MABS
2016, VOL. 8, NO. 6, 1045–1063
http://dx.doi.org/10.1080/19420862.2016.1190060

Taylor & Francis
Taylor & Francis Group

REPORT

OPEN ACCESS

Structural diversity in a human antibody germline library

Alexey Teplyakov, Galina Obmolova, Thomas J. Malia, Jinquan Luo, Salman Muzammil, Raymond Sweet, Juan Carlos Almagro, and Gary L. Gilliland

Janssen Research & Development LLC, Spring House, PA, USA

ABSTRACT

To support antibody therapeutic development, the crystal structures of a set of 16 germline variants composed of 4 different kappa light chains paired with 4 different heavy chains have been determined. All four heavy chains of the antigen-binding fragments (Fabs) have the same complementarity-determining region (CDR) H3 that was reported in an earlier Fab structure. The structure analyses include comparisons of the overall structures, canonical structures of the CDRs and the VH:VL packing interactions. The CDR conformations for the most part are tightly clustered, especially for the ones with shorter lengths. The longer CDRs with tandem glycines or serines have more conformational diversity than the others. CDR H3, despite having the same amino acid sequence, exhibits the largest conformational diversity. About half of the structures have CDR H3 conformations similar to that of the parent; the others diverge significantly. One conclusion is that the CDR H3 conformations are influenced by both their amino acid sequence and their structural environment determined by the heavy and light chain pairing. The stem regions of 14 of the variant pairs are in the 'kinked' conformation, and only 2 are in the extended conformation. The packing of the VH and VL domains is consistent with our knowledge of antibody structure, and the tilt angles between these domains cover a range of 11 degrees. Two of 16 structures showed particularly large variations in the tilt angles when compared with the other pairings. The structures and their analyses provide a rich foundation for future antibody modeling and engineering efforts.

ARTICLE HISTORY
Received 17 March 2016
Revised 9 May 2016
Accepted 11 May 2016

KEYWORDS
Antibody structure; CDR canonical structure; CDR H3; phage library; VH:VL packing

Introduction

At present, therapeutic antibodies are the largest class of biotherapeutic proteins that are in clinical trials.[1] The use of monoclonal antibodies as therapeutics began in the early 1980s, and their composition has transitioned from murine antibodies to generally less immunogenic humanized and human antibodies. The technologies currently used to obtain human antibodies include transgenic mice containing human antibody repertoires, cloning directly from human B cells, and in vitro selection from antibody libraries using various display technologies. Once a candidate antibody is identified, protein engineering is usually required to produce a molecule with the right biophysical and functional properties. All engineering efforts are guided by our understanding of the atomic structures of antibodies. In such efforts, the crystal structure of the specific antibody may not be available, but modeling can be used to guide the engineering efforts. Today's antibody modeling approaches, which normally focus on the variable region, are being developed by the application of structural principles and insights that are evolving as our knowledge of antibody structures continues to expand.

Our current structural knowledge of antibodies is based on a multitude of studies that used many techniques to gain insight into the functional and structural properties of this class of macromolecule. Five different antibody isotypes occur, IgG, IgD, IgE, IgA and IgM, and each isotype has a unique role in the adaptive immune system. IgG, IgD and IgE isotypes are composed of 2 heavy chains (HCs) and 2 light chains (LCs) linked through disulfide bonds, while IgA and IgM are double and quintuple versions of antibodies, respectively. Isotypes IgG, IgD and IgA each have 4 domains, one variable (V) and 3 constant (C) domains, while IgE and IgM each have the same 4 domains along with an additional C domain. These multimeric forms are linked with an additional J chain. The LCs that associate with the HCs are divided into 2 functionally indistinguishable classes, κ and λ. Both κ and λ polypeptide chains are composed of a single V domain and a single C domain.

The heavy and light chains are composed of structural domains that have ~110 amino acid residues. These domains have a common folding pattern often referred to as the "immunoglobulin fold," formed by the packing together of 2 anti-parallel β-sheets. All immunoglobulin chains have an N-terminal V domain followed by 1 to 4 C domains, depending upon the chain type. In antibodies, the heavy and light chain V domains pack together forming the antigen combining site. This site, which interacts with the antigen (or target), is the focus of current antibody modeling efforts. This interaction site is composed of 6 complementarity-determining regions (CDRs) that were identified in early antibody amino acid sequence analyses to

CONTACT Alexey Teplyakov ateplyak@its.jnj.com Janssen BioTherapeutics, Janssen R&D, Spring House, PA, USA
Supplemental material data for this article can be accessed on the publisher's website.

be hypervariable in nature,[2] and thus are responsible for the sequence and structural diversity of our antibody repertoire.

The sequence diversity of the CDR regions presents a substantial challenge to antibody modeling. However, an initial structural analysis of the combining sites of the small set of structures of immunoglobulin fragments available in the 1980s found that 5 of the 6 hypervariable loops or CDRs had canonical structures (a limited set of main-chain conformations).[3-4] A CDR canonical structure is defined by its length and conserved residues located in the hypervariable loop and framework residues (V-region residues that are not part of the CDRs). Furthermore, studies of antibody sequences revealed that the total number of canonical structures are limited for each CDR,[5-7] indicating possibly that antigen recognition may be affected by structural restrictions at the antigen-binding site. Later studies found that the CDR loop length is the primary determining factor of antigen-binding site topography because it is the primary factor for determining a canonical structure.[8,9] Additional efforts have led to our current understanding that the LC CDRs L1, L2, and L3 have preferred sets of canonical structures based on length and amino acid sequence composition. This was also found to be the case for the H1 and H2 CDRs.[8] Classification schemes for the canonical structures of these 5 CDRs have emerged and evolved as the number of depositions in the Protein Data Bank[10] of Fab fragments of antibodies grow.[8,11] Recently, a comprehensive CDR classification scheme was reported identifying 72 clusters of conformations observed in antibody structures.[12,13] The knowledge and predictability of these CDR canonical structures have greatly advanced antibody modeling efforts.

In contrast to CDRs L1, L2, L3, H1 and H2, no canonical structures have been observed for CDR H3, which is the most variable in length and amino acid sequence. Some clustering of conformations was observed for the shortest lengths; [12] however, for the longer loops, only the portions nearest the framework (torso, stem or anchor region) were found to have defined conformations.[14,15] In the torso region, 2 primary groups could be identified, which led to sequence-based rules that can predict with some degree of reliability the conformation of the stem region.[16,17] The "kinked" or "bulged" conformation is the most prevalent, but an "extended" or "non-bulged" conformation is also, but less frequently, observed. The cataloging and development of the rules for predicting the conformation of the anchor region of CDR H3 continue to be refined, producing new insight into the CDR H3 conformations and new tools for antibody engineering.[12,13,18]

Current antibody modeling approaches take advantage of the most recent advances in homology modeling, the evolving understanding of the CDR canonical structures, the emerging rules for CDR H3 modeling and the growing body of antibody structural data available from the PDB. Recent antibody modeling assessments show continued improvement in the quality of the models being generated by a variety of modeling methods.[19,20] Although antibody modeling is improving, the latest assessment [21] revealed a number of challenges that need to be overcome to provide accurate 3-dimensional models of antibody V regions, including accuracies in the modeling of CDR H3. The need for improvement in this area was also highlighted in a recent study [18] reporting an approach and results that may influence future antibody modeling efforts. One important finding of the antibody modeling assessments was that errors in the structural templates that are used as the basis for homology models can propagate into the final models, producing inaccuracies that may negatively influence the predictive nature of the V region model.

To support antibody engineering and therapeutic development efforts, a phage library was designed and constructed based on a limited number of scaffolds built with frequently used human germ-line IGV and IGJ gene segments that encode antigen combining sites suitable for recognition of peptides and proteins.[22] This Fab library is composed of 3 HC germlines, IGHV1-69 (**H1-69**), IGHV3-23 (**H3-23**) and IGHV5-51 (**H5-51**), and 4 LC germlines (all κ), IGKV1-39 (**L1-39**), IGKV3-11 (**L3-11**), IGKV3-20 (**L3-20**) and IGKV4-1 (**L4-1**). Selection of these genes was based on the high frequency of their use[23] and their cognate canonical structures that were found binding to peptides and proteins,[7] as well as their ability to be expressed in bacteria and displayed on filamentous phage. The implementation of the library involves the diversification of the human germline genes to mimic that found in natural human libraries.

The crystal structure determinations and structural analyses of all germline Fabs in the library described above along with the structures of a fourth HC germline, IGHV3-53 (**H3-53**), paired with the 4 LCs of the library have been carried out to support antibody therapeutic development. All 16 HCs of the Fabs have the same CDR H3 that was reported in an earlier Fab structure.[24] This is the first systematic study of the same VH and VL structures in the context of different pairings. The structure analyses include comparisons of the overall structures, canonical structures of the L1, L2, L3, H1 and H2 CDRs, the structures of all CDR H3s, and the VH:VL packing interactions. The structures and their analyses provide a foundation for future antibody engineering and structure determination efforts.

Results

Crystal structures

The crystal structures of a germline library composed of 16 Fabs generated by combining 4 HCs (**H1-69**, **H3-23**, **H3-53** and **H5-51**) and 4 LCs (**L1-39**, **L3-11**, **L3-20** and **L4-1**) have been determined. The Fab heavy and light chain sequences for the variants numbered according to Chothia are shown in Fig. S1. The four different HCs all have the same CDR H3 sequence, **ARYDGIYGELDF**. Crystallization of the 16 Fabs was previously reported.[25] Three sets of the crystals were isomorphous with nearly identical unit cells (Table 1). These include (1) **H3-23:L3-11** and **H3-23:L4-1** in $P2_12_12_1$, (2) **H3-53:L1-39**, **H3-53:L3-11** and **H3-53:L3-20** in $P6_522$, and (3) **H5-51:L1-39**, **H5-51:L3-11** and **H5-51:L3-20** in $P2_12_12_1$. Crystallization conditions for the 3 groups are also similar, but not identical (Table 1). Variations occur in the pH (buffer) and the additives, and, in group 3, PEG 3350 is the precipitant for one variants while ammonium sulfate is the precipitant for the other two. The similarity in the crystal forms is attributed in part to cross-seeding using the microseed matrix screening for groups 2 and 3.

Table 1. Crystal data, X-ray data, and refinement statistics.

Fab	H1-69:L1-39	H1-69:L3-11	H1-69:L3-20	H1-69:L4-1
PDB identifier	5I15	5I16	5I17	5I18
Crystal Data				
Crystallization Solution				
Buffer, pH	0.1 M MES- pH 6.5	0.1 M MES pH 6.5	0.1 M MES, pH 6.5	0.1 M HEPES, pH 7.5
Precipitant[1]	5 M Na Formate	25% PEG 3350	2.0 M Amm Sulfate	10% PEG 8000
Additive[1]		0.2 M Na Formate	5% MPD	8% EG
Space Group	$P3_121$	C2	P422	$P42_12$
Molecules/AU	1	2	2	1
Unit Cell				
a(Å)	129.2	212.0	152.5	120.0
b(Å)	129.2	55.1	152.5	120.0
c(Å)	91.8	80.3	123.4	64.2
β(°)	90.0	97.8	90.0	90.0
γ(°)	120.0	90.0	90.0	90.0
V_m (Å^3/Da)	4.67	2.44	3.77	2.39
Solvent Content (%)	74	50	67	48
X-Ray Data[2]				
Resolution (Å)	30-2.6 (2.7-2.6)	30.0-1.9 (1.95-1.9)	30.0-3.3 (3.4-3.3)	30-1.9 (2.0-1.9)
Measured Reflections	136,745 (8,650)	241,145 (16,580)	237,504 (15,007)	801,080 (19,309)
Unique Reflections	27,349 (1,730)	71,932 (5,198)	22,379 (1,590)	35,965 (2,194)
Completeness (%)	99.3 (98.7)	99.0 (97.3)	99.5 (96.8)	98.5 (82.8)
Redundancy	5.0 (5.0)	3.4 (3.2)	10.6 (9.4)	22.3 (8.8)
R_{merge}	0.048 (0.522)	0.044 (0.245)	0.086 (0.536)	0.093 (0.231)
$< I/\sigma >$	21.2 (3.9)	17.8 (4.7)	25.5 (4.5)	29.2 (8.1)
B-factor (Å^2)	60.5	33.2	61.0	19.6
Refinement				
Resolution (Å)	15-2.6	15-1.9	15-3.3	15-1.9
Number of Reflections	26,238	70,346	21,197	34,850
Number of All Atoms	3,224	6,975	6,398	3,695
Number of Waters	2	472	0	399
R-factor (%)	20.5	19.2	20.2	16.7
R-free (%)	24.1	22.2	24.7	21.3
RMSD				
Bond Lengths (Å)	0.006	0.005	0.005	0.008
Bond Angles (°)	1.2	1.1	1.0	1.1
Mean B-factor (Å^2)	65.3	34.4	80.1	20.0
Ramachandran Plot (%)				
Outliers	0.0	0.0	0.9	0.0
Favored	92.3	96.9	93.1	96.9

[1]Abbreviations: Amm, ammonium;EG, ethylene glycol; PEG, polyethylene glycol.
[2]Values for high-resolution shell are in parentheses.

