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)

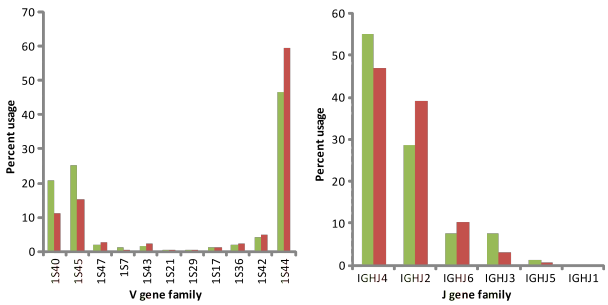

Fig. S1. Germline V and J usage for the BSA rabbit: Bar graphs represent V_H germline family usage (**left**) and J_H germline family usage (**right**) in the IgG repertoire from the BSA rabbit's peripheral blood B-cells (PBC's, N=3,324 unique, full length V_H amino acid sequences, red bars) or based 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).

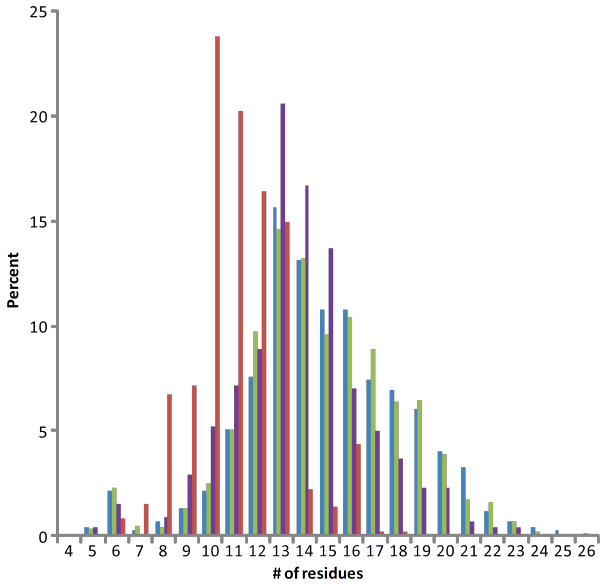

Fig. S2. CDR3 Length Distribution comparison: Bar graph representing CDRH3 length distribution in the IgG repertoire from $CD138^{+}$ bone marrow plasma cells (BM-PC, N=4,729 unique, full length amino acid sequences blue bars), peripheral B-cells (PBC's, N=2,788 unique, full length amino acid sequences, 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).

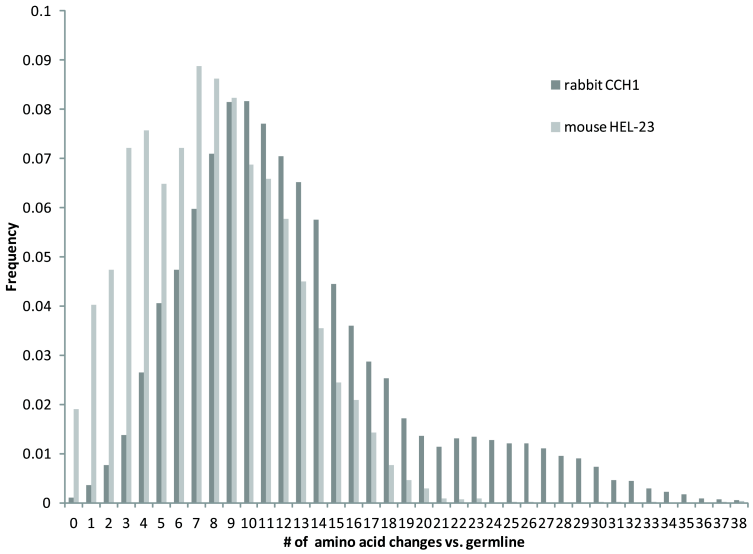

Fig. S3. Frequency of amino acid substitutions in V gene from BM-PCs relative to the rabbit and mouse germline V_H repertoires: The distribution of the number of mutations identified per sequence in unique full length VH sequences in the 454 database for bone marrow PCs (N=4,729 for CCH1, N=6,422 for 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.

