JOURNAL OF MOLECULAR RECOGNITION
J. Mol. Recognit. 2004; **17**: 332–338
DOI:10.1002/jmr.679

Length of the antibody heavy chain complementarity determining region 3 as a specificity-determining factor

Yvelise Barrios,[†] Pernilla Jirholt[‡] and Mats Ohlin*
Department of Immunotechnology, Lund University, Lund, Sweden

The antigen binding site of an antibody is made up of residues residing in six hypervariable loops of the heavy and light chains. In most cases several or all of these loops are required for the establishment of the antigen-binding surface. Five of these loops display a limited diversity in length and sequence while the third complementarity determining region (CDR) of the heavy chain is highly different between antibodies not only with respect to sequence but also with respect to length. Its extensive diversity is a key component in the establishment of binding sites allowing for the recognition of essentially any antigen by humoral immunity. The relative importance of its sequence vs its length diversity in this context is however, not very well established. To investigate this matter further we have used an approach employing combinatorial antibody libraries and antigen-specific selection in the search for CDRH3 length and sequence diversity compatible with a given antigen specificity, the major antigenic determinant on the tumour-associated antigen mucin-1. In this way we have now defined heavy chain CDR3 length as a critical parameter in the creation of an antigen-specific binding site. We also propose that this may reflect a dependence of a particular structure of this hypervariable loop, the major carrier of diversity in the binding site, for establishment of a given specificity. Copyright © 2004 John Wiley & Sons, Ltd.

Keywords: **antibody; antigen; complementarity determining region; diversity; loop; specificity**

Received 14 November 2003; revised 8 January 2004; accepted 11 March 2004

INTRODUCTION

Antibodies are proteins of immense importance not only for the protection of higher organisms against a hostile environment but also as tools in bioscience, biomedicine and biotechnology. Consequently, large efforts have been put into the development of technologies to raise specific antibodies against virtually any target. Studies aiming at understanding how an antigen-binding site, a paratope, of an antibody is constructed and how it evolves during the affinity maturation that perfects its binding characteristics are also well represented in the literature. Examples of such studies are investigations of the makeup of antibody repertoires recognizing multiple epitopes on individual proteins (Edwards *et al.*, 2003) as well as individual epitopes on model antigens or on biomedically more significant targets (as reviewed by Ohlin and Borrebeck, 1996; Ohlin and Zouali, 2003). The development of phage display (Smith, 1985; McCafferty *et al.*, 1990) and other combinatorial library technologies have greatly aided in these efforts and they provide a basis in molecular genetics for the study even of human antibody repertoires that only with substantial difficulties could be addressed by conventional hybridoma technology. Genome sequencing and the determination of the composition of the human immunoglobulin loci is an important aid for the understanding of how specific paratopes are constructed. The accumulation of antibody structures and even more importantly antibody–antigen structures, highlights critical aspects of the interaction of antibodies and their targets. Nevertheless, much remains to be understood regarding these matters. What determines the composition of a developing specific antibody repertoire? To what extent does the available repertoire govern the ability to raise specific immunity *in vivo* or select specific antibodies *in vitro* and how does the repertoire affect the nature of developing responses? How is the interaction with antigen determined by interplay between the centrally located third complementarity determining region of the heavy chain (CDRH3) and other hypervariable loops?

To investigate the last matter further we made use of a repertoire of antibody single-chain fragments (scFv) previously selected from a combinatorial library (Söderlind *et al.,* 2000) that is specific for the variable number of tandem repeats (VNTR)-epitope found in the tumour-associated

*Correspondence to: M. Ohlin, Department of Immunotechnology, Lund University, PO Box 7031, S-220 07 Lund, Sweden.
E-mail: mats.ohlin@immun.lth.se
[†]Present address: Immunology Section, Hospital Universitario de Canarias, Tenerife, Spain.
[‡]Present address: The Wallenberg Laboratory for Cardiovascular Research, Göteborg University, SE-413 45 Göteborg, Sweden.
Contract/grant sponsors: Wenner-Gren Foundations; Magnus Bergvall Foundation; John and Augusta Persson Foundation; Swedish Research Council; BioInvent International AB.