Fab	H3-23:L1-39	H3-23:L3-11	H3-23:L3-20	H3-23:L4-1
PDB identifier	5I19	5I1A	5I1C	5I1D
Crystal Data				
Crystallization Solution				
Buffer, pH	No Buffer	0.1 M Na Acetate, pH 4.5	0.1 M MES, pH 6.5	0.1 M HEPES, pH 7.5
Precipitant[1]	20% PEG 3350	2.0 M Amm Sulfate	16% PEG 3350	2.0 M Amm Sulfate
Additive[1]	0.2 M Li Citrate	5% PEG 400	0.2 M Amm Acetate	2% PEG 400
Space Group	$P4_12_12$	$P2_12_12_1$	$P6_222$	$P2_12_12_1$
Molecules/AU	1	2	1	2
Unit Cell				
a(Å)	96.6	60.9	121.5	62.7
b(Å)	96.6	110.6	121.5	111.0
c(Å)	105.4	158.9	160.4	160.0
β(°)	90	90	90	90
V_m (Å^3/Da)	2.60	2.82	3.60	2.90
Solvent Content (%)	53	56	66	57
X-Ray Data[2]				
Resolution (Å)	30-2.8 (2.9-2.8)	30-2.0 (2.1-2.0)	30-2.25 (2.3-2.25)	30-2.0 (2.1-2.0)
Measured Reflections	177,681 (12,072)	351,312 (8,634)	887,349 (59,919)	873,523 (49,118)
Unique Reflections	12,678 (899)	58,989 (2,870)	32,572 (2,300)	75,540 (5,343)
Completeness (%)	99.5 (97.4)	80.9 (54.2)	96.9 (94.8)	99.7 (96.9)
Redundancy	14.0 (13.4)	6.0 (3.0)	27.2 (26.1)	11.6 (9.2)
R_{merge}	0.091 (0.594)	0.066 (0.204)	0.086 (0.478)	0.094 (0.488)
$< I/\sigma >$	31.2 (5.1)	20.4 (4.6)	37.0 (10.4)	21.6 (5.0)
B-factor (Å^2)	42.8	27.1	33.7	29.4
Refinement				
Resolution (Å)	15-2.8	15-2.0	15-2.25	15-2.0

(*Continued on next page*)

Table 1. (*Continued*).

Fab	H3-23:L1-39	H3-23:L3-11	H3-23:L3-20	H3-23:L4-1
PDB identifier	5I19	5I1A	5I1C	5I1D
Number of Reflections	11,972	57,599	31,411	74,238
Number of All Atoms	3,234	6,948	3,472	7,210
Number of Waters	0	416	222	635
R-factor (%)	23.9	20.5	22.0	21.6
R-free (%)	31.5	25.5	26.6	25.1
RMSD				
Bond Lengths (Å)	0.009	0.010	0.005	0.008
Bond Angles (°)	1.3	1.3	1.0	1.1
Mean B-factor (Å^2)	48.4	36.7	47.7	46.4
Ramachandran Plot (%)				
Outliers	0.0	0.0	0.0	0.0
Favored	92.3	96.8	97.5	97.6

[1]Abbreviations: Amm, ammonium; PEG, polyethylene glycol.
[2]Values for high-resolution shell are in parentheses.

Fab	H3-53:L1-39	H3-53:L3-11	H3-53:L3-20	H3-53:L4-1
PDB indentifier	5I1E	5I1G	5I1H	5I1I
Crystal Data				
Crystallization Solution				
Buffer, pH	No buffer	0.1 M Na Acetate pH 4.5	0.1 M Na Acetate pH 4.5	0.1M MES, pH 6.5
Precipitant[1]	16% PEG 3350	25% PEG 3350	19% PEG 4000	17% PEG 3350
Additive[1]	0.2 M Amm Sulfate 5% Dioxane	0.2 M Li_2SO_4	0.2 M Amm Sulfate	0.2 M Na Formate, 5% MPD
Space Group	$P6_522$	$P6_522$	$P6_522$	$P3_1$
Molecules/AU	1	1	1	1
Unit Cell				
a(Å)	89.4	88.1	89.4	68.1
b(Å)	89.4	88.1	89.4	68.1
c(Å)	212.4	219.6	211.7	95.6
β(°)	90	90	90	90
γ(°)	120	120	120	120
V_m (Å^3/Da)	2.57	2.64	2.57	2.64
Solvent Content (%)	52	53	52	53
X-Ray Data[2]				
Resolution (Å)	30-2.7 (2.8-2.7)	30-2.3 (2.4-2.3)	30-2.2 (2.3-2.0)	30-2.5 (2.6-2.5)
Measured Reflections	297,367 (19,369)	333,739 (8,008)	381,125 (1,591)	137,992 (9,883)
Unique Reflections	14,402 (1,003)	21,683 (1,135)	24,323 (964)	16,727 (1,227)
Completeness (%)	99.6 (96.8)	93.8 (68.4)	95.3 (52.0)	98.6 (98.1)
Redundancy	20.6 (19.3)	15.4 (7.1)	15.7 (1.7)	8.2 (8.1)
R_{merge}	0.095 (0.451)	0.057 (0.324)	0.062 (0.406)	0.047 (0.445)
$< I/\sigma >$	38.3 (8.1)	36.7 (5.5)	36.2 (1.6)	31.6 (5.6)
B-factor (Å^2)	33.2	37.3	33.7	54.8
Refinement				
Resolution (Å)	15-2.7	15-2.3	15-2.2	15-2.5
Number of Reflections	13,583	20,255	24,962	15,811
Number of All Atoms	3,335	3,271	3,298	3,239
Number of Waters	88	70	71	21
R-factor (%)	19.1	29.8	22.8	25.0
R-free (%)	26.4	38.3	26.6	33.7
RMSD				
Bond Lengths (Å)	0.008	0.005	0.005	0.006
Bond Angles (°)	1.2	1.0	1.0	1.1
Mean B-factor (Å^2)	49.1	46.3	51.7	88.9
Ramachandran Plot (%)				
Outliers	0.2	0.2	0.2	1.2
Favored	96.7	97.1	96.5	90.9

[1]Abbreviations: Amm, ammonium; PEG, polyethylene glycol.
[2]Values for high-resolution shell are in parentheses.

Fab	H5-51:L1-39	H5-51:L3-11	H5-51:L3-20	H5-51:L4-1
PDB identifier	4KMT	5I1J	5I1K	5I1L
Crystal Data				
Crystallization Solution				
Buffer, pH	0.1 M CHES, pH 9.5	0.1 M Tris, pH 8.5	0.1 M CHES, pH 9.5	0.1 M Tris, pH 8.5
Precipitant[1]	1.8 M Amm Sulfate	25% PEG 3350	1.0 M Amm Sulfate	24% PEG 3350
Additive[1]	5% dioxane	0.2 M $MgCl_2$		0.2 M Amm Sulfate

(*Continued on next page*)

Table 1. (*Continued*).

Fab	H5-51:L1-39	H5-51:L3-11	H5-51:L3-20	H5-51:L4-1
PDB identifier	4KMT	5I1J	5I1K	5I1L
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
Molecules/AU	1	1	1	2
Unit Cell				
a(Å)	63.7	64.1	63.8	106.0
b(Å)	73.8	73.8	74.1	38.0
c(Å)	103.1	103.0	103.0	112.3
β(°)	90	90	90	100.4
V_m (Å^3/Da)	2.53	2.56	2.54	2.28
Solvent Content (%)	51	52	51	46
X-Ray Data[2]				
Resolution (Å)	30-2.1 (2.2-2.1)	30-2.5 (2.6-2.5)	30-1.65 (1.7-1.65)	30-1.95 (2.0-1.95)
Measured Reflections	131,839 (6,655)	120,521 (7,988)	246,750 (4,142)	320,324 (12,119)
Unique Reflections	27,026 (1,885)	17,286 (1,236)	53,058 (2,141)	61,554 (3,243)
Completeness (%)	93.6 (89.8)	99.7 (97.3)	89.8 (49.8)	94.4 (67.1)
Redundancy	4.9 (3.5)	7.0 (6.5)	4.7 (1.9)	5.2 (3.7)
R_{merge}	0.079 (0.278)	0.080 (0.281)	0.034 (0.131)	0.060 (0.395)
$< I/\sigma >$	16.8 (5.7)	21.1(6.9)	27.5 (5.8)	19.7 (3.1)
B-factor (Å^2)	26.0	27.0	21.6	31.4
Refinement				
Resolution (Å)	15-2.1	15-2.5	15-1.65	15-1.95
Number of Reflections	25,857	16,328	51,882	60,181
Number of All Atoms	3,676	3,454	3,814	7,175
Number of Waters	302	196	527	445
R-factor (%)	17.1	17.7	17.2	19.4
R-free (%)	22.0	25.8	19.7	25.8
RMSD				
Bond Lengths (Å)	0.006	0.009	0.005	0.009
Bond Angles (°)	1.0	1.3	1.3	1.3
Mean B-factor (Å^2)	25.2	38.2	20.0	19.5
Ramachandran Plot (%)				
Outliers	0.0	0.0	0.0	0.0
Favored	98.4	97.9	98.1	98.0

[1]Abbreviations: Amm, ammonium; PEG, polyethylene glycol.
[2]Values for high-resolution shell are in parentheses.

The crystal structures of the 16 Fabs have been determined at resolutions ranging from 3.3 Å to 1.65 Å (Table 1). The number of Fab molecules in the crystallographic asymmetric unit varies from 1 (for 12 Fabs) to 2 (for 4 Fabs). Overall the structures are fairly complete, and, as can be expected, the models for the higher resolution structures are more complete than those for the lower resolution structures (Table S1). Invariably, the HCs have more disorder than the LCs. For the LC, the disorder is observed at 2 of the C-terminal residues with few exceptions. Apart from the C-terminus, only a few surface residues in LC are disordered.

The HCs feature the largest number of disordered residues, with the lower resolution structures having the most. The C-terminal residues including the 6xHis tags are disordered in all 16 structures. In addition to these, 2 primary disordered stretches of residues are observed in a number of structures (Table S1). One involves the loop connecting the first 2 β-strands of the constant domain (in all Fabs except **H3-23:L1-39**, **H3-23:L3-11** and **H3-53:L1-39**). The other is located in CDR H3 (in **H5-51:L3-11**, **H5-51:L3-20** and in one of 2 copies of **H3-23:L4-1**). CDR H1 and CDR H2 also show some degree of disorder, but to a lesser extent.

CDR canonical structures

Several CDR definitions have evolved over decades of antibody research.[2-4] Depending on the focus of the study, the CDR boundaries differ slightly between various definitions. In this work, we use the CDR definition of North et al.,[12] which is similar to that of Martin [11] with the following exceptions: 1) CDRs H1 and H3 begin immediately after the Cys; and 2) CDR L2 includes an additional residue at the N-terminal side, typically Tyr.

CDR H1

The four HCs feature CDR H1 of the same length, and their sequences are highly similar (Table 2). The CDR H1 backbone conformations for all variants for each of the HCs are shown in Fig. 1. Three of the HCs, **H3-23**, **H3-53** and **H5-51**, have the same canonical structure, H1-13-1, and the backbone conformations are tightly clustered for each set of Fab structures as reflected in the rmsd values (Fig. 1B-D). Some deviation is observed for **H3-53**, mostly due to **H3-53:L4-1**, which exhibits a significant degree of disorder in CDR H1. The electron density for the backbone is weak and discontinuous, and completely missing for several side chains.

