

A central core structure in an antibody variable domain determines antigen specificity

Pernilla Jirholt¹, Leif Strandberg², Bo Jansson²,
Elias Krambovitis³, Eskil Söderlind²,
Carl A.K.Borrebaeck¹, Roland Carlsson²,
Lena Danielsson² and Mats Ohlin^{1,4}

¹Department of Immunotechnology, Lund University, S-220 07 Lund,

²BioInvent Therapeutic AB, S-223 70 Lund, Sweden and ³Department of Applied Biochemistry and Immunology, Institute of Molecular Biology and Biotechnology, GR-711 10 Heraklion, Greece

⁴To whom correspondence should be addressed.

E-mail: mats.ohlin@immun.lth.se

Antibody binding sites provide an adaptable surface capable of interacting with essentially any molecular target. Using CDR shuffling, residues important for the assembly of mucin-1 specific paratopes were defined by random recombination of the complementarity determining regions derived from a set of mucin-1 specific clones, previously selected from an antibody fragment library. It was found that positions 33 and 50 in the heavy chain and 32, 34, 90, 91 and 96 in the light chain were conserved in many of the clones. These particular residues seem to be located centrally in the binding site as indicated by a structure model analysis. The importance of several of these conserved residues was supported by their presence in a mouse monoclonal antibody with a known structure and the same epitope specificity. Several of these corresponding residues in the mouse monoclonal antibody are known to interact with the antigen. In conclusion, critical residues important for maintaining a human antigen-specific binding site during the process of *in vitro* antibody evolution were defined. Furthermore, an explanation for the observed restricted germline gene usage in certain antibody responses against protein epitopes is provided.

Keywords: antibody evolution/complementarity determining region/mucin-1/paratope/phage display

Introduction

Combinatorial chemistry has over the past few years expanded rapidly to address issues of specificity and characteristics of molecular interactions. Such a development has concerned both low molecular weight organic compounds and more complex biomolecular structures. Molecular evolution of antibodies has attracted much attention, since these molecules are valuable tools in various applications, such as diagnostics and therapeutics and also since they are useful as models for studies of biomolecular interactions. Although much has been learned about antibody–antigen recognition and antibody evolution in the past, much still remains to be elucidated. The evolution of the antibody repertoire in the primary and secondary immune response has been defined for a number of murine hapten-specific responses *in vivo* (Bothwell *et al.*, 1981; Berek and Milstein, 1987; Manser *et al.*, 1987) and molecular evolution of paratopes *in vitro* has also been used to study more complex antigen systems (Schier *et al.*, 1996;

McIntosh *et al.*, 1997). The multitude of different antibody–antigen interactions which, to different degrees, contribute to antigen binding makes it difficult to establish principles important for the evolution of every antibody, but some general principles are likely to emerge.

The construction of large genetic libraries and subsequent selection of molecular variants using, e.g., phage or ribosomal display methodologies, have proven invaluable when addressing issues related to *in vitro* antibody evolution. We have recently devised a system to create large human libraries by introducing variability from naturally occurring complementarity determining regions (CDR) into a single antibody framework region (FR) (Jirholt *et al.*, 1998; Söderlind *et al.*, 1999, 2000; Ohlin *et al.*, 2000) for the purpose of selecting human antibody specificities *in vitro*. From such a CDR implantation library we have selected specific antibody fragments recognising antigens of different types, including haptens, peptides and carbohydrates as well as proteins (Söderlind *et al.*, 2000). A fundamental characteristic of this library is its modularity, which is based on the identity of the FR of all its members. Thus, the CDR implantation library has the potential easily to evolve initially selected clones by shuffling of the selected CDR variability. In contrast to traditional DNA shuffling (Stemmer, 1994), this system permits directed targeting of variability to the CDR and it maintains important key framework structures essentially unaltered. This approach was chosen to minimize effects caused by modifications in the FR, since subtle effects caused by substitutions of variable domain residues residing far from the binding site are frequent (Daugherty *et al.*, 2000).

The aim of this study was to understand the characteristics of antibody binding site evolution *in vitro* and to define parameters important for this process. To investigate specifically events occurring in CDR during *in vitro* evolution of mucin-1 (MUC-1) specific paratopes, we decided randomly to recombine CDR from a set of MUC-1 specific single chain fragments (scFv), selected from the CDR implantation library, by CDR shuffling. This will create new variability, while still retaining a focus on the original specificity of the clones. For this purpose, two shuffled libraries were created using clones derived from a scFv-encoding library, n-CoDeR, specific for the tumour associated antigen MUC-1, as a source of CDR. These novel libraries were further selected on different MUC-1 peptides and retrieved clones were analysed for their genetic and reactivity characteristics.

Materials and methods

Reagents

The antigens used in all selections and screenings were synthetic peptides (Krambovitis *et al.*, 1998), containing different numbers of tandem repeats of the basic MUC-1 20-mer (PAHGVTSA PDTRPAPGSTAP) or fragments thereof. The peptides are denoted MUC-1₂, MUC-1₃ and MUC-1₅ depending on the number of repeats. Certain peptides were

Table I. The variability introduced into the different libraries

Library	Template for the CDR (except CDRH3)	Template for CDRH3
n-CoDeR	Natural B cell variability	Natural B cell variability
Shuffled library L-I	n-CoDeR selected 5 times on MUC-1 ₅	n-CoDeR selected 5 times on MUC-1 ₅
Shuffled library L-II	n-CoDeR selected 4 times on MUC-1 ₅	3:1 ratio of MVGAHALDI to VLAQQRMDV

biotinylated, either directly at their N-terminus or on a trialanine spacer located at their N-terminus. Peptide sequences are given using the standard one-letter code.

