High Specificity of Germline Antibodies Targeting Hapten: A Universally Intrinsic Feature

Yingjie Zhang, Weilin Wu, Qing Shen, Penghui Zhou, Jing Zhang, Yantong Pan, Xuezhi Yu, Kai Wen, Jianzhong Shen*, Zhanhui Wang*

State Key Laboratory of Veterinary Public Health and Safety, Beijing Key Laboratory of Detection Technology for Animal-Derived Food, College of Veterinary Medicine, China Agricultural University, 100193 Beijing, People's Republic of China

Corresponding author: Jianzhong Shen and Zhanhui Wang

Tel: +86-10-6273 4565

Fax: +86-10-6273 1032

E-mail: sjz@cau.edu.cn; wangzhanhui@cau.edu.cn

Abstract: Polyspecificity has been widely regarded as a fundamental feature of germline antibodies that enables broad antigen recognition with the limited size of the germline antibody repertoire. However, accumulating evidence indicates that germline antibodies against carbohydrate antigens can exhibit high specificity, thereby challenging this traditional paradigm. To investigate whether the polyspecificity paradigm applies to haptens (single-epitope antigens), we inferred and identified 14 germline antibodies targeting distinct haptens; unexpectedly, all exhibited intrinsically high specificity. To explore the underlying mechanisms, we constructed a comprehensive sequence-structure database of germline antibodies recognizing haptens, peptides, and proteins. Our results reveal that hapten-binding germline antibodies are characterized by shorter CDRH3 regions, enrichment of aromatic residues in CDRs, and fewer CDR-localized surface patches—features that may underlie their intrinsically high specificity. Conversely, germline antibodies recognizing peptides and proteins display longer CDRH3 regions, increased enrichment of polar and charged residues, and denser CDR-localized surface patches, providing structural determinants of their greater polyspecificity. In addition, germline gene usage biases—particularly V-gene selection and IGHV–IGKV/IGLV pairing patterns—collectively shape the antigen-binding properties of germline antibodies. Together, these findings represent the first large-scale, systematic analysis of germline antibody polyspecificity and have broad implications for fundamental immunology, vaccine development, and antibody discovery.

Keywords: Germline antibody, specificity, polyspecificity, surface patch, germline

gene, antigens.

1. Introduction

Antibodies are a critical component of the vertebrate adaptive immune response and are widely used in both basic research and clinical applications. Their primary function is to bind antigens via their antigen-binding fragments (Fab), while their crystallizable fragment (Fc) regions activate effector cells through Fc receptors, triggering antibody-dependent cellular cytotoxicity (ADCC) to eliminate pathogen-infected or malignant cells (*1*). The immune system generates millions of diverse germline B cells (naïve B cells) that have completed V(D)J recombination but have not yet undergone somatic hypermutation (SHM), thereby establishing a baseline antibody repertoire that evolves over time. Germline antibodies correspond to the B cell receptors (BCRs) expressed by these germline B cells and represent the first antibodies produced upon antigen exposure or infection, playing a crucial role in the primary immune response (*2*).

Given the importance of germline antibodies, an in-depth understanding of their binding properties is valuable for several reasons. First, studies of germline antibodies provide fundamental insights into the capacity of the initial antibody repertoire to recognize diverse antigens (*3*). Second, knowledge of germline binding characteristics is essential for rational immunogen design in vaccine development and for guiding the discovery of therapeutic and diagnostic antibodies (*4*, *5*). Finally, because germline antibodies serve as the precursors of affinity-matured antibodies, characterizing their binding properties helps elucidate how SHM shapes affinity and specificity, thereby

informing the engineering of monoclonal antibodies (mAbs) with improved performance (*6*).

Despite several prior studies evaluating the binding properties, structures, and dynamics of germline antibodies (*4*, *5*, *7–18*), their antigen-recognition specificity has long remained a subject of debate. Early work proposed that germline antibodies are inherently polyspecific—capable of recognizing a broad spectrum of distinct epitopes across diverse antigen classes, including viruses (*9*, *11*, *19*, *20*), proteins/peptides (*12*, *21*), bacterial polysaccharides (*8*, *22*), and haptens (*13*, *17*, *23–25*). According to the polyspecificity hypothesis, naïve B cells initially express germline antibodies with highly flexible antigen-binding sites that can accommodate numerous antigens, albeit typically with low affinity. This intrinsic polyspecificity has been considered a fundamental feature of the immune system, allowing a relatively limited germline antibody repertoire to recognize an enormously larger universe of potential antigens. However, more recent evidence challenges the polyspecificity paradigm, particularly in the context of carbohydrate antigens. For example, germline antibodies targeting the ganglioside GD2 have been shown to exhibit high specificity, as demonstrated by sugar microarray, protein microarray, and cell-binding assays (*6*). Similarly, germline antibodies from activation-induced cytidine deaminase (AID) knockout mice, as well as inferred germline counterparts of mAbs, displayed strong specificity toward the oligosaccharide Lewis Y (*5*). Thus, whether the polyspecificity paradigm extends to all antigen classes remains to be determined.

Several mechanisms have been proposed to account for the polyspecific binding

properties of germline antibodies. The first line of evidence stems from structural analyses. The solvent-accessible surface area (SASA) of the antibody antigen-binding site is typically ~2,800 Å^2, whereas the SASA of individual antigenic epitopes is markedly smaller, approximately 100–1,100 Å^2. This substantial size mismatch highlights the inherent potential of germline antibodies to engage a broad range of antigens (*26*, *27*). One illustrative example is the germline antibody 36-65, which recognizes three distinct 12-mer peptide antigens by engaging three separate paratopes within a single conformational state (*12*). This size asymmetry provides a fundamental physicochemical basis for polyspecific recognition. A second mechanism proposed to underlie germline antibody polyspecificity is the substantial conformational flexibility of their antigen-binding sites. Such flexibility enables antibody paratopes to adopt distinct binding-site topologies that exist in dynamic equilibrium and can be selectively stabilized through induced-fit interactions with structurally divergent antigens (*25*, *28*, *29*). Mechanistically, numerous structural investigations demonstrate that the complementarity-determining region (CDR) loops of germline antibodies exhibit markedly greater conformational heterogeneity compared with their affinity-matured descendants (*14–17*, *30*, *31*). During affinity maturation, SHM progressively rigidifies the binding site, increasing antigen-binding affinity while concomitantly reducing conformational plasticity and polyspecific potential (*32*, *33*). Accordingly, intrinsic CDR flexibility is considered a major contributor to polyspecific recognition. A third mechanism thought to contribute to germline antibody polyspecificity involves physicochemical interaction tendencies,

particularly surface-patch properties and charge imbalance. The surface characteristics of antibodies—most notably within the CDRs—govern intermolecular interactions and colloidal behavior and can promote polyspecific or even nonspecific binding (*34*, *35*). Low polyspecificity is therefore an important criterion for therapeutic antibody developability (*36*, *37*). In particular, the extent and distribution of CDR hydrophobicity, the presence of positively or negatively charged surface patches, and asymmetry in heavy- versus light-chain electrostatic charge have all been associated with poor developability (*37*). These surface-driven propensities thus provide another molecular explanation for the broad, and occasionally nonspecific, antigen recognition of germline antibodies.

Although numerous studies have described the polyspecificity of germline antibodies and proposed several underlying mechanisms, the available evidence remains limited in scale, and the polyspecificity paradigm has faced increasing scrutiny. A major source of inconsistency lies in the difficulty of accurately defining epitopes, which complicates the evaluation of antigen specificity versus polyspecificity. Hapten molecules, which generally have molecular weights below 1000 Da and are widely regarded as single-epitope antigens, represent ideal model systems for such assessments. In this study, we inferred and identified 14 germline antibodies targeting various haptens. Surprisingly, all identified germline antibodies exhibited intrinsically high specificity toward their respective hapten molecules. To explore the molecular basis of this specificity, we constructed a large-scale sequence–structure database comprising 1,614 germline antibodies recognizing

haptens, peptides, and proteins. Systematic analyses revealed that germline antibodies against the three antigen classes exhibit distinct sequence features, structural flexibilities, surface properties, and germline-usage biases, suggesting that these factors collectively shape specificity and polyspecificity in the germline repertoire. Overall, this study provides new insight into the polyspecificity paradigm of germline antibodies, with broad implications for fundamental immunology, vaccine design, and antibody discovery.

2. Results

3.1 Germline antibodies targeting haptens display high inherent specificity

The specificity and polyspecificity of germline antibodies have been extensively investigated; however, current findings remain inconclusive. A major reason for this inconsistency lies in the difficulty of accurately identifying epitopes, thereby complicating the evaluation of specificity versus polyspecificity. Hapten molecules, typically those compounds with molecular weights below 1000 Da and generally considered to contain a single epitope, serve as ideal model systems for such studies. Therefore, we selected antibodies targeting such haptens to systematically investigate the specificity and polyspecificity of germline antibodies.

In this work, we selected 23 mAbs previously produced by our group (*38–41*), targeting nine different haptens derived from six small molecules (Table S1). The germline antibody sequences of these mAbs were inferred and assembled using the IMGT/V-QUEST tool, totally yielding 19 unique germline antibodies, each

designated with the suffix "-GL" as shown in Table S2. Then, these 19 germline antibodies were expressed in HEK293 mammalian cells (Figure S1). To explore the specificity of germline antibodies and their corresponding mAbs, we characterized their binding affinities against a panel of structurally similar hapten-ovalbumin (OVA) conjugates, respectively. Lower affinity values correspond to higher binding affinity as shown in Figure 1A. The binding affinities of 10 mAbs against chloramphenicol (CAP) and 9 their corresponding germline antibodies were identified by 25 CAP structurally related hapten-OVA conjugates including haptens of thiamphenicol (TAP), and florfenicol (FF) (detailed hapten structure used for OVA conjugates shown in Figure S2). The results shown that these 10 mAbs against CAP exhibited different recognition spectra, binding to 4–14 conjugates, whereas, their corresponding germline antibodies exhibited significant narrow recognition spectra, with six recognize only 2–3 conjugates and other three failed to bind to all conjugates including CAP-OVA conjugates. Overall, germline antibodies against CAP exhibit lower binding affinity and a narrower recognition spectrum compared to their mature counterparts. Notably, germline antibodies CAP-5D11-GL and CAP-15A9-GL displayed high binding affinities of 5.4 ng/mL and 21 ng/mL to CAP, respectively, which are comparable to those affinities of their corresponding mAbs. In contrast, CAP-5D10-GL, CAP-2E4-GL1, CAP-2E4-GL2, and CAP-16B5-GL exhibited binding affinities ranging from 420 ng/mL to 10,000 ng/mL, approximately 2–4 orders of magnitude lower than those of their mAbs. The remaining three germline antibodies showed no detectable antigen recognition, likely

due to affinities below the detection threshold.

As shown in Figure 1B, mAbs against polymyxin B (PMB) derived from different haptens displayed distinct antigen-recognition behaviors (structurally related hapten shown in Figure S3). The three mAbs derived from H1 hapten including PMB-5B10, PMB-11D5, and PMB-13G6 recognized both PMB- and H1-OVA conjugates, whereas the two mAbs from H2 hapten (PMB-1B3 and PMB-10D5) recognized PMB- and H2-OVA conjugates. The same trend was observed for the germline antibody PMB-5B10-GL, which recognized both PMB-OVA and H1-OVA, with binding affinities of 31 ng/mL and 320 ng/mL, respectively—approximately 1–2 orders of magnitude lower than those of the corresponding mAbs with 3.4 ng/mL and 2.9 ng/mL. However, PMB-11D5-GL1 recognized only the H1-OVA conjugates, with binding affinity of 450 ng/mL, about two orders of magnitude lower than the mAb PMB-11D5 with a binding affinity of 1.7 ng/mL. Similarly, PMB-1B3-GL recognized only the H2-OVA conjugates, with binding affinity of 230 ng/mL, also about two orders of magnitude lower than its corresponding mAb of 4.3 ng/mL. In contrast, PMB-11D5-GL2 showed no detectable binding to any conjugates. Overall, the germline antibodies of PMB display reduced binding affinity and a narrower recognition spectrum compared to their mature counterparts.

Using 15 different sulfadiazine (SA)–OVA conjugates, the binding profiles of three SA mAbs and their germline antibodies were further evaluated. As shown in Figure 1C, the mAb SA-4D11 recognized all 15 structurally related hapten-OVA conjugates (structurally related hapten shown in Figure S4), SA-10E6 recognized 12, and

SA-4C7 recognized only 7. Among their germline antibodies, SA-10E6-GL bound to 8 hapten-OVA conjugates with binding affinity comparable to its mature mAb. In contrast, SA-4D11-GL recognized only its immunizing hapten-OVA conjugate, SA10-OVA, and its binding affinity (330 ng/mL) was reduced by approximately two orders of magnitude compared to that of the mature mAb of 3.6 ng/mL. SA-4C7-GL failed to recognize any of the SA conjugates, suggesting that its affinity remained below the detection limit.

