Molecular deconvolution of the monoclonal antibodies that comprise the polyclonal serum response

Yariv Wine[a,b,1], Daniel R. Boutz[b,c,1], Jason J. Lavinder[a,b,1], Aleksandr E. Miklos[b,d], Randall A. Hughes[b,d], Kam Hon Hoi[e], Sang Taek Jung[a,b,f], Andrew P. Horton[c,e], Ellen M. Murrin[a], Andrew D. Ellington[b,c,d,g], Edward M. Marcotte[b,c,g,2], and George Georgiou[a,b,e,h,2]

[a]Department of Chemical Engineering, [b]Institute for Cellular and Molecular Biology, [c]Center for Systems and Synthetic Biology, [d]Applied Research Laboratories, [e]Department of Biomedical Engineering, [g]Department of Chemistry and Biochemistry, and [h]Section of Molecular Genetics and Microbiology, University of Texas at Austin, Austin, TX 78712-1062; and [f]Department of Bio and Nano Chemistry, Kookmin University, Seoul 136-702, Korea

Edited by Allan Bradley, Wellcome Trust Sanger Institute, Hinxton, United Kingdom, and accepted by the Editorial Board January 9, 2013 (received for review August 8, 2012)

We have developed and validated a methodology for determining the antibody composition of the polyclonal serum response after immunization. Pepsin-digested serum IgGs were subjected to standard antigen-affinity chromatography, and resulting elution, wash, and flow-through fractions were analyzed by bottom-up, liquid chromatography–high-resolution tandem mass spectrometry. Identification of individual monoclonal antibodies required the generation of a database of IgG variable gene (V-gene) sequences constructed by NextGen sequencing of mature B cells. Antibody V-gene sequences are characterized by short complementarity determining regions (CDRs) of high diversity adjacent to framework regions shared across thousands of IgGs, greatly complicating the identification of antigen-specific IgGs from proteomically observed peptides. By mapping peptides marking unique V_H CDRH3 sequences, we identified a set of V-genes heavily enriched in the affinity chromatography elution, constituting the serum polyclonal response. After booster immunization in a rabbit, we find that the antigen-specific serum immune response is oligoclonal, comprising antibodies encoding 34 different CDRH3s that group into 30 distinct antibody V_H clonotypes. Of these 34 CDRH3s, 12 account for ∼60% of the antigen-specific CDRH3 peptide mass spectral counts. For comparison, antibodies with 18 different CDRH3s (12 clonotypes) were represented in the antigen-specific IgG fraction from an unimmunized rabbit that fortuitously displayed a moderate titer for BSA. Proteomically identified antibodies were synthesized and shown to display subnanomolar affinities. The ability to deconvolute the polyclonal serum response is likely to be of key importance for analyzing antibody responses after vaccination and for more completely understanding adaptive immune responses in health and disease.

antibody proteomics | antibody repertoire | serum immunoprofiling | B-cell response | humoral response

The first Nobel Prize in Medicine was awarded to Emil von Behring, who in collaboration with Kitasato Shibasaburo and Paul Ehrlich discovered serum antitoxins (1, 2). Remarkably, after more than 100 y of intense research in immunology, little is known about the clonality, relative concentrations, and binding properties of the monoclonal antibodies that constitute the antigen-specific Ig pool in serum.

At steady state, circulating antibodies are produced by terminally differentiated B lymphocytes (plasma cells) within the bone marrow, and thus cannot be accessed in living individuals (3). Although recent single B-cell cloning methods (4, 5) have led to the identification of peripheral antigen-specific B memory and/or antibody-secreting cells (plasmablasts), it is generally unknown whether the Igs encoded by peripheral blood B cells correspond to the antibodies present in circulation and especially whether they are present at physiologically relevant levels (i.e., at serum concentrations above K_D corresponding to >1 μg/mL for an average affinity of individual antibodies of 5 nM).

The proteomic deconvolution of serum Igs presents two major technical challenges: first, antibody genes in antigen stimulated B-lymphocytes are not simply encoded in the germline but are extensively diversified by somatic recombination, revision, and/or mutation. Therefore, the sequence database required for the interpretation of mass spectra is not available a priori (6, 7) and is completely different for each individual. Second, the antigen-specific antibody pool comprises a wide variety of Igs that display very high levels of amino acid identity within the framework regions. As a result, standard approaches for proteomic analysis by MS are confounded by this exceptionally high rate of identical sequence shared among Ig-derived peptides, which greatly complicates the task of confidently identifying individual variable (V) genes through peptide mapping. Advancements in sequencing and MS technologies have shown some success against these challenges. MS-based de novo sequencing approaches have been used for the identification of purified monoclonal antibodies (8). More recently the identification of a limited subset of antigen-specific antibodies in serum after very stringent enrichment to reduce the complexity of the antigen-specific polyclonal antibody pool to a limited set of Igs from humans and animals was reported (9–12). However, because of the inherent difficulties associated with the proteomic analysis of complex mixtures of antibodies, these studies had focused on the identification of only a small subset of the antigen-specific serum IgGs present in a fraction isolated after stringent affinity chromatography.

In contrast, complete understanding of how B-cell differentiation ultimately shapes humoral immunity requires addressing the more difficult problem of how to deconvolute the entire repertoire of antigen-specific antibodies in serum or in other secretions.

Here we describe the proteomic deconvolution of the serum-derived antigen-specific polyclonal antibody pool by combining NextGen sequencing of the immunoglobulin heavy chain variable region (V_H gene) repertoire with liquid chromatography–high-resolution tandem mass spectrometry (LC-MS/MS) (Fig. 1). Proteomic identifications of unique V_H-derived peptides (overwhelmingly from the CDR3 region of the V_H sequences) were used to determine the V_H repertoire of circulating antigen-specific antibodies, and identified V_H genes were shown to encode antibodies with subnanomolar antigen affinity.

Author contributions: Y.W., D.R.B., J.J.L., E. M. Marcotte, and G.G. designed research; Y.W., D.R.B., J.J.L., A.E.M., R.A.H., S.T.J., and E. M. Murrin performed research; Y.W., D.R.B., J.J.L., K.H.H., A.P.H., E. M. Marcotte, and G.G. analyzed data; and Y.W., D.R.B., J.J.L., A.D.E., E. M. Marcotte, and G.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. A.B. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

[1]Y.W., D.R.B., and J.J.L. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: marcotte@icmb.utexas.edu or gg@che.utexas.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1213737110/-/DCSupplemental.

IMMUNOLOGY

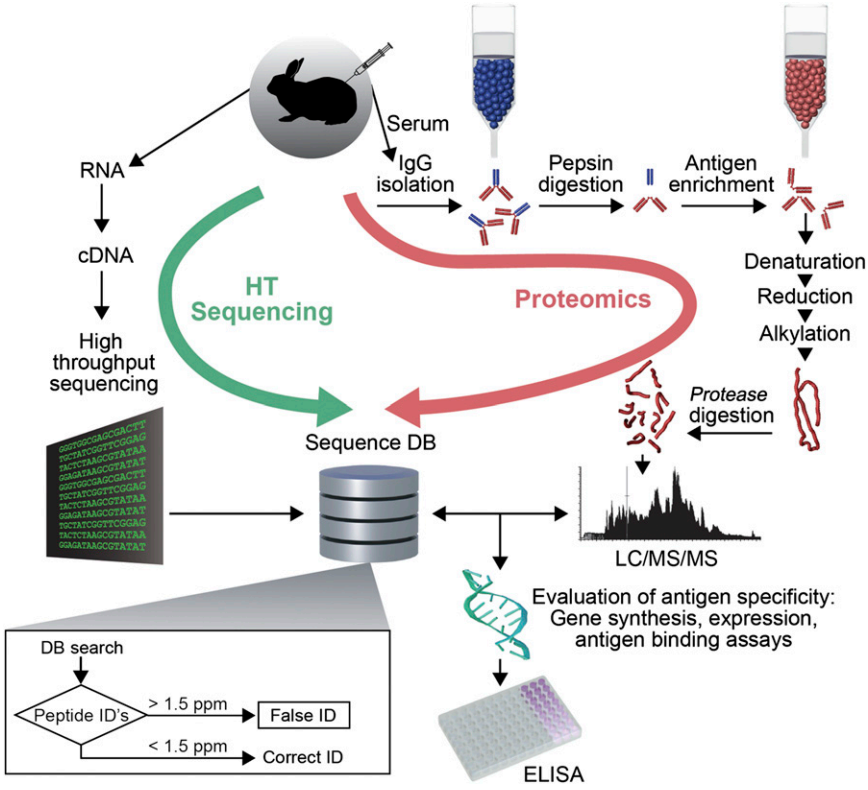

Fig. 1. Schematic of the workflow for serum Ig deconvolution. (*Left*) V gene repertoire sequencing pipeline: total RNA from desired B-cell subpopulations is reverse transcribed and amplified by 5′ RACE with IgG-specific (V_H) or Ig_κ/Ig_λ-specific (V_L) primers and sequenced by Roche 454 sequencing. Reads are processed bioinformatically to obtain a database of unique V genes and their relative transcript abundances. The V gene database is used to interpret the MS spectra. (*Right*) $F(ab)_2$ purification and proteomic pipeline: $F(ab)_2$ fragments from IgG are prepared and subjected to antigen-affinity chromatography. Proteins in the eluent, flow-through, and wash buffer are denatured, alkylated, proteolyzed, and resolved by high-resolution LC-MS/MS. Full-length V genes containing the identified *i*CDR3 peptides are then determined from the repertoire database. (*Lower*) Antibody production and validation: synthetic V_H genes are assembled into an scFv library using the V_L cDNA, then antigen-specific antibodies are isolated by two to three rounds of phage panning and characterized for antigen affinity.

Results

Overview of Experimental Approach. The first step for serum antibody deconvolution is the high-throughput sequencing of B lymphocyte cDNAs to generate a database of class-switched antibody V_H sequences in a particular individual (Fig. 1). In parallel, full-length IgGs are purified by protein A affinity chromatography and treated with pepsin to prepare $F(ab)_2$ fragments. Antigen-specific $F(ab)_2$ fragments are then isolated by affinity chromatography on immobilized antigen. Bound antibodies are eluted using standard immunoaffinity chromatography conditions (elution with pH 2.7 buffer), and the flow-through and the wash buffer from the column are also collected to identify weakly antigen-binding antibodies. The $F(ab)_2$ fragments are proteolytically digested, and the resulting peptides are resolved by high-resolution LC-MS/MS analysis on an Orbitrap Velos (Thermo Scientific). To ensure comprehensive coverage of the antibody repertoire, we performed three replicate injections per sample, using LC elution with a long, shallow gradient of 5–40% acetonitrile over 245 min, yielding an average of >100,000 MS2 fragmentation spectra per sample. Peptides found to be enriched >10-fold in the elution fraction compared with wash and flow-through were considered as corresponding to IgGs exhibiting a high degree of antigen specificity.

By far the highest diversity in antibodies occurs within the CDR3 region of the V_H (CDRH3), which is overwhelmingly responsible for antigen recognition (13) (Fig. 2*A*). Composed of the V(D)J join with its inherent junctional diversity, the CDRH3 specifies the V_H clonotype. The V_H clonotype is defined as the group of V_H sequences that share germ-line V and J segments, have identical CDRH3 lengths, and exhibit greater than 80% amino acid identity in the CDRH3 sequences (14, 15). The V_H clonotype is an important immunological concept because it accounts for antibodies that likely originate from a single B-cell lineage and may provide insight on the evolution of the antigen-specific response of that lineage. Therefore, we focused here on the high-confidence identification of CDRH3 peptides and of the corresponding clonotypes. As shown later, unique peptides derived from non-CDRH3 regions of Igs (i.e., from peptides containing all or portions of framework 1–3, CDR1, or CDR2 sequences) and corresponding to a single V gene in the database make only a minor contribution to the determination of the repertoire in immunized rabbits (contributing to <15% of the total identified clonotypes). Once the CDRH3s have been identified, the full-length sequences of corresponding V_H genes are determined from the NextGen V_H DNA database.

To generate antibodies, proteomically identified V_H sequences must be paired with the V_L genes. We therefore synthesized V_H domain DNAs and used these to construct phage displayed scFv libraries comprising the V_L cDNA repertoire. Functional scFv antibodies were isolated after two to three rounds of phage panning. Putative pairings of V_H and V_L genes were then expressed recombinantly as full-size IgGs in mammalian cells and isolated to characterize their antigen affinity and functionality.