Library construction and selection strategies

The original CDR-implantation library, n-CoDeR, had been selected using the MUC-1₅ peptide (Söderlind *et al.*, 2000). We used selected pools of MUC-1 specific clones to generate new variability, based on reassortment of initially selected CDR into new combinations, so-called CDR shuffling (Jirholt *et al.*, 1998). The output from rounds 5 and 4 of the original selections were chosen as sources of CDR variability for shuffling library I and II, respectively (Table I). In the case of shuffling library I (denoted L-I), all CDRH3 present in this pool of selected clones were incorporated into the new library. In contrast, in shuffling library II (denoted L-II), the variability in the CDRH3 was restricted to two different sequences (MVGAHALDI, VLAQQRMDV) found among the initially selected clones, at a 3:1 ratio. The shuffling procedure and library constructions were performed as described by Ohlin *et al.* (2000). In the following discussion, all clones originating from the L-I and L-II libraries are denoted sMuc followed by a clone number or Muc followed by a clone number, if they originate from the original library, n-CoDeR.

Antigen specific selection of the shuffled libraries was performed using biotinylated peptides and subsequent capture on streptavidin-coated paramagnetic beads. L-I was selected twice using a biotin-MUC-1₅ peptide at different concentrations ranging from pM to nM. In addition, L-I, L-II and the original library n-CoDeR were selected twice using a biotin-Ala₃-MUC-1₃ peptide, first at 25 and then at 5 nM. These two selections were followed by one selection using 0.75 nM biotin-MUC-1₂. Positive clones retrieved from the different selections were identified by ELISA and some clones were evaluated further.

Genetic analysis

Sequencing reactions were performed using the Big Dye Terminator kit (PE Biosystems, Warrington, UK) and the samples were analysed by Cybergene AB (Huddinge, Sweden). The CDR were compared at the nucleotide level to immunoglobulin germline gene sequences available in the VBASE germline directory (Tomlinson IM, MRC Centre for Protein Engineering, Cambridge, UK) using the FastA algorithm utilizing the Wisconsin sequence analysis package (Genetics Computer Group, Madison, WI). They were assigned to relevant germline gene loci according to the nomenclature described by Pallarès *et al.* (1998, 1999). Amino acid numbering is as stated by Al-Lazikani *et al.* (1997).

Production of monomeric scFv

The gene III segment was removed from the selected plasmids by *EagI/NotI* restriction digestion using conditions recommended by the supplier (New England Biolabs, Beverly, MA, USA), followed by ligation using T4 DNA ligase (Gibco BRL, Paisley, UK). The resulting plasmids were transformed into

Escherichia coli Top10F' for subsequent expression of soluble scFv fragments. The antibody fragments were purified from bacteria culture supernatant by affinity chromatography over a Ni-NTA agarose superflow column (Qiagen, Hilden, Germany). The monomeric form of the scFv was obtained after gel permeation chromatography on a Superdex 75 FPLC column in phosphate-buffered saline solution (PBS).

Analysis of reactivity characteristics

Reaction rate kinetics of monomeric scFv were evaluated by Biacore analysis. Briefly, streptavidin was coated on to CM-5 sensor chips (Biacore, Uppsala, Sweden) at low density (~300 RU) and biotin-Ala₃-MUC-1₃ was used to coat the available binding sites. Monomeric scFv was allowed to bind and both the association and dissociation phases were recorded at a flow rate of 20 µl/min. Reaction rate kinetic parameters were calculated using the BiaEvaluation 3.0 software. The fine specificity of the scFv fragments was determined by inhibition studies where the scFv was incubated with either a MUC-1₂ peptide, a 6-mer fragment containing the MUC-1 immunodominant epitope, APDTRP, a 9-mer fragment, SAPDTRPAP, containing this epitope or another 9-mer fragment, STAPPAHGV, within MUC-1. Residual binding was evaluated by Biacore analysis. The fine specificities of the scFv were compared with two MUC-1 specific monoclonal antibodies: HMFG-2 (Burchell *et al.*, 1984) and BC2 (Devine *et al.*, 1991).

Flow cytometry

To evaluate if the clones selected on the different peptides could bind MUC-1 on the surface of carcinoma cells, T47D/6 breast carcinoma cells were incubated with a mixture of 5 µg/ml monomeric scFv and 15 µg/ml anti-FLAG M2 antibody (Sigma-Aldrich, St. Louis, MO). The latter antibody was used to dimerize the scFv which all carry the FLAG epitope close to the C-terminus. The scFv-anti-FLAG M2 complexes bound to cells were detected using a rabbit anti-mouse IgG1 antibody labelled with R-phycoerythrin (Zymed, South San Francisco, CA). Flow cytometric analysis was performed on a FACScan cytometer (Becton Dickinson, Franklin Lakes, NJ). An scFv (AE11F/8) against cytomegalovirus glycoprotein B was used as a negative control.

Structure modelling

A homology model was generated by Molecular Simulations (San Diego, CA) for the sMuc159 scFv antibody using the program MODELER. Templates were from the Brookhaven Protein Databank, 1AQK was selected for the V_H domain and 2FB4 for the V_L domain.

