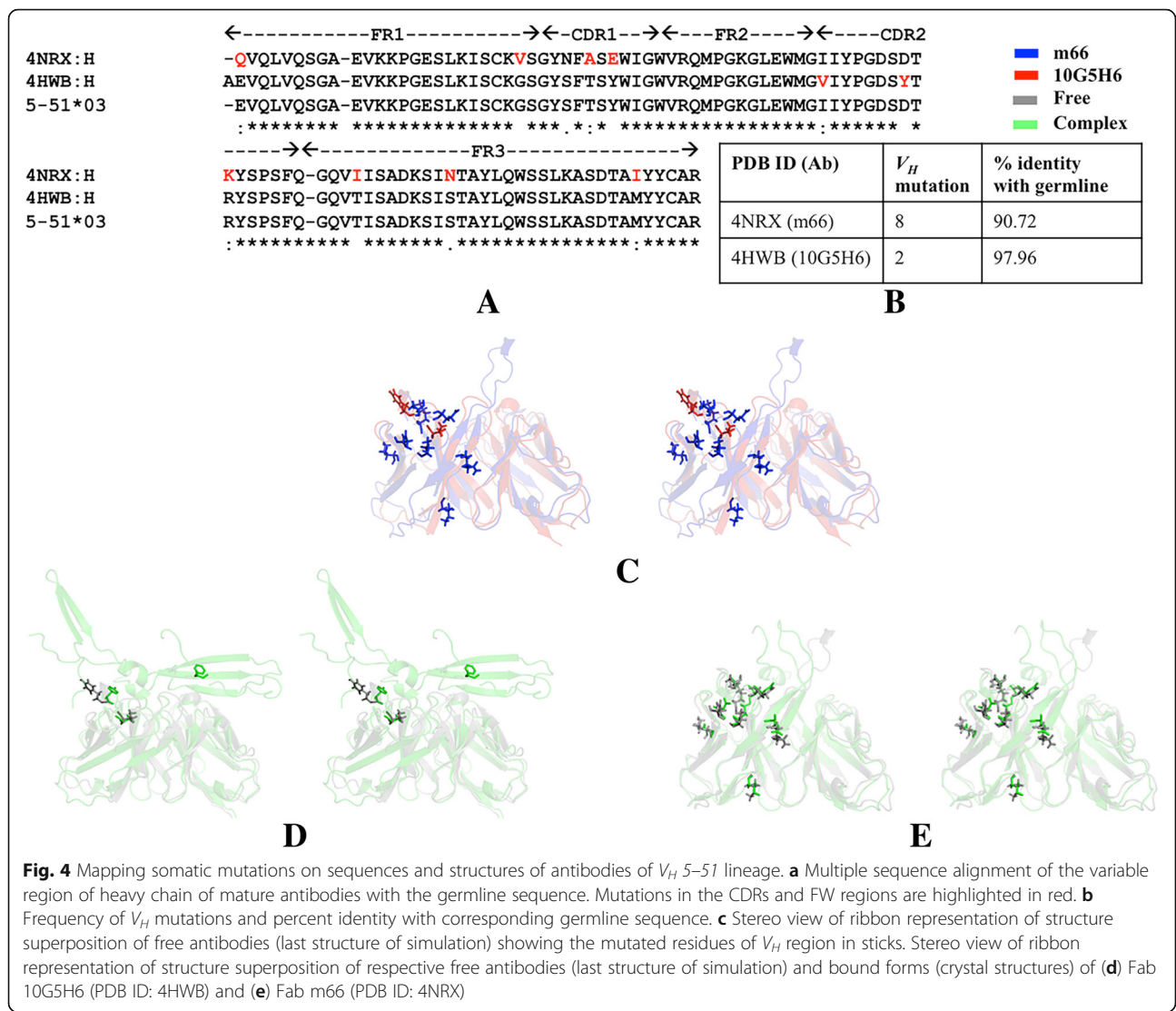


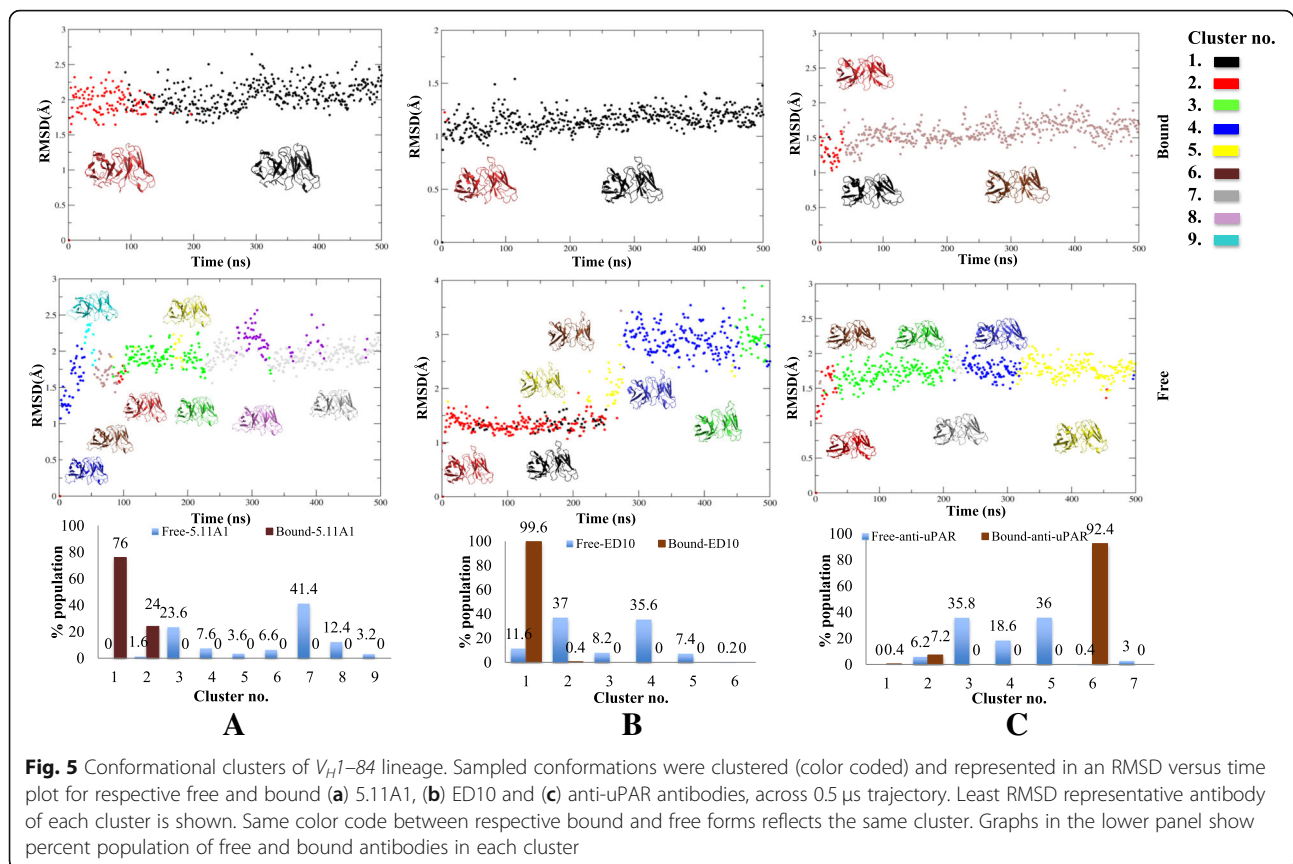
Fig. 4 Mapping somatic mutations on sequences and structures of antibodies of V_H 5-51 lineage. **a** Multiple sequence alignment of the variable region of heavy chain of mature antibodies with the germline sequence. Mutations in the CDRs and FW regions are highlighted in red. **b** Frequency of V_H mutations and percent identity with corresponding germline sequence. **c** Stereo view of ribbon representation of structure superposition of free antibodies (last structure of simulation) showing the mutated residues of V_H region in sticks. Stereo view of ribbon representation of structure superposition of respective free antibodies (last structure of simulation) and bound forms (crystal structures) of **(d)** Fab 10G5H6 (PDB ID: 4HWB) and **(e)** Fab m66 (PDB ID: 4NRX)

5.11A1 complex, - 53.38 kcal/mol for anti-uPAR complex and - 42.45 kcal/mol for Fab ED10 complex of V_H I-84 lineage (Fig. 8a). In V_H 5-51 lineage, average ΔG was - 92.48 kcal/mol for Fab m66 complex and - 68.28 kcal/mol for Fab 10G5H6 complex (Fig. 8b). Based on past studies, relative free binding energies were calculated by excluding entropy from the analysis as it is not considered to be predictive of accuracy [27, 28]. Hence, the values obtained for the antibody complexes are not exact match but relative values between different states that correlate with experimental values. The rule identified for governing affinity between complexes is largely dictated by the nature of ligand /antigen. In protein-protein systems studied till date, it is seen that affinity maturation contribute modestly to the binding energetics as compared to protein-hapten systems [25, 26]. For haptens, the increase in affinity is attributable to additional H-bonds, co-

operative binding or significant structural alterations while the same is not true for proteins [29]. Different sets of attractive interactions affect the binding energies by an increase in enthalpic gain [11].

Discussion

The study was aimed at understanding what factors shape antibodies coming from a common germline lineage to be able to bind significantly distinct antigens. While there are past reports that have analyzed maturation associated evolution, the contexts, however, were different from ours. Some reports are based on sequence analyses alone while some emphasize on dynamics of antibody evolution to a common epitope. The prime focus of our study was to strictly examine structural diversification of mature antibodies for a wide range of chemically distinct antigens. The approach of the study