All three mAbs targeting analginum (ANG) were derived from the same germline antibody, named ANG-3B3-GL (Table S2). As shown in Figure 1D, both the mature mAbs and ANG-3B3-GL recognized all five tested hapten-OVA conjugates (structurally related hapten shown in Figure S5). Importantly, the binding affinities of ANG-3B3-GL, ranging from 58 to 110 ng/mL, were comparable to those of the mature mAbs, which ranged from 8.9 to 55 ng/mL.

As shown in Figure 1E, warfarin (WAR)-15E9 and its germline antibody WAR-15E9-GL both recognized their immunizing hapten-OVA conjugate, WAR-A-OVA (hapten shown in Figure S5). However, a significant difference in binding affinity was observed: 1.6 ng/mL for the mature antibody and 310 ng/mL for the germline antibody, indicating an increase of more than two orders of magnitude in binding potency during affinity maturation.

Similarly, as shown in Figure 1F, both the mature antibody tolfenamic acid (TLF)-3B4 and the germline antibody TLF-3B4-GL recognized the TLF-OVA conjugate (hapten shown in Figure S5). The binding affinities of TLF-3B4 and

TLF-3B4-GL were similar and within the same order of magnitude of 22 ng/mL and 12 ng/mL.

In summary, 73.7% (14/19) of the inferred germline antibodies inherently exhibited detectable binding affinities toward their corresponding structurally related hapten-OVA conjugates. Across the panel, germline antibodies generally exhibited reduced binding affinities and a narrower recognition spectrum relative to their affinity-matured counterparts. Notably, several germline antibodies—including SA-10E6-GL, ANG-3B3-GL, and TLF-3B4-GL—exhibited inherently high antigen-binding affinities at the ng/mL level.

To investigate the polyspecificity of germline antibodies, we evaluated their binding profiles against a diverse panel of 42 hapten-OVA conjugates prepared by our group (Figure S3). This hapten library of structurally related and unrelated small molecule encompassed a wide range of chemical classes, including common human and veterinary antimicrobials (e.g., sulfonamides, quinolones, tetracyclines), illicit additives (e.g., melamine, ractopamine), mycotoxins (e.g., aflatoxin, deoxynivalenol), environmental contaminants (e.g., polychlorinated biphenyls, bisphenol A), rodenticides (e.g., warfarin, brodifacoum), anti-inflammatory agents (e.g., anandamide, tolfenamic acid), and immunosuppressants (e.g., rapamycin). These haptens span a broad range of molecular weights and structural motifs, making them ideal candidates for probing the polyspecificity of germline antibodies. As shown in Figure 1G, each of the 14 tested germline antibodies displayed exclusive recognition of their corresponding structurally related hapten-OVA conjugates and showed no

detectable polyspecificity toward structurally unrelated hapten-OVA conjugates in the library. This unexpected finding underscores the intrinsically high specificity of germline antibodies targeting hapten molecules, suggesting that high specificity may be an inherent property of such germline antibodies.

3.2 Hapten-binding germline antibodies exhibit short CDRH3 loops and long CDRL1 loops

According to our aforementioned findings, germline antibodies that recognize hapten molecules generally display high specificity, whereas those targeting protein antigens tend to exhibit more polyspecificity (*2*). To further investigate the recognition differences of germline antibodies toward various antigen types, we constructed a comprehensive sequence database of mouse mAbs by integrating entries from the SAbDab database (*42*) with 124 antibody sequences targeting hapten molecules generated in our laboratory. Germline antibody sequences were then inferred and assembled using the IMGT/V-QUEST tool. Based on antigen type, the resulting non-redundant germline antibody database comprised 276 germline antibodies recognizing haptens (small molecules), 52 recognizing peptides, and 1,286 recognizing proteins.

The variable heavy (VH) chain lengths of germline antibodies were first analyzed. As shown in Figure 2A, the average VH length of hapten group was 118.5 amino acids (aa), compared to 120.8 aa for peptide group and 119.0 aa for protein group. The VH length of peptide group was significantly longer than that of both hapten and protein group (****$p < 0.0001$), and protein group also exhibited a significantly

longer VH region than hapten group (**p < 0.01). Overall, VH length of germline antibodies followed the order: Peptide > Protein > Hapten. Since the CDRs of antibodies play a central role in antigen recognition and their lengths strongly influence recognition specificity (*43*), we further analyzed the CDR lengths of germline antibodies. As shown in Figure 2B, the average heavy chain CDRs (CDRH) lengths of germline antibodies targeting haptens, peptides, and proteins were 27.7 aa, 30.2 aa, and 28.2 aa respectively. Peptide group had significantly longer CDRH regions than those targeting haptens or proteins (****p < 0.0001), and protein group also had longer CDRH regions than hapten group (***p < 0.001). This trend is consistent with the VH length differences observed among the three groups. Detailed analysis of individual CDRs (Figures 2C–2E) revealed that the average length of CDRH1 in hapten germline antibodies was 8.2 aa, significantly longer than the 8.0 aa observed in both peptide and protein group (**p < 0.01; ****p < 0.0001). In contrast, CDRH2 was longest in the peptide group (8.0 aa), followed by the protein group (7.9 aa) and the hapten group (7.8 aa) (***p < 0.001, *p < 0.05). For CDRH3, the average lengths were 11.6 aa, 14.2 aa, and 12.3 aa for hapten, peptide, and protein group, respectively. These differences followed a similar trend to that observed for CDRH2, with peptide germline antibodies exhibiting significantly longer CDRH3 regions than both protein and hapten group (****p < 0.0001), and protein group having longer CDRH3 regions than hapten group (****p < 0.0001). Notably, the differences in CDRH2 and CDRH3 lengths among the three groups closely mirrored the overall CDRH length trends (Peptide > Protein > Hapten). Since the length variation between

different group in CDRH3 (0.9–2.6 aa) was substantially greater than that in CDRH1 and CDRH2 (0–0.2 aa), it can be inferred that CDRH3 is the principal determinant of the observed differences in overall CDRH length among the antibody groups.

As shown in Figure S4, although statistically significant, the differences in heavy chain framework region (FRH) and FRH1 lengths among the three groups were relatively small—averaging only 0.1 to 0.2 aa. These differences were markedly smaller than those observed for CDRH lengths, which ranged from 0.5 to 2.5 aa. Overall, the VH region of the peptide group was significantly longer than those of both the hapten and protein groups, and the protein group also exhibited a significantly longer VH region than the hapten group. The variability in VH length among the three groups is largely attributable to differences in CDRH lengths, with CDRH3 contributing the most to this variation.

Furthermore, the variable light (VL) chain lengths of germline antibodies were analyzed. As shown in Figure 2F, the average VL length of hapten group was 109.4 aa, followed by 108.5 aa for protein group, and 107.3 aa for peptide group. The VL length of hapten group was significantly greater than that of both peptide and protein group (****$p < 0.0001$ and **$p < 0.01$, respectively), and the VL of protein group was also significantly longer than that of peptide group (***$p < 0.001$). Overall, VL length of germline antibodies followed the order: Hapten > Protein > Peptide. As shown in Figure 2G, the average CDRL lengths of hapten, peptide, and protein group were 20.8, 18.8, and 19.8 aa, respectively. The light chain CDRs (CDRL) length of hapten group was significantly longer than that of both peptide and protein group

(****$p < 0.0001$), and protein group also had significantly longer CDRL regions compared to peptide group (***$p < 0.001$). These trends align with the observed differences in VL lengths among the three groups, suggesting that VL length variation is largely driven by differences in CDRL length. Further analysis of individual CDRL components (Figures 2H–2J) revealed that the mean CDRL1 lengths for hapten, peptide, and protein group were 8.8, 7.0, and 7.8 aa, respectively. CDRL1 was significantly longer in hapten group compared to both peptide and protein group (****$p < 0.0001$), while the difference between protein and peptide group was not statistically significant. In contrast, the CDRL2 length was consistently 3.0 aa across all groups, with no significant differences observed. For CDRL3, the mean lengths were 9.0 aa for hapten group, 8.7 aa for peptide group, and 8.9 aa for protein group, with hapten group again showing significantly greater length compared to peptide group (****$p < 0.0001$). The average length difference in CDRL1 ranged from 0.8 to 1.8 aa, which was substantially larger than the 0.0–0.3 aa difference observed for CDRL3. Accordingly, length variations in CDRL3 among the three groups were minor compared with those in CDRL1. Together, these results indicate that the differences in CDRL length among the three groups are primarily driven by variations in CDRL1. As shown in Figure S4, while minor statistically significant differences were observed in some light chain framework region (FRL) segments (FRL1–FRL4), the overall average differences ranged only from 0.1 to 0.2 aa, which is negligible compared to the 1.0 to 2.0 aa differences observed in CDRL length. This supports the conclusion that VL length differences among the three groups are primarily driven by

differences in CDRL, particularly CDRL1. Notably, these analyses reveal a consistent length complementarity pattern between VH and VL regions across antigen classes: germline antibodies targeting haptens exhibited the shortest VH regions and longest VL regions; the germline antibodies targeting peptide showed the opposite trend, with the longest VH and the shortest VL regions; and the germline antibodies targeting protein displayed intermediate lengths in both chains. Similar length complementary patterns extended to the CDR level.

Moreover, the CDR length distributions were further analyzed. As shown in Figure 2K–2P, CDRH3 and CDRL1 were the most diverse loops, exhibiting the most pronounced differences in length distribution across the three germline antibody groups. The most frequent CDRH3 lengths observed in hapten-, peptide-, and protein-binding germline antibodies were 12, 13, and 14 amino acids, with corresponding frequencies of 25.2%, 36.5%, and 23.9%, respectively. The peptide group displayed a high occurrence of long CDRH3 loops, with prominent frequency peaks at 16, 18, and 22 amino acids, whereas the hapten and protein groups lacked longer CDRs, consistent with previous reports for affinity-matured antibodies (*44*). The hapten, peptide, and protein groups exhibited distinct CDRL1 length distributions, with prominent peaks at 6 aa (22.5%, 74.5%, and 46.6%, respectively), 9 aa (13.3%, 7.8%, and 2.2%), and 11 aa (39.9%, 11.8%, and 19.2%). Notably, the most frequent CDRL1 length in the hapten group was 11 aa (39.9%), which is substantially longer than the predominant 6 aa length observed in the peptide (74.5%) and protein (46.6%) groups. Overall, the hapten group was enriched in longer CDRL1 loops, whereas the

peptide and protein groups were dominated by shorter loops, a pattern consistent with previous observations in mature antibodies (*27*). In contrast, the other CDRs exhibited similar length distributions, with the most frequent lengths being consistent across the three groups: 8 amino acids for CDRH1 and CDRH2, 3 for CDRL2, and 9 for CDRL3, each accounting for more than 60.1% of the sequences. In short, the protein and peptide groups exhibited a higher proportion of long CDRH3 loops, whereas the hapten group showed a greater prevalence of long CDRL1 loops.

Overall, we firstly found that VH lengths of germline antibodies followed the order Peptide > Protein > Hapten, whereas VL lengths displayed the opposite trend, Hapten > Protein > Peptide. The variations in VH length across the three groups are primarily driven by differences in CDRH length—particularly CDRH3—whereas the differences in VL length are governed by CDRL, especially CDRL1. The length complementary patterns observed between VH and VL regions are also reflected between CDRH and CDRL, as well as between CDRH3 and CDRL1.

3.3 Hapten-binding germline antibodies display enrichment of aromatic residues in their CDRs

The amino acid composition of CDR loops have a significant influence on the affinity, specificity and biophysical characteristics of an antibody (*45*).To further characterize the polyspecific recognition mechanism of germline antibodies targeting haptens, proteins and peptides, we analyzed the residue composition of their CDR profiles. As shown in Figure 3A–3C, CDRH1 across all three groups was predominantly enriched in phenylalanine (Phe), Gly, serine (Ser), threonine (Thr),

368 tryptophan (Trp), and tyrosine (Tyr) with frequency ranging from 1.6% to 4.4%, while
369 CDRH2 was enriched in Gly, Ile, asparagine (Asn), Ser, Thr, and Tyr with frequency
370 ranging from 1.5% to 5.6%. In contrast, CDRH3 was characterized by high
371 frequencies of aspartic acid (Asp), Phe, arginine (Arg), Trp, and Tyr ranging from
372 1.8% to 7.3%. Overall, aromatic residues (Phe, Trp, Tyr) and polar residues (Ser, Thr,
373 Asp, Asn, Arg) were highly accumulate in CDRH1, CDRH2, and CDRH3, accounting
374 for a major proportion of their composition (amino acid classification see Table S4).