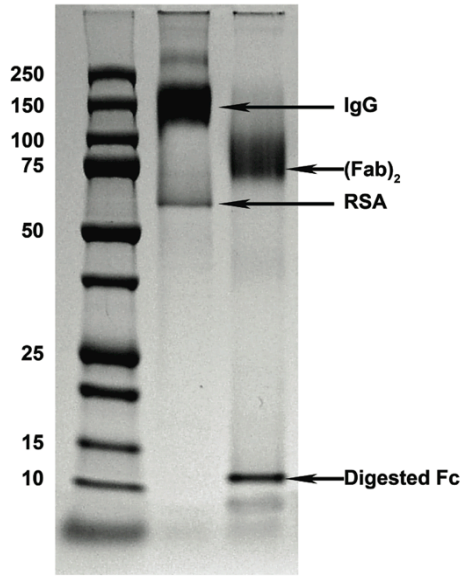

Fig. S4. SDS-PAGE (4-20%) of purified IgG from CCH immunized rabbit serum and F(ab)$_2$ product after pepsin digestion.

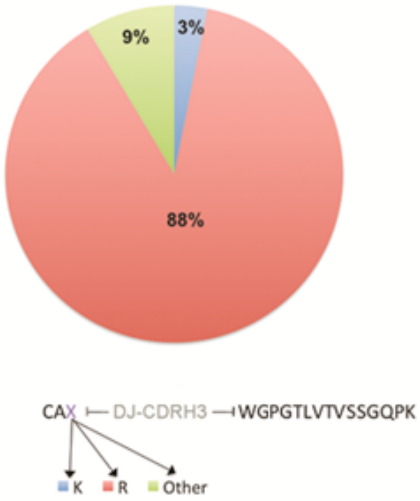

Fig. S5. Occurrences of tryptic sites (K/R) flanking the CDRH3 region. A set of 4,914 high confidence CCH1 V_H sequences was digested *in silico* using trypsin and the number of peptides containing the CDRH3 sequence was determined. Data were derived from the CCH rabbit V_H gene repertoire. X represents 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.

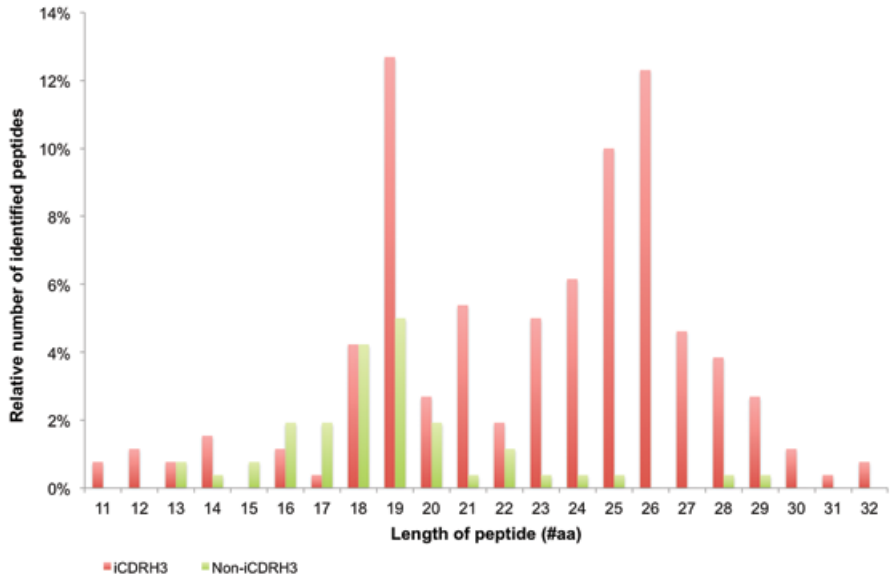

Fig. S6. **Distribution of peptide lengths identified by LC-MS/MS from CCH immunized rabbit serum.** *i*CDRH3 refers to peptides corresponding to n=1 CDRH3 in the transcript sequencing data and non-*i*CDRH3 refers to peptides corresponding to n>1 CDRH3 in the transcript sequencing data.