The CDR H1 structures with **H1-69** shown in Fig. 1A are quite variable, both for the structures with different LCs and for the copies of the same Fab in the asymmetric unit, **H1-69:L3-11** and **H1-69:L3-20**. In total, 6 independent Fab structures produce 5 different canonical structures, namely H1-13-1, H1-13-3, H1-13-4, H1-13-6 and H1-13-10. A major difference of **H1-69** from the other germlines in the experimental data set is

Table 2. Canonical structures.[1]

Pairs	PDB	CDR H1	CDR H2	CDR H3
H1-69		KASGGTFSSYAIS	GIIPIFGTAN	ARYDGIYGELDF
H1-69:L1-39	5I15	H1-13-4	H2-10-1	H3-12-1
H1-69:L3-11	5I16	H1-13-1/H1-13-1	H2-10-1/H2-10-1	H3-12-1/H3-12-1
H1-69:L3-20	5I17	H1-13-3/H1-13-6	H2-10-1/NA	H3-12-1/H3-12-1
H1-69:L4-1	5I18	H1-13-10	H2-10-1	H3-12-1
H3-23		AASGFTFSSYAMS	AISGSGGSTY	AKYDGIYDGIYGELDF
H3-23:L1-39	5I19	H1-13-1	H2-10-2	H3-12-1
H3-23:L3-11	5I1A	H1-13-1/H1-13-1	H2-10-2/H2-10-2	H3-12-1/H3-12-1
H3-23:L3-20	5I1C	H1-13-1	H2-10-2	H3-12-1
H3-23:L4-1	5I1D	H1-13-1/H1-13-1	H2-10-2/H2-10-2	H3-12-1/NA
H3-53		AASGFTVSSNYMS	VIYSGGSTY	ARYDGIYGELDF
H3-53:L1-39	5I1E	H1-13-1	H2-9-3	H3-12-1
H3-53:L3-11	5I1G	H1-13-1	H2-9-3	H3-12-1
H3-53:L3-20	5I1H	H1-13-1	H2-9-3	H3-12-1
H3-53:L4-1	5I1I	H1-13-1	H2-9-3	NA
H5-51		KGSGYSFTSYWIG	IIYPGDSDTR	ARYDGIYGELDF
H5-51:L1-39	4KMT	H1-13-1	H2-10-1	H3-12-1
H5-51:L3-11	5I1J	H1-13-1	H2-10-1	NA
H5-51:L3-20	5I1K	H1-13-1	H2-10-1	NA
H5-51:L4-1	5I1L	H1-13-1/H1-13-1	H2-10-1/H2-10-1	H3-12-1/H3-12-1
		CDR L1	**CDR L2**	**CDR L3**
L1-39		RASQSISSYLN	YAASSLQS	QQSYSTPLT
H1-69:**L1-39**	5I15	L1-11-1	L2-8-1	L3-9-cis7-1
H3-23:**L1-39**	5I19	L1-11-1	L2-8-1	L3-9-cis7-1
H3-53:**L1-39**	5I1E	L1-11-1	L2-8-1	L3-9-cis7-1
H5-51:**L1-39**	4KMT	L1-11-1	L2-8-1	L3-9-cis7-1
L3-11		RASQSVSSYLA	YDASNRAT	QQRSNWPLT
H1-69:**L3-11**	5I16	L1-11-1/L1-11-1	L2-8-1/L2-8-1	L3-9-cis7-1/L3-9-cis7-1
H3-23:**L3-11**	5I1A	L1-11-1/L1-11-1	L2-8-1/L2-8-1	L3-9-cis7-1/L3-9-cis7-1
H3-53:**L3-11**	5I1G	L1-11-1	L2-8-1	L3-9-cis7-1
H5-51:**L3-11**	5I1J	L1-11-1	L2-8-1	L3-9-cis7-1
L3-20		RASQSVSSSYLA	YGASSRAT	QQYGSSPLT
H1-69:**L3-20**	5I17	L1-12-2/L1-12-1	L2-8-1/L2-8-1	L3-9-cis7-1/L3-9-cis7-1
H3-23:**L3-20**	5I1C	L1-12-2	L2-8-1	L3-9-cis7-1
H3-53:**L3-20**	5I1H	L1-12-1	L2-8-1	L3-9-cis7-1
H5-51:**L3-20**	5I1K	L1-12-1	L2-8-1	L3-9-cis7-1
L4-1		KSSQSVLYSSNNKNYLA	YWASTRES	QQYYSTPLT
H1-69:**L4-1**	5I18	L1-17-1	L2-8-1	L3-9-cis7-1
H3-23:**L4-1**	5I1D	L1-17-1/L1-17-1	L2-8-1/L2-8-1	L3-9-cis7-1/L3-9-cis7-1
H3-53:**L4-1**	5I1I	L1-17-1	L2-8-1	L3-9-cis7-1
H5-51:**L4-1**	5I1L	L1-17-1/L1-17-1	L2-8-1/L2-8-1	L3-9-cis7-1/L3-9-cis7-1

[1]CDRs are defined using the Dunbrack convention [12]. Assignments for 2 copies of the Fab in the asymmetric unit are given for 5 structures. No assignment (NA) for CDRs with missing residues.

the presence of Gly instead of Phe or Tyr at position 27 (residue 5 of 13 in CDR H1). Glycine introduces the possibility of a higher degree of conformational flexibility that undoubtedly translates to the differences observed, and contributes to the elevated thermal parameters for the atoms in the amino acid residues in this region.

CDR H2

The canonical structures of CDR H2 have fairly consistent conformations (Table 2, Fig. 2). Each of the 4 HCs adopts only one canonical structure regardless of the pairing LC. Germlines **H1-69** and **H5-51** have the same canonical structure assignment H2-10-1, **H3-23** has H2-10-2, and **H3-53** has H2-9-3. The conformations for all of these CDR H2s are tightly clustered (Fig. 2). In one case, in the second Fab of **H1-69:L3-20**, CDR H2 is partially disordered (Δ55-60).

Although three of the germlines have CDR H2 of the same length, 10 residues, they adopt 2 distinctively different conformations depending mostly on the residue at position 71 from the so-called CDR H4.[26] Arg71 in **H3-23** fills the space between CDRs H2 and H4, and defines the conformation of the tip of CDR H2 so that residue 54 points away from the antigen binding site. Germlines **H1-69** and **H5-51** are unique in the human repertoire in having an Ala at position 71 that leaves enough space for H-Pro52a to pack deeper against CDR H4 so that the following residues 53 and 54 point toward the putative antigen.

Conformations of CDR H2 in **H1-69** and **H5-51**, both of which have canonical structure H2-10-1, show little deviation within each set of 4 structures. However, there is a significant shift of the CDR as a rigid body when the 2 sets are superimposed. Most likely this is the result of interaction of CDR H2 with CDR H1, namely with the residue at position 33 (residue 11 of 13 in CDR H1). Germline **H1-69** has Ala at position 33 whereas in **H5-51** position 33 is occupied by a bulky Trp, which stacks against H-Tyr52 and drives CDR H2 away from the center.

CDR L1

The four LC CDRs L1 feature 3 different lengths (11, 12 and 17 residues) having a total of 4 different canonical structure assignments. Of these LCs, **L1-39** and **L3-11** have the same

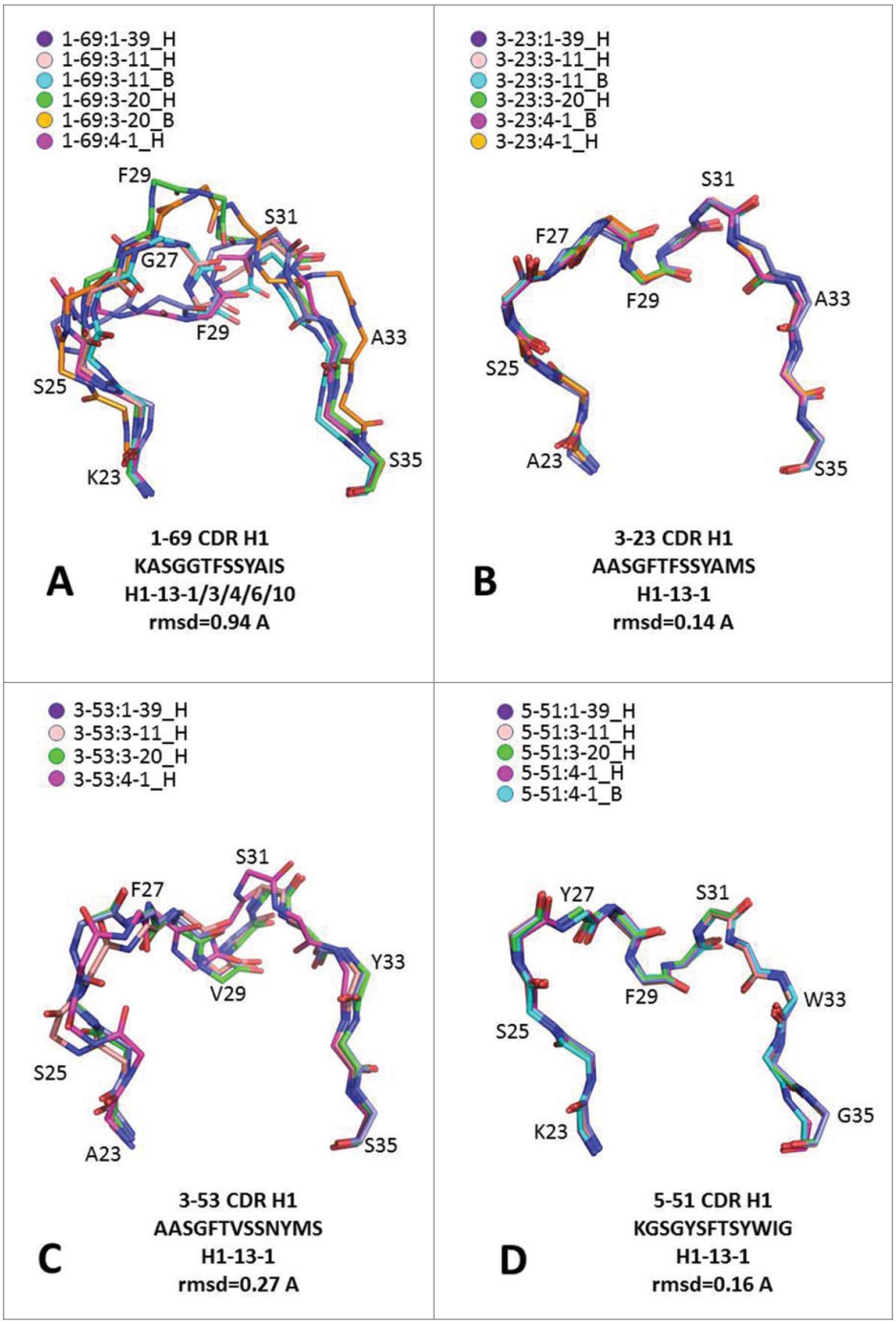

Figure 1. The superposition of CDR H1 backbones for all HC:LC pairs with heavy chains: (A) **H1-69**, (B) **H3-23**, (C) **H3-53** and (D) **H5-51**.

canonical structure, L1-11-1, and superimpose very well (Fig. 3A, B). For the remaining 2, **L3-20** has 2 different assignments, L1-12-1 and L1-12-2, while **L4-1** has a single assignment, L1-17-1.

L4-1 has the longest CDR L1, composed of 17 amino acid residues (Fig. 3D). Despite this, the conformations are tightly clustered (rmsd is 0.20 Å). The backbone conformations of the stem regions superimpose well. Some changes in conformation occur between residues 30a and 30f (residues 8 and 13 of 17 in CDR L1). This is the tip of the loop region, which appears to have similar conformations that fan out the structures because of the slight differences in torsion angles in the backbone near Tyr30a and Lys30f.

L3-20 is the most variable in CDR L1 among the 4 germlines as indicated by an rmsd of 0.54 Å (Fig. 3C). Two structures, **H3-53:L3-20** and **H5-51:L3-20** are assigned to canonical structure L1-12-1 with virtually identical backbone conformations. The third structure, **H3-23:L3-20**, has CDR L1 as L1-12-2, which deviates from L1-12-1 at residues 29-32, i.e., at the site of insertion with respect to the 11-residue CDR. The fourth member of the set, **H1-69:L3-20**, was crystallized with 2 Fabs in the asymmetric unit. The conformation of CDR L1 in these 2 Fabs is slightly different, and both conformations fall somewhere between L1-12-1 and L1-12-2. This reflects the lack of accuracy in the structure due to low resolution of the X-ray data (3.3 Å).