375 Furthermore, categorical statistical analysis revealed minimal differences among
376 the three groups in CDRH1, with variations limited mainly to tiny and charged
377 residues. In CDRH2, the hapten group exhibited a significantly higher proportion of
378 aromatic residues (10.3%), mainly contributed by Trp (1.7%) and Tyr (2.1%),
379 compared with the peptide and protein groups. In contrast, the peptide and protein
380 groups showed a greater abundance of polar residues (53.4% and 51.9%, respectively),
381 predominantly Ser (4.4% and 2.8%, respectively). In CDRH3, the proportion of tiny
382 and small residues decreased markedly across all three groups, reflecting an overall
383 enrichment of bulky residues. For example, small residues accounted for
384 57.1%–58.1% in CDRH1 and 72.6%–76.6% in CDRH2 across the three groups, but
385 this proportion dropped to 43.8%–45.2% in CDRH3. The hapten group exhibited a
386 significantly higher proportion of aliphatic residues (21.0%), primarily alanine (Ala,
387 1.8%), compared with the peptide and protein groups. In contrast, the peptide and
388 protein groups showed increased proportions of polar residues (41.0% and 33.4%,
389 respectively), largely contributed by serine (Ser, 1.4% and 0.8%, respectively),

relative to the hapten group. With respect to the three most enriched aromatic residues, the hapten and peptide groups contained significantly higher proportions of Phe (2.2% and 2.6%, respectively) and Trp (4.0% and 3.9%, respectively) than the protein group, whereas the protein group was significantly enriched in Tyr (7.3%). As a result, the total aromatic residue content did not differ significantly in the CDRH3 among the three groups. In summary, germline antibodies targeting haptens were enriched in hydrophobic residues within their CDRH regions, whereas those targeting peptides and proteins exhibited a greater enrichment of polar residues, similar to affinity-matured antibodies (*27*).

As shown in Figure 3D–3E, CDRL1 across all three groups was predominantly enriched in Asn, glutamine (Gln), Ser, Thr and Tyr with frequency ranging from 0.9% to 3.9%, while CDRL2 was enriched in Ala, Ser, Thr, Trp, and Tyr with frequency ranging from 0.6% to 4.5%. In contrast, CDRL3 was characterized by high frequencies of histidine (His), proline (Pro), Gln, Ser, Thr, Trp, and Tyr with frequency ranging from 1.5% to 5.6%. Similar to CDRH, aromatic residues and polar residues were highly accumulating in CDRL, accounting for a major proportion of their composition. Further categorical statistical analysis revealed significant differences among the three groups in CDRL1, CDRL2, and CDRL3. In CDRL1, both the hapten and protein groups exhibited significantly higher proportions of aromatic residues (15.0% and 13.8%, respectively), predominantly Tyr (3.5% and 3.2%), compared with the peptide group. In contrast, the protein group showed a higher abundance of polar residues (64.0%), mainly Ser (3.7%), as well as a greater

proportion of charged residues (11.5%) than the hapten group. In CDRL2, the peptide and protein groups exhibited significantly higher proportions of aromatic residues (7.7% and 13.4%, respectively) than the hapten group (4.3%), whereas the hapten group was enriched in polar residues (61.8%). Similar to CDRH3, the proportion of tiny and small residues was markedly reduced in CDRL3, indicating an accumulation of bulky residues. For example, small residues accounted for 62.2%–65.2% in CDRH1 and 73.7%–85.3% in CDRH2 across the three groups, but this proportion dropped to 41.4%–49.4% in CDRH3. In CDRL3, the hapten and peptide groups contained higher levels of aromatic residues (27.9% and 28.2%, respectively) than the protein group (24.6%), whereas the protein group exhibited a higher proportion of polar residues (54.5%). Overall, the three germline antibody groups exhibit distinct and complicated residues composition signatures across CDRL1–L3, with aromatic residues enriched in hapten- and peptide-binding germline antibodies, whereas protein-binding germline antibodies show higher proportions of polar and charged residues. Aromatic and polar residues in the CDRs facilitate hydrophobic contacts, electrostatic interactions, and other non-covalent forces that are essential for antigen recognition and specificity, while simultaneously providing a molecular basis for polyspecific binding (*35*). Thus, relative to the hapten group, the markedly higher proportions of polar residues in CDRH2, CDRH3, and CDRL1 in the peptide- and protein-binding groups may help explain their enhanced polyspecificity (*27*).

3.4 Hapten-binding germline antibodies demonstrate structural rigidity

The polyspecific binding properties of germline antibodies has been largely attributable to their highly flexible and adaptable binding pockets (*14–17*, *30*, *31*). To further characterize the polyspecific recognition mechanisms of germline antibodies against different classes of antigens, we constructed 3D models of these germline antibodies using AlphaFold3 (*46*) and calculated secondary-structure occupancies for each group using Define Secondary Structure of Proteins (DSSP). As shown in Figure 4A, the helix content of the hapten and protein groups was significantly higher than that of the peptide group (***$p < 0.001$). The sheet content of the hapten group was also significantly higher than that of the protein group (**$p < 0.01$). The coil contents of the hapten, peptide, and protein groups were 42.98%, 43.89%, and 43.36%, respectively. The peptide group showed a significantly higher coil proportion than both the hapten and protein groups (***$p < 0.001$ and **$p < 0.01$, respectively), and the protein group was significantly higher than the hapten group (*$p < 0.05$). Because differences in coil content—primarily located within the CDRs—directly influence overall structural flexibility, these results suggest that secondary-structure composition may be a critical factor contributing to germline antibody polyspecificity. This effect appears to be especially pronounced for germline antibodies targeting peptide and protein antigens.

Furthermore, we used the predicted local distance difference test (pLDDT) score to characterize structural flexibility, with lower scores indicating more flexible regions. As shown in Figure 4B, the pLDDT scores of the Fv regions in the hapten, peptide,

and protein groups were 96.0%, 95.7%, and 96.3%, respectively. The peptide group exhibited significantly lower pLDDT scores than both the hapten and protein groups (*$p < 0.05$ and ***$p < 0.001$, respectively), indicating higher structural flexibility. The hapten group also showed significantly lower scores than the protein group (***$p < 0.001$). Similar trends were observed for both VH and VL regions. As shown in Figure 4C, the pLDDT scores of the CDRs—and of each individual CDR—followed the same trends observed in the Fv regions. The pLDDT scores of CDRH3 in the hapten, peptide, and protein groups were 84.0%, 81.0%, and 86.2%, respectively, confirming that CDRH3 is the most flexible region within germline antibodies. Taken together, the 2D and 3D flexibility analyses indicate that germline antibodies recognizing peptides exhibit the higher structural flexibility, whereas those recognizing haptens and proteins display the greater structural rigidity.

3.5 Hapten-binding germline antibodies feature fewer CDR-localized surface patches

The surface properties of antibodies, particularly within their CDR regions, are key determinants of antigen-recognition specificity (*35*, *37*). Therefore, we performed an in-depth analysis and compare of the SASA and surface patches of the germline antibodies. Firstly, the SASA of the Fv region in each germline antibody group was calculated. As shown in Figure 5A, the average total SASA of the hapten, peptide, and protein groups was 10,458.8 $Å^2$, 10,578.8 $Å^2$, and 10,452.7 $Å^2$, respectively. The peptide group exhibited a larger SASA than both the hapten and protein groups (**$p < 0.01$ and ***$p < 0.001$, respectively), including increases in both hydrophilic and

hydrophobic SASA. When comparing the hapten and protein groups, the hapten group showed a significantly larger hydrophilic SASA (***$p < 0.001$), whereas the protein group displayed a significantly larger hydrophobic SASA (***$p < 0.001$). Furthermore, SASA values for the CDRs were calculated and analyzed. As shown in Figure 5B, the average total CDR SASA of the hapten, peptide, and protein groups was 2,209.1 Å^2, 2,303.0 Å^2, and 2,206.8 Å^2, respectively. Again, the peptide group exhibited significantly higher SASA than the hapten and protein groups (***$p < 0.001$), consistent with the trend observed for the total Fv SASA. Specifically, the hapten group exhibited the largest SASA in CDRH1, while the peptide group showed the largest SASA in CDRH2 and CDRH3. The average SASA of CDRH3 was 483.6 Å^2, 683.7 Å^2, and 551.5 Å^2 for the hapten, peptide, and protein groups, respectively—representing the largest SASA among all CDRH. For light chain, the hapten group exhibited the largest SASA in CDRL1 and CDRL2, while the peptide group showed the largest SASA in CDRL3. The average SASA of CDRL1 was 519.0 Å^2, 412.5 Å^2, and 456.2 Å^2 for the hapten, peptide, and protein groups, respectively—representing the largest SASA among all CDRL. Notably, CDRH3 and CDRL1 exhibited the largest magnitude and variability in surface area among the six CDRs, largely due to their diverse loop lengths. Moreover, the hapten group showed a larger CDRL1 area, whereas the protein group displayed a larger CDRH3 area, consistent with the CDR length distributions presented in Figure 2K–2P.

Furthermore, we analyzed hydrophobic, positive charged, and negative charged surface patches (SPs) across the entire Fv and CDR regions of these germline

antibodies using MOE. As shown in Figures 5C–5E, the hapten, peptide and protein groups exhibited average total SP counts of 21.8, 23.8, and 20.8, respectively. Germline antibodies in the peptide group showed significantly higher total SP numbers, total SP surface areas, and SP coverage percentages than the other two groups. In addition, the hapten group displayed significantly larger numbers, total areas, and proportions of SPs than the protein group. For hydrophobic SPs, the hapten, peptide, and protein groups exhibited average counts of 5.0, 5.9, and 4.4, and average SP areas of 398.0 $Å^2$, 462.7 $Å^2$, and 358.7 $Å^2$, respectively. The peptide group had significantly higher hydrophobic SP numbers, total areas, and proportions than both the hapten and protein groups, while the hapten group also exceeded the protein group in all three metrics. For positive charged SPs, the hapten, peptide, and protein groups exhibited average counts of 9.8, 10.4, and 9.3, with corresponding areas of 488.8 $Å^2$, 496.3 $Å^2$, and 462.6 $Å^2$. The peptide group contained significantly more positively charged SPs than the hapten and protein groups in terms of number, although differences in total area and proportion were not significant. The hapten group also exhibited significantly higher counts, total areas, and proportions than the protein group. For negative charged SPs, the hapten, peptide, and protein groups showed average counts of 7.0, 7.5, and 7.5, with areas of 376.7 $Å^2$, 427.3 $Å^2$, and 409.2 $Å^2$, respectively. Both the peptide and protein groups contained significantly more negative charged SPs than the hapten group. Overall, germline antibodies in the peptide group possessed the largest number of SPs and consistently ranked highest across all SP subclasses. In contrast, the hapten group contained more total,

hydrophobic, and positive charged SPs than the protein group, but exhibited relatively fewer negative charged SPs.

We further analyzed the SPs located in the vicinity of the CDRs. As shown in Figures 5F–5H, the hapten, peptide and protein groups exhibited average total CDR SP counts of 7.1, 7.9, and 7.6, and average CDR SP areas of 480.9 $Å^2$, 536.3 $Å^2$, and 507.8 $Å^2$, respectively. Germline antibodies in the peptide and protein groups showed significantly higher numbers, total areas, and proportions of CDR SPs than those in the hapten group, whereas no significant differences were observed between the peptide and protein groups. For hydrophobic SPs, the hapten, peptide, and protein groups exhibited average counts of 2.0, 2.4, and 1.8, with corresponding areas of 194.6 $Å^2$, 236.3 $Å^2$, and 185.8 $Å^2$. The peptide group displayed significantly higher hydrophobic SP numbers, total areas, and proportions than both the hapten and protein groups, while no significant difference was observed between the hapten and protein groups. For positive charged SPs, the hapten, peptide, and protein groups showed average counts of 3.7, 3.5, and 3.8, and average areas of 203.2 $Å^2$, 183.5 $Å^2$, and 209.8 $Å^2$, respectively. No significant differences were detected among the three groups. In contrast, for negatively charged SPs, the hapten, peptide, and protein groups exhibited average counts of 1.4, 2.0, and 2.0, with areas of 83.1 $Å^2$, 116.5 $Å^2$, and 112.2 $Å^2$, respectively. Both the peptide and protein groups contained significantly more negative charged SPs than the hapten group. Overall, germline antibodies in the peptide group contained more CDR SPs—particularly hydrophobic and negative charged SPs—than those in the other groups. The protein group also

exhibited more negative charged SPs than the hapten group, contributing to its relatively higher total number of CDR surface patches, consistent with the elevated proportions of polar and charged residues observed in their CDRs (Figure 3). Collectively, the peptide group showed the greatest abundance of SPs in both the Fv and CDR regions, indicating a strong potential for polyspecific recognition. Meanwhile, the greater density of CDR-localized SPs in the protein group suggests an additional structural basis for its polyspecificity.