Results

Library construction and selection

Selection and characterization of scFv clones specific for MUC-1 peptide had previously been performed (Söderlind *et al.*, 2000). Sequence analysis revealed that these antibody

fragments were largely of two types containing one of two CDRH3 (type A, VLAQQRMDV; type B, MVGAHALDI), but carrying variability in the other CDR. It was evident that the type B CDRH3 came to dominate the selected pools, in particular at later selection rounds. Both type A and type B scFv tended to use similar CDRH1 and CDRH2 in most specific clones and to some extent similar CDRL1 and CDRL3 (Table II). If very stringent selection procedures were employed, one specific scFv-clone (Muc46) also using the type B CDRH3 completely dominated among the selected clones. To study the evolution of these scFv, we decided to use the inherent capacity of the n-CoDeR library to allow rearrangement of the initially selected variability into new combinations. To create shuffling library L-I, all the CDR of clones retrieved after five selections on a MUC-1₅ peptide were independently amplified using PCR and these were subsequently assembled into new complete scFv-encoding genes using our CDR shuffling approach. For shuffling library L-II, CDRH1, CDRH2 and all light-chain CDR of clones retrieved after four selections on a MUC-1₅ were similarly amplified using PCR and assembled together with a 1:3 mixture of type A and B CDRH3, as above. When using these shuffled libraries, a variety of novel binders could be selected (Table II).

Characterization of binding properties

The selected scFv clones carrying type A and B CDRH3 obtained from both the original library and the shuffled libraries were not only positive for the peptides used in the selections. They were also able to bind MUC-1 expressed on the surface of the breast carcinoma cell line T47D/6, after cross-linking with the M2 anti-FLAG antibody (Figure 1). This binding was inhibited by soluble MUC-1₃ peptide, demonstrating the specificity of the binding. To characterize further the epitope recognized by these scFv clones, an inhibition assay was performed using different MUC-1 peptides. All investigated scFv clones were totally inhibited by a 9-mer peptide, containing the epitope (SAPDTRPAP). This is part of the repetitive fragment that is known to be immunodominant *in vivo* (Figure 2). The absence of reactivity against a shorter peptide (APDTRP) identified the epitope recognized by these scFv to be similar to that recognized by the mouse antibody HMFG-2, while being different from the epitope recognized by the mouse antibody BC2.

The common reactivity pattern shared between both type A and B clones may suggest a common epitope and this is in agreement with the many common features within CDRH1 and CDRH2 of the heavy chain of these both types of clones (see below). Also, the CDRH3 had similarities between the type A and B clones, which are identical in length. The presence of an arginine as the last residue of the V_H FR3 and an aspartate as their penultimate CDRH3 residue suggests that they take on similar structure at the base of the loop (Morea *et al.*, 1998).

When studying protein and in particular antibody evolution, the affinities of specific interactions are a major factor to consider. Following an investigation of the affinity constants of the selected clones, it was found that the type A clones had been affinity matured ~10-fold (Table III), a maturation process that mainly involved a reduction of the dissociation rate constant (data not shown). In contrast, the affinities of the type B clones were not substantially improved and many of them had a more rapid dissociation rate than the best type B clone (Muc41 and Muc46) derived from the very stringent selections made on the original library.

Residues defining the antigen binding site

In order to understand the evolution of the antibody paratope, we undertook a genetic study of a large array of MUC-1 specific scFv clones. As pointed out above, sequencing of the scFv genes demonstrated that the CDR used by the clones selected from the shuffled libraries carried a number of key residues that were also associated with specific recognition of this type of antigen in the context of the type A and B CDRH3. In particular, W33 and L50 within V_H were conserved among the type A clones and W33, V50 among the type B clones. In CDRL3, W91 was conserved in both types of clones and in position 96, the type A clones used leucine while the type B clones used tryptophan. Residue 90 within the V_L domain of type A clones was almost exclusively threonine, a residue that could be found here only after a mutation of the germline-encoded sequence. The importance of these particular positions for the binding site was logical in view of an independently made structural model, as they all seemed to be located close to the centre of the binding site, in proximity to the tip of CDRH3 (Figure 3).

Discussion

The present investigation demonstrates that it is possible to use rearrangement of initially selected CDR to diversify the paratope and enhance the binding properties of antibodies. Although the present investigation focuses on a single but biologically important specificity, it highlights certain residues which are centrally located in the binding site that are likely to be of importance for affinity maturation in general.

CDRH1

The presence of W33 in CDRH1 seems to dictate the dominant selection of germline locus 3-07 derived sequences for this particular CDR. This can be concluded since this is one of the few V_H3 sequences that carry a tryptophan in this position. In fact, 47/47 different MUC-1 specific clones carrying type A or B CDRH3 have W33 in the heavy chain, while the unselected n-CoDeR library carries this particular residue in only 6/45 randomly selected clones. The importance of this residue is underlined by the fact that CDRH1 from germline gene DP-47 (locus 3-23), was not selected despite its higher frequency in the unselected library, although the only difference between this sequence and the CDRH1 of germline locus 3-07 is an alanine in position 33. Further support for the importance of W33 is found in the structure model (Figure 3), indicating that the tryptophane side chains are extensively exposed to the environment.

Although W33 in the heavy chain was frequently found in the MUC-1 specific clones derived from n-CoDeR, it is not critical for this specificity. Other MUC-1 specific clones using other CDRH3 not carrying W33 were identified, although their exact epitope specificity may differ from the type A and B clones.