Abbreviations used: CDR, complementarity determining region; FR, framework region; H, heavy, L, light; PBS, phosphate buffered saline; scFv, single chain antibody fragment; VNTR, variable number of tandem repeats; VH, heavy chain variable; VL, light chain variable; wt, wild-type.

antigen mucin-1 (CD227). This clinically important epitope (Epenetos *et al.*, 2000) that is immunodominant in mice is found in a 20 amino acid-long repeat within the glycoprotein (Price *et al.*, 1998). Selection of human antibodies against this target has been difficult to achieve but lately strategies to select and evolve high-affinity human antibody fragments have been established (Jirholt *et al.*, 2001, 2002). The library (Söderlind *et al.*, 2000) from which such scFv were selected was originally created by the CDR shuffling methodology (Jirholt *et al.*, 1998) and as such carries CDRH1 and CDRH2 loops derived from different members of the VH3 gene family (in unmutated and somatically mutated form). Light chain hypervariable loops are mainly derived from members of the Vλ1 gene family. Gene fragments encoding these loops and a very diverse set of CDRH3, also derived from B-cell encoded rearranged immunoglobulin genes, have been assembled into complete genes carrying the well expressed and stable frameworks derived from the IGHV3-23 and IGLV1-47 germline genes. Thus the selected VNTR-specific scFv (Söderlind *et al.*, 2000) had been selected from a library carrying very extensive diversity (Borrebaeck and Ohlin, 2002). Repertoire and evolution studies defined among other things residues in CDRH1 and CDRH2 that were critical for the formation of the VNTR-specific paratope (Jirholt *et al.*, 2001) and mutations related to a high affinity character (Jirholt *et al.*, 2002). We now extend these studies to demonstrate the importance of CDRH3 length as a factor in the makeup of this specific paratope and discuss these findings in a broader context.

MATERIALS AND METHODS

Reagents

The pFAB5C.His phagemid vector (Ørum *et al.*, 1993) was kindly provided by Professor Jan Engberg (The Royal Danish School of Pharmacy, Copenhagen, Denmark). Biotinylated peptides were provided by Professor Elias Krambovitis. Non-biotinylated peptides were obtained from ThermoHybaid (Ulm, Germany). ScFv were purified as previously described by Ni-chelate chromatography making use of the N-terminal hexa-histidine tag, and subsequent gel filtration (Jirholt *et al.*, 2002).

Library constructions

A first library construction was similar to a previously described library successfully used for the evolution of mucin-1 VNTR-specific scFv (Jirholt *et al.,* 2002) with the exception that B-cell derived sequences were used to encode CDRH3. Briefly, sequences encoding CDRH1 and adjacent FR were amplified by PCR from phagemid vectors derived from the n-CoDeR library that had been selected on the mucin-1 peptide antigen. This gene population was enriched for two similar types of prototype clones specific for the epitope (Jirholt *et al.*, 2001). Thus, a substantial fraction of the sequences in this population was known to encode $W34_H$ (sequence numbering is according to the IMGT nomenclature; LeFranc, 2003) in CDRH1, a residue known to be involved in VNTR recognition by these scFv (Jirholt *et al.*, 2001). Sequences encoding CDRH2 and adjacent FR were amplified from the unselected n-CoDeR library population (Söderlind *et al.*, 2000), i.e. they represented a diverse set of unmutated and mutated sequences derived from genes belonging to the VH3 gene family. This amplified gene fragment population was, however, enriched for sequences carrying $L55_H$, a residue shown to be associated with VNTR recognition together with the prototype ARVLAQQRMDV-CDRH3 (Jirholt *et al.*, 2001), by using an upstream primer (AAGGGGCTGGAGTGGGTBKCWCT) that specifically incorporated the first two bases of residue 55_H encoding leucine. The use of the proofreading capable *Pfu* polymerase (Stratagene, La Jolla, CA, USA) in this PCR amplification, however, also ensured that a part of the CDRH2-population would include other residues at this site following 3′-editing of primers annealing to a template not encoding leucine in position 55_H. CDRH3 was amplified from cDNA obtained from mRNA derived from tonsil leukocytes that has been activated *in vitro* for 2 days with pokeweed mitogen. The up-stream primers AGAGCCGAGGACACTGCCGTGTATTACTGTGCNAGRGTNTTR and AGAGCCGAGGACACTGCCGTGTATTACTGTGCNAGRGT and downstream primer CGCTGCTCACGGTGACCAGTGTACCTTGGCCCCANACRTCCATYCT were used for this purpose. This procedure ensured amplification of CDRH3 sequences starting with ala-arg-val-leu or ala-arg-val and ending with arg-met-asp-val sequences found in many human VNTR-specific scFv (Jirholt *et al.*, 2001). Sequences encoding CDRL1 and CDRL2 and adjacent FR were amplified directly from the n-CoDeR library as this population of genes was known to contain at a high frequency ($\sim$60%) sequences of an origin known to be compatible (Jirholt *et al.*, 2001) with VNTR recognition (data not shown). Finally, the sequence encoding CDRL3, which also carries residues, in particular $W116_L$, known to be important for the specificity, was amplified from the vector encoding the VNTR-specific scFv sMuc159 that carries a prototype CDRH3 (Jirholt *et al.*, 2001). PCR products obtained in this way were assembled into an intact scFv-encoding gene library by overlap-extension PCR, as described previously (Jirholt *et al.*, 1998). The amplified genes were purified, cleaved with *Sfi*I and *Not*I and ligated into the pFAB5C.His phagemid vector. After electroporation of ligated DNA into *Escherichia coli* Top10F′ a library of 5×10^7 members was obtained. Phage stocks were prepared using R408 helper phage, and phages were concentrated by using a conventional PEG-precipitation protocol.