3.6 Germline gene usage biases shape the antigen-binding properties of germline antibodies targeting haptens, peptides and proteins

V(D)J germline gene recombination is a fundamental mechanism generating antibody diversity and shaping antigen-recognition specificity. To further investigate the potential molecular basis of polyspecificity in germline antibodies, we analyzed the germline gene usage profiles of antibodies targeting different classes of antigens. As shown in Figures 6A, the cumulative usage frequency of the top 20 IGHV and IGKV/IGLV germline genes accounted for over 50% of the repertoires in hapten and protein groups, while exceeding 80% in peptide group—likely a consequence of the relatively small sample size for peptide group. Notably, germline gene usage varied substantially across the three groups. As shown in Figures 6A and Table S4, the hapten group, exhibited the highest usage frequency of IGHV5-6-5 (8.4%), followed by IGHV3-2, IGHV1-76, IGHV1-9, and IGHV9-3-1, each exceeding 4.0%. The peptide group preferentially used IGHV5-9-3 (36.5%), with IGHV5-17, IGHV5-9, IGHV4-1, and IGHV5-6-3 also showing frequencies above 3.8%. In the protein group,

IGHV5-17 was the most enriched (14.3%), followed by IGHV14-3, IGHV1S81, IGHV1-26, and IGHV4-1, each with frequencies greater than 4.2%. As shown in Figures 6A and Table S5, the hapten group most frequently employed IGKV1-117 (15.8%), followed by IGLV1, IGKV1-110, IGKV1-133, and IGKV6-15, all exceeding 3.7%. The peptide group predominantly used IGKV11-125 (37.3%), with IGKV10-94, IGLV1, IGKV13-85, and IGKV6-15 also surpassing 5.9%. The protein group was most enriched for IGKV2-137 (14.5%), followed by IGKV5-48, IGKV4-59, IGKV10-96, and IGKV5-43, each exceeding 3.9%. Despite the distinct V-gene usage patterns observed among the three antigen groups, several shared features emerged. For example, members of the IGHV5 and IGHV1 germline gene families were frequently used across all three groups, while the IGKV1 and IGLV1 germline gene families were preferentially used in both the hapten and peptide groups. Regarding germline usage diversity, IGHV genes displayed greater diversity than IGKV/IGLV genes, with IGKV showing higher diversity than IGLV, reflecting the inherent complexity of their respective germline pools and recombination mechanisms (*2*). Moreover, the hapten and protein groups exhibited substantially greater germline usage diversity than the peptide group, likely influenced by sample size limitations in the latter.

Due to the limited J-gene repertoire, all three groups utilized all four IGHJ gene segments, although their usage frequencies differed. The hapten group showed the highest usage of IGHJ3 (37.6%), whereas the peptide and protein groups predominantly used IGHJ2, at 51.9% and 31.9%, respectively. For IGKJ/IGLJ gene

usage, IGKJ segments (IGKJ1, IGKJ2, IGKJ4, and IGKJ5) were employed at substantially higher frequencies than IGLJ segments (IGLJ1, IGLJ2, and IGLJ3) across all three groups. Among them, the hapten group exhibited the highest usage of IGLJ genes, with a combined frequency of 13.3% for IGLJ1–3, compared with 7.8% in the peptide group and only 2.2% in the protein group. In summary, the three groups exhibited distinct patterns of V gene and J gene usage.

Next, we analyzed the combinatorial usage frequencies of V–J germline gene combination in the VH and VL regions across the three groups. As shown in Figure 6B, the IGHV–IGHJ gene combination profiles varied markedly among the groups. The hapten group exhibited the broadest usage spectrum, with high-frequency combinations particularly involving IGHJ3 and IGHJ4, including IGHV1-76–IGHJ3 (3.6%) and IGHV5-6-5–IGHJ4 (2.9%). In contrast, the peptide group strongly favored IGHV5-9-3–IGHJ2 (34.6%) and IGHV5-9–IGHJ1 (11.5%), whereas the protein group was most enriched for IGHV5-17–IGHJ2 (14.0%) and IGHV14-3–IGHJ4 (5.8%). For IGKV–IGKJ combinations, the hapten group predominantly utilized IGKV1-117–IGKJ1 (7.4%) and IGKV1-110–IGKJ1 (4.8%). The peptide group preferentially used IGKV11-125–IGKJ2 (7.8%) and IGKV10-94–IGKJ1 (19.6%), while the protein group was enriched for IGKV2-137–IGKJ5 (14.2%) and IGKV5-48–IGKJ4 (6.0%). Due to the limited diversity of IGLV germline genes, no major differences were observed among groups in IGLV–IGLJ combinations. All three groups showed the highest usage of the IGLV1–IGLJ1 combination. In summary, the three groups exhibited distinct V–J germline combination patterns, particularly

within IGHV–IGHJ and IGKV–IGKJ repertoires.

Antibody VH–VL pairing patterns are key determinant of antigen-recognition specificity. Therefore, we examined the pairing frequencies of IGHV–IGKV/IGLV germline gene across the three groups. As shown in Figure 6C, the hapten group displayed broad pairing diversity, with IGKV1-110 and IGKV1-117 frequently pairing with multiple IGHV genes. Among these, IGHV1-76/IGKV1-133 (3.7%), IGHV5-6-5/IGKV1-117 (3.3%), and IGHV9-3-1/IGKV1-110 (2.2%) represented the most enriched VH–VL pairing patterns. In the peptide group, IGHV5-17 was predominantly paired with several IGKV germline genes, and the IGHV5-9-3/IGKV11-125 combination exhibited the highest usage frequency (29.4%). In the protein group, the most enriched VH–VL pairs were IGHV5-17/IGKV2-137 (14.3%), IGHV1S81/IGKV6-13 (5.9%), and IGHV14-3/IGKV14-100 (3.8%). In summary, the three group of germline antibodies exhibited significantly different in IGHV–IGKV/IGLV pairing patterns. These observed differences in individual V-gene usage, V–J recombination patterns, and IGHV–IGKV/IGLV pairing patterns likely reflect germline-encoded features that shape antigen recognition and may contribute to the polyspecific nature of germline antibodies.

The usage of specific V genes determines both the length and the residue composition of CDR1 and CDR2, thereby shaping the polyspecificity contributed by these regions in the VH and VL domains. As shown in Table S4 and Table S5, germline antibodies recognizing haptens preferentially employ IGKV/IGLV genes that encode substantially longer CDRL1 loops—most notably 9-, 10-, and 11-residue

CDRL1s. The most dominant IGKV1-117 (15.8%), as well as IGKV1-110 (9.9%) and IGKV1-133 (5.1%), all originate from IGHV1 germline gene family and encode an 11-residue CDRL1, consistent with the longer CDRL1 distribution observed for the hapten group (Figure 2N). Moreover, germline antibodies recognizing proteins preferentially utilize IGKV/IGLV germline genes enriched for distinct residue features in their CDRs. These include higher frequencies of Phe and Gly in CDRH1, increased isoleucine (Ile) in CDRH2, elevated Ser and Ile in CDRL1, and higher Ala and Tyr in CDRL2. Together, these V-gene-encoded sequence characteristics may contribute to the polyspecific binding profiles characteristic of germline antibodies targeting proteins.

Moreover, preferences in germline gene usage shape the structural properties of germline antibodies. For example, germline antibodies using the IGHV5-17/IGKV2-137 gene pair in protein group exhibited a coil content of 47.1%, which is significantly higher than the group-wide average of 43.4% (Table S6). A similar trend was observed in the pLDDT scores across individual regions (Table S7), indicating that this IGHV–IGKV/IGLV gene pair encodes germline antibodies with increased structural flexibility. As shown in Table S7, germline antibodies recognizing peptides employing the top five most frequently used IGHV–IGKV/IGLV gene pairs all displayed higher coil content than the corresponding group-wide average. In contrast, three of the top five pairs in the hapten group exceeded the average, whereas only two of the top five IGHV–IGKV/IGLV gene pairs in the protein group did so. Those results indicate that the germline gene usage shape the structural flexibility of

germline antibodies.

Furthermore, preferences in germline gene usage shape the surface properties of germline antibodies. For example, germline antibodies utilizing the IGHV5-17/IGKV11-125 gene pair exhibited higher SASA across all three antigen groups, a feature that was also reflected by increased total SP areas in both the Fv and CDR regions (Tables S8–S10). In the hapten group, antibodies encoded by the IGHV1-72/IGLV1 gene pair displayed elevated total CDR-localized SPs as well as increased numbers and areas of negative charged SPs (Table S10). Similarly, in the protein group, the IGHV1S81/IGKV5-48 gene pair was associated with higher total CDR SPs and an enrichment of negative charged SPs in both number and area (Table S10). Collectively, these results suggest that preferences in germline gene usage contribute to the emergence of distinct polyspecific recognition patterns in germline antibodies.

3. Discussion

Polyspecificity, the capacity of an antibody to recognize multiple distinct epitopes, remains an elusive concept, and the molecular features underlying this property have not yet been sufficiently defined (*2*, *5*, *6*). In this work, we performed the first large-scale functional characterization of germline antibodies derived from mAbs targeting haptens. Specifically, we inferred 19 germline antibodies from 23 mAbs, of which 14 retained detectable binding affinities toward their corresponding antigens. These results demonstrate that germline inference using the IMGT/V-QUEST tool is

an effective and reliable approach for reconstructing functional germline antibodies (*5*). We then systematically evaluated the polyspecificity of these germline antibodies using a diverse hapten-OVA conjugates array. Surprisingly, all identified germline antibodies exhibited intrinsically high specificity, with no evidence of polyspecific binding. This unexpected finding highlights that germline antibodies targeting haptens can possess a high degree of intrinsic specificity, challenging the long-held notion that germline antibodies are inherently polyspecific.

To investigate the molecular basis of germline antibody specificity, we constructed large-scale sequence-structure database containing 1614 germline antibodies targeting haptens, peptides and proteins. CDR length plays a critical role in shaping the nature of antigen-antibody binding (*43*). The characteristic CDR length patterns observed in hapten-, peptide-, and protein-binding germline antibodies may therefore be key determinants of antigen-binding specificity. In this study, we found that protein- and peptide-binding germline antibodies exhibit a significantly greater prevalence of long CDRH3 loops than hapten-binding germline antibodies, consistent with observations previously reported in mature antibodies (*44*). CDRH3 exhibits the greatest variability in length, sequence, and conformation, and is widely recognized as the principal determinant of antigen specificity (*1*, *47*, *48*). Antibodies with long CDRH3 loops tend to form the majority of antigen contacts, enabling recognition of diverse epitopes, whereas shorter CDRH3 loops often produce concave paratope in which other CDRs contribute more substantially to antigen engagement (*49*). Notably, extended CDRH3 loops are also considered negative developability features, as they are frequently

associated with polyspecific and nonspecific interactions (*37*). Accordingly, the high proportion of long CDRH3 loops in protein- and peptide -binding germline antibodies may underlie their polyspecific binding behavior. In contrast, we found that hapten-binding germline antibodies exhibit a significantly greater prevalence of long CDRL1 loops, consistent with previous observations in mature antibodies (*27*). Research on intrinsically oligoclonal multispecific antibody (IOMA)–class antibodies provides a relevant precedent, demonstrating that an extended CDRL1 can impart the conformational flexibility necessary to engage the gp120 N276 glycan (*50*). However, to date, the functional implications of long CDRL1 enrichment in hapten-specific germline antibodies have not been investigated, and this feature therefore warrants further systematic study. Collectively, these observations suggest that CDR length—particularly that of CDRH3 and CDRL1—plays a central role in shaping both specificity and polyspecificity of germline antibodies. Notably, the combination of shorter CDRH3 and longer CDRL1 observed in hapten-binding germline antibodies may underlie their intrinsically high specificity. Additionally, it is worth noting that the complementary length patterns observed between the VH and VL regions were also reflected at the level of the CDRs, particularly between CDRH and CDRL, as well as between CDRH3 and CDRL1. However, whether this length complementarity plays a functional role in recognition specificity remains to be elucidated.

