

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}

^aDepartment of Chemical Engineering, ^bInstitute for Cellular and Molecular Biology, ^cCenter for Systems and Synthetic Biology, ^dApplied Research Laboratories, ^eDepartment of Biomedical Engineering, ^fDepartment of Chemistry and Biochemistry, and ^gSection of Molecular Genetics and Microbiology, University of Texas at Austin, Austin, TX 78712-1062; and ^hDepartment 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 $>1 \mu\text{g/mL}$ 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.

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

²To 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.

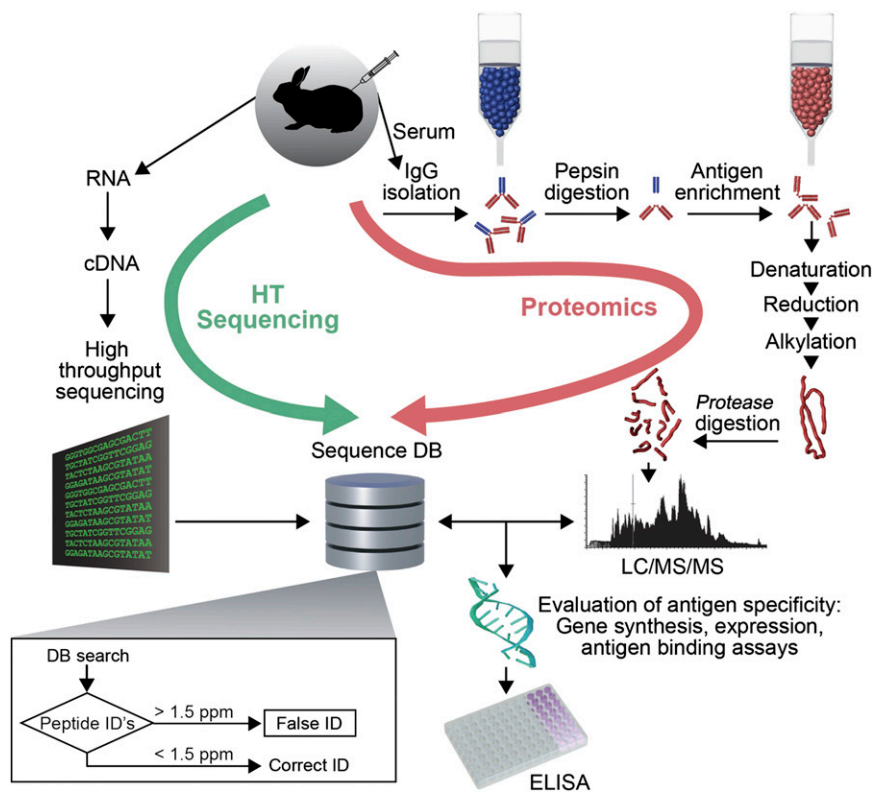


Fig. 1. Schematic of the workflow for serum Ig deconvolution. (Left) V gene repertoire sequencing pipeline: total RNA from desired B-cell subpopulations is reverse transcribed and amplified by 5' RACE with IgG-specific (V_H) or Ig_κ/Ig_λ-specific (V_L) primers and sequenced by Roche 454 sequencing. Reads 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 κ /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. 2A). 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

