SARS-CoV-2 antibodies recognize 23 distinct epitopic sites on the receptor binding domain

Jiansheng Jiang
National Institutes of Health https://orcid.org/0000-0003-0964-5481
Christopher Boughter
NIAID (National Institute of Allergy and Infectious Diseases)
javeed Ahmad
NIAID/NIH
Kannan Natarajan
National Institute of Allergy and Infectious Diseases https://orcid.org/0000-0002-6295-2571
Lisa Boyd
NIAID/NIH
Martin Meier-Schellersheim
National Institute of Allergy and Infectious Diseases https://orcid.org/0000-0002-8754-6377
David Margulies (dmargulies@niaid.nih.gov)
Molecular Biology Section/NIAID/NIH https://orcid.org/0000-0001-8530-7375

Article

Keywords:

Posted Date: May 18th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2800118/v1

Additional Declarations:
There is **NO** Competing Interest.

Tables 1 to 3 are available in the Supplementary Files section.

SARS-CoV-2 antibodies recognize 23 distinct epitopic sites on the receptor binding domain

Jiansheng Jiang[1,*], Christopher T. Boughter[2], Javeed Ahmad[1], Kannan Natarajan[1], Lisa F. Boyd[1], Martin Meier-Schellersheim[2], David H. Margulies[1,*]

[1] Molecular Biology Section, and [2] Computational Biology Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 10892, USA

*Corresponding Authors: jiangji@niaid.nih.gov, dmargulies@niaid.nih.gov

Abstract

The COVID-19 pandemic and SARS-CoV-2 variants have dramatically illustrated the need for a better understanding of antigen (epitope)-antibody (paratope) interactions. To gain insight into the immunogenic characteristics of epitopic sites (ES), we systematically investigated the structures of 340 Abs and 83 nanobodies (Nbs) complexed with the Receptor Binding Domain (RBD) of the SARS-CoV-2 spike protein. We identified 23 distinct ES on the RBD surface and determined the frequencies of amino acid usage in the corresponding CDR paratopes. We describe a clustering method for analysis of ES similarities that reveals binding motifs of the paratopes and that provides insights for vaccine design and therapies for SARS-CoV-2, as well as a broader understanding of the structural basis of Ab-protein antigen (Ag) interactions.

Introduction

Our ability to predict protein interactions is still very limited despite great progress in the application of computational methods for determining protein structures from amino acid sequence alone [1] [2]. This limitation is even more evident with regard to the interactions among highly variable immune receptor surfaces as dictated by Ab complementarity determining region (CDR) loops and the antigenic structures they bind. Accordingly, efforts directed toward providing systematic analyses or rational design strategies for Ab-Ag interactions need to incorporate experimentally determined structural data on specific Abs. Recent efforts in Ab design take advantage of segmental approaches [3] or extensive computational resources [4,5]. Such hindrances emphasize the importance of incorporating as much information on naturally occurring specific Ab-Ag structures as possible. Here, we report a systematic structural analysis, taking advantage of the thousands of structures of SARS-CoV-2-derived proteins, including spike and various Ab complexes that have been determined to further our understanding of the fundamental mechanisms of the pathogenesis and neutralization of SARS-CoV-2 in the context of the human immune system. Many Abs have been reported to have potent neutralizing activity, preventing spike interaction with the cellular receptor, angiotensin converting enzyme (ACE) 2. Several Abs have been developed as therapeutics and have variable efficacy against variants of concern (VOC). Our analysis of available structures may aid in understanding which Abs may be of value for emerging variants and contribute to evolving strategies for prophylaxis, treatment, and immunization.

Ab-protein antigen (Ab-Ag) interfaces have been a focus of immunologists and protein chemists for more than 80 years [6], not only because of the important role of Abs in defense against infection [7], but also due to the general interest in understanding protein-protein interactions [8]. High resolution structural analysis of protein-protein complexes, based initially on X-ray crystallography and more recently on cryogenic electron microscopy (cryo-EM), provides an objective basis for understanding not only the biophysical principles that determine affinity and specificity, but also for elucidating biological and evolutionary rules that govern immunological molecular recognition of foreign molecules and pathogens [9,10]. With an ever-expanding database of detailed Ab-Ag structures, great attention has been directed to the characterization of such molecular interfaces, particularly as an understanding of the rules of engagement might permit rationalization of the reactivity of existing Abs, the design of Abs with new binding activities, and strategies for design of immunogens that might elicit more broadly neutralizing Abs [11-13].

The widespread infectivity, variance, and molecular characterization of the SARS-CoV-2 virus have provided a wealth of information concerning the functional and structural biology of the immune response. At the beginning of the SARS-CoV-2 pandemic, many laboratories accomplished detailed structural characterization of anti-RBD Abs and nanobodies (Nbs, single domain antibodies), leading to a classification of Abs based on the location of their footprints on the RBD surface. Initially, four classes of Ab were categorized, based on the orientation of the RBD bound and whether the Ab blocks infectivity or binding to the cellular receptor, ACE2 [14] (***Supplementary Table 1***). A receptor binding motif (RBM) has been defined as those RBD residues that specifically interact with ACE2 [15]. Binding analysis of Nbs and human mAbs derived from

patients along with a limited number of protein structures assigned five surface regions of the RBD reflecting its antigenic anatomy [16]. Epitopic analysis was further extended by the definition of seven "communities" of Abs that bind to the RBD surface [17]. Recent analysis of anti-RBD antibodies in the context of evolving escape mutations has taken advantage of these earlier classification schemes[18-21].

Although these classification schemes have been valuable and adopted widely in the analysis of Abs as to how they bind to RBD and spike, particular Abs and Nbs may not be unambiguously classified (***Supplementary Figure 1***). The previous summaries were based on a relatively small number of available structures and focused on the relative superposition of the Abs in the complexes, rather than on a comparison of the epitopic contacts of the RBD surface. In particular, the original distinction between Class 1 and Class 2 seemed clear based on the initial structures. However, as more structural models became available, apparent inconsistencies arose. For example, Ahmad et al [22] determined that synthetic Nbs Sb16 and Sb45 contacted both Class 1 and Class 2 epitopic surfaces and approached the RBD from different angles. As more structures of Ab and Nb complexes are determined, it is apparent that an expansion of the initial classification scheme is warranted.

In this work, we focus on complexes of Abs and Nbs bound to the RBD of the spike protein to generate a comprehensive structural framework to further our understanding of Ab- and Nb-RBD recognition. Using a large database, we offer a structure-based classification exploiting quantitatively defined contacting amino acid residues on the RBD as well as a clustering analysis. These analyses reveal common characteristics of some 23 frequently contacted ES and the structural nature of the surfaces of the RBD that interact with Ab/Nb. We also systematically

analyze the molecular features that define these antibodies and, by applying a rigorous evaluation of the surface features of the RBD that are seen by Abs and Nbs, generate general insights into the fundamental nature of Ab-Ag recognition. This analysis should facilitate the characterization of new anti-RBD antibodies as they arise.

Results

Identification of epitopic sites (ES)

To identify common features of ES of the RBD, we systematically investigated structures of Abs (as Fabs and Fvs, Ab fragments that confer antigen binding activity) and of Nbs (as VHH or synthetic library-derived sybodies) in complex with the spike protein or its RBD as collected in the CovAbDab [23] and the protein data bank (PDB)[24,25]. Abs and Nbs that bind the SARS-CoV-2 RBD are summarized in ***Table 1***. As of 12/22/2022, a total of 6,746 Ab and 620 Nb sequences have been collected in the CovAbDab. Of the Abs, 6,321 are human, including those from vaccinees, and 390 derive from humanized mouse or phage display Ab libraries. For Nbs, 620 sequences derive from camelids (alpaca/camel/llama), of which 276 are from camelid-derived phage display libraries, some naïve, some immunized. Among these sequences, structural coordinates for only ~5% of the Abs and ~10% of Nbs were available in the PDB, and we compiled a non-redundant list of 340 Ab and of 83 Nb X-ray or cryo-EM structures (***Supplementary Table 2a & 2b***) which serve as the basis of our structural analysis.

Evaluation of the biophysical properties that contribute to protein-protein interactions may be based on different criteria, including calculation of free energy terms of interacting residues [26], measurement of shape complementarity (Sc [27]), and calculation of buried or accessible surface area [28-32]. We elected to simplify this analysis first by calculating interatomic contacts between Ab (paratopic) and Ag (epitopic) residues at the interface because the biophysical basis of binding (due to charge, hydrophobicity, hydrogen bonding and van der Waals interactions) is reflected in such contacts. We calculated distances between Ab and Ag interface residues with a cut-off of 5.0 Å (see Methods) and we plotted the numbers of Ab (paratope)

contacts as hits versus the residue number of the RBD (epitope) for the Ab heavy (H) (***Figure 1a***) and light (L) (***Supplementary Figure 2a***) chains individually, and also overall for both H and L chains together (***Supplementary Figure 2b***). We also plot the number of hits of the 83 Nbs to each RBD residue (***Figure 1c***). For 340 Abs, H chains contribute 5,623 contacts and L chains 3,107 (***Supplementary Table 3***). By comparison, for 83 Nbs, 1,836 contacts are observed. Thus, the number of contacts is ~25 per Ab and ~22 per Nb. Although the RBD residues bound by either Ab, H chain, or Nb are by and large, the same, the relative distribution of hits varies for several regions. In particular, the region from RBD residue 368 to 386 is recognized more frequently by Nbs, while other contiguous surfaces are seen equivalently (***Figure 1a & 1c***). The numbers of hits for Ab H chains are represented graphically as a heat map on the RBD surface in ***Figure 1b,*** and the heat maps for the Nbs are shown in ***Figure 1d***.

