



Published in final edited form as:

J Mol Biol. 2013 February 22; 425(4): 803–811. doi:10.1016/j.jmb.2012.11.037.

CDR-H3 Diversity is Not Required For Antigen Recognition by Synthetic Antibodies

Helena Persson^{a,1}, Wei Ye^a, Amy Wernimont^b, Jarrett J. Adams^a, Akiko Koide^c, Shohei Koide^c, Robert Lam^b, and Sachdev S. Sidhu^{a,*}

^aBanting and Best Department of Medical Research and Department of Molecular Genetics, University of Toronto, The Donnelly Centre, 160 College Street, Toronto, Ontario M5S 3E1, Canada.

^bStructural Genomics Consortium, MaRS Centre, South Tower, 101 College St., Suite 700, Toronto, Ontario M5G 1L7, Canada.

^cDepartment of Biochemistry and Molecular Biology, The University of Chicago, 929

Abstract

A synthetic phage-displayed antibody repertoire was constructed with equivalent chemical diversity in the third complementarity-determining regions of the heavy (CDR-H3) and light chains (CDR-L3), which contrasts with natural antibodies in which CDR-H3 is much more diverse than CDR-L3 due to the genetic mechanisms that generate antibody encoding genes. Surprisingly, the synthetic repertoire yielded numerous functional antibodies that contained mutated CDR-L3 sequences but a fixed CDR-H3 sequence. Alanine-scanning analysis of antibodies that recognized ten different antigens but contained a common CDR-H3 loop showed that, in most cases, the fixed CDR-H3 sequence was able to contribute favorably to antigen recognition, but in some cases, the loop was functionally inert. Structural analysis of one such antibody in complex with antigen showed that the inert CDR-H3 loop was nonetheless highly buried at the antibody-antigen interface. Taken together, these results show that CDR-H3 diversity is not necessarily required for the generation of antibodies that recognize diverse protein antigens with high affinity and specificity, and if given the chance, CDR-L3 readily assumes the dominant role for antigen recognition. These results contrast with the commonly accepted view of antigen recognition derived from the analysis of natural antibodies, in which CDR-H3 is presumed to be dominant and CDR-L3 is presumed to play an auxiliary role. Furthermore, the results show that natural antibody function is genetically constrained, and it should be possible to develop more functional synthetic antibody libraries by expanding the diversity of CDR-L3 beyond what is observed in nature.

Keywords

antibody library; phage display; protein engineering; affinity; specificity

*To whom correspondence should be addressed. sachdev.sidhu@utoronto.ca, Telephone: 001-416-946-0863 .

¹Current affiliation: Department of Immunotechnology, Lund University, BMC D13, SE-221 84 Lund, Sweden.

Protein Data Bank accession numbers

The coordinates and structure factors for the TDRD3:Fab-8-1 complex have been deposited into the RCSB Protein Data Bank under accession code 3PNW.

Introduction

In recent years, phage-displayed synthetic antibody libraries have proven to be a powerful alternative to natural antibodies for generating research affinity reagents¹ and potential therapeutics.² Synthetic libraries are constructed from scratch, thus enabling the incorporation of desirable design features such as stable, human frameworks that enhance antibody performance and lower the risk of immunogenicity for therapeutic applications. Moreover, diversity can be concentrated at positions most likely to enhance function without compromising structure, and facile methods can be employed for the optimization of affinity, specificity and stability.

In addition to these considerable practical advantages, synthetic antibodies offer an unprecedented opportunity to explore the principles underlying molecular recognition through direct experimentation rather than indirect analysis of natural antibodies. For example, we have used the synthetic approach to show that repertoires built on a single framework with diversity restricted to only four of the six complementarity-determining regions (CDRs) are sufficient for generating antibodies comparable to the best natural antibodies in terms of affinity and specificity.³ We have also shown that repertoires restricted to binary chemical diversity (Tyr and Ser) can generate highly functional antibodies⁴ and the abundance of Tyr residues is positively correlated with specificity.^{5,6} These findings run counter to commonly held beliefs inferred from the analysis of natural antibodies, which suggest that diverse frameworks and complex chemical diversity spread across the entire antigen-binding site are required for effective antigen recognition.⁷⁻¹¹ Moreover, these counterintuitive results show that further empirical experiments are needed to assess the validity of theories inferred from natural antibodies.

Here, we address another fundamental principle of antibody function: the dominant role of the third heavy chain CDR (CDR-H3) in antigen recognition. Natural antibodies provide abundant support for the notion that CDR-H3 is naturally predisposed to play the dominant role amongst the six CDRs.^{7,8,12-14} However, the question remains whether this dominance is a consequence of the genetics of the immune system which endow CDR-H3 with greater diversity than the other CDRs,^{15,16} or whether it arises from the central position of CDR-H3 within the antigen-binding site. In particular, CDR-H3 and the third CDR of the light chain (CDR-L3) occupy similar positions within the center of the paratope,^{10,17} and we constructed a synthetic library designed to ascertain whether CDR-H3 or -L3 is best suited as the dominant CDR in a repertoire free from the genetic biases that concentrate diversity in natural CDR-H3s. Surprisingly, our results suggest that CDR-L3 is more important than CDR-H3 for antigen recognition by this unbiased synthetic antibody repertoire.

Results

Library design and construction

We constructed a synthetic antigen-binding fragment (Fab) library (library F) designed to directly compare the contributions that CDR-H3 and -L3 make to antigen recognition. We used a single optimized human framework and diversified the three heavy-chain CDRs and CDR-L3 (Fig. 1a). Solvent accessible positions within CDR-H1 and -H2 were restricted to a binary diversity consisting of Tyr and Ser. Within CDR-H3 and -L3, we allowed more complex diversity, which was biased in favor of Tyr (25%), Ser (20%) and Gly (20%) but also included Ala (10%) and lesser quantities of five other amino acids (5% each of Phe, Trp, His, Val and Pro). We also introduced length diversity into CDR-H3 and -L3 by allowing loop lengths that are found within these regions of natural antibodies (Fig. 1b). The library was constructed using an anti-maltose binding protein Fab as the template. Library F contained 3×10^{10} unique clones and sequencing of the naïve library revealed the

incorporation of diversity in approximately 80% of the population within each CDR and the retention of template sequence in the remainder. Overall, 13%, 1%, 17% and 68% of the library members contained diversity in one, two, three or four CDRs, respectively (Fig. 1c).