CDRH2

The presence of L50 in CDRH2 in all type A clones clearly indicates this residue as critical for establishing this MUC-1 paratope. Its importance may partly be related to a direct antigen contact effect (MacCallum *et al.*, 1996) or to its importance for the paratope structure (for discussion, see below). The rarity of L50_H in the original library is probably a restricting factor giving the low frequency of unique type A versus type B clones (6:41 ratio) among the selected ones. In

Table II. Amino acid sequences of CDRH1-2 and CDRL1-3 of a MUC-1 specific scFv antibody fragments obtained from the original n-CoDeR library or from shuffled libraries

clone type ^a	clone	CDRH1 ^b		CDRH2 ^b		CDRL1 ^{b,c}		CDRL2 ^b		CDRL3 ^b	
		↓ SYWMS	GL ^d	↓	GL	↓ ↓	GL	XNXXRPS	GL	↓ ↓	GL
type A clones (selected from n-CoDeR library)	Muc23		3-07	VISYDGSNKYYADSVKG	3-30	SGSSSNIG---XNXVN	1-44	GDNN...	1-40	AXWDDXLXGXV	1-36/ 1-44
	Muc67	NT...	3-15	LM.....R.	3-30.3	S.T..	1-44	R.NQ...	1-47	.A...S.N.V.	1-47
	Muc201		3-07	L..GS.GSI.....A..	3-23	S.T..	1-44	G.GN...	1-40	.T...F.G.L.	1-47
type A clones (selected from shuffled library)	sMuc129		3-07	LV.....E.....	3-30/ 3-30.3	S.T..	1-44	D.NR...	1-51	.T...F.G.L.	1-47
	sMuc158	RH...	3-07	L.....HE.....M..	3-30	YIGG.A..	1-36/ 1-44	A.SN...	1-40	.T...F.G.L.	1-47
	sMuc159	N...	3-07	L..F.....A..	3-30	S.T..	1-44	G.NN...	1-40	.T...F.G.L.	1-47
type B clones (selected from n-CoDeR library)	Muc35		3-07	K.....	3-30.3	...H.....T.T..	1-44	D.NK...	1-51	.A...S.N.W.	1-36/ 1-44
	Muc41	R...	3-07	H.....	3-30	S.P..	1-44	G.IN...	1-40	.V...S.G.W.	1-47
	Muc46		3-07	T...H.....	3-30.3	V.---GHP..	1-44	D.NR...	1-51	.A.H.G.R.W.	1-47
type B clones (selected from shuffled library)	sMuc103		3-07	M..N.RD.....	3-30	V.---N.I..	1-36/ 1-44	G.SN...	1-40	.A...R.S.W.	1-47
	sMuc125		3-07	.T.....	3-30/ 3-30.3	...N.....YIGG.A..	1-36/ 1-44	G.SN...	1-40	.A...R.S.W.	1-47
	sMuc152	N...	3-07	T...H.G....	3-30/ 3-30.3	..G...V.---GHP..	1-44	G.SN...	1-40	.A...S.S.W.	1-47

^aThe sequences of CDRH3 are indicated by the clone type number (A, VLAQQRMDV; B, MVGAHALD). All variants of type A and a selection (6/41) of type B clones are shown.

^bThe sequences are compared with the most commonly used germline genes (CDRH1, locus 3-07; CDRH2, 3-30) or to a consensus sequence (CDRL1-3) as indicated in the header. X in the consensus sequence is used when the overall homology between all clones is <75%. Residues W33_H, L50_H, T32_L, N34_L, T90_L, W91_L and L96_L believed to be important for the paratope of type A clones, which are highlighted in Figure 3, are marked by arrows in the heading.

^cThe three additional residues found in CDRH1 of sMuc125 and 3Muc158 are most likely derived from a somatic mutation event occurring *in vivo* (Ohlin and Borrebaeck, 1998) or possibly through a PCR-related error as they are not encoded by any known germline gene.

^dGL loci according to Pallares *et al.*, 1998, 1999) carrying the most similar germline sequences; only genes considered to be functional were used in this comparison.

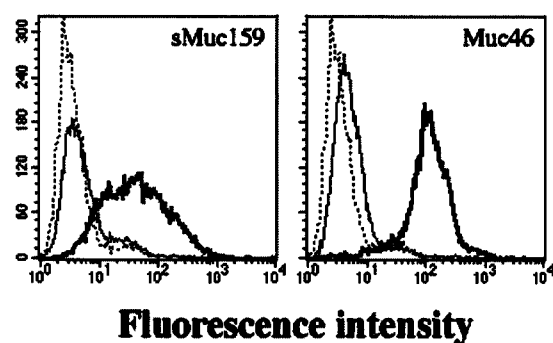


Fig. 1. Flow cytometric analysis of binding of sMuc159 (left) and Muc46 (right) to the breast carcinoma cell line T47D/6. Binding of the specific scFv is shown with a thick line and background staining using an irrelevant cytomegalovirus-specific scFv is shown with the broken line. The binding of the MUC-1-specific scFv was inhibited in the presence of soluble MUC-13 (thin line).

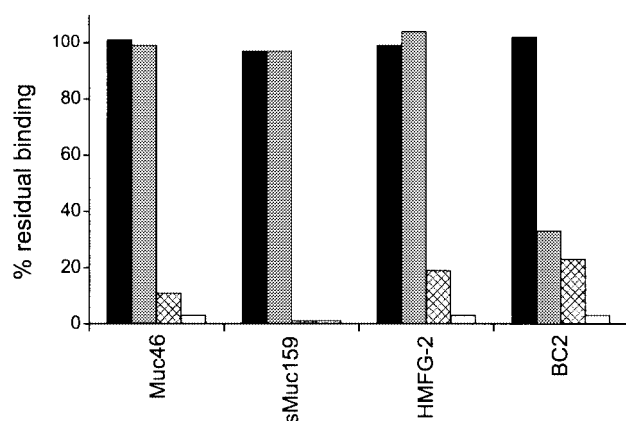


Fig. 2. Studies of the ability of synthetic peptides to inhibit the binding of antibodies (HMFG-2 and BC2) and scFv (Muc46 and sMuc159) to immobilized biotinylated MUC-13 peptide. The results represent inhibition with 10 μ g/ml of 9-mer peptides STAPPAHGV (black) and SAPDTRPAP (white), 6-mer peptide APDTRP (grey) and MUC-12 peptide (checked).