Several contiguous stretches of amino acids of the RBD that make Ab contact were apparent, although the frequency of hits varied considerably for different regions on the surface of the RBD. A fine-grained tabulation of regions of the RBD consisting of three to nine residues define each individual ES as shown in ***Table 2a***. Each of these ES may be assigned to either of the four major classes identified earlier or to the RBM recognized by the ACE2 receptor ***(Table 2b)***. These regions include distinct secondary structural features such as strands, loops, turns, and helices (***Supplementary Movie 1a***), and represent contacts seen by few (<0.3 %) to many (>10%) Abs. Consideration of the secondary structural features (loops, turns, or short β strands) and the accessible surface area prompts the identification of 23 distinct contiguous sites, including regions encompassing residues 404 to 421 that had been overlooked in previous studies. The hit numbers are not evenly distributed over the RBD surface, and it is difficult to distinguish which

binding sites belong to the previously defined Class 1 or Class 2 due to overlaps generated by the reduction of the three-dimensional surface to a two-dimensional plot. ***Figure 2a, b*** displays these ES on the RBD surface with the ES numbers for Abs (magenta) and Nbs (blue) respectively. The thickness of the putty cartoon indicates greater hit numbers. The computed accessible surface area (ASA) (see Methods) for each individual ES (***Table 2a*****)** ranged from ~100 $Å^2$ to more than 500 $Å^2$. The total buried surface area (BSA) is also computed for each of 340 Abs and 83 Nbs as in ***Supplementary Table 2a and 2b*** respectively. The values of BSA range from 106 $Å^2$ (PDB 6XDG) to 1112 $Å^2$ (PDB 7N64) for 340 Abs and from 444 $Å^2$ (PDB 7JVB) to 1412 $Å^2$ (7D2Z) for the 83 Nbs.

As an indication of the relative immunogenicity of each of the 23 ES, we tabulated the proportion of Abs and Nbs that recognized each site (***Figure 2c*****).** Approximately 7 to 11% of Ab H chains recognized ES11, 13, 16, 18, and 20, which represent ES contained within the previously defined Class 1 and Class 2 regions. In general, Nb recognition of specific ES was similar to that of Ab H chains, with the predominant recognition representing from 7 to about 10% of Nbs see ***Table 2a*** and ***Figure 2c***, falling within Class 2 and Class 4. Notable differences in the predominant ES recognized by Abs and Nbs are that ES8, 13, 16, and 18 are more frequently seen by Abs while ES4, 5, 6, 7, 11, and 20 are more frequently identified by Nbs. For example, ES16 was recognized by 10% of Abs and by 0.16% of Nbs. This difference may be explained since ES16 forms a solvent exposed convex structure which may not be conducive to recognition by Nbs. By contrast, ES4, 5, and 6 form a contiguous patch, recognized more frequently by Nbs, a region that is not exposed to solvent in the complete spike when the RBD is in the down position. Thus, Nbs may be better able to access such hidden surfaces, perhaps because of their relatively small size (12kD compared to ~25 or 50 kD for Fv and Fab respectively or ~150 kD for complete bivalent IgG, with

corresponding three-dimensional volumes) [33]. Alternatively, since many Nbs were identified based on binding to isolated RBD, some epitopes identified from such screens may be partially hidden in the complete spike protein. In comparing L chains with H chains, as shown in ***Figure 2d***, L chains generally contribute less to these ES. Nevertheless, L chains seem to preferentially contact ES7, 20 and 21. We note that some ES (e.g. ES7, 8, 9, and 23) could not be placed into the previous classification schemes and some sites overlap on Class 1 and Class 2 (i.e. ES12, 19, and 20). However, most of the 23 ES may be viewed within the four classes described by Barnes (***Table 2b***) [14]. In addition, the RBM of the RBD [15] may be defined in terms of the ES that overlap the ACE2-RBD interface (i.e. ES8, 11, 12, 13, 16, 18, 19, 20, 21, and 22 (***Table 2b***))***.*** With these 23 fine-grained ES, we extend the prior classification for Class 1 to now include ES8 and 9 (***Table 2b***). Each ES surface area or footprint is illustrated by a color map of the RBD surface (***Figure 2d, Supplementary Movie 1b***). The sum of these 23 ES covers as much as 70% of the total accessible surface area (ASA) of the isolated RBD, illustrating the breadth of the human antibody response to RBD.

Analysis of CDR loop contributions and epitope-paratope interactions

The CDRs in the hypervariable region of Abs play critical roles in recognizing antigens [9,34,35], and their variability in sequence and length facilitates interaction with distinct antigenic epitopes [36]. We tabulated the number of contacts for each CDR loop or non-CDR residues of 340 H chains and L chains and 83 Nbs to each of the 23 ES. The contact percentages are summarized in ***Figure 3a, 3b*** and ***3c*** respectively. The corresponding statistics are listed in ***Supplementary Table 3a, 3b*** *and* ***3c***. For Ab H chains (***Figure 3a***), CDR loops account for 82% of the contacts to

ES (CDR1=16%, CDR2=21%, CDR3=45%), while only 18% of the contacts are from non-CDR residues. Interestingly, CDR1 of H chains play a major role in binding to ES16. For Ab L chains (***Figure 3b***), CDR1 loops play a major role (40%) in binding to RBD while CDR3 represent only 25% of the contacts. One explanation for the reduced the role of the CDR3 loop of L chains might be that their average length (10 aa for 340 Abs) is generally shorter than that of H chain CDR3 (15 aa for 340 Abs), see ***Figure 3d***. For Nbs (***Figure 3c***), CDR represent 73% (CDR1=13%, CDR2=14%, CDR3=46%) of the contacts to the RBD surface, while 27% involve non-CDR residues. The average length of Nb CDR3 is 16 aa. Thus, for both Ab H chains and Nbs, CDR3 contributes the greater proportion of those residues that interact with the RBD, reflecting a major role for CDR3 in RBD recognition.

We plotted the frequency of particular amino acids used by Abs and Nbs (paratopic residues) that interact with particular ES of the RBD for Ab H chains (***Figure 4a***) and for Nbs (***Figure 4b***)**.** These are shown as heat maps. The residues listed on the top of the panel represent the most frequently contacting amino acids for the specific ES. The frequency of usage of each amino acid for Abs (pink) and Nbs (blue) is compared in ***Figure 4c***. Tyrosine (Y), serine (S), and arginine (R) are the three amino acids most preferred for binding any ES of RBD (***Figure 4c***). Previous analyses of paratopic preferences for a wide range of Abs recognized a high frequency of tyrosine usage [37]. We also observed that tryptophan is more frequently used in Nbs as compared with Abs (***Figure 4c***)**.** The usage of CDR3 amino acids is plotted in ***Figure 4d.*** To illustrate the predominance of particular paratopic residues of the Ab H chains that contact specific ES, we also grouped these as WebLogo plots [38] (***Supplementary Figure 3***).

Cluster analysis of epitopic sites and binding motifs

Having identified the sets of ES bound by each Ab and Nb (see ***Supplementary Table 2a,2b***), we then grouped the Abs and Nbs by computation of the similarity of the ES recognized (see Methods). Similarity of a pair of ES sets is a value between 0 and 1 reflecting recognition of completely different (0) or identical (1) sites. This clustering method compares ES sets on the RBD without visualization of graphic models. Assigning a similarity threshold of 0.85 (see Methods) results in the identification of 33 distinct, non-overlapping, clusters for Abs, designated A1 to A33 (***Supplementary Table 4a***) and 10 distinct clusters for Nbs, N1 to N10 (***Supplementary Table 4b***). Although Abs within a single cluster bind the same subset of ES, they may, or may not address the RBD from the same angle or utilize CDR of the same length or composition. These differences are illustrated in ***Figure 5a*** for clusters A1, A3, and A11 for H chains and in ***Figure 5b*** for clusters N1, N3, and N4 for Nbs. The members of nanobody cluster N4 reveal a similar orientation because they have the same conformation and length of CDR loops. Abs or Nbs within the same cluster recognize the same contiguous RBD surface and are expected to compete sterically.

CDR loops contain sequence motifs for epitope recognition [39-42]. To identify such motifs we analyzed a subset of interfaces from cluster A1, designated A1S1, that recognized ES with a similarity of ≥ 0.9. A1S1 consists of 28 members (cluster A1 has 56 members of similarity ≥ 0.85). All the members of A1S1 recognize the same ES set (ES8, 9, 12, 13, 16, 18, and 19) (***Figure 5c***), utilize the same CDR loops, and superpose well. Analysis of the residues of CDR1, 2, and 3 that contact the RBD indicated those residues that are preferentially utilized by this stringently selected cluster of Abs. For the binding motifs of CDR1, 2, and 3 of A1S1, the favored residues are summarized in a WebLogo plot (***Figure 5c***). Remarkably, Y, S, G, and T predominate for all

CDR except CDR3 which exploits R in most instances. Thus, application of a more stringent ES similarity score helps to identify the preferred binding motif utilized by the Ab of the same subgroup. This stringent grouping of Abs and Nbs, based on high similarity score of their respective ES, may prove a useful adjunct in structure prediction based on amino acid sequence and antibody competition.