V Gene Repertoire Analysis. Preliminary studies in mice revealed that the methodology described in Fig. 1 was able to successfully identify abundant, antigen-specific IgGs in the serum of immunized mice. However, the small amount of serum that is obtained from mice at killing (0.8–1.5 mL) resulted in low counts of V-gene informative peptides and poor MS identification statistics. Larger amounts of serum can easily be obtained from humans, as well as other comparatively larger animals such as rabbits, an animal model that has been used extensively for immunological studies for nearly a century. For the present study, a New Zealand white rabbit (*Oryctolagus cuniculus*) was immunized with *Concholepas concholepas* hemocyanin (CCH) in complete Freund's adjuvant (CFA), boosted with antigen in incomplete FA and killed 1 wk after the final boost (CCH rabbit). Additionally, to further validate our approach we performed serum IgG deconvolution on an unimmunized rabbit that, surprisingly, was found to exhibit a titer toward BSA (BSA rabbit; this observation occurred fortuitously because BSA was used as the generic blocking agent in our ELISA protocol). Unlike the BSA rabbit, the animal immunized with CCH did not exhibit any titer toward BSA. We prepared RNA samples from total peripheral B

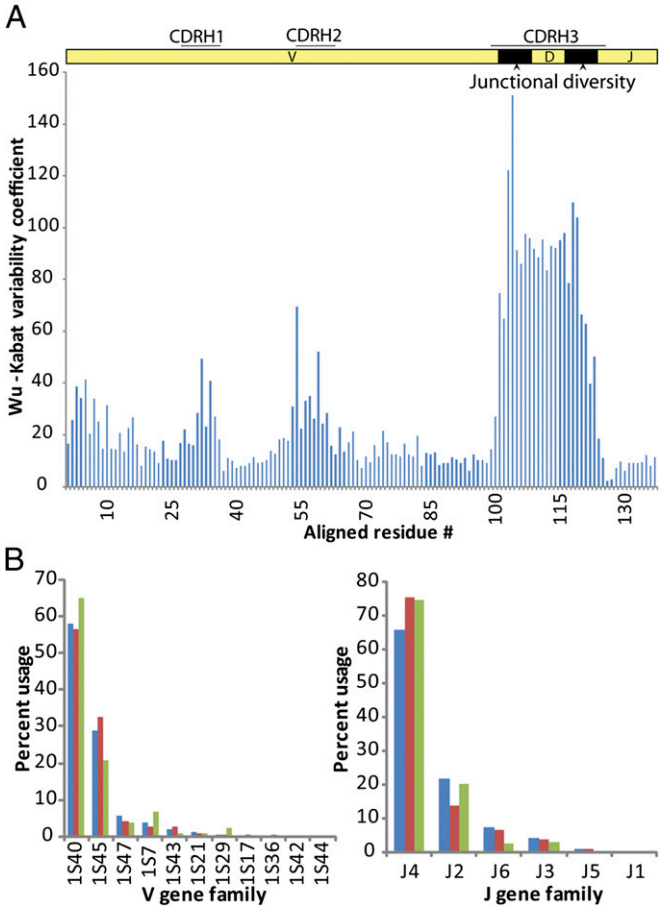

Fig. 2. Immunoglobulin heavy chain variable region (V_H gene) and circulating antibody repertoire characteristics. (*A*) Wu-Kabat variability plot representing the variability of the V_H genes is shown on a residue-by-residue basis; (*B*) bar graphs representing V_H germ-line family use (*Left*) and J_H germ-line family use (*Right*) in the V_H DNA repertoire determined by 454 sequencing of cDNA from $CD138^+$ bone marrow plasma cells (n = 4729 V_H genes, blue bars), peripheral B cells (n = 2788 V_H genes, red bars), or from antibody proteins identified by proteomic analysis of the serum affinity purified IgGs (n = 334, green bars) for the CCH rabbit.

cells (PBCs), total bone marrow cells, and $CD138^+$ bone marrow plasma cells (BM-PCs) isolated by magnetic sorting. First-strand cDNAs were generated using an oligo(dT) primer, and double-stranded products were amplified via 5′ RACE (16) using primers complementary to rabbit IgG CH1 (*SI Appendix*, Table S1). DNA amplification by 5′ RACE was preferable to using published FR1 and J region-specific primers designed for combinatorial library construction (17) because existing 5′ primer sets have not been validated for quantitative V gene amplification. In contrast, 5′ RACE circumvents the need for V gene-specific primers and provides a more accurate representation of the repertoire by avoiding biases introduced by the selection of PCR primer sets. V gene cDNA was sequenced using Roche 454 GS FLX Titanium (*SI Appendix*, Table S2). Germ-line V and J use were determined (Fig. 2*B*) as previously described (18). To reduce the impact of sequencing errors, the V_H and V_L protein sequence databases were compiled using sequences that occurred at $n \geq 2$ reads.

Consistent with earlier reports of limited germ-line V gene diversity in rabbits (19, 20), we found that 89% of the V_H genes in the IgG repertoire (CCH rabbit) were derived from only two germ-line V genes (1S40 and 1S45), and an overwhelming 75% contained the IGHJ4 segment. V gene and J gene use was highly similar in BM-PCs and PBCs. In the BSA rabbit, 86% of the V_H sequences were derived from 1S44, 1S45, and 1S40, whereas 55% contained the IGHJ4 and 28% contained the IGHJ2 J-segment (*SI Appendix*, Fig. S1). CDRH3 lengths in the class-switched IgG repertoire of an immunized rabbit were longer than in immunized mice (*SI Appendix*, Fig. S2) (21). The rabbit class-switched V genes displayed a bimodal distribution of amino acid substitutions relative to the germline, with one peak centered at 9-aa substitutions, slightly higher than in the mouse, and a second peak at 24- to 25-aa substitutions (*SI Appendix*, Fig. S3). Because gene conversion is an important mechanism for Ig diversification in the rabbit (22), it cannot be ascertained whether the highly mutated V gene population resulted from activation-induced deaminase (AID) mutagenesis or from recombination events with V pseudogenes.

MS Proteomic Detection of Ig-Identifying Peptides. IgG antibodies from a 2.5 mL rabbit serum sample were purified by Protein A affinity chromatography. $F(ab)_2$ fragments were prepared by pepsin digestion and purified by affinity chromatography on immobilized CCH or BSA yielding ∼0.7 and 0.2 mg of high-purity $F(ab)_2$ protein, respectively (>95% pure as determined by SDS/PAGE; typical results shown in *SI Appendix*, Fig. S4). The protein in the flow-through and wash fractions was also collected for LC-MS/MS analysis. The individual fractions were denatured, reduced, and alkylated with iodoacetamide to modify Cys residues before digestion with trypsin (*SI Appendix*, *SI Materials and Methods*). *In silico* analysis of the V gene database showed that digestion with trypsin should generate peptide fragments with enough coverage of the CDRH3 region and of lengths appropriate for MS detection to uniquely identify 91.4% of the putative antibody clones (*SI Appendix*, Fig. S5). Cleavage by chymotrypsin added less than 3% in observable V_H clone coverage, and a similar effect was calculated for other cleavage-specific proteases. This slight potential gain in coverage was diminished when potential missed cleavages were considered.

Peptides were analyzed by nanospray LC-MS/MS, and collected spectra were searched by Sequest against an in-house NextGen rabbit V_H+V_L sequence database (containing sequences with $n \geq 2$ reads) concatenated with the rabbit full protein-coding sequence database (OryCun2) and MaxQuant contaminants database (23, 24). Postsearch processing by the Percolator algorithm (25) generated a dataset of peptide-spectrum matches (PSMs) with an expected false-discovery rate <1%. False identifications were further controlled at the peptide level by accepting only those peptide identifications for which all PSMs exhibited an average deviation from the expected peptide mass of ≤1.5 ppm. Spectra were manually checked for consistency with the identified sequences, including the presence of modifications (static carbamidomethyl modification of Cys residues and dynamic oxidation of methionine to methionine-sulfoxide) and signature motifs such as the IGHJ-derived sequence, which gave a characteristic spectral pattern (Table 1).

Peptides identified by MS analysis were classified according to their cooccurrence with CDRH3 sequences in the V gene DNA sequence database. Peptides that uniquely identified specific CDRH3s were defined as *i*nformative of CDRH3 (*i*CDRH3) peptides; in contrast, peptides mapping to multiple CDRH3 sequences were defined as *noni*nformative (*ni*CDRH3). *i*CDRH3 peptides were considered antigen-specific if the frequency of spectral counts in the affinity chromatography elution fraction was 10-fold greater than in the combined wash and flow-through fractions. More than three-quarters of all peptides containing portions of a CDRH3 sequence were found to correspond to *i*CDRH3s (*SI Appendix*, Fig. S6) (75% and 85% of CDRH3-derived peptides in CCH and BSA rabbits, respectively). A small fraction of upstream peptides (i.e., from CDR1 or CDR2) mapping to nondegenerate V gene sequences (all with the same CDRH3) in the database were also identified, providing additional coverage and identification of antigen-specific clonotypes. By including the information deduced from these upstream *i*CDRH3 peptides, the number of antigen-specific *i*CDRH3s in CCH rabbit serum increased from 29 to 34. An additional 25 peptides

Table 1. Highest count *i*CDRH3 peptides from the CCH rabbit detected in the affinity chromatography elution fraction

Rank/ name	*i*CDRH3 MS sequence	Spectral counts (%)	Full CDRH3 transcript sequence (% oxidation products in parentheses)	Fractions where peptide was detected	No. somatic variants (from V_H gene repertoire)	V gene origin
1*	NVAGYLCAPAFNFR	115 (8.00)	ARNVAGYLCAPAFNFRSPGTLVTVSSGQPK	E	1	BM-PC
2*	NFKLWGPGTLVTVSSGQPK	88 (6.11)	ARNFKLWGPGTLVTVSSGQPK	E	3	BM-PC, PBC
3*	MDSHSDGFDPWGPGTLVSVSSGQPK	82 (5.70)	ARMDSHSDGFDPWGPGTLVSVSSGQPK (51%)	E	7	PBC
4	FTISSDNAQNTVDLK	64 (4.45)	AREGYGGYVGYMGLWGPGTLVTVSSGQPK	E+W+F	3	BM-PC, PBC
5*	VCGMDLWGPGTLVTVSSGQPK	50 (3.48)	ARNVYGASRVCGMDLWGPGTLVTVSSGQPK (43%)	E	2	BM-PC
6*	NPGGTSNLWGPGTLVTVSSGQPK	47 (3.27)	ARNPGGTSNLWGPGTLVTVSSGQPK	E	1	BM-PC
7*	KFNLWGPGTLVTVSSGQPK	44 (3.05)	ARDADDYRKFNLWGPGTLVTVSSGQPK	E	1	PBC
8	AFNLWGPGTLVTVSSGQPK	43 (3.00)	ARDVGYGNDNYRAFNLWGPGTLVTVSSGQPK	E+W+F	1	PBC
9*	SPSSGSSNLWGPGTLVTVSSGQPK	39 (2.71)	ARSPSSGSSNLWGPGTLVTVSSGQPK	E	2	BM-PC
10	NSGSASNLWGPGTLVTVSSGQPK	31 (2.16)	ARNSGSASNLWGPGTLVTVSSGQPK	E	2	BM-PC
11	GMDLWGPGTLVTVSSGQPK	29 (2.02)	AREDTYGDANTDYLYRGMDLWGPGTLVTVSSGQPK	E+W	3	BM-PC, PBC
12	NAGTASNLWGPGTLVTVSSGQPK	28 (1.95)	ARNAGTASNLWGPGTLVTVSSGQPK	E	1	BM-PC
13	GLTAADTATYFCAR	28 (1.95)	ARDGIDGNGYNDLNLWGPGTLVTVSSGQPK	E+F	1	BM-PC
14	ELTGNGIYALK	27 (1.88)	ARELTGNGIYALKLGGPGTLVTVSSGQPK	E	1	BM-PC
15	TSSTTVPLQMTSLTAADTATYFCGR	27 (1.88)	GRGYTDGMDLGGPGTLVTVSSGQPK	E	1	BM-PC

*i*CDRH3 peptide sequences, counts of affiliated mass spectra, and their frequencies relative to all spectral counts in the eluent (in parentheses), corresponding full-length CDRH3 sequences, and numbers of somatic variants deduced from the V_H DNA sequence database. BM-PC, V_H genes from the bone marrow PCs; PBC, V_H genes from peripheral B-cells. Percentages of spectral counts corresponding to oxidation products (oxidized L-Met) are shown in parentheses in the full CDRH3 sequences column. *i*CDRH3 peptides detected in the affinity chromatography eluent are marked as "E," in the wash buffer as "W," or in the flow-through as "F."

*V_H synthesized for phage panning and binding validation.

defining 83 full-length V-gene sequences were detected in the elution fraction but were overwhelmingly present in the wash and flow-through fractions (elution frequency:flowthrough + wash frequency ratio <10), indicative of very weakly binding/low-specificity and very low abundance antibodies (Fig. 3*A*; also *SI Appendix*, Fig. S7 for the BSA-specific *i*CDRH3 peptides). *SI Appendix*, Fig. S8 documents the high reproducibility among technical replicates. V-family and J-family gene use of full V_H sequences corresponding to identified *i*CDRH3 peptides were found to be consistent with the germ-line gene use data obtained from 454 sequencing (Fig. 2*B*) and BSA rabbits (*SI Appendix*, Fig. S1; also *SI Appendix*, Fig. S9 for CDRH3 statistics).

In total, our analysis of the CCH rabbit revealed that the antigen-specific polyclonal response is composed of ∼34 IgG antibodies that are classified into 30 different clonotypes [same V and J, same CDRH3 length, 80% aa homology (14, 15)]; 4 of the 34 antibodies that constitute the majority of the antigen-specific IgG repertoire in this animal differed by only 1–3 aa and therefore corresponded to clonally related antibodies (Fig. 3*B*; also *SI Appendix*, Fig. S10 for BSA). The top 12 *i*CDRH3 peptides accounted for ∼60% of all of the antigen-specific peptide counts, suggesting that the serum response was dominated by yet a smaller set of antibodies (sequence logo shown in *SI Appendix*, Fig. S11). The BSA rabbit data revealed a more restricted response comprising 18 distinct antibodies, as shown in *SI Appendix*, Fig. S10 and Table S3.

Comparison of the composition of the serum IgG response with the V gene repertoires obtained by NextGen sequencing of B cells in different compartments can provide useful insights on the dynamics of the humoral response. For example, in the CCH rabbit, 18 of 34 of the V_H sequences encoding the serum antibody repertoire correspond to BM-PC sequences in the sequencing database (database comprising sequences from BM and PBCs as detailed in *SI Appendix*, Table S2). The remainder (16 of 34) of the identified serum antibody repertoire map to PBC sequences in the database. Thus, 7 d after boost immunization nearly half of the serum antibodies seem to be expressed predominantly by plasma cells that had migrated into the bone marrow.