Selection and characterization of functional antibodies

To assemble a panel of diverse functional antibodies, we used library F to generate 168 antibodies against 39 diverse protein antigens. The functional antibody set was not enriched relative to the naïve set in terms of sequences with all four CDRs mutated, indicating that, in terms of antigen recognition capability, these heavily diversified clones do not hold a significant advantage over less diversified clones containing two or three mutated CDRs (Fig. 1c). Comparing diversity within each CDR amongst functional clones relative to naïve clones, diversity was unchanged in CDR-H2, was modestly enriched in CDR-H1 and -L3 but was modestly depleted in CDR-H3. Thus, contrary to the situation in natural antibodies, our results suggest that synthetic antibodies derived from this library use CDR-L3 more often than CDR-H3 for antigen recognition. In fact, almost 20% (32 of 168) of the functional Fabs, recognizing 16 different antigens, contained the template CDR-H3. In contrast, less than 5% (8 of 168) of the functional Fabs contained the template CDR-L3.

To assess specificity, we used enzyme-linked immunosorbant assays (ELISAs) against a diverse set of antigens and analyzed Fabs that recognized 14 different antigens but contained the same template CDR-H3 sequence (Fig. 2). For comparison, we also assessed the specificities of 11 Fabs that contained mutated CDR-H3 sequences. Regardless of having template or mutated CDR-H3, most of the analyzed binders showed high specificity for the cognate antigen. Six or two Fabs with template or mutated CDR-H3 sequences, respectively, exhibited detectable binding to at least one non-cognate antigen. Surprisingly, affinity analyses of pairs of Fabs recognizing the same antigen showed that in five out of eight cases, Fabs containing the template CDR-H3 loop exhibited higher affinities than those containing mutated CDR-H3 loops (Fig. 2). Moreover, for three antigens, we isolated high affinity Fabs (5-1, 6-1, 11-1) in which only CDR-L3 was mutated. Taken together, these results show that, for a significant fraction of antigens, it is possible to generate high affinity Fabs without the need for CDR-H3 diversity, and Fabs containing a fixed CDR-H3 sequence exhibit comparable affinity and specificity relative to those containing mutated CDR-H3 loops.

Alanine-scanning of CDR-H3

Having shown that many Fabs containing the same CDR-H3 loop can recognize diverse antigens, we used shotgun alanine-scanning to investigate the functional contributions of individual residues within CDR-H3.¹⁸ We scanned ten Fabs that contain the same CDR-H3 but recognize different antigens, and the wild-type/Ala ratio at each position was used to assess the contribution of each residue to the recognition of each antigen (Fig. 3).

For most of the Fabs, many of the CDR-H3 residues are important for antigen recognition, and in four cases (2-1, 7-1, 9-1, 10-1), more than two thirds of the loop appears to be important. At the other extreme, in two Fabs (8-1, 1-1), almost the entire CDR-H3 appears to be inert. Thus, in most cases, it appears that the same CDR-H3 residues participate in energetically favorable interactions with different antigens, but in some cases, it appears that the antigen-binding site functions with virtually no contributions from the CDR-H3 loop. Notably, as illustrated in Fig. 2, there is no correlation between affinity and the functional contributions of CDR-H3, as the two lowest affinity Fabs (2-1, 10-1) rely heavily on CDR-H3, while the highest affinity Fab (8-1) relies the least on CDR-H3. Moreover, the Fabs that exhibit significant non-specific binding (2-1, 4-1, 7-1) also appear to rely heavily on CDRH3, suggesting that specificity does not correlate with functional contributions from CDR-H3. In summary, it is surprising but clear that the common CDR-H3 sequence within

these Fabs participates in the recognition of numerous diverse antigens, and the Fabs are highly functional, with most exhibiting affinities in the single-digit nanomolar range.

Structural characterization of a Fab-antigen complex

To gain insights into how the fixed CDR-H3 loop functions in an antigen-binding site, we solved the crystal structure of Fab-8-1 in complex with its cognate antigen, the tudor domain of the human tudor domain-containing protein 3 (TDRD3) (Table 1, Fig. 4). Fab-8-1 was chosen for analysis because it exhibits the highest affinity amongst the 10 Fabs subjected to alanine-scanning, and it is unusual in that the CDR-H3 sequence appears to play virtually no functional role in antigen recognition (Fig. 3). Thus, we sought to better understand how high affinity binding is achieved without a functional CDR-H3, and also, to investigate the structural role of the CDR-H3 loop within the paratope.

The binding of Fab-8-1 to TDRD3 results in an extensive interface, with 836 or 765 Å² of surface area buried on the antibody paratope or the antigen epitope, respectively (Fig. 4). Although CDR-L1 and -L2 were not diversified in the Fab library, these loops make a few contacts in the interface and contribute a small proportion of the total buried surface area (Fig. 4a). Amongst the four CDRs that were diversified, CDR-H1 makes only a minor contribution. In contrast, CDR-H2, -H3 and -L3 dominate the paratope and make extensive contacts to the TDRD3 antigen (Fig. 4a, b). The large contribution of CDR-H3 to the interface is somewhat surprising, considering that this loop was not diversified in the Fab and the majority of its residues were found to be functionally inert by alanine-scanning analysis (Fig. 3).