To extend the utility of our ES definitions, we set out to determine broad biophysical trends common among the Abs that cluster to each ES region. Using the automated immune molecule separator (AIMS) software [43], a tool which characterizes immune molecules without structural knowledge, we analyzed similar SARS-CoV-2-specific Abs. With this we identified 11 clusters which are designated as AIMS1, AIMS2, etc ***(Figure 5d)***. Not all Abs in a single AIMS cluster bind the same ES. However, AIMS6 and AIMS7 overlap as subsets of cluster A1 and have a similarity score of 0.85.

Relation of ES and SARS-CoV-2 escape mutations

SARS-CoV-2 variants have evolved rapidly from Alpha, Beta, Delta, and Omicron with multiple mutations and deletions. The development of the latest Omicron subvariants can be traced from BA.1, BA.1.1, BA.2, BA.3, BA.4/5, and XBB.1 to XBB.1.5 and they incorporate as many as 30 mutations and deletions in their RBDs [44-46]. ***Table 6a*** lists the mutations in these variants and the ES to which they map. Subvariants marked "X" have different substitutions at a given position. ***Table 6b*** lists the major Omicron subvariants and their associated ES. (For example, XBB.1.5 has substitutions of P and S for V445 and G446, respectively, which are contained in ES11, and substitution of S and Q for F490 and R493, respectively, which are in ES19). Similarly, XBB.4

preserves the same substitutions, but also substitutes R for L452 in ES12. ***Figure 6a*** illustrates the location of these variants on the RBD surface for Omicron and their mutation sites are matched to one or more of the 23 ES. Strikingly, Omicron escape mutations are distributed throughout several distinct ES of the RBD (***Table 6a***, ***Figure 6a-d***), posing a formidable challenge in the design of new vaccines and therapeutic antibodies. Notably, mutations in ES3, 6, 9, 14, 15, and 23 have not yet been reported.

Our comprehensive analysis of RBD epitopes and their corresponding Ab paratopes offers the possibility of identifying currently approved SARS-CoV-2 therapeutic Abs that may be used to neutralize emerging SARS-CoV-2 variants and Omicron subvariants. The latest reported structures [41,47-49] describe some Abs that bind these subvariants. We can identify a number of Abs or Nbs that target particular ES sets that are either mutated or preserved in emerging variants. Those Abs/Nbs exhibiting multiple contacts to contiguous ES sites with concomitantly large buried surface area and high binding affinity deserve the greatest attention. Thus, using Ab/Nb structures already determined that target particular ES, we can model the effects of the variant mutations on antibody recognition.

Two examples illustrate this approach: the R346T RBD mutation in the subvariants BA.4, BA.5, BF.7 and XBB.1.5 lies within ES2 ***(Table 2a, Table 6a, Figure 6a)***, and those Abs that recognize ES2 may be further evaluated for their ability to bind the mutants that harbor the R->T substitution. ***Supplementary Table 5a*** lists a number of Abs and Nbs whose structures are known that interact with ES2, and analysis of several Abs which may potentially resist the escape mutation (***Supplementary Figure 5a***). Specifically, the emergency use authorized (EUA) mAb S309 (one of three Fab modeled in PDB 7JX3) (sotrovimab)) may have neutralizing potency when

277 combined with other antibodies to BA.1.1.529, BA.1, BA.2.75 subvariants [50,51]. A second is the
278 R486 mutation found in XBB.1 (R486S) and XBB1.5 (R486P) which is located in ES18 and 19 (F490
279 & R493). We identified a number of Abs and Nbs (***Supplementary Table 5b***) that have multiple
280 contacts with ES17, 18, and 19, such as for COVOX-45, which preserves those to P486 from the
281 main-chain of the CDR3 loop. Also, the nanobody Nb-2-67 makes multiple hydrogen bonds to
282 maintain contact with ES18 (***Supplementary Figure 5b***).

283 Our analysis of ES recognized by Abs and Nbs and the identification of specific ES affected
284 by mutations in VOC provides an explanation for the ineffectiveness of some Ab that have been
285 tested therapeutically. One example, Evushield™, which consists of two Abs, tixagevimab (AZD
286 8895) and cligavimab (AZD 1061) illustrates this point. These Ab have been studied by X-ray
287 crystallography (tixagevimab, PDB 7L7D, and cligavimab 7L7E [52]) and by cryo-EM [53]. By our
288 analysis, tixagevimab interacts with ES13, 16, 18, 19, and 20 and cligavimab with ES2, 10, 11, and
289 12. As shown in ***Table 6a,*** residues in every one of these ES are mutated in the Omicron variant.
290 This then explains the lack of beneficial effect of Evushield™ and supports a molecular basis for
291 the recent revision of its EUA by the FDA (https://www.fda.gov/drugs/drug-safety-and-
292 availability/fda-announces-evusheld-not-currently-authorized-emergency-use-us). This
293 highlights the importance of our analysis of ES bound by Abs and Nbs.

Discussion

The enormous world-wide effort to elucidate the mechanistic underpinnings of the immune response to SARS-CoV-2 has provided deep insight into aspects of the B cell and T cell responses to infection and immunization and has contributed to ongoing strategies for therapy and prevention. Here, we have taken advantage of the ever-increasing structural database of anti-SARS-CoV-2 Abs and Nbs to analyze the three-dimensional features that are described by X-ray and cryo-EM structures of Ab and Nb complexes with the RBD of the virus, either alone or in the context of the full spike protein. We have developed several analytical computational tools described in detail in the methods that allow the tabulation and analysis of molecular contacts and ES between the Abs/Nbs and the RBD. These provide a convenient avenue for querying and comparing the binding sites and interactions of particular Abs/Nbs and will support additional queries as the Cov-AbDab and PDB entries increase. This has permitted the categorization of the epitope-paratope interactions and molecular surface characteristics that lend themselves to recognition by Abs and the recurrent structural motifs of the CDR residues of the Abs/Nbs. This identification of 23 ES derives from evaluation of a large number of Ab/Nb-RBD and Ab/Nb-spike structures and their interface contacts, and thus surpasses analyses based on amino acid sequence or gross structural comparison alone. Our method of clustering ES sites with various stringencies, and independently of the antibodies that recognize them, offers an additional tool towards the goal of prediction of CDR sequences that recognize particular epitopic sites.

Of some 340 Abs and 83 Nbs, our analysis indicates that the 23 ES on the RBD characterized in part by secondary structural features may be recognized at different frequencies. This fine-grained analysis of the RBD surface reveals that as many as 10% of Abs may recognize common features such as those of ES16 as seen by Abs, or of ES11 as seen by Nbs.

Understanding the biophysical or structural characteristics of antigenic or immunogenic sites on protein antigens has been a subject of considerable interest for many years, beginning with efforts to understand common sites seen by heterogeneous Abs and further refined as monoclonal Abs have been studied [6,34,36,37,54]. Recent efforts have identified common motifs that human antibodies exploit to bind similar epitopes [55]. Consistent features of antigenic sites include hydrophobicity, accessibility, and segmental mobility as well as sequence dissimilarity to the Ab-producing organism (tolerance). Here we have taken the opportunity to investigate a large number of Abs and Nbs for which the antigenic site of a single protein is defined at high resolution by structural criteria. Although non-random factors may contribute to biases in the available database. several important consistent conclusions may be drawn: 1) common sites are recognized by a proportion of Abs or Nbs approaching 10%; 2) several major surfaces of the RBD have not been addressed by either Abs or Nbs; and 3) some sites are favored by either Abs (e.g, ES16 and ES18) or by Nbs (e.g., ES4 and ES5). This latter phenomenon may reflect germline VH gene preferences in the human (as suggested [56]) or the well-recognized characteristic of Nbs, whose relatively long CDR3 loops are capable of exploring concave surfaces [57].

Our analysis suggests that several regions of the RBD may be particularly important to incorporate into peptide-based immunogens (such as ES11, 13, 16, and 18) and that further generation vaccines might pay particular attention to new viral variants that affect these sites. Alternatively, Ab therapies may benefit from a focus on those reagents that recognize both common antigenic sites as well as those that are rarely identified. Although our analysis here has been confined to Abs/Nbs that recognize the RBD of the spike protein of SARS-CoV-2, this approach may, in principle, be applied to a variety Abs/Nbs directed against proteins of pathogenic organisms.

Methods

Databases

Covid antibodies and nanobodies were culled from the Coronavirus Antibody Database, Cov-AbDab (http://opig.stats.ox.ac.uk/webapps/covabdab/)[23] and coordinates of three-dimensional models were taken from the protein database (PDB)(https://www.rcsb.org/, https://rcsb.org/covid19/)[24,25].