Previous studies have shown that germline antibodies natural enrich of Tyr, Ser, and glycine (Gly) reflects their intrinsic capacity to mediate their polyspecificity property that could flexible and effective antigen engagement (*51*). Aromatic

717 residues—particularly Tyr and Trp, which are present at high proportions in the
718 CDRs—are frequently involved in antigen recognition (*52*, *53*). Polyspecific or
719 non-specific antibodies are frequently enriched in Phe, Trp, and Tyr in their CDRs,
720 residues that are known to possess mixed hydrophilic, hydrophobic, and aromatic
721 binding characteristics (*51*, *54–56*). This mixed chemical nature of CDRs allows these
722 residues to engage in a wide variety of interaction types, conferring a generally
723 "sticky" binding behavior (*44*). Furthermore, Ser was also commonly identified in
724 non-specific antibodies, which is likely explained by their polar, small and flexible
725 side chains allowing for increased promiscuity in structure and interactions (*55*, *57*).
726 In this work, we observed that germline antibody CDRs are enriched with bulky
727 residues, particularly aromatic residues such as Trp and Tyr. Notably, the protein
728 group accumulated significantly more Tyr and Ser in CDRH3, CDRL1, and CDRL3
729 than the hapten group, which may underlie enhanced polyspecificity of protein
730 germline antibody.

731 Conformational flexibility of germline antibodies is a major driver of their
732 polyspecificity, enabling the paratope to adopt distinct binding-site topologies.
733 Literature suggests that many polyspecific antibodies in the germline repertoire
734 (*58–60*) can be simultaneously improved in affinity and specificity through somatic
735 mutations that increase binding-pocket rigidity in the unbound state (*4*, *61*). Reducing
736 conformational flexibility within the binding site lowers the entropic penalty
737 associated with target engagement, thereby enhancing affinity. Increased rigidity also
738 constrains the promiscuity of interactions, which in turn improves binding specificity.

Thus, compared with hapten- and protein-binding germline antibodies, those recognizing peptides exhibit significantly higher structural flexibility, which may confer a higher degree of polyspecific recognition. However, this possibility has received little experimental investigation to date.

Hydrophobic, positive charged, and negative charged SPs are three of the five most important properties influencing therapeutic antibody developability, and they are often associated with significant polyspecific or even nonspecific tendencies (*37*). Structurally, antibody surfaces—especially within the CDRs—often contain hydrophobic patches and complementary charged patches that can act as promiscuous binding hotspots, enabling polyspecific or non-specific recognition (*62*). In this work, we comprehensively quantified the differences in hydrophobic, positive charged, and negative charged SPs of germline antibodies. We observed that germline antibodies targeting peptides exhibited the highest abundance of total SPs, as well as hydrophobic, positive charged, and negative charged SPs, in both the Fv and CDR regions, suggesting a strong propensity for polyspecific recognition. Meanwhile, the higher density of CDR-localized total SPs and negative charged SPs in protein-targeting germline antibodies suggests an additional structural basis for their polyspecificity. In contrast, hapten-targeting germline antibodies displayed substantially fewer CDR-localized total SPs and negative charged SPs than protein-binding germline antibodies, which may underlie their inherently high specificity. Taken together, we have connected potential polyspecific binding to the surface patch properties across a germline antibody library trying to discern the

feature of polyspecificity.

Germline gene segments—particularly IGHV genes—encode CDR1, CDR2, and the FR region, which are critical contributors to polyspecificity in germline antibodies (*4*). Furthermore, different classes of antigens tend to preferentially activate specific germline genes. For example, among the six commonly used human VH germline sequences, only IGHV1-2 confers the ability to recognize and accommodate structurally diverse lipopolysaccharide molecules circulating in the bloodstream (*63*). VRC01-class broadly neutralizing antibodies targeting the HIV Env receptor-binding site uniformly derive from the IGHV1-2*02 (*64*). Additional examples of germline-encoded features in protein-binding antibodies are well documented in our previous review (*2*). In this work, we observed that germline-usage bias—including the preferential selection of specific V(D)J gene segments, their recombination patterns, and the pairing of VH and VL chains—collectively determines CDR length and residue composition, conformational flexibility, and surface physicochemical properties, thereby shaping the polyspecificity of germline antibodies of mice. This study establishes a systematic framework for investigating antibody polyspecificity and elucidating its underlying molecular mechanisms. Importantly, it presents the first large-scale, systematic analysis and visualization of germline gene-usage landscapes, revealing distinct antigen-specific usage patterns and preferences.

4. Materials and Methods

2.1 Germline antibody analysis, expression, and validation

Germline antibody sequences were inferred and assembled using IMGT/V-QUEST within the IMGT database (*65*). The germline antibodies were subsequently expressed by transient transfection in HEK293 cells. Briefly, cells were suspended in expression medium and co-transfected with plasmids encoding the heavy- and light-chain genes, followed by the immediate addition of linear PEI MAX (Polysciences, Warrington, PA, USA). After 3 h at 37 °C, the cell density was adjusted with ExCell VPRO medium and cultures were maintained for seven days. Supernatants were harvested, filtered, and diluted 1:1 in PBS (pH 7.2) prior to loading onto a Protein A column on an ÄKTA Pure system. Bound IgG was eluted with 0.1 M citrate buffer (pH 3.0) into 1.0 M Tris-HCl (pH 9.0), followed by buffer exchange by dialysis. The identity and integrity of each germline antibody were confirmed by reduced and non-reduced sodium dodecyl sulfate-polyacrylamide gel electrophoresis SDS-PAGE.

Germline antibody binding affinity toward hapten-OVA conjugates was assessed using a noncompetitive ELISA (ncELISA). Binding affinity was expressed as the half-maximal effective concentration (EC_{50}), calculated from antibody dilution curves. Briefly, polystyrene 96-well microplates were coated with the conjugates (1 μL/mL) in coating buffer (100 μL/well) and incubated for 16 h at 4 ℃. Plates were washed three times with washing buffer and subsequently blocked with blocking buffer (300 μL/well) for 1 h at 37 ℃. Antibody samples (100 μL/well), with different

concentrations, were added and incubated for 30 min at 37 ℃. After washing, bound antibodies were detected with goat anti-mouse IgG-HRP (1:5000 in dilution buffer, 100 μL/well) for 30 min at 37 ℃, followed by three washes. TMB substrate (100 μL/well) was then added and developed for 15 min at room temperature before quenching with stop solution (50 μL/well). The absorbance was recorded at 450 nm. Curve fitting and statistical analysis were performed using a four parameters logistic model by OriginPro 8.5 (OriginLab Corporation, Northampton, MA) as follows:

$$Y = \frac{(A - D)}{[1 + (X/C)^B + D]}$$

where A is the asymptotic maximum, B is the curve slope at the inflection point, C is the concentration at which 50% of the mAb are bound to the conjugates (EC_{50}), D is the asymptotic minimum, and X is conjugates concentration.

2.2 The antibody sequence and amino acid composition analysis

CDR and FR annotations using IMGT numbering were obtained for all antibody sequences through the abYsis database (http://www.abysis.org/abysis/). The lengths of VH and VL regions, as well as the CDRs and FRs within VH and VL, were subsequently analyzed. Amino acid composition of VH and VL CDRs was assessed using the EMBOSS Pepstats tool (https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/). Germline gene usage, including V(D)J germline gene frequency, V-J germline gene combination frequency, IGHV-IGKV/IGLV combination frequency, and antibody mutation frequency, was analyzed using the IMGT/DomainGapAlign tool (https://www.imgt.org/).

2.3 Second structure analysis and pLDDT analysis

The three-dimensional structures of germline antibodies were predicted using AlphaFold3 (https://alphafoldserver.com/) (*66*). Per-residue predicted local distance difference test (pLDDT) scores were extracted to assess structural confidence. Secondary-structure probabilities were computed using DSSP (*67*), which assigns secondary-structure states based on a hydrogen-bond definition with an energy cutoff of −0.5 kcal mol^{-1}. In the DSSP output, α-helices, 310-helices, and π-helices were categorized as “Helix”; extended β-strands and β-bridges as “Sheet”; and and random coil as “Coil.”

2.4 SASA analysis and surface patch analysis

The surface properties of germline antibodies, including SASA and surface patches, were computed using the Protein Properties application in MOE 2022.02. For each model, 100 conformations were generated in which the framework was restrained and side chains were sampled using LowModeMD. Alternative protonation states were explored from pH 6.4 to 8.4 (centered at pH 7.4) using the Protonate-3D method (*68*, *69*). Hydrophobic patches were defined as regions in which a hydrophobic potential equal to or greater than that of a methyl group was maintained across > 50 $Å^2$ of surface area. Hydrophobic potential was calculated using the SLogP method on an atom-by-atom basis and mapped onto the molecular surface. Charged patches were defined as regions sustaining excess force-field charge (Amber10:EHT) over >40 $Å^2$ of surface area. CDR-located surface patches were computed similarly, by summing

842 the surface areas of hydrophobic patches comprising at least one CDR atom, and
843 averaging across the 100 conformations using a Boltzmann weighting centered at the
844 target pH (7.4).

845 **2.5 Statistical analysis**

846 Statistical analysis was performed using GraphPad Prism 8.0 (GraphPad Software,
847 Inc.). A two-tailed Mann-Whitney test was used for all statistical analyses (ns, Not
848 Significant; *p <0.05; **p <0.01; ***p <0.001; ****p <0.0001).

849

850 ## Conflict of interest statement

851 The authors have declared that no conflict of interest exists.

852

853 ## Acknowledgments

854 This work is supported by the National Natural Science Foundation of China (Grant
855 No. 32473085).

856

857 ## Appendix A. Supplementary data

858 Supplementary data associated with this article can be found in the online version
859 at https://.

860

References

1. H. W. Schroeder, L. Cavacini, Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology* **125**, S41–S52 (2010).

2. Y. Zhang, Q. Li, L. Luo, C. Duan, J. Shen, Z. Wang, Application of germline antibody features to vaccine development, antibody discovery, antibody optimization and disease diagnosis. *Biotechnology Advances* **65**, 108143 (2023).

3. S. Vajda, K. A. Porter, D. Kozakov, Progress toward improved understanding of antibody maturation. *Curr Opin Struct Biol* **67**, 226–231 (2021).

4. J. R. Willis, B. S. Briney, S. L. DeLuca, J. E. Crowe, J. Meiler, Human germline antibody gene segments encode polyspecific antibodies. *PLoS Comput Biol* **9**, e1003045 (2013).

5. A. T. DeLaitsch, J. R. Pridgen, A. Tytla, M. L. Peach, R. Hu, D. W. Farnsworth, A. K. McMillan, N. Flanagan, J. S. Temme, M. C. Nicklaus, J. C. Gildersleeve, Selective Recognition of Carbohydrate Antigens by Germline Antibodies Isolated from AID Knockout Mice. *J. Am. Chem. Soc.* **144**, 4925–4941 (2022).

6. E. Sterner, M. L. Peach, M. C. Nicklaus, J. C. Gildersleeve, Therapeutic Antibodies to Ganglioside GD2 Evolved from Highly Selective Germline Antibodies. *Cell Reports* **20**, 1681–1691 (2017).

7. M. Babor, T. Kortemme, Multi-constraint computational design suggests that native sequences of germline antibody H3 loops are nearly optimal for conformational flexibility. *Proteins* **75**, 846–858 (2009).

8. D. W. Evans, S. Muller-Loennies, C. L. Brooks, L. Brade, P. Kosma, H. Brade, S. V. Evans, Structural insights into parallel strategies for germline antibody recognition of lipopolysaccharide from Chlamydia. *Glycobiology* **21**, 1049–1059 (2011).

9. L. Scharf, A. P. West, S. A. Sievers, C. Chen, S. Jiang, H. Gao, M. D. Gray, A. T. McGuire, J. F. Scheid, M. C. Nussenzweig, L. Stamatatos, P. J. Bjorkman, Structural basis for germline antibody recognition of HIV-1 immunogens. *eLife* **5**, e13783 (2016).

10. L. Scharf, A. P. West, H. Gao, T. Lee, J. F. Scheid, M. C. Nussenzweig, P. J. Bjorkman, R. Diskin, Structural basis for HIV-1 gp120 recognition by a germ-line version of a broadly neutralizing antibody. *PNAS* **110**, 6049–6054 (2013).

11. S. Hoot, A. T. McGuire, K. W. Cohen, R. K. Strong, L. Hangartner, F. Klein, R. Diskin, J. F. Scheid, D. N. Sather, D. R. Burton, L. Stamatatos, Recombinant HIV Envelope Proteins Fail to Engage Germline Versions of Anti-CD4bs bNAbs.