A second library harbouring diversity in CDRH3 alone was constructed by amplification of two fragments of the VNTR-specific scFv sMuc159 (Jirholt *et al.*, 2001). One of the primers (ACCTTGGCCCCAAACGTCCATCYYVNNVNNASYAAVCACTCTCGCACAGT) used for this purpose incorporated diversity into the proposed apex of CDRH3 (residues 108, 109, 110, 113 and 114). Following assembly of these fragments and restriction enzyme cleavage, the scFv-encoding genes were ligated into the phagemid vector and transformed into *E. coli*. The library had a size of 6×10^5 members, 30-fold more than the theoretical size of the library. Phage stocks were prepared as described above with the exception that VSCM13 was used as the helper phage.

Antigen-specific selection

Selection of antigen-specific clones from the first library was carried out using biotinylated 40-mer peptide carrying two repeated units of the VNTR sequence (Jirholt *et al.*, 2002) at 10 nM in TPBSB (0.05% Tween 20, 3% bovine serum albumin in phosphate buffered saline, PBS). Streptavidin-coated paramagnetic beads (Dynabeads M-280, Dynal A/S, Oslo, Norway) were blocked with TPBSB and subsequently used to catch peptide-bound phages. The beads were washed seven times with TPBSB and four times with PBS. Phages were eluted with trypsin (0.5 mg/ml in PBS) for 30 min as trypsin cleaves the scFv-protein 3 fusion protein at a site between the phage and the scFv molecule. Aprotinin (final concentration 0.05 mg/ml) was added and eluted phages were allowed to infect TOP10F′ bacteria cultures ($OD_{600} = 0.5$). The selection was repeated once with 1 nM antigen. Specificity was confirmed by using antigen-specific ELISA [scFv-specific phages bound to the biotinylated peptide caught onto immobilized streptavidin were detected using peroxidase-labelled anti-M13 antibody (Amersham Bioscience, Uppsala, Sweden)]. Antigen-specific selection of the second library was carried out essentially as described above using a biotinylated 12-mer peptide (biotin-GVTSAPDTRPAP) derived from the VNTR at 10 nM. Two additional rounds of selection were performed on the output of the first selection of the second library. During these rounds of selection the stringency of selection was increased in the last one by reducing the concentration of the biotinylated 12-mer peptide antigen to 1 nM. Furthermore, a competition step was introduced in the second and third rounds by adding an excess (10 μM) of non-biotinylated antigen (SAPDTRPAP) to the phages, previously incubated with the biotinylated antigen for 5 min prior to addition of streptavidin-coated Dynabeads.

Analysis of epitope specificity

Epitope specificity was analysed by Biacore analysis. Briefly, streptavidin was coupled by conventional amine-coupling to a Biacore CM5 sensorchip according to the manufacturer's description (Biacore AB, Uppsala, Sweden). Biotin-binding sites were loaded with biotinylated peptide mimicking the VNTR epitope (biotin-GVTSAPDTRPAP). Purified scFv was diluted in running buffer [10 mM HEPES, 3.4 mM EDTA, 150 mM NaCl, 0.005% (v/v) P20 detergent (Biacore AB)] and mixed with peptides at a concentration of 1–10 μg/ml. Samples were run on the peptide-loaded sensorchip and binding with and without added peptide was compared.

Gene sequencing and analysis

Sequencing was carried out using the Big Dye Terminator Kit (PE Biosystems, Warrington, England), on an ABI 377 automated sequencing equipment (Applied Biosystems, Foster City, CA, USA). The origin of CDRH2 was determined by comparison using the FASTA algorithm with germline genes defined in the VBASE sequence directory (Cook and Tomlinson, 1995). Sequences are numbered and CDR are defined in accordance with the IMGT nomenclature (LeFranc, 2003).

