ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

The development of a single vector recombination system to make large phage antibody libraries.

Thesis submitted for the degree of

"Doctor Philosophiae"

CANDIDATE
Daniele Sblattero

SUPERVISOR
Andrew Bradbury

November 1999

SISSA - SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI

TRIESTE
Via Beirut 2-4

TRIESTE

ISAS - SISSA
INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

The development of a single vector recombination system to make large phage antibody libraries.

Thesis submitted for the degree of

"Doctor Philosophiae"

CANDIDATE
Daniele Sblattero

SUPERVISOR
Andrew Bradbury

November 1999

Declaration

The work described in this dissertation was carried out at the International School for Advanced Studies, Trieste Italy, between November 1995 and August 1999. All work reported arise from my own experiments and this work has not been submitted in whole or in part to any other university.

Daniele Sblattero

November 1999

Table of Contents

Chapter One: general introduction

Chapter Two: Materials and Methods

Chapter Three:
pDAN New Phage Display Vector

3.1 Introduction

3.2 Results

3.3 Discussion

Chapter Four: construction of immune libraries from coeliac disease patients

4.1 Introduction

4.2 Results

4.3 Discussion

Chapter Five:
Single bacteria making hundreds of different antibodies: the creation of teralibraries

Chapter One:
General introduction

1.1 Summary of the data presented

1.1.1 The aim of the reasearch

Phage display has been recently introduced as a means of making antibodies in vitro(Barbas, Kang et al. 1991; Griffiths, Malmqvist et al. 1993; Griffiths, Williams et al. 1994; Vaughan, Williams et al. 1996; Sheets, Amersdorfer et al. 1998). In general, the affinity of the antibodies isolated is proportional to the initial size of the library used for selection. This is based on both theoretical (Perelson and Oster 1979) and practical (Vaughan, Williams et al. 1996) considerations. In view of this, the creation of larger and larger libraries has become an important goal in the use of this technology in the selection of antibodies against any antigen.
The aim of my research were:
- the design and the validation of a new phagemid vector for phage display of antibody fragments (scFvs).
- The design of a new strategy for "one vector" *in vivo* recombination.
- The construction of a large naive scFv libray

1.1.2 Results obtained

The following chapters will describe all the steps taken in order to achieve the final goal: the construction of a large phage antibody library.
First of all there will be the description of the design of a new phagemid vector in which all the possible variables were considered from a theoretical point of view, and the best elements were chosen for use.
Secondly data confirming the quality of the vector will be presented. This was achieved first with the cloning of several mAb and then with the construction of two small immune libraries that allow the selection of several specific and functional scFv.
The third part will describe a method which uses a single vector to exploit the reversibility of cre catalysed recombination. This method involves the creation of a relatively small primary library ($7x10^7$ was used here) in a phagemid vector in which the VH and VL genes are separated by two non-homologous lox sites. The heavy and light chain genes in this primary library

are then recombined by infecting the phagemid particles into cre expressing bacteria at high multiplicity of infection (MOI). Under these conditions many different phagemid particles enter a single bacteria and the VH and VL genes are exchanged between different phagemids, creating many new VH/VL combinations, all of which are functional. This ends with the production of a large naive library of preaviously unattainable size that was validated by the selection of antibodies, with high affinity, against a large number of different protein antigens.

In order to fully appreciate the different aspects of experimental works and its underlying strategies some theoretical paragraphs are included in this first chapter.

1.2 Biology of filamentous phage

1.2.1 Filamentous Phage

Filamentous bacteriophages are a group of related viruses which only infect gram-negative bacteria and specifically adsorb to the tip of pili (Model and Russel 1988). They are long threadlike particles, with no organization into head or spikes. Their genetic material consists of a single stranded, closed circular DNA molecule. The most thoroughly investigated are the Ff phages, fd, f1 and M13, which infect *E. coli*. Ff denominates those filamentous phages which require F pili as the host receptor. They resemble each other so closely that is legitimate to consider them as slight mutations of basically the same phage. Except for a small number of base changes their DNA sequences are identical. Other Filamentous phage include IKE (Khatoon, Iyer et al. 1972) which infects bacteria bearing N pili and have probably evolved from a common ancestor as they have identical gene number and order and the genome display 55% homology (Peeters, Peters et al. 1985) with the Ff group.

1.2.2 The phage genome

The Ff phages contain a small and highly organized genome. The DNA include 6408 nucleotides in strain fd. Except for some discrete regions which are folded into "hairpin" loop the DNA has no secondary structure. The genome of the Ff phage encode 11 genes, two of these overlap and are

in-frame with larger genes. These genes are grouped on the DNA according to their function in the life cycle of the bacteriophage. One group encodes proteins required for DNA replication, (g2p, g5p and g10p) the second group encodes the proteins which make up the capsid (g3p, g6p, g7p, g8p and g9p) and the third group encodes three proteins involved in the membrane associated assembly of the phage (g1p, g4p and g11p). There is also the "intergenic region" a short strech of DNA which encodes no proteins. It contains the sites of origin for the synthesis of DNA as well as a hairpin region which is the site of initiation for the assembly of the phage particles (packaging signal).

1.2.3 Morphology and structure of phage particle

The wild type Ff phage particle is approximately 6.5 nm in diameter and 930 nm in length. The DNA is encased in a somewhat flexible cylinder composed of approximately 2700 copies of a major protein, the product of gene 8, p8. X-ray diffraction and other physical studies have given a fairly complete description of the p8 cylinder portion of the virion (Marvin, Hale et al. 1994; Overman and Thomas 1995; Williams, Glibowicka et al. 1995) Except for the amino terminal 5 amino acids, p8 appears to be present in an uninterrupted alpha helical conformation. The amino terminal portion of p8 is exposed on the outside of the particle, the carboxyl-terminal 10-13 residues form the inside wall of the particle and interact with the DNA. The axis of helical p8 in the phage particle is tilted approximately 20° to the long axis of the particle, gently wrapping around the long axis of the virus in a right-handed way.

The phage cylinder is capped by two structures, one made of 5 molecules each of p7 and p9, on the side that leaves the bacterial membrane first (Grant, Lin et al. 1981; Gailus and Rasched 1994) the arrangement of these proteins in the capsid and their interaction with p8 is not well known. The other end is composed of 3-5 molecules each of p3 and p6, and account for about 10-16 nm of the phage length. p3 and p6 appear to specifically recognize each other as they remain associated following disruption of the phage particle with certain detergents (Gailus and Rasched 1994; Endemann and Model 1995).

1.2.4 Gene 3 protein

Phage infection is mediated by the phage gene 3 protein (p3), a minor coat protein.

Three to five copies of p3 cap one end of the extended filamentous phage particle. The p3 is divided into three domains of 67 (D1), 131 (D2) and 150 (D3) amino acids, separated by glycine rich peptide linkers of 19 (G1) and 39 (G2) amino acids (see fig 1.1).

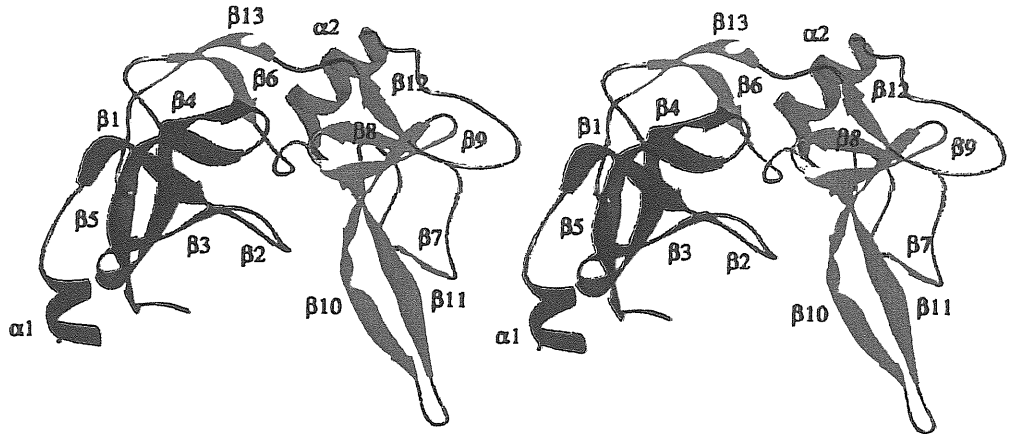


Fig 1.1 Gene 3 protein structure. Ribbon representatio of D1 (blue) domain and D2 domain (red) (Lubkowski, Hennecke et al. 1998) .

Deletion analysis led to the provisional assignment of different functions for the individual domains. The function assigned to the C-terminal domain is the anchoring of p3 in the phage particle. The transmembrane segment mediates incorporation of p3 into the host inner membrane, a prerequisite for incorporation into the nascent phage particle. No function has been assigned to the two glycine rich section in p3, but the region separating D2 and D3 is fairly tolerant of the insertion of foreign sequences (Smith 1985) and may simply have a tethering function. It has been observed, however, that they enhance the infectivity of the phage (Endemann, Gailus et al. 1993) probably by conferring flexibility in connecting the domains and adjusting the required distance between them.

The principal structural element of D1 are β sheets: the N-terminal residues 2-9 form an a helix, and the remainder of D1 is built of a single β-sheet folded into a distorted barrel. The barrel contains five antiparallel β strands and together with two additional β-strands from the D2 domain, form a

seven stranded antiparallel sheet. The larger D2 domain consists of eight β-strands with six of them arranged within a mixed β-sheet and the first and the last partecipating in the D1 structure. D1 and D2 associate in a horseshoe-like shape with in the central channel a large number of exposed aliphahtic and threonine residues. Thus, it is very likely that this central part of the molecule constitutes a binding site, probably for the pilus.

The N-terminus of D1 and the C-terminus of D3 (and thus the rest of the phage) are close together. But peptide or protein, which is displayed fused to the N-terminal is far from the inside of the horseshoe, where the interaction with the pilus may occur, which may explain why they do not interfere with the infection process. Infection may, however, be compromised through p3 folding problems caused by the fusion partner. Such interferences are alleviated in phagemid systems, where additional wild type p3 is provided in trans by a helper phage.

1.2.5 The infection and replication process

Infection of E. coli by the Ff phages is at least a two step process mediated by the p3. The two N-terminal domains of the proteins appear more directly involved in the infection process.

The first step involves the interaction of p3 with the tip of the F conjugative pilus. The intimate association and horseshoe shape of the D1-D2 domains suggest how the initial phase of the phage infection may proceed (Lubkowski, Hennecke et al. 1998). Several line of evidence point to the D2 domain as making the most important contacts with the pilus. Deng (Deng, Malik et al. 1999) shows that the construct p3-D1D2 and p3-D2 are able to bind to the F-pilus, competing with the wild type p3 in the intact virion and delaying infection by the wild type phage. The D2 domain is responsible for the absorption to the tip of the F-pilus and is capable of this without cooperation from the D1 domain. Any interaction with the D1 domain in the intact p3 plays no essential part in the binding of the D2 domain to the pilus. Furthermore in the absence of D2 only a low background infection is observed. Recent evidence obtained by mutational analysis of p3 showed, that in contrast to what may have been expected from the X-ray structure, the pilus binding surface may not be the inner face of the horseshoe, but an outer part of domain 2. This complements results obtained on the crystallisation of a p3 D1-C terminus tolA fusion protein, which showed that tolA interacts with domain 1 of p3 at the same site as

domain 2 interacts with domain 1, even though the structures of tolA and domain 2 are completely different.
Retraction of the pilus, presumably by depolymerization of pilin subunits into the inner membrane, brings the tip of the phage to the membrane surface.

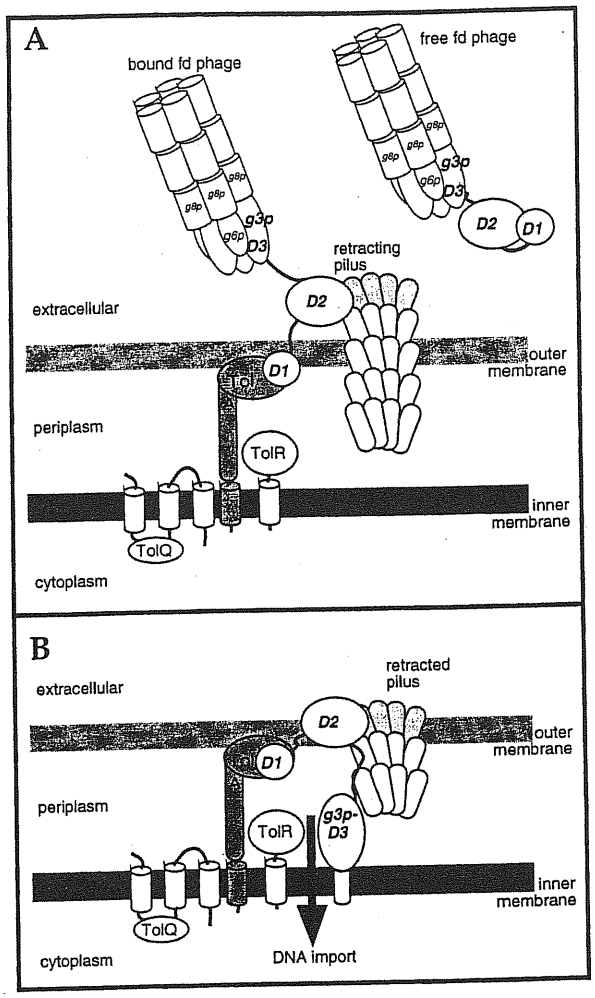


Fig 1.2 Model of infection of E. coli by filamentous phage (Riechmann and Holliger 1997)

The second step involves the integration of the p8 major capsid protein into the inner membrane together with the translocation of the DNA into the cytoplasm. Besides functional p3, penetration of the bacterial membrane requires the proteins TolQ, TolR and TolA that together with TolB form a complex in the E. coli inner membrane. The TolQRA complex appears to be preferentially located at junctions of the inner and outer membrane (Guihard, Boulanger et al. 1994) and is most probably the entry port for the phage into the host cell. Recent studies have shown that after D2 binds to the pilus the D1domain is free to interact with the TolQRA complex, and in particular with the C-terminal domain of TolA (Hollinger and Riechmann 1997;

Riechmann and Holliger 1997). It is still unknown how the interaction with TolA helps the phage to complete infection. Binding of the phage p3 to first the pilus and then TolA may therefore simply position and dock the phage particle over an entry port into the cytoplasm. Alternatively, the interaction between TolA and p3 D1 may locally disrupt the outer membrane structure to facilitate passage of viral material like DNA. The passage of DNA is probabily mediated by the formation of a channel by the p3 (Glaser-Wuttke, Keppner et al. 1989) alone or in conjunction with the inner membrane components of the TolQRA complex.

As the phage DNA is injected, the coat proteins are stripped from the DNA, which is immediately converted to a supercoiled double-stranded replicative form (RF) by bacterial enzymes. The RF is then replicated to form a pool of double-stranded DNA molecules. Host encoded enzymes and the viral gene 2 product are involved in this process. These molecules of double-stranded DNA serve as a template for transcription and translation from which all the phage proteins are synthesized. The production of phage proteins increases with the accumulation of these RF molecules. Capsid proteins, and other phage proteins involved in the assembly of the particle integrate into the cell envelope. Proteins involved in DNA replication remain in the cytoplasm. Late in infection when the phage specific single-stranded DNA binding protein p5, reaches the proper concentration, it switch from RF replication to single strand by binding to the newly formed viral strands. Inside the cell all the progeny of viral DNA strands become complexed with g5p. The sequestered DNA reaches the inner membrane where p5 is replaced by p8 as the DNA, capped with the minor coat proteins p7 and p9, is extruded through the periplasmic space and out of the cell (Russel 1991; Russel 1994). When the last part of the phage genome emerges from the cytoplasm, its protein coat is thought to be capped by a complex of p3 and p6. Hence p3 is the first viral protein to enter and the last to leave.

the assembly process is tolerated quite well by the host, as infected bacteria continue to grow and divide with a generation time approximately 50% longer than that for an uninfected bacteria. About 1000 phages particles are produced during the first generation following infection, after which the host produces approximately 100-200 phages per generation (Model and Russel 1988).

1.2.6 Filamentous phage as cloning vector

As a consequence of their structure and life cycle, the Ff phage have served as valuable tools for biological research. Since replication of DNA or assembly of the phage is not constrained by the size of the DNA, the Ff phage are excellent cloning vehicles (Smith 1988). The result of an insertion of foreign DNA into a not essential region merely results in a longer phage particles. Thus, large quantities of single-stranded DNA containing foreign DNA inserts ranging from few to thousands of nucleotides can be easily obtained. Filamentous phage were initially used to prepare single-stranded DNA for sequencing, as well as other manipulations, such as the creation of substrates to study DNA mismatch repair or reactions involved in recombination. By combining the features of plasmids and phage, new cloning vectors called phagemids have been constructed (Mead and Kemper 1988). They contain the replication origin and packaging signal of the filamentous phage together with the plasmid origin of replication and gene expression system. These vectors can be propagated as high copy number plasmids until super infection of the host with a helper phage such as M13K07, which provides all proteins required for packaging and export.

Several engineered strains of filamentous phage, including fd-TET (Zacher, Stock et al. 1980) have also been prepared, which confer antibiotic resistance on to the host cell. These strains can be propagated as plasmid in bacterial hosts growing on the appropriate antibiotic.

1.3 Immune system

1.3.1 Antibody structure

Five distinct classes of immunoglobulin molecule are recognized in most higher mammals, namely IgG, IgA, IgM, IgD, IgE, of which IgG is the most abundant in both mouse and human. They differ in size, charge, amino acid composition and carbohydrate content. The basic structure of all immunoglobulin molecules is a unit consisting of two identical light polypeptide chain and two identical heavy polypeptide chains, linked together by a combinations of disulphide bonds and non covalent interaction as shown in fig 1.3.

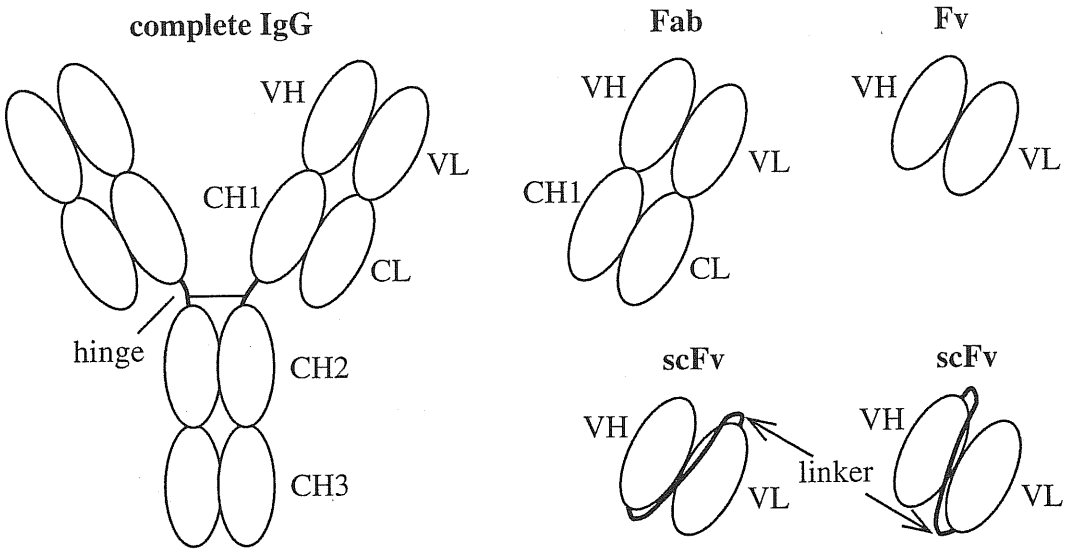


Fig 1.3 Structure of a complete Ig molecule and of the Fab, Fv and scFv fragments.

The plant protease papain cleaves the IgG molecule in the hinge region between the Cγ1 and Cγ2 domains, to give two identical Fab fragments and one Fc (crystallizable) fragment. These papain generated fragments have been of enormous value in structure/function studies on the antibody molecule because they separated the Fab region which binds to the antigen, from the Fc region, which mediates effector functions such as complement fixation and monocyte binding. The polypeptide chains of immunoglobulin are composed of domains of similar structure, each consisting of two stacked layers of beta sheets surrounding an internal space filled with hydrophobic aminoacid side chains, with terminal exposed loops. They are termed either constant (C) or variable (V) on the basis of the degree of sequence variation amongst different antibody molecule, which is focused on the three, hypervariable, exposed loops at the top of the variable domains.

The VH and VL domains consist of four regions of relatively conserved sequence (Kabat, Wu et al. 1987) called framework regions (FR1, FR2, FR3 and FR4) which form the scaffold for three regions of hypervariable sequences, these segments are located near amino acid position 30, 50 and 95. Because they create the antigen binding site, determining the shape of the binding site and so its specificity, these hypervariable regions are sometimes referred to as complementarity determining regions (CDRs). The CDRs, structurally form loops that vary in length as well in sequence. Analysis of the realationship between the sequence and the three dimensional structure of the antibody combining sites revealed that, except for VH CDR3, the other loops have a restricted number of main chain conformations or canonical structures. The canonical

structure formed in a particular loop is determined by its size and the presence of certain residues at key sites in the loop and in framework regions. The conformation of VH CDR3 shows some regularities, from which rules relating sequence to conformation can be stated (Tomlinson, Cox et al. 1995; Al-Lazikani, Lesk et al. 1997; Morea, Tramontano et al. 1998), but to a less complete degree than for the other 5 antigen binding loops.

1.3.2 V-gene maturation

B lymphocyte development has traditionally been divided into two phases. The first phase is thought to be independent of foreign antigen and terminates when Ig is expressed from rearranged H and L chains genes on the surface of immature B cells. In the second, antigen-dependent phase, the cells can be positively selected in the peripheral circulation and colonize the secondary organs. The enriched pool of binding clones is then driven to evolve antibodies of higher affinity for the antigen. Relatively low affinity antibodies displayed on memory B-cells are increased in affinity by somatic mutations of their antibody genes followed by antigen driven selection in a process termed affinity maturation (Milstein and Neuberger 1996)

Recombination of V genes starts in lymphoid progenitor cells, within the IgH locus of either the maternal or paternal chromosome 14. The recombinase proteins, RAG-1 and RAG-2 (Schatz, Oettinger et al. 1989) recognize and bind to conserved heptamer-spacer-nonamer recombination recognition sequences flanking all V gene segments. In B-lineage cells, rearrangements of one D gene to one J_H gene is followed by the addition of one of the numerous V_H genes (Tomlinson, Walter et al. 1992; Tomlinson, Cook et al. 1995) to the fused D-J_H segments. If successful, the resulting V_H-D-J_H rearrangement is transcribed as a single unit into RNA that is spliced to the constant region RNA (cμ) before translation into an Ig heavy (μ) chain in the pre-B cell. If the initial rearrangement yields a sequence that cannot be translated (non productive or abortive rearrangement) the rearrangement of the IgH locus proceeds on the other allele. The presence on the B-cell surface of a fully assembled μ heavy chain inhibits further rearrangements that might occur on the opposite allele (allelic exclusion).

Although light chain assembly may occur in the absence of a functional heavy chain, the process is greatly enhanced by successful heavy chain rearrangement. Light chain rearrangement begins when one of the 40 kappa light chains (Vκ) (Tomlinson, Cox et al. 1995) first rearranges to one

of the five joining Jκ genes. If kappa light chain rearrangement is unsuccessful on both alleles, one of the 30 lambda light chains (Vλ and Jλ) (Frippiat, Williams et al. 1995) will subsequently rearrange. This hierarchical sequence of k before λ may not be absolute as for example in fetal BM cells where λ light chains may occasionally rearrange first.

The mechanism leading to the assembly of the gene segments encoding the V regions of both immunoglobulin H and L chains is mostly elucidated: specific base sequences that act as joining signals (Seidman, Max et al. 1979; Max, Seidman et al. 1980) and two recombinases RAG1 and RAG2 (Schatz, Oettinger et al. 1989) responsible for the recombination process, have been identified. Joining is directed by specific recombination signal sequences (RSS), which flank the coding sequence of every germline gene segment. Each RSS consists of a conserved palindromic heptamer sequence separated from a conserved nonamer sequence by 12 or 23 bp spacer. Only segments with different sized spacers may be joined according to the 12/23 rule.

1.3.3 CDR3 diversity

The presence of multiple V (D) and J gene segments, their recombination and their inaccuracy of recombination contribute to the generation of diversity of an antibody repertoire.

The CDR3 is in direct contact with the antigen and is the most variable portion of the Ig molecule. Its extraordinary diversity- up to 10^{14} different peptides are possible (Sanz 1991; Yamada, Wasserman et al. 1991) results from several mechanism (fig 1.4).

The CDR3 encompasses the 3' end of V_H, all of D, and the 5' end of J_H. It often contain N nucleotides, which are randomly inserted at both the V_H-D and D-J_H junctions by the enzyme terminal deoxytransferase (TdT) (Alt and Baltimore 1982). The introduction of junctional nucleotides is developmentally regulated; N insertions are found in 68% of fetal B cells, 86% of neonatal B cells, and in 91% to 100% of mature adult B cells. Another source of diversity in CDR3 is random deletion by nucleases of the terminal nucleotides of rearranging V_H, D and J_H genes. Recent evidences (Corbett, Tomlinson et al. 1997) obtained by sequencing the D locus and by comparison of the complete sequence of this with a data base of rearranged sequences suggest that DIR segments, inverted D segments, minor D segments or D-D joins are not used in repertoire of natural antibodies. This analysis suggest that D segments can be used in all three reading frames

even if this use is highly biased. At junctional sites in which no nucleotide deletions occur there may be template "p" nucleotides. These sequences, usually only one or two bases long, are complementary copies of the last nucleotides of the coding region they abut. They may be used to repair asymmetrical breaks in the hairpin ends of V genes undergoing recombination. All of these variables contributes to the CDR3 antibody "fingerprint" even before somatic mutations that further distinguish and individualize V genes. They results in a CDR3 sequence that is unique to each rearrangement, and that therefore indentifies individual B cells or clonal B-cell expansions.