Software

All analyses were performed with our **EPI (E**pitope-**P**aratope **I**nteraction) software package of mixed scripts of C-shell, perl and python. **EPI** software is available at https://github.com/jiangj-niaid/EPI/. Contact distances were calculated based on scripts from **CNS 1.3** (http://cns-online.org/v1.3/) [58], using a cut-off of 5.0 Å. Buried surface area (BSA) [31,59,60] was calculated with **PISA** (**P**roteins, **I**nterfaces, **S**tructures and **A**ssemblies [31]), and accessible surface area (ASA) [32,61-64] with **CNS.**

The clustering method used in EPI is based on the ES (i.e. RBD binding sites) not amino acid sequences of Abs or Nb2. The numbers of ES (1-23) are converted to a corresponding string of 23 letters from “a” to “w” and the similarity between sets of ES is computed using the Normalized Edit Distance that was developed from Hamming Distance[65] and Levenshtein Distance [66]. A similarity of 1 indicates that the two strings or two ES sets are identical; a similarity of 0 indicates that the two strings or ES sets are completely different. The similarity is then calculated for pairwise combinations of all Abs or Nbs based on their ES sets. Abs or Nbs can be clustered by imposing a similarity

threshold. For 340 Abs we tested similarity thresholds from 0.50 to 0.99 at 0.05 intervals and found that a similarity threshold of 0.85 yielded 33 clusters that covered all Ab without overlap between clusters. For 83 Nbs a similarity threshold of 0.85 yielded 10 clusters. We also provide a program with which users can make inquire for a particular ES combination, PDB ID, antibody name, or Class 1-4 designation, at a given similarity threshold.

The AIMS analysis package used for biophysical clustering of antibody sequences can be found at https://github.com/ctboughter/AIMS, including generalized Jupyter Notebooks and a Python-based GUI for the replication of the results presented herein or for the application of this analysis to novel datasets. Detailed descriptions of the foundational concepts critical for this analysis and the instructions for use can be found at https://aims-doc.readthedocs.io.

Figures for structural models are generated by using PyMOL [67] (https://pymol.org/2/). Sequence logo figures were generated with WebLogo (https://weblogo.berkeley.edu/) [38]. Sequence alignments were made with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) [68]. Graphic plots were generated with Prism 9 (https://GraphPad.com).

Data availability

All data generated for analysis in this study has been published on GitHub at https://github.com/jiangj-niaid/RBD-SARS2/.

Acknowledgements

This research was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, NIH.

Author contributions

JJ conceived the project, wrote programs, analyzed and discussed data, prepared figures. CTB contributed to program scripts, analyzed and discussed data, and prepared figures. JJ, CTB, JA, KN, LFB, MM-S and DHM discussed data, and wrote and revised the paper.

References

1 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org:10.1038/s41586-021-03819-2

2 Yin, R., Feng, B. Y., Varshney, A. & Pierce, B. G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci* **31**, e4379 (2022). https://doi.org:10.1002/pro.4379

3 Aguilar Rangel, M. *et al.* Fragment-based computational design of antibodies targeting structured epitopes. *Sci Adv* **8**, eabp9540 (2022). https://doi.org:10.1126/sciadv.abp9540

4 Fischman, S. & Ofran, Y. Computational design of antibodies. *Curr Opin Struct Biol* **51**, 156-162 (2018). https://doi.org:10.1016/j.sbi.2018.04.007

5 Chidyausiku, T. M. *et al.* De novo design of immunoglobulin-like domains. *Nat Commun* **13**, 5661 (2022). https://doi.org:10.1038/s41467-022-33004-6

6 Tiselius, A. & Kabat, E. A. An Electrophoretic Study of Immune Sera and Purified Antibody Preparations. *J Exp Med* **69**, 119-131 (1939). https://doi.org:10.1084/jem.69.1.119

7 Lu, L. L., Suscovich, T. J., Fortune, S. M. & Alter, G. Beyond binding: antibody effector functions in infectious diseases. *Nat Rev Immunol* **18**, 46-61 (2018). https://doi.org:10.1038/nri.2017.106

8 Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **93**, 13-20 (1996). https://doi.org:10.1073/pnas.93.1.13

9 Davies, D. R. & Cohen, G. H. Interactions of protein antigens with antibodies. *Proc Natl Acad Sci U S A* **93**, 7-12 (1996). https://doi.org:10.1073/pnas.93.1.7

10 Burley, S. K. *et al.* Electron microscopy holdings of the Protein Data Bank: the impact of the resolution revolution, new validation tools, and implications for the future. *Biophys Rev*, 1-21 (2022). https://doi.org:10.1007/s12551-022-01013-w

11 Mendoza, P., Lorenzi, J. C. C. & Gaebler, C. COVID-19 antibody development fueled by HIV-1 broadly neutralizing antibody research. *Curr Opin HIV AIDS* **16**, 25-35 (2021). https://doi.org:10.1097/COH.0000000000000657

12 Uversky, V. N. & Van Regenmortel, M. H. V. Mobility and disorder in antibody and antigen binding sites do not prevent immunochemical recognition. *Crit Rev Biochem Mol Biol* **56**, 149-156 (2021). https://doi.org:10.1080/10409238.2020.1869683

13 Regep, C., Georges, G., Shi, J., Popovic, B. & Deane, C. M. The H3 loop of antibodies shows unique structural characteristics. *Proteins* **85**, 1311-1318 (2017). https://doi.org:10.1002/prot.25291

14 Barnes, C. O. *et al.* SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682-687 (2020). https://doi.org:10.1038/s41586-020-2852-1

15 Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221-224 (2020). https://doi.org:10.1038/s41586-020-2179-y

16 Dejnirattisai, W. *et al.* The antigenic anatomy of SARS-CoV-2 receptor binding domain. *Cell* **184**, 2183-2200 e2122 (2021). https://doi.org:10.1016/j.cell.2021.02.032

17 Hastie, K. M. *et al.* Defining variant-resistant epitopes targeted by SARS-CoV-2 antibodies: A global consortium study. *Science* **374**, 472-478 (2021). https://doi.org:10.1126/science.abh2315

18 Greaney, A. J. *et al.* The SARS-CoV-2 Delta variant induces an antibody response largely focused on class 1 and 2 antibody epitopes. *PLoS Pathog* **18**, e1010592 (2022). https://doi.org:10.1371/journal.ppat.1010592

19 Greaney, A. J. *et al.* Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nat Commun* **12**, 4196 (2021). https://doi.org:10.1038/s41467-021-24435-8

20 Starr, T. N. *et al.* SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature* **597**, 97-102 (2021). https://doi.org:10.1038/s41586-021-03807-6

21 Lubin, J. H. *et al.* Structural models of SARS-CoV-2 Omicron variant in complex with ACE2 receptor or antibodies suggest altered binding interfaces. *bioRxiv* (2021). https://doi.org:10.1101/2021.12.12.472313

22 Ahmad, J. *et al.* Structures of synthetic nanobody-SARS-CoV-2 receptor-binding domain complexes reveal distinct sites of interaction. *J Biol Chem* **297**, 101202 (2021). https://doi.org:10.1016/j.jbc.2021.101202

23 Raybould, M. I. J., Kovaltsuk, A., Marks, C. & Deane, C. M. CoV-AbDab: the coronavirus antibody database. *Bioinformatics* **37**, 734-735 (2021). https://doi.org:10.1093/bioinformatics/btaa739

24 Sussman, J. L. *et al.* Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr D Biol Crystallogr* **54**, 1078-1084 (1998). https://doi.org:10.1107/s0907444998009378

25 Berman, H. M. *et al.* The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58**, 899-907 (2002). https://doi.org:10.1107/s0907444902003451

26 Kastritis, P. L., Rodrigues, J. P. & Bonvin, A. M. HADDOCK(2P2I): a biophysical model for predicting the binding affinity of protein-protein interaction inhibitors. *J Chem Inf Model* **54**, 826-836 (2014). https://doi.org:10.1021/ci4005332

27 Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J Mol Biol* **234**, 946-950 (1993). https://doi.org:10.1006/jmbi.1993.1648

28 Thornton, J. M., Singh, J., Campbell, S. & Blundell, T. L. Protein-protein recognition via side-chain interactions. *Biochem Soc Trans* **16**, 927-930 (1988). https://doi.org:10.1042/bst0160927

29 Laskowski, R. A. *et al.* PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends Biochem Sci* **22**, 488-490 (1997). https://doi.org:10.1016/s0968-0004(97)01140-7

30 Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res* **33**, D266-268 (2005). https://doi.org:10.1093/nar/gki001

31 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797 (2007). https://doi.org:10.1016/j.jmb.2007.05.022

32 Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* **55**, 379-400 (1971). https://doi.org:10.1016/0022-2836(71)90324-x

33 Ingram, J. R. *et al.* Anti-CTLA-4 therapy requires an Fc domain for efficacy. *Proc Natl Acad Sci U S A* **115**, 3912-3917 (2018). https://doi.org:10.1073/pnas.1801524115

34 Kabat, E. A., Wu, T. T. & Bilofsky, H. Attempts to locate residues in complementarity-determining regions of antibody combining sites that make contact with antigen. *Proc Natl Acad Sci U S A* **73**, 617-619 (1976). https://doi.org:10.1073/pnas.73.2.617

35 Chothia, C. *et al.* Conformations of immunoglobulin hypervariable regions. *Nature* **342**, 877-883 (1989). https://doi.org:10.1038/342877a0

36 Alzari, P. M., Lascombe, M. B. & Poljak, R. J. Three-dimensional structure of antibodies. *Annu Rev Immunol* **6**, 555-580 (1988). https://doi.org:10.1146/annurev.iy.06.040188.003011

37 Padlan, E. A. Structural basis for the specificity of antibody-antigen reactions and structural mechanisms for the diversification of antigen-binding specificities. *Q Rev Biophys* **10**, 35-65 (1977). https://doi.org:10.1017/s0033583500000135

38 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Research* **14**, 1188-1190 (2004).