We compared the details of the molecular interactions that TDRD3 makes with the fixed CDR-H3 and the mutated CDR-L3 (Fig. 4c). CDR-L3 makes numerous van der Waals contacts through the aromatic/cyclic residues His^{107L}, Pro^{109L}, Phe^{110L}, Tyr^{113L} and Trp^{114L}. CDR-L3 also forms hydrogen bonds with TDRD3 through the backbone amides of Phe^{110L} and Trp^{114L} to the carbonyl of Gly⁵⁹⁵ or the side chain hydroxyl of Trp⁵⁹⁷, respectively. Consequently, the CDR-L3 main chain shows exquisite shape complementarity with the antigen surface, allowing the main chain to mediate hydrogen bonds. Likewise, the functionally inert CDR-H3 also makes numerous van der Waals contacts through the aromatic/cyclic residues Pro^{111.3H}, Tyr^{112.3H}, Phe^{112.2H} and Trp^{113H}, and possibly, through the aliphatic region of Lys^{111.2H} (not defined in the electron density). Moreover, the Arg^{109H} side chain of CDR-H3 forms a salt bridge to the Glu⁵⁹⁹ side chain of TDRD3. It is noteworthy that, 8 aside for Trp^{113H}, all of the residues in CDR-H3 that make contact with TDRD3 are functionally inert as assessed by alanine-scanning (Fig. 3).

To further define the role of CDR-H3 in antigen recognition, we examined the contacts made by the most conserved residues identified in our alanine-scanning analysis: Trp^{113H} and, to a lesser extent, Thr^{107H} and Gly^{112H} (Fig. 3, 4d). Trp^{113H} makes only a few van der Waals contacts to the antigen (Fig. 4c), and instead, primarily contacts residues on the adjacent CDR-L3 loop. CDR-L3 residues His^{107L}, Trp^{114L} and Phe^{116L} bury the side chain of Trp^{113H} in a hydrophobic pocket and the side chain of Gln^{105L} forms a hydrogen bond with the main chain carbonyl of Trp^{113H}. Interestingly, neither Thr^{107H} nor Gly^{112H} make contact with TDRD3. However, both residues are positioned in close proximity to Trp^{113H}, suggesting that they contribute indirectly to antigen recognition by either stabilizing the Trp^{113H} side chain rotamer or by stabilizing the short 310-helix that positions Trp^{113H} in contact with CDR-L3. Thus, the role of CDR-H3 in antigen recognition appears to be primarily structural, as the functionally important residues within this loop are primarily involved in facilitating the conformation of CDR-L3 required for antigen engagement.

Discussion

Using a synthetic antibody library with equivalent chemical diversity incorporated in both CDR-H3 and CDR-L3, we show that CDR-L3 is better able to contribute to functional antigen-binding sites than is CDR-H3. For almost half (16 of 39) of the antigens studied, the repertoire yielded functional antibodies that contained diversified CDR-L3 loops but retained a fixed CDR-H3 sequence. This observation of light chain driven antigen recognition contradicts previous studies of natural antibodies that have shown binding^{7,8,12-14} CDR-H3,¹⁹ but we show that CDR-L3 can readily dominate antigen recognition in synthetic antibodies that are not subject to these constraints.

Alanine-scanning analysis of ten unique antibodies with a common CDR-H3 showed that, in most cases, the fixed loop contributes energetically to the recognition of diverse antigens. The conformation of CDR-H3 strongly depends on its chemical and structural environment, or in other words, its interactions with other residues within the antibody and within the antigen.^{7,20-22} Our results show that the context dependence of CDR-H3 structure also extends to CDR-H3 function. The structure of one antibody shows how the common CDR-H3 loop establishes numerous contacts with the antigen and also supports the structure of CDR-L3, and it would be interesting to solve additional structures to compare and contrast the role of the same loop in different antibodies. Taken together, our results show that diversity in CDR-H3 is not required for high affinity recognition of diverse protein antigens, as a single fixed CDR-H3 can interact with many antigens and can thus contribute to the formation of diverse antigen-binding sites.

An important implication of our findings is that rules inferred from studies of natural antibodies are far from absolute, and when given the chance, it is clear that CDR-L3 can play a much more important role than that prescribed by natural sequence diversity. Although it has long been known that the light chain contributes to antigen recognition by natural antibodies¹⁹ and light chain engineering has been employed to create antibodies with novel binding characteristics,²³⁻²⁵ we show that CDR-L3 not only contributes to antigen recognition but can in fact dominate the antigen-binding site. In practical terms, these results indicate that synthetic antibody repertoires designed to incorporate diversity in CDR-L3 beyond that endowed by nature may prove to be better than natural antibody repertoires as resources for generating antibodies with novel functions.

Materials and Methods

Library construction

Library F was designed for the display of Fabs on the surface of M13 bacteriophage in a bivalent format,²⁶ as described,²⁷ using an optimized human framework (Fig. S1a) and mutagenic oligonucleotides designed to mutagenize CDR-H1, -H2, -H3 and -L3 (Fig. S1b). The highly diverse positions in the oligonucleotides for mutagenesis of CDR-H3 and -L3 were synthesized using a custom Trimer Phosphoramidite Mix (Glen Research, Sterling, VA) containing codons for nine amino acids in the following molar ratios: 25% of Tyr, 20% of Ser, 20% of Gly, 10% of Ala, and 5% each of Phe, Trp, His, Pro and Val.

Selection and characterization of Fab-phage

Phage from library F were cycled through rounds of binding selection with antigen coated on 96-well Maxisorp Immunoplates (NUNC, Rochester, NY), as described.^{3,28} After three to five rounds of selection, phage were produced from individual clones grown in a 96-well format and the culture supernatants were used in phage ELISAs to detect specific binding clones. Clones that bound to antigen but not to bovine serum albumin (BSA, Sigma-Aldridge, St. Louis, MO) were subjected to DNA sequence analysis.