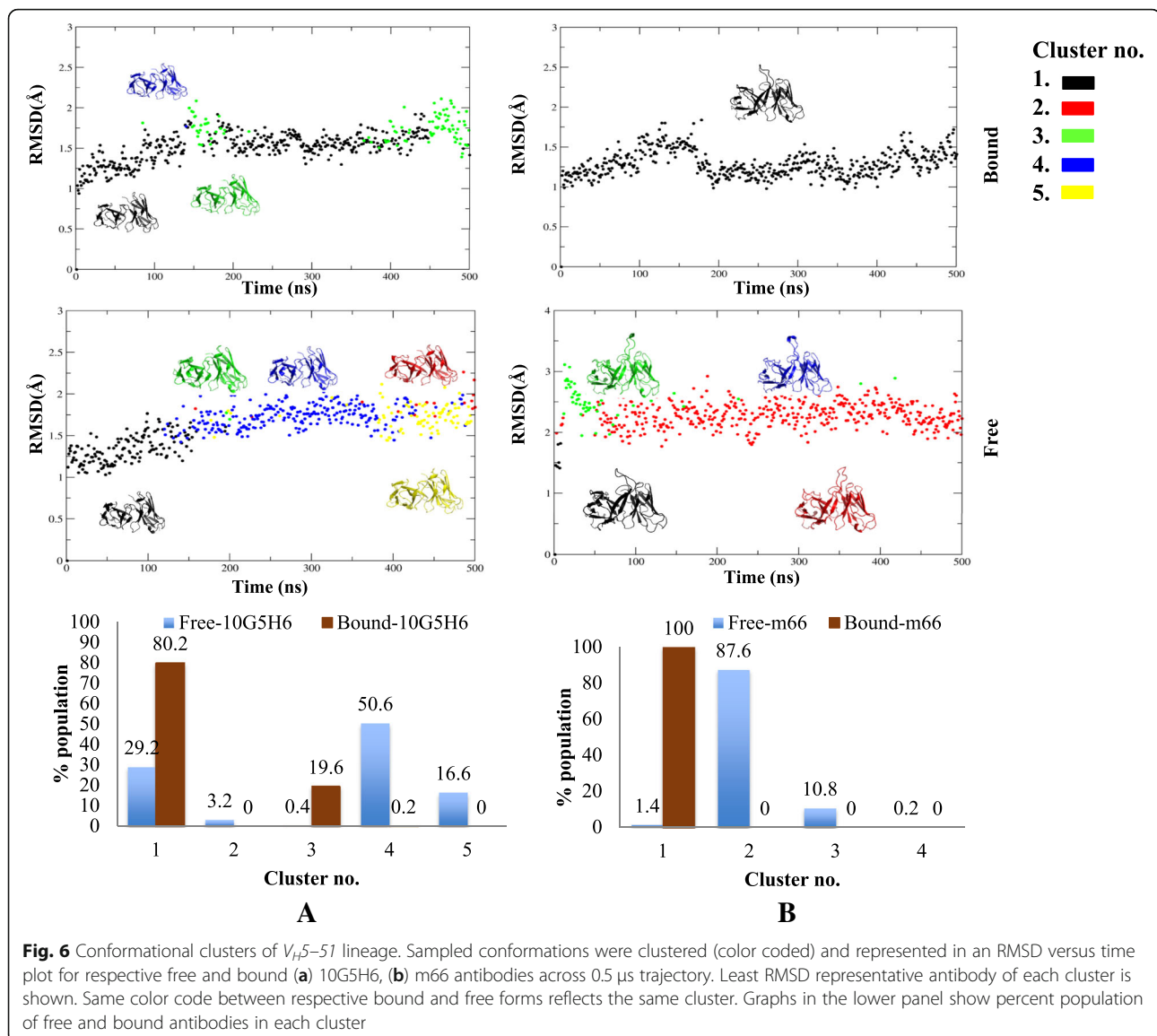


is data-driven, where available crystallographic structures have been analyzed and inferences drawn are validated by characterizing dynamics within lineages.

Analyses of static structural data from 35 (mouse) and 13 (human) IGHV families coupled with different light chain genes revealed that antibodies of a certain descent undergo topological tailoring of their binding pockets mediated by all the CDR loops to accommodate distinct antigens. The contribution of individual loop, however, varies across lineages. The dataset transcends a wide range of antigens covering hapten (24% in mouse and 5% in human), peptide (27% in mouse and 21% in human), sugar (5% in mouse and 5% in human), protein (41% in mouse and 69% in human) and nucleic acid (3% in mouse) (Fig. 1) that are typical in pathogens. Hence, conceptually our study is of credence as the diversity of epitopes facilitates a realistic understanding of host-pathogen interaction.

Sequence examination of two lineages indicates that different number and nature of mutations differentiate the antibodies from the common germline precursor. During affinity maturation, repeated rounds of mutations result in various intermediate stages of sequence diversification; hence the antibodies presumably represent different stages of evolution each following an isolated maturation pathway [30]. If the extent of

maturation of the reported structures could be investigated, additional insights could have been drawn. Sequence variations within the hypervariable regions shift the canonical structure framework relative to individual CDR loops to accommodate distinct antigens and maintain complementarity of interacting surfaces by reducing entropic cost [11, 31]. Further conformational landscape reveals that despite sequence variation led structural re-arrangement, the overall backbone geometry is similar. It can be envisaged that a common structural imprint of the germline is inherited. This suggests that while affinity maturation is highly stochastic, the evolution of the antibody repertoire is shaped by structural constraints. Examination of paratope sheds light on how one conformational state is naturally selected from the available spectrum and optimized to favor antigen binding. It is the optimization that brings about divergence in the paratope of individual antibody; evident from non-overlapping paratope topologies. Studies that reported maturation associated pre-ordering of the antibody combining site to favor antigen binding corroborates these findings [24, 32–35]. Present concepts on protein structure and function postulate coexistence of several functional states even for a highly specific enzyme. It is the population of a distinct conformation that determines the specificity of the molecule [36]. The