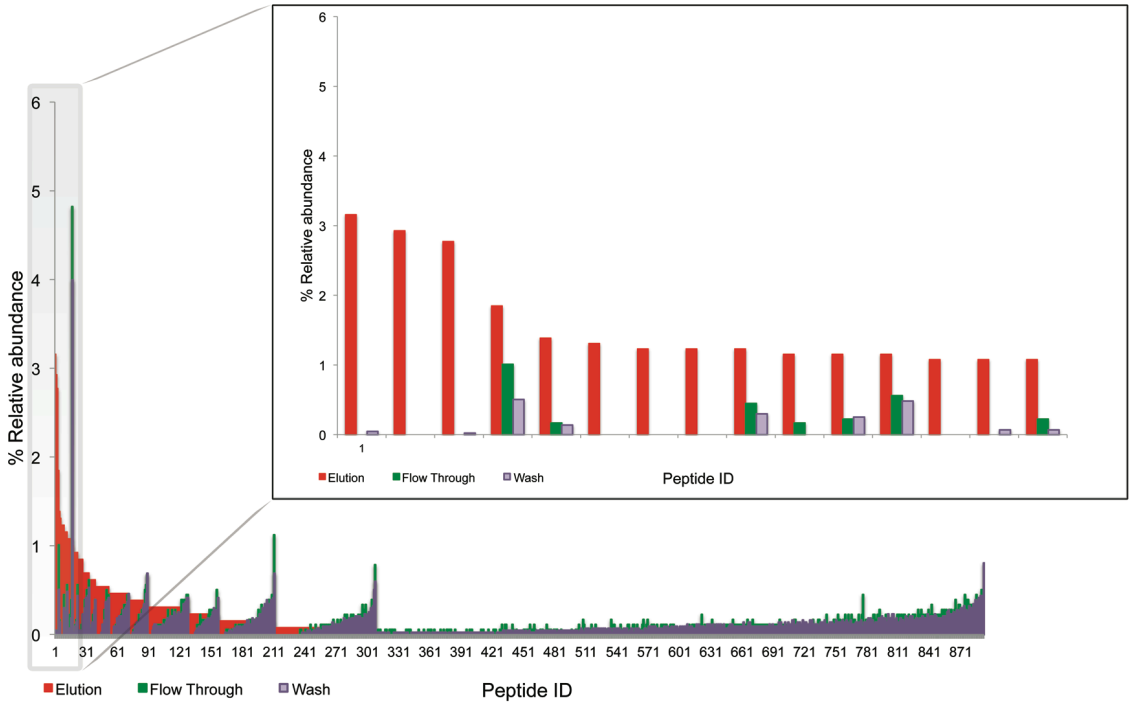

Fig. S7. Histogram showing frequencies of identified *i*CDRH3 peptides in the antigen affinity chromatography elution fraction (BSA rabbit). **Insert:** magnified histogram of the top 15 highest count peptides in the antigen affinity chromatography elution fraction. A total of 18 *i*CDRH3 peptides were 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.

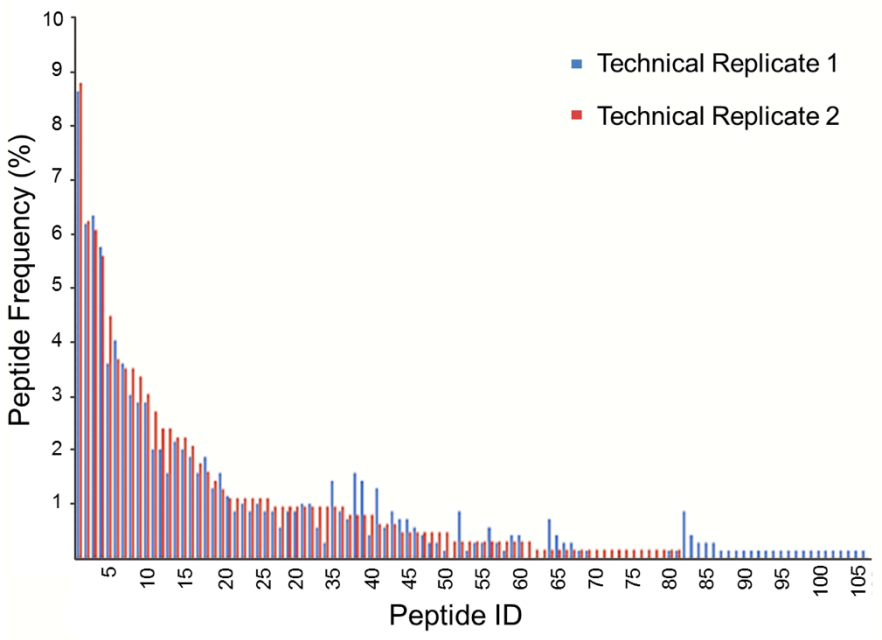

Fig. S8. **Histogram demonstrating the reproducibility of CCH immunized rabbit sample and LC-MS/MS analysis.** Peptides identified by LC-MS/MS derived from antigen (CCH) enrichment $F(ab)_2$ from affinity chromatography elution. Technical replicates of elution samples were prepared in parallel and independently,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