Construction and Characterization of Serum MAbs. To evaluate whether the identified V_H genes encoded proteins that recognize the antigen, it was first necessary to identify the V_L domains to which they pair. The in vivo V_H:V_L pairing problem cannot be universally solved (i.e., for all identified V_H sequences) by proteomic approaches for two reasons: (*i*) the abundances of V_H and V_L chains do not correlate because an excess of V_L chains are secreted in the serum, and V_H chains can pair with more than one V_L; and (*ii*) because of the lower sequence complexity of V_L chains relative to V_H, higher proportions of V_L peptides share partial sequence identity, resulting in increased ambiguity in PSMs and peptide-sequence mappings. Hence, the proteomic problem of distinguishing false-positive identifications increases significantly, and the confidence in the analysis is weak.

We therefore addressed the V_H:V_L pairing problem by synthesizing select V_H genes and then performing two to three rounds of phage panning of scFv libraries constructed with amplified V_L cDNA (i.e., a V_L chain shuffled library for each selected V_H). Phage panning of a fixed V_H with a library of V_L genes is an established method for identifying functional pairs that bind antigen with high affinity (26). The V_H genes corresponding to seven of the most abundant proteomically identified *i*CDRH3s were synthesized by automated DNA synthesis (18). In instances where more than one full-length V_H gene in the database corresponded to an *i*CDRH3 (owing to additional somatic mutations within the V gene), the most common somatic variant was selected for synthesis. The libraries were confirmed to be of sufficient size to cover the entire V_L gene repertoire deduced from DNA sequencing, as estimated by rarefaction analysis (*SI Appendix*, Fig. S12).

For the antigen-specific full-length scFvs isolated by panning (*SI Appendix*, Table S4), DNA sequencing revealed that the synthetic V_H genes paired with one or, in the case of V_H gene 6, two clonally related V_L domains (*SI Appendix*, Fig. S13). The V_H and V_L genes were inserted into vectors encoding rabbit H and L chains, respectively, and cotransfected into HEK293 cells to produce the respective IgGs. The recombinant antibodies were found to display subnanomolar K_D for the antigen by competitive ELISA (Fig. 4) and shown to be effective for

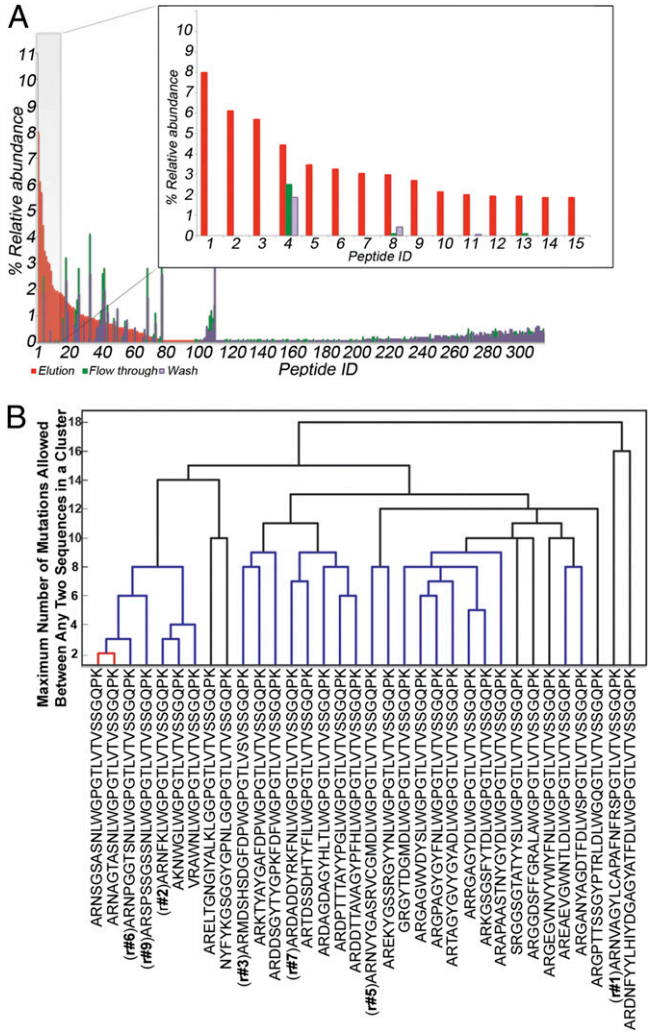

Fig. 3. Identified *i*CDRH3 peptides from affinity chromatography and alignment of corresponding CDRH3s. (*A*) Histogram showing frequencies of identified informative peptides corresponding to a unique clonotype (V genes with same VH, JH CDRH3 sequence and 80% homology in the CDRH3) in the antigen-affinity chromatography elution, flow-through, and wash fractions. (*Inset*) Magnified histogram of the top 15 highest count unique peptides detected in the antigen-affinity chromatography elution, flow-through, and wash fractions. Peptide IDs are ranked by relative abundance in elution. Identified peptides in the affinity chromatography elution fraction that are found overwhelmingly in the flow-through and wash buffer fractions likely correspond to antibodies that bind antigen very weakly or nonspecifically. (*B*) Pairwise alignment of CCH-immunized rabbit CDRH3s in the antigen-specific serum IgG repertoire and observed exclusively in the affinity chromatography elution. The dendrogram shows hierarchical clustering (based on pairwise sequence alignments at the amino acid level) of CDRH3 sequences detected in the elution at >10-fold higher number of counts relative to the affinity chromatography wash and flow-through. (Numbered sequences represent VH synthesized for binding validation.)

immunoprecipitation of CCH from mock mixtures (*SI Appendix*, Fig. S14). Certain proteomically identified V_H genes (2, 5, 7) did not yield specific binders by phage panning. This is not surprising, because it is well established that recombinant rabbit antibody fragments are particularly difficult to express in bacteria (27). Panning phage display of libraries using synthetic V_H genes from the BSA library also showed significant enrichment after two rounds.

Discussion

We report a general strategy for the molecular deconvolution of the monoclonal antibodies that constitute the polyclonal response to antigen. We found that the serum IgG response in an immunized rabbit is oligoclonal, comprising 34 individual antibodies with high antigen selectivity that grouped into 30 distinct clonotypes of antigen-specific antibodies likely to have been derived from different progenitor B cells (*SI Appendix*, Fig. S15). An unimmunized animal that fortuitously had a titer to BSA contained 18 antigen-specific antibodies (12 clonotypes).

The first step in the analysis pipeline involved the determination of the V gene repertoire using NextGen sequencing. In the rabbit we find that the class-switched repertoire is dominated by the use of two to three germ-line V_H families and by two germ-line J_H families [J4 and J2 in both CCH and BSA rabbits (Fig. 2*B* and *SI Appendix*, Fig. S1)]. The analysis of the V gene repertoire further revealed that a subpopulation of rabbit V genes contains a large number of amino acid substitutions relative to the germline, likely a consequence of the gene conversion processes that occur during V gene diversification in rabbits. The deconvolution of the serum IgGs was made possible by LC/MS-MS shotgun proteomic analysis using an individually derived B-cell V gene sequence database, combined with strict filtering criteria for the confident identification of peptide sequences, including the use of a high mass accuracy filter (≤1.5 ppm). Although we have provided evidence that our analysis correctly captures the key features of the serological IgG repertoire, we do acknowledge that as a method that relies on LC-MS, it is subject to the experimental constraints of shotgun proteomic analyses (28).

Collectively, the identification of the repertoire of antigen-specific antibodies in serum leads to several interesting observations. First, the serum response in these animals seemed to be oligoclonal (i.e., neither is it highly diverse nor does it comprise only of a handful different antibodies). Second, V_H use in the circulating antibodies was entirely consistent with the V_H gene repertoire in the animal as determined by NextGen sequencing, highlighting the significance of the cellular repertoire in shaping humoral immunity (Fig. 2*B* and *SI Appendix*, Fig. S1). Third, it is interesting that most of the antigen-specific V_H clonotypes identified proteomically correspond to a single V gene or at most to only a few somatic variants. However, this may not always be the case, especially when antibodies are generated in response to persistent or recurrent infections (29–32). In those instances, detection of peptides from CDR1 and CDR2 might be used to identify the dominant somatic variant(s) in serum. Fourth, as

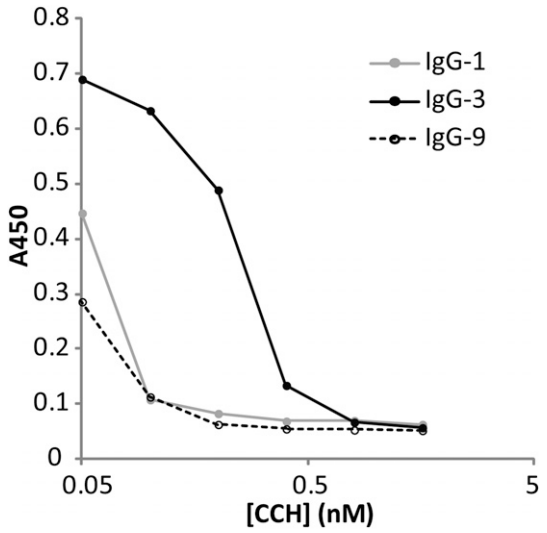

Fig. 4. Competitive ELISA of full-length IgG containing the V_H genes corresponding to selected abundant *i*CDRH3s (Table 1) identified in the CCH immunized rabbit. Antibodies identified using the proposed methodology (Fig. 1) were shown to display subnanomolar affinities. The expression yield of the rabbit IgG-6 antibody in HEK-293 cells was too low for accurate quantitative affinity measurements.

expected, many of the proteomically identified CDRH3s corresponded to V_H genes isolated from BM-PCs. Approximately half of the iCDRH3s, however, were found to map only to PBCs and may be derived from recently activated plasmablasts in transit to the bone marrow. We note that because the formation of plasmablasts in the course of B-cell expansion is a consequence of asymmetric division that should also give rise to B memory cells, highly exhaustive DNA sequencing of the peripheral B memory V gene repertoire should be sufficient to yield all of the Ig sequences found in circulation at steady state (33). Fifth, the CCH rabbit data revealed that two of the top ranked *i*CDRH3s in terms of spectral counts display evidence of oxidative modifications, as shown in Table 1. Although it is known that amino acid oxidation can occur during sample preparation, it is tempting to speculate that the oxidative modification of certain circulating IgGs may have resulted from in vivo posttranslational modifications rather than processing artifacts, because plasma cells experience high levels of oxidative stress (34), and the $t_{1/2}$ of circulating Igs in serum is >1 wk, resulting in extensive exposure of antibodies to oxidizing conditions. L-Methionine oxidation has also been observed repeatedly in recombinantly expressed therapeutic antibodies from CHO cells (35, 36). Additional studies will be needed to determine the extent of L-Methionine oxidation during sample processing and the frequency of in vivo modification.

At present the molecular deconvolution of antigen-specific serum antibodies requires a sample size of ∼3–5 mL of whole blood, an amount that is easily obtainable from most laboratory animals down to rats and also from humans. With further advances in the sensitivity of MS, it may prove possible to also analyze the Ig serum composition from a single mouse, including genetically homogenous transgenic animals displaying well-characterized defects in B-cell development. Importantly, the approach we have developed may be extended to the analysis of the serological response in humans after vaccination or related to pathologic states.

Materials and Methods

All materials used in this study, including vendor source are provided in *SI Appendix*, *SI Materials and Methods*.

Rabbit immunization, serum IgG isolation and sample preparation for LC-MS/MS measurements, proteomics data analysis, and all related computational analyses are described in *SI Appendix*, *SI Materials and Methods*.

Validation and characterization of proteomically identified serum antibodies including synthetic gene synthesis, recombinant IgG cloning, and ELISA methodologies are provided in detail in *SI Appendix*, *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Bob Glass for assistance with rabbit immunization and bone marrow isolation, Dr. Sai Reddy for help with flow cytometry, Dr. J. Borrok for initial experiments, Dr. Scott Hunicke-Smith for assistance with Next-Gen DNA sequencing, Constantine Chrysostomou for assistance in bioinformatics analysis, Chhaya Das for recombinant IgG expression, Dr. Greg Ippolito for reading the manuscript, Prof. Itai Benhar for important input and comments on the manuscript, and Prof. Brent L. Iverson for useful discussions. Funding for this work was provided by the Clayton Foundation (G.G.), Welch Foundation Grant F1515 (to E. M. Marcotte), Defense Advanced Research Projects Agency (G.G. and A.D.E.), and National Institutes of Health (NIH) Grants 5 RC1DA028779 (to G.G. via a subcontract from University of Chicago) and GM 076536 (to E. M. Marcotte). J.J.L. was supported by a postdoctoral fellowship by Cancer Prevention and Research Institute of Texas. The Linear Trap Quadrupole (LTQ) Orbitrap Velos MS was purchased with generous support by the NIH Western Research Center of Excellent in Biodefense (NIH Grant 5U54AI057156) and the Texas Institute for Drug and Diagnostics Development (TI-3D).