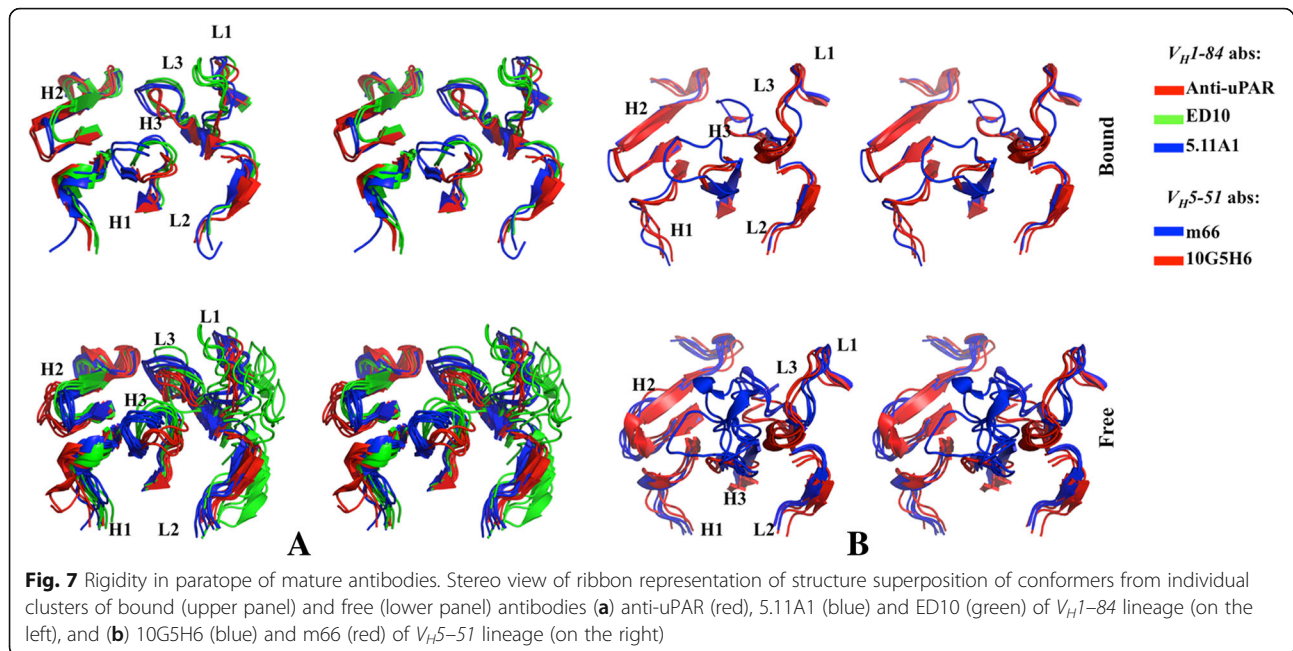
observed homogeneity in bound antibody indicates selection of a functional state, thereby demonstrating narrowing down of specificity window leading to molecular divergence during affinity maturation [35, 37–39]. An analogy to this can be quoted in Darwin’s words from *The Origin of Species*, “from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved” [40].

It may be noted that the observations made from simulation studies provide an interesting perspective to the interrogation of structural changes in sibling antibodies that target different antigens. Whether the antigen is self or foreign has no bearing on our observation as the study primarily dwells with identifying structural principles of interaction governing evolution. Due to practical limitations, conclusions have been drawn from simulations of 5

complexes from 2 IGHV lineages. Subjecting more complexes to simulation would increase the confidence of the interpretation. Additionally, the hypothesis derived from the study, as is the case with all computational analyses, has to be further validated experimentally. The best experimental follow-up of this study would be to make recombinant germline antibodies and determine their three dimensional structures by X-ray diffraction in antigen bound and unbound states. Subsequently, simulation studies can be conducted to sample conformational space of the germline complexes and compare them with the mature counterparts to decipher structural diversification in the lineage.

Conclusion

The computational analysis of global data followed by dynamics illustrates the molecular mechanism of how



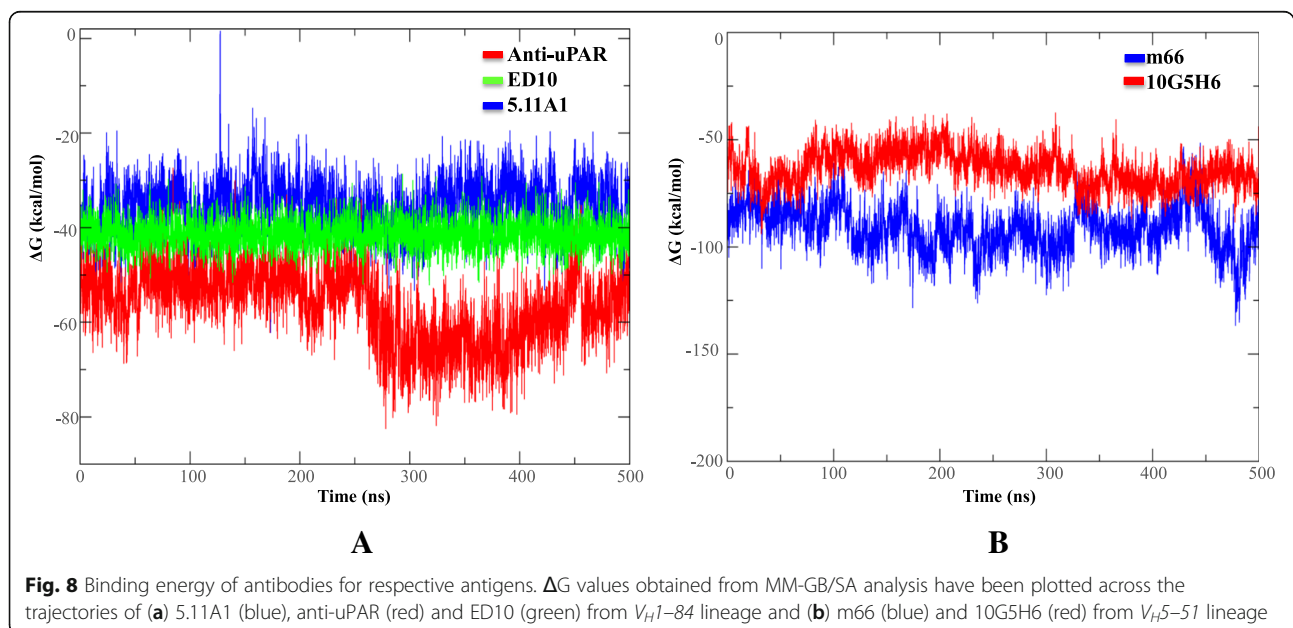
modulation in the structure is paramount to biomolecular recognition. Data from various computational approaches used in this study support the existence of different routes of maturation. Structural adaptation of the paratope despite conservation of overall backbone architecture is a pivotal finding of the study. The results of the study present an interesting implication of molecular evolution leading to the generation of antibody diversity. Additionally, our analysis displays characteristic that can perpetuate the antibody diversity model [41] whereby the 4th level of

diversity is attained after somatic diversification of a lineage leading to recognition of diverse antigens.