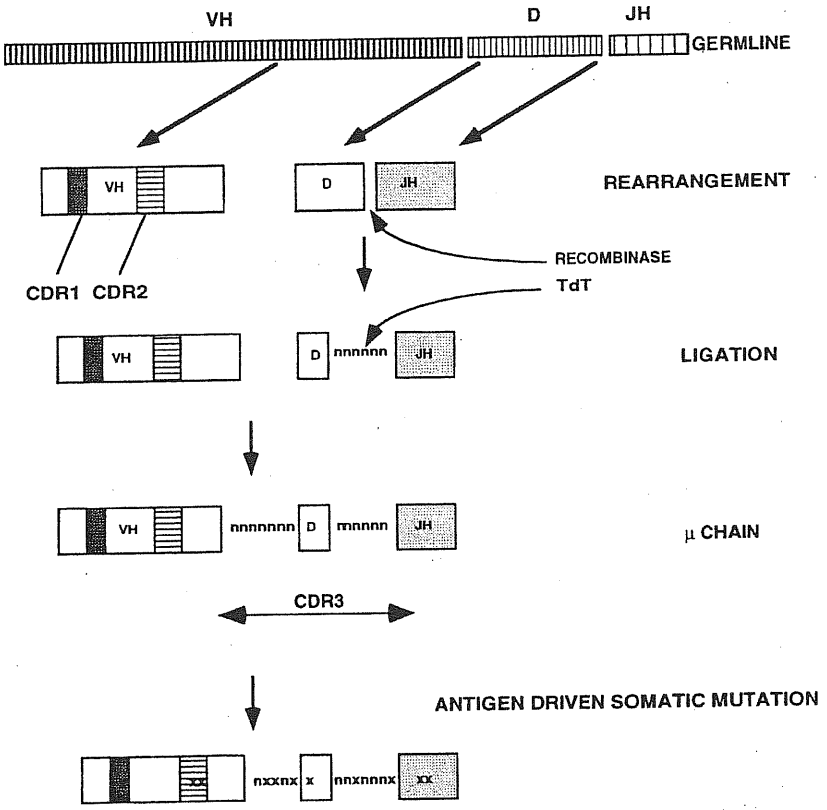


Fig. 1. 4 Formation of CDR3. During V gene rearrangement variable deletions of the end of V_H, D, and J_H, combined with non template nucleotide (n) insertions at V_H-D and D-J_H junctions contribute to the CDR3. During antibody maturation, somatic mutations (x) further diversify the CDR fingerprint. (Stewart and Schwartz 1994)

Interestingly analysis of large numbers of rearranged genes indicates that few germline segments dominate the mature repertoire (Stewart, Huang et al. 1993) (Cox, Tomlinson et al. 1994; Suzuki, Pfister et al. 1995) and that the sequence diversity encoded by these segments is focused on residues at the centre of the antigen binding site (Tomlinson, Walter et al. 1996). With

somatic hypermutation, diversity spreads to regions at the periphery of the binding site that are highly conserved in the primary repertoire. This complementarity has probably been selected by evolution as an efficent strategy for searching sequence space (Tomlinson, Walter et al. 1996; Ignatovich, Tomlinson et al. 1997)

1.4 Phage display technology

1.4.1 Phage display

The origin of phage display dates to the mid-80s when G.P. Smith first expressed a foreign segment of protein on the surface of bacteriophage M13 virus particle (Smith 1985). Smith demonstrated that phages displaying the antigen could be affinity purified against an immobilised cognate antibody, allowing more than 1000-fold enrichment of fusion phage from a background of phage particles displaying no antigen. In general if foreign DNA encoding a polypeptide A is cloned downstream of the gene 3 or gene 8 leader sequence, it will be translated and exposed at the N-terminus of the mature p3 or p8, such a polypeptide A can be expressed in such a way that it does not impede the normal function of either p8 or p3. If an antibody recognising A (anti-A) is now fixed to a solid support, phage displaying A can be selected from a background of billions of other phage displaying polypeptides which are unable to bind A

From these first experiments two important concepts emerged:

a) using recombinant DNA technology, it should be possible to build collection of billions of different phage displaying different polypeptides on their surface termed "phage library". This library provides a valuable resource, from which useful polypeptides can be isolated.

b) the methodology provides a direct physical link between phenotype and genotype. The fusion phage is, therefore, a self-replicating entity with its phenotype physically linked to its genotype. That is, every displayed molecule has an addressable tag via the DNA encoding that molecule. The fact that the gene encoding A is found within the phage displaying A, means that cloning of the gene for A occurs simultaneously with the selection of phage displaying A. This permits the continuation of the phage display cycle (fig 1.5) until a population which is homogenous in its binding properties is obtained (all bind anti-A). The cloning of the gene encoding A allows

identification of the protein A sequence and also permits the subsequent recombinant manipulation or expression of A for further experiments.

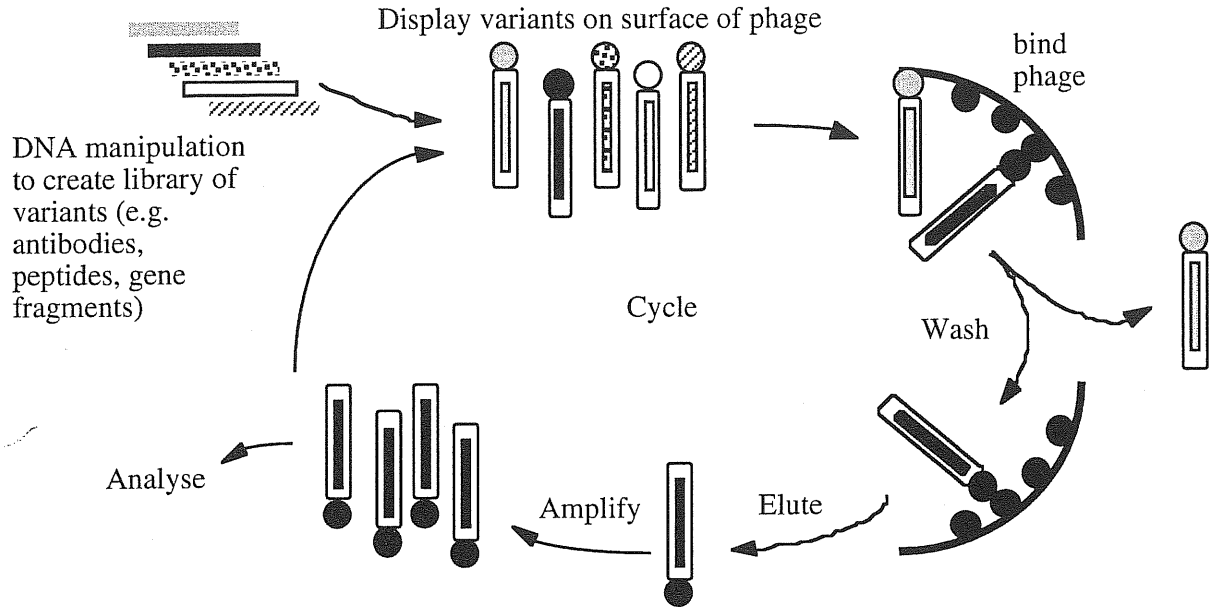


Fig 1. 5 The phage display cycle

The phage display cycle starts with the creation of diversity at the DNA level. This is translated into phenotypic diversity by the display of different polypeptides on the surface of the phage. The application of selection pressure, manifested by cycles of binding to a ligand (here depicted in solid phase), washing and elution allow the selection of phage displaying polypeptides which bind to the ligand

1.4.2 Vector systems used for phage display

Three broad classes of vector systems used for phage display have been described. In the first, the type 3 or type 8 vectors, phage DNA is directly modified and all phage proteins should be recombinant. Such vectors are termed polyvalent for this reason. While p3 display has been used for many peptide libraries and a few proteins, display on p8 in this system is limited to peptides smaller than six amino acids, as longer ones cause a severe display bias (Iannolo, Minenkova et al. 1995). It has proved difficult to work with phage vectors, because of the low transfection rates and instability of the phage genome to inserts: there is a tendency for the phage to delete extraneous DNA. For this reason the 3+3 vectors are generally preferred for protein display. These are phagemid vectors which contain an Ff origin of replication (which ensures packaging into viral particles) as well as a copy of p3 which is used to display the recombinant polypeptide. These

vectors can be propagated as plasmids. When display is required, bacteria containing the phagemid are infected with helper phage, which provide all the necessary phage proteins for the creation of selectable phagemid particles. The origin of replication in the phagemid is packaged in preference to that of the helper phage and a proportion (1-10%) (Clackson and Wells 1994) of the phage particles will display recombinant p3. Of these, most will display single copies, so-called monovalent display. The remaining p3 molecules are derived from the helper phage. As a result such vector systems are both phenotypically (wild type and recombinant p3 is present) and genotypically (both helper and phagemid genomes are present) heterogenous. A similar 8+8 system has been used for the display of some peptide libraries.

The 33 and 88 systems are intermediate, being phage based and containing both wild type g3 or g8 in association with recombinant g3 or g8 cloned within the phage genome. In practice only the 88 system has been used for some peptide libraries (Bonnycastle, Mehroke et al. 1996) phage produced using this system are phenotypically heterogenous, but genotypically homogenous. They suffer from some of the same problems as the phage vectors and have not yet been used for protein display.

1.4.3 Phage display applications.

One of the most impressive aspects of phage display is the variety of uses for the technology. A few of the many applications includes:

a)Phage display of natural peptides.

- mapping epitopes of monoclonal, polyclonal antibodies and patient sera
- generating immunogens

b) Phage display of random peptides

- mapping epitopes of monoclonal and polyclonal antibodies
- identify peptide ligands for proteins involved in interaction
- mapping substrates sites for proteases and kinases

c) Phage display of proteins and protein domains

- direct evolutions of proteins
- isolation of high affinity antibodies
- cDNA expression screening

Protein displayed	**vector type**	**Reference**
cDNA libraries (i.e. random clones)	pJuFo	(Crameri and Suter, 1993; Crameri et al., 1994)

genomic libraries (Staphylococcus and Streptococcus genomes)	3+3, 8+8	(Jacobsson and Frykberg, 1995; Jacobsson and Frykberg, 1996; Jacobsson and Frykberg,1998)
Protein fragments		
ß galactosidase	3	(Parmley and Smith, 1988)
bluetongue virus VP5 and NS1 proteins	3	(Du Plessis et al., 1995; Wang et al., 1995)
p53	3	(Petersen et al., 1995)
RNA polymerase II	3	(Petersen et al., 1995)
cytokeratin 19	3	(Petersen et al., 1995)
PM/Scl	3	(Bluthner et al., 1996)
ZAG and MAG from streptococci	8+8	(Jacobsson et al., 1997)
myb	3+3	(Kiewitz and Wolfes, 1997)
PAI-1	3+3	(van Zonneveld et al., 1995; van Meijer M et al., 1996)
Small constrained peptide domains		
hybrid rop protein constrained peptide library	3	(Santiago Vispo et al., 1993)
cytochrome b562		(Ku and Schultz, 1995)
Tendamistat constrained peptide library	33	(McConnell and Hoess, 1995)
Protein A (E, D, A and B domains or B domain alone)	3+3	(Djojonegoro et al., 1994; Kushwaha et al., 1994; Nord et al., 1995; Nord et al., 1997)
Knottins	3+3	(Smith et al., 1998)
Protease inhibitors		
Alzheimer's amyloid ß-protein precursor Kunitz domain	3+3	(Dennis and Lazarus, 1994; Dennis and Lazarus, 1994; Dennis et al., 1995)
human plasminogen-activator inhibitor 1 (PAI-1)	3+3	(Pannekoek et al., 1993)
Kunitz domain libraries, BPTI	3, 88	(Markland et al., 1991; Roberts et al., 1992; Roberts et al., 1992)
Ecotin	3+3	(Wang et al., 1995)
cystatin	3+3	(Tanaka et al., 1995)
Proteases		
Prostate specific antigen	3+3	(Eerola et al., 1994)
Trypsin	3+3, 33, 8+8, 88	(Corey et al., 1993; Wang et al., 1996)
Enzymes		
glutathione transferase A1-1	3+3	(Widersten and Mannervik, 1995; Hansson et al., 1997)
Staphylococcal nuclease	3, 3+3	(Ku and Schultz, 1994; Light and Lerner,1995)
Alkaline phosphatase	pJuFo, 3, 3+3	(McCafferty et al., 1991; Crameri and Suter, 1993; Maenaka et al., 1996)
ß lactamase	3	(Soumillion et al., 1994; Vanwetswinkel et al., 1995)
lysozyme	3+3	(Maenaka et al., 1996)
Inactive phospholipase A2	pJuFo	(Crameri and Suter, 1993)
Cell surface receptor fragments		
CD4, e.c. and individual domains	3, 3+3	(Chiswell and McCafferty, 1992; Abrol et al., 1994; Krykbaev et al., 1997)

FceR1 a-chain, e.c. and individual domains	3, 3+3	(Robertson, 1993; Scarselli et al., 1993)
PDGF receptor, e.c domain	3	(Chiswell and McCafferty, 1992)
Hormones, interleukins and bioactive peptides		
Human growth hormone	3+3	(Bass et al., 1990; Lowman and Wells, 1993)
C5a	pJuFo	(Hennecke et al., 1997)
Thymosin ß4		(Rossenu et al., 1997)
Transforming growth factor alpha	3	(Tang et al., 1997)
Tumour necrosis factor	3+3	(Clackson and Wells, 1994)
Interleukin-2 (IL-2)	3+3	(Buchli et al., 1997; Vispo et al., 1997)
Interleukin-3 (IL-3)	3, 3+3	(Gram et al., 1993; Merlin et al., 1997)
Interleukin-6 (IL-6)	3+3	(Cabibbo et al., 1995)
Interleukin-8 (IL-8)	3+3	(Clackson and Wells, 1994)
Insulin like growth factor (IGF) binding protein	3+3	(Lucic et al., 1998)
Heregulin beta domain	3+3	(Ballinger et al., 1998)
Epidermal growth factor (EGF)	3	(Souriau et al., 1997)
Ciliary neurotrophic factor	3+3	(Saggio et al., 1995)
Atrial natriuretic peptide and derivatives	3+3	(Cunningham et al., 1994; Li et al., 1995 Jin et al., 1996)
Bone morphogenetic protein (2A)		(Liu et al., 1996)
Antibodies and derivatives		
Minibody (61aa, 4 beta sheets)	3, 3+3	(Pessi et al., 1993; Martin et al., 1996)
FAb	3, 3+3, 8+8	(Hoogenboom et al., 1991; Kang et al., 1991; Huse et al., 1992; Orum et al., 1993; Geoffroy et al., 1994; Griffiths et al., 1994)
ScFv	3, 3+3, 8+8	(McCafferty et al., 1990; Clackson et al., 1991; Marks et al., 1991; Marks et al., 1992; Griffiths et al., 1993; de Kruif et al., 1995; Kretzschmar and Geiser, 1995; Vaughan et al., 1996)
CH3	3+3	(Atwell et al., 1997)
VH domains	3+3	(Davies and Riechmann, 1996)
Dromedary heavy chain antibodies	3+3	(Lauwereys et al., 1998)
Toxins		
Ricin B-chain, complete or domain 2	3	(Swimmer et al., 1992; Lehar et al., 1994)
Fungal ribotoxin reAsp f I/a	3+3	(Crameri and Suter, 1993)
B. thuringiensis CryIA(a) toxin	3+3	(Marzari et al., 1997)
Nucleic acid binding proteins		
HIV tat	3+3	(Hoffmann and Willbold, 1997)
U1A protein	3+3	(Laird-Offringa and Belasco, 1996)

Zinc finger proteins, domains and libraries	3, 3+3	(Choo and Klug, 1994; Choo and Klug, 1994; Jamieson et al., 1994; Rebar and Pabo, 1994; Choo and Klug, 1995; Wu et al., 1995)
Miscellaneous		
T cell receptor alpha chain	3	(Onda et al., 1995)
peptostreptococcal protein L		(Gu et al., 1995)
Pseudomonas aeruginosa protein F	3	(Kermani et al., 1995)

Tab. 1.1 Proteins which have been displayed on phage are indicated as well as the vector system used

Some of the applications listed above will be described in the following paragraphs

1.4.4 Epitope mapping using peptide display libraries

Peptides were the first foreign sequences to be displayed in p3 (Smith 1985), and initial libraries displayed linear peptides on the phage surface. More recently, constrained libraries (Luzzago, Felici et al. 1993; Hoess, Mack et al. 1994) have been generated in which a random peptide sequence is inserted, generally between fixed cysteine residues. It is argued that constrained peptides may better approximate the structure recognized in the native protein.

One of the motivation behind Smith's original experiments was to use phage display as a means of identifying clones that were expressing epitopes recognized by a given antibody. Traditionally, epitope mapping of protein antigens has relied heavily on physical chemical analysis, all methods that are labor intensive and in general are not amenable to high-throughput analysis. As an alternative to this methods, phage display can be use to localize the antigenic epitope relatively quickly. This has been done successfully for several antibodies (Parmley and Smith 1988; Parmley and Smith 1989; Cwirla, Peters et al. 1990; Stephen and Lane 1992; Balass, Heldman et al. 1993; Jellis, Cradick et al. 1993; Stephen, Helminen et al. 1995) that recognise myohemerythrin, p53 and HIV. A discontinuous epitope on human ferritin has been mapped using information derived from peptide selected from a phage library but this did require detailed knowledge of the crystal structure of ferritin (Luzzago, Felici et al. 1993). In another case, a discontinuous epitope on *Bordetella pertussis* toxin could not be mapped using the peptide resulting from a selection of a library with monoclonal

antibody (Felici, Luzzago et al. 1993). No consensus sequence were identified and, furthermore, mice immunized with the peptides failed to produce antibodies against the toxin.

Usually, one or more different motifs are obtained, although this is not always the case. Such motifs may be similar to a sequence found within the original protein, if the identity of the target protein is unknown, screening the sequences of selected peptides against protein databases may allow the identification of the protein recognised, although the degeneracy of the motif usually obtained can make this difficult, and experimental confirmation is always required. When the motif has no homologue within the starting protein, it is termed a 'mimotope', and is thought to mimic a conformational or discontinuous epitope. This probably occurs by displaying the epitopic amino acids found close to one another on the surface of the protein (but distant in primary sequence), in a linear fashion, such that they mimic the spatial arrangement found on the surface of the target, e.g. ferritin.

Peptide libraries have also been used to identify antigenic epitopes present in the sera of patients suffering from infectious diseases. Although similar conceptually to the use of monoclonal antibodies, it is practically more difficult, as individual patient sera contain many antibodies able to bind epitopes which have no relevance to the disease under study. For this reason selection on sera from a number of different patients with the same disease is followed by depletion or negative screening using sera from normal or diseased controls, this leading to the identification of disease 'phagotopes' as they have termed (Cortese, Felici et al. 1994). Such phagotopes can often induce resistance to infectious disease agents when used as immunogens (Meola, Delmastro et al. 1995), as well as providing important information on the immune response.

1.4.5 Other application of peptide libraries

Random peptide libraries have been used with selector molecules other than antibodies to determine binding site of these proteins. These include biotin (Saggio and Laufer 1993), pancreatic RNAse (Smith, Schultz et al. 1993), calmodulin, the endoplasmic chaperone BiP (Blond-Elguindi, Cwiria et al. 1993), the src SH3 domain (Sparks, Quilliam et al. 1994) and many more. In all cases, peptide mimics with predicted binding properties were generated. It is often difficult to obtain pure protein preparations with which to do peptide phage selections, especially proteins which maintain their native structure. In some cases this has been overcome by expressing receptors

(e.g. eythropoietin or thymopoietin (Livnah, Stura et al. 1996; Cwirla, Balasubramanian et al. 1997)) on the surface of cells and using them for selection. An alternative is to express proteins on the surface of bacteria, as has been done for ß adrenergic receptors (Marullo, Delavier-Klutchko et al. 1989) and to use these as living columns to purify binding phage (Bradbury, Persic et al. 1993).

The library approach can also be used to identify and characterize protease substrates. A random peptide library is inserted between a binding domain (human growth hormone was used (hGH)) and p3. Phage captured on an affinity support (hGH receptor) are then treated with a protease, cleavage of the peptide sequence result in phage release. Phage containing sequence sensitive to cleavage can be amplified and used for further rounds of selection for the identification of a optimal peptide substrate. Alternatively, resistant peptide can be identified by elution of phages still bound after protease treatment. This approach has identified substrate sequence sensitive to subtilisin, factor Xa (Matthews and Wells 1993) and furin (Matthews, Goodman et al. 1994).

1.4.6 Displaying other proteins

The permissiveness of p3 to the display of foreign proteins is remarkable (see table 1.1). Over fifty different proteins have been cited in the literature, and these include secreted proteins, the extracellular domains of membrane proteins, intracellular proteins, DNA binding proteins, enzymes, toxins and a number of small scaffolds derived from many sources. Any protein which can be secreted into the periplasmic space of E. coli can probably be displayed on phage. The main reason to display a protein on phage is to select mutants of that protein which have different properties, usually an increase in affinity or biological activity, although decreases in affinity which allow the identification of interaction residues, can also be selected for (Jespers, Jenne et al. 1997). This process requires a number of different steps: the creation of the library of mutants from which selection occurs, selection and screening. Targeted mutations have been used most frequently and with the greatest success: the selection of a BPTI derivative which binds human neutrophil elastase with an affinity $3.6x10^6$ times greater (final affinity 1pM) than the wild type (from a library in which five residues were mutated) being the most striking example (Roberts, Markland et al. 1992). Similarly, an ecotin variant with an affinity for urokinase-type plasminogen activator 2800 higher than the wild type was

selected from a targeted library in which only two residues were mutated (Wang, Yang et al. 1995).

1.4.7 Phage display of antibody

The use of phage display to select antibodies recognising specific antigens is arguably the most successful use of phage display. Correctly folded antibody fragments, Fv, scFv (single-chain Fv) or Fab are expressed by routing the nascent antibody chain(s) to the periplasm of the bacterium (Better, Chang et al. 1988; Skerra and Pluckthun 1988) where the intradomain disulphide bridge between the two β-sheets is formed and the VH and VL will pair. Periplasmic expression in the bacterium resembles the natural production route in the endoplasmic reticulum (ER) of the lymphocytes.

Antibody phage display is accomplished by fusing the coding sequence of the antibody variable (V) region to the amino terminus of the phage protein p3. When antibody V-genes are cloned into phage display vectors, based on either phage or phagemid, functional antibody fragments are expressed on the surface of infective particles while the encoding genes reside within the phage particles. Two approaches have been successfully demonstrated: single chain Fv fragments fused to the amino-terminus of p3 (McCafferty, Griffiths et al. 1990) and heterodimeric Fab fragments (Hoogenboom, Griffiths et al. 1991). Heterodimeric Fab fragments can be assembled on the surface of phage by linking one chain to the phage coat protein and secreting the other into the bacterial periplasm, where the two chains associate.

The linkage between antibody genotype and phenotype allows the enrichment of antigen specific phage antibodies, using immobilized or labelled antigen. Phage that display a relevant antibody will be retained on a surface coated with antigen while non-adherent phages will be washed away, bound phages can be recovered from the surface, reinfected into bacteria and regrown for further enrichment and eventually for binding analysis.

1.4.8 Phage antibody libraries

With phage display, antibodies can be made completely *in vitro*, by-passing the immune system and the immunization procedure, allowing *in vitro* "tailoring" of the affinity and specificity of the antibody. The concepts

behind phage antibodies library selection are identical to those used for any of the other biological molecular diversity techniques: the creation of diversity, followed by a series of recursive cycles of selection on antigen, each of which involves binding, washing, elution and amplification and finally analysis of selected clones.

There are two main routes to antibodies libraries via phage display

The first is to select antibodies from libraries prepared from immune donors e.g. immunized animal or, in some instances, human immune B-cells. Immune libraries have two main advantages: firstly they are highly biased towards V-genes that encode antibodies against the immunogen (especially if IgG specific primers are used), which means that relatively small (10^5 clones) libraries can be successfully screened; and secondly, many of the genes will encode both affinity matured antibodies (increasing the number of high affinity antibodies in the library) and highly specific antibodies. This method gives access to more and sometimes better antibodies than working with hybridomas. For example, from an immune murine phage antibody library, Chester et al. (Chester, Begent et al. 1994) identified an anti-Carcinoembryonic antigen (CEA) antibody with an affinity that was higher than ever obtained with hybridoma technology; Andersen et al. (Andersen, Stryhn et al. 1996) select a Fab specifically recognizing an MHC/peptide complex that was pre-determined and used for the immunization.

Disadvantages of this strategy, however, include the time required to immunize animals, the unpredictability of immune response to the antigen of interest and the lack of immune response to some antigens (self antigens or toxic molecules). Moreover the scrambling of heavy and light chains which occours in library construction makes it very difficult to discriminate "original" VH-VL combination from de novo *in vitro* formed pairs. The likelihood that original pairs may be retrieved is very low, although the success of the method indicates this may not be important if one is interested only in obtaining binding antibodies. In particular, most of the binding energy and specificity is provided by the V_H domain, rather than the V_L domain, as shown by a number of co-crystallographic studies (reviewed in (Davies and Cohen 1996)) in which both the number of V_H residues, as well as the area of the V_H domain, in contact with antigen, is almost always greater than the VL domain. Furthermore, in selection from very large naïve phage antibody libraries (Vaughan, Williams et al. 1996), different antigens always selected different V_H genes, whereas some VL genes were found in scFvs binding to different antigens.

Phage display has been used extensively to study the characteristics of self antibodies in autoimmune diseases. This work has been most extensively carried out with thyroid disease (Portolano, McLachlan et al. 1993; Hexham, Partridge et al. 1994; Portolano, Prummel et al. 1995; McIntosh, Asghar et al. 1996; McIntosh, Asghar et al. 1997; McIntosh, Asghar et al. 1997), with the antibodies selected from such libraries having similar specificities to those found in patients' serum and with the high affinities characteristic of immune libraries. Similar experiments have been done with systemic lupus erythematosis (Barbas, Ditzel et al. 1995; Marchbank and Deutscher 1995; Roben, Barbas et al. 1996), Sjogren's syndrome (Suzuki, Takemura et al. 1997), paraneoplastic encephalomyelitis (Graus, Verschuuren et al. 1998) and myasthenia gravis (MG) (Graus, de Baets et al. 1997; Graus, de Baets et al. 1997).

The second route to antibody libraries begins with naive or semi-synthetic libraries. In both cases, the ultimate goal is to assemble libraries of such enormous size and diversity that high affinity antibodies of any specificity can be selected (Marks, Hoogenboom et al. 1991; Griffiths, Williams et al. 1994; Vaughan, Williams et al. 1996; Sheets, Amersdorfer et al. 1998).