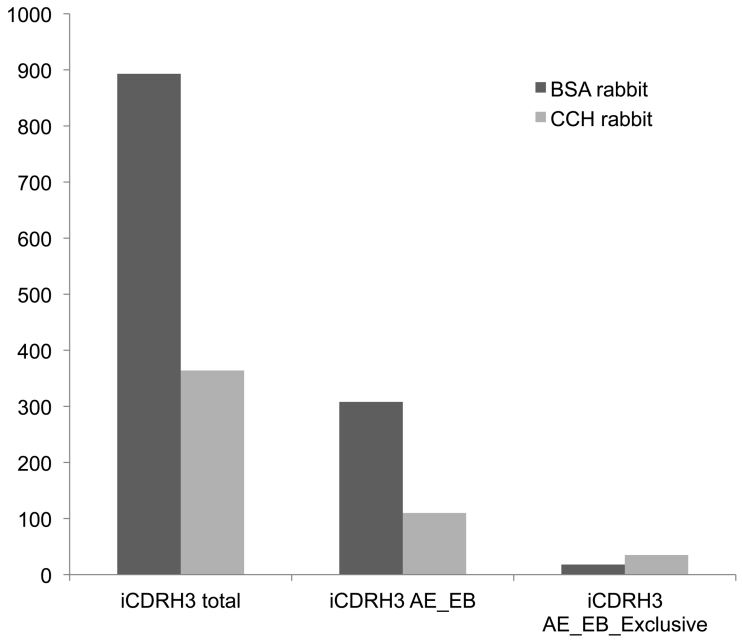

Fig. S9. Number of distinct *i*CDRH3 peptides identified by LC-MS/MS derived from antigen (CCH, BSA) enrichment of $F(ab)_2$ from affinity chromatography elution fractions. *i*CDRH3 total corresponds to the total number of informative CDRH3 peptides observed in all fractions (flow through, wash and elution); *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.

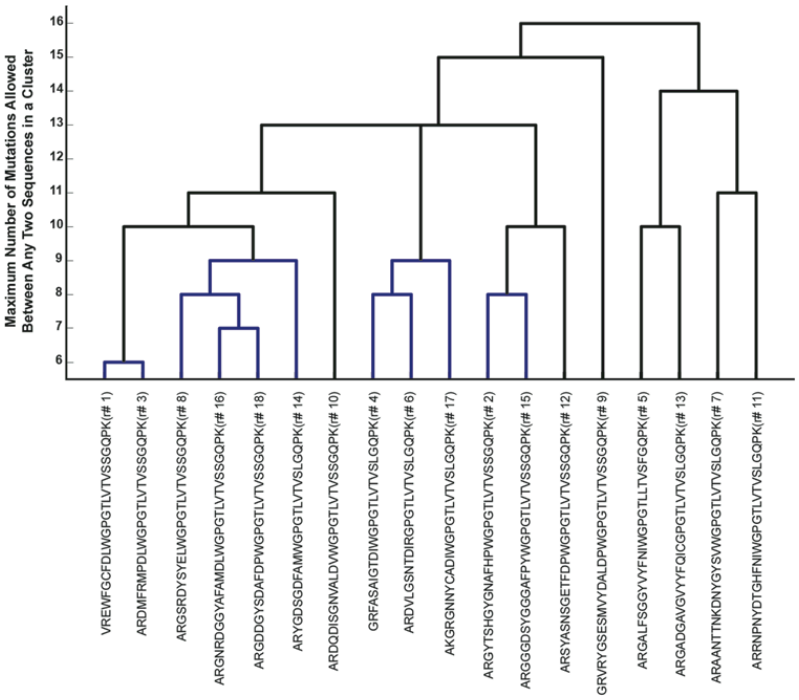

Fig. S10. Pairwise alignment of CDRH3s defined by *i*CDRH3s peptides from an unimmunized rabbit showing a high BSA titer and observed exclusively in the affinity chromatography elution. The dendrogram shows hierarchical clustering of CDRH3 sequences based on pairwise sequence alignments at the amino 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.