1. Browning CH (1955) Emil Behring and Paul Ehrlich: Their contributions to science. *Nature* 175(4457):570–575.
2. Kantha SS (1991) A centennial review; the 1890 tetanus antitoxin paper of von Behring and Kitasato and the related developments. *Keio J Med* 40(1):35–39.
3. Radbruch A, et al. (2006) Competence and competition: The challenge of becoming a long-lived plasma cell. *Nat Rev Immunol* 6(10):741–750.
4. Scheid JF, et al. (2009) A method for identification of HIV gp140 binding memory B cells in human blood. *J Immunol Methods* 343(2):65–67.
5. Wrammert J, et al. (2008) Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* 453(7195):667–671.
6. Dekker LJ, et al. (2011) An antibody-based biomarker discovery method by mass spectrometry sequencing of complementarity determining regions. *Anal Bioanal Chem* 399(3):1081–1091.
7. de Costa D, et al. (2010) Sequencing and quantifying IgG fragments and antigen-binding regions by mass spectrometry. *J Proteome Res* 9(6):2937–2945.
8. Bandeira N, Pham V, Pevzner P, Arnott D, Lill JR (2008) Automated de novo protein sequencing of monoclonal antibodies. *Nat Biotechnol* 26(12):1336–1338.
9. Cheung WC, et al. (2012) A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat Biotechnol* 30(5):447–452.
10. Reddy S, et al. Rapid Isolation of Monoclonal Antibodies from Animals. Patent 20110312505 (Filed May 17, 2011).
11. Lindop R, et al. (2011) Molecular signature of a public clonotypic autoantibody in primary Sjögren's syndrome: A "forbidden" clone in systemic autoimmunity. *Arthritis Rheum* 63(11):3477–3486.
12. Sato S, et al. (2012) Proteomics-directed cloning of circulating antiviral human monoclonal antibodies. *Nat Biotechnol* 30(11):1039–1043.
13. Murphy K, Travers P, Walport M, eds (2007) *Janeway's Immunobiology* (Garland Science, New York).
14. Poulsen TR, Jensen A, Haurum JS, Andersen PS (2011) Limits for antibody affinity maturation and repertoire diversification in hypervaccinated humans. *J Immunol* 187 (8):4229–4235.
15. Moody MA, et al. (2011) H3N2 influenza infection elicits more cross-reactive and less clonally expanded anti-hemagglutinin antibodies than influenza vaccination. *PLoS ONE* 6(10):e25797.
16. Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* 85(23):8998–9002.
17. Rader C, et al. (2000) The rabbit antibody repertoire as a novel source for the generation of therapeutic human antibodies. *J Biol Chem* 275(18):13668–13676.
18. Reddy ST, et al. (2010) Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat Biotechnol* 28(9):965–969.
19. Knight KL (1992) Restricted VH gene usage and generation of antibody diversity in rabbit. *Annu Rev Immunol* 10:593–616.
20. Knight KL, Becker RS (1990) Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: Implications for the generation of antibody diversity. *Cell* 60(6):963–970.
21. Wu TT, Johnson G, Kabat EA (1993) Length distribution of CDRH3 in antibodies. *Proteins* 16(1):1–7.
22. Becker RS, Knight KL (1990) Somatic diversification of immunoglobulin heavy chain VDJ genes: Evidence for somatic gene conversion in rabbits. *Cell* 63(5):987–997.
23. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass-spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5(11):976–989.
24. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26(12):1367–1372.
25. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 4(11):923–925.
26. Marks JD, et al. (1992) By-passing immunization: Building high affinity human antibodies by chain shuffling. *Biotechnology (N Y)* 10(7):779–783.
27. Popkov M, et al. (2003) Rabbit immune repertoires as sources for therapeutic monoclonal antibodies: The impact of kappa allotype-correlated variation in cysteine content on antibody libraries selected by phage display. *J Mol Biol* 325(2):325–335.
28. Mallick P, Kuster B (2010) Proteomics: A pragmatic perspective. *Nat Biotechnol* 28(7): 695–709.
29. Corti D, et al. (2011) A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* 333(6044):850–856.
30. Wu X, et al.; NISC Comparative Sequencing Program (2011) Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* 333(6049):1593–1602.
31. Scheid JF, et al. (2011) Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* 333(6049):1633–1637.
32. Walker LM, et al.; Protocol G Principal Investigators (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 477(7365):466–470.
33. Barnett BE, et al. (2012) Asymmetric B cell division in the germinal center reaction. *Science* 335(6066):342–344.
34. Slifka MK, Antia R, Whitmire JK, Ahmed R (1998) Humoral immunity due to long-lived plasma cells. *Immunity* 8(3):363–372.
35. Boyd D, Kaschak T, Yan B (2011) HIC resolution of an IgG1 with an oxidized Trp in a complementarity determining region. *J Chromatogr B Analyt Technol Biomed Life Sci* 879(13-14):955–960.
36. Houde D, Peng Y, Berkowitz SA, Engen JR (2010) Post-translational modifications differentially affect IgG1 conformation and receptor binding. *Mol Cell Proteomics* 9(8):1716–1728.

Supporting Information

Molecular Deconvolution of the Monoclonal Antibodies that Comprise the Polyclonal Serum Response

Yariv Wine[a,b,1], Daniel R. Boutz[b,c,1], Jason J. Lavinder[a,b,1], Aleksandr E Miklos[b,e], Randall A Hughes[b,e], Kam Hon Hoi[f], Sang Taek Jung[a,b], Andrew P Horton[c,f], Ellen M Murrin[a], Andrew D Ellington[b,c,d,e], Edward M Marcotte[b,c,d,2] & George Georgiou[a,b,f,g,2]

SI Materials and Methods

Rabbit immunization and sample preparation

Rabbit CCH: A New Zealand white rabbit was immunized with 100 μg *Concholepas concholepas* hemocyanin (CCH, Pierce, IL, USA) in 2 ml of 1:1 saline and complete Freund's adjuvant (CFA). At days 14 and 28, a booster immunization with the same volume of antigen in incomplete Freund's Adjuvant (IFA) was administered. The animal was sacrificed on day 35, at which point femoral bone marrow (BM) cells were isolated and approximately 100 ml blood was collected into heparin tubes. Blood aliquots of 20 ml were gently layered over 20 ml of Histopaque 1077 (Sigma, MO, USA) and centrifuged in a swinging bucket rotor at 400*g*, 45 min at 25°C (Beckman Coulter). The serum was removed from the top of the gradient and stored at -20° C. Peripheral blood cells (PBC) were isolated from the intermediate layer. Each collected tissue (BM and PBC) was processed as previously described (1), with the exception that PBC's did not require red blood cell lysis after gradient centrifugation. $CD138^+$ cells were isolated as previously described using rat anti-mouse CD138 antibody-clone 281-2 (BD Pharmingen, CA, USA), which had been shown (2) to be cross-reactive to rabbit CD138. Cell fractions isolated as described herein (PBCs or $CD138^+$ cells) were centrifuged at 930*g*, 5 min at 4°C. Cells were then lysed with TRI reagent (Ambion, TX, USA) and total RNA was isolated according to the manufacturer's protocol in the Ribopure RNA isolation kit

(Ambion). RNA concentrations were measured with an ND-1000 spectrophotometer (Nanodrop, DE, USA).

Rabbit BSA: A New Zealand white rabbit was screened for high titer against BSA before any immunization regime was applied (Day-0). From this high titer rabbit (titer >1:10^5), approximately 15 ml blood was collected into heparin tubes. Blood samples were gently layered over 15 ml of Histopaque 1077 (Sigma, MO, USA) and centrifuged in a swinging bucket rotor at 400*g*, 45 min at 25°C (Beckman Coulter). The serum was removed from the top of the gradient and stored at -20° C. PBC's were isolated from the intermediate layer. Collected PBC's were processed as previously described (1), with the exception that PBC's did not require red blood cell lysis after gradient centrifugation. PBC's were centrifuged at 930*g*, 5 min at 4°C. Cells were then lysed with TRI reagent (Ambion, TX, USA) and total RNA was isolated according to the manufacturer's protocol in the Ribopure RNA isolation kit (Ambion). RNA concentrations were measured with an ND-1000 spectrophotometer (Nanodrop, DE, USA).

Amplification and high-throughput sequencing of V_H and V_L gene repertoires

Approximately 0.5 µg of ethanol precipitated RNA was used for first-strand cDNA synthesis according to the manufacturer's protocol for 5' RACE using the SMARTer RACE cDNA Amplification kit (Clontech, CA, USA). The cDNA reaction was diluted into 100 µl of Tris-EDTA buffer and stored at -20°C. 5' RACE PCR amplification was performed on the first strand cDNA to amplify the V_H repertoire with the kit-provided, 5' primer mix and 3' rabbit IgG-specific primers RIGHC1 and RIGHC2 (**Table S1**). The rabbit V_L repertoire was amplified via 5' RACE, using a 3' primer mix specific for both the V_κ and V_λ rabbit constant regions. The V_L primer mix comprised 90% RIGκC and 10% RIG$_\lambda$C (**Table S1**) to approximate known ratios of light chain isotypes in rabbits. Reactions were carried out in a 50 µl volume by mixing 35.25 µl H_2O, 5 µl 10X Advantage-2 PCR buffer (Clontech), 5 µl 10X Universal Primer A mix (Clontech), 0.75 µl Advantage-2 polymerase mix (Clontech), 2 µl cDNA, 200 nM V_H or V_L primer mix, and 200 µM dNTP mix. PCR conditions were: 95 °C for 5 min, followed by 30 cycles

of amplification (95 °C for 30 sec, 60 °C for 30 sec, 72 °C for 2 min), and a final 72 °C extension for 7 min. The PCR products were gel-purified to isolate the amplified V_H or V_L DNA (~500 bp). 100 ng of each 5' RACE amplified V_H or V_L DNA was processed for Roche GS-FLX 454 DNA sequencing according to the manufacturer's protocol.

All 454 data were first processed using the sequence quality and signal filters of the 454 Roche pipeline and then subjected to bioinformatics analysis that relied on homologies to conserved framework regions and V germline gene identification using IMGT/HighV-Quest Tool (3). The numbers of full-length V sequences, unique V genes and CDRH3s are summarized in **Table S2**. Additional filters were applied for full repertoire database construction as follows: (i) Length cutoff*:* full-length sequences were filtered by aligned amino acid lengths > 70 residues and aligned framework 4 region lengths > 2 residues; (ii) Stop codons: aligned amino acid sequences containing stop codons were removed; (iii) Finally for the purposes of germline VDJ, CDR3 length and amino acid composition analyses, only sequences with n≥2 reads were considered.

Protein A purification and pepsin digestion

2.5 ml of rabbit serum was diluted 4-fold in PBS and IgG proteins were purified by affinity chromatography using 2.5 ml of protein A agarose (Pierce, IL, USA) packed in an Econo-Pac chromatography column (Biorad, CA, USA). Diluted serum was recycled 6 times through the protein A affinity column in gravity mode. The column was washed with 10 column volumes (CV) of PBS and IgG was eluted using 5 CV of 100 mM glycine pH 2.7 and immediately neutralized with 1M Tris-HCl pH 8.5. IgG in the flow through, wash and elution was determined by non-reducing SDS-PAGE in 4-20% gels (Biorad, CA, USA). Protein A purified serum IgG was digested to produce $F(ab)_2$ fragments using immobilized pepsin (Pierce, IL, USA). Briefly, 500 µl of immobilized pepsin agarose was equilibrated with 20 mM sodium acetate, pH 4.5 mixed with approximately 10 mg of purified IgG in 1.5 ml 20 mM sodium acetate, pH 4.5. Digestion was allowed to proceed for 7 hours, shaking vigorously at 37 °C. The pepsin-agarose

beads were separated by applying the reaction solution to an Ultrafree centrifugal filter column (Millipore, MA, USA) and the degree of digestion was evaluated by non-reducing 4-20% SDS-PAGE (**Fig. S4**)

Antigen enrichment

Affinity chromatography for the isolation of antigen-specific IgG-derived $F(ab)_2$ was carried out by coupling *Concholepas concholepas* hemocyanin protein (Pierce, IL, USA) for the CCH immunized rabbit and BSA (Sigma, MO, USA) for the BSA rabbit onto N-hydroxysuccinimide (NHS)-activated agarose according to the manufacturer's protocol (Pierce, IL, USA). Briefly, for the CCH immunized rabbit $F(ab)_2$, 100 mg of CCH in 50 ml of PBS was incubated with 1 g NHS-activated agarose and for the BSA rabbit, 0.5 mg of BSA in 10 ml of PBS was incubated with 50 mg NHS-activated agarose at 4°C overnight, rotating end-over-end. The coupled agarose beads were washed with PBS, unreacted NHS groups were blocked with 1M ethanolamine, pH 8.3, (Sigma, MO, USA) for 60 min at room temperature, washed with PBS and packed into a 15 ml and 5 ml Econo-Pac chromatography column (CCH and BSA coupled beads respectively, Biorad, CA, USA). $F(ab)_2$ fragments were applied to the CCH and BSA affinity column in gravity mode, with the flow-through collected and reapplied to the column 5 times. The column was subsequently washed with 10 CV of PBS, eluted using 100 mM glycine pH 2.7 and immediately neutralized with 1 M Tris-HCl pH 8.5. The flow-through, wash, and elution fractions were collected for subsequent analysis.

Trypsin digestion

Trypsin digestion in the presence of TFE was carried out by as follows: Protein fractions from the affinity chromatography steps above (fractions from elution, flow through and wash buffer) were incubated at 37 °C for 60 min in a reaction solution that consisted of (final concentrations): 50% v/v TFE, 50 mM ammonium bicarbonate and 2.5 mM DTT.

Denatured, reduced $F(ab')_2$ were then alkylated by incubation with 32 mM iodoacetamide (Sigma, MO, USA) for 1 hour at room temperature and quenched by addition of 7.7 mM DTT for 1 hour at room temperature. Samples were diluted with water to reach a final TFE concentration of 5% v/v. Trypsin digestion was carried out with a ratio of 1:75 trypsin:protein and incubating at 37 °C for 5 hours. Trypsin was inactivated by lowering the pH with 1% v/v formic acid.

Sample preparation for LC-MS/MS

Trypsin digested $F(ab)_2$ peptides were concentrated by SpeedVac centrifugation and the trypsinized $F(ab)_2$ solution was applied to Hypersep C-18 spin tips (Thermo Scientific, IL, USA), washed 3x with 0.1% v/v formic acid and eluted using 60% v/v acetonitrile, 0.1% v/v formic acid. Eluted peptides were lyophilized and re-suspended in 100 µl 5% v/v acetonitrile, 0.1% v/v formic acid. Subsequently, samples were subjected to a 10 kDa MWCO spin column (Millipore, MA, USA) and the flow-through containing the peptides was collected.