The primary immune response involves a large array of IgM antibodies that recognize a variety of antigens. The murine naive repertoire has been estimated to contain $<5x10^8$ different B-lymphocytes while the human repertoire may be 100 to 1000 times bigger (Winter, Griffiths et al. 1994). This array of antibodies may be cloned as a naive repertoire of rearranged genes, by harvesting the V-genes from the IgM mRNA of B-cells of unimmunized human donors, isolated from peripheral blood lymphocytes (PBLs), bone marrow, spleen cells, or from animal sources. The VH genes can be amplified either from IgM mRNA (Marks, Hoogenboom et al. 1991) or total mRNA (Vaughan, Williams et al. 1996). Libraries made from IgG mRNA perform very poorly by comparison as the V-genes cloned are biased to previous antigenic exposure in the donor and are therefore, lacking in diversity. In fact, in an early paper on phage antibody libraries (Marks, Hoogenboom et al. 1991), two libraries were made, one using IgM primers and another with IgG primers. Interestingly only the IgM derived library yielded binders to the antigens used for selection.

Providing the repertoire is diverse, the greatest advance of this approach is: one library can be used for all antigens; high affinity human antibodies can be isolated; antibodies to self, non-immunogenic or toxic

substances can be generated; and antibody generation takes two-four rounds of selection (two weeks).
The first of such "single pot" repertoires was made from the PBLs of two healthy human volunteers and contained $3x10^7$ clones. From this pool antibodies to over 25 different antigens were isolate. The antigens were either foreign, such as bovine serum albumin, and hapten, or self, such as thyroglobulin, CD4 etc. Recently, much larger scFv repertoire have been made by "brute force" cloning (Vaughan, Williams et al. 1996) yielding $1.4x10^{10}$ independent antibody clones. From this library, antibodies with affinities typical for a secondary immune response were isolated (average K_a of around 10^8 M^{-1}. Thus, neither immunization nor affinity maturation are needed for the generation of high affinity (human) antibodies.

However there are also some disadvantages in particular the largely unknown and uncontrollable contents of these naive libraries, and the lower affinity of the antibodies where smaller-sized repertoire are used.

Library performance is demonstrably improved by increased size and diversity; however, other factors including expression levels, folding, and toxicity to *E. coli* may all reduce the functional repertoire size. The advantages of synthetic V-gene repertoires is that they can be designed and tuned to optimize library function. To construct a synthetic antibody library, V-genes are assembled by introducing a predetermined level of randomization in the CDRs (or possibly also bordering FR regions) into, ideally germline V-genes segments (Hoogenboom and Winter 1992). Several synthetic libraries have been designed with the heavy chain CDR3 sequence encoded by oligonucleotide primers encoding a stretch of randomised amino acids residues (Griffiths, Williams et al. 1994; Nissim, Hoogenboom et al. 1994; de Kruif, Boel et al. 1995). In fact, from structural studies, it has become apparent that five of the six CDR regions (all but the CDR3 of VH) have limited structural variation, and frequently follow a certain canonical fold (see par 1.3.3). Thus the CDR3 of VH is the most diverse loop, in composition length and structure, and this is the region that should be partially or completely randomized using oligonucleotide-directed mutagenesis or PCR-based methods.

Antibodies with affinities comparable to those obtained using traditional hybridoma technology can be selected from large naive antibody libraries, and the affinity of these can be further increased, to levels

unattainable in the immune system, by using the selected antibodies as the basis for subsequent libraries and selection.
Although a number of different phage(mid) antibody libraries have been published, the number which are 'naive' is relatively small (see table 1.2)

In general, the affinity of the antibodies selected is proportional to the size of the library, with Kds ranging from $10^{-6/7}$ for the smaller libraries (Marks, Hoogenboom et al. 1991; Nissim, Hoogenboom et al. 1994) to 10^{-9} for the larger ones (Griffiths, Williams et al. 1994; Vaughan, Williams et al. 1996; Sheets, Amersdorfer et al. 1998), a finding which is in line with theoretical considerations (Perelson and Oster 1979). Antibodies selected from immunized libraries tend to have higher affinities for the antigen used for immunization from an equivalent library size.

Table 1.2: Published antibody phage libraries

library	theoretical diversity	construction	V gene source	antibody form	vector	best affinity
Marks 91	$3x10^7$	natural naive PCR μVH assembled with PCR Vλ &Vκ	human PBL	scFv	phagemid	10^6-10^7
Hoogenboom 92	$2x10^7$	synthetic 49 VH, CD3 5 or 8, 1 JH 1 Vλ	synthetic human V genes	scFv	phagemid	$7x10^6$
Orum 93	$5x10^6$	natural immunised (factor VII) PCR fd, PCR VL, assemble	mouse spleen	Fab	phagemid	10^8-10^9
Barbas 93	10^7	natural immunised (HIV) PCR fd, PCR Vκ, assemble	human BM	Fab	phagemid	?
Williamson 93	$3x10^6$	natural immunised PCR fd, PCR Vκ, assemble	human BM	Fab	phagemid	?
Griffiths 94	$6x10^{10}$	synthetic 49 VH, CDR3 4-12, 1 JH 26 Vκ CDR3 8-10 (1-3) 21 Vλ CDR3 8-13 (0-5)	synthetic human V genes	Fab	phage	$3.8x10^9$
Nissim 94	$3x10^8$	synthetic 49 VH, CDR3 4-12, 1 JH 1 Vλ	synthetic human V genes	scFv	phagemid	10^6-10^7
Kruif 95	$3x10^8$	synthetic 49 VH, CDR3 6-15, 1 JH CDR3 conservatively randomised for CDR >9 4 Vκ 3Vλ	synthetic human V genes	scFv	phagemid	10^7
De haart 99	$3.7x10^{10}$	Natural Ig specific primers, PCR V regions, clone VH and VL repertoires and assemble	human spleen PBL	Fab	phagemid	$2.7x10^9$
Sheets 98	$7x10^9$	Natural Ig specific primers, PCR V regions, clone VH and VL repertoires and assemble	human spleen PBL	scFv	phagemid	$7x10^9$
Vaughn 96	$1.4x10^{10}$	natural PCR VH (from hexamer primed, therefore all isotypes) cloned (10^8) PCR Vκ and Vλ cloned with upstream GS linker (10^7). Both reamplified and PCR assembled.	human PBL-15, BM-24, tonsil-4	scFv	phagemid	$3x10^{10}$

1.5 Other display technologies

1.5.1 T4 phage display

Phage T4 offers some unique features that can be exploited for display of peptides on the capsid surface and for their potential use as multicomponent vaccines. The exterior of bacteriophage T4 capsid is coated with two outer capsid proteins *Hoc* and *Soc* at symmetrical positions on the icosaedron (160 copies of Hoc and 960 copies of Soc per capsid particle). Both these proteins are nonessential for phage infectivity and viability and assemble onto the capsid surface after completion of capsid assembly, elimination by mutation of one or both proteins does not affect phage productivity, viability, or infectivity. Apparently, these proteins provide additional stability to T4 phage under adverse conditions such as exterme pH or osmotic shock. Recently (Jiang, Abu-Shilbayeh et al. 1997) these proteins have been used to develop a phage display system which allowed in-frame fusion of foreign DNA at a unique cloning site in the 5' end of *hoc* and *soc*. A DNA fragment corresponding to the 36-aminoacid PorA peptide from *Neisseria meningitidis* was cloned into the display vector to generate fusions at the N-terminus of *Hoc* and *Soc*. The porA-Hoc and PorA-Soc fusion proteins retained their ability to bind to the capsid surface, and bound peptide was displayed in an accessible form as shown by its reactivity with specific monoclonal antibodies in ELISA. Furthermore the T4 phage system could has also been used to display polypeptides on the C-terminus of *Soc* (Ren, Lewis et al. 1996).

1.5.2 λ phage display

The λ phage has also been used to display peptides and proteins fused to both the bacteriophage tail protein V (Maruyama, Maruyama et al. 1994; Dunn 1996; Kuwabara, Maruyama et al. 1997) and to the capsid protein D (Sternberg and Hoess 1995; Mikawa, Maruyama et al. 1996). The lambda D protein is a particularly attractive system for display of cDNA expression libraries, since a variety of proteins or proteins domains large as β-galactosidase have been successfully displayed as fusion to its N or C termini. Recently a cDNA expression library expressed on the lambda surface as fusion to the viral capsid protein D was constructed (Santini, Brennan et al. 1998) and the performance was compared with two libraries

made as N-terminal fusions to III and VIII capsid proteins of M13. The results demonstrate the great potential of the lambda display system for constructing complex cDNA libraries for natural ligand discovery.

1.5.3 Bacterial display

Besides filamentous phage, the bacterial cell itself can also be used as a display vehicle. The display of foreign epitopes on the surface of bacteria has been carried out on both gram positive (Hansson, Stahl et al. 1992; Medaglini, Pozzi et al. 1995; Samuelson, Hansson et al. 1995) and gram negative bacteria (Hofnung 1991; Georgiou, Poetschke et al. 1993; Georgiou, Stathopoulos et al. 1997). Foreign epitopes and/or proteins can be exposed on the surface of the bacterial cell using outer membrane bacterial proteins or other cellular appendages. To be displayed on the bacterial surface, the cloned epitope/protein have to cross the inner as well as the outer membrane and this is possible if they contain a secretory leader and if they have a structure compatible with translocation. Some of the surface exposed regions loops of the outer membrane proteins are permissive for the display of epitopes.

LamB was the first prototype protein of this type developed as a vector for cell surface display. In bacterial membranes it exibits two biological activities: it is involved in the entry of maltose and it serves as a surface receptor for λ phages. The site LamB153 identified by (Charbit, Boulain et al. 1986) can accepts inserts of a wide variety of sequences up to a size of 55-60 aminoacid residues without loss of lamB functions. In this form of bacterial display the inserted epitope/protein has a constrained structure as its ends are embedded within the carrier protein, and the display is limited by insert size. An *E. coli* libray displaying random sequences on the bacterium surface has been constructed (Brown 1992) by inserting randomized oligonucleotide (11 codons in length) in the permissive site of lamB. This library has been used to identify engineered proteins that confer the bacterium the ability to adhere to iron oxide and not to adhere to other metal oxides.

Foreign epitopes has also been displayed using a chimeric protein, this being composed of a C-terminal OmpA fragment and the signal sequence of the lipoprotein Lpp. Lpp directs the chimeric proteins to the outer membrane and OmpA ensures that the inserted peptide faces the external side of the membrane (Francisco, Earhart et al. 1992; Francisco, Campbell et al. 1993; Francisco, Stathopoulos et al. 1993). In this display system the exposed

peptide is not constrained, its N-terminal end is fused to the OmpA domain and its C-terminal part is free in the extracellular medium.

1.5.4 Ribosome display

A number of evolutionary methods are currently being developed that can accelerate natural evolution of biological macromolecules to a matter of days. All of them have to fulfill two basic requirements: to couple genotype and phenotype for selection and to introduce diversification between selection rounds. Nucleic acids, where the molecules are simultaneously genotype and phenotype, have been evolved and selected for physical properties or for binding to target molecules (Irvine, Tuerk et al. 1991). In contrast, most of the methods used of proteins as a carrier of the phenotype have been based on living cells directly or indirectly by producing phages or viruses.

In vivo approaches to the evolutionary methods are limited by transformation efficiency, and the repeated construction of libraries with more than 10^9 to 10^{10} indipendent members is quite laborious. This limitation can be overcame by using *in vitro* systems based on cell-free translation. Displaying single chain antibodies on procaryotic (Hanes and Plückthun 1997) and eucaryotic (He and Taussig 1997) ribosomes has recently been demonstrated to be a promising cell-free alternative to the *in vivo* selection techniques. This novel methodology was realized due to the following experimental facts: 1) *in vitro* translation of mRNA lacking a stop codon results in a stable complex of the ribosome with both nascent polypeptide and its encoding message and 2) the nascent polypeptide can attain its functionally active conformation in the ribosome-bound state.

The ability of the ribosome display method to provide antibody selection has been shown initially with model systems containing some amount of a target single chain antibody diluted with an overwhelming excess of competing antibody. 10^4-10^5 fold enrichment of a specific combining site was obtained in a single cycle, with further enrichment in subsequent cycles. This is equal or greater than that reported for phage display. Recently the construction of murine antibody libraries from spleen tissue of non immunized (Makeyev, Kolb et al. 1999) mice has been reported. Library from immunized (Hanes, Jermutus et al. 1998) mice with a variant of the GCN4 zipper has also been constructed. It has been shown that, from this library, several scFvs that bind the antigen could be isolated. Because the proteins does not have to be eluted from the ligand, as the RNA

can be easily isolated from the bound ribosome complexes, the binding of the antibody to the target is not an upper limit to the affinity that can be selected. Another advantage of ribosome display is that libraries containing 10^{12} or more independent functional members can be prepared easily and repeatedly after each randomization step and used for affinity selection. In ribosome display, proteins can also mutate during selection when DNA polymerase without proof reading function are used in PCR.

1.5.5 Puromycin bonding

Another example of an *in vitro* approach to link genotype and displayed phenotype has been used by (Nemoto, Miyamoto-Sato et al. 1997). In this study an *in vitro virus* was constructed in which a genotype molecule (mRNA) binds to the phenotype molecule (protein) through puromycin in a cell-free translation system and will be amplified through RT-PCR. The life cycle of *in vitro virus* is a test tube version of that of retrovirus. The protein can be regarded as a coat protein of the virus, *in vitro* virus parasite to a cell-free test tube.

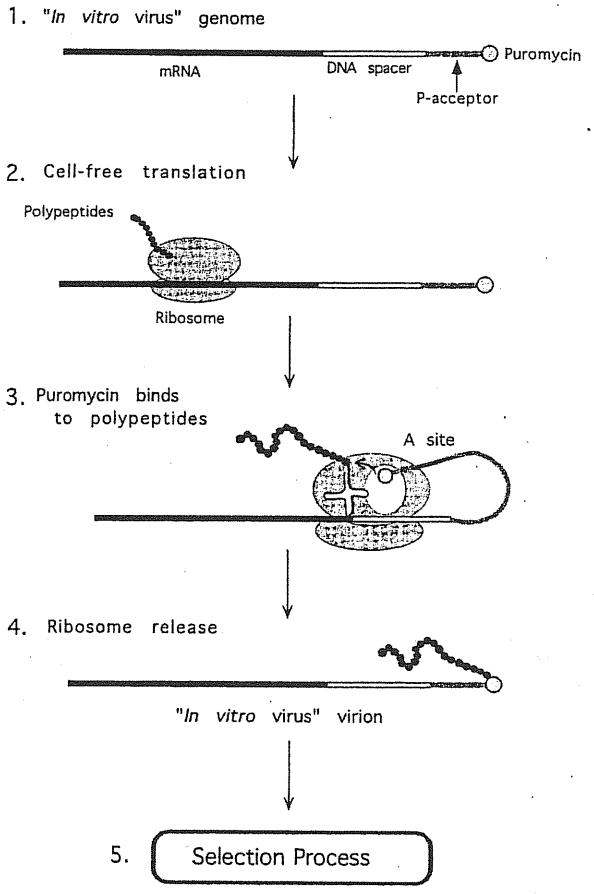


Fig 1.6 Scheme for the construction of the *in vitro* virus virion (Nemoto, Miyamoto-Sato et al. 1997).

Puromycin is a very powerful inhibitor of the growth of a cell, through its hindrance of chain elongation. Its structure resemble the 3' end of an aminoacyl-tRNA molecule, and thus is readly capable of entering the ribosomal A site to be transferred to nascent polypeptide chains by peptidyl transferase. Covalent bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal of the encoded protein in a cell-free translation system using rabbit reticulocyte lysates occurs and this forms the bases for coupling of phenotype and genotype (see fig 1.6). In (Nemoto, Miyamoto-Sato et al. 1997) the N-terminal fragment of human tau protein was displayed and bonding efficiency was 10% indicating that a population of *in vitro* viruses could achieve a diversity of 10^{12} variants.

1.6 designing a phage display vector

1.6.1 Parameters to consider for a new display vector

A number of different phage(mid) antibody libraries have been published, to date, (see tab 1.2) all these libraries have been made using the filamentous phage, Ff (which comprises fd, f1 and M13). These phages present several advantages compare to other viral or *in vitro* systems described above: a well known biology in term of infection and replication process; a stable structure that allow strong treatments for example during the selection procedure, the possibility to use phage or phagemid system to vary the display valency and many more.
Although many essential requisites have been well defined, in the creation of a library a number of different choices can still be made. As can be seen in tab 1.3 the variables to consider include: a) phage biology related points b) structure and function of the antibody and c) general cloning and molecular biology choices. The advantages and the problems related to each of these choices will be discussed in the following paragraphs.

1	The display protein used (p3, p6 or p8)
2	Phage or phagemid
3	Antibody form (Single chain Fv (scFv) or Fab)
4	How to assemble and clone the V regions
5	For scFvs: the order of V regions and the linker sequence
6	Primers used to amplify the V genes

Tab 1.3 Possible variables to consider during the construction of a phage antibody library

1.6.2 Phage or phagemid?

A number of different phage and phagemid vectors have been created and successfully used to display antibodies and other proteins. From the point of view of phage antibody libraries, the choice of phagemid or phage has a number of considerations. From a practical point of view it is easier to work with phagemids than phage (preparing DNA, transfection efficiency), and as a result it is far easier to create large libraries in phagemid than phage. This advantage is offset, however, by the slightly greater difficulty in using phagemid libraries in selection (the cycles of recovery are longer because of the need for helper phage superinfection).

Another advantage of phagemid vectors concerns their relative resistance to deletions of extraneous genetic material. Filamentous phage vectors in general have a tendency to delete unneeded DNA. Phagemids suffer far less from this disadvantage and as a result are more stable.

The fact that phage carry 3-5 copies of the recombinant protein, whereas with phagemid 1-10% of particles contain a single copy of the displayed protein (Clackson and Wells 1994), means that when selecting with phage, avidity effects allow the selection of particles with lower affinities. It also means that it is more difficult to improve the affinity of proteins displayed on phage as such avidity effects overcome any small changes in affinity brought about by mutation. Another practical advantage of phagemid libraries is that it is easier to produce soluble proteins if an amber stop codon is inserted between the displayed protein and p3 (see fig 3.3) (Hoogenboom, Griffiths et al. 1991).

1.6.3 The display protein

Although, p3, p6 and p8 have all been used to display proteins (see tab. 1.1), p3 is the display protein par excellence, having been used to display large numbers of different proteins, in fact the permissivity of this protein is remarkable.

P8 is the major coat protein, found in 2700 copies per phage, X-ray diffraction analysis and model building experiments (Malik, Terry et al. 1996) reveal no structural impediment to the display of large peptides (up to 5 kDa) as fusion protein on the surface of the phage. P3 is the phage protein involved in bacterial infection and is present in 3 to 5 copies at one end of the phage. The site at the N-terminal after the leader has become the site of choice for display of foreign proteins. P6 is another minor coat protein found at the same end of the phage as p3, in contrast to p3 the C-terminal

rather then N-terminal is exposed, this make the protein suitable for the creation of cDNA libraries.

Where a direct comparison has been made between the display efficiency of p3 and p8, it was found that despite the fact that Ff contains 2700 copies of p8 and only 3-5 copies of p3, there was more scFv expressed in the p3 format, indicating that p8 is a rather sensitive site for proteins of this size (Kretzschmar and Geiser 1995). This is confirmed by experiments in p8, which have attempted to define the length limits of insertion within p8, where it was found that phage viability is affected by peptide length while peptide sequence plays a minor "tuning" role. Most peptide of six residues are tolerated whatever their sequence, while only 40% of phages carrying an octapeptide can form infective particles, this fraction drops to 20% and 1% for insertions of 10 or 16 amino acids long. These number can be increased by overexpressing leader peptidase protein (Malik, Terry et al. 1998). This could overcome problems related to a reduced susceptibility to leader peptidase due to unfavorable secondary or tertiary structure of the inserted sequence.

For all these reasons, p3 has been the display protein used for all published antibody libraries. Display within p3 appears to be limitless judging by the large numbers of proteins that have been expressed at its N terminus, but some constraints have been found for peptides with positive charges close to the signal sequence. or proteins containing long hydrophobic regions These appear to interfere with correct insertion of p3 into the *E. coli* inner membrane, or problem in translocation, thus blocking assembly and extrusion of the phage.

1.6.4 Antibody form

The variable portion (Fv) of an antibody comprising the variable heavy and light chains (VH and VL) is the smallest portion of the molecule that consistently maintains the binding specificity of the whole antibody. The choice of which form of antibody fragment to use in a phage antibody library is on the whole dictated by practical considerations.

Fabs form consist of two chains, the VH+CH1 and the VL+CL which need to assemble. In the case of bacterial expression system the assembly will occur in the periplasmic space. In general, Fabs are more difficult to assemble, more likely to be degraded, have lower yields as soluble fragments, and are more likely to cause problems of DNA instability in the phage due to the larger size of the encoding DNA. All these problems arise

from the fact that Fabs contain two protein chains which need to assemble, each one of which is the same size as the single scFv chain. However, Fabs also have advantages: large libraries have been made using *in vivo* recombination and Fabs do not appear to suffer from the problem of dimerisation which afflict scFvs. This make easier the selection on the base of affinity and not avidity, and also measure of affinity of isolated clones (de Haard, van Neer et al. 1999).

scFvs on the other hand are single proteins with the two V regions joined by a flexible linker (Bird and Walker 1991). scFvs do not suffer from many of the problems described above, but the main limitation is that they have the tendency to form dimers in which the VH of one scFv interacts with the VL of another, this problem being aggravated by shorter linkers. The formation of such dimeric antibodies has been exploited to create dimeric antibodies (termed diabodies) in which each member of the dimer expresses a different antibody (Holliger, Prospero et al. 1993). Although, this problem can be reduced by increasing the length of the linker to more than 20 amino acids, it has also been exploited to create phage antibody libraries in which each phage carries two antibody specificities (McGuinness, Walter et al. 1996).

1.6.5 Assembling and cloning V-genes

To create phage antibody libraries VH regions need to be coexpressed with VL regions, a goal which involves placing both genes on a single plasmid. V-regions may be assembled using 3 different approaches.

The original method for scFv format was PCR assembly. In this method V regions are amplified with regions of overlap, either to a separately amplified linker region, or to each other, in such a way that mixing the two V regions recreates a linker region joining the two V genes. With a number of amplification cycles without the addition of external primers VH regions are joined to VL and subsequently the full scFv is amplified with the addition of external primers and other cycles of PCR. This method has the main advantage that the assembled chains are cloned in a single step rather than separate VH and VL fragments. However this procedure can be difficult and is sometimes unsuccessful, first of all it require a careful titration of the amounts of VH and VL, furthermore, it is impossible to assess the degree of recombination which has occurred, since residual VH and VL fragments are usually observed.

An alternative method involves the assembly of the library by cloning. Cloning involves the separate amplification and cloning of VH and VL regions into separate vectors and the joining of these by cloning steps. This increases the number of cloning steps and so the possibility of introducing bias and loseing diversity. However there is the advantage that it is relatively straightforward to create separate VH and VL libraries with diversities of 10^{7-8}, this providing a readily accessible source of V regions for further PCR or cloning.

The third method of joining heavy and light chains involves the use of recombination systems. One phage library in the Fab format has been constructed in this fashion (Griffiths, Williams et al. 1994) all the system will be described in more details in the following paragraphs and in chapter 5.

During all the amplification and cloning steps for the creation of an antibody library one of the major concern is about the risk to reduce the diversity of the original VH and VL repertoire. Considering the cloning procedure this involve restriction enzymes digestion of the PCR fragments. It is important that the enzymes used do not cut frequently within the V genes, otherwise those V regions will be unclonable and diversity will be lost. All the V genes germline sequences information are now available and so it is more easy to analyze the frequency of cutting of the restriction enzymes and determine which are the useful one to use.

1.6.6 V-region order and linker sequence

Analysing the scFvs characteristics in more detail two main points are essential for the the construction of a functional scFv protein: V region order and the sequence of the linker peptide.

scFvs have two possible conformations: VH-VL or VL-VH. The first scFvs and all libraries published to date have been made using the VH-VL format. Although few direct comparisons have been made (Ayala, Balint et al. 1995; Luo, Mah et al. 1995) between these two formats, one theoretical consideration suggests that the VL-VH format may be better: VL chains are generally more soluble than VH chains, and as such may be more likely to induce the formation of soluble scFv-p3 fusion protein if they enter the periplasmic space before the VH chain. A number of authors have reported that the presence of positive charges within 20-30 amino acids downstream from the signal peptide con inhibit protein translocation to the periplasm (Li, Beckwith et al. 1988; Boyd and Beckwith 1990). VH regions have more

charged residues than VL, in some cases a simple reversion of the VH-VL order led to a soluble scFv compared to an originally insoluble one.

Furthermore, the use of the VL-VH format has allowed the use of a very short tag sequence, based on the FLAG epitope, at the N terminus of the light chain (Knappik and Plückthun 1994). This would be particularly useful to visualise full length scFvs in Western blots, although it has not been systematically shown that the addition of this tag at the N terminus of VL, at a site very close the antibody binding site, does not alter antibody reactivity.

The second point to consider is the linker sequence. The peptide linker used to join V domains should be flexible and long enough to allow pairing of the two domains to form an intact antigen binding cavity. Considering the length of the linker several studies show that direct linking of the domains or the use of linker peptides which are too short to allow pairing of the domains on the same chains favor pairing between domains in two adjacent chains. In addition to the length of the linker the amino acid composition has a crucial role in the design of a viable linker. Many commonly used linkers peptides are rich in the small gly-ser residues. These residues are optimal since hydrophilic amino acid allow hydrogen bonding to the solvent and glycines provide the necessary flexibility.

Many different linkers have been used in single antibodies expressed as scFv see table 1.4. however, although some comparisons have been made for single scFvs, no direct comparison has been made between these different linkers in the creation of phage antibody libraries. In fact, the only linker which has been used extensively in scFv libraries is the (Gly4-Ser)3 linker. This has a number of known disadvantages, including the tendency to shorten and form dimeric scFvs.

Linkers used in scFvs	length	order	**Comments**
GGGGSGGGGSGGGGS	15	VH/VL	Linker used in many scFv libraries so far (VH/VL format). Has a tendency to make scFv dimers
GGGGSGGGGSGGGGSGGGGS	20	VL/VH	Longer GS linker which has been shown to reduce dimerisation in the VL/VH format
GSTSGSGKPGSGEGSSKG	18	VL/VH	218: based on the original GS linker and claimed to be less susceptible to proteolysis and aggregation
SGGSTSGSGKPGSGEGSSGS	20	VL/VH	220: based on the 218 linker but 2 aa longer to reduce dimer formation
PGGNRGTTTTRRPATTTGSSPGPTQSHY	28	VH/VL	Natural linker from fungal cellulase, forms monomers, but contains an internal protease site

ASTSSGGGGSITSYSIHYTKLSGGGGSEL	29	VH/VL	loxP wild type recombination site, flanked by extra amino acids, used to make a single functional scFv
GQPKSSPSVTLFPPSSNG	18	VL/VH	
GSTSGSGKSSEGKG	14	VL/VH	Each of these has been used for isolated scFvs. On the whole they tend to be short and so will probably suffer from problems of dimerisation
GGSGSGGSGSGGSGS	15	VL/VH	
GGGGTGGGGTGGGGT	15	VH/VL	
EGKSSGSGSESKEF	14	VL/VH	
GGSGGSGGSGGSGG	14	VL/VH	

Tab 1.4 List of linker sequences used to clone single scFv or to create libraries.