39 Robinson, S. A. *et al.* Epitope profiling using computational structural modelling demonstrated on coronavirus-binding antibodies. *PLoS Comput Biol* **17**, e1009675 (2021). https://doi.org:10.1371/journal.pcbi.1009675

40 Wang, Y. *et al.* A large-scale systematic survey reveals recurring molecular features of public antibody responses to SARS-CoV-2. *Immunity* **55**, 1105-1117 e1104 (2022). https://doi.org:10.1016/j.immuni.2022.03.019

41 Liu, L. *et al.* An antibody class with a common CDRH3 motif broadly neutralizes sarbecoviruses. *Sci Transl Med* **14**, eabn6859 (2022). https://doi.org:10.1126/scitranslmed.abn6859

42 Tan, T. J. C. *et al.* Sequence signatures of two public antibody clonotypes that bind SARS-CoV-2 receptor binding domain. *Nat Commun* **12**, 3815 (2021). https://doi.org:10.1038/s41467-021-24123-7

43 Boughter, C. T. *et al.* Biochemical patterns of antibody polyreactivity revealed through a bioinformatics-based analysis of CDR loops. *Elife* **9** (2020). https://doi.org:10.7554/eLife.61393

44 Wang, Q. *et al.* Antibody evasion by SARS-CoV-2 Omicron subvariants BA.2.12.1, BA.4 and BA.5. *Nature* **608**, 603-608 (2022). https://doi.org:10.1038/s41586-022-05053-w

45 Tuekprakhon, A. *et al.* Antibody escape of SARS-CoV-2 Omicron BA.4 and BA.5 from vaccine and BA.1 serum. *Cell* **185**, 2422-2433 e2413 (2022). https://doi.org:10.1016/j.cell.2022.06.005

46 Nutalai, R. *et al.* Potent cross-reactive antibodies following Omicron breakthrough in vaccinees. *Cell* **185**, 2116-2131 e2118 (2022). https://doi.org:10.1016/j.cell.2022.05.014

47 Cao, Y. *et al.* Rational identification of potent and broad sarbecovirus-neutralizing antibody cocktails from SARS convalescents. *Cell Rep* **41**, 111845 (2022). https://doi.org:10.1016/j.celrep.2022.111845

48 Cao, Y. *et al.* Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. *Nature* (2022). https://doi.org:10.1038/s41586-022-05644-7

49 Mannar, D. *et al.* SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein-ACE2 complex. *Science* **375**, 760-764 (2022). https://doi.org:10.1126/science.abn7760

50 McCallum, M. *et al.* Structural basis of SARS-CoV-2 Omicron immune evasion and receptor engagement. *Science* **375**, 864-868 (2022). https://doi.org:10.1126/science.abn8652

51 Wu, Y. *et al.* Lineage-mosaic and mutation-patched spike proteins for broad-spectrum COVID-19 vaccine. *Cell Host Microbe* **30**, 1732-1744 e1737 (2022). https://doi.org:10.1016/j.chom.2022.10.011

52 Dong, J. *et al.* Genetic and structural basis for SARS-CoV-2 variant neutralization by a two-antibody cocktail. *Nat Microbiol* **6**, 1233-1244 (2021). https://doi.org:10.1038/s41564-021-00972-2

53 Parzych, E. M. *et al.* DNA-delivered antibody cocktail exhibits improved pharmacokinetics and confers prophylactic protection against SARS-CoV-2. *Nat Commun* **13**, 5886 (2022). https://doi.org:10.1038/s41467-022-33309-6

54 Benjamin, D. C. *et al.* The antigenic structure of proteins: a reappraisal. *Annu Rev Immunol* **2**, 67-101 (1984). https://doi.org:10.1146/annurev.iy.02.040184.000435

55 Shrock, E. L. *et al.* Germline-encoded amino acid-binding motifs drive immunodominant public antibody responses. *Science* **380**, eadc9498 (2023). https://doi.org:10.1126/science.adc9498

56 Barnes, C. O. *et al.* Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828-842 e816 (2020). https://doi.org:10.1016/j.cell.2020.06.025

57 Arbabi-Ghahroudi, M. Camelid Single-Domain Antibodies: Promises and Challenges as Lifesaving Treatments. *Int J Mol Sci* **23** (2022). https://doi.org:10.3390/ijms23095009

58 Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-921 (1998). https://doi.org:10.1107/s0907444998003254

59 Chothia, C. & Janin, J. Principles of protein-protein recognition. *Nature* **256**, 705-708 (1975). https://doi.org:10.1038/256705a0

60 Chen, J., Sawyer, N. & Regan, L. Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci* **22**, 510-515 (2013). https://doi.org:10.1002/pro.2230

61 'NACCESS', computer program. (Department of Biochemistry and Molecular Biology, University College, London., 1993).
62 Jones, S. & Thornton, J. M. Analysis of protein-protein interaction sites using surface patches. *J Mol Biol* **272**, 121-132 (1997). https://doi.org:10.1006/jmbi.1997.1234
63 Fraczkiewicz, R. & Braun, W. Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *J. Comp. Chem.* **19**, 319-333 (1998).
64 Ribeiro, J., Rios-Vera, C., Melo, F. & Schuller, A. Calculation of accurate interatomic contact surface areas for the quantitative analysis of non-bonded molecular interactions. *Bioinformatics* **35**, 3499-3501 (2019). https://doi.org:10.1093/bioinformatics/btz062
65 Hamming, R. W. Error detecting and error correcting codes. *The Bell System Technical Journall* **29**, 147-160 (1950).
66 Levenshtein, V. I. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* **10**, 707-710 (1966).
67 The PyMOL Molecular Graphics System, Version 2.5.4 Schrödinger, LLC.
68 Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* **50**, W276-279 (2022). https://doi.org:10.1093/nar/gkac240
69 Xu, J. *et al.* Nanobodies from camelid mice and llamas neutralize SARS-CoV-2 variants. *Nature* **595**, 278-282 (2021). https://doi.org:10.1038/s41586-021-03676-z
70 Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983). https://doi.org:10.1002/bip.360221211

FIGURE LEGENDS

Fig. 1. Number of contacts to RBD by Abs and Nbs. a Total number of contacts to each of the indicated RBD residues summed from all available X-ray and cryo-EM structures from Ab H chains. **b** Graphic depiction of number of contacts illustrated as footprint on the RBD and as putty heat map of RBD cartoon backbone. Top, inner face, and side views of RBD are shown. **c** Total number of contacts as in **a**, but for Nb contacts. **d** Surface footprint and putty heat map of Nb contacts as in **b**.

Fig. 2. Distribution of Abs and Nbs on RBD surface. a Putty heat map of H chain of antibody with the definition of ES. The thickness of putty represents the number of hits. **b** Putty heat map of Nb with the definition of ES. **c** Distribution of Abs/Nbs on ES of RBD surface (percentage, %). Magenta represents Ab, blue represents Nb. **d** Comparison of antibody H chains and L chains on ES of RBD surface (by hit numbers). **e** ES surface area or footprint is illustrated by a color map of the RBD surface.

Fig. 3. Distribution of CDR loops of contacts to RBD surface over ES. **a** Antibody H chains are plotted (percentage). Pie graph indicates the composition of CDR1 (16%, orange), CDR2 (21%, marine blue), CDR3 (45%, purple) and non-CDR (18%, gray) respectively. **b** Antibody L chains are plotted (percentage). Pie graph indicates the composition of CDR1 (40%, orange), CDR2 (10%, marine blue), CDR3 (25%, purple) and non-CDR (25%, gray) respectively. **c** Nanobody chains are plotted (percentage). Pie graph indicates the composition of CDR1 (13%, orange), CDR2 (14%, marine blue), CDR3 (46%, purple) and non-CDR (27%, gray) respectively. **d** Average length (in amino acids (aa)) of CDR loops extracted from the sequences (CovAbDab, [23], as of 12/20/2022) and used in this study. The averages are over 340 antibodies and 83 nanobodies respectively.

Fig. 4. Distributions of amino acids of Abs/Nbs over ES. **a** Heat map of amino acids of Ab H chains on each ES, magenta indicates the frequency of the amino acids. **b** Heat map of amino acids of nanobody on each ES, blue indicates the frequency of the amino acids. Top triplets of amino acids are those most frequently observed amino of Abs/Nbs on each ES. **c** The usage of amino acids of antibody H chains (magenta) and nanobody (blue) in

interacting with RBD is plotted in descending order (percentage). YSR are most frequently observed amino acids both for Ab and Nb. **d** Usage of amino acids in CDR3 loops (purple for Abs; light blue for Nbs). W of Nbs has relatively higher percentage in comparison to Abs both overall and for CDR3 loop.