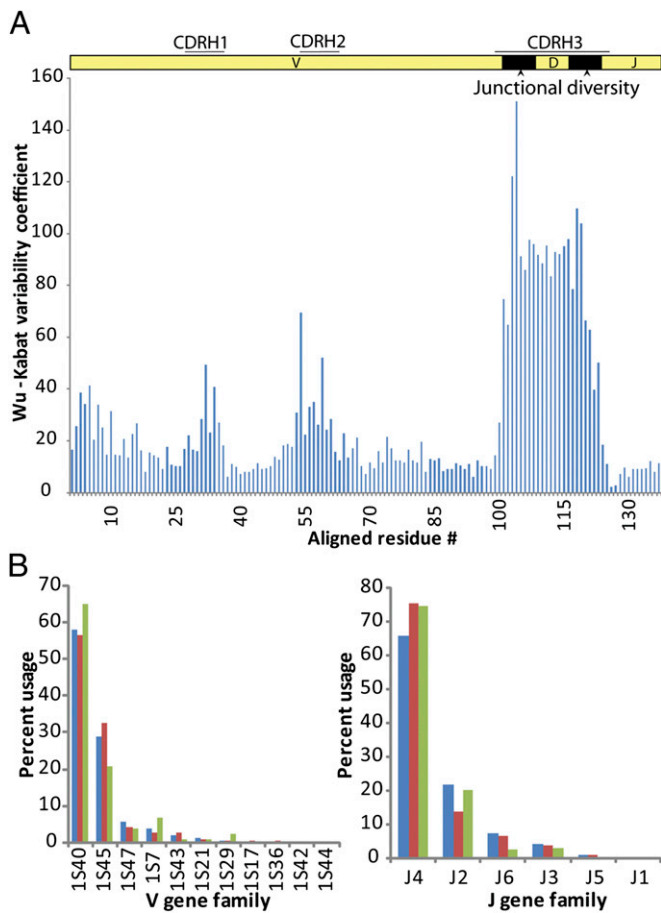


Fig. 2. Immunoglobulin heavy chain variable region (V_H) gene and circulating antibody repertoire characteristics. (A) Wu-Kabat variability plot representing the variability of the V_H genes is shown on a residue-by-residue basis; (B) bar graphs representing V_H germ-line family use (Left) and J_H germ-line family use (Right) in the V_H gene 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. 2B) 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 informative of CDRH3 (*i*CDRH3) peptides; in contrast, peptides mapping to multiple CDRH3 sequences were defined as noninformative (*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>i</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)	ARNVAGYLCAPAFNFRSPGTLTVSSGQPK	E	1	BM-PC
2*	NFKLWGPGLTVSSGQPK	88 (6.11)	ARNFKLWGPGLTVSSGQPK	E	3	BM-PC, PBC
3*	MDSHSDGDFPWGPGTLVSVSSG QPK	82 (5.70)	ARMDSHSDGDFPWGPGTLVSVSSGQPK (51%)	E	7	PBC
4	FTISSDNAQNTVDLK	64 (4.45)	AREGYGGYVGYMGLWGPGLTVSSGQPK	E+W+F	3	BM-PC, PBC
5*	VCGMDLWGPGLTVSSGQPK	50 (3.48)	ARNVYGASRVCGMDLWGPGLTVSSGQPK (43%)	E	2	BM-PC
6*	NPGGTSNLWGPGLTVSSGQPK	47 (3.27)	ARNPGGTSNLWGPGLTVSSGQPK	E	1	BM-PC
7*	KFNLWGPGLTVSSGQPK	44 (3.05)	ARDADDYRKFNLWGPGLTVSSGQPK	E	1	PBC
8	AFNLWGPGLTVSSGQPK	43 (3.00)	ARDVGYGNDNYRAFNLWGPGLTVSSGQPK	E+W+F	1	PBC
9*	SPSSGSSNLWGPGLTVSSGQPK	39 (2.71)	ARSPSSGSSNLWGPGLTVSSGQPK	E	2	BM-PC
10	NSGSASNLWGPGLTVSSGQPK	31 (2.16)	ARNSGSASNLWGPGLTVSSGQPK	E	2	BM-PC
11	GMDLWGPGLTVSSGQPK	29 (2.02)	AREDTYGDANTDYLYRGMDLWGPGLTVS SGQPK	E+W	3	BM-PC, PBC
12	NAGTASNLWGPGLTVSSGQPK	28 (1.95)	ARNAGTASNLWGPGLTVSSGQPK	E	1	BM-PC
13	GLTAADTATYFCAR	28 (1.95)	ARDGIDNGYNDLNLWGPGLTVSSGQPK	E+F	1	BM-PC
14	ELTGNGIYALK	27 (1.88)	ARELTGNGIYALKLGGPGLTVSSGQPK	E	1	BM-PC
15	TSSTTVPLQMTSLAADTATYFCGR	27 (1.88)	GRGYTDGMDLGGPGLTVSSGQPK	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. 3A; 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. 2B) 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. 3B; 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