1.7 Strategies to create large primary antibody répertoires

1.7.1 Site-specific recombination as tool to increase repertoire size

Most site specific recombination systems in bacteria and yeast fall into one of two large families (Craig 1988; Stark, Boocock et al. 1992) that carry out a similar repertoire of genetic rearrangements, including integration and excision of bacteriophage genomes into and out of host chromosomes, resolution of transposition intermediates, regulation of gene expression and plasmid copy number by inversion of DNA segments. The Integrase family, also known as the "tyrosine recombinases" include over 100 members that share a conserved catalytic domain responsible for cleavage and ligation of DNA substrates. The second large family of site specific recombinases includes the resolvase enzymes from the γδ and Tn3 transposon and the Gin and Hin invertases. Integrase family members cleave their DNA substrate by a series of staggered cuts, during which the recombinase became covalently linked to the DNA through a catalytic tyrosine residue. Recombination take place by exchanging one set of strands to yield a Holliday structure intermediate, followed by the exchange of a second set of strands to resolve the intermediate into recombinant product. This reaction results in a precisely defined recombination between appropriate target sequences (site-specific recombination). Certain site specific recombinases, belonging to the lambda integrase family of proteins, require no other additional accessory proteins to catalyse recombination of their target sequence, which are typically 30-40 bp. Since target size are sufficently large that is unlikely to occur naturally in any eucariotic/procariotic genome these enzymes offer the potential of a site specific recombination system for use in any species (Kilby, Snaith et al. 1993).

To overcome the transformation limit of bacteria, recombination has been proposed as an alternative to the use of cloning to create large libraries (Waterhouse, Griffiths et al. 1993; Geoffroy, Sodoyer et al. 1994; Tsurushita, Fu et al. 1996). In the first two cases, cre recombinase is used to recombine VH with VL in the Fab (Waterhouse, Griffiths et al. 1993) or scFv formats (Tsurushita, Fu et al. 1996), while in the latter, lambda recombinase is used to recombine Fabs (Geoffroy, Sodoyer et al. 1994). Important parameters to be considered when choosing between this systems are: a) the availability and inducibility of the recombinase activity; b) the irreversible or leaky character of the vector association; c) the rapid identification and easy recovery of true recombinants. The two used systems will be described and discussed in the following paragraphs.

1.7.2 The *ATT* system

This system take advantages of the site-specific integration of bacteriophage λ (Craig 1988; Miller and Feiss 1988) A circular form of bacteriophage λ integrates into the bacterial cromosome during the establishment of lysogeny. Two sequences are involved in intermolecular recombination, a 240 bp phage encoded *attP*-specific site and a 23 bp *attB* site in the bacterial cromosome. The phage encoded protein Int and E. coli encoded protein IHF are both required for this process. Two different att sites, *attL* (99 bp) and *attR* (164 bp), are created upon recombination. The presence of a phage encoded excision factor (Xis) is absolutely essential to reverse the integration mechanism.

A system of recombination using the *att* system is described in (Geoffroy, Sodoyer et al. 1994). Two different plasmids were created. One is the prototype of the phagemid family, carrying *attP*, the cat promoter, two replication origins ColE1 and f1 a resistance to ampicillin and the heavy chain of an antibody. The second plasmid is the prototype of the plasmid family, carrying *attB*, the cat gene, a distinct bacterial replication origin (P15), a resistance to kanamycin and the light chain.

The functionality of the system was tested with a monoclonal antibody anti HIV gp160. The plasmid carrying the light chain was transformed and the bacteria were subsequently infected with the phagemid particles carrying the heavy chain. The recombination was induced in a specific recombination strain expressing lambda recombinase, and upon recombination a new phagemid was obtained carrying both antibody chains and resistent to chloramphenicol thanks to the functional association of the cat resistance

gene with its promoter. The absence of the Xis protein prevents any reverse recombination.

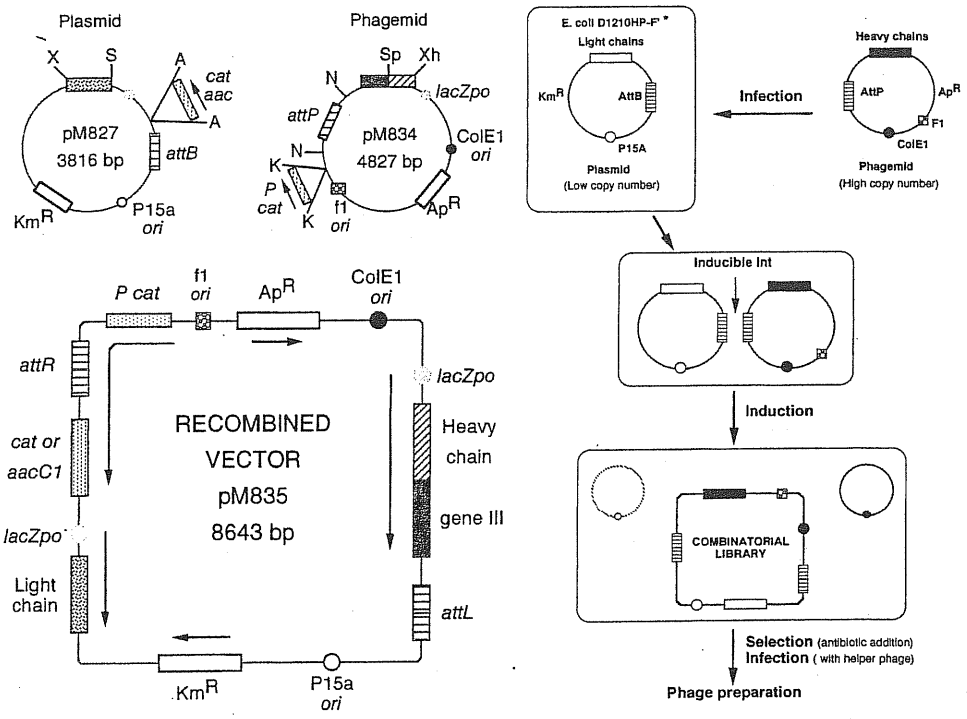


Fig 1.7 Protocol and vectors used for recombination (Geoffroy, Sodoyer et al. 1994).

Althougth the system seems extremly easy it suffers from several problems such as the large size of the recombined plasmid that limit the packaging efficency compared to the small plasmid phagemid present; problems related with the induction of recombinase through heat shock; low expression of the genetic markers and others. These problems have limited published reports on the use of this system to a single antibody. A large library was recently reported at an International meeting (2nd Phage Club Meeting, Maastricht, The Netherlands, May 1998), but it did not prove possible to select antibodies recognising antigens from this library.

1.7.3 *loxP /cre* site- specific recombination in E. coli cells

Probably the most used site-specific recombination system is *loxP/cre*. Analysing the properties of fragments of the P1 bacteriophage genome, Sternberg (Sternberg 1981) isolated a fragment of DNA which is necessary for recombination (locus of recombination, hence *lox*) and a P1 genes whose product causes recobination (hence cre) both in cis and in trans. The

loxP/cre system may be involved in several processes *in vivo* (Adams, Bliska et al. 1992). Its proposed role include cyclization of newly injected linear P1 DNA, the occasional integration of the cyclic prophage into the host chromosome, the promotion of the transition from early to late replicative forms and the maintenance of plasmid copy number. The experimental evidence for the last role is the clearest. The P1 intracellular plasmid form is maintained at a copy number of about 1 per chromosome, yet the frequency of plasmid loss in 10^{-5}. Removal of either the cre gene or the *loxP* site increase loss by 40 fold

The sequence of the *loxP* sites reveals that it contains a 13bp inverted repeat separated by an 8 bp spacer region.

Recently the three-dimensional structure at 2.4Å resolution of a covalent intermediate in the site specific reaction has been solved (Guo, Gopaul et al. 1997). Each Cre molecule contact the outermost 15 base pairs of one *loxP* half site, which include the 13-base-pair inverted repeat sequence and the first two base pairs of the central strand-excange region. The Cre recombinase makes specific cleavages in the *loxP* site resulting in the formation of a characteristic staggered cut in the spacer region. The cleavage result in the covalent attachment of both the bound enzymes to the released 3' phosphate groups.

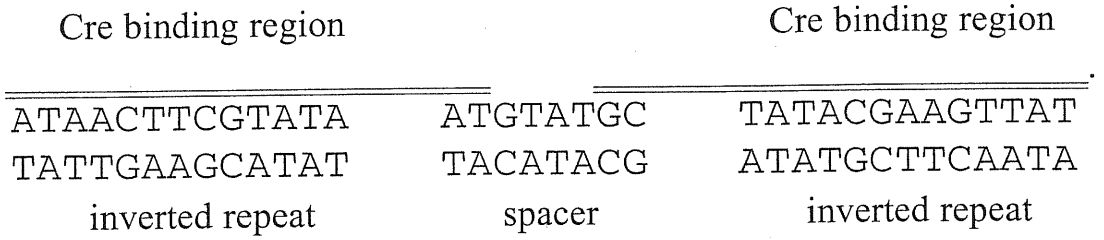


Fig 1.7 Sequence of the *loxP* site. The inverted repeats and the spacer region are indicated. Lines above the site indicate the region to which Cre binds

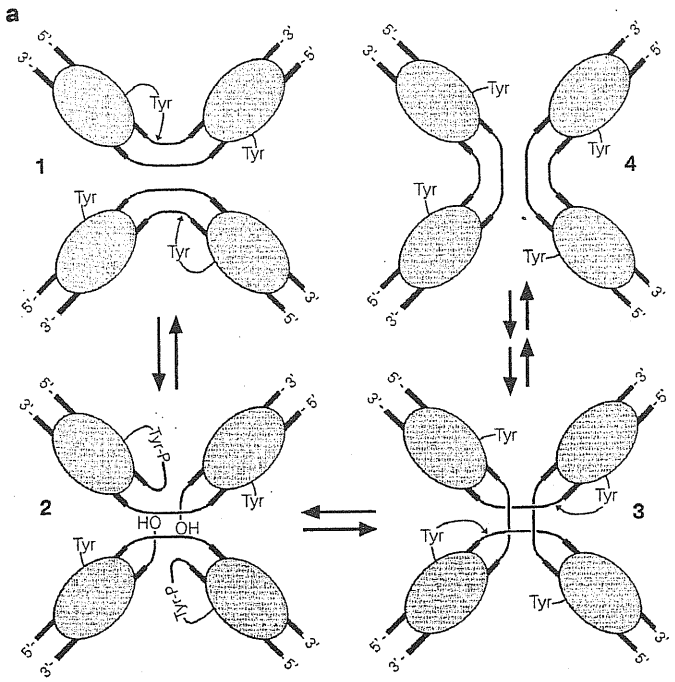


Fig 1.8 *loxP*/cre site specific recombination reaction based on studies in the lambda integrase family and on the work of (Guo, Gopaul et al. 1997). 1) A single recombinase molecule binds to each of the inverted repeats so that when synapsis occurs between the target sites, four protomers are involved. 2) conserved Tyr 324 cleaves each substrate to form covalent 3'-phosphotyrosine intermediates in an antiparallel arrangement. 3) strand transfer reaction between 5' hydroxil groups released and the phosphotyrosines yielding a Halliday-junction intermediate. 4) A second round of clevage and strand exchange gives recombinant products (Guo, Gopaul et al. 1997).

The mechanism of Cre mediated recombination was further investigated by mutagenic studies of the *loxP* site. Small number of changes in the inverted repeats do not affect recombination efficency, while mutation in the spacer region have profound effects on recombination (Hoess and Abremski 1984; Hoess, Wierzbicki et al. 1986). This is presumably a consequence of the mechanism of the recombinase; the single strands of DNA generated by steggered cut in the spacer region must be complementary for efficent recombination. Only pair of sites having identity in the central 8bp of the core region are proficient for recombination; sites having non identical core sequences (heterospecific lox sites) do not efficently recombine with each other. Thus the core region of the lox site determines the specificity of recombination with a lox partner. For example, a variant hetero specific lox site "*loxY*" would not be able to recombine with the canonical *loxP* site, but would be proficient for *loxY* x *loxY*

recombination. Several mutated core sequences have bee isolated and each of this mutant sites have 100% efficent recombination with themselves devoid of any recombinat activity with *loxPWT* site. One of the most used site is the *loxP*511 in which the mutation is an A instead of a G as the seventh base in the spacer. A functional lox site was recently also found in the yeast genome and it has been termed Fas lox (Sauer 1992).

1.7.4 Using *loxP*/Cre recombination as a strategy for making large phage antibody repertoire

Large phage antibody library has only been made using cre recombinase to recombine Fabs (Griffiths, Williams et al. 1994). This is schematically illlustrated in figure 1.9, which shows the recombination which occurs within a single bacteria containing a heavy chain, VHA, which is infected by a phage carrying a light chain, VLA. After recombination the functional VHA/VLA combination is produced, as well as the three non-functional forms illustrated.

To make this recombined library two primary synthetic libraries were first prepared: 1) a VH library of 10^8 VH+CH1 genes cloned into a plasmid vector with the VH+CH1 domains flanked by two recombination signals (*loxP* wild type and *loxP511*). 2) a VL library of $8x10^5$ VL+CL light chain genes made in fd phage containing a dummy VH+CH1 flanked by the same recombination signals.

The library was then made by infecting bacteria containing the plasmid heavy chain library with phage containing the light chain library. The recombination signals are only able to recombine in a homologous fashion, i.e. *loxP* with *loxP* and *loxP*511 with *loxP*511, and as a result the dummy heavy chain can be exchanged for the library heavy chain, as illustrated in figure 1.9, to create the library.

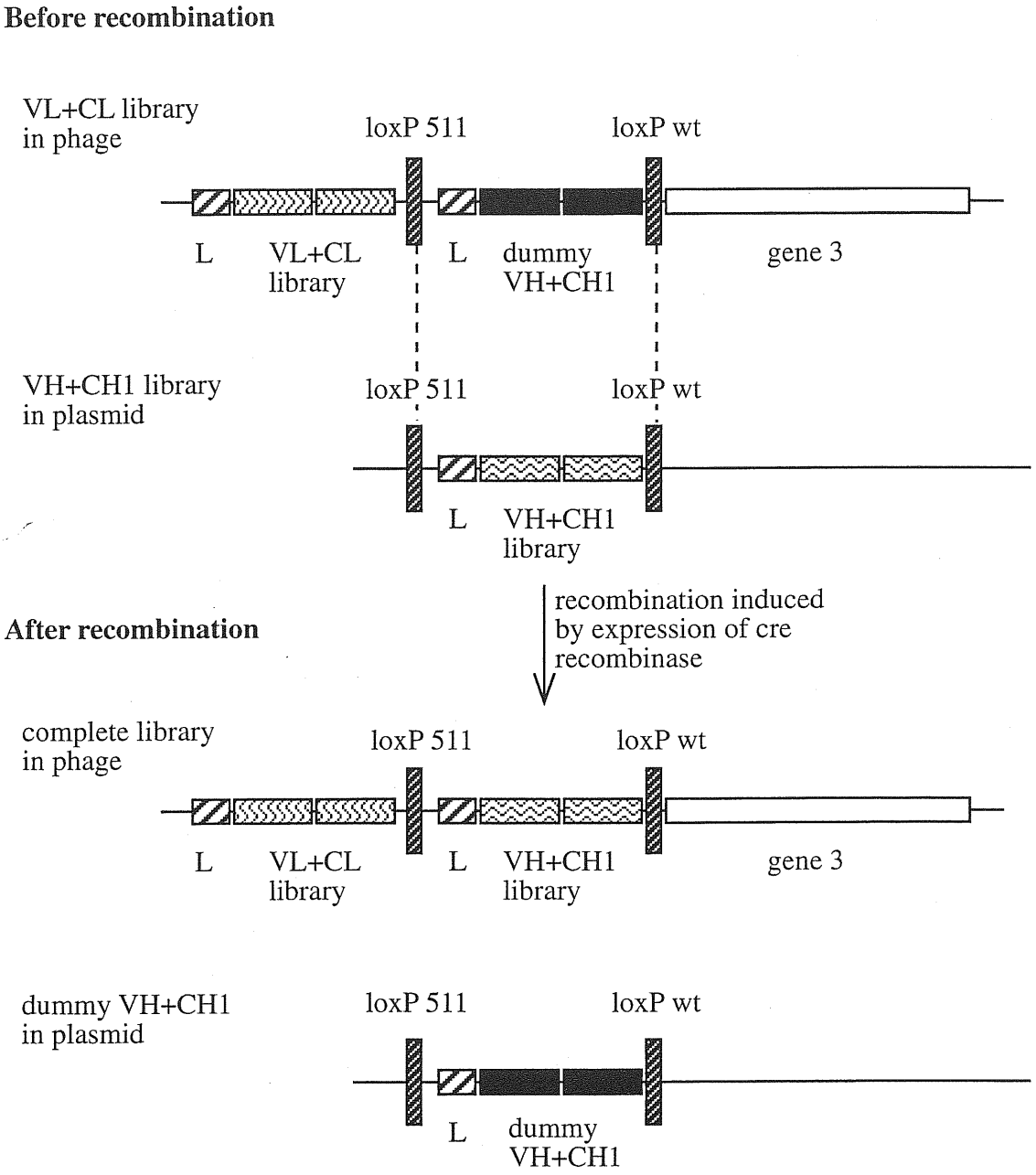


Fig. 1.9 Model of recombination to create library according to (Griffiths, Williams et al. 1994)

Antibodies with nanomolar affinities were isolated from this library. As mentioned above, this permits recombination between a single heavy chain gene and a single light chain gene, but does not permit recombination between multiple V genes.

Although requiring far fewer transfections than the other methods, this system does suffer from a number of problems:

a) the recombination system is reversible, and as a result the library will be 'contaminated' with the starting phage (containing dummy VH chains) and plasmids (containing the VH+CH1 library or the dummy VH+CH1) at relatively high levels;
b) there are by-products of the recombination process which consist of double plasmids;
c) it has been found that most plasmids, the pUC19 based plasmid used here included, although they do not contain an Ff origin of replication, can nevertheless be incorporated into phage particles with variable efficiencies. As a result, the final library contains a mixture of many different genetic elements, so reducing the effective functional diversity;
d) the use of a phage to produce the final genotypic and phenotypic diversity results in genetic instability caused by deletion of antibody genes at relatively high frequencies, which is a characteristic of the use of phage as a cloning vehicle;
e) recombination occurs between a single phage DNA and a single plasmid DNA molecule, recombination between multiple different plasmids is not possible. As a result, each bacteria produces a single novel recombined antibody. This limits the size of the library obtained to the number of bacteria used in the recombination step, which reaches a practical limit of $5x10^{12}$ for 10 liters). To obtain larger libraries, more bacteria need to be used.
f) The products of recombination are a final product, and cannot be recombined again. As a result there is no facility to recombine after selection, a procedure which allows one to access a diversity far greater than the actual library and close to the theoretical maximum diversity (the product of the number of VH genes and VL genes)

1.7.5 Other example of Cre mediated recombination of scFv

LoxP/cre system has been used in other model experiments. Tsurishita (Tsurushita, Fu et al. 1996) created a new phage display vector for *in vivo* recombination of scFv. Also in this case, a plasmid and a phagemid vectors were constructed. The VH was cloned in the plasmid vector, the VL and gene 3 in the phagemid vector. Cre recombinase was cloned in a third vector under an inducible promoter. All the vectors have different origins or replication and antibiotic resistances. After infection with the phagemid particles and induction of the Cre there is the recombination and formation of two new plasmids: one carrying the complete scFv and gene 3, the second a byproduct of recombination. In this case the *loxP* site is included

between the VH and VL and works as part of the linker sequence. Almost all of the limitations described above apply also in this case, in particular the recombination is permitted between a single heavy and light genes. No big library has been constructed with this system.

Two other applications of *in vivo* Cre recombination are described in (Davies and Riechmann 1995; Fisch, Kontermann et al. 1996) in both cases the *loxP* site has been inserted in the middle of the coding sequence, in one case VH chain and in the second a peptide sequence. In the first case the *loxP* sequence is retained in the coding sequence, in the second it is spliced out.

1.7.6 Problems for recombination

Altough the examples described above showed that in theory and in practice *in vivo* recombination is a useful and powerful tool to create diversity in a population of DNA sequence many problems still need to be overcome to use recombination induced within cells to create diversity. The first problem is how the different DNAs which are substrates for recombination are introduced into the cell which perform the recombination, and the second is how the different DNA substrates for recombination are maintained within the cell for sufficiently long for recombination to take place.

The experiments described in the paragraphs above use infection to introduce one recombination substrate into a cell in which the other recombination substrate is already present. This has an efficiency of entry which approaches 100%. However, the full power of *in vivo* recombination is not exploited using this system, because recombination is only induced between a single phage and a single plasmid within a single cell; i.e. between two DNA substrates. It would be far more powerful if more than two DNA substrates could be used.

It is known that bacteria containing a Ff filamentous phage are resistant to subsequent superinfection by other Ff phage (e.g. see page 191 of (Boeke, Model et al. 1982) and (Marvin and Hohn 1969)). This resistance arises from an inhibition of the formation of the F' pilus, which is the receptor on the bacterial cell surface exploited by Ff phage to enter the bacteria. This inhibition of F' pilus synthesis is induced by the gene 3 protein, which is present on the phage and is rapidly produced following infection. This has biological sense, as it is to the advantage of a phage which has infected a bacteria to inhibit infection by other phage, as in this way it increases its

own progeny. As the phagemid used to produce phage antibodies also contains gene 3, it may seem logical that it would not be possible to infect a bacterium with more than one infectious particle, of the same kind. However, it is well known that bacteria containing phagemids (introduced either by infection or by transfection) are amenable to subsequent efficient infection by Ff phage (e.g. helper phage) if p3 production is inhibited, by for example, using glucose to turn off the activity of the promoter. Furthermore, it has been shown that while superinfection cannot occur, co-infection can occur, although the number of different phage shown to enter a single bacterium is limited to no more than 3 (Marvin and Hohn 1969). This is in line with the two to three pili expressed by a bacterium at any one time (see (Marvin and Hohn 1969)). This suggests that, although never exploited previously, it may be possible to use co-infection of infectious particles to introduce more than one nucleic acid molecule into a cell, although this may be limited to no more than 3. This does not, therefore, solve the second problem, that of the maintenance of DNA substrates within cells for sufficiently long to permit recombination to occur. This has usually been overcome by putting the substrates for recombination on different plasmids containing origins of replication belonging to different incompatibility groups (as described above for both the Cre and the ATT system) and different antibiotic resistances. This strict limitation of the number of potential recombination substrates requires that recombination substrates be specifically designed, and does not fully exploit the potential for recombination.

Chapter Two:
Materials and Methods

Abbreviations

aa	Amino Acids
Amp	ampicillin
bp	base pair
cDNA	complementarity DNA
cfu	colony forming units
DMSO	Dimethylsulfoxide
EDTA	ethylenediamine tetraacetic acid
IPTG	isopropylthio-β-D-galactoside
isoAA	isoamyl alcohol
Kan	Kanamycin
LB	Luria-Bertani
PAGE	polyacrilamide gel electrophoresis
PBS	Phosphate buffer saline
PEG	polyethylene glycol
pfu	plaque forming units
rcf	relative centrifugal field
SDS	Sodium dodecyl sulfate

2.1 Bacterial strains

DH5αF' (Gibco BRL): F'/*endA1 hsdR17* (r_K^- m_K^+) *supE44 thi-1 recA1 gyrA* (Nal^r) *relA1 D (lacZYA-argF)U169 deoR* (F80*dlacΔ(lacZ)M15*) was used for standard phage propagation,

BS1365 *BS591 F' kan (BS591: recA1 endA1 gyrA96 thi-1 D lacU169 supE44 hsdR17 [lamda1mm434 nin5 X1-cre]* (Sauer and Henderson 1988))

HB2151 *(K12, ara Δ(lac-pro), thi/F' proA$^+$B$^+$, lackI^{q}ZΔM15)* a non-suppresser strain was used to make soluble scFv.

TG1 *supE hsd Δ 5 thi Δ (lac-proAB)F'(traD36proAB$^+$lacI^{q}lacZΔM15)*

2.2 Oligonucleotides

The sequences of the oligonucleotides used in this thesis are described in tab 2.1

Polylinker Back	TTCgAAAgCTTgCCAAATTCTATTTCAAggAgACAgTCATAAT gAAATACCTATTgCCTACggCAgCCgCTggATTgTTATTACTC gCAgCAAgCCACgCgTTCgCCTCCggAggT
Polylinker For	TTCAACAgTAgCggCCgCCTAATggTgATggTgATggTgAgTA CTATCCAggCCCAgCAgTgggTTTgggATTggTTTgCCgCTAg CTgAggAgACggTgACCTCCggAggCgAAC
VH PTL Back	ggagggtcgaccataacttcgtataatgtatactatacgaagt tatcctcgagcggta
VH PT1 For	ccaggcccagcagtgggtttgggattggtttgccgcta
VH PT2 FOR	tggtgatggtgagtactatccaggcccagcagtgggtttg
VL PTL For	accgctcgaggataacttcgtatagtatacattatacgaagtt atggtcgaccctcc
VL PT1 Back	cgctggattgttattactcgcagcaagcggcgcgcatgcc
VL PT2 Back	tacctattgcctacggcagccgctggattgttattactc
VH12 218 Back	**ggCTCgAgCAAAggT**CAggTCCAgCTKgTRCAgTCTgg
218 PT For	ACCggAACCTggTTTCCCAgA**ACCgCTggTCgACCC**
218 PT Back	gggAAACCAggTTCCggTgAA**ggCTCgAgCAAAggT**
VK 2/4 218 For	**ACCgCTggTCgACCC**TTTgATCTCCASCTTggTCC
VH12 GS Back	**ggAggCggCggTAgC**CAggTCCAgCTKgTRCAgTCTgg
GS PT For	gCCTCCggAgCCTCCg**CCACCAgAACCTCCACCgCC**
GS PT Back	ggAggCTCC**ggAggCggCggTAgC**
VK 2/4 GS For	**CCACCAgAACCTCCACCgCC**TTTgATCTCCASCTTggTCC
Gene 3 Back	CATTAggCggCCgCTACTgTTgAAAgTTgTTTAgC
Gene 3 For	AATACCCAAAAgAATTCgCATgCTTAAgACTCCTT
VL seq	tggtgatggtgatggtgagt
VH seq	ataatgaaatacctattgcc
D1.3 VHCDR	gtagtcaagcctataatctctctc
D1.3 VLCDR	caacatttttggagtactcct
gene3LOX	gttgaattcataacttcgtatagcatacattatacgaagttat gcatgcttaagactcct

Tab 2.1 Oligonucleotides used.