PLOS Pathogens **9**, e1003106 (2013).

12. D. K. Sethi, A. Agarwal, V. Manivel, K. V. S. Rao, D. M. Salunke, Differential Epitope Positioning within the Germline Antibody Paratope Enhances Promiscuity in the Primary Immune Response. *Immunity* **24**, 429–438 (2006).

13. F. E. Romesberg, B. Spiller, P. G. Schultz, R. C. Stevens, Immunological origins of binding and catalysis in a Diels-Alderase antibody. *Science* **279**, 1929–1933 (1998).

14. G. J. Wedemayer, P. A. Patten, L. H. Wang, P. G. Schultz, R. C. Stevens, Structural Insights into the Evolution of an Antibody Combining Site. *Science, New Series* **276**, 1665–1669 (1997).

15. P. A. Patten, N. S. Gray, P. L. Yang, C. B. Marks, G. J. Wedemayer, J. J. Boniface, R. C. Stevens, P. G. Schultz, The Immunological Evolution of Catalysis. *Science* **271**, 1086–1091 (1996).

16. H. P. Nguyen, N. O. L. Seto, C. R. MacKenzie, L. Brade, P. Kosma, H. Brade, S. V. Evans, Germline antibody recognition of distinct carbohydrate epitopes. *Nat Struct Mol Biol* **10**, 1019–1025 (2003).

17. R. Jimenez, G. Salazar, J. Yin, T. Joo, F. E. Romesberg, Protein dynamics and the immunological evolution of molecular recognition. *Proc Natl Acad Sci U S A* **101**, 3803–3808 (2004).

18. S. Éliás, C. Wrzodek, C. M. Deane, A. C. Tissot, S. Klostermann, F. Ros, Prediction of polyspecificity from antibody sequence data by machine learning. *Front. Bioinform.* **3** (2024).

19. K. A. K. Finton, D. Friend, J. Jaffe, M. Gewe, M. A. Holmes, H. B. Larman, A. Stuart, K. Larimore, P. D. Greenberg, S. J. Elledge, L. Stamatatos, R. K. Strong, Ontogeny of Recognition Specificity and Functionality for the Broadly Neutralizing Anti-HIV Antibody 4E10. *PLOS Pathogens* **10**, e1004403 (2014).

20. X. Xiao, W. Chen, Y. Feng, Z. Zhu, P. Prabakaran, Y. Wang, M.-Y. Zhang, N. S. Longo, D. S. Dimitrov, Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: Implications for evasion of immune responses and design of vaccine immunogens. *Biochemical and Biophysical Research Communications* **390**, 404–409 (2009).

21. V. Manivel, N. C. Sahoo, D. M. Salunke, K. V. S. Rao, Maturation of an Antibody Response Is Governed by Modulations in Flexibility of the Antigen-Combining Site. *Immunity* **13**, 611–620 (2000).

22. H. P. Nguyen, N. O. L. Seto, C. R. MacKenzie, L. Brade, P. Kosma, H. Brade, S.

V. Evans, Germline antibody recognition of distinct carbohydrate epitopes. *Nat Struct Mol Biol* **10**, 1019–1025 (2003).

23. R. Adhikary, W. Yu, M. Oda, R. C. Walker, T. Chen, R. L. Stanfield, I. A. Wilson, J. Zimmermann, F. E. Romesberg, Adaptive Mutations Alter Antibody Structure and Dynamics during Affinity Maturation. *Biochemistry* **54**, 2085–2093 (2015).

24. J. Yin, S. E. Andryski, A. E. Beuscher, R. C. Stevens, P. G. Schultz, Structural evidence for substrate strain in antibody catalysis. *Proc Natl Acad Sci U S A* **100**, 856–861 (2003).

25. L. C. James, P. Roversi, D. S. Tawfik, Antibody multispecificity mediated by conformational diversity. *Science* **299**, 1362–1367 (2003).

26. E. A. Padlan, C. Abergel, J. P. Tipper, Identification of specificity-determining residues in antibodies. *The FASEB Journal* **9**, 133–139 (1995).

27. G. Raghunathan, J. Smart, J. Williams, J. C. Almagro, Antigen-binding site anatomy and somatic mutations in antibodies that recognize different types of antigens. *J Mol Recognit* **25**, 103–113 (2012).

28. T. Khan, D. M. Salunke, Structural elucidation of the mechanistic basis of degeneracy in the primary humoral response. *J Immunol* **188**, 1819–1827 (2012).

29. J. Yin, A. E. Beuscher, S. E. Andryski, R. C. Stevens, P. G. Schultz, Structural Plasticity and the Evolution of Antibody Affinity and Specificity. *Journal of Molecular Biology* **330**, 651–656 (2003).

30. J. Zimmermann, E. L. Oakman, I. F. Thorpe, X. Shi, P. Abbyad, C. L. Brooks, S. G. Boxer, F. E. Romesberg, Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics. *Proceedings of the National Academy of Sciences* **103**, 13722–13727 (2006).

31. G. J. Wedemayer, L. H. Wang, P. A. Patten, P. G. Schultz, R. C. Stevens, Crystal structures of the free and liganded form of an esterolytic catalytic antibody1. *Journal of Molecular Biology* **268**, 390–400 (1997).

32. I. F. Thorpe, C. L. Brooks, Molecular evolution of affinity and flexibility in the immune system. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8821–8826 (2007).

33. T. Li, M. B. Tracka, S. Uddin, J. Casas-Finet, D. J. Jacobs, D. R. Livesay, Rigidity Emerges during Antibody Evolution in Three Distinct Antibody Systems: Evidence from QSFR Analysis of Fab Fragments. *PLOS Computational Biology* **11**, e1004327 (2015).

34. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**,

419–434 (2019).

35. H. Ausserwöger, M. M. Schneider, T. W. Herling, P. Arosio, G. Invernizzi, T. P. J. Knowles, N. Lorenzen, Non-specificity as the sticky problem in therapeutic antibody development. *Nat Rev Chem* **6**, 844–861 (2022).

36. N. E. Kaleli, M. Karadag, S. Kalyoncu, Phage display derived therapeutic antibodies have enriched aliphatic content: Insights for developability issues. *Proteins* **87**, 607–618 (2019).

37. M. I. J. Raybould, C. Marks, K. Krawczyk, B. Taddese, J. Nowak, A. P. Lewis, A. Bujotzek, J. Shi, C. M. Deane, Five computational developability guidelines for therapeutic antibody profiling. *Proc Natl Acad Sci U S A* **116**, 4025–4030 (2019).

38. Y. Zhang, C. Duan, Q. Li, Y. Bai, B. Dong, Y. Tang, M. He, C. Hao, K. Wen, J. Shen, Z. Wang, Fluorescence polarization immunoassay based on fragmentary hapten for rapid and sensitive screening of polymyxins in human serum. *Sensors and Actuators B: Chemical* **370**, 132404 (2022).

39. Z. Wang, R. C. Beier, Y. Sheng, S. Zhang, W. Jiang, Z. Wang, J. Wang, J. Shen, Monoclonal antibodies with group specificity toward sulfonamides: selection of hapten and antibody selectivity. *Anal Bioanal Chem* **405**, 4027–4037 (2013).

40. H. Li, B. Dong, L. Dou, W. Yu, X. Yu, K. Wen, Y. Ke, J. Shen, Z. Wang, Fluorescent lateral flow immunoassay for highly sensitive detection of eight anticoagulant rodenticides based on cadmium-free quantum dot-encapsulated nanospheres. *Sensors and Actuators B: Chemical* **324**, 128771 (2020).

41. Y. Zhang, J. Mi, W. Wu, J. Fei, B. Lv, X. Yu, K. Wen, J. Shen, Z. Wang, Investigation of Antibody Tolerance in Methanol for Analytical Purposes: Methanol Effect Patterns and Molecular Mechanisms. *Advanced Science* **11**, 2402050 (2024).

42. J. Dunbar, K. Krawczyk, J. Leem, T. Baker, A. Fuchs, G. Georges, J. Shi, C. M. Deane, SAbDab: the structural antibody database. *Nucleic Acids Research* **42**, D1140–D1146 (2014).

43. Y. Barrios, P. Jirholt, M. Ohlin, Length of the antibody heavy chain complementarity determining region 3 as a specificity-determining factor. *Journal of Molecular Recognition* **17**, 332–338 (2004).

44. A. V. J. Collis, A. P. Brouwer, A. C. R. Martin, Analysis of the Antigen Combining Site: Correlations Between Length and Sequence Composition of the Hypervariable Loops and the Nature of the Antigen. *Journal of Molecular Biology* **325**, 337–354 (2003).

45. M. L. Chiu, G. L. Gilliland, Engineering antibody therapeutics. *Curr Opin Struct*

Biol **38**, 163–173 (2016).

46. J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 1–3 (2024).

47. D. Kuroda, H. Shirai, M. Kobori, H. Nakamura, Structural classification of CDR-H3 revisited: a lesson in antibody modeling. *Proteins* **73**, 608–620 (2008).

48. B. North, A. Lehmann, R. L. Dunbrack, A new clustering of antibody CDR loop conformations. *J Mol Biol* **406**, 228–256 (2011).

49. Y. Tsuchiya, K. Mizuguchi, The diversity of H3 loops determines the antigen-binding tendencies of antibody CDR loops. *Protein Sci* **25**, 815–825 (2016).

50. K.-M. A. Dam, H. B. Gristick, Y. E. Li, Z. Yang, P. N. P. Gnanapragasam, A. P. West, M. S. Seaman, P. J. Bjorkman, Mapping essential somatic hypermutations in a CD4-binding site bNAb informs HIV-1 vaccine design. *Cell Reports* **44** (2025).

51. S. Birtalan, Y. Zhang, F. A. Fellouse, L. Shao, G. Schaefer, S. S. Sidhu, The Intrinsic Contributions of Tyrosine, Serine, Glycine and Arginine to the Affinity and Specificity of Antibodies. *Journal of Molecular Biology* **377**, 1518–1528 (2008).

52. L. Lo Conte, C. Chothia, J. Janin, The atomic structure of protein-protein recognition sites. *J Mol Biol* **285**, 2177–2198 (1999).

53. I. S. Mian, A. R. Bradwell, A. J. Olson, Structure, function and properties of antibody binding sites. *J Mol Biol* **217**, 133–151 (1991).

54. C. R. Rupakheti, B. Roux, F. Dehez, C. Chipot, Modeling induction phenomena in amino acid cation– $$\pi $$ interactions. *Theor Chem Acc* **137**, 1–6 (2018).

55. R. L. Kelly, D. Le, J. Zhao, K. D. Wittrup, Reduction of Nonspecificity Motifs in Synthetic Antibody Libraries. *J Mol Biol* **430**, 119–130 (2018).

56. C.-H. Luan, T. M. Parker, D. C. Gowda, D. W. Urry, Hydrophobicity of amino acid residues: Differential scanning calorimetry and synthesis of the aromatic analogues of the polypentapeptide of elastin. *Biopolymers* **32**, 1251–1261

(1992).

57. B. S, F. Rd, S. Ss, The functional capacity of the natural amino acids for molecular recognition. *Molecular bioSystems* **6** (2010).

58. D. Jain, D. M. Salunke, Antibody specificity and promiscuity. *Biochem J* **476**, 433–447 (2019).

59. Z.-H. Zhou, Y. Zhang, Y.-F. Hu, L. M. Wahl, J. O. Cisar, A. L. Notkins, The broad antibacterial activity of the natural antibody repertoire is due to polyreactive antibodies. *Cell Host Microbe* **1**, 51–61 (2007).

60. H. Mouquet, M. C. Nussenzweig, Polyreactive antibodies in adaptive immune responses to viruses. *Cell. Mol. Life Sci.* **69**, 1435–1445 (2012).

61. A. G. Schmidt, H. Xu, A. R. Khan, T. O'Donnell, S. Khurana, L. R. King, J. Manischewitz, H. Golding, P. Suphaphiphat, A. Carfi, E. C. Settembre, P. R. Dormitzer, T. B. Kepler, R. Zhang, M. A. Moody, B. F. Haynes, H.-X. Liao, D. E. Shaw, S. C. Harrison, Preconfiguration of the antigen-binding site during affinity maturation of a broadly neutralizing influenza virus antibody. *Proceedings of the National Academy of Sciences* **110**, 264–269 (2013).