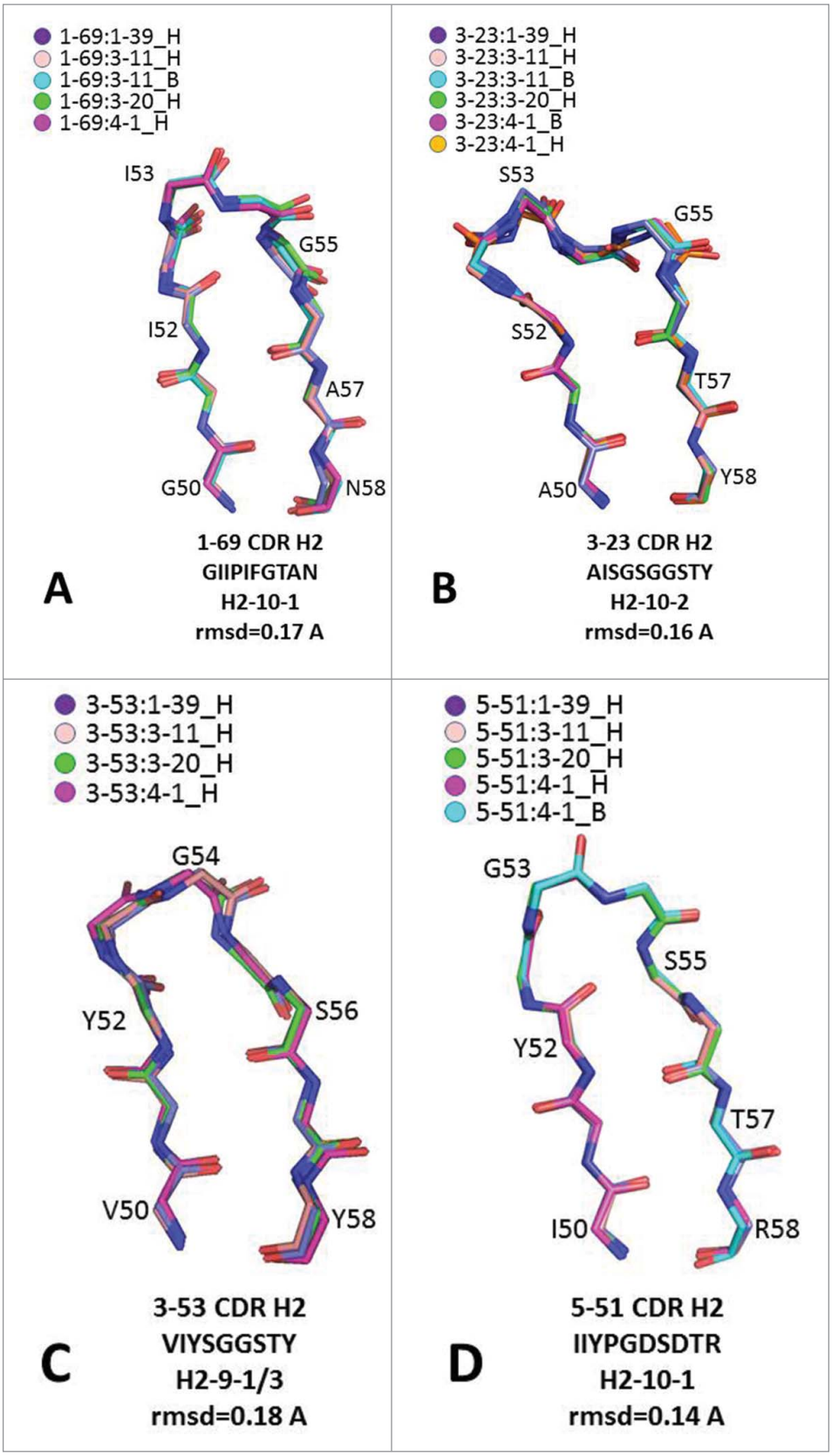

Figure 2. The superposition of CDR H2 backbones for all HC:LC pairs with heavy chains: (A) **H1-69**, (B) **H3-23**, (C) **H3-53** and (D) **H5-51**.

CDR L2

All four LCs have CDR L2 of the same length and canonical structure, L2-8-1 (Table 2). The CDR L2 conformations for each of the LCs paired with the 4 HCs are clustered more tightly than any of the other CDRs (rmsd values are in the range 0.09-0.16 Å), and all 4 sets have virtually the same

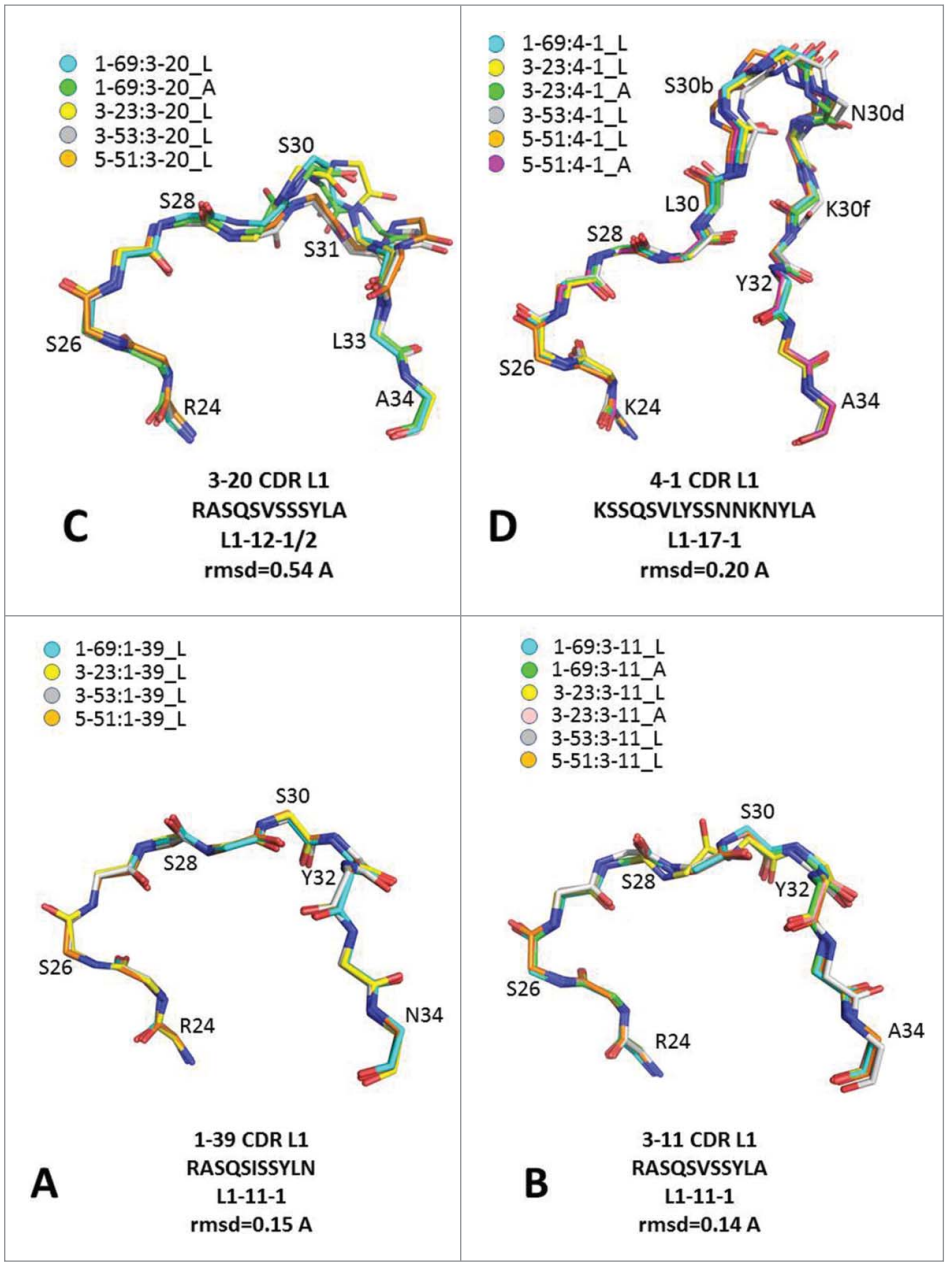

Figure 3. The superposition of CDR L1 backbones for all HC:LC pairs with light chains: (A) **L1-39**, (B) **L3-11**, (C) **L3-20** and (D) **L4-1**.

conformation despite the sequence diversity of the loop. No significant conformation outliers are observed (Fig. 4).

CDR L3

As with CDR L2, all 4 LCs have CDR L3 of the same length and canonical structure, L3-9-cis7-1 (Table 2). The conformations of CDR L3 for **L1-39, L3-11**, and particularly for **L320**, are not as tightly clustered as those of **L4-1** (Fig. 5). The slight conformational variability occurs in the region of amino acid residues 90-92, which is in contact with CDR H3.

CDR H3 conformational diversity

As mentioned earlier, all 16 Fabs have the same CDR H3, for which the amino acid sequence is derived from the anti-CCL2 antibody CNTO 888.[24] The loop and the 2 β-strands of the CDR H3 in this 'parent' structure are stabilized by H-bonds between the carbonyl oxygen and peptide nitrogen atoms in the 2 strands. An interesting feature of these CDR H3 structures is the presence of a water molecule that interacts with the peptide nitrogens and carbonyl oxygens near the bridging loop connecting the 2 β-strands. This water is present in both the bound (**4DN4**) and unbound (**4DN3**) forms of CNTO 888. The stem region of CDR H3 in the parental Fab is in a 'kinked' conformation,[27,14] in which the indole nitrogen of Trp103 forms a hydrogen bond with the carbonyl oxygen of Leu100b. The carboxyl group of Asp101 forms a salt bridge with Arg94. These interactions are illustrated in Fig. S2.

Despite having the same amino acid sequence in all variants, CDR H3 has the highest degree of structural diversity and

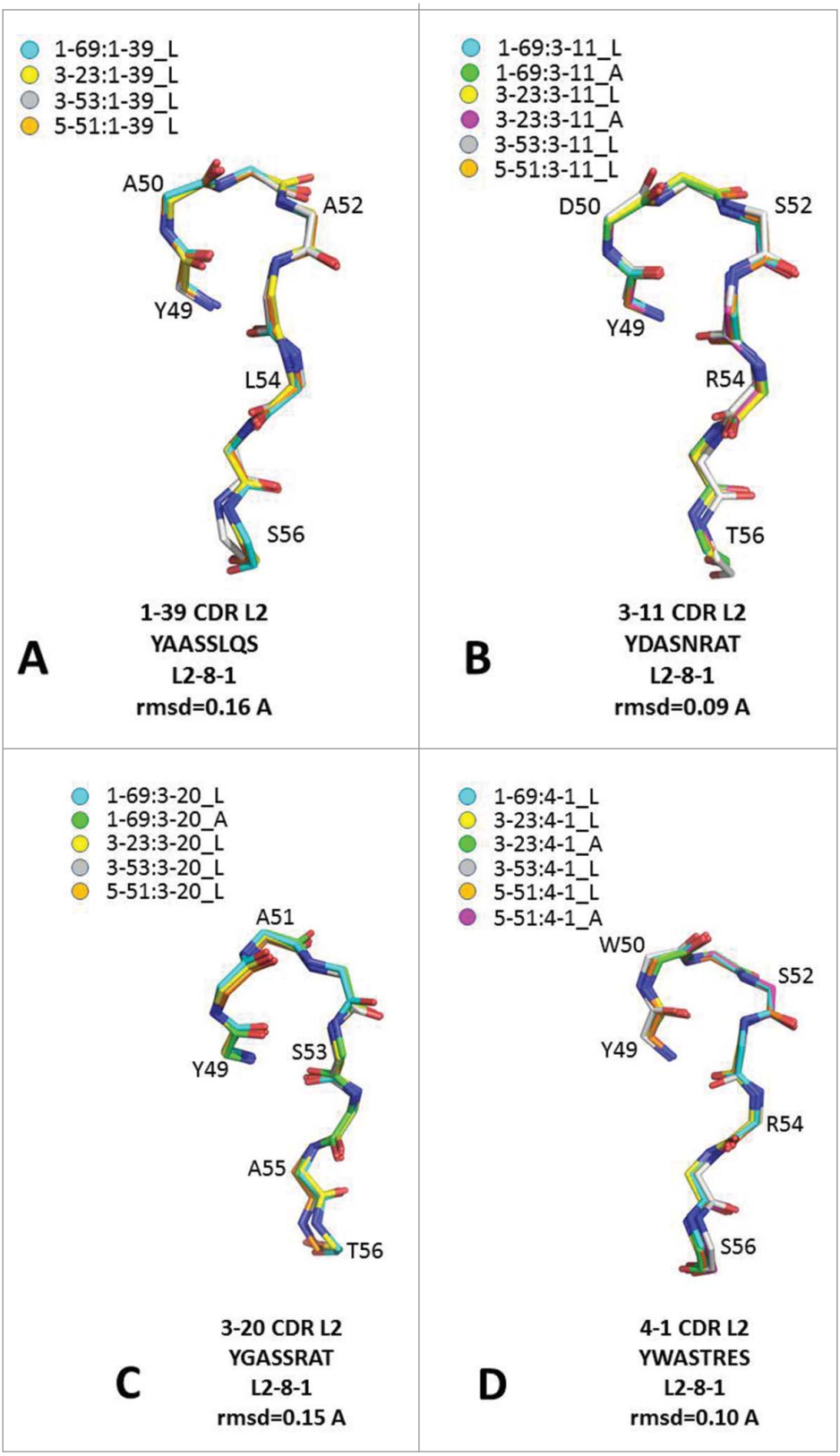

Figure 4. The superposition of CDR L2 backbones for all HC:LC pairs with light chains: (A) **L1-39**, (B) **L3-11**, (C) **L3-20** and (D) **L4-1**.

disorder of all of the CDRs in the experimental set. Three of the 21 Fab structures (including multiple copies in the asymmetric unit), **H5-51:L3-11**, **H551:L3-20** and **H3-23:L4-1** (one of the 2 Fabs), have missing (disordered) residues at the apex of the CDR loop. Another four of the Fabs, **H3-23:L1-39**, **H3-53:L1-39**, **H3-53:L3-11** and **H3-53:L4-1** have missing side-chain atoms. The variations in CDR H3 conformation are illustrated in Fig. 6 for the 18 Fab structures that have ordered backbone atoms.