2.3 V-genes PCR primer design

The back (5') primers were located in the first 21-23 nucleotides of framework 1. The original DNA based primer set (Marks, Tristem et al. 1991) was analysed using the MacVector program (IBI) and aligned with all the functional V-genes in the V BASE sequence directory (Tomlinson, Williams et al. 1996). The primers were optimised empirically (changing bases and introducing degeneration) as described below.

The for (3') primers were located at the 3' end of the human germ line J segments. As knowledge on the number and sequences of the J genes has not changed greatly recently, these primers are similar to those previously described, with the exceptions described below.
The two primers specific for the IgG and IgM regions were located at aa 114-120 at the 5' end of the CH1 exon.
The oligos were further analyzed using Amplify 1.2 (Engels 1993) to ensure that no primer pair formed primer dimers. All the oligos used for amplification (back and for) in table 3.1 have a constant part added to the 5' end to introduce restriction sites in order to make them compatible with our cloning vector pDAN3/5.

TŤA TCC TCG AGC GGT ACC---	VH back
GAT TGG TTT GCC GCT AGC---	VH for
AGC AAG CGG CGC GCA TGC C---	VL back
GAA GTT ATG GTC GAC CCT CCG GAT---	VL for

2.4 Preparation of lymphocytes from peripheral blood

1) Mix 20 ml of heparinized blood (blood can also be treated with EDTA, citrate, acid citrate dextrose (ACD) or citrate phophate dextrose (CPD)) with an equal volume of BSS and divide equally into two 50 ml Falcon tubes (i.e. 20ml in each tube).
2) Underlay with 15 ml Ficoll Paque (Pharmacia 10-A-001-07). This should be done carefully to avoid mixing the Ficoll Paque with the blood. The opposite (i.e. layer diluted blood on top of Ficoll Paque) can also be done.
3) Spin for 30 min at 350g at room temperature in a Beckman table centrifuge or equivalent.
4) The PBL's are found at the interface as a whitish band. In order from the top, will be found: plasma, lymphocytes and eythrocytes, Ficoll Paque, granulocytes and red cells. Carefully remove the plasma with a pasteur pipette leaving the lymphocyte layer undisturbed
5) Transfer the lymphocyte layer to another 50 ml tube (the samples can be pooled) and dilute with at least 3 volumes of BSS. Avoid taking any of the Ficoll Paque layer which will contain granulocytes. Resuspend the cells gently using a pasteur pipette.
6) Spin at 350g rpm at room temperature for 10 minutes.
7) Remove the supernatant. Resuspend in 6-8ml BSS. Repeat step 6.
8) The cells in the pellet are now ready for RNA preparation.

BSS (balanced salt solution)
Mix 9 volumes 150mM NaCl with 1 volume solution A:

solution A

Anhydrous D glucose	0.1%,	1g/L
CaCl2.2H2O	50µM,	0.0074g/L
MgCl2.6H2O	980µM,	0.1992g/L
KCl	5.4mM,	0.4026g/L
Tris	145mM,	17.565g/L

2.5 RNA extraction (Chomczynski and Sacchi 1987).
1) Add 1ml of solution D (4M guanidine thiocyanate, 25mM sodium citrate pH7, 0.5% sarcosyl, ß mercaptoethanol 100mM) for each 0.1g tissue or cells at room temperature. Approximately 10^7 lymphocytes should be obtained from 20ml of blood. This should be resuspended in 1ml of solution D.
2) Transfer the homogenizate to a polypropylene tube and add for each ml of lysis buffer 0.1ml of 2M sodium acetate (pH4.2), 1ml phenol (saturated pH4.3), and 0.2ml of CHCl3:isoAA (49:1).
3) Vortex 10sec and incubate 15min on ice.
4) Centrifuge 10000 rpm at 4°C for 20min.
5) RNA is present in the aqueous phase whereas DNA and proteins are in the interphase and phenol phase. Transfer the aqueous phase to a new tube and precipitate with 1 volume isopropanol at -20°C for 2 hours.
6) Centrifuge 7000 rpm at 4°C for 20min.
7) Wash the RNA pellet with 70% ethanol, recentrifuge and remove last traces of ethanol.
8) Resuspend in 50-100µl water (formamide or TE can also be used).
9) Quantitate the RNA at OD260 (0D 1 = 40µg RNA).

solution D: to make 50ml:
23.63g Guanidine Thiocyanate
1.25ml Sodium citrate 1M
1.25ml Sarcosyl 20%
water to 50ml
put at 60°C and add 375µl ß mercaptoethanol
Store at room temp in the dark (foil), stable for 1 month.

2.6 cDNA synthesis

1) In an eppendorf tube add the following in order:
-1µl random primers (100ng/µl) (2pmol specific or 500ng oligo dT12-18, can also be used)
-2µl of total RNA (1-5µg, 500ng mRNA can also be used)
-9µl sterile distilled water
2) Heat to 70°C for 10min and chill quickly on ice.
3) Centrifuge briefly in eppendorf to collect contents to bottom of the tube.
4) Add:
-4µl 5X first strand buffer (250mM Tris-HCl pH8.3 at room temperature)
-2µl 100mM DTT
-1µl 10mM dNTP mix (10mM of each of dATP, dGTP, dCTP and dTTP at pH7)
5) Add 1µl (200 units) Superscript II (Gibco BRL RNase H^- MLM reverse transcriptase, cat. no. 1806-014), or any other commercial reverse transcriptase, and mix gently by pipetting up and down.
6) Incubate 25°C for 10 minutes (omit this step if using gene specific or oligo dT). This allows the oligos to be slightly extended prior to the 42°C incubation step which would denature them.
7) Incubate 42°C for 50min.
8) Inactivate at 70°C for 10min.

2.7 V-genes PCR

1) Mix in the following order:

cDNA	2µl
H2O	29µl
10x PCR buffer	5µl
dNTP 2mM each	5µl
MgCl2 25mM	3µl

The reaction can be irradiated on a short wave transilluminator (to destroy contaminating DNA) for five minutes at this stage.

For primer (10pmol/µl)	2.5µl
Back primer (10pmol/µl)	2.5µl
Taq polymerase (5u/µl)	1µl

2) Put the tube in the PCR machine and use the following program
94°C - 5'

94°C - 1' |
60°C - 1' | 30 cycles
72°C - 1' |

72°C - 10'
4°C - 24 h

2.8 Pull through PCR of amplified V regions

V regions amplified from cDNA can be reamplified to increase the amount available for cloning as well as to add extra restrictions sites etc. at each end. As the starting template is a PCR fragment this amplification tends to be extremely efficient.

1) Mix in the following order:
-10µl PCR buffer
-5µl purified VH/VL band
-10µl VH or VK pull through primer mix (PTL and PT1)
-6µl MgCl2 25mM
-10µl dNTP 2mM
-ddw to 100µl
cover with mineral oil (50-100µl)

2) Put the tube in the PCR machine and use the following program:

94°C - 5'

94°C - 1' |
60°C - 1' | 30 cycles
72°C - 1' |

72°C - 10'
4°C - 24 h

3) When the tube has reached 94°C, add 0.5µl Taq polymerase (or Vent polymerase) and leave to carry out program.

2.9 Fingerprinting PCR bands

To make a good library many V regions are required. When amplifying V regions from cDNA it is not obvious from an examination of the amplified

band whether a single V region (which may be a contaminant) or many are present. This can be resolved by carrying out fingerprinting of the V regions using restriction enzymes with 4bp recognition sequences. Each individual V region will give a 'fingerprint'. A diverse collection of V regions will give a smear when cut with such enzymes (caused by an overlap of many different fingerprints), whereas a restricted collection of V regions will give a number of discrete bands. Fingerprinting may be carried out on V regions amplified from cDNA, or a complete library (in which case many different V regions are present), or on V regions amplified from individual clones (each one of which should be different in a good library).

1) For fingerprinting the V regions derived from cDNA, add 10µl of the PCR mix directly to 20µl of the enzyme mix described below. This can be done in a single eppendorf tube rather than a 96 well plate. After digestion run adjacent to 10µl of the non digested PCR fragment. A similar fingerprint can be done with the positive control which contains a single V region.

2) If fingerprinting n clones, make a restriction enzyme mix containing:

- (n+1) x 17.8µl water
- (n+1) x 2µl restriction enzyme buffer (New England Biolabs buffer 2)
- (n+1) x 0.1 or 0.2µl restriction enzyme (0.1µl for HaeIII HC and 0.2µl for BstNI)
- (n+1) x 0.2µl BSA (10mg/ml) required for BstNI, optional for HaeIII

enzymes which can be used:

BstNI (NEB buffer 2+BSA)

HaeIII (NEB buffer 2). This is a high concentration preparation.

3) Add 20µl of the mix to each well of a 96 well plate.

4) Add 10µl of PCR mix from different clones to each well.

5) Cover the wells with sellotape (scotch tape) and put to incubate at the appropriate temperature (BstNI 60°C, HaeIII 37°C) for 2-3 hours.

6 Add 5µl 6X gel loading buffer to each well and load on 2% Metaphor/Nusieve TBE gel. Run 60V 60-120'.

2.10 Creating pDAN3 (scFv display vector) and pDAN3-scFv derivatives

To make pDAN3, a new polylinker was cloned into pUC119 using HindIII and NotI by overlap PCR of two long oligonucleotides Polylinker Back and polylinker For. This introduced the bacterial leader sequence, a polycloning cassette containing the restriction sites, the SV5 tag (Hanke, Szawlowski et al. 1992), a His_6 tag and an amber stop codon (see figures 3.3).

Gene 3 was amplified from fdtet using the primers Gene 3 Back and Gene3 For. The fragment was then digested with NotI and EcoRI and cloned in the vector pUC 119 plus Polylinker. The 5' end of mature gene 3 was inserted downstream of the amber stop codon using these sites.

The D1.3 scFv was assembled (in the order VL-VH) from D1.3 scFv (VH-VL order) using the V region specific primers (Krebber, Bornhauser et al. 1997).

1) VH and Vk genes were amplified from previously created D1.3 scFv.
2) All V genes were gel purified and approximately 200ng were used as templates for further amplification to add a region of overlap in the scFv linker as well as long tails to facilitate restriction enzyme digestion. The primers VL back PT1 and VL for PTL were used to reamplify VL genes and VH for PT1 and VH back PTL for the VH genes. Amplified bands were gel purified as below.
3) The D1.3 scFv was assembled by mixing equal amounts (50-200ng) of VH and VL genes and performing assembly essentially as described in (Krebber, Bornhauser et al. 1997): 8 cycles of PCR without primers followed by 25 cycles in the presence of VL back PT2 and VH for PT2. Cycling parameters were 94°C for 1 min (denaturation), 60°C for 1 min (annealing) and 72°C for 1' 30" (extension).
4) The amplified scFv was digested with BssHII and NheI and ligated into BssHII / NheI cut pDAN3. The ligation mix was electroporated into electrocompetent DH5αF' and plated on 2XTY 100μg/ml ampicillin / 1% glucose plates, clones were confirmed by sequencing.
The Y13-259 scFv was assembled (in the order VL-VH) from Y13-259 scFv (VH-VL order) (Werge, Bradbury et al. 1992). Other pDAN-scFvs were cloned, either as assembled scFv PCR fragments, or as individual VH and VL genes amplified to incorporate BssHII and SalI for VL and XhoI and NheI for VH.

2.11 Introducing a second *lox* sequence: creating pDAN5 D1.3 vector

To introduce the *loxP*WT site at the end of gene3 a new oligo named gene3lox was designed. The oligo has the 3' homologous to gene3, then the *loxPWT* sequence, and the 5' is homologous to the vector. Gene3 was amplified with the primers gene3 back and gene3lox. The PCR band was gel purified, digested with Not I and EcoRI and cloned in the vector pDAN in

which the gene3 has been excised using the same enzymes. After ligation and transformation positive colonies were confirmed by DNA fingerprinting and sequencing.

2.12 Creation of the D1.3 vectors and model recombination

Two scFvs which contained either D1.3 VH or D1.3 VL with irrelevent partner chains were created (VL/D1.3-VH/X and VL/Y-VH/D1.3) by PCR cloning. Recognition of lysozyme by D1.3 scFv was shown to require the presence of both D1.3 heavy and light chains; single D1.3 chains associated with irrelevent partner chains were non-functional.

The strategy of combinatorial infection and *in vivo* recombinaton was applied following the scheme:

1) 10ml E. coli BS1365, which expresses cre recombinase constitutively was grown to OD550 0.5 at 37°C in 2XTY 100μg/ml kanamycin/1% glucose. all the steps were performed simultaneously with DH5αF' bacteria

2) Equal amounts of phagemid particles containing the two scFv genes (VL/D1.3-VH/X and VL/Y-VH/D1.3) were added to the bacteria at an MOI of 20:1 ($5x10^{10}$ of each phagemid particles added to $5x10^{9}$ bacteria). This was left for 30 minutes at 37°C without shaking to allow infection to occur.

3) Ampicillin was added to 100μg/ml and bacteria were grown overnight at 30°C. Recombination occurs during this period.

4) After overnight growth, bacteria were diluted 1:20 in 10ml of the same growth medium in order to have an OD600 of 0.1 or less. The culture was grown at 37 °C to OD600 0.5 and M13 K07 helper phage added at MOI of 20:1.

5) This was left for 30 minutes at 37°C without shaking to allow infection to occur.

6) The culture was grown for 6-18 hours at 30°C in the presence of 1% glucose, centrifuged at 4000 rpm for 15 mins and the supernatant taken. This step prepares phagemid particles which do not display scFv.

7) 10ml DH5aF' was grown to OD550 0.5 in 2XTY at 37°C.

8) Phagemid particless prepared in step 6 were added to the DH5aF' at MOI less than 1, left for 30 minutes at 37°C and plated on 2XTY 100μg/ml ampicillin/1% glucose plates. This step couples phenotype to genotype, in the absence of this step, the displayed scFv may not necessarily correspond to the scFv gene within the phagemid particles.

9) Phagemid particles displaying scFv were made from 96 individual colonies. The culture supernatat was tested in ELISA using the protocold

described in the preavious chapter. 48 colonies were dissolved in 50 ml of 2XTY medium and 1 ml was used in PCR reaction with the primers D1.3VHCDR and D1.3VLCDR.

2.13 Primary (or immune) library construction by standard cloning techniques

1) 40 different samples of human peripheral blood lymphocytes (or lymphocytes from CD patient) were prepared by density gradient centrifugation on Ficoll Hypaque (Pharmacia).

2) Total RNA was prepared from these lymphocytes by acid guanidinium thiocyanate, phenol chloroform extraction and isopropanol precipitation (Chomczynski and Sacchi 1987).

3) cDNA was prepared using SuperScript II RNase H- Reverse Transcriptase (Gibco BRL) with random hexamers starting with 1-5μg of total RNA in a final volume of 20μl following instructions provided with the SuperScript.

4) IgM VH genes (IgA for immune library) were first amplified from 0.5μl cDNA reaction, using IgMfor and the individual VHback primers described in chapter 3. Reaction volumes were 20 μl, using 0.5 μl of cDNA reaction, 10 pmol of each primer, 200 μM dNTPs, 2μl 10X PCR buffer, and 0.5 μl (2.5 U) of Taq DNA polymerase (Perkin Elmer). Cycling parameters were 94°C for 1 min (denaturation), 55°C for 1 min (annealing) and 72°C for 1' (extension) for thirty cycles. All 20μl were loaded on a 1.5% agarose gel and gel purified using the Qiagel purification kit (Qiagen). Subsequently, VH genes were reamplified using the VHfor mix of primers and the VHbackPTL in 50μl volumes using 1μl of purified VH (other parameters as before).

5) Vλ and Vk genes were similarly amplified (using individual VLback primers with the mix of VL for primers) from random primed cDNA. V genes were gel purified and used as templates for further amplification to add a region of overlap in the scFv linker as well as long tails to facilitate restriction enzyme digestion. The primers VL back PT1 and VL for PTL were used to amplify VL genes

6) The scFv library was assembled by mixing equal amounts (200-500ng) of VH and VL genes and performing assembly essentially as described in (Krebber, Bornhauser et al. 1997): 8 cycles of PCR without primers followed by 25 cycles in the presence of VL back PT2 and VH for PT1. Cycling parameters were 94°C for 1 min (denaturation), 60°C for 1 min (annealing) and 72°C for 1' 30" (extension).

7) The amplified scFv were digested with BssHII and NheI and ligated into BssHII/NheI cut pDAN5. The ligation mix was electroporated into electrocompetent DH5αF' and plated on 2XTY 100μg/ml ampicillin / 1% glucose plates to obtain a primary library.
8) The colonies were scraped up in 2XTY 10% glycerol and frozen down in 1ml aliquots.

2.14 Recombination and creation of the secondary library

To induce recombination, the same protocol applied for the model recombination of D1.3 was used, only the volume of medium and the number of phagemid particles was scale up.
1) $3x10^{12}$ phagemid particles obtained from the primary library were added to 20 ml of exponentially growing BS1365 (10^{10} cells) in phagemid particles excess with a MOI of 300:1. After overnight recombination, bacteria were diluted 1/20 into 400ml and phagemid particles prepared by standard techniques (Marks, Griffiths et al. 1992).
2) As these phagemid particles arise from bacteria containing many different scFvs, there is no coupling between phenotype and gentoype. This was overcome by infecting 1 liter of DH5αF' ($5x10^8$ cells) at OD600 0.5 at an MOI ≤1 and growing the culture overnigth at 30 °C
3) Phagemid particles were purified from the culture with double peg precipitation and further purified by cesium chloride density centrifugation as described in (Smith and Scott 1993) and resuspended in 20ml. This constitutes the final antibody phagemid library which was used for selections. Aliquotes of $5x10^{12}$ phagemid particles in 15% glycerol were stored at -80°C.

2.15 Assessment of diversity in individual cells

1) After infection at high MOI, and growth overnight, cre expressing bacteria were plated out on 2XTY 100μg/ml kanamycin / 100μg/ml ampicillin / 1% glucose plates to isolate individual colonies. These contain multiple recombined V genes (if recombination has been successful) and it is necessary to prepare phagemid particles from them to isolate single scFv genes (each phagemid particle can only contain one scFv gene).
2) Phagemid particles were prepared from these individual colonies by growing such colonies to OD600 0.5 in 1ml 2XTY 100μg/ml kanamycin/

100µg/ml ampicillin / 1% glucose at 37°C, infecting with M13K07 and using the techniques described above to produce phagemid particles.
3) These phagemid particles were used to infect DH5αF' bacteria, grown to an OD600 0.5, at an MOI less than 1. The resulting colonies contain single phagemids, and the complete population of colonies represent the diversity of the phagemids contained within the single starting bacteria isolated in step 1.
4) Individual VH and VL chains present in each phagemid were amplified by PCR using PT2 primers and fingerprinted with BstNI or sequenced.

2CDR) CDR3 fingerprinting was performed by amplification from these colonies with the primers hv2 annealing in the FR3 and with gene3FOR.
3CDR) The PCR fragment was used as templete in a single round of extension for 20' at 72°C with labeled VHseq primer.
4CDR) 3 µl were loaded in a sequnecing gel and run as a normal sequence.

2.16 Selection of phage-antibody libraries by panning in 'immunotubes'

1) To a 75 x12 mm Nunc-immunotube (Maxisorp; Cat. No. 4-44202) add 2.0 ml PBS and up to 0.1 mg/ml of antigen stock. Leave overnight at 4 °C to coat (see footnote a).
2) Next day wash the tube 2x with PBS-Tween-20 (0.1%), 2x with PBS[c] (simply pour solution in and pour out again immediately).
3) For pre-blocking, fill tube to brim with PBS containing 2% Marvel (2% MPBS)[b]. Cover with parafilm and incubate at room temperature for at least 30 min (ideally 2 hours) to block.
4) Prepare phage mix: 5-800 µl of PEG concentrated phage (~10^{11}-10^{12} TU) +500µl PBS and 1 ml of 4% MPBS, leave for 30'-1 h at room temperature. Store remainder of phage prep at 4°C.
5) Wash tube 2x with PBS-Tween-20 (0.1%), 2x with PBS.
6) Transfer phage mix from step 4 to the Immunotube. Seal tube with a cap or with parafilm. Incubate 30 min at room temp on under and over turntable and then stand for at least a further 1.5 hrs at room temp.
7) Wash tubes with 20 washes PBS-Tween-20 (0.1%), then 20 washes PBS. Each washing step is performed by pouring buffer in and out immediately. This is best achieved using a wash bottle.

8) Elute phage from tube by adding 1 ml 100 mM triethylamine (make fresh: 140 µl per 10 ml water; must be pH 12). Rotate the tube for 10 min on an under and over turntable. Phage viability decreases with longer elution times.
9) Transfer solution to an eppendorf tube with 0.5 ml 1.0 M Tris-HCl, pH 7.4 and mix by inversion. It is necessary to neutralise the phage eluate immediately after elution.
10) Transfer to ice for later re-infection.
11) Store selected phage/phagemid particles at 4°C. Displayed antibody/p3 fusion protein is gradually proteolysed during storage, so we generally perform our phage rescues and selections on two consecutive days.

a Sometimes coating for 2 hours at 37°C can replace the overnight incubation.
b Casein (Sigma, C5890) can be used as an alternative to skimmed milk powder. Use at 0.5% w/v and incubate at 50°C to dissolve. The resulting solution is cloudy. Use undiluted.
c. sometimes problems with 'stickyness' of antigens is a problem, in which case polyreactive clones may be selected from the repertoire. In that case inclusion of Tween-20 (0.05-0.1%) in ALL incubation steps (in selection itself and in all washes) may help to remove these binders, and favour the specific ones.

2.17 Rescuing phagemid libraries
1) The inoculum size should be 10x library size in number of bacteria at the start, but should not exceed 0.05 O.D.600nm.
Transfer the inoculum into 10 ml of 2 x TY, 100 µg/ml ampicillin, 2% glucose, in a 50ml Falcon tube, or 20ml in a 250ml flask. Generally make an innoculation of 10µl of concentrated bacterial stock.
2) Grow with shaking (270 rpm) for 1,5-2,5 hours at 37°C, to an O.D.600nm of 0.5.
3) When an O.D.600nm of appr. 0.5 is reached (2 x 10^{10} bacteria in total), transfer 5 ml (2 x 10^{9} bacteria) to a 50-ml Falcon tube containing an appropriate amount of helper phages. For example, of the M13-K07 stock of 10^{12} pfu/ml add 40 µl. This will give a ratio of phages: bacteria between 10:1 and 20:1. Leave at 37°C for 30 minutes, standing, in a watherbath, with occasional agitation.
4) For a <u>first</u> rescue of a large repertoire, also shake at 37°C for 30 min, at 100 rpm. (In all other cases skip this step).

5) After the infection event, spin the cells for 10 min. at 3.000 rcf. Remove the supernatant.
Resuspend the bacterial pellet in 10-20 ml of 2xTY, 100 µg/ml ampicillin, 25 µg/ml kanamycin. Transfer to a 50ml Falcon (10ml) or 250ml flask (20ml). Grow with shaking (270 rpm) overnight at 30°C (scFv)
6) Spin the culture in a 50-ml Falcon tube for 20 minutes at 3.000 rcf to pellet the bacteria.
7) To the supernatant add 1/5th of the volume of PEG solution (20 % Polyethylene glycol 6000, 2.5 M NaCl) and leave on ice for 1 hour.
8) Pellet phage by spinning for 15 min., 3.000 rcf at 4°C. Throw away the supernatant. Remove the traces of supernatant. Resuspend the pellet in 1.0 ml PBS with a blue 1-ml filter-tip; transfer to 1.5 ml eppendorf tube.
9) Spin in microcentrifuge (2 min, max speed) to remove the remaining bacteria. Transfer supernatant to a new tube.
10) Optional step : Repeat steps 6-8 (this if many contamination bacteria are visible) Add 150 µl PEG solution to the supernatant; leave on ice for 10-20 minutes; Spin phage down (5 min, max speed), remove supernatant carefully and resuspend the pellet in 1 ml PBS with a blue 1-ml filter-tip. Remove bacteria again by spinning (2 min, max speed).
11) The phage is now ready for selection. Phage can be stored at 4°C without much loss of titer; the antibodies may however proteolytically be removed by contaminating proteases, and, for selection, phages should be used within a week.
The standard yield is about 2-10 x 10^{12} phages from a 25 ml culture.

2.18 Growth and rescue of phage(mid) particles in 96-well micotitre plates.

The following methods are suitable for growth of large numbers of clones for preliminary screening for binding activity by methods such as ELISA. The cultures do not aerate very well, so the yield of phage is lower than usual. We have not, however, found this to be a problem for screening purposes.
1) Decide on the number of clones from the titration plates from the selections in "immunotubes" to analyse by phage ELISA and soluble antibody ELISA. Toothpick colonies into 150 µl 2xTY, 100 µg/ml ampicillin, 2% glucose in 96-well flat-bottomed plates (Costar, Cat No. 3595) and grow with shaking (270 r.p.m.) overnight at 30°C. It is convenient to cushion the

plate with foam inside a plastic box, such as those provided by Boehringer for enzyme storage.
2) Next day, use a 96-well transfer device to inoculate (twice) [or pipet 2 µl per well] from this plate to a fresh 96-well plate containing 150 µl 2xTY, 100 µg/ml ampicillin, 1% glucose per well. Use round-bottomed 96-well plates (Costar). Grow 2,5 hrs, 37°C, shaking.To the wells of the master plate, add 50 µl 60 % glycerol per well and after use store at -70°C.
3) To each well add 50 µl 2xTY, 100 µg/ml ampicillin, 1% glucose containing $2x10^9$ pfu M13K07 phage (0.1 µl of a $2x10^{13}$ pfu/ml stock per well). The ratio of phage to bacterium should be between 20:1. Stand 30 min at 37°C.
4) Spin 600 rcf (faster will crack the plates) for 10 min; then remove supernatant with multichannel pipette or suction device.
5) Resuspend pellet in 150 µl 2 x TY, 100 µg/ml ampicillin, 25 µg/ml kanamycin. Grow overnight, 30°C (scFv) , shaking.
6) Next day, spin at 600 rcf (faster will crack the plates) for 10 min and use 50 µl supernatant per well for phage ELISA.