Methods

Data retrieval and compilation

Coordinate files of Ab-Ag complexes were retrieved from RCSB PDB (www.rcsb.org/pdb). Data so obtained were filtered and subjected to mining as mentioned in the result section. CDRs were identified using Kabat numbering system [42]. Coordinate files of complexes



were retrieved and segregated based on source of antibody into two groups, human and mouse. While more than 90% of the retrieved antibodies were obtained from immunized house mouse, the strain of the mouse was not considered as a criterion for selection of the antibodies as no two individuals have identical genetic make up, hence the immune repertoire would also vary.

Identification of germline origin of the antibodies and their clustering

Candidate sequences were queried for germline genes in IMGT Database using Ig BLAST tool 1.3.0 at NCBI with default settings [43, 44]. The antibodies were then clustered based on common germline V_H origin. Data sharing no common origin, were discarded while the rest were grouped.

Structural analysis

Structure based sequence alignment of all the lineages were performed in Chimera 1.11.2 [45]. One antibody was randomly chosen as reference with which other structures were matched. RMSDs of CDRs were noted. Contacts between the complexes were noted from PDBsum [46]. For multi-subunit antigen, the bonds between antibody and each chain of the antigen are added and reported. In complexes where PDBsum did not fetch any interaction information, PISA 1.48 [47] was used to identify the contacts.

Selection of system

Three mature antibodies, anti-uPAR, Fab 5.11A1 and Fab ED10 that bound to uPAR, CD28 and DNA respectively of V_H1-84 lineage from mouse formed a system (PDB ID: 3BT2, 1YJD, 2OK0 respectively). Two of four antibodies, Fab 10G5H6 bound with ectodomain D3 of IL-13 and antibody m66 bound with gp41 MPER (Membrane Proximal External Region) peptide (PDB ID: 4HWB, 4NRX respectively) of V_H5-51 lineage from human comprised of the other system and were chosen for simulation. Since in human, the other 2 antibodies Fab 2558 and Fab CH58 also bound to peptides, therefore these were excluded in the study to maintain distinctness of epitope. Only molecules directly interacting with the antibody were retained, rest were deleted. For antibody, only Fv (fragment variable) region constituting of CDRs and framework regions was retained. Antigens from each of the complexes were removed to generate free form of antibodies.

Sequence analysis

Multiple sequence analysis of the antibodies and their corresponding germline V_H -gene was performed online using CLUSTAL omega (default settings) because of accuracy of alignment [48]. Mutations were identified from the

alignment. Percent identity matrix was generated from the alignment to obtain identity between the sequence with germline counterpart. Canonical classes of the CDRs were assigned using strict Chothia SDR templates (<http://www.bioinf.org.uk/abs/chothia.html>) [12, 31].

Molecular dynamics simulation

Simulation set-up

All the starting heteromeric structures were provided as input to tLEaP module in AMBER14 package to generate topology and coordinate files [49]. Molecular mechanics parameters were assigned using ff12SB force-field [50]. The molecules were explicitly solvated using TIP3P water box with box edges lying 10 Å from the outermost atoms of the proteins in all directions. Charge of the system was neutralized with monovalent counter ions, Na⁺ or Cl⁻. Prior to subjecting to simulation, energy minimization was performed for 5000 steps with steepest descent for first 2500 steps followed by conjugate gradient for rest. If steric clashes persisted, minimization cycle was increased. Systems were heated to 300 K during a 14 ps dynamics simulation using the NVT ensemble. The temperature of the system was constrained using Langevin dynamics temperature coupling with a time step of 2 fs. Pressure was equilibrated to 1 atm over a period of 10 ps using isotropic position scaling keeping the temperature constant at 300 K. A third equilibration was run for 100 ps to stabilize the system. Production MD run was conducted using the NPT ensemble for 0.5 μs at 300 K and 1 atm for each system. Snapshots were saved at an interval of 10 ps. All the MD simulations were performed using Sander and a parallel CUDA version of PMEMD from AMBER14 [51, 52]. All simulations were performed in-house using High Performance Computing (HPC) facility with NVIDIA K20X GPUs.

Analysis of MD trajectories

MD trajectories were visualized in Visual Molecular Dynamics (VMD) [53] and analyzed using CPPTRAJ module of AMBER14 [54]. Structure superpositions of the frames were done using PyMOL 1.8.0.6. To analyze the extent of conformational sampling and to compare conformers obtained, snapshots extracted at an interval of 1 ns from the 0.5 μs trajectories were clustered together. Clustering was implemented between respective free and bound forms by subjecting to k-mean clustering with 1.5 Å radius from the centroid using the kclust utility in MMTSB (Multi-scale Modeling Tools for Structural Biology) suite [55]. The k-means clustering tries to find k groups from n data points depending on their nearest mean to each cluster [56]. Conformational landscape obtained from clustering are depicted in an RMSD vs time plot where different colors correspond to different

clusters. Highly flexible residues far from the binding pocket (if any) were removed during RMSD calculation. To decipher overall structural correlation between all the conformers, their backbone atoms were aligned and a heatmap was generated with the pairwise RMSD scores using pheatmap package of R [57]. Somatic mutations identified from sequence analysis were mapped in the frames. For contact assignment, H-bond interactions were noted. Side-chain hydrogen bond distance cut-off was set to 3.5 Å between donor/acceptor atoms for frames extracted at 100 ps. Molecular mechanics generalized Born surface area (MM-GB/SA) was used for a relative comparison of free energy of interaction between the complexes [58]. The method was implemented using MMPBSA.py utilizing generalized Born (GB) implicit-solvent model [58]. Free energy of binding was calculated according to the following equation:

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{receptor}} - \Delta G_{\text{ligand}}$$

Additional files

Additional file 1: Table S1. Mouse antibodies of common germline VH (variable region of heavy chain) lineages complexed with antigens. PDB IDs of complexes tabulated. (XLSX 34 kb)

Additional file 2: Table S2. Human antibodies of common germline VH (variable region of heavy chain) lineages complexed with antigens. PDB IDs of complexes tabulated. (XLSX 30 kb)

Additional file 3: Figure S1. Conformational heterogeneity in mouse data. Stereo images of structure superposition of mouse antibodies of a lineage showing variability in CDRs. **Figure S2.** Conformational heterogeneity in human data. Stereo images of structure superposition of human antibodies of a lineage showing variability in CDRs. **Figure S3.** Contact analysis in mouse data. Stacked bar diagram representing number of H-bonds formed by CDRs of H and L chains of mouse antibody with bound antigens. PDB IDs of antibody complexes belonging to germline VH lineages are represented along Y-axis and number of H-bonds formed by each CDR loop is represented along X-axis. **Figure S4.** Contact analysis in human data. Stacked bar diagram representing number of H-bonds formed by CDRs of H and L chains of human antibody with bound antigens. PDB IDs of antibody complexes belonging to germline VH lineages are represented along Y-axis and number of H-bonds formed by each CDR loop is represented along X-axis. **Figure S5.** Heatmap of RMSD between conformers and crystal structures of antibodies of V_H1-84 origin. Pair wise structural comparison of crystal structures and all conformers (bound and free) obtained after clustering of antibodies of V_H1-84 origin from mouse, plotted along X-axis and Y-axis. Names of all conformers end with a number to represent the clusters. Crystal structures are named as bound-5.11A1 (PDB ID: 1YJD), bound-ED10 (PDB ID: 2OK0) and bound-anti-uPAR (PDB ID: 3BT2). Result is shown as a measure of RMSD in a gradient from blue (low) to red (high). **Figure S6.** Heatmap of RMSD between conformers and crystal structures of antibodies of V_H5-51 origin. Pair wise structural comparison of crystal structures and all conformers (bound and free) obtained after clustering of antibodies of V_H5-51 origin from human, plotted along X-axis and Y-axis. Names of all conformers end with a number to represent the clusters. Crystal structures are named as bound-m66 (PDB ID: 4NRX) and bound-10G5H6 (PDB ID: 4HWB). Result is shown as a measure of RMSD in a gradient from blue (low) to red (high). **Table S3.** H-bond (above 30 % occupancy across trajectory) of antibody complexes of mouse V_H1-84 lineage. **Table S4.** H-bond (above 30 % occupancy across trajectory) of antibody complexes of human V_H5-51 lineage. (PDF 8090 kb)

Abbreviations

Ab: antibody; Ab-Ag: antibody-antigen; Ag: antigen; CDR: Complementarity Determining Region; Gp: glyco-protein; IL: Interleukin; MD: Molecular dynamics; MPER: Membrane Proximal External Region; PDB: Protein Data Bank; PISA: Proteins, Interfaces, Structures and Assemblies; RMSD: Root Mean Square Deviation; VDJ: Variable Diversity Junction; V_H : Variable region of heavy chain; V_L : Variable region of light chain

Acknowledgements

None.

Funding

We acknowledge financial support from Department of Science & Technology and Department of Biotechnology, Government of India, to Dinakar M. Salunke and Council of Scientific and Industrial Research for fellowship to Harmeet Kaur.

Availability of data and materials

The structural data used in the study are available in Protein Data Bank (PDB). The IDs associated with the structural data are 1TPX, 2ADF, 1TET, 1NCC, 1WEJ, 2BDN, 2Q8A, 2VDO, 25C8, 1A3R, 2NR6, 4K2U, 1FNS, 1A2Y, 1E6J, 1NMB, 3LEY, 3RVV, 3G5Y, 1KCS, 1F90, 1BAF, 1C12, 1CF8, 2AJV, 3CFD, 1QLE, 1TZH, 2QHR, 2OTW, 1CLZ, 2AEP, 3OKD, 3PHO, 2G5B, 4HLZ, 1EJO, 1IND, 3LS4, 3CFB, 1KFA, 4HZL, 1YQV, 1QOY, 1YNK, 2BJM, 1CT8, 1NAK, 1DQJ, 4GAG, 1OB1, 2ZPK, 2OK0, 3BT2, 1YJD, 1IGJ, 1EGJ, 2VXT, 1F3D, 4DW2, 1FL3, 3IFP, 1GGI, 2ADJ, 2H1P, 3FFD, 3LIZ, 2OR9, 3HNS, 3OOR, 1MFE, 3VW3, 1MPA, 1MH5, 4ALA, 4FFV, 2YPV, 1V7M, 1WZ1, 3FO9, 3IET, 1SM3, 1Q72, 1M7D, 1UM5, 2DDQ, 2FR4, 2F58, 1OSP, 2J5L, 1QKZ, 2HKF, 3SGE, 1CBV, 3I50, 3VOW, 1JRH, 2R0W, 3RKD, 3IFO, 3O41, 3IGA, 4OII, 2I9L, 4AG4, 4BKL, 4DGI, 4ETQ, 3QUM, 4F2M, 1KNO, 2HRP, 4C83, 2GSI, 1CU4, 4DGV, 3U2S, 3ZTN, 2NYY, 2QQK, 4HHA, 4G7V, 3GBN, 2NXZ, 2DD8, 4DN4, 2CMR, 2JB6, 4JZO, 4HJ0, 4MWF, 2FX7, 3W9E, 3NPS, 4LST, 4HKX, 3D85, 4AL8, 3MXW, 3HAE, 3KDM, 3HI6, 3SO3, 2VXS, 1H0D, 3BN9, 4FP8, 4JFZ, 3MLX, 3HI1, 2YK1, 4JY4, 1IKF, 3BDY, 3DVG, 3IDX, 3S0B, 2H9G, 3K2U, 3NH7, 1OP5, 3H42, 3UJI, 4HPO, 4NRX, 4HWB, 3LSX, 2F5B, 3THM, 3TWC, 4G6F, 1Q1J. All data generated or analyzed during this study are included in this published article and its Additional file 1, Additional file 2 and Additional file 3.