In 10 of the 18 Fab structures, **H1-69:L1-39**, **H1-69:L3-11** (2 Fabs), **H1-69:L4-1, H3-23:L3-11** (2 Fabs), **H3-23:L3-20**,

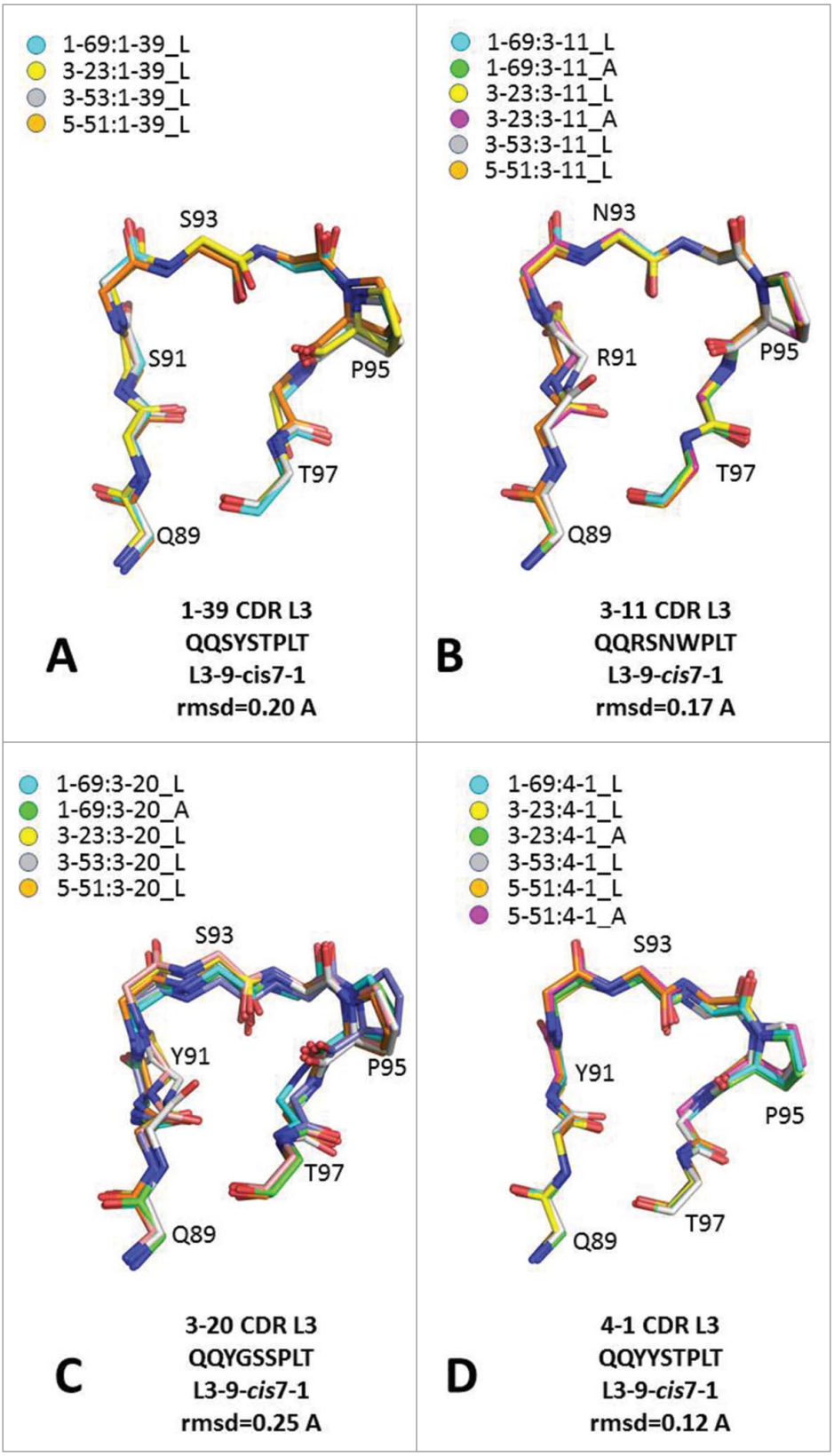

Figure 5. The superposition of CDR L3 backbones for all HC:LC pairs with light chains: (A) **L1-39**, (B) **L3-11**, (C) **L3-20** and (D) **L4-1**.

H3-53:L3-11, **H3-53:L3-20** and **H5-51:L1-39**, the CDRs have similar conformations to that found in **4DN3**. The bases of these structures have the 'kinked' conformation with the H-bond between Trp103 and Leu100b. A representative CDR H3 structure for **H1-69:L1-39** illustrating this is shown in Fig. 7A. The largest backbone conformational deviation for the set is at Tyr99, where the C=O is rotated by 90° relative to that observed in **4DN3**. Also, it is worth noting that only one of these structures, **H1-69:L4-1**, has the conserved water molecule in CDR H3 observed in the **4DN3** and **4DN4** structures. In

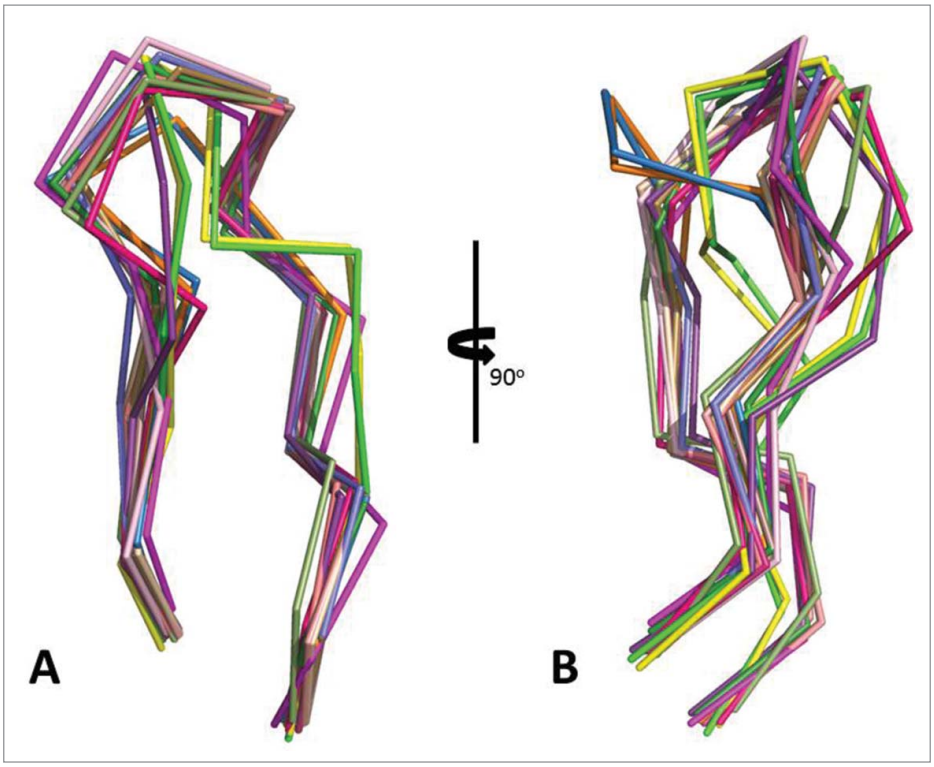

Figure 6. Ribbon representations of (A) the superposition of all CDR H3s of the structures with complete backbone traces. (B) The CDR H3s rotated 90° about the *y* axis of the page. The structure of each CDR H3 is represented with a different color.

fact, it is the only Fab in the set that has a water molecule present at this site. The CDR H3 for this structure is shown in Fig. S3.

The remaining 8 Fabs can be grouped into 5 different conformational classes. Three of the Fabs, **H3-23:L1-39**, **H3-23:L4-1** and **H3-53:L1-39**, have distinctive conformations. The stem regions in these 3 cases are in the 'kinked' conformation consistent with that observed for **4DN3**. The five remaining Fabs, **H5-51:L4-1** (2 copies), **H1-69:L3-20** (2 copies) and **H3-53:L4-1**, have 3 different CDR H3 conformations (Fig. S4). The stem regions of CDR H3 for the **H5-51:L4-1** Fabs are in the 'kinked' conformation while, surprisingly, those of the **H1-69:L3-20** pair and **H3-53:L4-1** are in the 'extended' conformation (Fig. 7B).

VH:VL domain packing

The VH and VL domains have a β-sandwich structure (also often referred as a Greek key motif) and each is composed of a 4-stranded and a 5-stranded antiparallel β-sheets. The two domains pack together such that the 5-stranded β-sheets, which have hydrophobic surfaces, interact with each other bringing the CDRs from both the VH and VL domains into close proximity. The domain packing of the variants was assessed by computing the domain interface interactions, the VH:VL tilt angles, the buried surface area and surface complementarity. The results of these analyses are shown in Tables 3, 4 and S2.

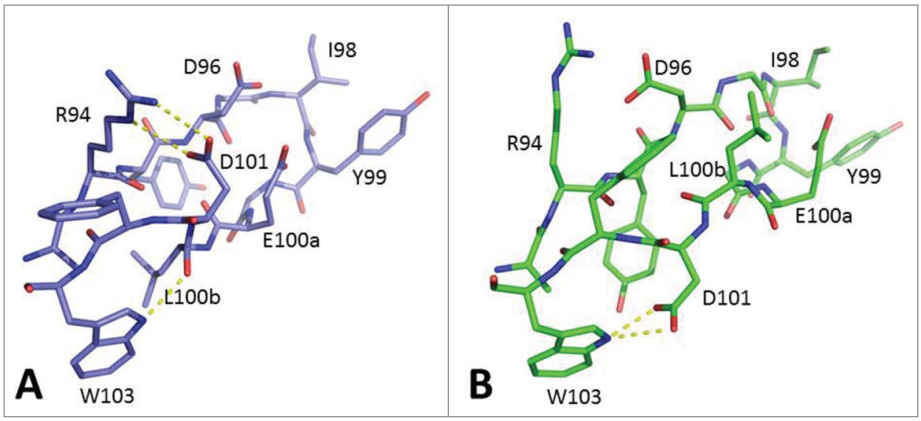

Figure 7. A comparison of representatives of the "kinked" and "extended" structures. (A) The "kinked" CDR H3 of **H1-69:L3-11** with purple carbon atoms and yellow dashed lines connecting the H-bond pairs for Leu100b O and Trp103 NE1, Arg94 NE and Asp101 OD1, and Arg94 NH2 and Asp101 OD2. (B) The "extended" CDR H3 of **H1-69:L3-20** with green carbon atoms and yellow dashed lines connecting the H-bond pairs for Asp101 OD1 and OD2 and Trp103 NE1.

Table 3. Differences in VH:VL tilt angles.