2.19 Induction of soluble antibody fragments in 96-well plates

This method is based on that of (De Bellis and Schwartz 1990) and relies on the low levels of glucose present in the starting medium being metabolised by the time the inducer (IPTG) is added.
1) Inoculate 150 µl 2xTY, 100 µg/ml ampicillin, 1% glucose in 96-well plates (Costar 3595) and grow with shaking (270 r.p.m.) overnight at 30°C.
2) Use a 96-well transfer device to transfer small inocula (twice) or pipet 2 µl from this plate to a second 96-well plate containing 100 µl fresh 2 x TY, 100 µg/ml ampicillin, 0.1% glucose per well. Grow at 37°C, shaking, until O.D.600nm is approximately 0.9, (about 2-3 hrs), or alternatively at 30°C for 3-4 hours. To the wells of the master plate, add 50 µl 60% glycerol per well to 15% final, and store at -70°C after use.
3) Add 50 µl 2 x TY, 100 µg/ml ampicillin, 3mM IPTG (final concentration 1 mM IPTG; stock is 1M in water, stored at -20°C. Continue shaking at 30°C for a further 16 to 24 hrs.
4) Spin at 600 rcf for 10 min and use 50 µl supernatant in ELISA.

2.20 ELISA for detection of soluble or phage antibody fragments

1) Coat plate (Falcon 3912) with 100 µl per well of protein antigen used for selections. [10 µg/ml is standard; sometimes more is required, i.e. 3 mg/ml for lysozyme]. Coating is in PBS, and occasionally in 100 mM sodium hydrogen carbonate, pH 9.6. Leave overnight at 4°C. Note for the specificity ELISA, 8 different antigens will be tested.

2) Rinse wells 2x with PBS-Tween-20 (0.1%), 2x with PBS, and block with

120 µl per well of 2% Marvel/PBS (MPBS), for at least 30 min at room temperature. Wash by submersing the plate into buffer and removing the air bubbles in the wells by agitation.

3) Rinse wells 3x with PBS-Tween-20 (0.1%), 3x with PBS, then add 50 µl 4% MPBS to all wells.

4) Add 50 µl culture supernatant containing soluble antibody fragment or phage antibody to the appropriate wells. [Phage supernatants can be concentrated if required by PEG precipitation.] Mix by pipetting up and down, leave appr. 1.5 hrs at room temp.

5) Discard solution, and wash out wells 3x with PBS-Tween-20 (0.1%) and 3x with PBS.

6sol) <u>For soluble antibody detection :</u> Pipette 100 µl of SV5 (Hanke, Szawlowski et al. 1992) antibody in 2% MPBS into each well. Final concentration of SV5 should be in the 1 µg/ml range. Use 1/2 diluted hybridoma supernantant, or appr. 100-fold dilution of 'concentrated' stock. Incubate at room temp. for 1 hr.

6ph) <u>For phage antibody detection :</u> Add 100µl anti fd-phage HRP conjugated (1/5000 dilution in 2% MPBS) to each well. Incubate for 1hr at room temperature.

7) Discard antibody, and wash out wells with 3x with PBS-Tween-20 (0.1%) and 3x with PBS.

8sol) <u>For soluble antibody detection :</u> Pipette 100 µl of 1:1000 dilution (in 2% MPBS) of anti-mouse antibody (peroxidase-conjugated anti-mouse immuno-globulins. Incubate at room temp. for 1 hr.

9) Discard 2nd antibody, and wash wells 3x with PBS-Tween-20 (0.1%) and 3x with PBS.

10) Add 100 ml of TMB (Sigma)

11) Leave in the dark at room temp. for 2-30 min (sometimes longer).

12) Quench by adding 50-100 µl stop solution, 2 N H_2SO_4).

13) Read at 450 nm (TMB).

2.21 Proteins used for selection

Proteins used for selection were kindly provided by Min Park (FLAP endonuclease, Rad52), Scott Peterson (Ku70/80, Cyclin D, cdk2, cdc25A and cdc25C), Tom Peat (PIGS 10 and 12B), Tracy Ruscetti (PARP 85kDa fragment and DNA binding domain) and Michal Novak (tau). Human serum albumin, Guinea pig tTG, β-lactoglobulin were purchased from Sigma. Water insoluble fractions (prolamins) from wheat, rye, barley, oats, rice, maize, millet and soy were obtained by centrifugation of an overnight extract in 70% ethanol (200mg/1ml). The supernatant was recovered and stored at –20 °C. Human tTG was obtained by amplifying cDNA from an intestinal biopsy with specific primers and cloning into pET28.b (Novagen). The tTG was purified by Ni-NTA chromatography using non denaturing conditions (Quiagen).

2.22 Preparation of soluble scFv

1) Phages from individual colonies were used to infect HB2151 (non-suppressor) E. coli strain. Bacteria were grown in 2xYT amp medium at 37°C to 0.5 O.D.
2) induce with 0.1 mM IPTG and incubate at 25°C for an additional 5h.
3) The periplasmic scFv fraction was prepared with osmotic shock. Pelleted bacteria were resuspended in PPB buffer (200mg/ml sucrose, 1mM EDTA, 30 mM TrisHCl pH 8) and leaved on ice for 20 minutes.
4) After centrifugation the supernatant was collected and the cells resuspended in 5mM $MgSO_4$ buffer for 20 minutes
5) The solution was centrifuged and both supernatants were pooled and dialysed against PBS.

2.23 Batch Purification on Ni^{2+} NTA Agarose

All the following steps are carried out at 4°C or on ice.
1) Pre-equilibrate 100 µl 50% slurry of NTA-agarose (Qiagen cat. no. 30210) - add 2 ml of PBS pH 7.5. Spin down at 150 rcf for 1' and resuspend in 2 ml of PBS pH 7.5 containing 2% BSA (filtered through 0.2µM to remove any globs of undisolved BSA). Pre-block NTA agarose for 1-2 hrs with end-over-end mixing. Spin again, wash NTA agarose 2-3 times with 2 ml of PBS pH 7.5 (check OD_{280} of wash to ensure removal of unbound BSA) and resuspend finally in 100 µl of same buffer.

2) To 2 ml periplasmic extract add 100 µl pre-equilibrated NTA-agarose. Mix end-over-end in the cold room for 45'.
3) Spin down NTA-agarose 150 rcf for 1' and resuspend in 2 ml PBS pH 7.5. Repeat spin and resuspension at least twice to remove unbound protein.
4) Spin again 500 rpm for 1' and resuspend in PBS pH 7.5. Either carry out next steps using a BIO-Rad poly-prep chromatography column (cat. no. 731-1550), or use the spinning to wash the column. If a column is used, allow to settle in column for 15', then snap off neck of column and run through excess buffer.
5) Wash column or pellet with 2-4 ml of PBS pH 7.5 checking OD_{280} at end of wash and washing for longer if necessary - allow OD_{280} to drop to 0.001.
6) Wash column or pellet with 2 ml of PBS, 10% glycerol pH 7.5. Again check OD_{280}.
7) Wash column or pellet with 2 ml of PBS, 30mM imidazole, 10% glycerol pH 7.5. Check OD_{280} of wash.
8) Elute scFv with 50 µl PBS, 250mM imidazole, 10% glycerol pH 7.5 . Analyse sample by SDS-PAGE/ western blotting.
9) If required dialyse the sample into PBS, to remove the imidazole (use dialysis or the Centricon system). Add Sodium Azide to 0.05% for storage at 4°C or in aliquots, frozen.

2.24 Competitive ELISA

Prior to conventional ELISA, 100µl of individual scFvs, diluted 1:1 with MPBS, were incubated with increasing amounts of purified α-gliadin (100 nM to 100 µM). ScFv binding was revealed using the SV5 tag (Hanke, Szawlowski et al. 1992) and anti-mouse HRP (Dako).

2.25 Western blotting

Western blotting was performed according to (Sambrook, Fritsch et al. 1989) using soluble scFvs prepared as described above as primary antibody. 10 mg of flour extracts were blotted onto nitrocellulose following SDS PAGE using standard protocols. After blocking with MPBS the nitrocellulose sheets were sequentially incubated with soluble scFv, anti SV5 tag monoclonal antibody supernatant (Hanke, Szawlowski et al. 1992) and anti mouse Ig goat antibodies conjugated with alkaline phosphatase diluted

1:5000 (Dako). The positive bands were revealed by the chromogenic substrates BCIP and NBT (Boehringer).

2.26 Dot blot assay

1) Squares of Immobilon-P membrane (Millipore, IPVH00010) were wetted in 100% methanol for 15 seconds. They were washed with distilled water for 2 minutes and the h-tTG was absorbed as a spot (0.1-1 µg/spot) onto the squares.
2) After 15 minutes, the Immobilon-P free binding sites were blocked with PBS-3% Tween 20 for 30 minutes and then the squares were washed, allowed to dry, and stored at 4°C. Such coated membranes were found to be stable for at least four months.
3) Serum samples were diluted 1:50 in PBS-0.3% Tween 20 and were incubated for 5 minutes at RT. Bacterial supernatant containing anti tTG scFv was used after dialysis in PBS.
4) The squares were washed 3 times with PBS-0.3% Tween 20, and incubated for 5 minutes at RT with phosphatase-conjugated anti-human IgA and IgG immunoglobulines (IgA Sigma A-3062; IgG Sigma A-8542), diluted 1:2500 in PBS-0.3% Tween 20. ScFv were washed in the same way but incubated first with anti SV5 and after washing again with anti mouse phosphatase conjugated.
5) The squares were washed and the immunocomplexes were revealed by substrate solution. The strips were dried and examined for the results, which are expressed as a colorimetric reaction.

2.27 Immunoflorescence on umbilical cord

Serum IgA anti-endomysium antibody was measured by means of the indirect immunofluorescence, using cryostat sections of human umbilical cord. Briefly, the sections were incubated with the subject's serum diluted 1:5, for 30 min. After washing, sections were incubated with fluorescein-labeled goat antihuman IgA antibodies for 30 min. The same analysis was performed using dyalized bacterial supernatant containing scFv ant tTG using anti SV5 as secondary antibody and anti mouse fluorescein labeled to reveal. The slides were washed, mounted in aqueous mounting medium and examined by fluorescent microscopy.

2.28 Affinity determination

scFv dissociation equilibrium constants (K_d) were calculated from the association (k_{on}) and dissociation (k_{off}) rate constants determined using surface plasmon resonance in a BIAcore $^{TM}2000$ instrument (Biacore AB, Uppsala, Sweden). Calculation of Kd values was performed by fitting the data according to a single-site model, using the BIAevaluation 3 software. Approximately 200 RU units of each antigen was immobilised on CM5 sensor chips (BIAcore). ScFv were used at 50-300nM.

2.29 Sequencing

Sequencing was carried out using the Epicentre Sequitherm Excel II kit and analysed using specific labelled primers annealing within the SV5 tag region and the leader sequence. Sequences were analysed on a Likor 4000L automatic sequencer. The identity of the different V genes were analysed by submitting the sequence to VBASE (Tomlinson, Williams et al. 1996). 20 scFvs were sequenced for the primary and 24 for the recombined libraries, 21 different selected scFvs were sequenced, and in the single cell analysis 35 different scFvs were sequenced.

2.30 Induction of epitopes in *E coli* and fixation

E coli 4DJ (Bradbury, Persic et al. 1993) were grown in 2xTY at 30°C to OD 600nm 0.5 IPTG was added to a final concentration of 0.01 mM and after 1 hour induction the bacteria were harvested and washed with PBS. The bacterial pellet was resuspended in 10 volumes of PBS and formaldehyde added to a final concentration of 1.5%. After 90 minutes of incubation with stirring at room temperature, bacteria were washed and inactivated by stirring at 80°C for 5 min. After a final wash in PBS were resuspended in 10 pellet volumes.

2.31 Phage affinity chromatography on fixed *E. coli* cells

A suspension of fixed cell (300 µl equivalent to 30 µl of bacterial pellet) was washed in PBS and saturated in 300 µl MPBS for 1 hour at room temperature, mixing end over end. The fixed cells were then incubated with 300 µl of phage suspension (containing 10^{10} phages) in MPBS for 1 hour at room temperature mixing end over end. After incubation the fixed cells were washed, the phage were eluted and titrated.

Chapter Three:
pDAN New Phage Display Vector

3.1 Introduction

3.1.1 Construction of an improved phage antibody display system.

As described in details in chapter one for a profitable use of recombinant antibody technology and in general of phage display two basic requirements are needed: a) a reliable cloning procedure for the immune repertoire and b) a robust, stable and tight controlled vector.
In this chapter both these needs will be considered, first there will be the description of the design and test of a new set of primers to amplify human V-genes, and second the description of the construction of a new phagemid vector, pDAN. Several phage/phagemid vectors have been designed and successfully used for the creation of antibody libraries. In order to construct an improved vector and design a more efficient cloning procedure the first step in this work was to theoretically analyse several possible variables and subsequently practically test the ones that seem to fit better our model of the phage display system . The main points considered were

1 Antibody form (Single chain Fv (scFv) or Fab)
2 For scFvs: the linker and the order of V regions
3 Restriction enzymes used for cloning V-region
4 Tag sequence

3.1.2 Primers for the amplification of natural V genes

The creation of large V region libraries displayed on filamentous phage (Winter, Griffiths et al. 1994) by V region PCR (Orlandi, Gussow et al. 1989) would be considerably facilitated by a set of V region primers which was able to amplify all rearranged V regions. While the goal of amplifying all possible rearranged V genes, with their associated somatic mutations, may be ambitious, recent sequence information (Tomlinson, Walter et al. 1992; Williams and Winter 1993; Cook, Tomlinson et al. 1994; Cook and Tomlinson 1995; Frippiat, Williams et al. 1995; Mattila, Schugk et al. 1995;

Williams, Frippiat et al. 1996) makes the design of a minimal set of primers able to amplify all germline V gene sequences, and consequently most rearranged V regions, feasible. A number of sets of primers, based on the sequences of V regions at either the DNA level (Larrick, Danielsson et al. 1989; Marks, Hoogenboom et al. 1991; Marks, Tristem et al. 1991; De Boer, Chang et al. 1994; Dziegiel, Nielsen et al. 1995; Watkins, Davis et al. 1995) or the protein level (Welschof, Terness et al. 1995), have been used to amplify V regions and subsequently construct phage antibody libraries. We have found, however, that using computer programs (in silico) these are unable to amplify all V genes, although all these sets clearly give prominent PCR bands when used in amplifications.

Since these primers were designed, the sequences of all functional germline V regions have been catalogued into an easily accessible and user friendly database, V BASE (Tomlinson, Williams et al. 1996). Using this database we have designed a small set of primers which is able to recognise 100% of functional V-genes with high efficiency. This compares very favorably with the lower percentages picked up by the complete primer sets previously reported which did not have access to this database. We have tested our primers one by one in PCRs using cDNA derived from both adult and umbilical cord lymphocytes and show that all (bar one) are functional, produce minimal primer dimer formation and give the amplification patterns which correlate with the known V gene expression of these two diverse V region sources

Mouse antibodies can also be displayed on phage and a series of primers which will pick up almost all mouse V genes has been described (Krebber, Bornhauser et al. 1997). These are likely to be very useful for the cloning of hybridoma V regions and for making antibody libraries from immunised mice.

3.2 Results

3.2.1 Primer design strategy

A series of initial oligos were designed and screened against the functional V regions using the MacVector program. They were then modified by trial and error until all V-genes were recognized with at least 16 bp homology at the 3' end (see Table 3.1). Degenerate nucleotides were introduced at no more than three positions to extend the number of sequences recognized by a single primer, with all primer preparations (VK9back being the only exception), containing no more than a total of eight different variants. Degenerate nucleotides were introduced only if the total number of V genes recognised by a primer preparation exceeded 50% of the total number of primer variants in that preparation. New primers were also designed to amplify some V gene families which remained unrecognised (e.g. VH14 back for the VH2 family and VL4back for the Vλ4 and Vλ5 families). The purpose was not to design family specific primers, but a complete set of primers which is able to amplify all V genes. Inevitably, however, there is some correspondence between primers and V gene families: on the whole each back primer amplifies one specific family, or a subset of a specific family, although a couple (VH22, which recognises families 1, 3, 5; and VK9, which recognises families 2, 3, 4, 6) are more diverse. The situation is more complex for the Vλ primers which are more heterogeneous and especially family Vλ 1 which requires 3 different primers for 5 members. These theoretical gene specific amplifications are based on the stringent criteria described above. However, we expect each primer to amplify more V genes than we have indicated: e.g. primer VK1back will probably also amplify VK4 (21 of 23 bases being matched) although primer VK9back should amplify it better, being perfectly matched.

The VHfor primers contain an additional degeneration in primer VH1/2for at aa position 110, which allows the single primer to amplify both J regions.

By introducing a degeneration at position 104, the number of VKfor primers was reduced by one. The two Vlfor primers were designed ex novo, one (VL1/2for) for J genes 1, 2, 3 and the other (VL7for) (Bauer and Blomberg 1991) for the J 7 gene. The previously described Vλ 4/5 primers were not included as they recognize pseudogenes (Dariavich, Lefranc et al. 1987), a result we have been able to confirm by being unable to obtain amplification from any V gene source using this primer.

PCR primers back (A) and for (B) for human V-genes chain

(A) back primers (5')

VH	**Oligo name**	**V gene family recognised**
CAG GTG CAG CTG CAG GAG TCS G	VH4back	4
CAG GTA CAG CTG CAG CAG TCA	VH5back	6
CAG GTG CAG CTA CAG CAG TGG G	VH6back	4 (DP63)
GAG GTG CAG CTG KTG GAG WCY	VH10back	3
CAG GTC CAG CTK GTR CAG TCT GG	VH12back	1
CAG RTC ACC TTG AAG GAG TCT G	VH14back	2
CAG GTG CAG CTG GTG SAR TCT GG	VH22back	1, 3, 5, 7
Vλ		
CAG TCT GTS BTG ACG CAG CCG CC	VL1back	1
TCC TAT GWG CTG ACW CAG CCA C	VL3back	3
TCC TAT GAG CTG AYR CAG CYA CC	VL38back	3
CAG CCT GTG CTG ACT CAR YC	VL4back	1, 4, 5, 9
CAG DCT GTG GTG ACY CAG GAG CC	VL7/8back	7, 8
CAG CCW GKG CTG ACT CAG CCM CC	VL9back	1, 5, 9, 10
TCC TCT GAG CTG AST CAG GAS CC	VL11back	3 (DPL16)
CAG TCT GYY CTG AYT CAG CCT	VL13back	2
AAT TTT ATG CTG ACT CAG CCC C	VL15back	6
Vκ		
GAC ATC CRG DTG ACC CAG TCT CC	VK1back	1
GAA ATT GTR WTG ACR CAG TCT CC	VK2backts	3, 6
GAT ATT GTG MTG ACB CAG WCT CC	VK9back	2, 3, 4, 6
GAA ACG ACA CTC ACG CAG TCT C	VK12back	5

(B) for primers (3')

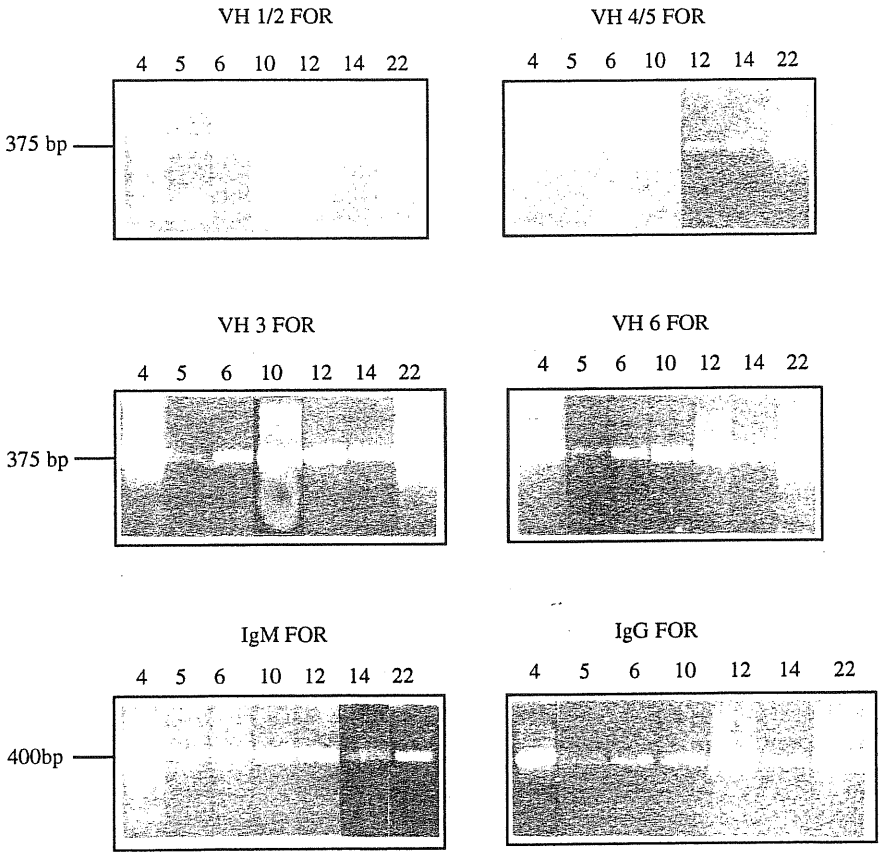
VH		
TGA GGA GAC RGT GAC CAG GGT G	VH1/2for	JH1, JH2
TGA GGA GAC GGT GAC CAG GGT T	VH4/5for	JH4, JH5
TGA AGA GAC GGT GAC CAT TGT	VH3for	JH3
TGA GGA GAC GGT GAC CGT GGT CC	VH6for	JH6
GGT TGG GGC GGA TGC ACT CC	IGMfor	CH1 Cμ
SGA TGG GCC CTT GGT GGA RGC	IGGfor	CH1 Cg
Vλ		
TAG GAC GGT SAS CTT GGT CC	VL1/2for	Jl1, Jl2, Jl3
GAG GAC GGT CAG CTG GGT GC	VL7for	Jl7
Vκ		
TTT GAT TTC CAC CTT GGT CC	VK1for	Jk1
TTT GAT CTC CAS CTT GGT CC	VK2/4for	Jk2, Jk4
TTT GAT ATC CAC TTT GGT CC	VK3for	Jk3
TTT AAT CTC CAG TCG TGT CC	VK5for	Jk5

Tab 3.1 The sequences of the primers used, their names and the V gene families they recognise are indicated

3.2.2 PCR analysis of VH repertoire

These oligos were then tested for their ability to amplify human V-regions. We used cDNA derived from two different sources: PBLs from 3 healthy donors and PBLs derived from 40 umbilical cord blood samples. Although there are fifty one different V region genes, their frequency of use in different lymphoid tissues is not random. For example, one member of the VH3 family, VH26, is found in 15% of cord blood cDNA clones (Mortari, Newton et al. 1992; Mortari, J.Y. et al. 1993)and 6-10% of unselected adult cDNA clones (Stewart, Huang et al. 1993) rather than the 2% which would be expected. In general, these V gene sources express two different overlapping repertoires of V genes, with the fetal repertoire having an over-representation of VH6, VH4 and some members of the VH3 families (Schroeder, Hillson et al. 1987; Mortari, J.Y. et al. 1993), while the adult repertoire, although being more variable between different individuals, has a preponderance of similar VH3 family members and a reduction of VH6 (with the general order VH3>VH5>VH2>VH1>VH4>VH6 (Davidkova, Pettersson et al. 1997) or VH3>VH4>VH1>VH5>VH6>VH2 (Guigou, Cuisinier et al. 1990). All possible combinations of back and for primers for VH, VK and Vl were individually tested for PCR amplification and were shown to be functional (figs 3.1 and 3.2). The intensity of the PCR bands obtained with these primers in general reflects what is known about the number of V genes recognized by each primer and to the different repertoire of V-genes usage of adult PBLs and cord blood lymphocytes, although it cannot be excluded that part of this variation may be due to differing abilities of primers to amplify their respective templates. The primer VH5back (which amplifies the single VH6 gene), for example, gives stronger bands in cord blood samples than in adult, reflecting the fact that VH6 is five fold over expressed in cord blood relative to adult blood (Van Es, Raaphorst et al. 1993). Similarly, VH4back (which amplifies VH4) gives stronger bands with most cord samples than adult, again reflecting the greater expression in cord blood, while the primers VH10back and VH22 back (which amplify VH3) give very strong bands in both cord and adult samples, confirming the high percentage of this gene family in both samples (Davidkova, Pettersson et al. 1997).

a)



b)