LC-MS/MS measurements

The resulting peptides from antigen enriched trypsin digested $F(ab)_2$ were loaded onto an Acclaim PepMap C18 Column (Dionex, IL, USA) interfaced to a Ultimate3000 RSLCnano UHPLC system (Dionex) and separated using a 5-40% acetonitrile gradient over 245 minutes. Peptides were eluted onto an LTQ Orbitrap Velos mass spectrometer (Thermo Scientific) using a Nano-spray source. The LTQ Orbitrap Velos was operated in data dependent mode with a target value of 5e5 ions for parent ion (MS1) scans collected at 60,000 resolution. Ions with charge >+1 were selected for fragmentation by collision-induced dissociation with a maximum of 20 MS2 scans per MS1. Dynamic exclusion was activated, with a 45-s exclusion time for ions selected more than twice in a 30-s window.

Proteomics Data Analysis

The resulting spectra were searched against a protein sequence database consisting of the in-house rabbit V_H and V_L sequences obtained as described above, concatenated with the rabbit full protein-coding sequence database (OryCun2) and a database of common protein contaminants compiled by the Max Planck Institute of Biochemistry (www.maxquant.org). Peptide-spectrum matches were identified by SEQUEST (Proteome Discoverer 1.3, Thermo Scientific). Only V_H and V_L sequences with ≥2 reads were included in the search. The search specified full tryptic peptides with up to 2 missed tryptic cleavages allowed. A precursor mass tolerance of 10 ppm was used, with fragment mass tolerance set to 0.5 Da. Carbamidomethylation of cysteine residues by iodoacetamide was selected as a static modification, while oxidized methionine was allowed as a dynamic modification. Following the search, peptides were filtered using Percolator (4) through Proteome Discoverer, with an applied FDR of <1% determined against a reverse-sequence decoy database. For cases where multiple high-confidence matches were scored for the same spectrum (with maximum ΔCN of 0.05 allowed), the top-match was selected unless it showed an absolute mass error >2ppm and was superseded by a lower-ranking match with absolute mass error at least 0.5 ppm more accurate than the top-ranked match.

To control false discovery rates at the peptide level, the average mass deviation (AMD) was calculated for each set of spectra identified as the same peptide, with modifications considered as unique from unmodified peptides, and only peptide identifications with an average <1.5 ppm across all PSMs were included in the final dataset. For these purposes, all observed masses were first recalibrated according to the method of Cox *et al* (5). The recalibrated mass errors were then averaged across all high-confidence spectra identified in order to calculate a given peptide's AMD.

An examination of V gene sequences suggested that isobaric peptides with an isoleucine-leucine swap were more common than in standard proteomes and thus required special

consideration, as two peptides differing exclusively in this way will generate identical spectra by MS and cannot be differentiated. We therefore considered all Iso/Leu sequence variants as a single group, and mapped the group to all CDRH3s associated with any of the group members, which preserved the possibility that even if an individual sequence cannot be uniquely identified, the associated CDRH3 might still be unique. For other isobaric pairings (e.g., Asp/Gly-Gly, Gln/Gly-Ala) and ambiguous identifications where MS/MS spectral differences can distinguish between pairings, we considered only the top-ranked PSM as determined by the SEQUEST-Percolator pipeline.

Multiple V_H sequence alignment

Multiple sequence alignments of the *i*CDRH3 sequences with the V gene database were carried out using MUSCLE (6) as implemented in Geneious (www.geneious.com). Full V_H sequences with the highest *i*CDRH3 counts were chosen if the number of reads for the sequence was higher compared to the next sequence in the alignment.

Pairwise sequence alignments for every potential pairing of CDRH3 sequences, were calculated using the Needleman-Wunsch global alignment algorithm (7). Sequences were then clustered using complete linkage hierarchical clustering employing the pairwise alignment scores as the distance measures. The sequences were organized into a dendrogram in which the maximum number of mutations found in each cluster at each cluster level was determined and plotted as the y-axis (**Fig. 3a and Fig. S7** for CCH and BSA rabbits respectively).

Construction of synthetic antibody genes.

CCH rabbit: Synthetic gene construction was carried out as described previously (1) with the following modifications: The coding sequences for the selected V_H genes were designed using the GeneFab software component of our in-house protein fabrication

automation (PFA) platform (8). After reverse translation of the primary amino acid sequences for each V_H using an *E. coli* class II codon table for CCH rabbit, the coding sequences were built with a polyglycine-serine linker $(GGGGS)_2$ at the C-terminus for overlap reassembly scFv construction. A 5' SfiI restriction endonuclease site was added to facilitate cloning of the scFv constructs into the pAK200 phage display vector (9). The V_H genes were aligned using the sequence encoding the common $(GGGGS)_2$ linker sequence and a universal randomly generated stuffer sequence was applied to the ends of the V_H sequences to ensure that all of the constructs were of the same length. The Sfi I-V_H-$(GGGGS)_2$ genes were synthesized from overlapping oligonucleotides using a modified thermodynamically balanced inside-out nucleation PCR (10). The 80-mer oligonucleotides necessary for the construction of the various scFv genes were designed using the GeneFab software with a minimal overlap of 30 nucleotides between oligonucleotide fragments. The oligonucleotides were synthesized using standard phosphoramidite chemistry at a 50 nmol scale using a Mermade 192 oligonucleotide synthesizer (Bioautomation, TX, USA) using synthesis reagents from EMD Chemical and phosphoramidites from Glen Research. All of the oligonucleotide liquid-handling operations necessary for assembling the various genes were done on a Tecan Evo 200 workstation (Tecan, CA, USA) with reagent management and instrument control done through the FabMgr software component of the PFA platform (8). The gene assembly PCRs were performed using KOD-Hotstart polymerase using buffers and reagents supplied with the enzyme (Novagen, MA, USA).

BSA rabbit: Synthetic gene construction was carried out as described herein (see CCH rabbit) with the following modifications: reverse translation of the primary amino acid sequences were carried out by DNA2.0 gene design (11) and DNA sequences were synthesized by integrated Device Technologies (IDT, CA, USA).

Combinatorial V_L chain shuffling of selected V_H as phage displayed scFv

CCH rabbit: The V_L libraries were prepared by amplifying PBC's and BM-PC's cDNA in a reaction containing: 40.25 µl H_2O, 5 µl 10X Advantage-2 buffer, 2 µl cDNA, 0.75 µl Advantage-2 polymerase mix, 1 µl 10 mM dNTP mix, 0.5 µl 100 uM RLR1/RLR2 equimolar degenerate primer mix, and 0.5 µl 100 uM FLR1 degenerate primer. The PCR program used for V_H amplification described above was used. The PCR product (~400 bp) was gel-purified and quantified with the ND-1000 spectrophotometer.

DNA encoding each of the synthetic V_H genes was heated, hybridized, and treated with the SURVEYOR mutation detection kit (Transgenomic, NE, USA) according to the manufacturer's protocol. The undigested full-length product for each V_H reaction was gel-purified and quantitated in the ND-1000 spectrophotometer. scFv overlap reassembly PCR libraries were prepared in reactions containing: 100 ng of full-length synthetic V_H gene DNA, 50 ng each of gel-purified V_L PCR product from BM-PC and PBC's, 5 µl 10X Thermopol buffer (NEB, MA, USA), 0.5 µl Taq DNA polymerase (NEB), 200 uM dNTP mix, 1 uM rabbit V_H forward primer, 1 uM OE-R primer and filled to 50 µl final volume with ddH_2O. The PCR thermocycle program was: 94 °C for 1 min, 25 cycles of amplification (94 °C for 15 sec, 60 °C for 15 sec, 72 °C for 2 min), and a final 72 °C extension for 5 min. The overlap PCR product (~750 bp) was gel-purified twice and digested with Sfi I (NEB) and ligated into the pAK200 phage display vector (12). The ligation product was transformed into XL1-Blue or Jude1 *E.coli* (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI*qZΔ*M15* Tn*10* (Tetr)])(13) to give 7 separate libraries comprising between 10^6-10^7 transformants each. Rarefaction analysis(14) and species richness estimation(15) on the BM-PC CCH rabbit V_L high-throughput sequencing data revealed that the V_L repertoire encoded by bone marrow $CD138^+$ cells consisted of an estimated 10,252 unique CDRL3 (**Figure S12**). Therefore, a library comprising approximately 10^5 clones should capture 99% of the repertoire. Thus, each library was at least one order of magnitude larger than required to capture even the rarest CDRL3 clones.

BSA rabbit: For the high BSA titer rabbit without antigen immunization, V_L libraries were prepared as described except that the V_L libraries were prepared by amplification of only PBC cDNA.

Phage panning

Cells for the seven CCH scFv libraries and 4 BSA scFv libraries, each comprising a synthetic V_H gene joined to the amplified V_L cDNA library, were scraped from agar plates containing LB+ chloramphenicol (35 μg/ml) + 1% w/v glucose and then diluted into 25 ml of 2YT growth media supplemented with chloramphenicol (35 μg/ml) + tetracycline (10 μg/ml) + 1% w/v glucose to a final OD_{600} ~ 0.1. Cells were grown at 37°C, with shaking at 250 RPM until reaching log phase growth (OD_{600} ~ 0.5), and then infected with 100 MOI of M13KO7 helper phage (16) and incubated without mixing at 37°C for 1 hour. The cells were pelleted and resuspended in 25 ml of fresh 2YT media + chloramphenicol (35 μg/ml) + kanamycin (35 μg/ml) + 1% w/v glucose + 0.5 mM IPTG. Cultures were grown at 25 °C, with shaking at 250 RPM overnight (~14 hours). The cells were pelleted by centrifugation and phages were isolated from the supernatant by PEG-NaCl precipitation. For panning, immunotubes were coated overnight at 4 °C with either BSA or antigen CCH resuspended in PBS at 50 μg/ml concentration and then blocked for 2 hours at room temperature with 2% milk dissolved in PBS. CCH libraries panning included additional immunotubes with 3% BSA in PBS as well (blocking solutions were alternated during sequential rounds of panning).

CCH libraries: Phage-scFv (dissolved in PBS) were diluted into 2% milk to input 10^{13} phage into each of two BSA-coated, blocked immunotubes and rotated end-over-end at room temperature for 1.5 hours. One immunotube of the depleted phage-scFv was then directly transferred into a CCH-coated blocked immunotube and the other to a BSA-coated, blocked immunotube.

BSA libraries: Phage-scFv (dissolved in PBS) were diluted into 2% milk to input 10^{13} phage into each of two 2% milk coated tubes and rotated end-over-end at room temperature for 1.5 hours. One immunotube of the depleted phage-scFv was then directly transferred into a BSA-coated blocked immunotube and the other to a 2% milk blocked immunotube.

For both CCH and BSA libraries, each immunotube was subsequently rotated at room temperature for 2 hours for binding of the phage-scFv. The immunotubes were then washed 6X with 4 ml PBST (0.05% v/v Tween 20) and 4X with 4 ml PBS. Elution was accomplished using 1 ml 100 mM triethylamine, rotating at room temperature for 8 min and then the solution was immediately transferred to a 2 ml microcentrifuge tubes containing 700 µl 1.5 M Tris-HCl pH 8.0. Subsequently, 250 µl of Tris-HCl pH 8.0 was added directly into the emptied immunotube to neutralize any residual elution solution. Both elution fractions (700 µl and the residual 250 µl) were used to infect 12 ml of log phase *E.coli* XL1-Blue or Jude1 cells, with 3 ml of the culture placed in the neutralized immunotubes to capture remaining bound phage. After 1 hour at 37 °C, the infected culture was plated onto LB agar plates containing chloramphenicol (35 µg/ml) + 1% w/v glucose for titering both the BSA-specific elution and the CCH-specific elution. For the CCH libraries, selectivity was determined as the titer of the CCH-specific elution divided by the titer of the BSA-specific elution (**Table S4**) and for the BSA libraries the selectivity was determined against the 2% milk elution. The entire CCH-specific and BSA-specific elution solutions (~12 ml infected culture spun down and resuspended in 2 ml 2YT) were spread onto large LB-chloramphenicol-glucose plates and incubated overnight 37 °C. Colonies were scraped and cells were resuspended and used for subsequent rounds of phage amplification and panning.

Phage scFv ELISAs

To evaluate binding of clones obtained by phage panning, single colonies from each V_H-V_L library were inoculated into 150 µl 2YT media + chloramphenicol (35 µg/ml) +

tetracycline (10 µg/ml) + 1% w/v glucose to a final OD_{600} of ~0.5 in a 96 well round bottom plate. Each culture was then infected with 100 MOI of M13KO7 helper phage and incubated at 37 °C for 1 hour. Cells were then pelleted by centrifugation and resuspended in 25 ml 2YT media + chloramphenicol (35 µg/ml) + kanamycin (35 µg/ml) + 1% w/v glucose + 0.5 mM IPTG. Phage displaying scFv antibodies was produced by growing the cells at 25 °C with shaking at 250 RPM overnight (~14 hours). Cells were pelleted by centrifugation and 50 µl of supernatant was transferred to ELISA plates previously coated with CCH (10 µg/ml overnight at 4 °C) and BSA and blocked with 2% milk in PBS (2 hours, room temperature). An equal volume of 2% milk in PBS was added to each well and phage-scFv were allowed to bind with gentle shaking for 1 hour. After binding, ELISA plates were washed 3x with PBST and incubated with 50 µl of anti-M13-HRP secondary antibody (1:5000, 2% milk in PBS) for 30 min, 25 °C. Plates were washed 3x with PBST, then 50 µl Ultra TMB substrate (Thermo Scientific) was added to each well and incubated 25 °C for 5 min. Reactions were stopped using equal volume of 1M H_2SO_4 and absorbance was read at 450 nm (BioTek, VT, USA).