	H1-69:L1-39	H1-69:L3-11	H1-69:L3-20	H1-69:L4-1	H3-23:L1-39	H3-23:L3-11	H3-23:L3-20	H3-23:L4-1	H3-53:L1-39	H3-53:L3-11	H3-53:L3-20	H3-53:L4-1	H5-51:L1-39	H5-51:L3-11	H5-51:L3-20	H5-51:L4-1
H1-69:L1-39	0	2.1	8.9	1.1	4.2	3.0	9.5	1.5	3.3	3.6	3.1	1.6	1.8	2.9	2.4	5.2
H1-69:L3-11		0	7.3	2.9	2.5	2.0	8.4	1.3	2.6	2.9	3.2	1.8	3.9	4.6	4.4	5.0
H1-69:L3-20			0	9.2	5.0	8.7	7.4	7.6	8.9	8.6	9.4	7.9	10.5	10.1	11.0	9.7
H1-69:L4-1				0	4.6	3.9	10.1	1.8	4.4	4.7	4.1	2.3	1.6	2.5	2.3	6.2
H3-23:L1-39					0	4.0	8.0	2.8	4.8	4.8	5.4	3.6	6.0	6.2	6.6	6.6
H3-23:L3-11						0	9.3	3.0	2.1	2.9	3.3	3.3	4.6	5.8	5.0	5.2
H3-23:L3-20							0	8.9	7.9	7.0	7.6	7.9	10.5	9.7	10.7	6.2
H3-23:L4-1								0	3.6	3.8	3.7	1.5	3.2	3.7	3.8	5.6
H3-53:L1-39									0	1.0	1.6	2.9	4.6	5.3	4.8	3.1
H3-53:L3-11										0	1.3	2.9	4.8	5.2	5.0	2.3
H3-53:L3-20											0	2.5	3.8	4.2	3.9	2.2
H3-53:L4-1												0	2.9	3.0	3.3	4.2
H5-51:L1-39													0	1.9	0.6	5.8
H5-51:L3-11														0	1.9	5.7
H5-51:L3-20															0	5.8
H5-51:L4-1																0

Table 4. VH:VL surface areas and surface complementarity.

Chain Pairs	PDB	Contact surfaceVH (Å^2)	Contact surfaceVL (Å^2)	Interface(Å^2)	Surface complementarity
H1-69:L1-39	5I15	727	771	749	0.743
H1-69:L3-11	5I16	802	870	836	0.762
H1-69:L3-20	5I17	713	736	725	0.723
H1-69:L4-1	5I18	729	736	733	0.734
H3-23:L1-39[1]	5I19	795	817	806	0.722
H3-23:L3-11	5I1A	822	834	828	0.725
H3-23:L3-20	5I1C	670	698	684	0.676
H3-23:L4-1	5I1D	743	770	757	0.708
H3-53:L1-39[1]	5I1E	698	719	709	0.712
H3-53:L3-11[1]	5I1G	747	758	753	0.690
H3-53:L3-20	5I1H	743	735	739	0.687
H3-53:L4-1	5I1I	689	693	691	0.711
H5-51:L1-39	4KMT	761	808	785	0.728
H5-51:L3-11[2]	5I1J	648	714	681	0.717
H5-51:L3-20[2]	5I1K	622	643	633	0.740
H5-51:L4-1	5I1L	790	792	791	0.704

[1]Some side chain atoms in CDR H3 are missing.
[2]Residues in CDR H3 are missing: YGE in **H5-51:L3-11**, GIY in **H5-51:L3-20**.

VH:VL interface amino acid residue interactions

The VH:VL interface is pseudosymmetric, and involves 2 stretches of the polypeptide chain from each domain, namely CDR3 and the framework region between CDRs 1 and 2. These stretches form antiparallel β-hairpins within the internal 5-stranded β-sheet. There are a few principal inter-domain interactions that are conserved not only in the experimental set of 16 Fabs, but in all human antibodies. They include: 1) a bidentate hydrogen bond between L-Gln38 and H-Gln39; 2) H-Leu45 in a hydrophobic pocket between L-Phe98, L-Tyr87 and L-Pro44; 3) L-Pro44 stacked against H-Trp103; and 4) L-Ala43 opposite the face of H-Tyr91 (Fig. 8). With the exception of L-Ala43, all other residues are conserved in human germlines. Position 43 may be alternatively occupied by Ser, Val or Pro (as in **L4-1**), but the hydrophobic interaction with H-Tyr91 is preserved. These core interactions provide enough stability to the VH:VL dimer so that additional VH-VL contacts can tolerate amino acid sequence variations in CDRs H3 and L3 that form part of the VH:VL interface.

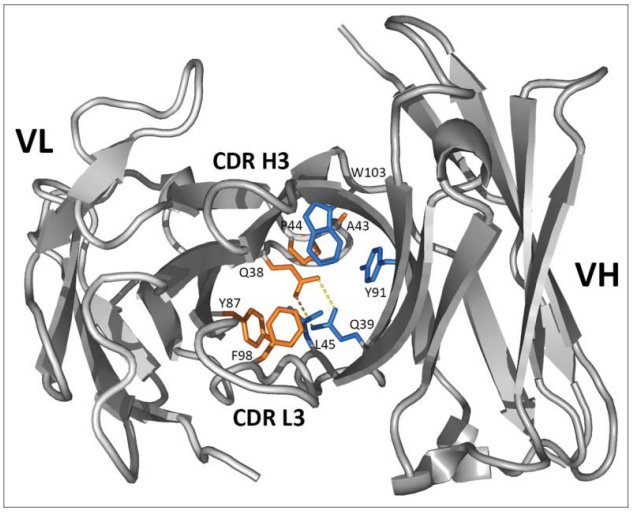

Figure 8. The conserved VH:VL interactions as viewed along the VH/VL axis. The VH residues are in blue, the VL residues are in orange.

In total, about 20 residues are involved in the VH:VL interactions on each side (Fig. S5). Half of them are in the framework regions and those residues (except residue 61 in HC, which is actually in CDR2 in Kabat's definition) are conserved in the set of 16 Fabs. The side chain conformations of these conserved residues are also highly similar. One notable exception is H-Trp47, which exhibits 2 conformations of the indole ring. In most of the structures, it has the χ_2 angle of $\sim$80°, while the ring is flipped over ($\chi_2 = -100°$) in **H5-51:L3:11** and **H5-51:L3-20**. Interestingly, these are the only 2 structures with residues missing in CDR H3 because of disorder, although both structures are determined at high resolution and the rest of the structure is well defined. Apparently, residues flanking CDR H3 in the 2 VH:VL pairings are inconsistent with any stable conformation of CDR H3, which translates into a less restricted conformational space for some of them, including H-Trp47.

VH:VL tilt angles

The relative orientation of VH and VL has been measured in a number of different ways.[28-31] Presented here are the results of 2 different approaches for determining the orientation of one domain relative to the other.

The first approach uses ABangles,[31] the results of which are shown in Table S2. The four LCs all are classified as Type A because they have a proline at position 44, and the results for each orientation parameter are within the range of values of this type reported by Dunbar and co-workers.[31] In fact, the parameter values for the set of 16 Fabs are in the middle of the distribution observed for 351 non-redundant antibody structures determined at 3.0 Å resolution or better. The only exception is **HC1**, which is shifted toward smaller angles with the mean value of 70.8° as compared to the distribution centered at 72° for the entire PDB. This probably reflects the invariance of CDR H3 in the current set as opposed to the CDR H3 diversity in the PDB.

The second approach used for comparing tilt angles [28,21] involved computing the difference in the tilt angles between all pairs of structures. For structures with 2 copies of the Fab in the asymmetric unit, only one structure was used. The

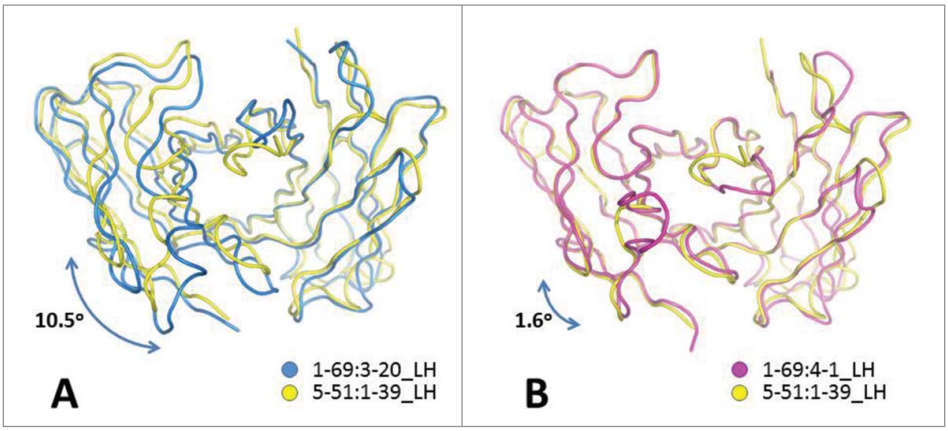

Figure 9. An illustration of the difference in tilt angle for 2 pairs of variants by the superposition of the VH domains of (A) **H1-69:L3-20** on that of **H5-51:L1-39** (the VL domain is off by a rigid-body roatation of 10.5°) and (B) **H1-69:L4-1** on that of **H5-51:L1-39** (the VL domain is off by a rigid-body roatation of 1.6°).

differences between independent Fabs in the same structure are 4.9° for **H1-69:L3-20**, 1.6° for **H1-69:L3-11**, 1.4° for **H3-23:L4-1**, 3.3° for **H3-23:L3-11**, and 2.5° for **H5-51:L4-1**. With the exception of **H1-69:L3-20**, the angles are within the range of 2-3° as are observed in the identical structures in the PDB.[21] In **H1-69:L3-20**, one of the Fabs is substantially disordered so that part of CDR H2 (the outer β-strand, residues 55-60) is completely missing. This kind of disorder may compromise the integrity of the VH domain and its interaction with the VL. Indeed, this Fab has the largest twist angle HC2 [31] within the experimental set that exceeds the mean value by 2.5 standard deviations (Table S2).

The differences in the tilt angle are shown for all pairs of V regions in Table 3. They range from 0.6° to 11.0°. The smallest differences in the tilt angle are between the Fabs in isomorphous crystal forms. The largest deviations in the tilt angle, up to 11.0°, are found for 2 structures, **H1-69:L3-20** and **H3-23:L3-20**, that stand out from the other Fabs. One of the 2 structures, **H1-69:L3-20**, has its CDR H3 in the 'extended' conformation; the other structure has it in the 'kinked' conformation. Two examples illustrating large (10.5°) and small (1.6°) differences in the tilt angles are shown in Fig. 9.

VH:VL buried surface area and complementarity

The results of the PISA contact surface calculation [32] and surface complementarity calculation [33] are shown in Table 4. The interface areas are calculated as the average of the VH and VL contact surfaces. Six of the 16 structures have CDR H3 side chains or complete residues missing, and therefore their interfaces are much smaller than in the other 10 structures with complete CDRs (the results are provided for all Fabs for completeness). Among the complete structures, the interface areas range from 684 to 836 Å^2. Interestingly, the 2 structures that have the largest tilt angle differences with the other variants, **H3-23:L3-20** and **H1-69:L3-20**, have the smallest VH:VL interfaces, 684 and 725 Å^2, respectively. **H3-23:L3-20** is also unique in that it has the lowest value (0.676) of surface complementarity.

Stability of germline pairings

Melting temperatures (T_m) were measured for all Fabs using differential scanning calorimetry (Table 5). It appears that for each given LC, the Fabs with germlines **H1-69** and **H3-23** are substantially more stable than those with germlines **H3-53** and **H5-51**. In addition, **L1-39** provides a much higher degree of stabilization than the other 3 LC germlines when combined with any of the HCs. As a result, the T_m for pairs **H1-69:L1-39** and **H3-23:L1-39** is 12-13° higher than for pairs **H3-53:L3-20**, **H3-53:L4-1**, **H5-51:L3-20** and **H5-51:L4-1**.

These findings correlate well with the degree of conformational disorder observed in the crystal structures. Parts of CDR H3 main chain are completely disordered, and were not modeled in Fabs **H5-51:L3-20** and **H5-51:L3-11** that have the lowest T_ms in the set. No electron density is observed for a number of side chains in CDRs H3 and L3 in all Fabs with germline **H3-53**, which indicates loose packing of the variable domains. All those molecules are relatively unstable, as is reflected in their low T_ms.

Table 5. Melting temperatures for the 16 Fabs.[1]

	L3-20	L4-1	L3-11	L1-39	HC average
H1-69	73.6	74.8	75.6	80.3	76.1
H3-23	74.8	75.2	4.8	81.5	76.6
H3-53	68.4	68.0	71.5	73.9	70.5
H5-51	68.4	68.4	71.9	77.0	71.4
LC average	71.3	71.6	73.5	78.2	

[1]Colors: blue ($T_m < 70°C$), green ($70°C < T_m < 73°C$), yellow ($73°C < T_m < 78°C$), orange ($T_m > 78°C$).