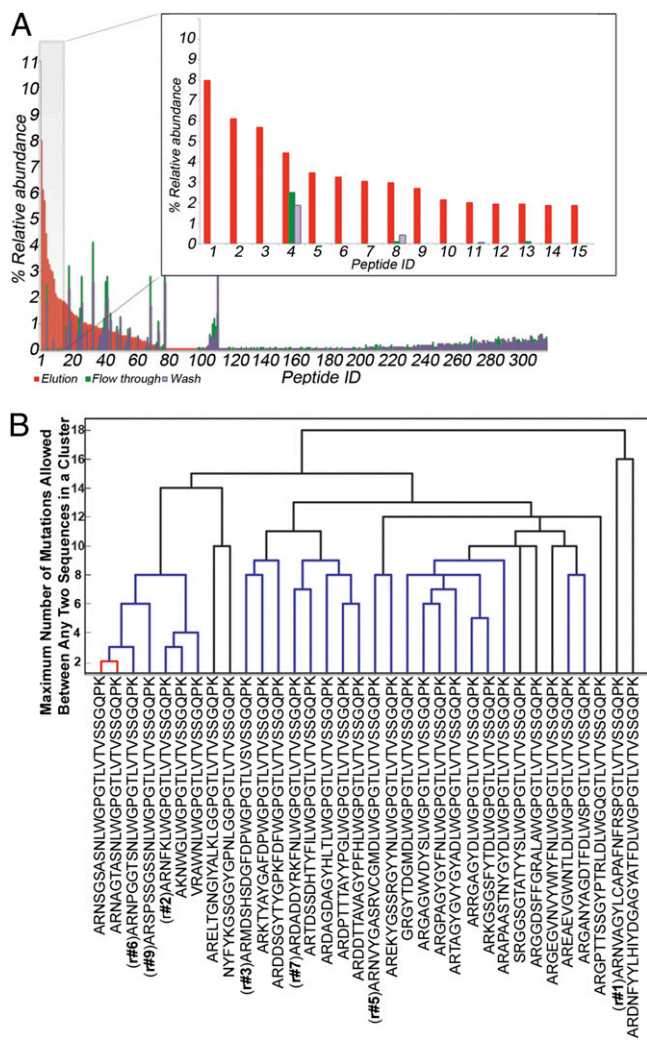


Fig. 3. Identified *i*CDRH3 peptides from affinity chromatography and alignment of corresponding CDRH3s. (A) Histogram showing frequencies of identified informative peptides corresponding to a unique clonotype (V genes with same V_H , J_H CDRH3 sequence and 80% homology in the CDRH3) in the antigen-affinity 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 V_H 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 [J_4 and J_2 in both CCH and BSA rabbits (*Fig. 2B* 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. 2B* 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

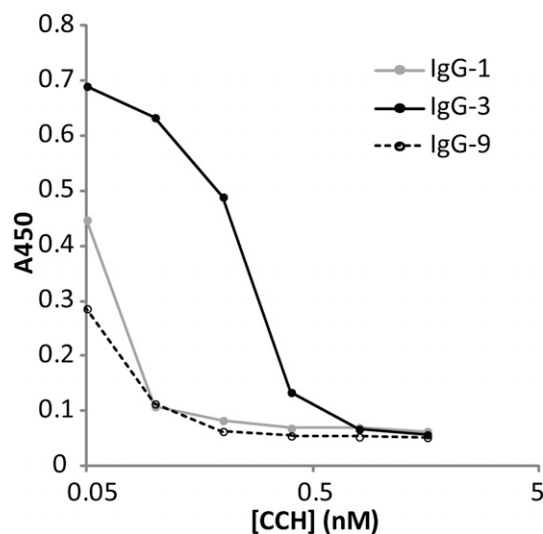


Fig. 4. Competitive ELISA of full-length IgG containing the V_H genes corresponding to selected abundant *i*CDRH3s (Table 1) identified in the CCH immunized rabbit. Antibodies identified using the proposed methodology (*Fig. 1*) were shown to display subnanomolar affinities. The expression yield of the 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 iCDRH3s 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 homoge-

nous 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 Excellence in Biodefense (NIH Grant 5U54AI057156) and the Texas Institute for Drug and Diagnostics Development (TI-3D).

- Browning CH (1955) Emil Behring and Paul Ehrlich: Their contributions to science. *Nature* 175(4457):570–575.
- 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.
- Radbruch A, et al. (2006) Competence and competition: The challenge of becoming a long-lived plasma cell. *Nat Rev Immunol* 6(10):741–750.
- 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.
- Wrarmert J, et al. (2008) Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* 453(7195):667–671.
- 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.
- 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.
- Bandeira N, Pham V, Pevzner P, Arnett D, Lill JR (2008) Automated de novo protein sequencing of monoclonal antibodies. *Nat Biotechnol* 26(12):1336–1338.
- Cheung WC, et al. (2012) A proteomics approach for the identification and cloning of monoclonal antibodies from serum. *Nat Biotechnol* 30(5):447–452.
- Reddy S, et al. Rapid Isolation of Monoclonal Antibodies from Animals. Patent 20110312505 (Filed May 17, 2011).
- 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.
- Sato S, et al. (2012) Proteomics-directed cloning of circulating antiviral human monoclonal antibodies. *Nat Biotechnol* 30(11):1039–1043.
- Murphy K, Travers P, Walport M, eds (2007) *Janeway's Immunobiology* (Garland Science, New York).
- 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.
- 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.
- 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.
- 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.
- 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.
- Knight KL (1992) Restricted VH gene usage and generation of antibody diversity in rabbit. *Annu Rev Immunol* 10:593–616.
- 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.
- Wu TT, Johnson G, Kabat EA (1993) Length distribution of CDRH3 in antibodies. *Proteins* 16(1):1–7.
- 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.
- Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass-spectrometric data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5(11):976–989.
- 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.
- 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.
- Marks JD, et al. (1992) By-passing immunization: Building high affinity human antibodies by chain shuffling. *Biotechnology (N Y)* 10(7):779–783.
- 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.
- Mallick P, Kuster B (2010) Proteomics: A pragmatic perspective. *Nat Biotechnol* 28(7):695–709.
- 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.
- 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.
- Scheid JF, et al. (2011) Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* 333(6049):1633–1637.
- Walker LM, et al.; Protocol G Principal Investigators (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 477(7365):466–470.
- Barnett BE, et al. (2012) Asymmetric B cell division in the germinal center reaction. *Science* 335(6066):342–344.
- Slifka MK, Antia R, Whitmire JK, Ahmed R (1998) Humoral immunity due to long-lived plasma cells. *Immunity* 8(3):363–372.
- 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.
- 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.