62. H. Ausserwöger, G. Krainer, T. J. Welsh, N. Thorsteinson, E. de Csilléry, T. Sneideris, M. M. Schneider, T. Egebjerg, G. Invernizzi, T. W. Herling, N. Lorenzen, T. P. J. Knowles, Surface patches induce nonspecific binding and phase separation of antibodies. *Proceedings of the National Academy of Sciences* **120**, e2210332120 (2023).

63. M. Sangesland, A. S. Yousif, L. Ronsard, S. W. Kazer, A. L. Zhu, G. J. Gatter, M. R. Hayward, R. M. Barnes, M. Quirindongo-Crespo, D. Rohrer, N. Lonberg, D. Kwon, A. K. Shalek, D. Lingwood, A Single Human VH-gene Allows for a Broad-Spectrum Antibody Response Targeting Bacterial Lipopolysaccharides in the Blood. *Cell Rep* **32**, 108065 (2020).

64. J. Jardine, J.-P. Julien, S. Menis, T. Ota, O. Kalyuzhniy, A. McGuire, D. Sok, P.-S. Huang, S. MacPherson, M. Jones, T. Nieusma, J. Mathison, D. Baker, A. B. Ward, D. R. Burton, L. Stamatatos, D. Nemazee, I. A. Wilson, W. R. Schief, Rational HIV Immunogen Design to Target Specific Germline B Cell Receptors. *Science* **340**, 711–716 (2013).

65. X. Brochet, M.-P. Lefranc, V. Giudicelli, IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Research* **36**, W503–W508 (2008).

66. M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A.

C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, eabj8754 (2021).

67. W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

68. P. Labute, LowModeMD—Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops. *J. Chem. Inf. Model.* **50**, 792–800 (2010).

69. Protonate 3D: Assignment of Macromolecular Protonation State and Geometry. https://server.ccl.net/cca/documents/proton/proton.htm.

Figures Legends

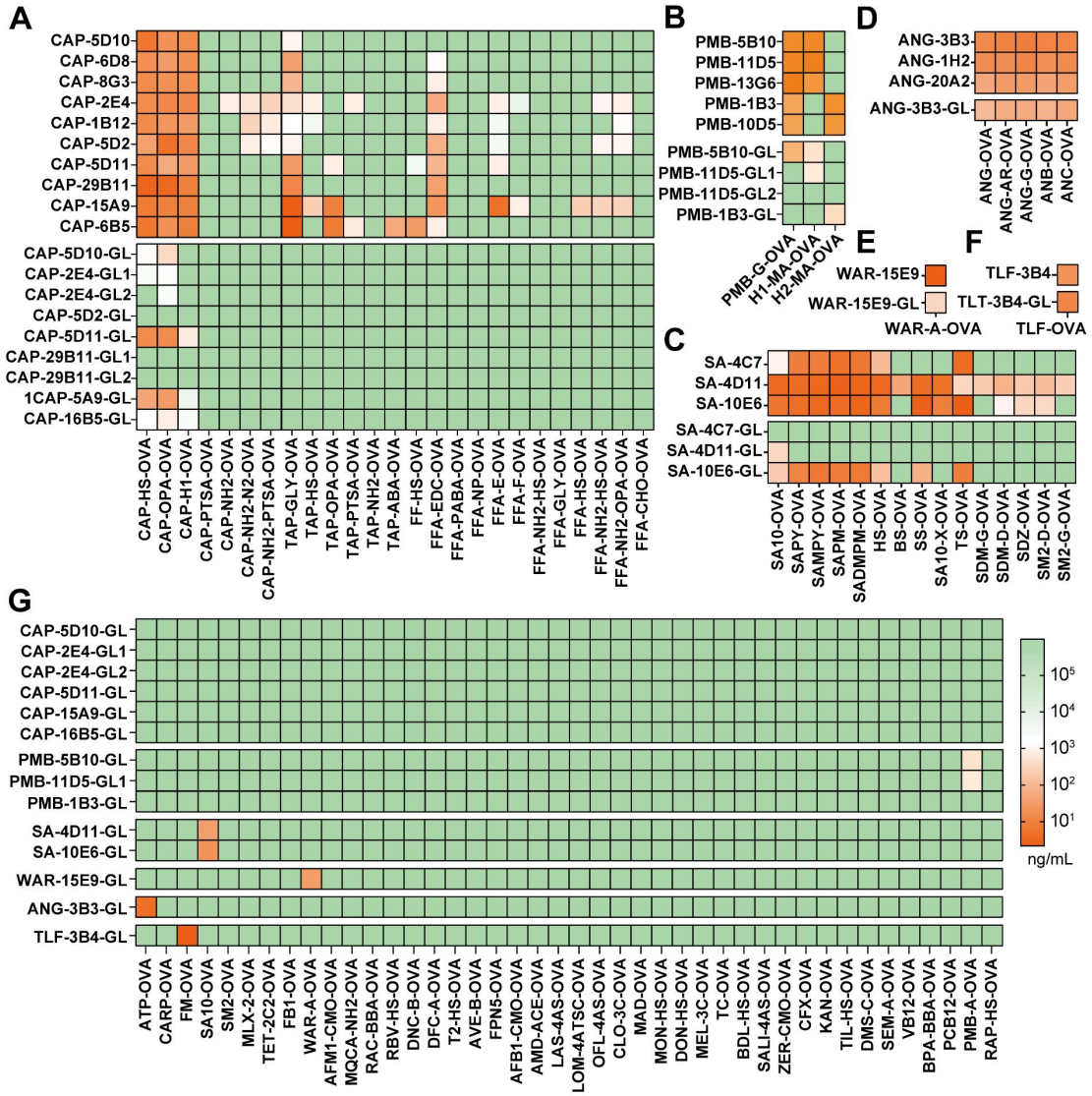
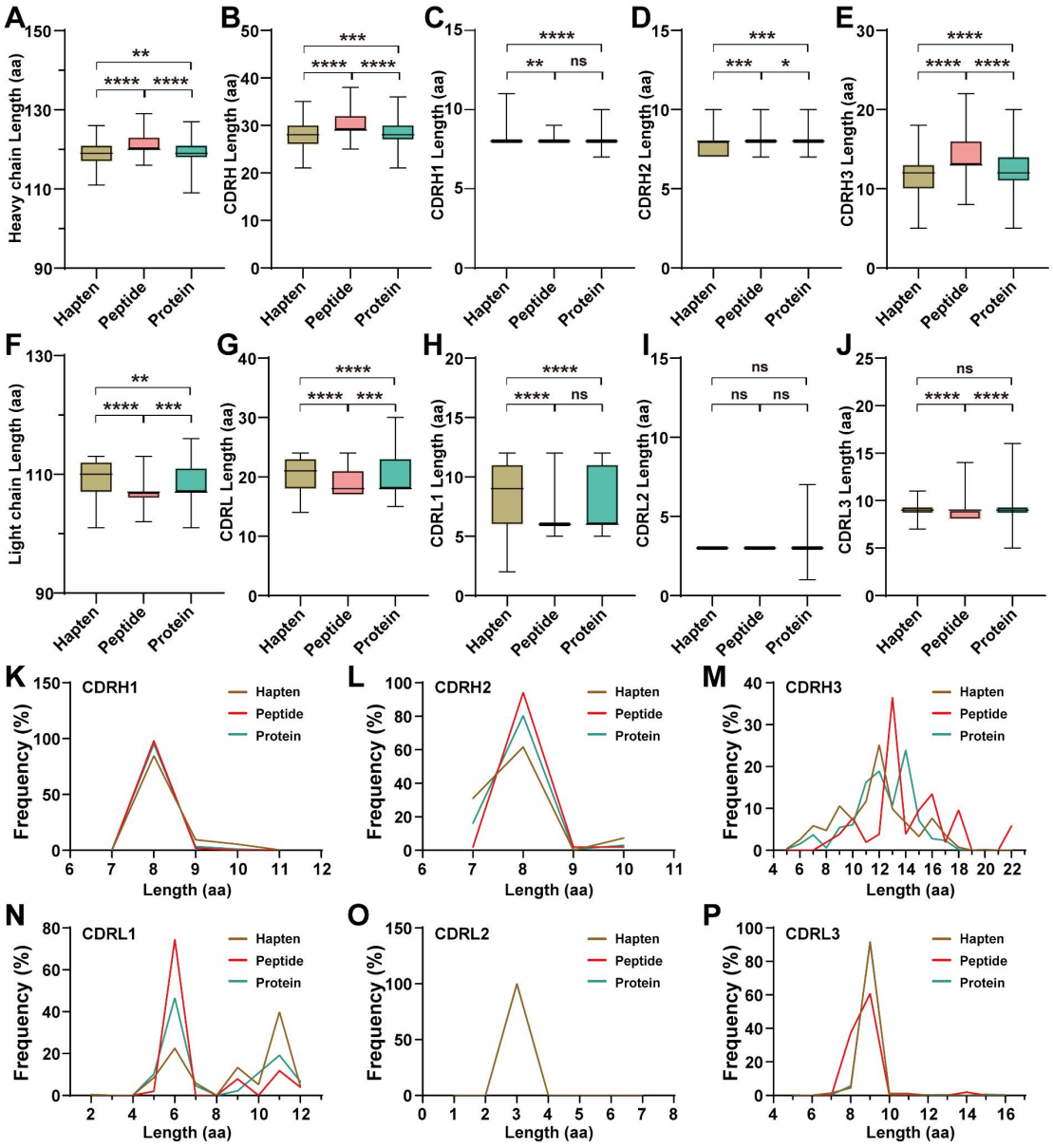


Figure 1. The binding specificity identification of germline antibodies and their corresponding mAbs. (A–F) The binding affinities of 19 germline antibodies and their corresponding mature monoclonal antibodies (mAbs) targeting chloramphenicol (CAP), polymyxin B (PMB), sulfadiazine (SA), analginum (ANG), warfarin (WAR), and tolfenamic acid (TLF) were evaluated against panels of structurally related hapten-OVA conjugates (hapten structure shown in Figure S2–S5). As binding affinity increases—i.e., as the ng/mL value decreases—the color shifts from green to red. (G) The polyspecificity of 14 hapten-binding germline antibodies was evaluated against a

1101 diverse panel of 42 hapten-OVA conjugates (hapten structure shown in Figure S6).

1102



1103

1104 **Figure 2.** The sequence length analysis of germline antibodies targeting haptens,

1105 peptides and proteins. Length analysis of (A) VH, (B) CDRH, (C) CDRH1, (D)

1106 CDRH2, (E) CDRH3, (F) VL, (G) CDRL, (H) CDRL1, (I) CDRL2, and (J) CDRL3.

1107 Length distribution analysis of (K) CDRH1, (K) CDRH2, (M) CDRH3, (N) CDRL1,

1108 (O) CDRL2, and (P) CDRL3.