To further assess specificity, phage ELISAs were used to measure binding to a panel of antigens, as described.⁵ Competitive phage ELISAs were used to determine IC50 values, defined as the concentration of soluble antigen that blocked 50% of the phage binding to immobilized antigen. A fixed, sub-saturating concentration of phage was pre-incubated for 2 h with serial dilutions of antigen and then transferred to antigen-coated plates, which were incubated, washed, developed and read, as described.⁵

Kinetic affinity analysis

Fab proteins were purified from *Escherichia coli*, as described.²⁹ Binding kinetics were determined by surface plasmon resonance using a ProteON XPR36 (BioRAD) with antigen immobilized on GLC chips at a density sufficient to produce ~100 response units when saturated with Fab. Serial dilutions of Fab proteins were injected, and binding responses were corrected by subtraction of responses on a blank flow cell. For kinetic analysis, a 1:1 Langmuir model of global fittings of k_{on} and k_{off} was used. The K_D values were estimated from the ratios of k_{on} and k_{off} .

Shotgun alanine-scanning

Combinatorial alanine scanning was performed as described.¹⁸ A mutagenic oligonucleotide was used to construct libraries, one for each of the ten Fabs analyzed, in which codons within the CDR-H3 sequence (positions 107-113) were replaced with a degenerate codon that encoded for equal proportions of wild-type, Ala and, in some cases, two additional amino acids. Phage from each library were cycled through two or three rounds of binding selection as described above. Approximately 50 binding clones were sequenced from each library and the wild-type/Ala ratio was determined for unique clones at each varied position.

Crystallization, structure determination and refinement

DNA encoding the tudor domain of TDRD3 or the anti-TDRD3 Fab-8-1 were cloned into the pET28-MHL (GenBank accession EF456735) or pCW-LIC (GenBank accession EF460848) vectors, respectively. The resulting expression vectors were used to produce recombinant protein in *E.coli* BL21 (DE3) codon plus strain (Stratagene). TDRD3 was purified using metal affinity chromatography on a Nichelating open column followed by size exclusion chromatography on a pre-packed HiLoad 26/60 Superdex 75 pg size exclusion column (GE Life Sciences). Fab-8-1 was purified using a protein A-sepharose open column followed by cation-exchange chromatography on a HiTrap SP HP column (GE Life Sciences). Purified TDRD3 protein was mixed with Fab-8-1 protein at a 2:1 molar ratio, and the TDRD3:Fab-8-1 complex was purified to homogeneity by size exclusion chromatography on a HiLoad 16/60 Superdex 75 pg size exclusion column (GE Life Sciences).

The TDRD3:Fab-8-1 complex was crystallized at a concentration 24 mg/ml using the sitting drop vapor diffusion method at 18 °C. The reservoir solution contained 16% PEG 3350, 0.05 citric acid, 0.05M BIS-Tris propane, pH 5. Using a nylon loop,³⁰ crystals were passed through the reservoir solution containing 20% glycerol, flash-frozen and stored in liquid nitrogen until data collection.³¹ A 2.05 Å resolution dataset was collected at beamline 19ID of the Advanced Photon Source and reduced using the HKL3000 suite of programs.³² The structure was solved by molecular replacement with the program PHASER³³ and coordinates from PDB³⁴ entries 2HFF,³⁵ modified on the FFAS03 server,³⁶ and 3PMT. The asymmetric unit contained eight heterotrimers, each of which comprised light and heavy antibody chains and TDRD3. Manual model adjustments were performed with COOT.³⁷ REFMAC³⁸ and PHENIX³⁹ were used for refinement. Model geometry was validated on the MOLPROBITY server.⁴⁰

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We gratefully acknowledge Nick Jarvik for assistance in protein purification and affinity analysis and Nicolas Economopoulos, Jonathan Olsen and Eric MacKinnon for performing some of the phage display selections presented in the study. This work was supported by funds from the Canadian Institutes for Health Research (MOP-93725 to S.S.S.), the U.S. National Institutes of Health (U01-GM094588 and U54-HG006436 to S.S.S. and S.K.) and The Swedish Research Council (524-2008-617 to H.P.)

Abbreviations

BSA	bovine serum albumin
CDR	complementarity-determining region
CDR-Hn	heavy-chain CDR 1, 2 or 3 (where n=1, 2, or 3)
CDR-Ln	light-chain CDR 1, 2 or 3 (where n=1, 2, or 3)
ELISA	enzyme-linked immunosorbant assay
Fab	antigen-binding fragment
PBS	phosphate-buffered saline
TDRD3	tudor domain-containing protein 3

References

1. Harel Inbar N, Benhar I. Selection of antibodies from synthetic antibody libraries. *Archives of Biochemistry and Biophysics*. 2012; 526:87–98. [PubMed: 22244834]
2. Sidhu SS, Fellouse FA. Synthetic therapeutic antibodies. *Nat Chem Biol*. 2006; 2:682–688. [PubMed: 17108986]
3. Fellouse FA, Esaki K, Birtalan S, Raptis D, Cancasci VJ, Koide A, et al. High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J Mol Biol*. 2007; 373:924–940. [PubMed: 17825836]
4. Fellouse FA, Li B, Compaan DM, Peden AA, Hymowitz SG, Sidhu SS. Molecular recognition by a binary code. *J Mol Biol*. 2005; 348:1153–1162. [PubMed: 15854651]
5. Birtalan S, Zhang Y, Fellouse FA, Shao L, Schaefer G, Sidhu SS. The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. *J Mol Biol*. 2008; 377:1518–1528. [PubMed: 18336836]
6. Birtalan S, Fisher RD, Sidhu SS. The functional capacity of the natural amino acids for molecular recognition. *Mol Biosyst*. 2010; 6:1186–1194. [PubMed: 20383388]
7. Almagro JC. Identification of differences in the specificitydetermining residues of antibodies that recognize antigens of different size: implications for the rational design of antibody repertoires. *J Mol Recognit*. 2004; 17:132–143. [PubMed: 15027033]
8. MacCallum RM, Martin AC, Thornton JM. Antibody-antigen interactions: contact analysis and binding site topography. *J Mol Biol*. 1996; 262:732–745. [PubMed: 8876650]
9. Padlan EA, Abergel C, Tipper JP. Identification of specificitydetermining residues in antibodies. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 1995; 9:133–139. [PubMed: 7821752]
10. Wilson IA, Stanfield RL. Antibody-antigen interactions. *Curr Opin Struct Biol*. 1993; 3:113–118.
11. Wilson IA, Stanfield RL. Antibody-antigen interactions: new structures and new conformational changes. *Current opinion in structural biology*. 1994; 4:857–867. [PubMed: 7536111]