Table III. Reaction rate constants of MUC-1 specific scFv as assessed by surface plasmon resonance analysis

Clone type ^a	Clone	K_A (10^6 M ⁻¹)
Type A clones (selected from original library)	Muc23	1
	Muc67	3
	Muc201	5
	sMuc129	13
Type A clones (selected from shuffled library)	sMuc158	13
	sMuc15	38
	sMuc103	33
Type B clones (selected from original library)	Muc41	21
	Muc46	10

^aThe clones contain the VLAQQRMDV (type A) or the MVGAHALDI (type B) CDRH3 sequences.

order for this residue to appear in this position, the germline sequence had to undergo a somatic mutation. Only one germline gene belonging to the V_H3 family locus (3-43) uses a leucine in this position. The fact that this germline gene is never used in the CDRH2 by the selected clones may be due to its low frequency in the library or that it carries other residues incompatible with a MUC-1 specific paratope.

In the studied CDRH2, mutated sequences originating from loci 3-30 or 3-30.3 were preferred. However, one clone

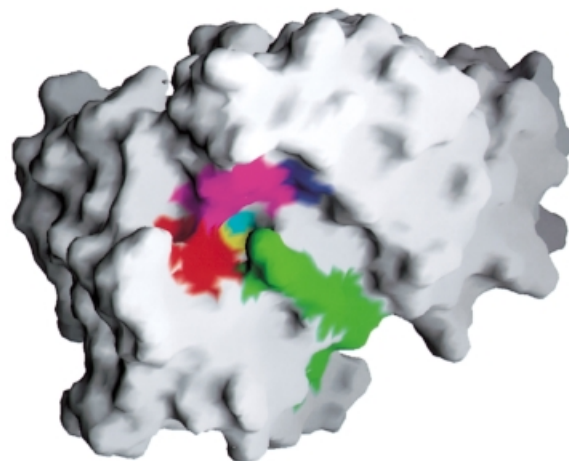


Fig. 3. Model of sMuc159 scFv antibody. The amino acid side chains in CDRH3 are coloured green, W33_H red, L50_H yellow, T32_L blue, W91_L magenta and L96_L cyan. Residues N34_L and T90_L are not visible in this projection

(Muc201) used a locus 3-23 derived CDRH2, but also in this case the sequence carried a substitution to leucine in the critical position 50 (Table II). The relative rarity of DP-47 (locus 3-23) derived CDRH2 being accepted among the selected clones is probably related to the fact that codon 50 in this germline gene requires two point mutations to mutate to a leucine codon. This will be a much rarer event than the single mutation required to substitute the valine found in 3-30/3-30.3 for leucine. Consequently, the events needed to create the leucine in position 50 will give that locus 3-23 derived CDRH2 will not be selected at a high frequency despite their high frequency within the unselected n-CoDeR library.

CDRL1 and CDRL3

Additional sequence restrictions were observed for the light chain. It was evident that there was an increased frequency of sequences deriving from the locus 1-44 in the CDRL1 and a decreased frequency of sequences deriving from the locus 1-47 in the MUC-1 specific clones, when compared with random, unselected clones derived from the n-CoDeR library (n-CoDeR, 47% 1-44 and 31% 1-47; MUC-1 specific clones, 96% 1-44 and 0% 1-47). This suggests that the residues 32_L and/or 34_L are important, since these are the only residues differing between 1-44 and 1-47 gene products. Both of these residues are believed to be located close to other important residues, as defined above and to the tip of the CDRH3 loop and N34_L are conserved among the selected clones. Also, for type A clones T32_L, which is a highly exposed residue, was conserved to a great extent, while the type B clones permitted variability in this position.

Not only features of CDRL1 but also CDRL3 were important for establishing the paratope. The importance of either tryptophan or leucine in position 96_L is pinpointed by their relatively low frequencies in the unselected n-CoDeR library (9/36 and 5/36, respectively). These low frequencies stems from the fact that there is only a single allele found in the $J\lambda$ gene segment loci which encodes tryptophan and none encodes leucine. Therefore, leucine 96_L have to originate either through the recombination event itself or subsequent somatic hypermutation taking place in the B cell from which this CDR was derived, while tryptophan can originate from a single germline gene segment or through the above-mentioned processes. We

believe that restrictions in 91_L and 96_L may be due to either their involvement in antigen binding or to structural constraints, as outlined below. Furthermore, type A clones were significantly restricted by the use of an unusual T90_L residue. The side chain of this residue in general does not contact antigen and it is not accessible to the environment (MacCallum *et al.*, 1996), suggesting that sequence restriction at this site is related to structural constraints. This is strongly supported by the finding that this residue is one of those playing an important role in directly determining the conformation of the hyper-variable loops (Chothia *et al.*, 1998).

Overall paratope considerations

As seen in the modelled scFv (Figure 3), several of the conserved residues in the MUC-1 specific clones appear to be located close to the centre of the binding site. Some of these positions have also been shown to be important contributors to the paratope in other studies. Residue 33 of the VH domain is often important for the recognition of antigen and it is also frequently a contact residue (MacCallum *et al.*, 1996). A tryptophan in this position has been shown to be important for the recognition of photodamaged DNA (Kobayashi *et al.*, 1999). Similarly, residues 50_H, 32_L, 91_L and 96_L frequently interact with the antigen, whereas 90_L and 34_L rarely do so.