RESULTS AND DISCUSSION

A range of VNTR-specific human scFv have previously been selected from the n-CoDeR library (Söderlind *et al.*, 2000) or evolved from such selected primary clones (Jirholt *et al.*, 2001, 2002). We thus have a good view of selectable repertoires recognising this target. In fact, many of these clones showed characteristic features. These included specific sequence usage in CDRH1 (in particular $W34_H$), CDRH2 (residue $L55_H$ or $V55_H$ depending on CDRH3 sequence and/or residue 116_L) and CDRL3 (in particular $W107_L$ and $L116_L$ or $W116_L$) as well as the use of two related prototype CDRH3 sequences with identical length and a somewhat similar sequence [Ala-Arg-hydrophobic-hydrophobic-small-X-X-X-hydrophobic-Asp-hydrophobic (where X denotes unrelated residues); Jirholt *et al.*, 2001]. Specific scFv selected directly from the naive library recognized the target peptide with an affinity of $1–21 \times 10^6$ M^{-1} (Jirholt *et al.*, 2001), but variants could be affinity-matured more than 20-fold in a process accompanied by specific mutations in residue 32_H and 62_H within CDRH1 and CDRH2, respectively (Jirholt *et al.*, 2002). These results suggested the involvement of several of the hypervariable loops, including CDRH3, in the establishment of the VNTR-specific paratope. At least the first two and the last two residues of the CDRH3 motif are encoded by the VH and JH genes, respectively, and they are commonly found in antibodies carrying a so-called bulged-type CDRH3 stem (Morea *et al.*, 1998). This suggests that the base of CDRH3 in both cases takes on a similar structure while the apex of the loop may differ between these mucin-1-specific scFv. Importantly, the two clone types display a very similar recognition of antigen, as demonstrated by their pattern of recognition of mutated peptide sequences (Fig. 1). Thus, the observed similarities in sequence correlated with similarities in fine specificity. To further explore the observed restricted CDRH3 diversity of these clones and to define the interdependence of the various parts of the binding site, we created a new library using the CDR shuffling technology that carried diversity in several of the hypervariable loops. This was done through primer design and selection of template in a way that also maintained, at a high frequency, some sequence features associated with this VNTR-specificity, like those found in CDRH1 ($W34_H$) and in CDRH2 ($L55_H$), in the light chain, and also maintained the bulged-type CDRH3 stem. Sequencing of randomly picked clones from this library demonstrated that the library was diverse and it also showed that the incorporated CDRH3 tended to be highly diverse in length [range five to 25 residues, average 17.5 ($n = 8$)] and compared well with average human CDRH3 (average length 15 residues; Wu *et al.*, 1993; Yamada *et al.*, 1991; Zemlin *et al.*, 2003). Selection of this library on the 40-mer mucin-1 peptide (representing two VNTR) resulted in the retrieval of specific clones already after two rounds of selection (7/22 tested clones were peptide-specific at this stage). Sequence characteristics among the selected VNTR-specific clones in the, within the unselected library, highly diversified CDRH2 were very