12. Xu JL, Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity*. 2000; 13:37–45. [PubMed: 10933393]
13. Zemlin M, Klinger M, Link J, Zemlin C, Bauer K, Engler JA, et al. Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. *J Mol Biol*. 2003; 334:733–749. [PubMed: 14636599]
14. Kabat EA, Wu TT. Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementaritydetermining regions to binding of antibody-combining sites. *J Immunol*. 1991; 147:1709–1719. [PubMed: 1908882]
15. Alt FW, Blackwell TK, Yancopoulos GD. Development of the primary antibody repertoire. *Science*. 1987; 238:1079–1087. [PubMed: 3317825]
16. Tonegawa S. Somatic generation of antibody diversity. *Nature*. 1983; 302:575–581. [PubMed: 6300689]
17. Padlan EA. Anatomy of the antibody molecule. *Mol Immunol*. 1994; 31:169–217. [PubMed: 8114766]
18. Weiss GA, Watanabe CK, Zhong A, Goddard A, Sidhu SS. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc Natl Acad Sci U S A*. 2000; 97:8950–8954. [PubMed: 10908667]
19. Nemazee D. Receptor editing in lymphocyte development and central tolerance. *Nat Rev Immunol*. 2006; 6:728–740. [PubMed: 16998507]
20. James LC, Roversi P, Tawfik DS. Antibody multispecificity mediated by conformational diversity. *Science*. 2003; 299:1362–1367. [PubMed: 12610298]
21. Kuroda D, Shirai H, Kobori M, Nakamura H. Structural classification of CDR-H3 revisited: a lesson in antibody modeling. *Proteins*. 2008; 73:608–620. [PubMed: 18473362]
22. Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. Conformations of the third hypervariable region in the VH domain of immunoglobulins. *Journal of molecular biology*. 1998; 275:269–294. [PubMed: 9466909]
23. Boström J, Yu SF, Kan D, Appleton BA, Lee CV, Billeci K, et al. Variants of the antibody herceptin that interact with HER2 and VEGF at the antigen binding site. *Science*. 2009; 323:1610–1614. [PubMed: 19299620]
24. Schaefer G, Haber L, Crocker LM, Shia S, Shao L, Dowbenko D, et al. A two-in-one antibody against HER3 and EGFR has superior inhibitory activity compared with monospecific antibodies. *Cancer cell*. 2011; 20:472–486. [PubMed: 22014573]
25. van den Beucken T, van Neer N, Sablon E, Desmet J, Celis L, Hoogenboom HR, et al. Building novel binding ligands to B7.1 and B7.2 based on human antibody single variable light chain domains. *Journal of molecular biology*. 2001; 310:591–601. [PubMed: 11439026]
26. Lee CV, Sidhu SS, Fuh G. Bivalent antibody phage display mimics natural immunoglobulin. *J Immunol Methods*. 2004; 284:119–132. [PubMed: 14736422]
27. Fellouse FA.; Sidhu, SS. Making antibodies in bacteria. In: Howard, GC.; Kaser, MS., editors. *Making and Using Antibodies*. CRC Press; Boca Raton, FL: 2007. p. 157-180.
28. Sidhu SS, Li B, Chen Y, Fellouse FA, Eigenbrot C, Fuh G. Phage-displayed antibody libraries of synthetic heavy chain complementarity determining regions. *J Mol Biol*. 2004; 338:299–310. [PubMed: 15066433]
29. Colwill K, Gräslund S. A roadmap to generate renewable protein binders to the human proteome. *Nature methods*. 2011; 8:551–558. [PubMed: 21572409]
30. Teng TY. Mounting of Crystals for Macromolecular Crystallography in a Freestanding Thin-Film. *Journal of Applied Crystallography*. 1990; 23:387–391.
31. Hope H. Cryocrystallography of Biological Macromolecules - a Generally Applicable Method. *Acta Crystallographica Section B-Structural Science*. 1988; 44:22–26.
32. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution - from diffraction images to an initial model in minutes. *Acta Crystallographica Section D-Biological Crystallography*. 2006; 62:859–866.

33. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *Journal of Applied Crystallography*. 2007; 40:658–674. [PubMed: 19461840]
34. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Research*. 2000; 28:235–242. [PubMed: 10592235]
35. Lee CV, Hymowitz SG, Wallweber HJ, Gordon NC, Billeci KL, Tsai SP, et al. Synthetic anti-BR3 antibodies that mimic BAFF binding and target both human and murine B cells. *Blood*. 2006; 108:3103–3111. [PubMed: 16840730]
36. Jaroszewski L, Rychlewski L, Li ZW, Li WZ, Godzik A. FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Research*. 2005; 33:W284–W288. [PubMed: 15980471]
37. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography*. 2010; 66:486–501.
38. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D-Biological Crystallography*. 2011; 67:355–367.
39. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallographica Section DBiological Crystallography*. 2002; 58:1948–1954.
40. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D-Biological Crystallography*. 2010; 66:12–21.
41. Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol*. 2003; 27:55–77. [PubMed: 12477501]
42. Biochemistry, N. C. o. t. I. U. o. Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984. *Biochem J*. 1985; 229:281–286. [PubMed: 4038268]