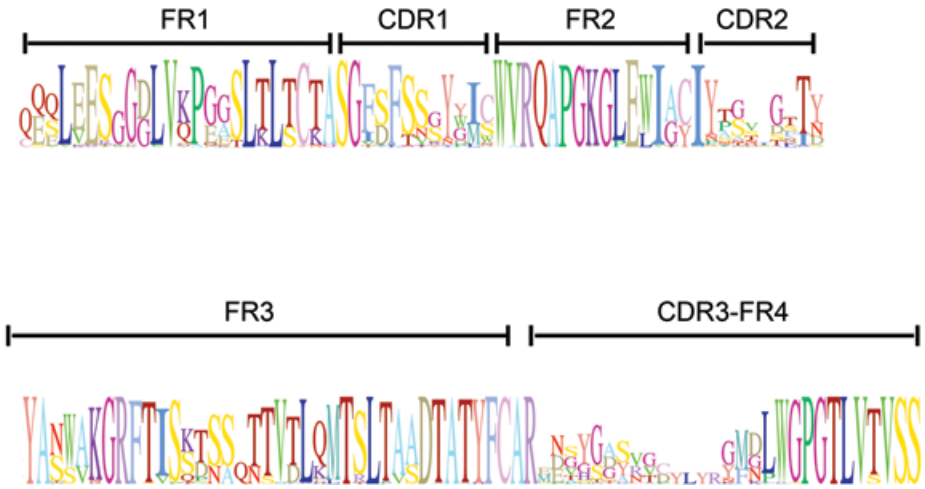

Fig. S11. V_H sequence logo defined by the top 12 *i*CDRH3 (CCH rabbit) highest count peptides. V_H genes corresponding to *i*CDRH3s were determined from the reconstructed V gene repertoire.

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

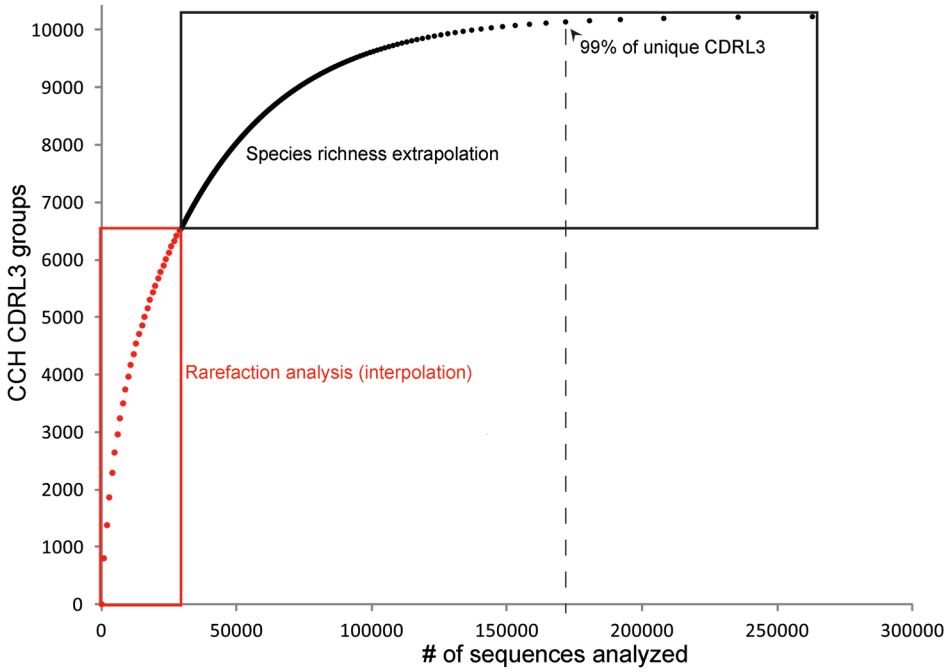

Fig. S12. Rarefaction analysis (subsampling and interpolation) and species richness analysis (extrapolation) of the CDRL3 repertoire in rabbit CCH. The rarefaction analysis was completed using the Vegan package(14) as implemented in the statistical environment R, by subsampling ~30,000 reads at 1000 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).**