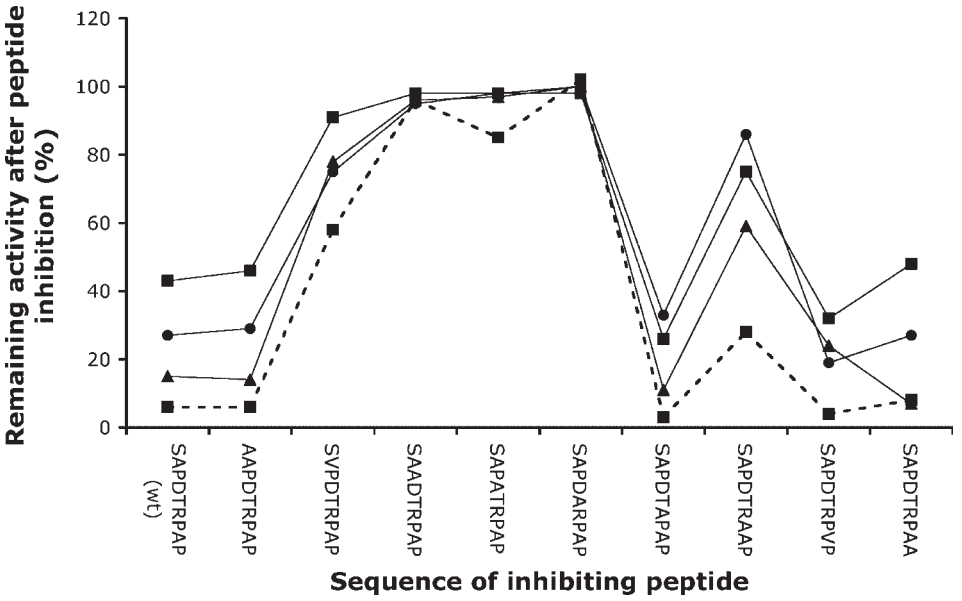

Figure 1. Remaining binding of scFv to the mucin-1 VNTR immobilized via a biotin–streptavidin interaction onto a Biacore chip after incubation with soluble peptides. Peptides inhibiting the binding were the wild-type (wt) VNTR sequence or alanine-scanned variants thereof. In cases where the wt sequence carried an alanine residue, this residue was modified by incorporation of valine. ScFv included Muc201 (squares) and the affinity matured variant thereof, sMuc159 (triangles), both of which use the ARVLAQQRMDV CDRH3, as well as Muc46 (circles), which carries the ARMVGAHALDI CDRH3 (Jirholt *et al.*, 2001). Inhibiting peptides were used at 1 μg/ml (solid lines). The lower affinity (Jirhol *et al.*, 2001) scFv Muc201 was also tested for inhibition at a peptide concentration of 10 μg/ml (dashed line).

similar with respect to the germline gene origin of CDRH2 (IGHV3-30/3-30-3 and IGHV3-23; Table 1) to that found in clones selected in previous studies (Jirholt *et al.*, 2001). Furthermore, despite the diverse composition of CDRH3 in the unselected library, the selected, specific clones all had a CDRH3 with a length of 11 residues (Table 1), mimicking the length but not necessarily the sequence of CDRH3 of previously selected clones (Jirholt *et al.*, 2001). Thus, despite the fact that longer CDRH3 dominated the library, specific paratopes all retained the length-characteristics of the originally selected clones. It is highly unlikely that the observed length restriction is a result of a specific composition of CDRH3 with a length of 11 residues as sequence composition and variability of CDRH3 of other lengths does not preclude or even disfavour the presence of residues found in prototype CDRH3 (Zemlin *et al.*, 2003, including supporting figures available at the publisher's web-site: www.sciencedirect.com). We could thus confirm the features that had previously been shown to characterize a VNTR-specific paratope (Jirholt *et al.*, 2001) and extended this finding to also include the length of CDRH3 as a specificity-determining factor.

To further explore the importance of tolerated diversity in CDRH3 we created a second library carrying some targeted diversity within the proposed tip of the CDRH3 while maintaining the rest of the VNTR-specific sequence intact. Although the diversity of the unselected library was not as extensive as expected from the primer design, it could be confirmed that antigen-specific selection did not strongly bias the residues found in the different positions of the apex of the 11-residue long CDRH3 (Table 2), except in one case. In position 108, the larger aromatic residue phenylalanine seemed to be disfavoured over the purely hydrophobic residues leucine and valine. This particular residue is not located close to the centre of the binding site of an affinity-matured variant of a prototype VNTR-specific clone as evaluated by structure modelling (1KU4.pdb). Thus, the precise basis for this sequence restriction, i.e. if it is related to direct interaction with antigen or an indirect effect on the structure of the paratope, is not known. Despite this slight sequence restriction, it appeared that the VNTR-specificity is to a larger extent dependent on the length of the CDRH3 than on the precise sequence characteristics of the apex of the hypervariable loop.

This study raises several questions. For instance, how general are these observations, and what is the role of CDRH3 in the make-up of a specific paratope? It is generally believed that CDRH3 takes on a key role in the formation of the binding site and in some cases isolated CDRH3 derived from a specific antibody may even be sufficient to form an antigen-binding structure (Bourgeois *et al.*, 1998; Levi *et al.*, 1993). Furthermore, a recent study demonstrated that one-third of scFv selected for specificity for BlyS from a large library employed a given D segment sequence, suggesting that it carried features making CDRH3 created from it particularly able to bind to this antigen (Edwards *et al.*, 2003). Indeed, diversity located within CDRH3 is sufficient to create specific binders *in vivo* against essentially any large protein (Xu and Davis, 2000). However, such restricted repertoires fail to respond to some antigens like polysaccharides (Xu and Davis, 2000), suggesting that CDRH3 diversity as such is not enough to realize the full potential of antibody diversity (Ohlin and Zouali, 2003). In fact, the involvement, or not, of other parts of the variable domains in the recognition of an epitope is often a reflection of their ability to form critical interactions

Table 1. Sequence characteristics of clones specific for mucin-1 VNTR in comparison to prototype scFv (Jirholt *et al.*, 2001). The germline origin of CDRH2 among the unselected may differ as a consequence of the CDR-shuffling approach (Jirholt *et al.*, 1998) used to establish the library. However, selected CDRH2 sequences had one of two possible origins as determined at the nucleotide level (in particular the highly related germline genes IGHV3-30 and IGHV3-30-3 were selected) in accordance with previously defined scFv (Jirholt *et al.*, 2001, 2002). The length, but not the sequence of CDRH3 was maintained despite extensive length diversity in the unselected library. Note that residues 105–107 and 114–117 are essentially kept constant by the primer design used to prepare the CDRH3-encoding gene sequence