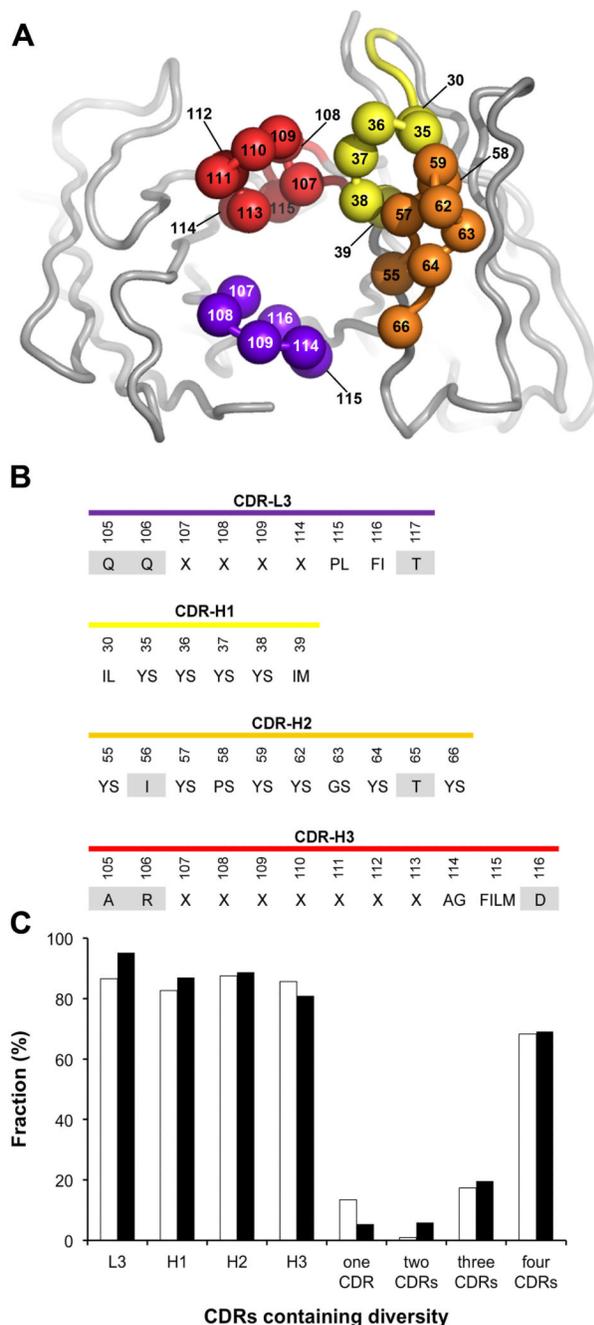


Fig. 1. Library design and characterization. (a) The backbones of the heavy and light chain variable domains are shown as tubes. The frameworks are colored grey and the CDR loops are colored as follows: CDR-L3 (purple), CDR-H1 (yellow), CDR-H2 (orange) CDR-H3 (red). Spheres colored according to the CDR coloring scheme represent positions that were diversified. The figure was generated using PyMOL (<http://www.pymol.org/>) with crystal structure coordinates (Protein Data Bank entry 1FVC). (b) CDR diversity design. Positions shaded in grey were fixed as the parental sequence, and at each diversified position, the allowed amino acids are denoted by the single-letter code. *X* denotes a mixture of nine amino acids (Y, S, G, A, F, W, H, P or V) introduced at proportions described in Materials

and Methods. The lengths of CDR-L3 and -H3 were varied by replacing the positions denoted by *X* with 3-7 or 1-17 degenerate codons, respectively. Residue numbering is according to the IMGT scheme.⁴¹ (c) Actual CDR diversity in naïve and functional Fabs. The fractions of Fab-phage containing diversity within a particular CDR or containing a given number of mutated CDRs are shown for 104 unique naïve Fabs (white bars) and 168 unique functional Fabs selected for binding to 39 different antigens (black bars).