Fig. 5. Identity of the similarity of the ES and clustering of Abs/Nbs (see Supplementary Table 4). a Illustration of three antibody clusters: A1, A3 and A11, each identifies a specific ES combination. Superimposed are members of the cluster on the RBD (only HV domains are shown for clarity). **b** Illustration of three nanobody clusters: N1, N3 and N4. RBD is presented as gray surface, magenta indicates the binding areas (footprints) of ES of RBD. **c** A subset of the Cluster-A1, named A1S1, ES=(8,9,13,16,18,19) with similarity≥0.90, shows a strong binding motif on CDR loops. The members (28) of A1S1 are superposed on the RBD on the left panel. On the right top panel are shown the contacts between CDR loops and the binding sites (ES8-9, ES13, ES16, and ES18-19). On the right below panel, WebLogo plots show the amino acids from Abs binding to ES8-9, 13, and ES18-19 respectively. Y,S,G from CDR2 are favor binding to ES13 (RBD residues from 455-460); S,S,N,T of CDR1 favor binding to ES16 (RBD residues 455-459); and R and Y of CDR3 are favor binding to ES18-19 (RBD residues 485-491). **d** Clustering using AIMS [43]. Here AX1 and AX2 are "principle components" of biophysical properties, or "mature information".

Fig. 6. Ilustration of locations of variant mutations and associated ES on RBD surface. a Alpha variants. **b** Beta variants. **c** Delta variants. **d** Omicron variants and subvariants.

TABLE LEGENDS

Table 1. Summary of sequences and structures of anti-SARS-CoV-2 antibodies and nanobodies.

The sequences and origin/source are collected in CovAbDab [23], as of 12/20/2022. The number of structures of antibodies and nanobodies in complex with RBD or spike protein are downloaded from PDB.

*Unique non-redundant structures determined either by X-ray or cryo-EM as listed in the PDB.

#This includes two sequences/structures from mice engineered to express llama Nb genes [69].

Table 2. Definitions of Epitopic Sites (ES) seen by Abs and Nbs. a RBD residue range for each ES is indicated, along with the amino acid sequence, secondary structural features (as determined by DSSP [70]), accessible surface area (ASA) (see Methods) of the contacting residues, and percentage of Ab H chains and Nbs. **b** correlation of ES with Class definitions by Barnes [14] and with receptor binding motif (RBM) [15].

Table 3. Relation of ES to SARS2-CoV-2 escape mutations. **a** Major mutations with the main lineage of variants of SERS-Cov-2 and corresponding ES site. "X" column indicates the amino acids substitution of the sub-variants of Omicron. **b** Latest mutations in the major lineage of sub-variants of Omicron and corresponding ES site.

Figures

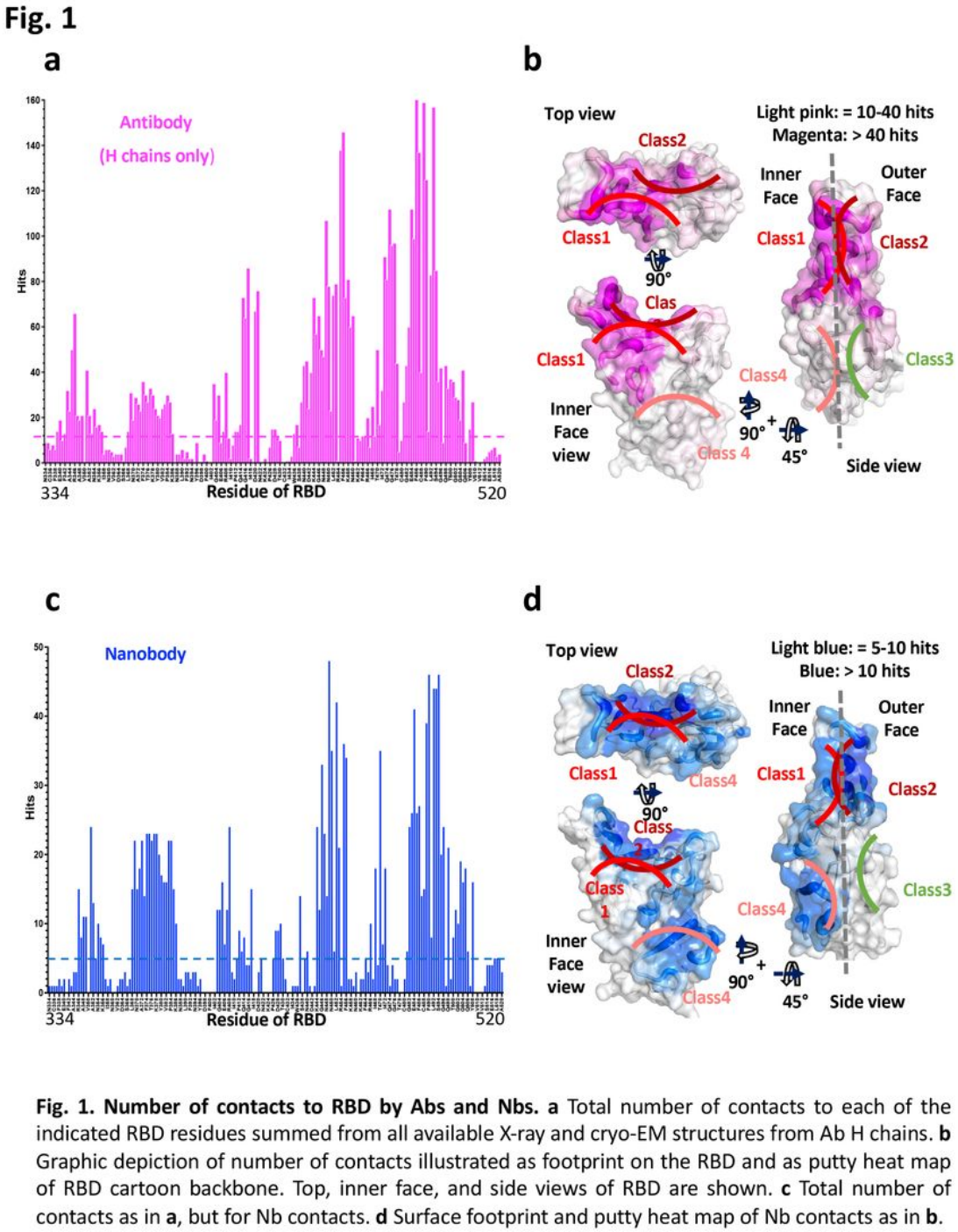

Fig. 1. Number of contacts to RBD by Abs and Nbs. a Total number of contacts to each of the indicated RBD residues summed from all available X-ray and cryo-EM structures from Ab H chains. **b** Graphic depiction of number of contacts illustrated as footprint on the RBD and as putty heat map of RBD cartoon backbone. Top, inner face, and side views of RBD are shown. **c** Total number of contacts as in **a**, but for Nb contacts. **d** Surface footprint and putty heat map of Nb contacts as in **b**.

Figure 1

See image above for figure legend.