		Residue number (IMGT)																				
	CDRH2 origin (IGHV)	FRH2	CDRH2								CDRH3										
		55	56	57	58	59	60	61	62	63	105	106	107	108	109	110	113	114	115	116	117
Prototype sequence sMuc159	3–30	L	I	S	F	D	G	S	N	K	A	R	V	L	A	Q	Q	R	M	D	V
Prototype sequence Muc201	3–23	•	•	•	G	S	•	G	S	I	•	•	•	•	•	•	•	•	•	•	•
Clones selected from phage-displayed library																					
sMuc-C2-A3	3–23	•	•	•	G	S	•	G	S	I	•	•	•	A	G	A	N	•	•	•	•
sMuc-C2-B3	3–30–3	V	•	•	Y	•	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•
sMuc-C2-B4	3–30–3	•	•	•	Y	•	•	•	R	•	•	•	•	V	S	F	Y	•	•	•	•
sMuc-C2-C3	3–30	V	•	•	Y	•	E	D	K	•	•	•	•	•	G	G	L	•	•	•	•
sMuc-C2-D2	3–30–3	V	•	•	Y	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	R
sMuc-C2-D4	3–23	•	•	R	G	S	•	D	T	T	•	•	•	•	•	•	•	•	•	•	•
sMuc-C2-E1	3–30–3	•	•	•	Y	•	•	•	S	•	•	•	•	•	G	A	T	•	•	•	•

Table 2. Frequencies of different amino acids at the tip of the CDRH3 loop (residues 108–110, 113, 114) of two prototype VNTR-specific scFv sequences (Jirholt *et al.*, 2001) and of sequences found in the unselected ($n = 16$) and antigen-selected ($n = 17$) repertoires

IMGT position	108			109			110			113			114		
Prototype sequences:															
Muc201	L			A			Q			Q			R		
Muc46	V			G			A			H			A		
Library diversity:	FVL			SAGT			Any			Any			RKEG		
	Amino acid	Unselected frequency (%)	Selected frequency (%)	Amino acid	Unselected frequency (%)	Selected frequency (%)	Amino acid	Unselected frequency (%)	Selected frequency (%)	Amino acid	Unselected frequency (%)	Selected frequency (%)	Amino acid	Unselected frequency (%)	Selected frequency (%)
	F	50	0	S	81	59	N	63	65	N	88	65	K	94	82
	V	12	18	A	6	18	R	13	0	T	6	6	E	6	0
	L	38	82	G	13	18	G	6	0	S	6	6	R	0	18
				T	0	6	I	6	0	Q	0	12			
							D	6	0	L	0	6			
							F	6	6	M	0	6			
							H	0	18						
							S	0	6						
							Y	0	6						