Authors' contributions

Conceived and designed the experiments: DMS, HK. Performed the experiments: HK. Analysed the data: HK, NS, DM, DMS. Wrote the manuscript: HK, DMS. Manuscript is reviewed, discussed and approved by all authors.

Authors' information

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Regional Centre for Biotechnology, Biotech Science Cluster, Faridabad, Haryana 121001, India. ²Manipal Academy of Higher Education, Manipal, Karnataka 576104, India. ³National Institute of Immunology, New Delhi, Delhi 110067, India. ⁴International Centre for Genetic Engineering and Biotechnology, New Delhi, Delhi 110067, India.

Received: 17 August 2018 Accepted: 14 November 2018

Published online: 19 December 2018

References

- Tonegawa S. Somatic generation of antibody diversity. *Nature*. 1983; 302(5909):575–81.
- Victoria GD, Nussenzweig MC. Germinal centers. *Annu Rev Immunol*. 2012; 30:429–57.
- Kepler TB, Liao HX, Alam SM, Bhaskarabhatla R, Zhang R, Yandava C, et al. Immunoglobulin gene insertions and deletions in the affinity maturation of HIV-1 broadly reactive neutralizing antibodies. *Cell Host Microbe*. 2014;16(3): 304–13.
- Reason DC, Zhou J. Codon insertion and deletion functions as a somatic diversification mechanism in human antibody repertoires. *Biol Direct*. 2006;1:24.
- Cauerhff A, Goldbaum FA, Braden BC. Structural mechanism for affinity maturation of an anti-lysozyme antibody. *Proc Natl Acad Sci U S A*. 2004; 101(10):3539–44.
- Mesin L, Ersching J, Victoria GD. Germinal Center B Cell Dynamics. *Immunity*. 2016;45(3):471–82.
- Brooks CL, Muller-Loennies S, Borisova SN, Brade L, Kosma P, Hiramata T, et al. Antibodies raised against chlamydial lipopolysaccharide antigens reveal convergence in germline gene usage and differential epitope recognition. *Biochemistry*. 2010;49(3):570–81.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235–42.
- Kieber-Emmons T, Foster MH, Williams WW, Madaio MP. Structural properties of a subset of nephritogenic anti-DNA antibodies. *Immunol Res*. 1994;13(2–3):172–85.
- Yin J, Mundorff EC, Yang PL, Wendt KU, Hanway D, Stevens RC, et al. A comparative analysis of the immunological evolution of antibody 28B4. *Biochemistry*. 2001;40(36):10764–73.
- Willis JR, Brinye BS, DeLuca SL, Crowe JE Jr, Meiler J. Human germline antibody gene segments encode polyspecific antibodies. *PLoS Comput Biol*. 2013;9(4):e1003045.
- Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, et al. Conformations of immunoglobulin hypervariable regions. *Nature*. 1989; 342(6252):877–83.
- Vargas-Madrado E, Lara-Ochoa F, Almagro JC. Canonical structure repertoire of the antigen-binding site of immunoglobulins suggests strong geometrical restrictions associated to the mechanism of immune recognition. *J Mol Biol*. 1995;254(3):497–504.
- Saphire EO, Parren PW, Pantophlet R, Zwick MB, Morris GM, Rudd PM, et al. Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *Science*. 2001;293(5532):1155–9.
- Breden F, Lepik C, Longo NS, Montero M, Lipsky PE, Scott JK. Comparison of antibody repertoires produced by HIV-1 infection, other chronic and acute infections, and systemic autoimmune disease. *PLoS One*. 2011;6(3):e16857.
- Kidd MJ, Chen Z, Wang Y, Jackson KJ, Zhang L, Boyd SD, et al. The inference of phased haplotypes for the immunoglobulin H chain V region gene loci by analysis of VDJ gene rearrangements. *J Immunol*. 2012;188(3):1333–40.
- Wang Y, Jackson KJ, Gaeta B, Pomat W, Siba P, Sewell WA, et al. Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants. *Immunogenetics*. 2011; 63(5):259–65.
- Collins AM, Wang Y, Roskin KM, Marquis CP, Jackson KJ. The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Philos Trans R Soc Lond B Biol Sci*. 2015;370(1676): 20140236.
- Foote J, Winter G. Antibody framework residues affecting the conformation of the hypervariable loops. *J Mol Biol*. 1992;224(2):487–99.
- Abhinandan KR, Martin AC. Analysis and prediction of VH/VL packing in antibodies. *Protein Eng Des Sel*. 2010;23(9):689–97.
- Lara-Ochoa F, Almagro JC, Vargas-Madrado E, Conrad M. Antibody-antigen recognition: a canonical structure paradigm. *J Mol Evol*. 1996;43(6):678–84.
- Finlay WJ, Almagro JC. Natural and man-made V-gene repertoires for antibody discovery. *Front Immunol*. 2012;3:342.
- Midelfort KS, Hernandez HH, Lippow SM, Tidor B, Drennan CL, Witttrup KD. Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J Mol Biol*. 2004;343(3):685–701.
- Wedemayer GJ, Patten PA, Wang LH, Schultz PG, Stevens RC. Structural insights into the evolution of an antibody combining site. *Science*. 1997; 276(5319):1665–9.