Discussion

This is the first report of a systematic structural investigation of a phage germline library. The 16 Fab structures offer a unique look at all pairings of 4 different HCs (**H1-69**, **H3-23**, **H3-53**, and **H5-51**) and 4 different LCs (**L1-39**, **L3-11**, **L3-20** and **L4-1**), all with the same CDR H3. The structural data set taken as a whole provides insight into how the backbone conformations of the CDRs of a specific heavy or light chain vary when it

is paired with 4 different light or heavy chains, respectively. A large variability in the CDR conformations for the sets of HCs and LCs is observed. In some cases the CDR conformations for all members of a set are virtually identical, for others subtle changes occur in a few members of a set, and in some cases larger deviations are observed within a set. The five variants that crystallized with 2 copies of the Fab in the asymmetric unit serve somewhat as controls for the influence of crystal packing on the conformations of the CDRs. In four of the 5 structures the CDR conformations are consistent. In only one case, that of **H1-69:L3-20** (the lowest resolution structure), do we see differences in the conformations of the 2 copies of CDRs H1 and L1. This variability is likely a result of 2 factors, crystal packing interactions and internal instability of the variable domain.

For the CDRs with canonical structures, the largest changes in conformation occur for CDR H1 of **H1-69** and **H3-53**. The other 2 HCs, **H3-23** and **H5-51**, have canonical structures that are remarkably well conserved (Fig. 1). Of the 4 HCs, **H1-69** has the greatest number of canonical structure assignments (Table 2). **H1-69** is unique in having a pair of glycine residues at positions 26 and 27, which provide more conformational freedom in CDR H1. Besides IGHV1-69, only the germlines of the VH4 family possess double glycines in CDR H1, and it will be interesting to see if they are also conformationally unstable.

Having all 16 VH:VL pairs with the same CDR H3 provides some insights into why molecular modeling efforts of CDR H3 have proven so difficult. As mentioned in the Results section, this data set is composed of 21 Fabs, since 5 of the 16 variants have 2 Fab copies in the asymmetric unit. For the 18 Fabs with complete backbone atoms for CDR H3, 10 have conformations similar to that of the parent, while the others have significantly different conformations (Fig. 6). Thus, it is likely that the CDR H3 conformation is dependent upon 2 dominating factors: 1) amino acid sequence; and 2) VH and VL context. More than half of the variants retain the conformation of the parent despite having differences in the VH:VL pairing. This subset includes 2 structures with 2 copies of the Fab in the asymmetric unit, all of which are nearly identical in conformation. This provides an internal control showing a consistency in the conformations. The remaining 8 structures exhibit "non-parental" conformations, indicating that the VH and VL context can also be a dominating factor influencing CDR H3. Importantly, there are 5 distinctive conformations in this subset. This subset also has 2 structures with 2 Fab copies in the asymmetric unit. Each pair has nearly identical conformations providing an internal check on the consistency of the conformations. Interestingly, as described earlier, these 2 pairs differ in the stem regions with the **H1-69:L3-20** pair in the 'extended' conformation and **H5-51:L4-1** pair in the 'kinked' conformation. The conformations are different from each other, as well as from the parent.

The CDR H3 conformational analysis shows that, for each set of variants of one HC paired with the 4 different LCs, both "parental" and "non-parental" conformations are observed. The same variability is observed for the sets of variants composed of one LC paired with each of the 4 HCs. Thus, no patterns of conformational preference for a particular HC or LC emerge to shed any direct light on what drives the conformational differences. This finding supports the hypothesis of Weitzner et al.[18] that the H3 conformation is controlled both by its sequence and its environment.

In looking at a possible correlation between the tilt angle and the conformation of CDR H3, no clear trends are observed. Two variants, **H1-69:L3-20** and **H3-23:L3-20**, have the largest differences in the tilt angles compared to other variants as seen in Table 3. The absolute VH:VL orientation parameters for the 2 Fabs (Table S2) show significant deviation in HL, LC1 and HC2 values (2-3 standard deviations from the mean). One of the variants, **H3-23:L3-20**, has the CDR H3 conformation similar to the parent, but the other, **H1-69:L3-20**, is different.

As noted in the Results section, the 2 variants, **H1-69:L3-20** and **H3-23:L3-20**, are outliers in terms of the tilt angle; at the same time, both have the smallest VH:VL interface. These smaller interfaces may perhaps translate to a significant deviation in how VH is oriented relative to VL than the other variants. These deviations from the other variants can also be seen to some extent in VH:VL orientation parameters in Table S2, as well as in the smaller number of residues involved in the VH:VL interfaces of these 2 variants (Fig. S5). These differences undoubtedly influence the conformation of the CDRs, in particular CDR H1 (Fig. 1A) and CDR L1 (Fig. 3C), especially with the tandem glycines and multiple serines present, respectively.

Pairing of different germlines yields antibodies with various degrees of stability. As indicated by the melting temperatures, germlines **H1-69** and **H3-23** for HC and germline **L1-39** for LC produce more stable Fabs compared to the other germlines in the experimental set. Structural determinants of the differential stability are not always easy to decipher. One possible explanation of the clear preference of LC germline **L1-39** is that CDR L3 has smaller residues at positions 91 and 94, allowing for more room to accommodate CDR H3. Other germlines have bulky residues, Tyr, Arg and Trp, at these positions, whereas **L1-39** has Ser and Thr. Various combinations of germline sequences for VL and VH impose certain constraints on CDR H3, which has to adapt to the environment. A more compact CDR L3 may be beneficial in this situation.

At the other end of the stability range is LC germline **L3-20**, which yields antibodies with the lowest T_ms. While pairings with **H3-53** and **H5-51** may be safely called a mismatch, those with **H1-69** and **H3-23** have T_ms about 5-6° higher. Curiously, the 2 Fabs, **H1-69:L3-20** and **H3-23:L3-20**, deviate markedly in their tilt angles from the rest of the panel. It is possible that by adopting extreme tilt angles the structure modulates CDR H3 and its environment, which apparently cannot be achieved solely by conformational rearrangement of the CDR. Note that most of the VH:VL interface residues are invariant; therefore, significant change of the tilt angle must come with a penalty in free energy. Yet, for the 2 antibodies, the total gain in stability merits the domain repacking.

Overall, the stability of the Fab, as measured by T_m, is a result of the mutual adjustment of the HC and LC variable

domains and adjustment of CDR H3 to the VH:VL cleft. The final conformation represents an energetic minimum; however, in most cases it is very shallow, so that a single mutation can cause a dramatic rearrangement of the structure.

In summary, the analysis of this structural library of germline variants composed of all pairs of 4 HCs and 4LCs, all with the same CDR H3, offers some unique insights into antibody structure and how pairing and sequence may influence, or not, the canonical structures of the L1, L2, L3, H1 and H2 CDRs. Comparison of the CDR H3s reveals a large set of variants with conformations similar to the parent, while a second set has significant conformational variability, indicating that both the sequence and the structural context define the CDR H3 conformation. Quite unexpectedly, 2 of the variants, **H1-69:L3-20** and **H3-53:L4-1**, have the 'extended' stem region differing from the other 14 that have a 'kinked' stem region. Why this is the case is unclear at present. These data reveal the difficulty of modeling CDR H3 accurately, as shown again in Antibody Modeling Assessment II.[21] Furthermore, antibody CDRs, H3 in particular, may go through conformational changes upon binding their targets,[34] making structural prediction for docking purposes an even more difficult task. Fortunately, for most applications of antibody modeling, such as engineering affinity and biophysical properties, an accurate CDR H3 structure is not always necessary. For those applications where accurate CDR structures are essential, such as docking, the results in this work demonstrate the importance of experimental structures. With the recent advances in expression and crystallization methods, Fab structures can be obtained rapidly.

The set of 16 germline Fab structures offers a unique dataset to facilitate software development for antibody modeling. The results essentially support the underlying idea of canonical structures, indicating that most CDRs with germline sequences tend to adopt predefined conformations. From this point of view, a novel approach to design combinatorial antibody libraries would be to cover the range of CDR conformations that may not necessarily coincide with the germline usage in the human repertoire. This would insure more structural diversity, leading to a more diverse panel of antibodies that would bind to a broad spectrum of targets.

Materials and methods

Fab production, purification and crystallization

The production, purification and crystallization of the Fabs reported in this article were described previously.[25] Briefly, the 16 Fabs were produced by combining 4 different HC and 4 different LC germline constructs. The human HC germlines were IGHV1-69 (**H1-69**), IGHV3-23 (**H3-23**), IGHV3-53 (**H3-53**) and IGHV5-51 (**H5-51**) in the IMGT nomenclature.[35] The human LC germlines were IGKV1-39 (**L1-39**), IGKV3-11 (**L3-11**), IGKV3-20 (**L3-20**) and IGKV4-1 (**L4-1**) corresponding to O12, L6, A27 and B3 in the V-BASE nomenclature.[36] CDR H3 of the anti-CCL2 antibody CNTO 888 with the amino acid sequence **ARYDGIYGELDF** [24] was used in all Fab constructs. The J region genes were IGHJ1 for the HC and IGKJ1 for the LC for all Fabs. Human IgG1 and κ constant regions were used in all Fab constructs. A 6xHis tag was added to the C-terminus of the HC to facilitate purification.

The Fabs were expressed in HEK 293E cells and purified by affinity and size-exclusion chromatography.[37] For crystallization, the Fabs were dialyzed into 20 mM Tris buffer, pH 7.4, with 50 mM NaCl and concentrated to 12-18 mg/mL. Automated crystallization screening was carried out using the vapor diffusion method at 20°C with an Oryx4 (Douglas Instruments) or a Mosquito (TTP Labtech) crystallization robot in a sitting drop format using Corning 3550 plates. Initial screening was carried out with an in-house 192-well screen optimized for Fab crystallization and the Hampton 96-well Crystal Screen HT (Hampton Research). For the majority of the Fabs, the crystallization protocol employed microseed matrix screening [38,39] using self-seeding or cross-seeding approaches.[25] A summary of the final crystallization conditions for each of the Fabs is presented in Table 1.

X-ray data collection

For 13 of the Fab crystals, X-ray data collection was carried out at Janssen Research and Development, LLC using a Rigaku MicroMax™-007HF microfocus X-ray generator equipped with a Saturn 944 CCD detector and an X-stream™ 2000 cryo-cooling system (Rigaku), and for the remaining 3, X-ray data collection was carried out at the Advanced Photon Source (APS) synchrotron at Argonne National Laboratory using the IMCA 17-ID beamline with a Pilatus 6M detector. For X-ray data collection, the Fab crystals were soaked for a few seconds in a cryo-protectant solution containing the corresponding mother liquor supplemented with 17-25% glycerol (Table S1). The crystals for which data were collected in-house were flash cooled in the stream of nitrogen at 100 K. Crystals sent to the APS were flash cooled in liquid nitrogen prior to shipping them to the synchrotron. Diffraction data for all variants were processed with the program XDS.[40] X-ray data statistics are given in Table 1.

Structure determination

A summary of the methods used in the structure solution and refinement of the 16 Fabs is presented in Table S1. Twelve of the structures were solved by molecular replacement with Phaser [41] using different combinations of search models for the VH, VL and constant domains. Four of the structures, **H3-53:L1-39**, **H3-53:L3-11**, **H5-51:L1-39** and **H5-51:L3-11**, were solved by direct replacement followed by rigid body refinement with REFMAC.[42] All structures were refined using REFMAC. Model adjustments were carried out using the program Coot.[43] The refinement statistics are given in Table 1. Other crystallographic calculations were performed with the CCP4 suite of programs.[44] The structural figures were prepared using the PyMOL Molecular Graphics System, Version 1.0 (Schrödinger, LLC).

Structural analysis

The canonical structure assignments (Table 2) were made using PyIgClassify, an online canonical structure classification tool (http://dunbrack2.fccc.edu/pyigclassify/) [13] that uses the rules set forth by Dunbrack and coworkers.[12]

The conformational variability within the CDRs was assessed by calculating the root-mean-square deviation (rmsd) from the average structure that was generated after superposition of all structures of the set using the main-chain atoms of the CDR in question. The rmsd was calculated for all main-chain atoms (N, CA, C, O) of the CDR.