Expression and purification of recombinant rabbit IgGs

V_H sequences derived from the rabbit immunized with CCH and paired V_L genes identified following screening of the respective phage scFv libraries, were cloned into the rabbit IgG expression vectors pFUSEss-CHIg-rG*03 and pFUSE2ss-CLIg-rk1 (Invivogen, NY. USA) respectively, and the DNA was electroporated into DH10B cells. 120 µg each purified, sequence verified pFUSEss-CHIg and pFUSE2ss-CLIg vectors were co-transfected into HEK 293F cells following the Freestyle MAX expression system instructions (Invitrogen, NY, USA). HEK 293F cells were grown for 6 days after transfection and medium was harvested by centrifugation and IgG was purified by a protein-A agarose (Pierce, IL, USA) chromatography column.

IgG affinities for CCH were determined by competitive ELISA using different concentrations of IgG in a serial dilution of antigen, ranging from 1.6 nM to 0.05 nM in

the presence of 1% milk in PBS. The concentrations of IgG used were chosen based on the signal given in an initial indirect ELISA in which a dilution series of each IgG was analyzed, with the IgG concentrations analyzed being in the linear range of the initial ELISA. Each sample was incubated overnight at room temperature to equilibrate. Plates were coated overnight at 4 °C with 10 µg/mL of CCH in 50 mM carbonate buffer, pH 9.6. Coated plates were washed three times in PBST and blocked with 2% milk in PBS for two hours at room temperature. Equilibrated samples were then added to the block plate and incubated for one hour at room temperature. After binding, ELISA plates were washed 3x with PBST and incubated with 50 µl of anti-rabbit IgG-HRP secondary antibody (Sigma, MO, USA) (1:5,000, 2% milk in PBS) for 30 min, 25 °C. Plates were washed 3x with PBST, then 50 µl Ultra TMB substrate (Thermo Scientific) was added to each well and incubated 25 °C for 5 min. Reactions were stopped using equal volume of 1M H_2SO_4 and absorbance was read at 450 nm (BioTek, VT, USA).

Immunoprecipitation

15 mL of an overnight Jude1 *E. coli* culture grown in LB media with glucose was collected by centrifugation at 4000g for 10 minutes and resuspended in 1 mL of PBS. The cell suspension was lysed by sonication and cleared by centrifugation. The lysate was then depleted against 100 µl of Protein A resin (Thermo Scientific, IL, USA) mixing end-over-end for one hour at 4 °C, spun down for 3 minutes at 3000g, and the supernatant collected and the protein concentration determined. 300 µg of CCH were mixed with 6 mg of E. coli cell lysate protein and 10 µg of the IgG-2 rabbit antibody in PBST (PBS+ 0.05% Tween 20) buffer also containing 5 mM EDTA and 0.5 mM PMSF (IP buffer), giving a total volume of 600 µl. After overnight incubation at 4° C, 100 µl of Protein A resin washed with IP buffer 3X was added to the mock immunoprecipitation solution above and rotated end-over-end for 1 hour at room temperature. The resin was then spun down at 3000g and washed with 1 mL of PBST. After six washes in PBST, the resin was resuspended in 60 µl of SDS gel-loading buffer with 100 mM DTT and heated

at 95 °C for 10 minutes. The mix was then centrifuged through an Ultrafree MC filter (Millipore, MA, USA) and the filtrate was loaded on a 4% SDS-PAGE gel.

SUPPLEMENTARY FIGURES AND TABLES

Table S1. Primer sequences for PCR amplification of V_L and V_H genes.

Primer Name	Sequence	Description of use
RIGHC1	CAGTGGGAAGACTGACGGAGCCTTAG	Rabbit IgG CH1 reverse V_H primer mix (equimolar)
RIGHC2	CAGTGGGAAGACTGATGGAGCCTTAG	Rabbit IgG CH1 reverse V_H primer mix (equimolar)
RIGκC1	TGGTGGGAAGAKGAGGACAGTAGG	Rabbit Igκ reverse primer mix (90% of mix)
RIGκC2	TGGTGGGAAGAKGAGGACACTAGG	Rabbit Igκ reverse primer mix (5% of mix)
RIGκC3	TGGTGGGAAGAKGAGGACAGAAGG	Rabbit Igκ reverse primer mix (5% of mix)
RIGλC1	CAAGGGGGCGACCACAGGCTGAC	Rabbit Igλ reverse primer mix (equimolar)
RIGλC2	GTGAAGGAGTGACTACGGGTTGACC	Rabbit Igλ reverse primer mix (equimolar)
RIGλC3	GAGGGGGTCACCGCGGGCTGAC	Rabbit Igλ reverse primer mix (equimolar)
RLR1	GATGACGATGCGGCCCCCGAGGCCTTGATTTCYACMTTGGTGCCAG	Rabbit VL repertoire reverse primer mix (equimolar)
RLR2	GATGACGATGCGGCCCCCGAGGCCTYGACSACCACCTCGGTCCCTC	Rabbit VL repertoire reverse primer mix (equimolar)
FLR1	GGTGGTGGTGGTAGCGGTGGTGGTGGCAGCGMNNHHGWDMTGACCCAGACTS	Rabbit VL repertoire forward primer
VHF-QQL	GGCCCAGCCGGCCATGGCTCAGCAGCTGGAAG	scFv gene forward primer(s)
VHF-QEQ	GGCCCAGCCGGCCATGGCTCAGGAACAGCTG	scFv gene forward primer(s)
VHF-QSL	GGCCCAGCCGGCCATGGCTCAGTCTCTGGAAG	scFv gene forward primer(s)
OE-R	GATGACGATGCGGCCCCCGAG	scFv gene reverse primer

Table S2. Total reads and numbers of unique full V gene amino acid sequences and CDRH3 amino acid sequences obtained from different B cell samples for the CCH rabbit and the BSA rabbit. Numbers in parentheses indicate the number of unique sequences that have n≥2 reads.

Animal	Source	454 reads	Unique full v region sequences	Unique CDR3 sequences
Rabbit-CCH	PBC V_H	106,325 (36,044)	81,862 (11,589)	21,120 (4,725)
	PBC (CD138+) V_H	21,945 (17,236)	7,116 (2,407)	2,235 (1,038)
	BM (CD138+) V_H	31,136 (1,6821)	19,044 (4,729)	8,541 (3,146)
	BM (CD138+) V_L	29,998 (22,353)	12,120 (4,475)	6,521 (3,792)
Rabbit-BSA	PBC (CD138+) V_H	28,931 (19,512)	12,743 (3,324)	5,285 (2,231)

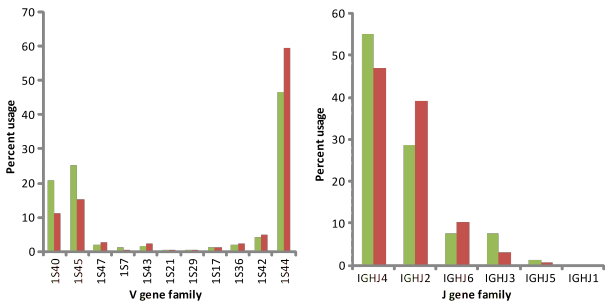

Fig. S1. Germline V and J usage for the BSA rabbit: Bar graphs represent V_H germline family usage (**left**) and J_H germline family usage (**right**) in the IgG repertoire from the BSA rabbit's peripheral blood B-cells (PBC's, N=3,324 unique, full length V_H amino acid sequences, red bars) or based on the family usage of full V-genes corresponding to *i*CDRH3 peptides identified by LC-MS/MS derived from antigen (BSA) enrichment $F(ab)_2$ from affinity chromatography elution (N= 455 indicate the total unique CDRH3 as identified by *i*CDRH3 peptides used for proteomic analysis, green bars).

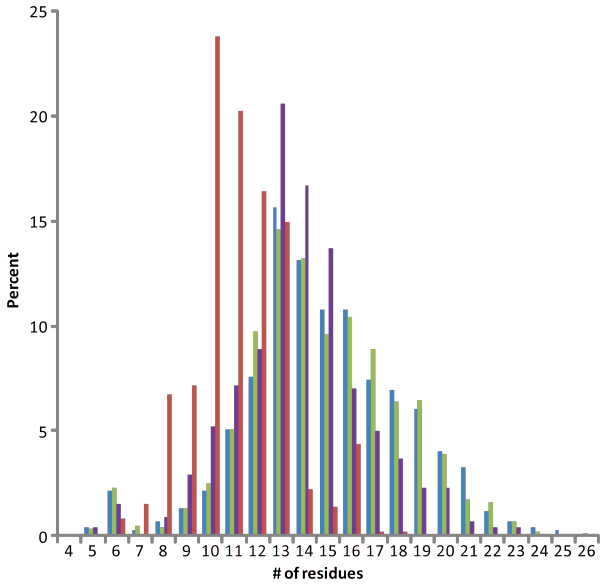

Fig. S2. CDR3 Length Distribution comparison: Bar graph representing CDRH3 length distribution in the IgG repertoire from $CD138^{+}$ bone marrow plasma cells (BM-PC, N=4,729 unique, full length amino acid sequences blue bars), peripheral B-cells (PBC's, N=2,788 unique, full length amino acid sequences, green bars) from a rabbit immunized with CCH, peripheral B-cells (PBC's, purple bars, N=3,324 unique, full length amino acid sequences) from an unimmunized rabbit with high titer to BSA (titer $>1{:}10^{5}$) and BM-PC's (N=6,422 unique, full length amino acid sequences, red bars) from a mouse immunized with Hen Egg White Lysozyme (HEL).

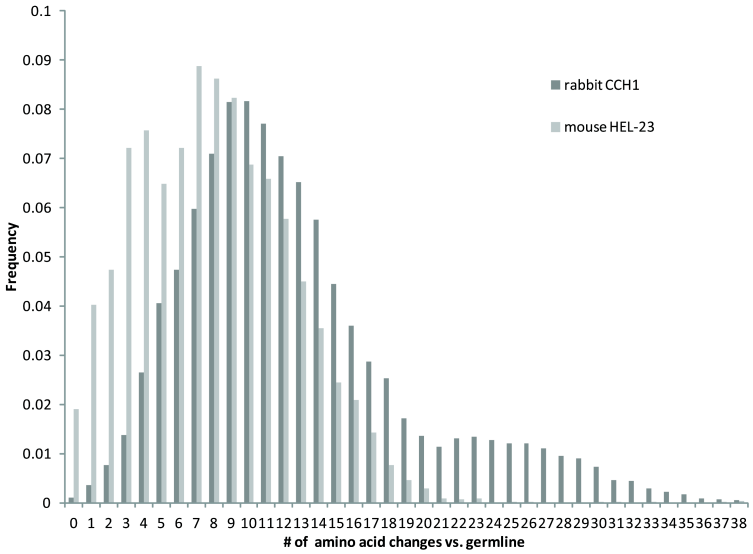

Fig. S3. Frequency of amino acid substitutions in V gene from BM-PCs relative to the rabbit and mouse germline V_H repertoires: The distribution of the number of mutations identified per sequence in unique full length VH sequences in the 454 database for bone marrow PCs (N=4,729 for CCH1, N=6,422 for HEL-23). These mutations are reflective of both SHM and gene conversion events as identified by IMGT based upon nucleotide sequence alignment with the set of germline V_H genes in each respective species.

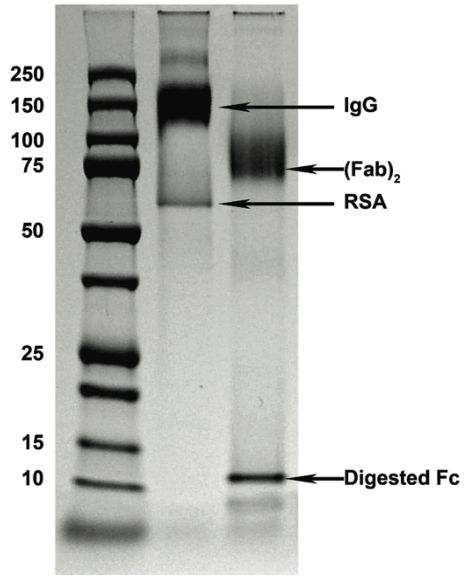

Fig. S4. SDS-PAGE (4-20%) of purified IgG from CCH immunized rabbit serum and F(ab)$_2$ product after pepsin digestion.

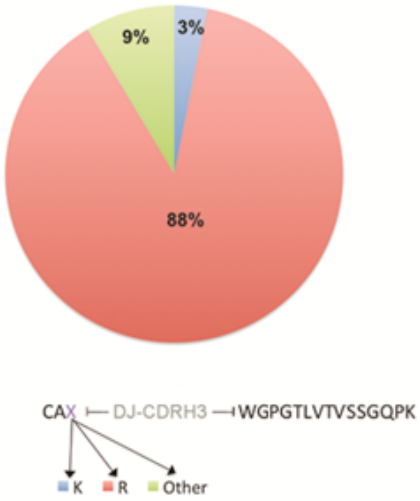

Fig. S5. Occurrences of tryptic sites (K/R) flanking the CDRH3 region. A set of 4,914 high confidence CCH1 V_H sequences was digested *in silico* using trypsin and the number of peptides containing the CDRH3 sequence was determined. Data were derived from the CCH rabbit V_H gene repertoire. X represents the potential trypsin cleavage site in rabbit V_H N-terminal to the CDRH3, which contains the amino acid R/K in 91% of the instances.