Apart from directly interacting with the antigen, structural requirements may also cause restrictions in some of the positions discussed above (Chen *et al.*, 1999). Replacing V_L residues 91_L, 96_L and 100b_H with other residues of various sizes has demonstrated that only residues up to a certain limit of van der Waals radius will be accepted at the bottom of the binding pocket in order to maintain an intact binding site. In the selected type B clones from the present study the small valine residue [normalized van der Waals radius, 3.00 (Fauchere *et al.*, 1988)] in position 50_H co-exists with W96_L (normalized van der Waals radius, 8.08). The type A clones employ L50_H (normalized van der Waals radius, 4.00) and are thus restricted to using a smaller residue such as L96_L as the penultimate residue of CDRL3 in order for the paratope to remain intact. In fact, we suggest that the L50_H, L96_L, T90_L and N34_L in

proximity to the apex of the CDRH3 and the centre of the binding site were selected as a unit so as to adopt an appropriate structure of the binding site to fit the antigen in question.

Comparison with other MUC-1 specific antibodies

Creation of a suitable MUC-1 specific paratope can be solved in ways different from the ones selected by the type A and B clones (Henderikx *et al.*, 1998). However, a MUC-1 specific mouse antibody, SM3, for which the structure (PDB: 1SM3) has been determined (Dokurno *et al.*, 1998), displays several features common to the specificity described here. The epitope of SM3 has previously been defined as the APDTRP sequence (Price *et al.* 1998) but the antibody makes contact with residues outside of that sequence in a way suggesting that it recognizes an epitope similar to that recognized by the type A and B clones. Several of the key residues in the combining site are identical between the mouse and human paratopes. In particular, F27_H, Y32_H, W33_H and W91_L are shared. The W96_L residue is also shared by type B clones and SM3. In addition, N31_H used by SM3 is used by some (e.g. by sMuc159) but not all type A and B clones. In other MUC-1 specific clones, residue 31_H is frequently a serine, which is the amino acid encoded by the unmutated germline CDRH1 sequence derived from locus 3-07 (Table II).

The five residues F27_H, N31_H, Y32_H, W33_H and W91_L together account for not less than 43.6% of the buried molecular surface area of SM3. These residues alone make contacts (van der Waals interactions or hydrogen bonds) with eight out of nine residues within the sequence SAPDTRPAP (contacted residues in upper case) and thus seem to be critical for the establishment of the binding site. When comparing the structure of SM3 with the model of sMuc159, the similar localization of the side chains of F27_H, N31_H, Y32_H, W33_H and W91_L is evident (Figure 4). We suggest that these antigen-binding sites recognize the epitope in a similar manner. Other residues as well as the length of several CDR differ between SM3 and the scFv, so the paratopes are able to establish a functional binding site in different ways but based on these common residues.

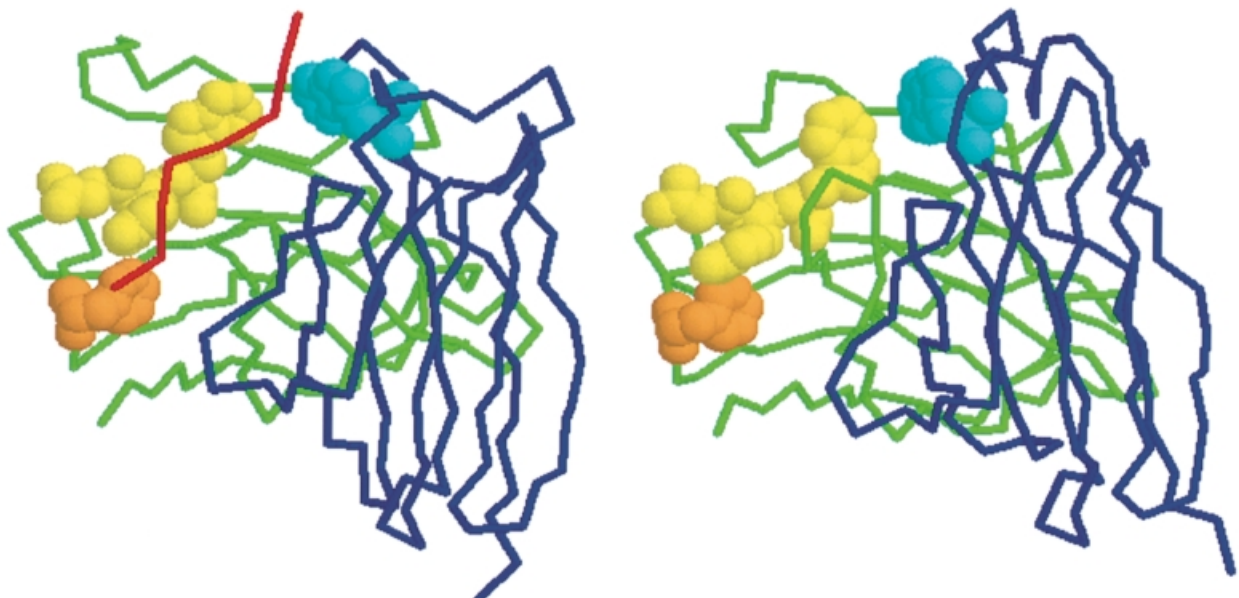


Fig. 4. The structure of the SM3 Fv (left) and the model of sMuc159 (right). V_H, V_L and the antigen (SM3 only) are shown in green, blue and red, respectively. The common, important residues F27_H (orange), N31_H, Y32_H, W33_H (all in yellow) and W91_L (cyan) are highlighted.