Fig. 2

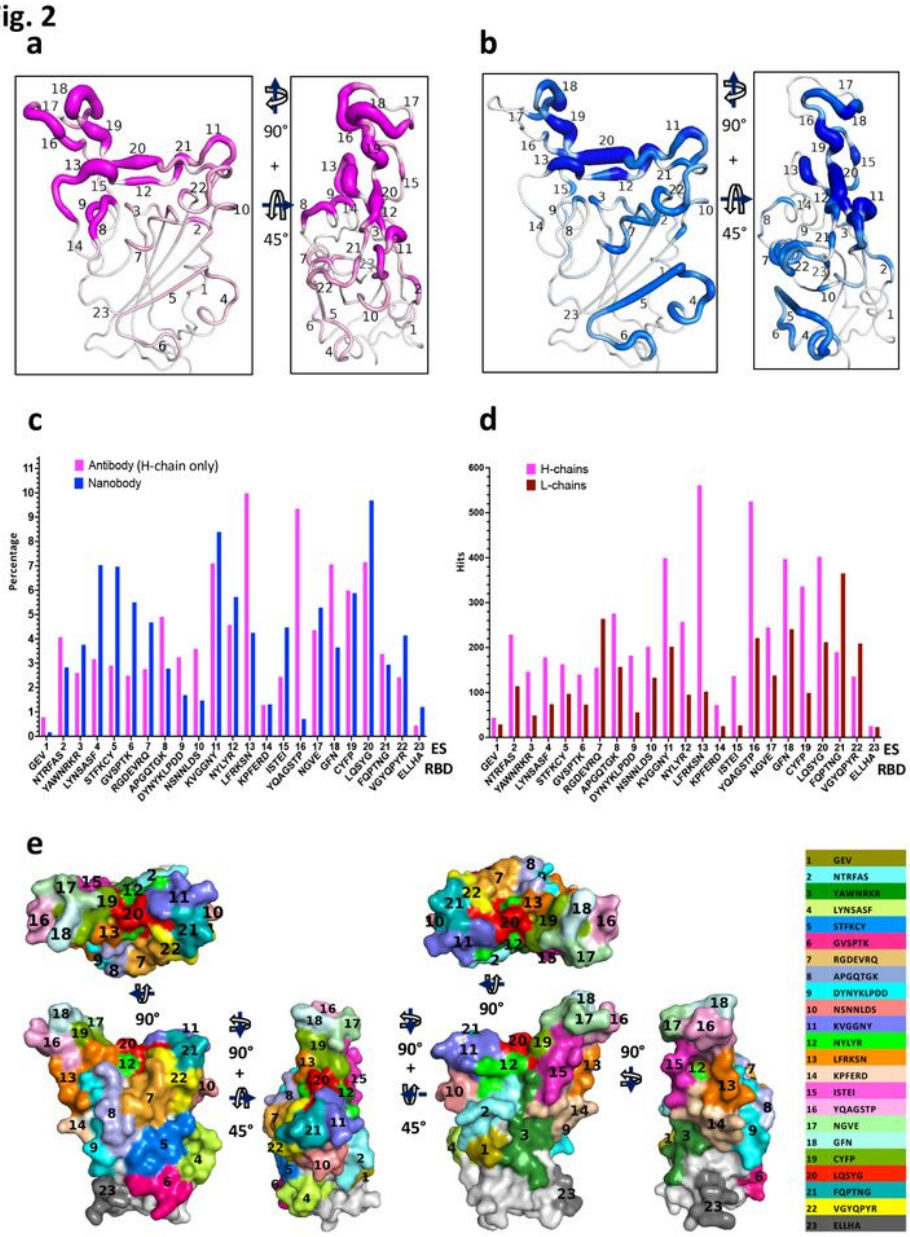

Fig. 2. Distribution of Abs and Nbs on RBD surface. a Putty heat map of H chain of antibody with the definition of ES. The thickness of putty represents the number of hits. **b** Putty heat map of Nb with the definition of ES. **c** Distribution of Abs/Nbs on ES of RBD surface (percentage, %). Magenta represents Ab, blue represents Nb. **d** Comparison of antibody H chains and L chains on ES of RBD surface (by hit numbers). **e** ES surface area or footprint is illustrated by a color map of the RBD surface.

Figure 2

See image above for figure legend.

Fig. 3

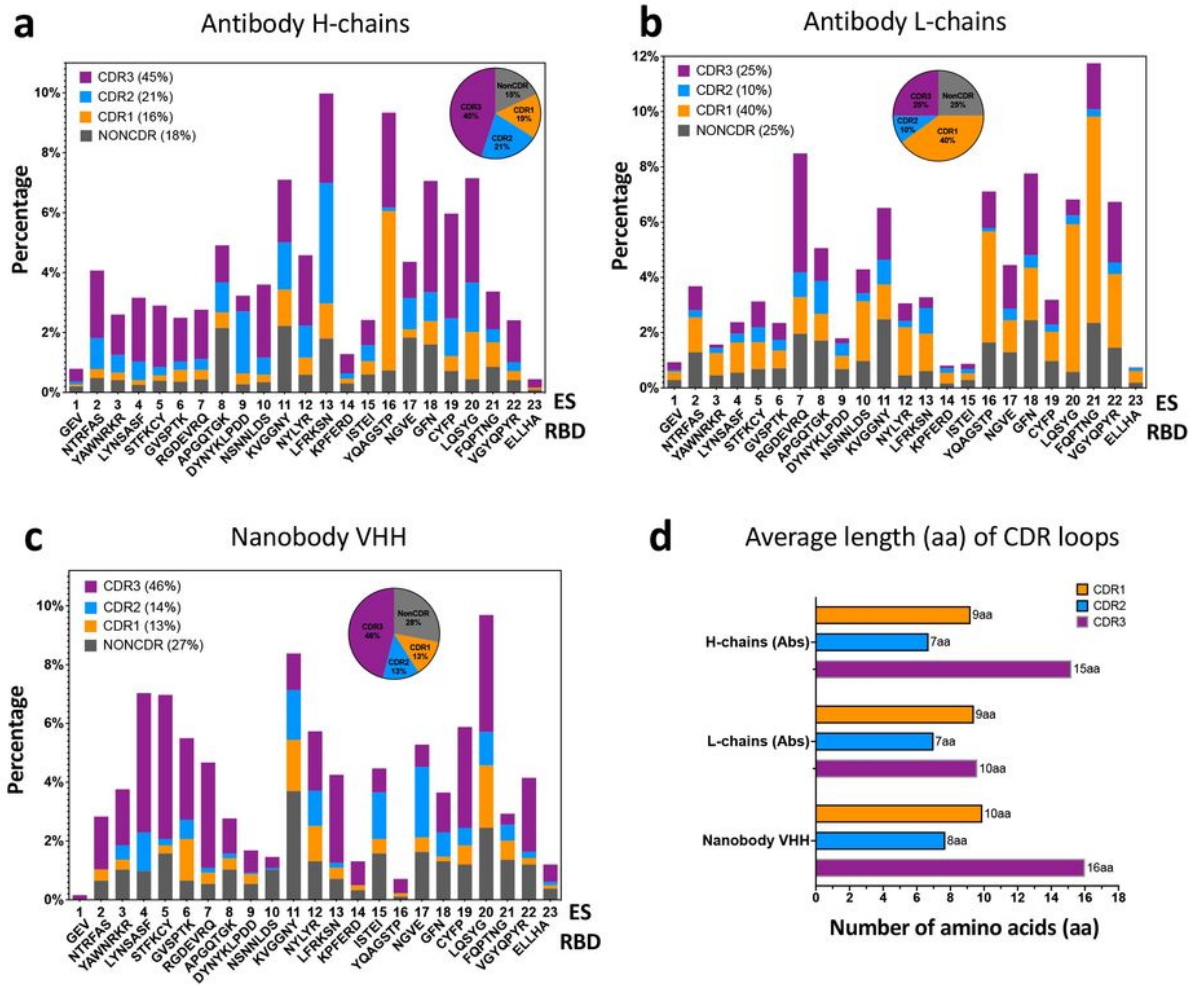

Fig. 3. Distribution of CDR loops of contacts to RBD surface over ES. **a** Antibody H chains are plotted (percentage). Pie graph indicates the composition of CDR1 (16%, orange), CDR2 (21%, marine blue), CDR3 (45%, purple) and non-CDR (18%, gray) respectively. **b** Antibody L chains are plotted (percentage). Pie graph indicates the composition of CDR1 (40%, orange), CDR2 (10%, marine blue), CDR3 (25%, purple) and non-CDR (25%, gray) respectively. **c** Nanobody chains are plotted (percentage). Pie graph indicates the composition of CDR1 (13%, orange), CDR2 (14%, marine blue), CDR3 (46%, purple) and non-CDR (27%, gray) respectively. **d** Average length (in amino acids (aa)) of CDR loops extracted from the sequences (CovAbDab, [23], as of 12/20/2022) and used in this study. The averages are over 340 antibodies and 83 nanobodies respectively.

Figure 3

See image above for figure legend.