- De Genst E, Handelberg F, Van Meirhaeghe A, Wynck S, Loris R, Wyns L, et al. Chemical basis for the affinity maturation of a camel single domain antibody. *J Biol Chem*. 2004;279(51):53593–601.
- Sundberg EJ, Andersen PS, Schlievert PM, Karjalainen K, Mariuzza RA. Structural, energetic, and functional analysis of a protein-protein interface at distinct stages of affinity maturation. *Structure*. 2003;11(9):1151–61.
- Hou T, Wang J, Li Y, Wang W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model*. 2011;51(1):69–82.
- leong P, Amaro RE, Li WW. Molecular dynamics analysis of antibody recognition and escape by human H1N1 influenza hemagglutinin. *Biophys J*. 2015;108(11):2704–12.
- Cho S, Swaminathan CP, Yang J, Kerzic MC, Guan R, Kieke MC, et al. Structural basis of affinity maturation and intramolecular cooperativity in a protein-protein interaction. *Structure*. 2005;13(12):1775–87.
- Adhikary R, Yu W, Oda M, Walker RC, Chen T, Stanfield RL, et al. Adaptive mutations alter antibody structure and dynamics during affinity maturation. *Biochemistry*. 2015;54(11):2085–93.
- Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol*. 1987;196(4):901–17.
- Romesberg FE, Spiller B, Schultz PG, Stevens RC. Immunological origins of binding and catalysis in a Diels-Alderase antibody. *Science*. 1998;279(5358): 1929–33.
- Jimenez R, Salazar G, Yin J, Joo T, Romesberg FE. Protein dynamics and the immunological evolution of molecular recognition. *Proc Natl Acad Sci U S A*. 2004;101(11):3803–8.
- Thielges MC, Zimmermann J, Yu W, Oda M, Romesberg FE. Exploring the energy landscape of antibody-antigen complexes: protein dynamics, flexibility, and molecular recognition. *Biochemistry*. 2008; 47(27):7237–47.
- Zimmermann J, Oakman EL, Thorpe IF, Shi X, Abbyad P, Brooks CL 3rd, et al. Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics. *Proc Natl Acad Sci U S A*. 2006;103(37): 13722–7.
- Ma B, Nussinov R. Protein dynamics: conformational footprints. *Nat Chem Biol*. 2016;12(11):890–1.
- Terzyan S, Ramsland PA, Voss EW Jr, Herron JN, Edmundson AB. Three-dimensional structures of idiotypically related Fabs with intermediate and high affinity for fluorescein. *J Mol Biol*. 2004;339(5):1141–51.
- Sagawa T, Oda M, Ishimura M, Furukawa K, Azuma T. Thermodynamic and kinetic aspects of antibody evolution during the immune response to hapten. *Mol Immunol*. 2003;39(13):801–8.
- Manivel V, Sahoo NC, Salunke DM, Rao KV. Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity*. 2000;13(5):611–20.
- Darwin C, R. *The Origin of Species*. The Harvard Classics. New York: P.F. Collier & Son; 1858.
- Khan T, Salunke DM. Structural elucidation of the mechanistic basis of degeneracy in the primary humoral response. *J Immunol*. 2012;188(4): 1819–27.
- Kabat EA, Wu TT, Bilofsky H, Reid-Miller M, Perry H. Sequence of proteins of immunological interest. Bethesda: National Institutes of Health; 1983.
- Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res*. 2013;41(Web Server issue):W34–40.
- Giudicelli V, Chaume D, Lefranc MP. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res*. 2005;33(Database issue):D256–61.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605–12.
- Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM. PDBsum: a web-based database of summaries and analyses of all PDB structures. *Trends Biochem Sci*. 1997;22(12):488–90.
- Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372(3):774–97.

48. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Mol Syst Biol.* 2011;7:539.
49. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, et al. The Amber biomolecular simulation programs. *J Comput Chem.* 2005; 26(16):1668–88.
50. Ponder JW, Case DA. Force fields for protein simulations. *Adv Protein Chem.* 2003;66:27–85.
51. Gotz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput.* 2012;8(5):1542–55.
52. Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J Chem Theory Comput.* 2013;9(9): 3878–88.
53. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14(1):33–8 27-8.
54. Roe DR, Cheatham TE 3rd. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput.* 2013;9(7):3084–95.
55. Feig M, Karanicolas J, Brooks CL 3rd. MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph Model.* 2004;22(5):377–95.
56. Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley Symposium on Mathematical Statistics and Probability. University of California Press; 1967.
57. Team RC. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2013.
58. Miller BR 3rd, McGee TD Jr, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA. py: An efficient program for end-state free energy calculations. *J Chem Theory Comput.* 2012;8(9):3314–21.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