Fab	CDR-L3	CDR-H1	CDR-H2	CDR-H3	K _d /IC ₅₀ (nM)	Specificity ELISA signal (absorbance, 450 nm)													
						1	2	3	4	5	6	7	8	9	10	11	12	13	14
WT	SSSSSII	SSSSSYGVYF	TVRGSKKP	YFSSGWAM		0.02	0.00	0.01	0.00	0.02	0.06	0.05	0.04	0.04	0.05	0.17	0.01	0.00	0.01
151	PYVW-WLI	YYSYM	SPYSGY	TVRGSKKP	10*	1.34	0.01	0.01	0.00	0.03	0.01	0.00	0.07	0.01	0.16	0.00	nd	nd	
12	SSSY-SLI	YYSYM	SPYSGY	TVRGSKKP	<10*	2.39	0.01	0.02	0.03	0.03	0.00	0.01	0.03	0.01	nd	0.02	nd	nd	
283	YFVWALI	YYSVI	SPYSGY	TVRGSKKP	60	1.39	0.06	0.01	0.00	0.11	0.15	0.00	0.16	0.03	0.06	0.01	0.07	0.05	
22	H F - - - S P I	SSSSII	YYSYGVY	TVRGSKKP	200	0.06	2.32	0.02	0.04	0.03	0.00	0.00	0.04	0.02	nd	0.02	0.10	0.06	
338	W F W - - S P I	SSSSII	YYSYGVY	TVRGSKKP	9	0.09	0.22	1.68	0.01	0.06	0.02	0.00	0.26	0.01	0.09	0.01	0.04	0.03	
32	W S Y - - Y L I	YYSYM	SPYSGY	TVRGSKKP	7	0.23	0.03	2.10	0.11	0.14	0.01	0.02	0.07	0.06	nd	0.05	0.02	0.02	
485	W P - - - W P I	YYSYM	SPYSGY	TVRGSKKP	9*	0.47	0.05	0.07	2.04	0.22	0.13	0.05	0.41	0.08	0.19	0.04	nd	nd	
42	W V - - - W P I	YYSYM	SPYSGY	TVRGSKKP	5	0.08	0.22	0.24	1.78	0.07	0.03	0.01	0.25	0.04	nd	0.03	nd	nd	
84	W P W Y A P P F	SSSSII	SSSSYGVY	TVRGSKKP	250*	0.26	0.05	0.06	0.06	2.44	0.02	0.03	0.16	0.03	0.12	0.03	nd	nd	
52	W P P - - S P I	SSSSII	SSSSYGVY	TVRGSKKP	40	0.03	0.00	0.03	0.02	2.43	0.01	0.01	0.06	0.26	nd	0.01	nd	nd	
61	P W - - - P S L F	SSSSII	SSSSYGVY	TVRGSKKP	9	0.02	0.00	0.00	0.00	0.01	1.59	0.00	0.03	0.00	0.00	0.00	nd	nd	
62	W P - - - S P I	SSSSII	SSSSYGVY	TVRGSKKP	40	0.03	0.00	0.00	0.00	1.20	0.00	0.00	0.00	0.00	0.00	0.01	nd	nd	
71	Y V A - - H P I	SSSSII	SSPYSGY	TVRGSKKP	6	0.24	0.03	0.08	0.03	nd	0.07	0.20	0.22	nd	0.14	0.06	0.28	0.27	
72	W P - - - S P I	SSSSII	SSSSYGVY	TVRGSKKP	9	0.13	0.00	0.01	-0.01	nd	0.01	0.57	0.07	nd	nd	0.01	0.10	0.13	
85	H S P F - W L F	SSSSII	SSSSYGVY	TVRGSKKP	1	0.04	0.01	0.01	0.00	0.03	0.02	0.00	0.28	0.01	0.04	0.02	0.03	0.01	
91	P S P - - S P I	SSSSII	SSPYSGY	TVRGSKKP	<10*	0.05	0.01	0.01	0.01	0.03	0.02	0.00	0.10	0.05	nd	0.01	nd	nd	
104	F S P - - H L I	YYSVI	SSSSYGVY	TVRGSKKP	40	0.19	0.05	0.16	0.03	0.04	0.08	0.05	0.11	nd	0.26	nd	nd	nd	
111	K Y F W - M P I	SSSSII	SSSSYGVY	TVRGSKKP	2	0.67	0.04	0.08	0.05	nd	0.07	0.03	0.19	nd	2.34	0.54	0.38	0.50	
112	W P - - - S P I	YYSVI	SSPYSGY	TVRGSKKP	60	1.04	0.08	0.10	0.05	nd	0.06	0.20	0.14	nd	1.94	1.72	1.16	1.37	
121	SSSY-SLI	YYSYM	SPYSGY	TVRGSKKP		0.16	0.07	0.26	0.04	nd	0.10	0.02	0.12	nd	0.05	0.20	0.10	0.15	
122	H W P - - W L F	YYSYM	SSPYSGY	TVRGSKKP		0.04	-0.01	0.00	0.02	nd	0.00	-0.02	0.01	nd	0.05	0.27	0.02	0.06	
131	P H W - - W L F	YYSYM	SSPYSGY	TVRGSKKP		0.05	-0.01	0.01	-0.01	nd	0.01	0.00	0.07	nd	0.03	0.32	0.28	0.05	
132	K C Y - - Y S L F	SSSSII	SSSSYGVY	TVRGSKKP		0.04	-0.01	-0.01	-0.01	nd	0.01	-0.01	0.06	nd	-0.01	0.02	0.04	0.07	
141	K T V - - V S L F	SSSSII	SSSSYGVY	TVRGSKKP		0.01	-0.01	-0.01	-0.02	nd	0.00	-0.01	-0.01	nd	-0.02	0.04	0.00	0.06	
142	W S A - - W P I	YYSYM	H S P S G S Y	H G A W W A G C		0.01	-0.02	0.01	0.02	nd	-0.01	-0.01	-0.01	nd	-0.01	0.01	0.00	0.28	

Fig. 2. Sequences, affinities and specificities of antigen-binding Fabs. CDR positions that were randomized in the library are shown for Fabs selected for binding to 14 different antigens (numbered from 1 through 14). CDRs shaded in *grey* indicate sequences of the template Fab (WT) used for library construction. Tyr, Ser, Gly and Ala are shown in yellow, red, green or blue, respectively. Dashes indicate gaps in the alignment. Binding affinities to cognate antigens were determined by surface plasmon resonance (KD) or by competitive phage ELISA (IC50, marked with asterisk). Specificity ELISA signals are colored blue for cognate antigen or as follows for non-cognate antigens: 1.0 < dark yellow; 0.2 < light yellow < 1.0; 0.2 > white. *ND* indicates that the value was not determined. Fab names shaded in grey indicate that the CDR-H3 loops were subjected to shotgun alanine-scanning analysis (see Fig. 3). The numbers denote the following antigens (described with gene name and Uniprot identifier): (1) HUWE1 (Q7Z6Z7), (2) PA0623 (Q915S9), (3) ACTB (P60709), (4) Tank2Parp (Q9H2K23), (5) MCMBP (Q9BTE3-2), (6) CBX3 (Q13185), (7) G9A (Q96KQ7), (8) TDRD3 (Q9H7E2), (9) MEN1 (O00255-2), (10) TcTex1 (P631723), (11) GRM5 (P41594), (12) SH3PXD2A (Q5TCZ1), (13) CD2AP (Q9Y5K6), (14) SH3D19 (Q5HYK7).