1109

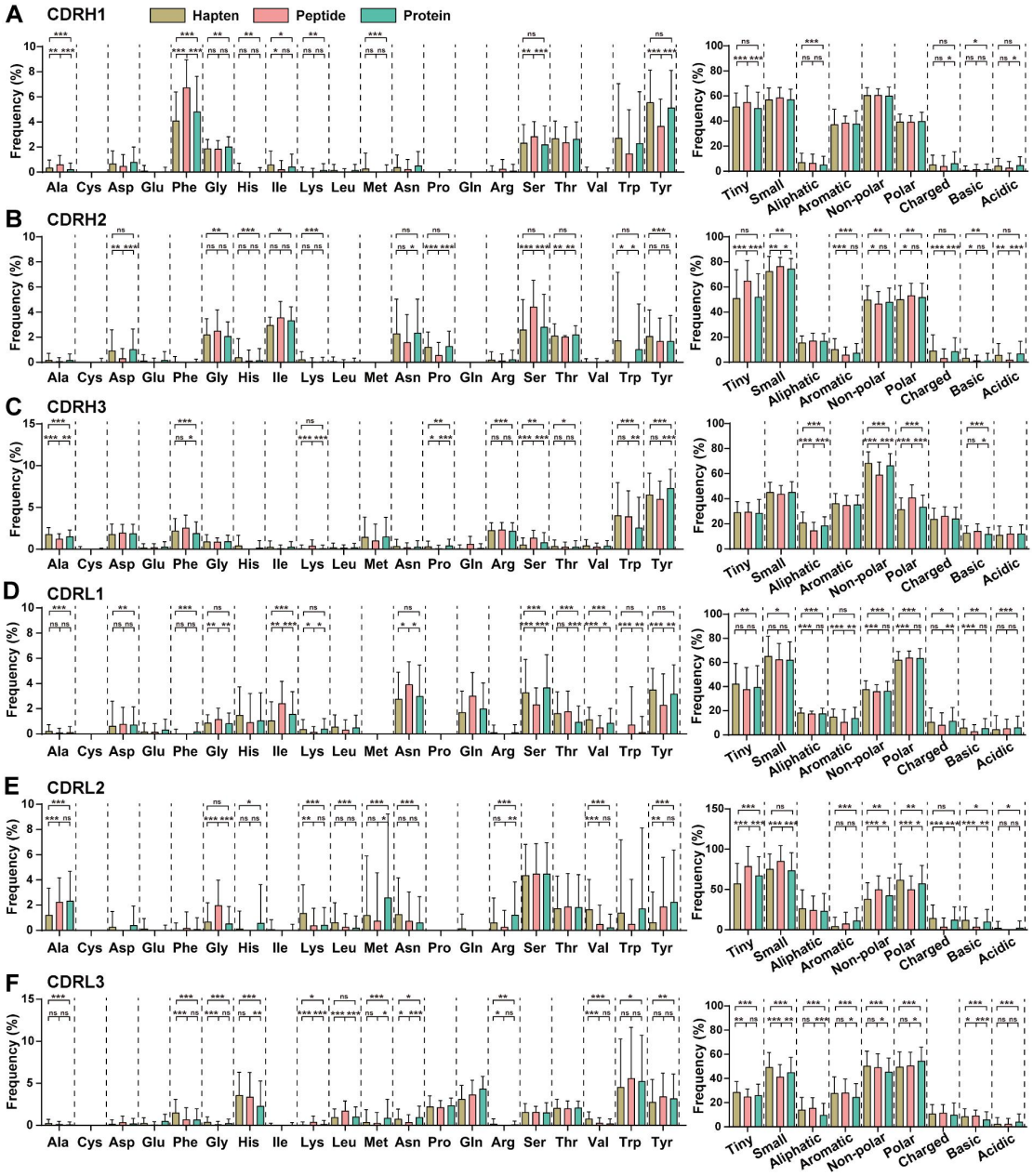


Figure 3. CDR residue composition analysis of germline antibodies targeting haptens, peptides and proteins. Analysis of 20 residue composition and categorical statistical analysis of (A) CDRH1, (B) CDRH2, (C) CDRH3, (D) CDRL1, (E) CDRL2, and (F) CDRL3. Residue classification shown in Table S4.

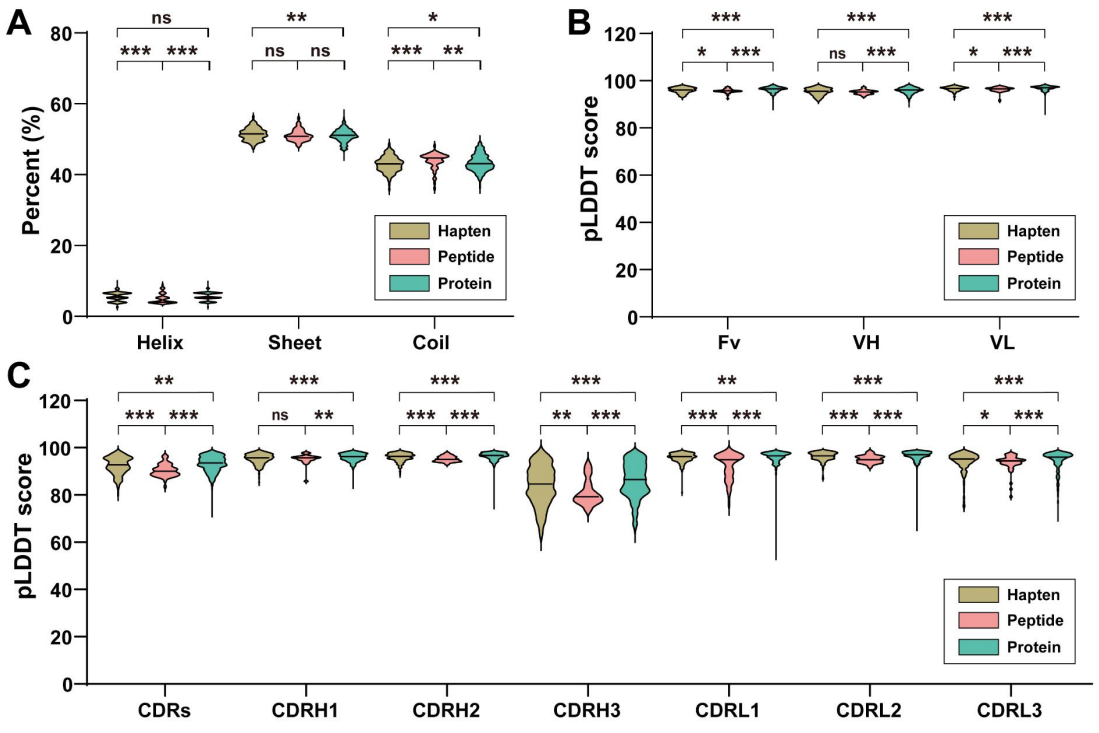


Figure 4. Structural flexible analysis of germline antibodies targeting haptens, peptides and proteins. (A) The secondary-structure occupancies analysis of germline antibodies calculated by DSSP. α-helices, 310-helices, and π-helices were categorized as “Helix”; extended β-strands and β-bridges as “Sheet”; and random coil as “Coil.” (B–C) The predicted local distance difference test (pLDDT) score analysis of germline antibodies predicted by AlphaFold3.

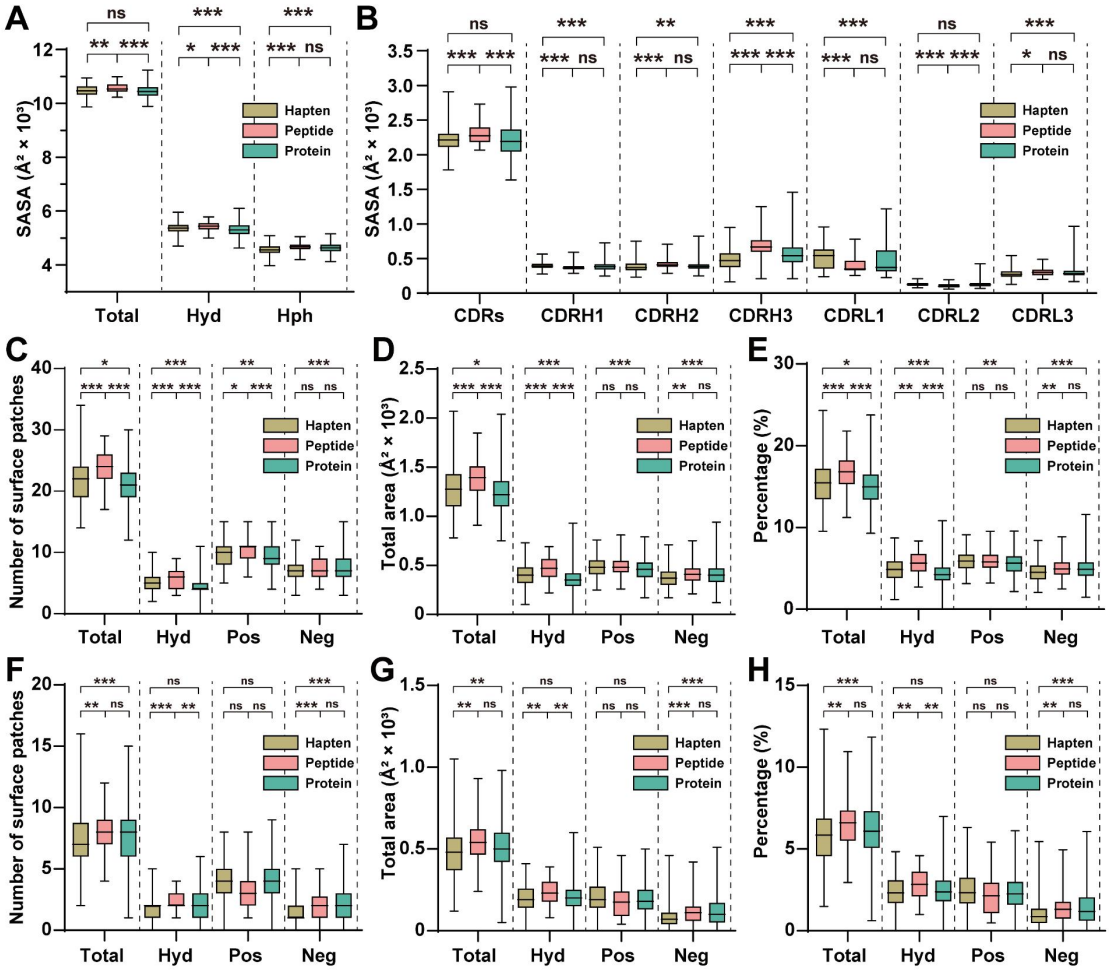


Figure 5. Surface properties analysis of germline antibodies targeting haptens, peptides and proteins. (A) The total solvent-accessible surface area (SASA), hydrophobic SASA and hydrophilic SASA analysis of germline antibodies. (B) SASA analysis of total CDRs and individual CDR regions in germline antibodies. (C–E) The number, total area and percentage of surface patches (SPs) in Fv region of germline antibodies. (F–H) The number, total area and percentage of SPs located in the vicinity of the CDRs of germline antibodies. Total: total SPs; Hyd: hydrophobic SPs; Pos: positive charged SPs; Neg: negative charged SPs.

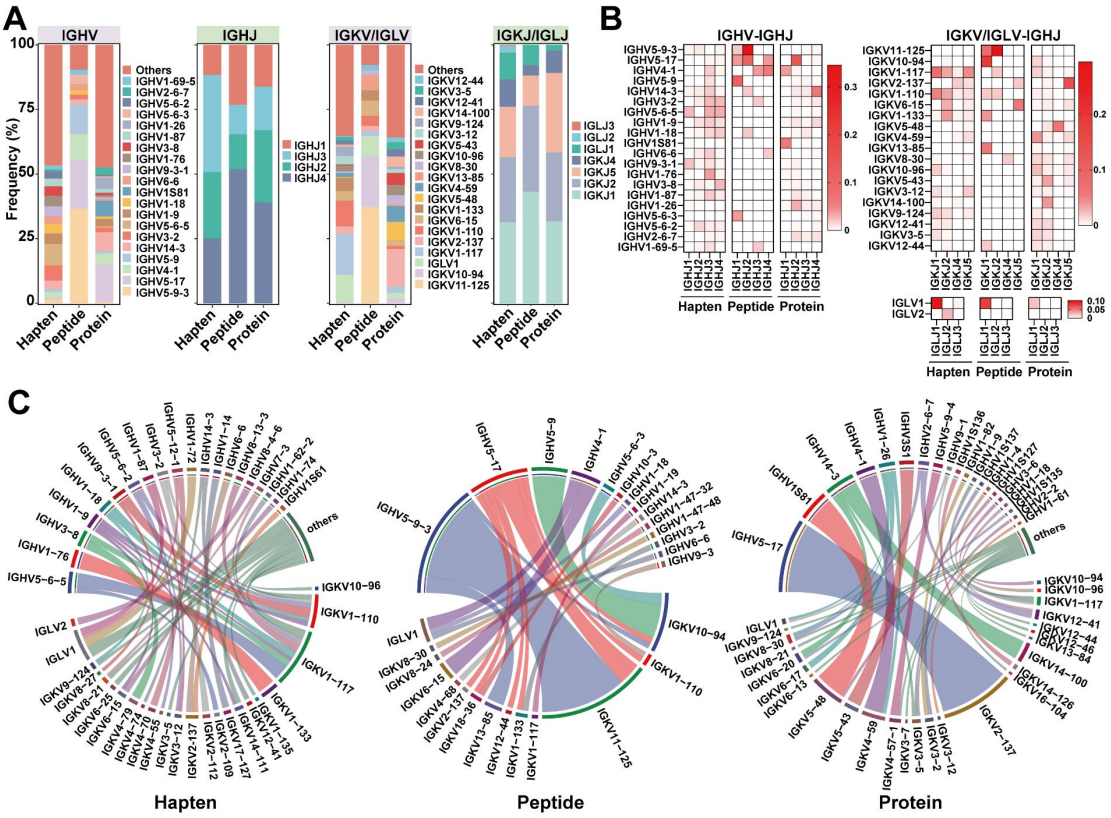


Figure 6. Germline genes analysis of germline antibodies targeting haptens, peptides and proteins. (A) The V and J germline gene frequency analysis of germline antibodies. (B) The V-J germline gene combination frequency analysis of germline antibodies. (C) The IGHV-IGKV/IGLV germline gene paired frequency analysis of germline antibodies.