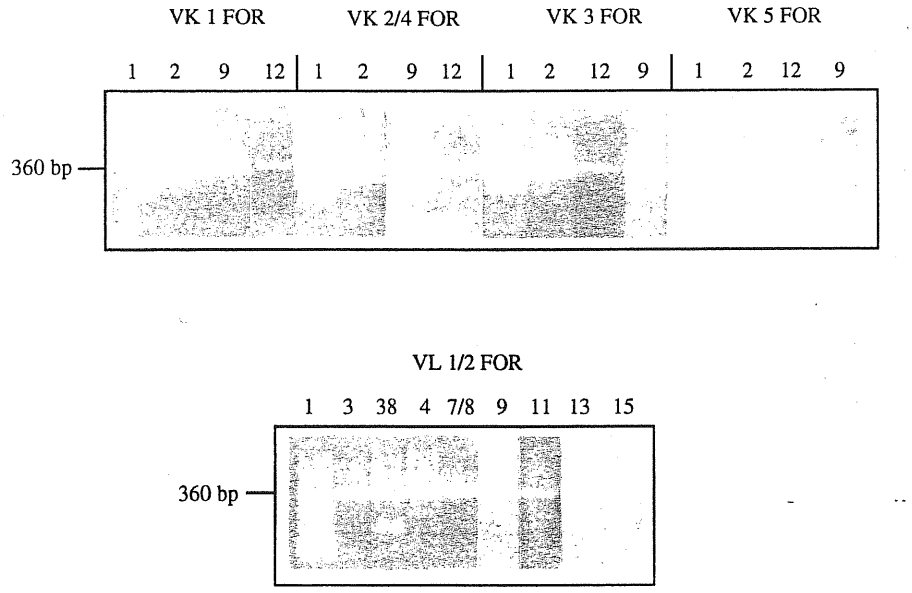
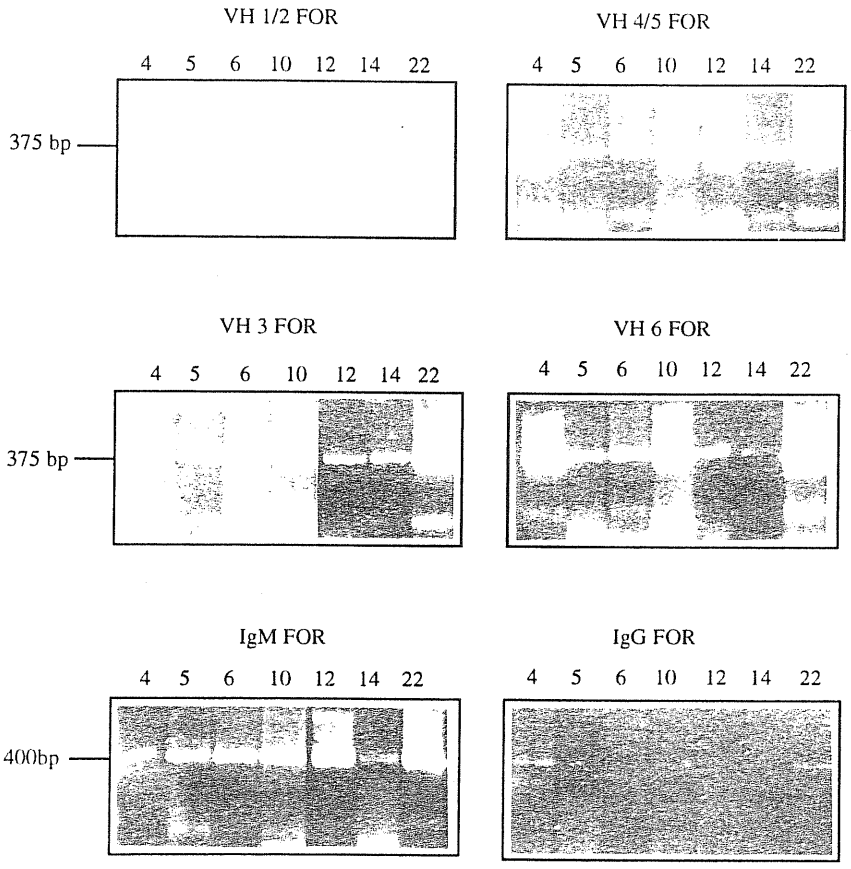


Fig 3.1 PCR amplification of adult PBL blood V genes

All possible back and for primer combinations were used in separate PCR reactions. The for (3') primer is indicated above each gel, while the back (5') primer is indicated above each lane. The size of the expected PCR product is indicated. A) VH primers, B) VK and Vλ primers.

a)



b)

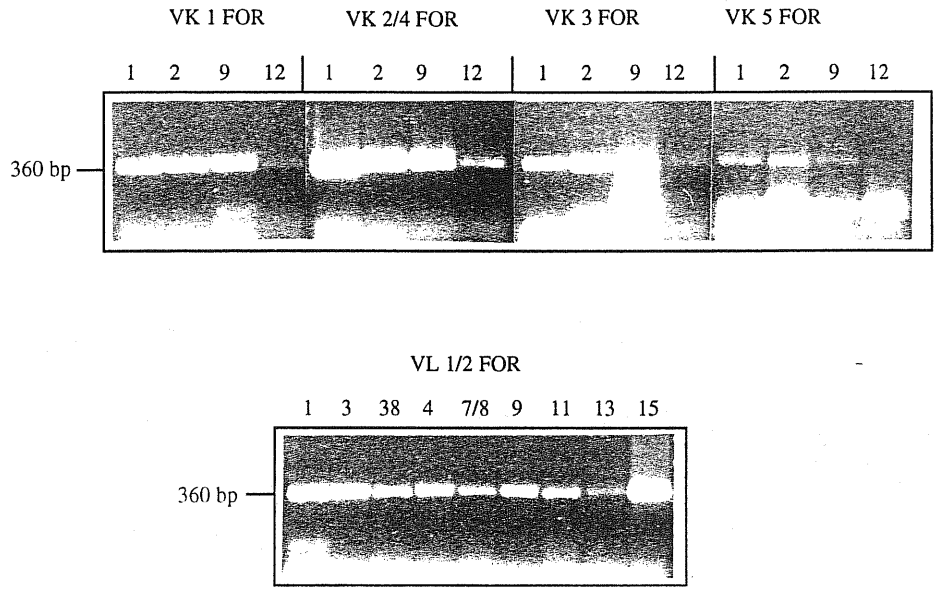


Fig 3.2 PCR amplification of cord blood lymphocyte V genes
See legend to figure 3.1.

3.2.3 PCR analysis of V light repertoire

There are no clear differences between the amplification of cord and adult samples using the VK primers. The VK12back primer which was designed to amplify the single member VK5 family gives reduced amplification in both cord and adult samples, probably reflecting low expression of this gene.

The Vλ primers also give similar intensity bands with a few exceptions: VL1back, which is specific for two members of the Vλ1 family, and VL15back, which is specific for the Vλ6 family, both give stronger signals in cord blood than adult, while VL13back, which amplifies the Vλ2 gene family, gives a stronger signal in adult than cord. Although repeated several times we were unable to obtain any amplification with the VL7for primer which was designed to amplify the recently described Vλ7 gene (Bauer and Blomberg 1991) (data not shown). This probably reflects the known low level of expression of this gene (Bauer and Blomberg 1991).

3.2.4 Construction of pDAN: a new phagemid display vector.

Based partially on previous vector structures and on a series of theoretical hypotheses a new phagemid vector named pDAN was designed. Several variables were tested both simultaneously and singly in different vector constructs in order to select the best one for the costruction of the final vector to use as a standard for our antibody library.

The well known phagemid pUC119 was used as backbone vector. The first step was the creation of a new polylinker sequence. This was achieved by partial annealing and extension of two long oligonucleotides: polylinker back and For (seeTab 2.1). The resulting DNA fragment was cloned after digestion with Hind3 and NotI in pUC119.

The polylinker as can be see in fig 3.3 has the following features: a S/D sequence followed by a leader sequence based on the PelB sequence (Lei, Lin et al. 1987), which was modified at the C terminal according to the signal sequence matrices (von Heijne 1986) in order to be more easily cleaved by reducing the number of charged amino acids known to alter the export of the protein (Li, Beckwith et al. 1988; Boyd and Beckwith 1990; Johansson, Nilsson et al. 1993) and to be compatible with the restriction enzymes used for cloning the scFv. After the leader sequence a polycloning cassette was inserted. This includes besides the BssHII - NheI sites, that are used to clone the scFvs, an NruI site. A frequently phenomenon is the contamination of antibody libraries with uncut recipient vector (Courtney,

Williams et al. 1995; Johansen, Albrechtsen et al. 1995). Normally antibody free vectors have a growth advantage over scFv encoding ones (de Bruin, Spelt et al. 1999) and cause problems during enrichment of antigen-binding antibodies sequences. This problem can partially be solved by including the sequence of a restriction enzyme in the middle of the polycloning cassette and by adding the enzyme in the ligation reaction. This procedure will dramatically reduce the background that could be due to uncut vector. In fact, the vector will be linearized and so its transformation efficiency will be greatly reduced. The enzyme NruI was selected because is an enzyme which cuts rarely in V genes and was shown to cut very well even in the ligation buffer, and the cut vector is almost impossible to religate.

```
AAGCTTGCCAAATTCTATTTCA AGGAGAC AGTCATA
HindIII                 S/D

 M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  S    G  A    H  A
ATGAAATACCTATTGCCTACGGCAGCCGCTGGATTGTTATTACTCGCAGCAAGC G GCGCG CATGCC
PelB leader                                                BssHII  end leader

 S  G     V  T  V  S  S
TCCGGA G GTCACCGTCTCCTCA
 NruI   BstEII  (end of JH)
   frame shift

 A  S
GCTAGC
NheI

 G  K  P  I  P  N  P  L  L  G  L  D   S  T
GGCAAACCAATCCCAAACCCACTGCTGGGCCTGGAT AGTACT
SV5 tag                               ScaI

 H  H  H  H  H  H  (Q)
CACCATCACCATCACCAT TAG
His6               amber

 A   A   A   T   V   E   S   C   L   A
GCG GCC GCT ACT GTT GAA AGT TGT TTA GC...
  NotI      | gene 3 start

  V  F  S  T  F  A  N  I  L   R   N   K   E   S  stop
TGTATTTTCGACGTTTGCTAACATACTG CGT AAT AAG GAG TCT TAA GCA TGC GAA TTC
......geneIII.................................AflII  SphI  EcoRI
```

Fig 3.3 Structure of the polylinker and gene 3 fragment cloned in the pUC119 vector.

The polylinker sequence is also out of frame respect to the gene 3 protein and so the vector will not produce p3 unless a scFv is cloned. After the NheI cloning sites there is the SV5 tag (Hanke, Szawlowski et al. 1992) followed by the His6 tag and the amber stop codon. The vector was completed by cloning the gene 3 sequence (amplified from fdtet) as a NotI - EcoRI PCR fragment after the stop codon. The sequence of the polycloning cassette was confirmed by sequencing on both strands.

3.2.5 Restriction enzymes and Tags

A number of tags have already been tested and used in phage display (or expression) vectors, after a series of considerations we selected a 12 amino acid sequence known as SV5 (Hanke, Szawlowski et al. 1992) for several reasons. It can be used at N-terminal, C-terminal and in the middle of proteins; there is a very well characterised mAb available that has high affinity and could be used for Western blotting, immunoprecipitation and immunofluorescence, but most importantly the epitope, deriving from a simian virus protein, give no crossreactivity with eucaryotic or bacterial proteins. This is, together with the low affinity, one of the limiting factors of the widely used 9E10 anti myc mAb

The polylinker also includes the His6 tag, useful for immobilized metal affinity chromatography (IMAC) purification of the soluble scFvs that can be produced thanks to the presence of an amber stop codon.

The restriction enzymes used as cloning sites in the vector were selected after a complete analysis of the frequency of cutting in the V-regions of almost all the 6bp or more cutters. A limited set of enzymes was subsequently obtained (see tab 3.2) and from this list 4 enzymes were selected: BssHII and SalI for the cloning of the VL regions and XhoI plus NheI as cloning sequences for the VH regions. Besides having low cutting frequencies these enzymes were chosen because each pair could work in the same restriction buffer allowing a single step digestion, so reducing the possibility of losing material during the purification procedures.
An additional enzymes site include ScaI which is found between SV5 and the His6 tag.

Enzyme	Recognition site	337 V_H	80 D/JH	168 VL	12 JL
*Afl*II	C/TTAAG	0	0	2	0
*Age*I	A/CCGGT	0	0	8	0
*Asc*I	GG/CGCGCC	0	0	0	0
*BspE*I	T/CCGGA	12	0	2	0
***Bss*HII**	**G/CGCGC**	**0**	**0**	**0**	**0**
*Bst*BI	TT/CGAA	1	0	0	0
*Mlu*I	A/CGCGT	1	0	0	0
***Nhe*I**	**G/CTAGC**	**0**	**0**	**0**	**0**
*Nru*I	TCG/CGA	0	0	0	0
*Pac*I	TTAAT/TAA	0	0	0	0
*Pml*I	CAC/GTG	1	0	0	0
***Sal*I**	**G/TCGAC**	**0**	**0**	**0**	**0**
*Sfi*I	GGCCN4/NGGCC	0	0	0	0
*Sna*BI	TAC/GTA	0	0	0	0
*Spl*I	C/GTACG	2	0	1	0
*Srf*I	GCCC/GGGC	0	0	1	0
***Xho*I**	**C/TCGAG**	**0**	**0**	**2**	**0**

Tab 3.2: Frequency of rare restriction sites in human germline V-genes (Tomlinson, Williams et al. 1996). In bold are signed the four enzymes selected for cloning the scFvs.

3.2.6 Construction of 3 different linker sequences

The functionality of the new pDAN vector was tested by cloning and expressing several well characterised mAb. These V-regions were first assembled and then cloned in the scFv format. As the linker sequence is one of the key features to achieve good expression and a functional scFv we evaluated the behaviour of 3 different linker sequences that have different and well defined characteristics.

The first is the classical 15 aa sequence (Bird, Hardman et al. 1988) that is composed by tree repeats of the sequence (Gly4-Ser) and is so called GS. We modified the original linker DNA sequence in order to have the three repeats encoded by different codons to avoid the occurrence of incorrect overlaps during assembly PCR and possible recombination problems due to repetitions in the sequence. The sequence of the BspEI site (that is a relatively rare cutter) was also included to allow, eventually, the separate cloning of the VH and VL sequence.

The second linker is a modified version of the so called 218 (Whitlow, Bell et al. 1993) that after our changes we call 220. The original linker is 18 aa long and has a reduced aggregation tendency and is stable to proteolysis. The sequence was partially modified: first length was increased to 20 aa by adding at the N-terminus glycine and serine residues this, besides increasing the length, was done to introduce the sequence for the restriction sites used for cloning (BspEI and SalI). The C-terminus was also modified changing the last 3 residues for the same reason. The new residues introduced were Gly and Ser and Thr which provide a general flexibility as well as incorporating the restriction enzymes sites used (KpnI and XhoI).

The third linker used is based on the sequence of the loxP recombination site. To design this linker all the possible 6 frames available for the wild type *loxP* site and the mutated *loxP511*site (which will not recombine with the wild type *loxP* (Hoess, Wierzbicki et al. 1986)) were analysed (see fig 3.4). From the analysis a translation of *loxP511* (ITSYNVYYTKL) was identified which had only a single basic amino acid (to reduce the possibility of proteolysis), lacked stop codons and was the least hydrophobic. This sequence is different from that used in the work of (Tsurushita, Fu et al. 1996) that was based on loxP511 but with a reverse sequence.

The original 11 amino acids length of the linker was extended to 21 aa by adding 5 residues at both ends. As for the 220 linker the sequence was designed to encode Gly-Ser-Thr residues and to include the sequence of the restriction enzymes sites. The sequence of the linker as used in the scFvs is given in figure 3.5

loxP511

```
ATA ACT TCG TAT AAT GTA TAC TAT ACG AAG TTA T
TAT TGA AGC ATA TTA CAT ATG ATA TGC TTC AAT A
I   T   S   Y   N   V   Y   Y   T   K   L   X>
 *   L   R   I   M   Y   T   I   R   S   Y>
  N   F   V   *   C   I   L   Y   E   V   X>

   <L   K   T   Y   H   I   S   Y   S   T   I
 <Y   S   R   I   I   Y   V   I   R   L   *
   <V   E   Y   L   T   Y   *   V   F   N
```

loxPWT

```
ATA ACT TCG TAT AAT GTA TGC TAT ACG AAG TTA T
TAT TGA AGC ATA TTA CAT ACG ATA TGC TTC AAT A
I   T   S   Y   N   V   C   Y   T   K   L   X>
 *   L   R   I   M   Y   A   I   R   S   Y>
  N   F   V   *   C   M   L   Y   E   V   X>

   <L   K   T   Y   H   I   S   Y   S   T   I
 <Y   S   R   I   I   Y   A   I   R   L   *
    <V   E   Y   L   T   H   *   V   F   N
```

Fig 3.4 Analysis of the possible frames of the translated *loxP511* and *loxPWT*, * indicate a stop codon. The nucleotide difference between the two sequence is indicated in bold.

Gly-Ser linker

```
 G  G  G  G  S  G  G  G  G  S  G  G  G  G  S
GGCGGTGGAGGCAGCGGCGGTGGCGGCTCCGGAGGCGGTGGCAGC
                            BspEI
```

220 linker

```
 S  G  G  S  T  S  G  S  G  K  P  G  S  G  E  G  S  S  G  T
TCCGGAGGGTCGACCAGCGGTTCTGGGAAACCAGGTTCCGGTGAAGGCTCGAGCGGTACC
 BspEI   SalI                                    XhoI   Kpn I
```

lox511 linker

```
 S  G  G  S  T   I  T  S  Y  N  V  Y  Y  T  K  L   S  S  S  G  T
TCCGGAGGGTCGACC ATAACTTCGTATAATGTATACTATACGAAGTTAT CCTCGAGCGGTACC
 BspEI   SalI                loxP511                XhoI   KpnI
```

Fig 3.5. The DNA and deduced amino acid sequences of the three linkers used to create mini-libraries is indicated. Restriction sites used for cloning V regions are indicated below the DNA sequence. The sequence of the loxP511 recognition site is in italics, restriction sites used for cloning are underlined, and the inverted repeats in the loxP511 recognition sequence are double underlined. Antibodies V-regions were cloned in the order VL/VH.

3.2.7 Display of functional scFvs

The mAbs selected to test the different linker variants were: D1.3: anti lysozyme (Mariuzza, Jankovic et al. 1983), Y13-259 (Furth, Davis et al. 1982) anti P21ras and GL30 (Sblattero, Not et al. 1999) an anti A-gliadin antibody (see chapter 4).

The variable regions of the first two were amplified with the mouse specific primers (Krebber, Bornhauser et al. 1997) from mAbs previously cloned as scFvs, the GL30 was instead cloned directly from the hybridoma cell line.

The scFvs were assembled and cloned as BssHII-NheI fragment to give a final vector as seen in fig 3.6.

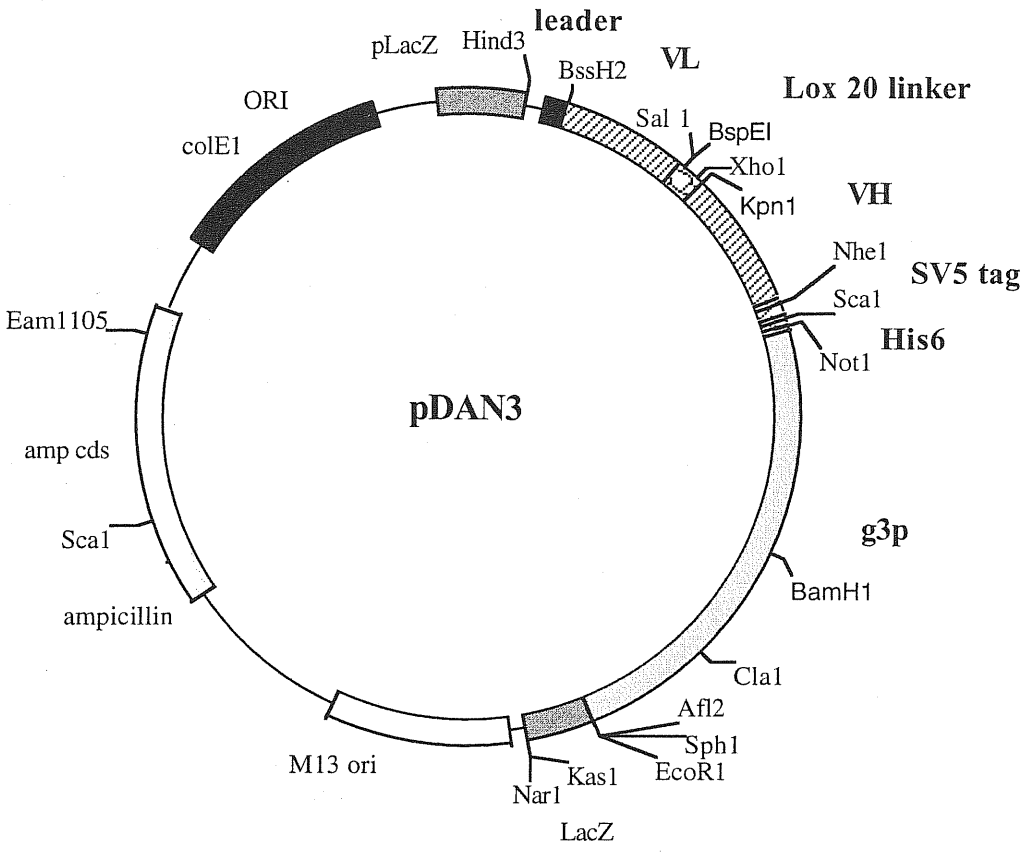


Fig 3.6 Plasmid map of the Vector pDAN3. In Bold are marked the leader sequence, the VL linker VH sequences, the tags SV5 and His6 and the gene 3

After cloning positive colonies were sequenced to confirm homology with the parental mAb. Phagemid particles were produced (in DH5αF') from all three antibody clones using M13KO7 as helper phage and and the presence of the recombinant protein was tested by Western blotting and ELISA. All the scFv express the correct band in WB revealed both with an anti gene 3 mAb (Tesar, Beckmann et al. 1995) and anti SV5 mAb (Hanke, Szawlowski et al. 1992). These results confirm the correct processing of the leader sequence, and the functionality of the new tag sequence, both at the C-terminal when the scFv was produced as a soluble scFv protein as well as in

the middle of the scFv-gene3 fusion protein when incorporated in the phagemid particle.

When the phages were tested in ELISA all the clones showed good signals against the related protein and low background against unrelated control proteins. This confirms that, besides being expressed, the scFv is also functional when displayed on the phage. scFvs with the lox linker pDAN-D1.3LOX and pDAN-Y13LOX were compared with the original vector pHEN1, which contains a Gly-Ser linker. In this case the effect of inversion of the V regions order from VH-VL in pHEN to VL-VH in pDAN and the presence of a different linker were tested simultaneously.
For the D1.3 scFv (see fig 3.7) the pDAN scFv give a signal 3 times higher signal than that of pHEN vector. In the case of Y13-259 both ELISA and living column experiments showed positive signals for pDAN although lower than the original pHEN1 vector see fig 3.8

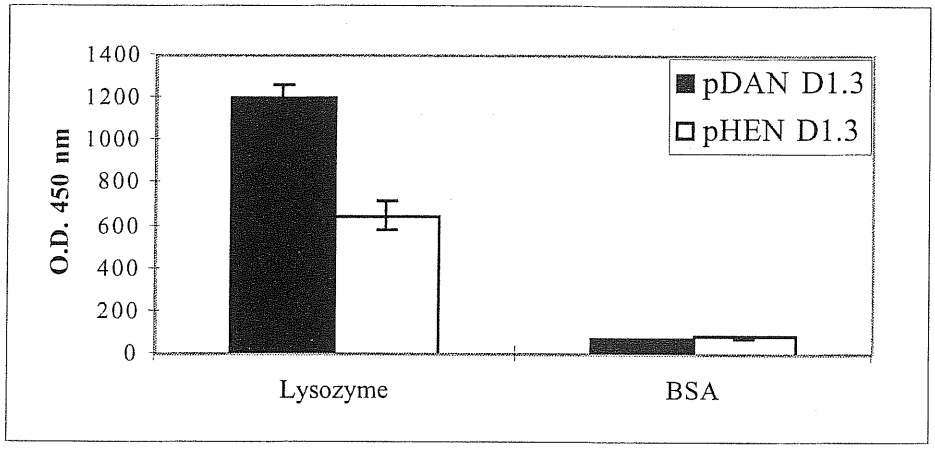
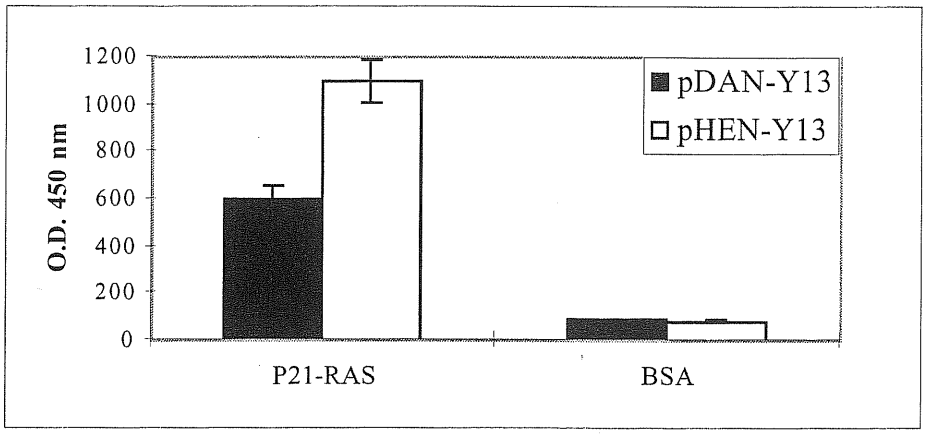


Fig 3.7 Comparison by phage ELISA between pDAN and pHEN1. Phage particles from bacterial supernatant from pDAN-D1.3 with the lox linker and pHEN1-D1.3 with the Gly-Ser linker were compared in ELISA. Background binding was determined against the unrelated antigen BSA

a) phage ELISA



b) living column

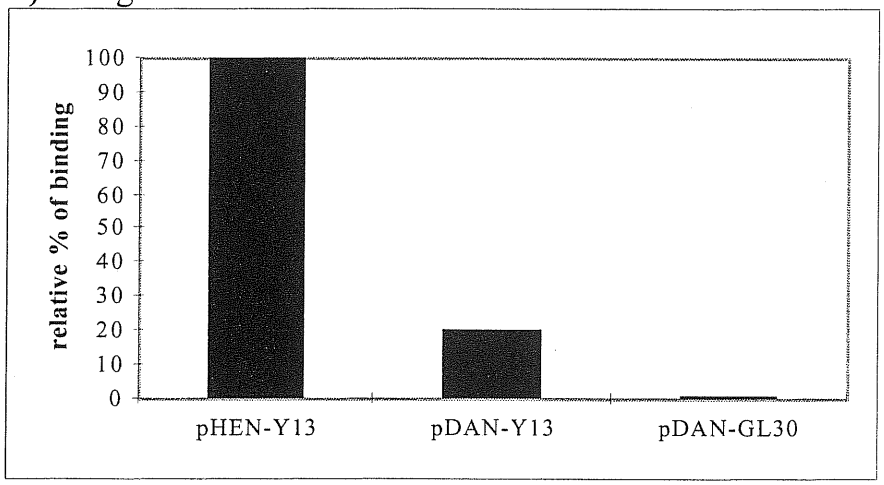


Fig 3.8 Comparison between pDAN-Y13-lox and pHEN1-Y13 in **a)** Phage ELISA results BSA-P21ras was used as positive control, BSA as negative control for the determination of the unspecific binding; in **b)** Living column model selection results (see chapter 2 for details), GL30 (anti αgliadin scFv) was used as negative control.