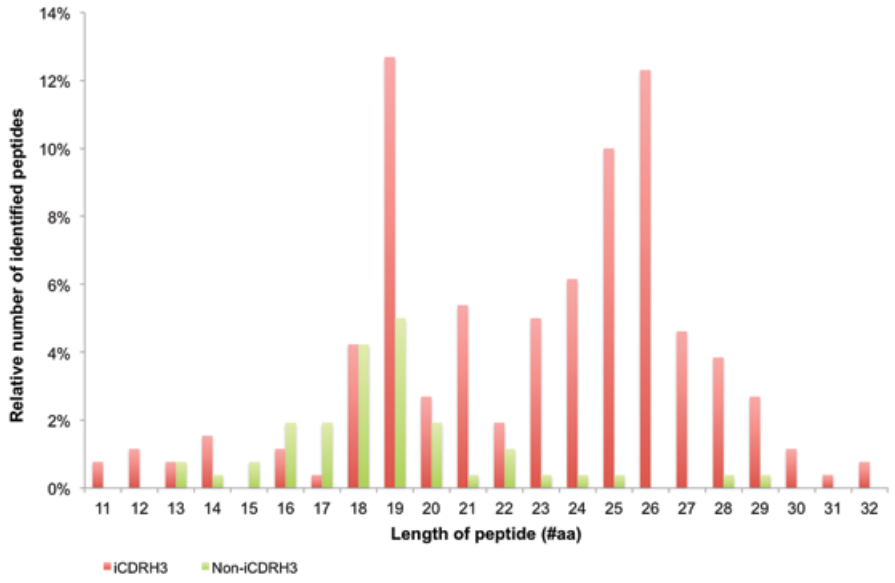

Fig. S6. **Distribution of peptide lengths identified by LC-MS/MS from CCH immunized rabbit serum.** *i*CDRH3 refers to peptides corresponding to n=1 CDRH3 in the transcript sequencing data and non-*i*CDRH3 refers to peptides corresponding to n>1 CDRH3 in the transcript sequencing data.

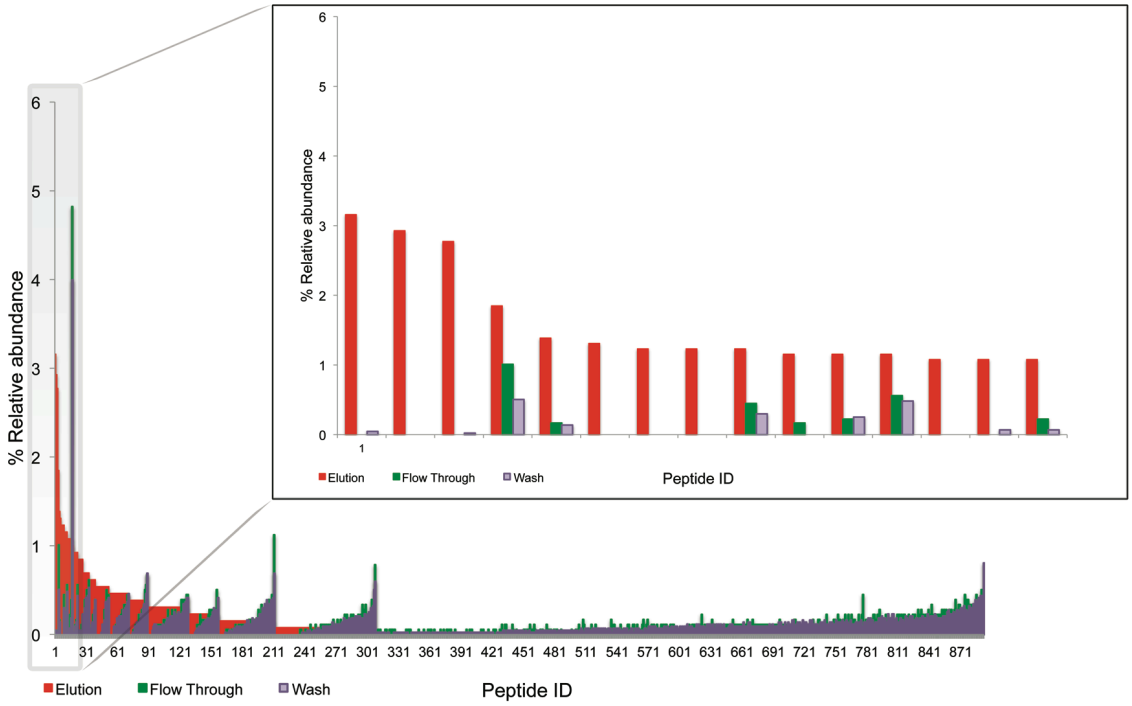

Fig. S7. Histogram showing frequencies of identified *i*CDRH3 peptides in the antigen affinity chromatography elution fraction (BSA rabbit). **Insert:** magnified histogram of the top 15 highest count peptides in the antigen affinity chromatography elution fraction. A total of 18 *i*CDRH3 peptides were found exclusively in the elution fraction. Peptide IDs are ranked by relative abundance in elution. Peptides with low abundance in the affinity chromatography elution fraction and are found overwhelmingly in the flow through and wash buffer fractions likely correspond to antibodies that bind the antigen very weakly or non-specifically.

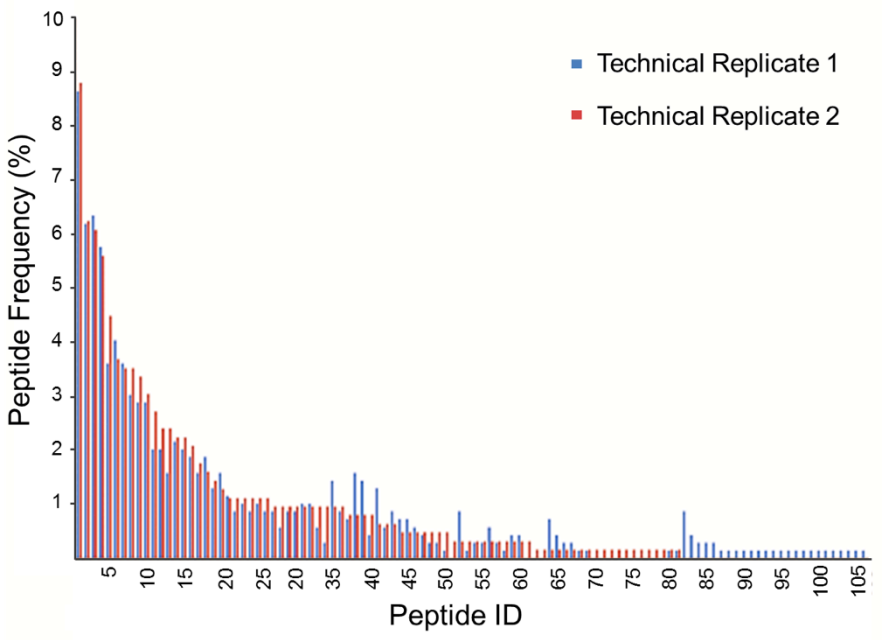

Fig. S8. **Histogram demonstrating the reproducibility of CCH immunized rabbit sample and LC-MS/MS analysis.** Peptides identified by LC-MS/MS derived from antigen (CCH) enrichment $F(ab)_2$ from affinity chromatography elution. Technical replicates of elution samples were prepared in parallel and independently, according to the proteomic pipeline (**Fig. 1**) for preparation and sequence assignment of information-rich CDRH3 peptides.

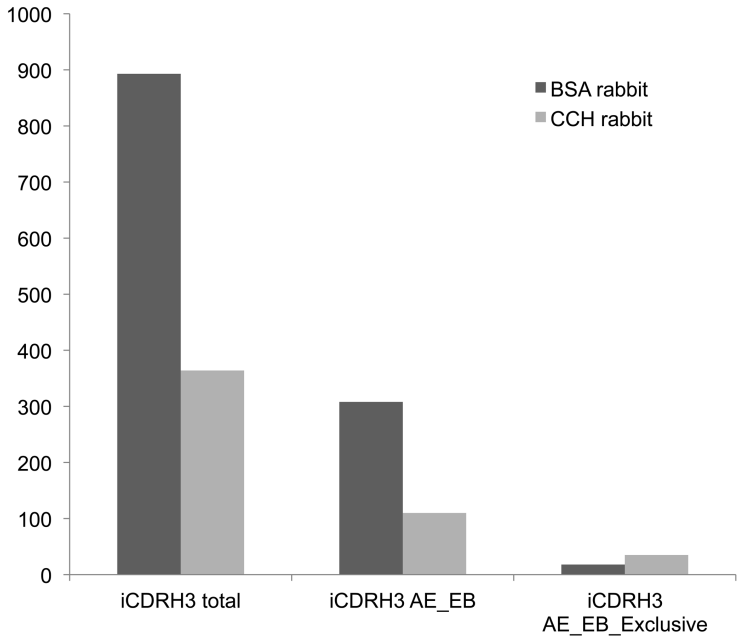

Fig. S9. Number of distinct *i*CDRH3 peptides identified by LC-MS/MS derived from antigen (CCH, BSA) enrichment of $F(ab)_2$ from affinity chromatography elution fractions. *i*CDRH3 total corresponds to the total number of informative CDRH3 peptides observed in all fractions (flow through, wash and elution); *i*CDRH3 AE_EB correspond to peptides identified in the elution fraction regardless whether they appear in the flow through and wash fractions; *i*CDRH3 AE-EB_Exclusive are LC-MS/MS identified peptides which were observed exclusively in the elution fractions.

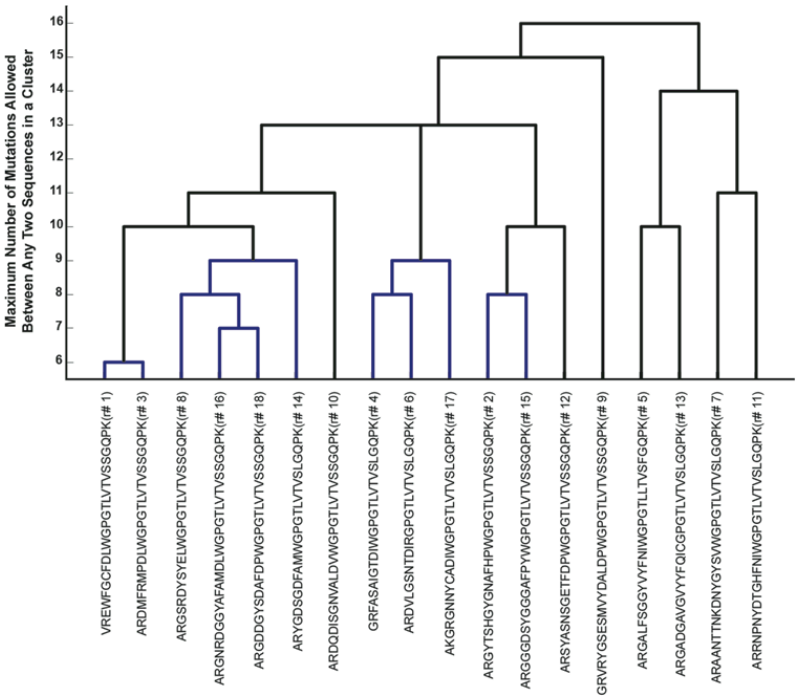

Fig. S10. Pairwise alignment of CDRH3s defined by *i*CDRH3s peptides from an unimmunized rabbit showing a high BSA titer and observed exclusively in the affinity chromatography elution. The dendrogram shows hierarchical clustering of CDRH3 sequences based on pairwise sequence alignments at the amino acid level. At each level of hierarchy, CDRH3 sequences are successively clustered into groups by increasing the maximum number of mutations allowed between any two aligned sequences in a cluster. The most similar CDRH3 sequences cluster first at a stringent threshold, which allows only 6 mutations between any pair of aligned sequences. The majority of the remaining CDRH3 sequences are grouped into clusters only at a very permissive threshold of 10 or 12 amino acid differences.

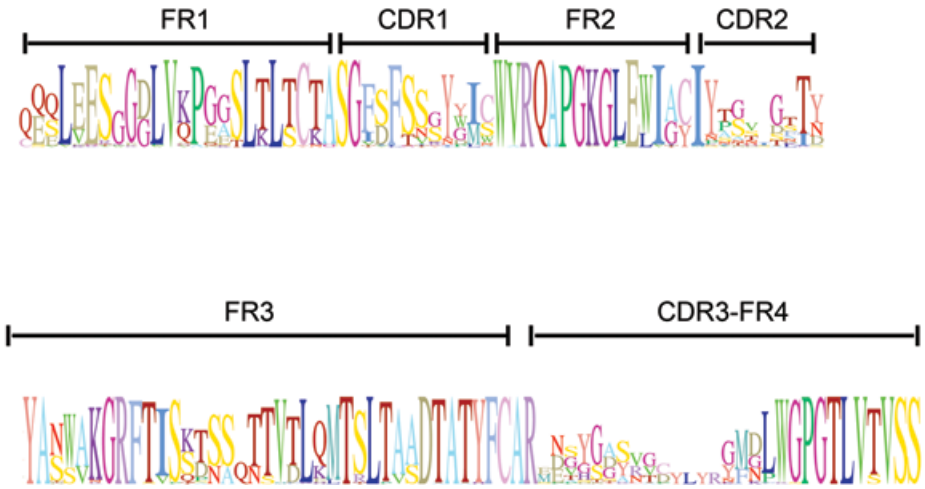

Fig. S11. V_H sequence logo defined by the top 12 *i*CDRH3 (CCH rabbit) highest count peptides. V_H genes corresponding to *i*CDRH3s were determined from the reconstructed V gene repertoire.