Fig. 4

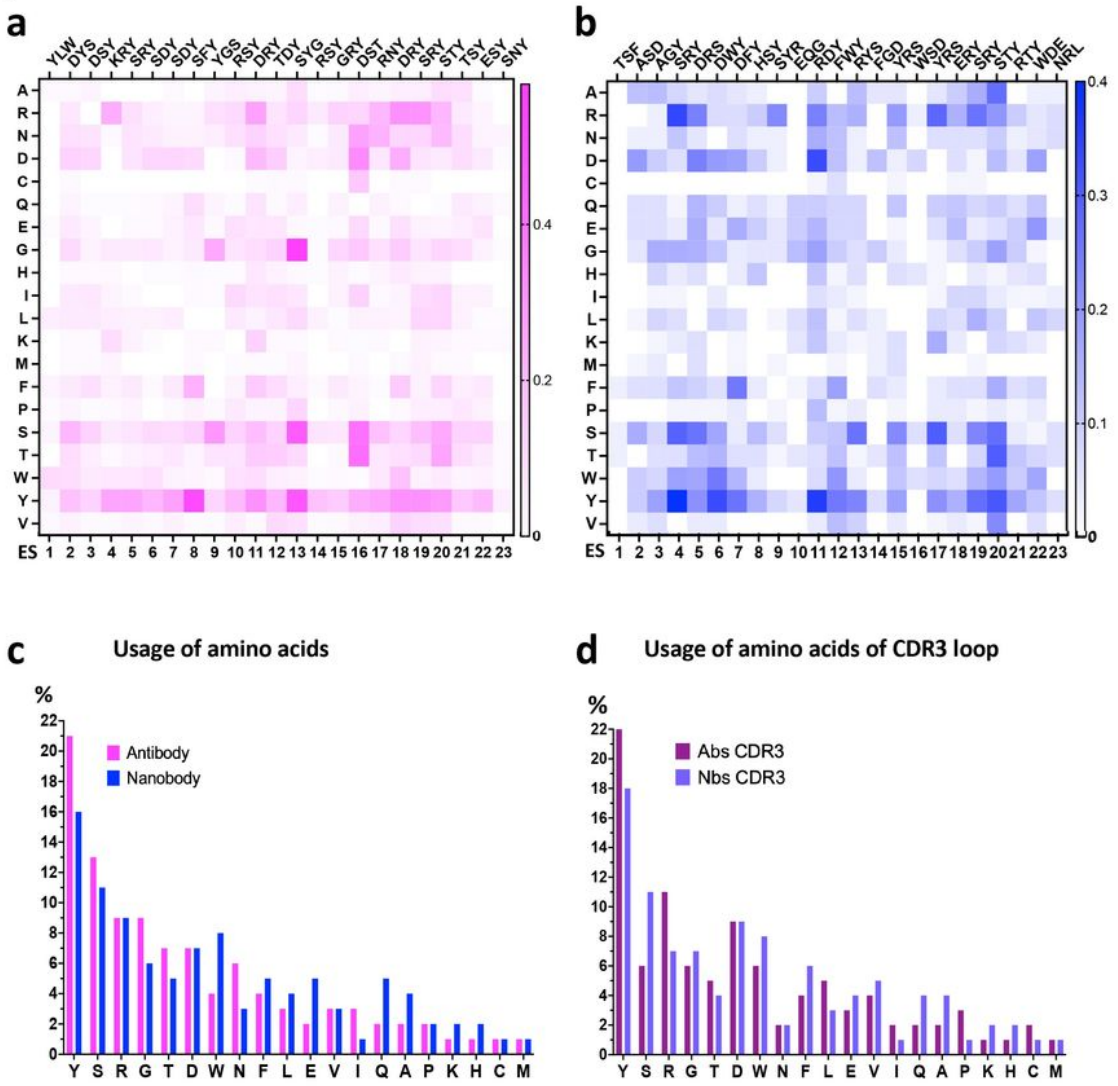

Fig. 4. Distributions of amino acids of Abs/Nbs over ES. **a** Heat map of amino acids of Ab H chains on each ES, magenta indicates the frequency of the amino acids. **b** Heat map of amino acids of nanobody on each ES, blue indicates the frequency of the amino acids. Top triplets of amino acids are those most frequently observed amino of Abs/Nbs on each ES. **c** The usage of amino acids of antibody H chains (magenta) and nanobody (blue) in interacting with RBD is plotted in descending order (percentage). YSR are most frequently observed amino acids both for Ab and Nb. **d** Usage of amino acids in CDR3 loops (purple for Abs; light blue for Nbs). W of Nbs has relatively higher percentage in comparison to Abs both overall and for CDR3 loop.

Figure 4

See image above for figure legend.

Fig. 5

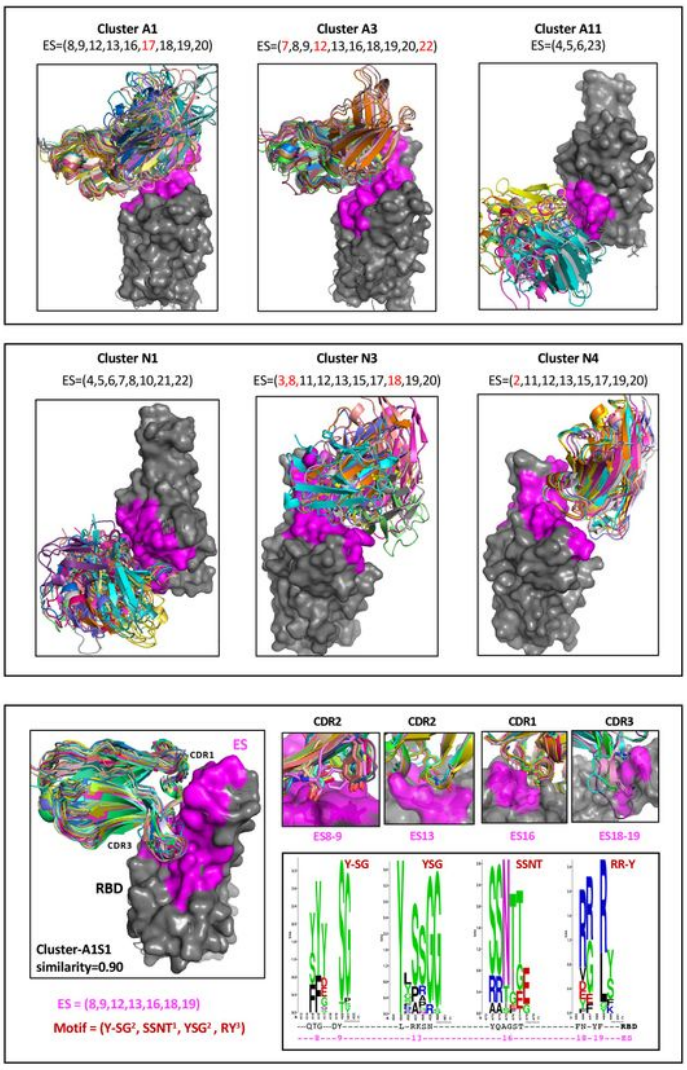

Fig. 5. Identity of the similarity of the ES and clustering of Abs/Nbs (see Supplementary Table 4). a Illustration of three antibody clusters: A1, A3 and A11, each identifies a specific ES combination. Superimposed are members of the cluster on the RBD (only HV domains are shown for clarity). **b** Illustration of three nanobody clusters: N1, N3 and N4. RBD is presented as gray surface, magenta indicates the binding areas (footprints) of ES of RBD. **c** A subset of the Cluster-A1, named A1S1, ES=(8,9,13,16,18,19) with similarity≥0.90, shows a strong binding motif on CDR loops. The members (28) of A1S1 are superposed on the RBD on the left panel. On the right top panel are shown the contacts between CDR loops and the binding sites (ES8-9, ES13, ES16, and ES18-19). On the right below panel, WebLogo plots show the amino acids from Abs binding to ES8-9, 13, and ES18-19 respectively. Y,S,G from CDR2 are favor binding to ES13 (RBD residues from 455-460); S,S,N,T of CDR1 favor binding to ES16 (RBD residues 455-459); and R and Y of CDR3 are favor binding to ES18-19 (RBD residues 485-491). **d** Clustering using AIMS [43]. Here AX1 and AX2 are "principle components" of biophysical properties, or "mature information".

Figure 5

See image above for figure legend.

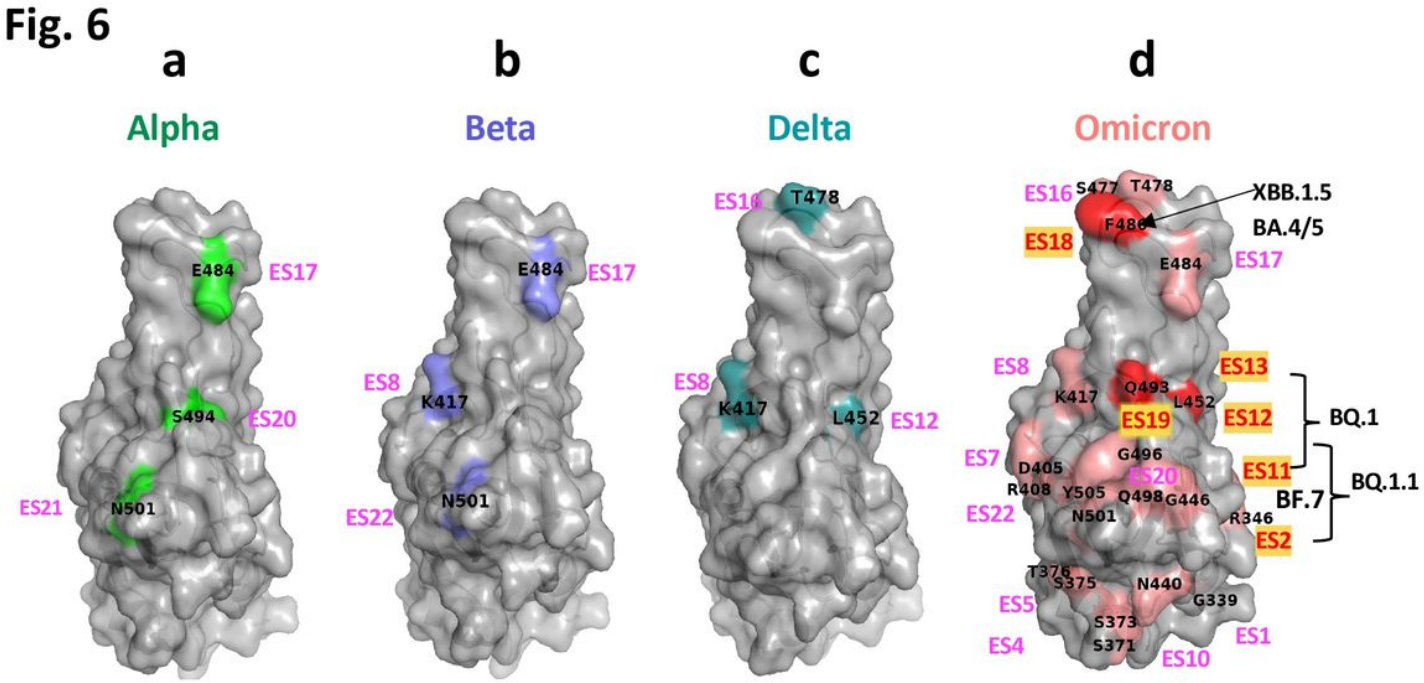

Fig. 6. Ilustration of locations of variant mutations and associated ES on RBD surface. a Alpha variants. **b** Beta variants. **c** Delta variants. **d** Omicron variants and subvariants.

Figure 6

See image above for figure legend.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- jiangSupplementaryFiguresTablesCBv7.1.pdf
- jiangmovies1aES2ndStru.mov
- jiangmovies1bESsurface.mov
- jiangsupplTable2a.csv
- Tables.pdf
- jiangsupplTable2b.csv