with the antigen. For instance, in the case of recognition of 17β-estradiol, a set of independently raised antibodies create a critical hydrogen bonding centre through the use of specific light chain residues, providing the basis for a particular light chain sequence restriction among the estradiol-specific antibodies (Lamminmäki, 2003), while still allowing for extensive diversity in the heavy chain variable domain including CDRH3. In responses towards cytomegalovirus glycoprotein B epitope AD-1, it is the VH gene that seems to be the specificity-defining sequence while both CDRH3 and the light chain may be more diverse (Ohlin *et al.*, 1994). Similarly, the response to two rotavirus antigens is confined to a limited number of V genes but not D genes (Weitkamp *et al.*, 2003). In contrast, the recognition of *Cryptococcus neoformans* glucuronoxylomannan (Nakouzi and Casadevall, 2003) and *Neisseria meningitidis* PorA (Ihle *et al.*, 2003) involves both V gene and CDRH3 length restrictions. Finally, the repertoire selected against a single protein, BlyS, was not only enriched for certain CDRH3 sequences based on D-gene involvement, as outlined above, but also for scFv encoded by certain V genes (Edwards *et al.*, 2003). Consequently the recognition of this target, despite the fact that it most likely displayed multiple epitopes, strongly favoured the recruitment of particular types of sequence features in specific binders. In conclusion, the preferred recognition of different antigens is very different with respect to the type of sequence restrictions that are observed. In the case of our mucin-1 specificity, we hypothesize that the restriction in CDRH3 length is related to a proper positioning of key determinants either in the CDRH3 itself or in key contacts in the part of the VH domain encoded by the VH gene and the light chain (Jirholt *et al.,* 2001). It is less likely that it is the assembled VH gene that requires a specific CDRH3 length for functionality (Martin *et al.*, 2003) since clones with other specificities but carrying CDRH1 and CDRH2 with the same origin as those described here, carry CDRH3 of other lengths (data not shown). Thus, CDRH3 length is associated with specificity and not with the stability of the scFv as such. The fact that quite extensive diversity is allowed at least at the apex of the CDRH3 loop may suggest that these residues as such are not directly involved in antigen contact but rather that their spacing allows for positioning of other residues, e.g. at the base of the CDRH3 in positions appropriate for antigen binding. Alternatively the main point of interaction between antigen and CDRH3 may be through atoms of the hypervariable loop backbone rather than through atoms of the side chains. Nevertheless, it confirms the notion that length restriction in CDRH3 is a factor contributing to a specific paratope. The great diversity of CDRH3 length and sequence has up to now prevented the development of general rules for classification of this loop into precise groups, so called canonical structures, that have already been defined for other hypervariable loops in antibodies (Al-Lazikani *et al.*, 1997). Some things are known though, mainly in relation to the structure of the base of the CDRH3 loop. Indeed, our CDRH3 cloning strategy assured that CDRH3 loops in the library most likely take on the bulged type structure (Morea *et al.*, 1998) as determined by the presence of $R106_H$ and $D116_H$. As this field continues to evolve it will be interesting to learn whether or not CDRH3 loop length restrictions in many specific antibody repertoires are related to a requisite of the apex of this loop to take on a given, but as yet undefined canonical structure for the establishment of a given specificity. If this is so, libraries designed to evolve a given antibody often ought to be made so as to maintain a given canonical loop structure, in order for the library to contain functional members at a high frequency. In all, we conclude that CDRH3 length restrictions may be encountered in specific immune responses and that CDRH3 length diversity and restriction is a factor to consider when designing optimal libraries for molecular evolution of antigen-specific paratopes.

Acknowledgements

We are grateful to Virginie Coste for technical assistance. This study was supported by grants from the Wenner-Gren Foundations, the Magnus Bergvall Foundation, the John and Augusta Persson Foundation, The Swedish Research Council, and BioInvent International AB.

REFERENCES

Al-Lazikani B, Lesk AM, Chothia C. 1997. Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **273**: 927–948.

Borrebaeck CAK, Ohlin M. 2002. Antibody evolution beyond nature. *Nat. Biotechnol.* **20**: 1189–1190.

Bourgeois C, Bour JB, Aho LS, Pothier P. 1998. Prophylactic administration of a complementarity-determining region derived from a neutralizing monoclonal antibody is effective against respiratory syncytial virus infection in BALB/c mice. *J. Virol.* **72**: 807–810.

Cook GP, Tomlinson IM. 1995. The human immunoglobulin VH repertoire. *Immunol. Today* **16**: 237–242.

Edwards BM, Barash SC, Main SH, Choi GH, Minter R, Ullrich S, Williams E, Du Fou L, Wilton J, Albert VR, Ruben SM, Tristan J, Vaughan TJ. 2003. The remarkable flexibility of the human antibody repertoire; isolation of over one thousand different antibodies to a single protein, BlyS. *J. Mol. Biol.* **223**: 103–118.

Epenetos AA, Hird V, Lambert H, Mason P, Coulter C. 2000. Long term survival of patients with advanced ovarian cancer treated with intraperitoneal radioimmunotherapy. *Int. J. Gynecol. Cancer* **10**: 44–46.

Ihle Ø, Beckstrøm KJ, Michaelsen TE. 2003. Cloning, sequencing and expression of immunoglobulin variable regions of murine monoclonal antibodies specific for the P1.7 and P1.16 PorA protein loops of *Neisseria meningitidis. Scand. J. Immunol.* **57**: 453–462.

Jirholt P, Ohlin M, Borrebaeck CAK, Söderlind E. 1998. Exploiting sequence space: shuffling in vivo rearranged CDR into a master framework. *Gene* **215**: 471–476.

Jirholt P, Strandberg L, Jansson B, Krambovitis E, Söderlind E, Borrebaeck CAK, Carlsson R, Danielsson L, Ohlin M. 2001. A central core structure in an antibody variable domain determines antigen specificity. *Prot. Eng.* **14**: 67–73.