The contact surface areas of the VH and VL domains at the VH:VL inteface were computed with the CCP4 program PISA.[32] The surface complementarity of the VH and VL domains was computed using the CCP4 program SC.[33]

VH:VL tilt angles

The orientation of the VH domain with respect to the VL domain was assessed using 2 different approaches. The first approach calculates the 6 VH:VL orientation parameters that describe the VH:VL relationship according to Dunbar and coworkers [31] using a script downloaded from the website (http://opig.stats.ox.ac.uk/webapps/abangle). The six parameters include 5 angles, HL, H1, H2, L1 and L2, and a distance, dc. These parameters are derived by first defining 2 planes, one for each domain, based on core residues in the domains. The distance between the planes, dc, is determined along a vector between the planes that is used to establish a consistent coordinate system. The torsion angle between the domains, HL, is much like the VH:VL packing angle defined by Abhinandan and Martin.[30] The tilt of one domain relative to the other is defined by the HC1 and LC1 angles, and the twist of one domain relative to the other is defined by the HC2 and LC2 angles.

The second approach calculates the difference in the tilt angle between pairs of Fvs, which reflects the relative orientation between the VH and VL domains.[21] The difference with respect to the reference structure is calculated by sequential root-mean-square superposition of the VL and VH domains using β-sheet core Cα positions (Chothia numbering scheme): 3–13, 18–25, 33–38, 43–49, 61–67, 70–76, 85–90, 97–103 for VL; 3–7, 18–24, 34–40, 44–51, 56–59, 67–72, 77–82a, 87–94, 102–110 for VH. The κ angle in the spherical polar angular system (ω, ϕ, κ) of the latter transformation is the difference in the tilt angle.

Differential scanning calorimetry

DSC experiments were performed on a VP-capillary DSC system (MicroCal Inc., Northampton, MA) in which temperature differences between the reference and sample cell are continuously measured and calibrated to power units. Samples were heated from 10°C to 95°C at a heating rate of 60°C/hour. The pre-scan time was 15 minutes and the filtering period was 10 seconds. The concentration used in the DSC experiments was about 0.4 mg/mL in phosphate-buffered saline. Analysis of the resulting thermograms was performed using MicroCal Origin 7 software. Melting temperature of proteins was determined by deconvolution of the DSC scans using non-2 state model in the MicroCal Origin 7 software. Scans were deconvoluted using a non-2 state model with either 1-step transition or 2-step transition depending on the number of resolved peaks observed in a scan.

Accession numbers

Atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers 4KMT, 5I15, 5I16, 5I17, 5I18, 5I19, 5I1A, 5I1C, 5I1D, 5I1E, 5I1G, 5I1H, 5I1I, 5I1J, 5I1K and 5I1L.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

References

1. Reichert JM. Antibodies to watch in 2016. MAbs 2016; 8:197-204; PMID:26651519; http://dx.doi.org/10.1080/19420862.2015.1125583
2. Wu TT, Kabat EA. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J Exp Med 1970; 132:211-250; PMID:5508247; http://dx.doi.org/10.1084/jem.132.2.211
3. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. J Mol Biol 1987; 196:901-917; PMID:3681981; http://dx.doi.org/10.1016/0022-2836(87)90412-8
4. Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, et al. Conformations of immunoglobulin hypervariable regions. Nature 1989; 342:877-883; PMID:2687698; http://dx.doi.org/10.1038/342877a0
5. Chothia C, Lesk AM, Gherardi E, Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G. Structural repertoire of the human VH segments. J Mol Biol 1992; 227:799-817; PMID:1404389; http://dx.doi.org/10.1016/0022-2836(92)90224-8
6. Tomlinson IM, Cox JP, Herardi GE, Lesk AM, Chothia C. The structural repertoire of the human V kappa domain. EMBO J 1995; 14:4628-4638; PMID:7556106
7. Vargas-Madrazo E, Lara-Ochoa F, Almagro JC. Canonical structure repertoire of the antigen-binding site of immunoglobulins suggests strong geometrical restrictions associated to the mechanism of immune recognition. J Mol Biol 1995; 254:497-504; PMID:7490765; http://dx.doi.org/10.1006/jmbi.1995.0633
8. Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. J Mol Biol 1997; 273:927-948; PMID:9367782; http://dx.doi.org/10.1006/jmbi.1997.1354
9. Collis AV, Brouwer AP, Martin AC. Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. J Mol Biol 2003; 325:337-354; PMID:12488099; http://dx.doi.org/10.1016/S0022-2836(02)01222-6
10. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000; 28:235-242; PMID:10592235; http://dx.doi.org/10.1093/nar/28.1.235
11. Martin AC, Thornton JM. Structural families in loops of homologous proteins: automatic classification, modelling and application to antibodies. J Mol Biol 1996; 263:800-815; PMID:8947577; http://dx.doi.org/10.1006/jmbi.1996.0617
12. North B, Lehmann A, Dunbrack RL. A new clustering of antibody CDR loop conformations. J Mol Biol 2011; 406:228-256; PMID:21035459; http://dx.doi.org/10.1016/j.jmb.2010.10.030

13. Adolf-Bryfogle J, Xu Q, North B, Lehmann A, Dunbrack RL. PyIgClassify: a database of antibody CDR structural classifications. Nucleic Acids Res 2015; 43:D432-D438; PMID:25392411; http://dx.doi.org/10.1093/nar/gku1106
14. Shirai H, Kidera A, Nakamura H. Structural classification of CDR-H3 in antibodies. FEBS Lett 1996; 399:1-8; PMID:8980108; http://dx.doi.org/10.1016/S0014-5793(96)01252-5
15. Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. Antibody structure, prediction and redesign. Biophys Chem 1997; 68:9-16; PMID:9468606; http://dx.doi.org/10.1016/S0301-4622(96)02266-1
16. Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. Conformations of the third hypervariable region in the VH domain of immunoglobulins. J Mol Biol 1998; 275:269-294; PMID:9466909; http://dx.doi.org/10.1006/jmbi.1997.1442
17. Kuroda D, Shirai H, Kobori M, Nakamura H. Structural classification of CDR-H3 revisited: a lesson in antibody modeling. Proteins 2008; 73:608-620; PMID:18473362; http://dx.doi.org/10.1002/prot.22087
18. Weitzner BD, Dunbrack RL, Gray JJ. The origin of CDR H3 structural diversity. Structure 2015; 23:302-311; PMID:25579815; http://dx.doi.org/10.1016/j.str.2014.11.010
19. Almagro JC, Beavers MP, Hernandez-Guzman F, Maier J, Shaulsky J, Butenhof K, Labute P, Thorsteinson N, Kelly K, Teplyakov A, et al. Antibody modeling assessment. Proteins 2011; 79:3050-3066; PMID:21935986; http://dx.doi.org/10.1002/prot.23130
20. Almagro JC, Teplyakov A, Luo J, Sweet R, Kodangattil S, Hernadez-Guzman F, Gilliland GL. Second Antibody Modeling Assessment (AMA-II). Proteins 2014; 82:1553-1562; PMID:24668560; http://dx.doi.org/10.1002/prot.24567
21. Teplyakov A, Luo J, Obmolova G, Malia TJ, Sweet R, Stanfield RL, Kodangattil S, Almagro JC, Gilliland GL. Antibody modeling assessment II. Structures and models. Proteins 2014; 82:1563-1582; PMID:24633955; http://dx.doi.org/10.1002/prot.24554
22. Shi L, Wheeler JC, Sweet RW, Lu J, Luo J, Tornetta M, Whitaker B, Reddy R, Brittingham R, Borozdina L, et al. De novo selection of high-affinity antibodies from synthetic fab libraries displayed on phage as pIX fusion proteins. J Mol Biol 2010; 397:385-396; PMID:20114051; http://dx.doi.org/10.1016/j.jmb.2010.01.034
23. de Wildt RM, Hoet RM, van Venrooij WJ, Tomlinson IM, Winter G. Analysis of heavy and light chain pairings indicates that receptor editing shapes the human antibody repertoire. J Mol Biol 1999; 285:895-901; PMID:9887257; http://dx.doi.org/10.1006/jmbi.1998.2396
24. Obmolova G, Teplyakov A, Malia TJ, Grygiel TL, Sweet R, Snyder LA, Gilliland GL. Structural basis for high selectivity of anti-CCL2 neutralizing antibody CNTO 888. Mol Immunol 2012; 51:227-233; PMID:22487721; http://dx.doi.org/10.1016/j.molimm.2012.03.022
25. Obmolova G, Malia TJ, Teplyakov A, Sweet R, Gilliland GL. Protein crystallization with microseed matrix screening: application to human germline antibody Fabs. Acta Crystallogr 2014; F70:1107-1115; PMID:25084393; http://dx.doi.org/10.1107/S2053230X14012552
26. Tramontano A, Chothia C, Lesk AM. Framework residue-71 is a major determinant of the position and conformation of the 2nd hypervariable region in the VH domains of immunoglobulins. J Mol Biol 1990; 215:175-182; PMID:2118959; http://dx.doi.org/10.1016/S0022-2836(05)80102-0
27. Mas MT, Smith KC, Yarmush DL, Aisaka K, Fine RM. Modeling the anti-CEA antibody combining site by homology and conformational search. Proteins 1992; 14:483-498; PMID:1438186; http://dx.doi.org/10.1002/prot.340140409
28. Stanfield RL, Takimoto-Kamimura M, Rini JM, Profy AT, Wilson IA. Major antigen-induced domain rearrangements in an antibody. Structure 1993; 1:83-93; PMID:8069628; http://dx.doi.org/10.1016/0969-2126(93)90024-B
29. Narayanan A, Sellers BD, Jacobson MP. Energy-based analysis and prediction of the orientation between light- and heavy-chain antibody variable domains. J Mol Biol 2009; 388:941-953; PMID:19324053; http://dx.doi.org/10.1016/j.jmb.2009.03.043
30. Abhinandan KR, Martin AC. Analysis and prediction of VH/VL packing in antibodies. Protein Eng Des Sel 2010; 23:689-697; PMID:20591902; http://dx.doi.org/10.1093/protein/gzq043
31. Dunbar J, Fuchs A, Shi J, Deane CM. ABangle: characterising the VH-VL orientation in antibodies. Protein Eng Des Sel 2013; 26:611-620; PMID:23708320; http://dx.doi.org/10.1093/protein/gzt020
32. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol 2007; 372:774-797; PMID:17681537; http://dx.doi.org/10.1016/j.jmb.2007.05.022
33. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. J Mol Biol 1993; 234:946-950; PMID:8263940; http://dx.doi.org/10.1006/jmbi.1993.1648
34. Rini JM, Schulze-Gahmen U, Wilson IA. Structural evidence for induced fit as a mechanism for antibody-antigen recognition. Science 1992; 255:959-965; PMID:1546293; http://dx.doi.org/10.1126/science.1546293
35. Lefranc MP. Nomenclature of the human immunoglobulin genes. Curr Protoc Immunol 2001; Appendix 1P:1-37; PMID:18432650; http://dx.doi.org/10.1002/0471142735.ima01ps40.
36. Tomlinson IM, Williams SC, Ignatovitch O, Corbett SJ, Winter G. 1998. V BASE Sequence Directory, MRC Centre for Protein Engineering Cambridge, UK.
37. Zhao Y, Gutshall L, Jiang H, Baker A, Beil E, Obmolova G, Carton J, Taudte S, Amegadzie B. Two routes for production and purification of Fab fragments in biopharmaceutical discovery research: papain digestion of mAb and transient expression in mammalian cells. Protein Expr Purif 2009; 67:182-189; PMID:19442740; http://dx.doi.org/10.1016/j.pep.2009.04.012
38. D'Arcy A, Villard F, Marsh M. An automated microseed matrix-screening method for protein crystallization. Acta Crystallogr 2007; D63:550-554; PMID:17372361; http://dx.doi.org/10.1107/S0907444907007652
39. Obmolova G, Malia TJ, Teplyakov A, Sweet R, Gilliland GL. Promoting crystallization of antibody-antigen complexes via microseed matrix screening. Acta Crystallogr 2010; D66:927-933; PMID:20693692; http://dx.doi.org/10.1107/S0907444910026041
40. Kabsch W. XDS. Acta Crystallogr 2010; D66:125-132; PMID:20124692; http://dx.doi.org/10.1107/S0907444909047337
41. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr 2007; 40:658-674; PMID:19461840; http://dx.doi.org/10.1107/S0021889807021206
42. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr 2011; D67:355-367; PMID:21460454; http://dx.doi.org/10.1107/S0907444911001314
43. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr 2010; D66:486-501; PMID:20383002; http://dx.doi.org/10.1107/S0907444910007493
44. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr 2011; D67:235-242; PMID:21460441; http://dx.doi.org/10.1107/S0907444910045749