Table S3. Highest count *i*CDRH3 peptides and oxidative post-translational modifications in the BSA rabbit. *i*CDRH3 peptides from unimmunized rabbit, showing high BSA titer, detected by MS; frequencies of *i*CDRH3's relative to all spectral counts in the eluent; full length CDRH3 sequences and number of somatic variants deduced from the V_H DNA sequence database (all sequences derived from peripheral B cells). *i*CDRH3 peptides detected in the affinity chromatography eluent are marked as 'E', wash buffer are marked as 'W' or flow through as 'F'; *V_H synthesized for phage panning.

Rank/ Name	*i*CDRH3 MS sequence	Peptide freq (%)	Full CDRH3 transcript sequence	Identified in affinity chromatography fraction (Elution – E, Wash –W, Flow Through – F)	# Somatic Variants
1	LVTPGTPLTLTCTVSGFSLSSFDMSWVR	3.16	ARASVGNSHDIWGPGTLVTVSLGQPK	E+W	1
*2	EWFGCFDLWGPGTLVTVSSGQPK	2.93	VREWFGCFDLWGPGTLVTVSSGQPK	E	1
3	DSKLWGPGTLVTVSSGQPK	2.78	ARYGNLYRDSKLWGPGTLVTVSSGQPK	E+W	3
4	NTYAGGIDAGLTR	1.85	ARNTYAGGIDAGLTRLDLWGQGTLVTVSSGQPK	E+W+F	17
5	LVTPGTPLTLTCTVSGFSLSSYAMNWVR	1.39	ARVGDDYGDDAFDPWGPGTLVTVSSGQPK	E+W+F	2
6	GYTSHGYGNAFHPWGPGTLVTVSSGQPK	1.31	ARGYTSHGYGNAFHPWGPGTLVTVSSGQPK	E	1
*7	MPDLWGPGTLVTVSSGQPK	1.23	ARDMFRMPDLWGPGTLVTVSSGQPK	E	1
*8	FASAIGTDIWGPGTLVTVSLGQPK	1.23	GRFASAIGTDIWGPGTLVTVSLGQPK	E	3
9	VTSPTTEDTATYFCGR	1.23	GRGRSSHTSIHGFDIWGPGTLVTVSLGQPK	E+W+F	2
10	ITSPTTEDTATYFCSR	1.16	SRGDSGGWDAFSAIWGPGTLVTVSLGQPK	E+F	1
11	LVTPGTPLTLTCTASGFSLSTYHMGWVR	1.16	ARRNPNYDTGHFNIWGPGTLVTVSLGQPK	E+W+F	3
12	NVSPANWDYFDLWGPGTLVTVSSGQPK	1.16	VRRNVSPANWDYFDLWGPGTLVTVSSGQPK	E+W+F	1
13	GLEWIGMIDSTAGTYYASWAK	1.08	ARGALFSGGYVYFNIWGPGTLLTVSFGQPK	E	1
14	DVLGSNTDIWGPGTLVTVSLGQPK	1.08	ARDVLGSNTDIWGPGTLVTVSLGQPK	E+W	2
15	VTSLTTDDTATYFCAR	1.08	ARSANGAAGKGFDIWGPGTLATVSLGQPK	E+W+F	1

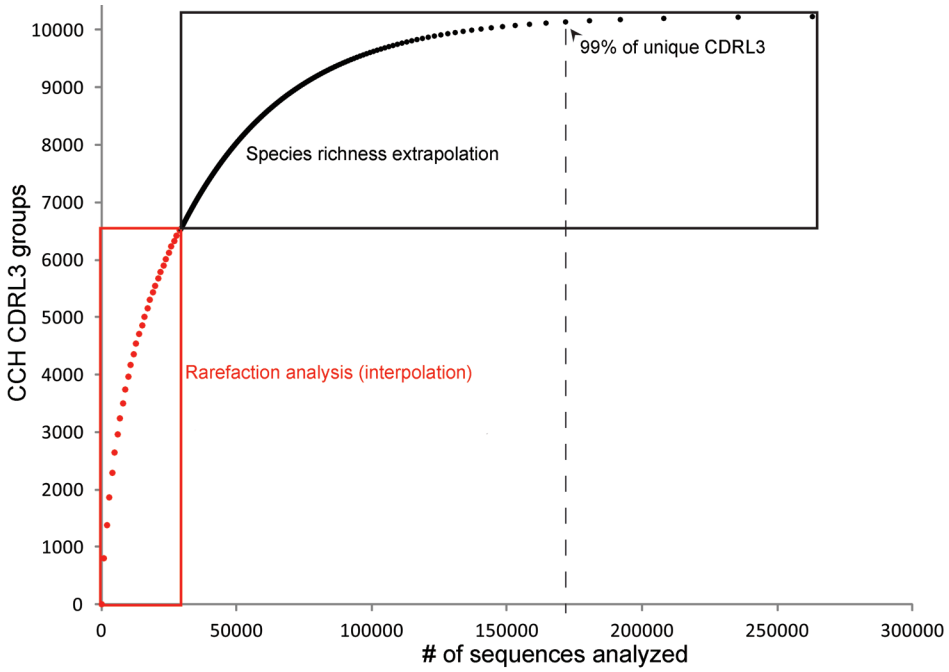

Fig. S12. Rarefaction analysis (subsampling and interpolation) and species richness analysis (extrapolation) of the CDRL3 repertoire in rabbit CCH. The rarefaction analysis was completed using the Vegan package(14) as implemented in the statistical environment R, by subsampling ~30,000 reads at 1000 read increments. A Chao(15) estimate of 10,252 unique CDRL3 sequences represents the estimated CDRL3 repertoire size. Species richness estimation was used to determine the number of clones that the libraries require to sample >99% of all unique CDRL3. This analysis showed that a library size of ~10^6 would be sufficient to capture >99% of the asymptotic estimate of CDRL3 diversity.

Table S4. Phage titers and polyclonal phage binding selectivity (phage titer against CCH/phage titer against BSA – note: this rabbit showed no titer towards BSA) after different rounds of phage panning scFv libraries comprising each of the proteomically identified V_H genes paired with a combinatorial library of V_L cDNA from that animal (CCH rabbit).

	Round 1		**Round 2**		**Round 3**	
Library	Titer	Selectivity	Titer	Selectivity	Titer	Selectivity
1	$4x10^4$	0.4	$5x10^4$	25	$6x10^3$	>4
3	$4x10^4$	0.4	$2x10^6$	>2000	$2x10^6$	>1300
6	$7x10^4$	1.4	$5x10^4$	12.5	$1x10^7$	400
9	$7x10^4$	1.4	$7x10^6$	1167	$3x10^8$	32,000

Fig. S13. **V_H-V_L sequences of functional scFvs obtained by phage panning of libraries of proteomically identified VH genes paired with the VL repertoire (CCH rabbit).**

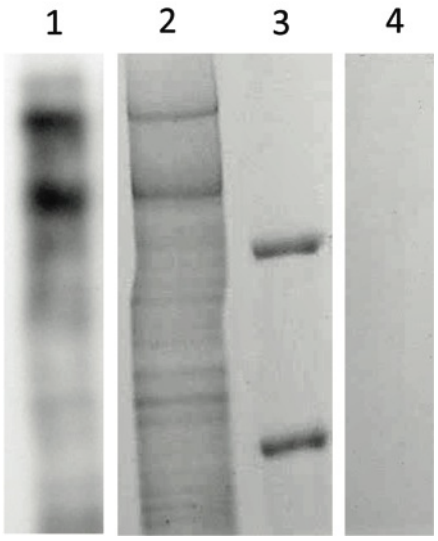

Fig. S14. Mock-immunoprecipitation of CCH from cell lysate demonstrates the utility of the IgG-3 antibody in pulling down the CCH antigen and its proteolytic fragments. Lane 1 – CCH antigen detected by Western blotting using affinity purified polyclonal $F(ab)_2$. The two chains of CCH, CCH-A, CCH-B, and proteolytic fragments are evident in the commercial CCH preparation. Lane 2 –SDS-PAGE of a mock immunoprecipitation of CCH mixed with a 20-fold excess of *E. coli* cell lysate using the IgG-3 antibody (V_H and V_L sequences in **Fig. S13**). $1/3^{rd}$ volume of the total elution from the IP was loaded on a 4% gel. Lane 3 –250 kD and 150 kD protein standards. Lane 4 – control IP without CCH-specific antibody

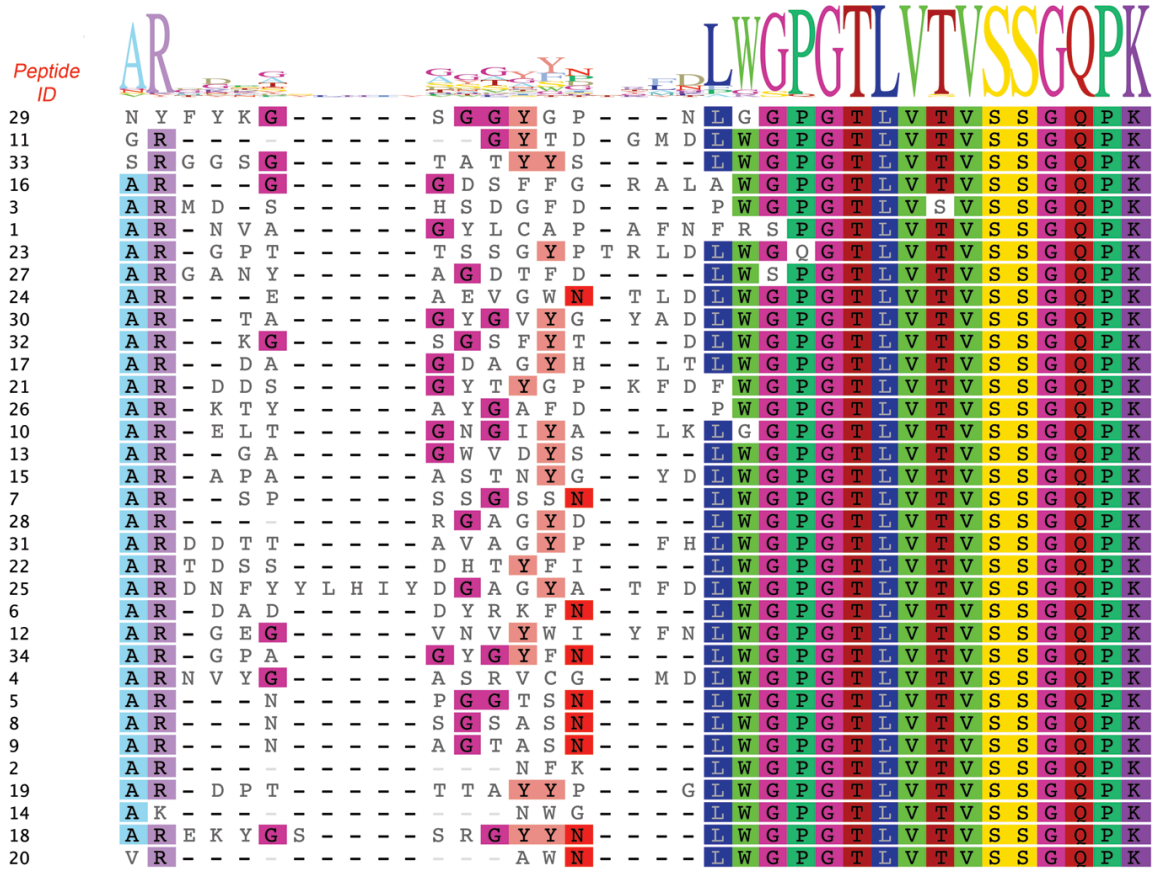

Fig. S15. Multiple sequence alignment and sequence logo of the CDRH3s identified from *i*CDRH3 peptides found exclusively in the affinity chromatography elution fraction.

REFERENCES

1. Reddy ST*, et al.* (2010) Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat Biotechnol* 28(9):965-969.
2. Wang H, Moore S, & Alavi MZ (1997) Expression of syndecan-1 in rabbit neointima following de-endothelialization by a balloon catheter. *Atherosclerosis* 131(2):141-147.
3. Lefranc M-P (2008) IMGT®, the International ImMunoGeneTics Information System® for Immunoinformatics. *Mol Biotechnol* 40(1):101-111.
4. Kall L, Canterbury JD, Weston J, Noble WS, & MacCoss MJ (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Meth* 4(11):923-925.
5. Cox J, Michalski A, & Mann M (2011) Software Lock Mass by Two-Dimensional Minimization of Peptide Mass Errors. *J Am Soc Mass Spectrom* 22(8):1373-1380.
6. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-1797.
7. Needleman SB & Wunsch CD (1970) A general method applicable to search for similarities in amino acid sequence of 2 proteins. *J Mol Biol* 48(3):443-453.
8. Cox JC, Lape J, Sayed MA, & Hellinga HW (2007) Protein fabrication automation. *Protein Sci* 16(3):379-390.
9. Hayhurst A*, et al.* (2003) Isolation and expression of recombinant antibody fragments to the biological warfare pathogen Brucella melitensis. *J Immunol Methods* 276(1-2):185-196.
10. Gao X, Yo P, Keith A, Ragan TJ, & Harris TK (2003) Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences. *Nucleic Acids Res* 31(22):e143.
11. Villalobos A, Ness J, Gustafsson C, Minshull J, & Govindarajan S (2006) Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* 7(1):285.
12. Krebber A*, et al.* (1997) Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J Immunol Methods* 201(1):35-55.
13. Bullock WO, Fernandez JM, & Short JM (1987) XL1-Blue - a high-efficiency plasmid transforming recA *escherichia-coli* strain with beta-galactosidase selection. *Biotechniques* 5(4):376-379.
14. Dixon P (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* 14(6):927-930.

15. Chao A, Colwell RK, Lin C-W, & Gotelli NJ (2009) Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* 90(4):1125-1133.
16. Vieira J & Messing J (1987) Production of single-stranded plasmid DNA. *Methods Enzymol* 153:3-11.