Jirholt P, Borrebaeck CAK, Ohlin M. 2002. Antibody evolution *in vitro* from the centre to the periphery–applied to a human

antibody fragment recognising the tumour-associated antigen mucin-1. *J. Mol. Biol.* **318**: 407–416.
Lamminmäki U. 2003. Towards improved hapten binders: structural characterization and molecular engineering of an anti-17b-estradiol antibody Fab fragment. *Ann. Unive. Turk., Ser. AI*, thesis Vol. **305**.
LeFranc M-P. 2003. IMGT, the international ImMunoGeneTics database®. *Nucl. Acids Res.* **31**: 307–310.
Levi M, Sällberg M, Rudén U, Herlyn D, Maruyama H, Wigzell H, Marks J, Wahren B. 1993. A complementarity-determining region synthetic peptide acts as a miniantibody and neutralizes human immunodeficiency virus type 1 *in vitro*. *Proc. Natl Acad. Sci. USA* **90**: 4374–4378.
Martin DA, Bradl H, Collins TJ, Roth E, Jack H-M, Wu GE. 2003. Selection of Ig μ heavy chains by complementarity-determining region 3 length and amino acid composition. *J. Immunol.* **171**: 4663–4671.
McCafferty J, Griffiths AD, Winter G, Chiswell DJ. 1990. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**: 552–554.
Morea V, Tramontano A, Rustici M, Chothia C, Lesk AM. 1998. Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J. Mol. Biol.* **275**: 269–294.
Nakouzi A, Casadevall A. 2003. The function of conserved amino acids in or near the complementarity determining regions for related antibodies to *Cryptococcus neoformans* glucuronoxylomannan. *Mol. Immunol.* **40**: 351–361.
Ohlin M, Borrebaeck CAK. 1996. Characteristics of human antibody repertoires following active immune responses *in vivo*. *Mol. Immunol.* **33**: 583–592.
Ohlin M, Zouali M. 2003. The human antibody repertoire to infectious agents: implications for disease pathogenesis. *Mol. Immunol.* **40**: 1–11.
Ohlin M, Owman H, Rioux JD, Newkirk MM, Borrebaeck CAK. 1994. Restricted variable region gene usage and possible rheumatoid factor relationship among human monoclonal antibodies specific for the AD-1 epitope on cytomegalovirus glycoprotein B. *Mol. Immunol.* **31**: 983–991.
Ørum H, Andersen PS, Øster A, Johansen LK, Riise E, Bjørnvad M, Svendsen I, Engberg J. 1993. Efficient method for constructing comprehensive murine Fab antibody libraries displayed on phage. *Nucl. Acids Res.* **21**: 4491–4498.
Price MR, Rye PD, Petrakou E, Murray A, Brady K, Imai S, Haga S, Kiyozuka Y, Schol D, Meulenbroek MFA, Snijdewint FGM, von Mensdorff-Pouilly S, Verstraeten RA, Kenemans P, Blockzjil A, Nilsson K, Nilsson O, Reddish M, Suresh MR, Koganty RR, Fortier S, Baronic L, Berg A, Longenecker MB, Hilkens J, Boer M, Karanikas V, McKenzie IFC, Galanina OE, Simeoni LA, Ter-Grigoryan AG, Belyanchikov IM, Bovin NV, Cao Y, Karsten U, Dai J, Allard WJ, Davis G, Yeung KK, Hanish F-G, Lloyd KO, Kudryashov V, Sikut R, Sikut A, Zhang K, Baeckström D, Hansson GC, Reis CA, Hassan H, Bennett EP, Claussen H, Norum L, Varaas T, Kierulf B, Nustad K, Ciborowski P, Konitzki WM, Magarian-Blader J, Finn OJ, Hilgers J. 1998. Summary report on the ISOBM TD-4 workshop: analysis of 56 monoclonal antibodies against the MUC1 mucin. *Tumor Biol.* **19**(Suppl. 1): 1–20.
Smith GP. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**: 1315–1317.
Söderlind E, Strandberg L, Jirholt P, Kobayashi N, Alexeiva V, Åberg AM, Nilsson A, Jansson B, Ohlin M, Wingren C, Danielsson L, Carlsson R, Borrebaeck CAK. 2000. Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nat. Biotechnol.* **18**: 852–856.
Weitkamp J-H, Kallewaard N, Kusuhara K, Bures E, Williams JV, LaFleur B, Greenberg HB, Crowe JE Jr. 2003. Infant and adult human B cell responses to rotavirus share common immunodominant variable gene repertoires. *J. Immunol.* **171**: 4680–4688
Wu TT, Johnson G, Kabat EA. 1993. Length distribution of CDRH3 in antibodies. *Proteins* **16**: 1–7.
Xu JL, Davis MM. 2000. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* **13**: 37–45.
Yamada M, Wasserman R, Reichard BA, Shane S, Caton AJ, Rovera G. 1991. Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood B lymphocytes. *J. Exp. Med.* **173**: 395–407.
Zemlin M, Klinger M, Link J, Zemlin C, Bauer K, Engler JA, Schroeder HW Jr, Kirkham PM. 2003. 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.* **334**: 733–749.