General structural aspects

The critical residues of SM3 and type A and type B clones have all been brought together by different pathways. F27_H, N31_H, Y32_H and W33_H are all encoded by the germline gene IGHV6S1 (GenBank: X03398) from which the SM3 antibody probably originated. In contrast, the CDR implantation approach used to construct the antibody library brought naturally occurring residues from different genes belonging to the VH3 family into a predefined framework, which is also of VH3 origin. The shuffled CDR sequences included VH positions 31–33 and the framework encoded among other things the F27H residue. As previously discussed, the predominance of CDRH1 derived from locus 3-07 relates to its specific use of W33_H, in contrast to most other human germline genes. Y32_H is a common feature among many antibodies belonging to the VH3 family and, as noted above, some variability seems to be tolerated at position 31_H. W91_L is a common feature of most mouse and some human lambda chain CDRL3. The requirement for W91_L probably is a restricting factor limiting the origin of the CDRL3 in type A and B MUC-1 specific clones to sequences from a few germline genes.

Based on this evaluation, we suggest that the human and mouse antibody repertoires may recognize an antigen in a similar manner, provided that the germline has members displaying similar or identical residues in key positions. It has long been known that small defined antigens such as *p*-azophenylarsonate (Siekevitz *et al.*, 1983), phenyloxazalone (Kartinen *et al.*, 1984) and (4-hydroxy-3-nitrophenyl)acetyl (Bothwell *et al.*, 1981) give rise to restricted antibody responses in mice. Recently, it has become apparent that also the human and mouse immune systems recognize what may be individual protein epitopes by a restricted repertoire of germline genes (Ohlin *et al.*, 1994; Binley *et al.*, 1996; Ohlin and Borrebaeck, 1996; Ikematsu *et al.*, 1998; Goldbaum *et al.*, 1999; Lavoie *et al.*, 1999; van den Brink *et al.*, 2000). The basis for immunodominance in certain immune responses is probably due to the existence of germline members fitting well to a particular epitope (Seidl *et al.*, 1999). In fact, it was recently demonstrated that also two independently raised antibodies specific for the same epitope on influenza virus hemagglutinin, despite the presence of many other sequence dissimilarities, shared many critical contact residues in their interaction with antigen, (Fleury *et al.*, 2000). We believe that the same effect is observed here and that the described critical residues in the similar CDRH1, CDRH2 and CDRL1 of most type A and B clones are important factors in determining this particular epitope specificity.

In conclusion, this study shows the power of using initially selected clones for rearrangement of CDR and to study amino acid residues important for assembly of the paratope. By using a stepwise evolutionary process, general information about which residues are likely to be important is gained. Owing to the identical framework regions of the clones, information about important residues within the CDR can be defined. Also, the existence of multiple clones which have similar or identical CDRH3 makes it possible to define important key residues for other CDR using sequence similarity studies. The conservation of residues 33_H, 50_H, 32_L and/or 34_L, 90_L, 91_L and 96_L may define the minimal conservation necessary for maintaining a functional binding site in association with a particular type of CDRH3, around which variability can be introduced to evolve the binding site. These residues would comprise at least a part

of the central specificity-determining structure created during rearrangement of gene segments into complete variable region genes. In agreement with our observations, Hawkins *et al.* (1993) proposed that a band of residues in the middle of the paratope makes up a critical centre crucial for the interaction with antigen. Similarly, it has been suggested (Tomlinson *et al.*, 1996; Ignatovich *et al.*, 1997) that the somatic hypermutation process *in vivo* mainly introduces variability around such a core structure in order to evolve its function, while keeping the core intact. This knowledge can together be used for evolution of antigen specific clones in a more directed fashion *in vitro*, giving a high functionality of the created combinatorial libraries and may reduce the size of libraries needed to find optimal clones. Furthermore, we suggest a molecular basis for restrictions in germline usage in specific antibody repertoires. The library approach employed may also be used to define more precisely the basis for such restrictions observed even with more complex antigens.

Acknowledgements

The authors thank Anna-Maria Larsson, Vessela Alexeiva and Anna Nilsson for their assistance with the library constructions and selections and Jessica Frygeli for DNA sequencing of the clones. This study was supported financially by grants from NUTEK, TFR and the European Union, grant BIOT4CT-950252.