Fab	Residue number and sequence												
	T107	V108	R109	G110	S111	K111.1	K111.2	P111.3	Y112.3	F112.2	S112.1	G112	W113
1-1	3.8	3.7	6.8	1.2	0.7	0.7	1.2	0.7	0.9	0.4	0.3	1.2	38
2-1	>41	1.6	7.5	20	2.4	3.0	12	13	8.5	>39	1.6	13	>41
3-1	0.7	7.0	3.3	1.4	1.3	2.0	1.7	0.8	>45	45	47	>48	47
4-1	>15	2.0	14	1.5	1.5	9.0	5.0	6.5	>15	>14	2.0	14	>14
5-1	>41	1.7	35	1.0	1.3	7.3	1.6	1.2	1.1	1.5	5.8	>41	7.2
6-1	>40	3.4	16	0.5	1.9	16	3.6	1.1	3.6	>40	1.1	0.7	>40
7-1	>15	>15	>15	14	1.5	3.0	1.2	6.5	>13	>11	4.0	>15	>15
8-1	5.1	0.9	3.6	1.5	1.0	2.4	1.2	0.8	1.0	0.6	0.4	4.4	>43
9-1	>11	5.0	>12	>12	1.4	>8	>12	>12	>12	>12	3.0	11	>12
10-1	>10	0.3	>10	2.3	0.7	>9	>10	>10	>10	>10	4.0	>10	>10

Fig. 3. Shotgun alanine-scanning analysis of CDR-H3 loops. The wild-type/Ala ratios for the residues in CDR-H3 following selection for binding to antigen are shown for ten Fabs that contain an identical CDR-H3 loop but recognize different antigens (see Fig. 2). Residues predicted to be important for antigen binding (wild-type/Ala > 4) are shaded *grey*. In cases, where no alanine mutations were found a lower limit is given.

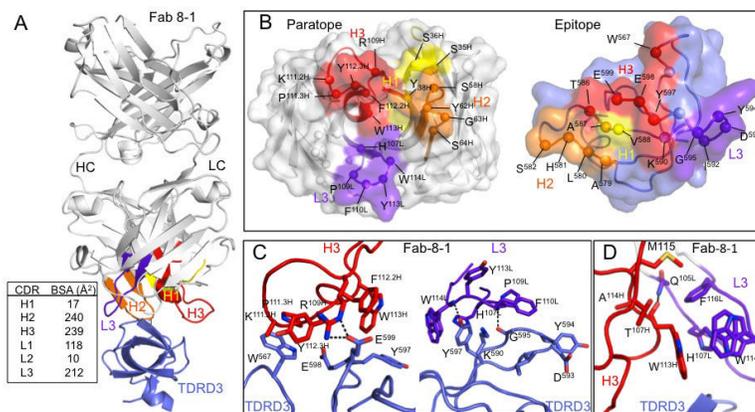


Fig. 4. The crystal structure of the TDRD3:Fab-8-1 complex. (a) Overall structure of the TDRD3:Fab-8-1 complex. TDRD3 is colored blue. Fab-8-1 is colored grey, or as follows: CDR-H1 (yellow), CDR-H2 (orange), CDR-H3 (red), CDR-L3 (purple). Inset is the buried surface area (BSA) contributions of each CDR to the interface. (b) The structural paratope and epitope. Fab-8-1 (left) and TDRD3 (right) are shown in an open book view as molecular surfaces. Residues within 4.5 Å of the cognate ligand are represented by spheres. Fab-8-1 paratope residues are colored yellow, orange, red or purple if they contact TDRD3 and reside within CDR-H1, -H2, -H3 or -L3, respectively. TDRD3 epitope residues are colored yellow, orange, red or purple if they contact CDR-H1, -H2, -H3 or -L3, respectively. (c) Interactions between TDRD3 and CDR-H3 (left) or CDR-L3 (right). (d) Interactions between CDRH3 and CDR-L3. Dashed lines represent hydrogen bonds or ionic bonds.

Table 1

Data collection and refinement statistics for the TDRD3:Fab-8-1 complex (3PNW)

<i>Data collection</i>	
Space group	P1
Cell dimensions	
$a \times b \times c$ (Å)	76.7, 93.7, 159.9
$\alpha \times \beta \times \gamma$ (°)	81.0, 82.8, 90.1
Resolution (Å)	50.0-2.05 (2.12-2.05) ^a
R_{sym}^b (%)	8.6 (41.4) ^a
$\ \sigma I$	9.5 (2.0) ^a
Completeness (%)	97.4 (93.1) ^a
Redundancy	2.5 (2.3) ^a
Total reflections	266201
Unique reflections	25457
<i>Refinement statistics</i>	
Resolution (Å)	30.0-2.05 (2.10-2.05) ^a
$R_{\text{work}} / R_{\text{free}}^c$ (%)	22.3/ 26.4 (32.6/ none) ^d
Reflections used	266095 (18210) ^a
No. of protein atoms	29758
No. of waters	1275
rmsd bond lengths (Å)	0.014
rmsd bond angles (°)	1.3
Average <i>B</i> -factors	
Protein (Å ²)	55.6
Solvent (Å ²)	47.5
<i>Ramachandran plot statistics</i> ^e	
Most favored (%)	91.8
Additional allowed (%)	7.7
Generously allowed (%)	0.3
Disallowed (%)	0.3

^aValues in parentheses represent the highest resolution shell.

^b $R_{\text{Sym}}(I) = \frac{\sum_{hk\ell} \sum_i |I_i(hk\ell) - \langle I(hk\ell) \rangle|}{\sum_{hk\ell} \sum_i I_i(hk\ell)}$ where the summations are over *i* observations of each reflections and all *hkℓ*. $\langle I(hk\ell) \rangle$ is the average intensity of the *i* observations. $R_{\text{work}} = |F(\text{obs}) - F(\text{calc})|/F(\text{obs})$

^c R_{free} was calculated for 2162 reflections selected in thin resolution shells (SFTOOLS, B. Hazes, University of Alberta) and not used in the refinement.

^dNo free reflections were selected within the high resolution shell.

^eRA Laskowski, MW MacArthur, DS Moss and JM Thornton (1993) *JAppl Crystallogr* 26: 283-291