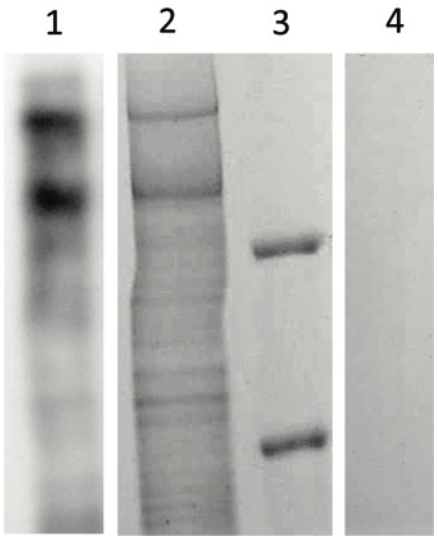

Fig. S14. Mock-immunoprecipitation of CCH from cell lysate demonstrates the utility of the IgG-3 antibody in pulling down the CCH antigen and its proteolytic fragments. Lane 1 – CCH antigen detected by Western blotting using affinity purified polyclonal $F(ab)_2$. The two chains of CCH, CCH-A, CCH-B, 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

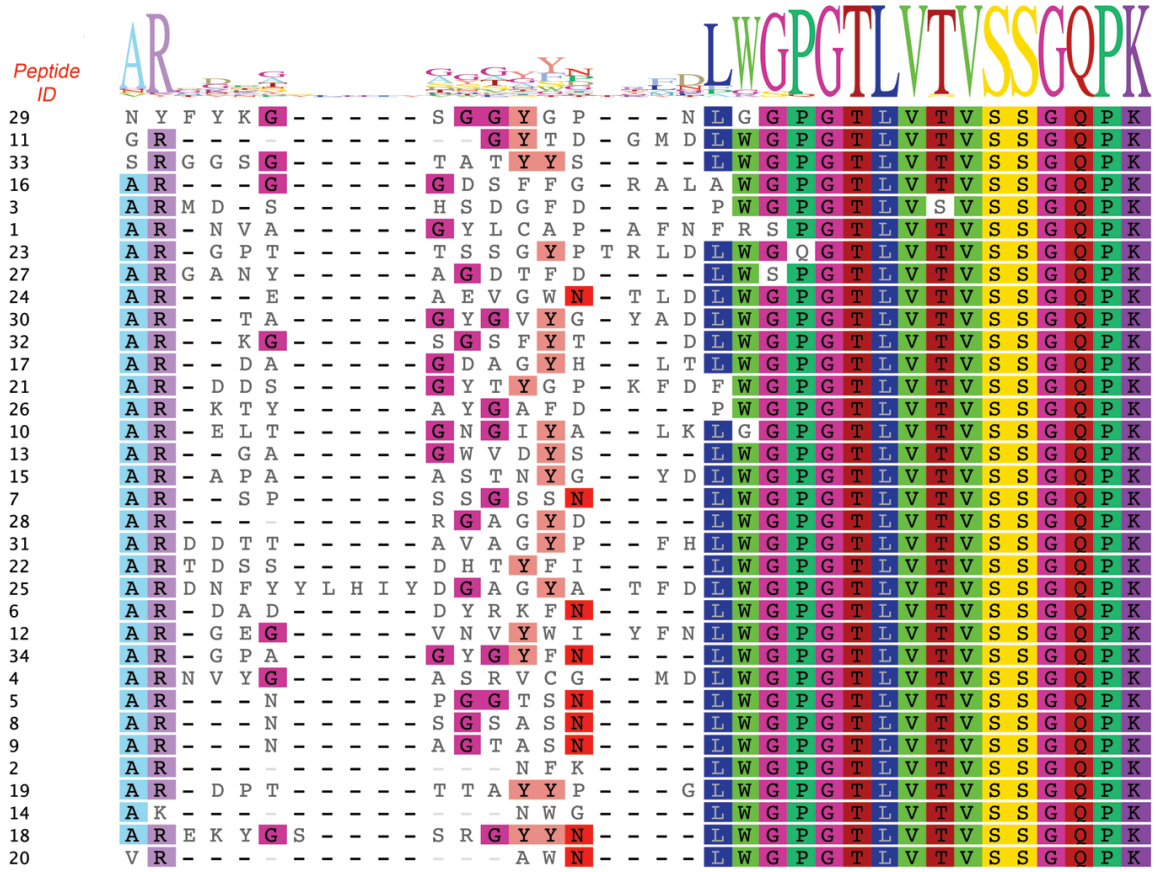

Fig. S15. Multiple sequence alignment and sequence logo of the CDRH3s identified from *i*CDRH3 peptides found exclusively in the affinity chromatography elution fraction.

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.