References

- Al-Lazikani, B., Lesk, A.M. and Chothia, C. (1997) *J. Mol. Biol.*, **273**, 927–948.
- Bates, P.A., Dokurno, P., Freemont, P.S. and Sternberg, M.J.E. (1998) *J. Mol. Biol.*, **284**, 549–555.
- Berek, C. and Milstein, C. (1987) *Immunol. Rev.*, **96**, 23–41.
- Binley, J.M., Ditzel, H.J., Barbas, C.F., III, Sullivan, N., Sodroski, J., Parren, P.W. and Burton, D.R. (1996) *AIDS Res. Hum. Retroviruses*, **12**, 911–924.
- Bothwell, A.L., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. and Baltimore, D. (1981) *Cell*, **24**, 625–637.
- Burchell, J., Taylor-Papadimitriou, J., Granowska, M. and Britton, K. (1984) *Behring Inst. Mitt.*, **74**, 87–93.
- Chen, G., Dubrawsky, I., Mendez, P., Georbiou, G. and Iverson, B.L. (1999) *Protein Eng.*, **12**, 349–356.
- Chothia, C., Gelfand, I. and Kister, A. (1998) *J. Mol. Biol.*, **278**, 457–479.
- Daugherty, P.S., Chen, G., Iverson, B.L. and Georgiou, G. (2000) *Proc. Natl Acad. Sci. USA*, **97**, 2029–2034.
- Devine, P.L., Layton, G.T., Clark, B.A., Birrell, G.W., Ward, B.G., Xing, P.X. and McKenzie, I.F. (1991) *Biochem. Biophys. Res. Commun.*, **178**, 593–599.
- Dokurno, P. *et al.* (1998) *J. Mol. Biol.*, **284**, 713–728.
- Fauchere, J.L., Charton, M., Kier, L.B., Verloop, A. and Pliska, V. (1988) *Int. J. Pept. Protein Res.*, **32**, 269–278.
- Fleury, D., Daniels, R.S., Skehel, J.J., Knossow, M. and Bizebard, T. (2000) *Proteins*, **40**, 572–578.
- Goldbaum, F.A., Cauerhff, A., Velikovskiy, C.A., Llera, A.S., Riottot, M.M. and Poljak, R.J. (1999) *J. Immunol.*, **162**, 6040–6045.
- Hawkins, R.E., Russell, S.J., Baier, M. and Winter, G. (1993) *J. Mol. Biol.*, **234**, 958–964.
- Henderikx, P., Kandilogiannaki, M., Petrarca, C., von Mensdorff-Pouilly, S., Hilgers, J.H.M., Krambovitis, E., Arends, J.W. and Hoogenboom, H.R. (1998) *Cancer Res.*, **58**, 4324–4322.
- Ignatovich, O., Tomlinson, I.M., Jones, P.T. and Winter, G. (1997) *J. Mol. Biol.*, **268**, 69–77.
- Ikematsu, W., Kobarg, J., Ikematsu, H., Ichiyoshi, Y. and Casali, P. (1998) *J. Immunol.*, **161**, 2895–2905.
- Jirholt, P., Ohlin, M., Borrebaeck, C.A.K. and Söderlind, E. (1998) *Gene*, **215**, 471–476.
- Kartinen, M., Griffiths, G.M. and Milstein, C. (1984) *Ann. Immunol.*, **135C**, 143–148.
- Kobayashi, H., Kato, J., Morioka, H., Stewart, J.D. and Ohtsuka, E. (1999) *Protein Eng.*, **12**, 879–884.
- Krambovitis, E., Hatzidakis, G. and Barlos, K. (1998) *J. Biol. Chem.*, **273**, 10874–10879.
- Lavoie, T.B., Mohan, S., Lipschultz, C.A., Grivel, J.C., Li, Y., Mainhart, C.R., Kam-Morgan, L.N., Drohan, W.N. and Smith-Gill, S.J. (1999) *Mol. Immunol.*, **36**, 1189–1205.

- MacCallum,R.M., Martin,A.C.R. and Thornton,J.M. (1996) *J. Mol. Biol.*, **262**, 732–745.
- McIntosh,R.S., Asghar,M.S., Kemp,E.H., Watson,P.F., Gardas,A., Banga,J.P. and Weetman,A.P. (1997) *J. Clin. Endocrinol. Metab.*, **82**, 3818–3825.
- Manser,T., Wysocki,L.J., Margolies,M.N. and Gefter,M.L. (1987) *Immunol. Rev.*, **96**, 141–162.
- Morea,V., Tramontano,A., Rustici,M., Chothia,C. and Lesk,A.M. (1998) *J. Mol. Biol.*, **275**, 269–294.
- Ohlin,M. and Borrebaeck,C.A.K. (1996) *Mol. Immunol.*, **33**, 583–592.
- Ohlin,M. and Borrebaeck,C.A.K. (1998) *Mol. Immunol.*, **35**, 233–238.
- Ohlin,M., Owman,H., Rioux,J.D., Newkirk,M.M. and Borrebaeck,C.A.K. (1994) *Mol. Immunol.*, **31**, 983–991.
- Ohlin,M., Jirholt,P., Thorsteinsdottir,H.B. and Borrebaeck,C.A.K. (2000) In Zanetti,M. and Capra,J.D. (eds), *The Antibodies*, vol. 6. Harwood, Amsterdam, pp. 81–104 (in press).
- Pallarès,N., Frippiat,J.P., Giudicelli,V. and Lefranc,M.-P. (1998) *Exp. Clin. Immunogenet.*, **15**, 8–18.
- Pallarès,N., Lefebvre,S., Contet,V., Matsuda,F. and Lefranc,M.-P. (1999) *Exp. Clin. Immunogenet.*, **16**, 3 6–60.
- Price,M.R. *et al.* (1998) *Tumor Biol.*, **19** (Suppl. 1), 1–20.
- Schier,R., Balint,R.F., McCall,A., Apell,G., Larrick,J.W. and Marks,J.D. (1996) *Gene*, **169**, 147–155.
- Seidl,K.J., Wilshire,J.A., MacKenzie,J.D., Kantor,A.B., Herzenberg,L.A. and Herzenberg,L.A. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 2262–2267.
- Siekevitz,M., Huang,S.Y. and Gefter,M.L. (1983) *Eur. J. Immunol.*, **13**, 123–132.
- Stemmer,W.P. (1994) *Nature*, **370**, 389–391.
- Söderlind,E., Ohlin,M. and Carlsson,R. (1999) *Immunotechnology*, **4**, 279–285.
- Söderlind,E. *et al.* (2000) *Nature/BioTechnology*, in press.
- Tomlinson,I.M., Walter,G., Jones,P.T., Dear,P.H., Sonnhammer,E.L. and Winter,G. (1996) *J. Mol. Biol.*, **256**, 813–817.
- van den Brink,E.N., Turenhout,E.A.M., Davies,J., Bovenschen,N., Fijnvandraat,K., Ouweland,W.H., Peters,M. and Voorberg,J. (2000) *Blood*, **95**, 558–563.

Received June 8, 2000; revised October 31, 2000; accepted November 9, 2000