3.2.8 Creation of 3 minilibraries with the different linkers

The three linkers described above when tested with single and well characterised mAb were functional although with marked differences between each other. In order to evaluate the behaviour of the linkers not just with a single scFv but with a population of different V-genes small scFv libraries were created with each of the three linkers. cDNA was prepared from a pool of PBL of three healthy donors. The libraries were constructed not with the complete set of primers but just amplifying the VH regions with the primer VH12 that is specific for the VHI family and the VL region with the primer VK1 that is specific for the VKI family. The scFvs were cloned

through a two step cloning procedure, first the VLs were cloned and a small library was created, this was used as recipient vector for the VH cloning to obtain the final libraries consisting of 10^5 clones. The diversity was confirmed by DNA fingerprinting of 15-20 clones for each library. Phages were then produced by growing the culture at different temperature (30°C and 37°C). The display levels were assessed by Western blots developed with the anti SV5 antibody and anti gene 3 protein. Comparing the results the *loxP511* linker showed display levels as good as, or even better, than the two other more widely used linkers, see fig 3.9. GS and 220 linkers give a very bad display level with almost no full length protein visible for 220. Analysing the *lox P511* linker the ratio between the intensity of the bands of full length protein and p3 protein show that roughly the display level could be estimated to be at least 10% of the total protein present that is lower than what we have found with monoclonal (e.g. D1.3) but in accordance with what was found by others (Clackson and Wells 1994). Moreover the display rate for the *lox P511* linker remain almost the same growing at 37°C or 30°C, this was not true for the other two linkers that have an evident decrease in the level of display at 37°C (data not shown).

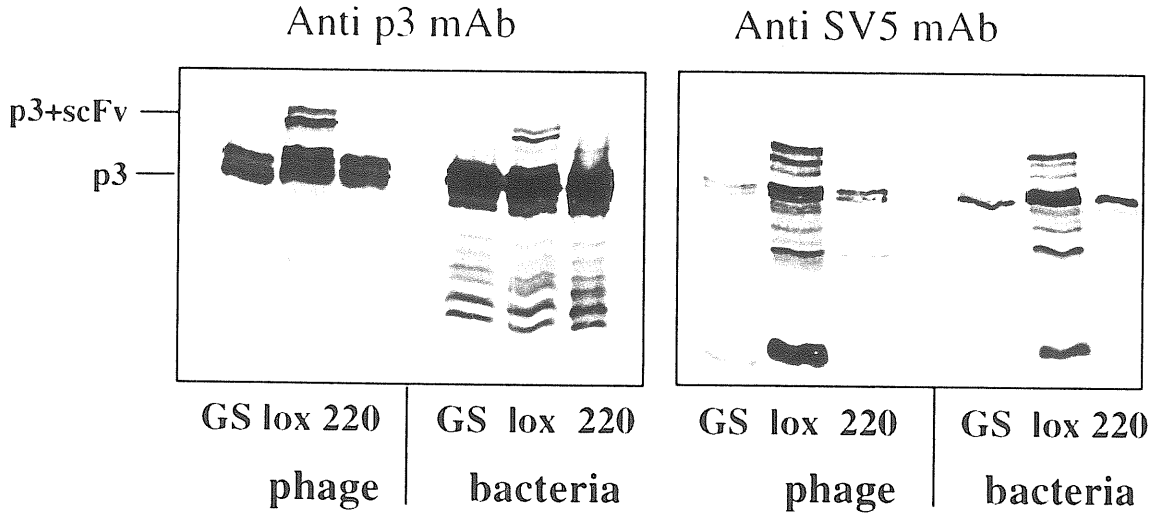


Fig 3.9 Comparison for the display level of the recombinat g3p-scFv protein with three different linkers. Left side: Phagemid particles were produced O/N at 30 °C, Peg precipitated and a comparable number was loaded for the three linkers. Right side: 10 µl of bacteria cells with the three different phagemids, after O/N growth at 30 °C, were loaded. The western was reavealed with the anti p3 antibody and with the anti SV5.

3.3 Discussion

3.3.1 design of a new set of PCR primers

While satisfactory phage antibody libraries can be made using the published primers, it is clear that a library will be more diverse the more different V-genes are included. Previous primer sets analysed using our criteria (16 bases homology at the 3' end) recognized 59.5% (Marks), 62.8% (Welschof), 62.8% (Watkins) 37% (de Boer) and 86% (Dziegiel that use the highest number of primers almost double compared to the other sets) of all the V-genes (VH, VK and Vλ). By relaxing the criteria (12 bases homology at the 3' end), these percentages could be increased by up to 5% in some cases. Analyzing the three gene sets separately there are some interesting differences.

The VH back primers have a high percentage of recognition, 74.5% (Marks), 80.3% (Welschof), 78.4% (Watkins) 41% (de Boer) and 98% (Dziegiel) but some V-genes used at high frequencies (Tomlinson, Walter et al. 1992; Griffiths, Williams et al. 1994) are not picked up. DP47, for example, is picked up only by Welschof and Dziegiel (with a specifically designed primers) while DP14 is picked up using the stringent criteria by Marks, and by all except Welschof using the relaxed criteria. We have designed a specific oligo (VH6back) for DP63, this being the third most commonly used V gene in the repertoire of naive antibodies, and picked up only by de Boer, Watkins and Dziegiel.

The percentage of recognition is lower for the VK back primers 57.5% (Marks), 54.9% (Welschof), 47.5% (Watkins) 60% (de Boer) and 90% (Dziegiel). In this case the frequently used V-gene DPK21 is missed by all but Dziegiel.

The percentage recognizing Vλ genes was particulary low 36.6% (Marks), 43.3% (Welschof), 56.7% (Watkins) and 60% (Dziegiel) - de Boer did not present λ specific primers) with some families not picked up at all (e.g. 8 and 10 or 7 and 9 for Dziegiel), and some high frequency Vλ genes also excluded (eg DLP23 for Marks and Watkins).

By choosing a stringent test of primability (12 or 16 bp at the 3' end) we may well have underestimated the ability of the other primer sets to amplify sub-optimal matches. However, the purpose of our investigation was to create a primer set which was able to optimally amplify all V genes, and in choosing this stringent test of primability we feel that we have done this. It is true that the Marks primers are able to amplify DP47 from genomic DNA

(Hoogenboom and Winter 1992), however, we feel that the full diversity of DP47 based V genes is more likely to be amplified by a perfectly matched primer rather than one which has a mismatch 11 bases from the 3' end.

3.3.2 Designing of a new phage display vector

Simultaneously with the design of the extended set of primers an improved phage display vector - pDAN was constructed. This vector together with the cloning procedure proved to be robust and reliable, first allowing the cloning and expression in a functional way of several hybridoma V-regions, and then in a library set up experiment. All the variables described in Tab 1.3 were first considered theoretically and the choices made were tested practically with the new vector.

3.3.3 Antibody form: stability, secretion and solubility

The first choice made in the new vector regards the antibody form to use: as described in the introduction in chapter 1, there are essentially two possible choices: scFv and Fab. After consideration about the pros and cons the scFv format was selected. The use of Fabs shows difficulties in assembly, high degradation, low yields of soluble protein and DNA instability. On the other hand, scFvs also presents drawbacks. We analysed the problems of secretion and solubility of the scFv in detail as these are essential prerequisites for the function of the protein. These problems were reduced by introducing some changes with respect to the previous vectors: first, a partially new leader sequence based on the PelB sequence was constructed to maximize secretion into the periplasmic space and the cleavage probability of the leader peptide from the fusion protein. Secretion into the periplasmic space is controlled by, in addition to the leader sequence, the first few amino acids of the secreted protein. Charged residues have been shown to inhibit the correct secretion of protein (Li, Beckwith et al. 1988; Boyd and Beckwith 1990; Ayala, Balint et al. 1995). For historical reasons all scFvs have been assembled in the VH-VL format, but there is evidence (Ayala, Balint et al. 1995; Luo, Mah et al. 1995) which suggests that the reverse orientation is a better choice, mainly for the higher solubility of the light chain. Considering also that light chains also have fewer charged residues in the first amino acids we decide to reverse the order of the scFv to the VL-linker-VH. These changes, with the choice of linker (see below)

seem to give display levels which are at least as good as, if not better, than more "traditional" pHEN1 display vector.

3.3.4 scFvs linker sequence

The main problem of scFvs is the tendency to form dimers and higher molecular weight forms. This is probably the result of an intermolecular association of VHs and VLs due to the linker peptide which do not allow pairing of the domains on the same chain. The linker sequence should be long enough to span the 35Å gap between the C-terminus of the VL and the N-terminus of the VH. It must also provide flexibility, for this reason the linker have an underlying sequence of alternating gly and ser residues. Several studies have demonstrated that these problems of dimer formation could be partially overcome by the use of longer linker. As a reference the classical 15 amino acids (Gly4-Ser) linker was used. Then in order to test the effect of linker length the well characterised 218 linker was selected as base sequence and was then modified and the length extended to 20 amino acids. The display level obtained was much better compared to the GS linker. Recently, the use of a recombination signal (*loxP*) as a scFv linker and the use of the *cre/lox* recombination system to switch variable regions indicates that, at least in principle, it should be possible to make libraries using recombination with scFvs (Tsurushita, Fu et al. 1996). To verify the possibility to exploit the recombination for the creation of large libraries a detailed analysis of the *loxP* sequence was performed. A new sequence, corresponding in part to one of the possible translations of the *loxP511* recombination sequence plus 5 amino acids at both ends was tested. The new linker is 21 aa long, has reduced proteolytic susceptibility, gave a display level comparable or even better than that of the 220 linker and all scFv tested were revealed to be functional. These results opened the way for the creation of libraries that will be described in detail in chapter 4 and 5.

3.3.5 Tags and restriction enzymes

Recombinant scFvs have been selected against almost all kind of peptides, proteins, and haptens. Besides the consideration regarding the affinity of the selected antibody, one of the limiting factor for their use was the lacking of an epitope tag which could be used in all applications, from western blotting to immunoprecipitation to intracellular antibodies. The most commonly used Myc epitope and the mAb which recognize it 9E10, suffer,

for example, from crossreactivity with the intracellular endogenous myc protein that limits its use (Fan, Villegas et al. 1998). A bibliographic study of the different tags which have been used, revealed a paper (Hirst, Fisher et al. 1994) which directly compared three commonly used linkers: 9E10, SV5 and HA. In this study, SV5 was found to be superior by a number of criteria including: no cross reaction with bacterial or eukaryotic proteins, high affinity, functional in WesternBlot, Immuno precipitation, Immuno fluorescence, and the ability to function at any site within a protein.

The vector and especially the restriction enzymes are in fact designed also to be compatible for the subcloning of the isolated scFvs in a series of intracellular expression vectors described by Persic et al., (Persic, Righi et al. 1997; Persic, Roberts et al. 1997)

Chapter Four: construction of immune libraries from coeliac disease patients

4.1 Introduction

4.1.1 Construction of an immune library

The results of chapter 3 shows that the vector pDAN3 with the lox linker is indeed functional when used with very well characterized mAbs. Furthermore, in a mini-library experiment it seems to perform better then the other linkers. Nevertheless these results need a definitive confirmation with the isolation and characterization of scFv from a real library. Phage antibody libraries made from immune donors can generate antibodies with high affinity and specificity. As seen before it is very easy to isolate antibodies from relatively small libraries (10^5). For this reason we constructed a small immune library and use this as a model to definitively validate the phagemid vector. The creation of a small library of this type requires less work than the creation of a large library, and should give a series of results which would be representative for a large naive library.
In particular the small library could provide information on: cloning and selection procedure, but especially the stability of the vector system in terms of deletion or recombination of V-genes and functionality of the selected scFvs. The ScFvs can be tested as immunological reagents in classical Wester blot, Dot blot and immunofluorescence assay. Besides these practical consideration the creation of a library from an (auto)immune source also has the advantage that it may be useful to understand the humoral response in the disease state chosen.

4.1.2 Coeliac disease

Coeliac disease (CD) is characterised by intestinal mucosal injury and malabsorption. The pathogenesis appears to involve dietary exposure to wheat gluten and similar proteins in rye, barley and possibly oats (for a review see (Goggins and Kelleher 1994)), with gliadins, specific antigenic

determinants found in glutens, playing a prominent role. Ingestion of gluten by sensitive individuals induces a T cell infiltrate, crypt cell hyperplasia, and villous atrophy in the small intestine, resulting in diarrhea and malabsorption. The hypotesis that CD is a T cell-mediated immunological disease has been supported by the observation that the large majority of patients expresses the HLA-DQ2 (DQ(α1*0501,β1*02) and/or DQ8 molecules. Furthermore, HLA-DQ restricted, gluten specific T cell have been isolated from small intestin of CD patients (Lundin, Gjertsen et al. 1994; Lundin, Scott et al. 1994). The disease is characterised by the presence of specific and polyreactive antibodies recognising gliadins, food proteins, such as lactoglobulin and an endomysial auto-antigen, recently identified as being tissue transglutaminase (tTG) (Dieterich, Ehnis et al. 1997). These antibody levels are increased by exposure to glutens, and decrease on a gluten-free diet. The role of tTG in the disease is still unknown, although it has been proposed that this enzyme is involved in the activation of gliadin peptides (Molberg, McAdam et al. 1998) leading to their toxicity, and antibodies recognising tTG are produced by *in vitro* culture of biopsies treated with gliadin (Picarelli, Maiuri et al. 1996).

The high titre of specific antibodies found in these patients make them an ideal source of lymphocytes for the construction of a small immune library. For this reason several small libraries were constructed from PBLs and intestinal lymphocytes and selected against tTG, gliadin and food antgiens.

4.2 Results

4.2.1 Construction and selection of a library from PBLs

The scFv library was constructed starting from the peripheral blood lymphocytes (PBL) of a previously untreated CD patient with high titres of anti-gliadin and endomysial antibodies. VH and VL chains were amplified from PBL cDNA by PCR using V gene oligonucleotides (Sblattero and Bradbury 1998). VH and VL amplificates were assembled and cloned into the phagemid vector pDAN3 to obtain a primary library of $2x10^6$ individual clones. Thirty clones underwent scFv PCR amplification and fingerprinting with BstNI to test heterogeneity. All the clones were full length and with different fingerprints, confirming the diversity and integrity of the library.

Antibodies were selected from the library using purified α-gliadin, β-lactoglobulin, guinea pig tTG and cloned human tTG, using normal washing procedure (Marks, Hoogenboom et al. 1991) or long (30 minute) washes that should have higher stringency. After two or three rounds of selection, 96 individual phages from each selected population were tested by ELISA against the antigen used for selection, human serum albumin and the other antigens used. This was performed to examine the possibility that the antibodies selected were polyreactive, a known feature of immunoglobulins from CD patients. The results of the screening are reported in Table 4.1.

Antigen	Rounds of selection	Low stringency Positive clones	High Stringency Positive clones	different monoreactive clones
β Lactoglobulin	2	85/96	20/96	10
α Gliadin	2	80/96	25/96	8
Hum/GP tTG	3	0/96	0/96	0

Tab 4.1 Results of selection. The number of positive clones and the number of different monoreactive clones is reported.

After two rounds of selection using low stringency washes 80 clones out of 96 to α-gliadin and 85 out of 96 to β-lactoglobulin were positive but polyreactive. By using high stringency washes, 20 (α-gliadin) and 25 (β-lactoglobulin) clones out of 96 were positive and all were monoreactive. To

test the diversity, the isolated scFvs were fingerprinted by PCR amplification and digestion with BstNI. 10 different fingerprints for α-gliadin and 8 for β-lactoglobulin were found.
Despite a number of attempts, using both low stringency and high stringency washes, different coating procedure with or without activation with Ca^{++}, no ELISA positive clones were ever identified after selection on either guinea pig or human tTG. This is in contrast to the result obtained when a large human scFv library (Sheets, Amersdorfer et al. 1998) was used: after two cycles, 50% of clones were positive to both tTGs confirming that antibodies could be selected against these antigens. Furthermore, cloned human tTG proved to be highly antigenic and specific in ELISA using sera of CD and non CD patients (Sblattero, Berti et al. 1999).

4.2.2. Specificity of anti-gliadin antibodies

When tested by ELISA, all the scFvs isolated from the CD patient showed reactivity to α-gliadin, but not to human serum albumin or other irrelevant antigens, indicating that these are mono-, and not polyreactive antibodies. As reference, the monoclonal antibodies GL30, raised against α-gliadin, was used. All the scFvs showed in repeated experiments reproducible O.D. levels, presumably reflecting different affinities (or avidities) for the antigen. Among the others, clone n. 4 showed an O.D. comparable to scFv derived from mAb GL30, obtained from hyperimmunized mouse. Representative results are summarized in fig 4.1

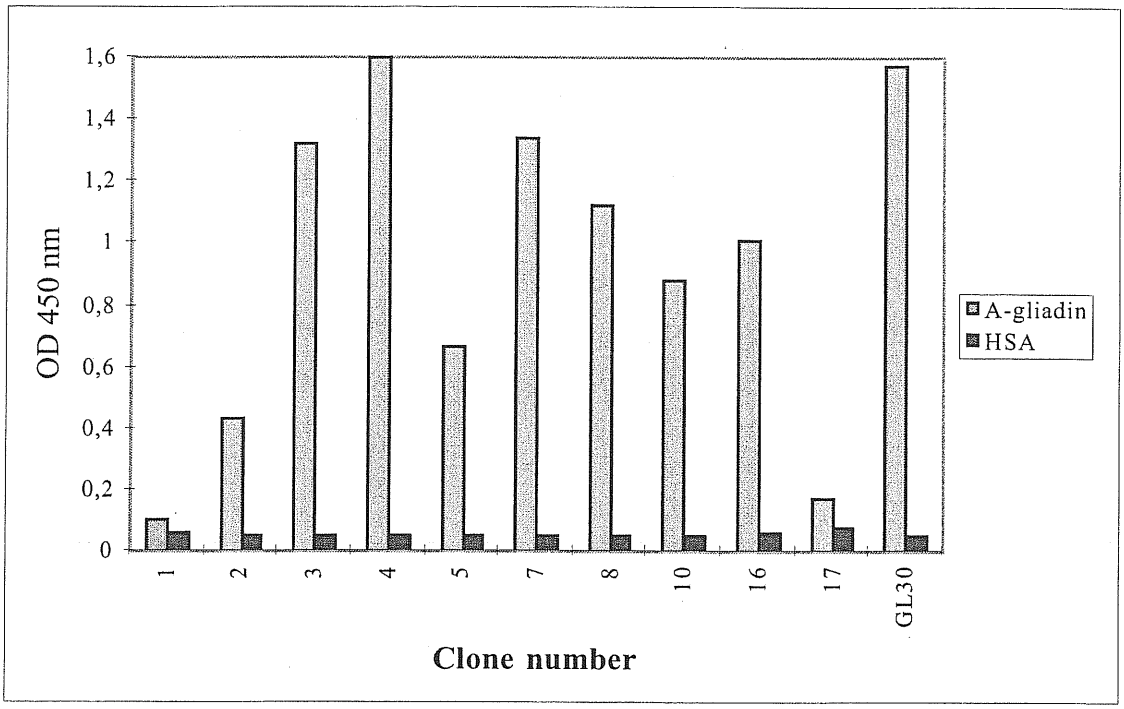


Fig 4.1 ELISA reactivity of the isolated clones against α-gliadin and HSA. GL30 is a scFv derived from a monoclonal Ab raised against α-gliadin and it is used as reference Antigens were coated at 10 μg/ml.

In competition experiments (see figure 4.2 for a typical curve), binding to α-gliadin adsorbed to the ELISA well (10 μg/ml) was inhibited only by soluble α-gliadin and not by other irrelevant antigens (BSA-HSA, data not shown), again showing the specificity of the interaction. In the example reported in figure 4.2, half maximum inhibition for scFv n°8 was obtained at a concentration of free α-gliadin of +-200 nM. Complete inhibition was not observed probably because higher amount of soluble α-gliadin was required.

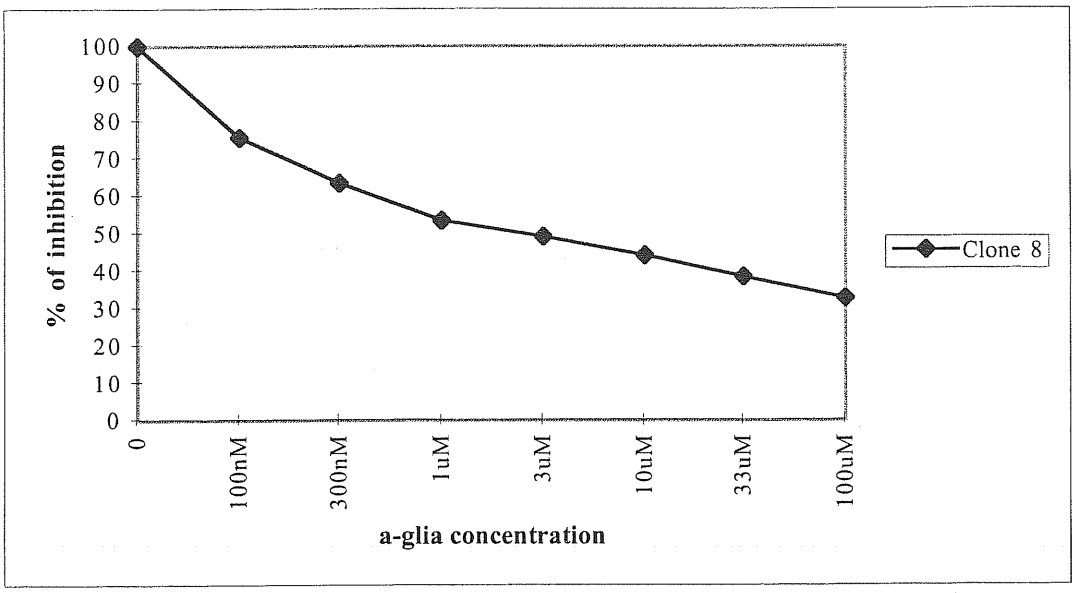


Fig 4.2 Typicl curve of competition experiment. In this case the binding of scFv number 8 was inhibited with soluble α-gliadin (from 100 μM to 100 nM). Half maximum inhibition could be calculated +- 200 nM.

4.2.3 Crossreactivity to toxic flour: ELISA and Western blot

The 10 anti-α-gliadin scFvs described above were also evaluated for reactivity to barley and rye flour extracts (figure 4.3). As expected, all the clones reacted to a different extent to wheat flour extracts in ELISA and did not react to BSA. Very poor or no reactivity was also found against rye extracts while 6 clones out of 10 strongly recognised the barley extract, indicating the recognition of antigenic determinants shared by the two cereals.

To test if the isolated scFvs can be used as immune reagents Western blotting was also performed. The monoclonal GL30 and all 10 isolated scFvs were produced as soluble protein and His tag purified using IMAC from a periplasmic extract and used as primary antibodies. The GL30 scFv (derived from a hybridoma) showed a broad reactivity to many wheat, barley and rye protein bands, but interestingly not oats, whose toxicity in CD is uncertain (figure 4.3). This was interpreted as indicating that GL30 recognised an antigenic determinant shared by all the toxic cereals tested. This is in contrast to the results obtained with the anti-gliadin scFvs obtained from the coeliac patient, each of which recognised a different complex pattern of bands, some of which were in common. The number of bands identified was always lower than that recognised by GL30 (figure 4.4), suggesting recognition of more restricted, but nevertheless shared, epitopes.

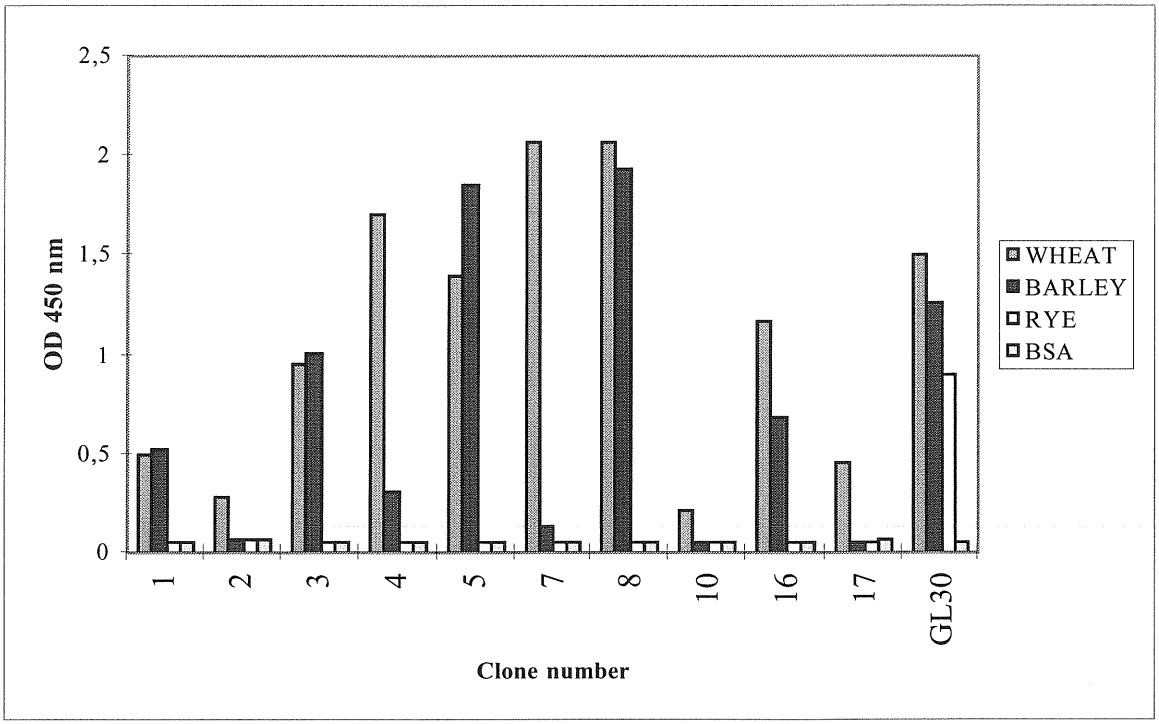


Fig 4.3 Comparison of the ELISA O.D. values of 10 scFv clones against a-gliadin to protein extracts from wheat, barley, rye and BSA. The reference scFv derived from mAb GL30 raised against α-gliadin.

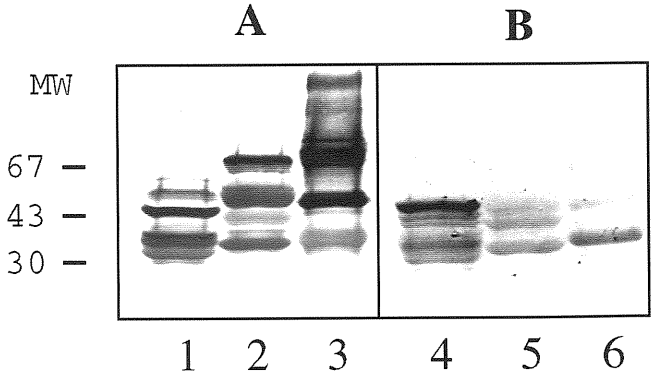


Fig 4.4 Western blotting of mAb GL30 (A) and scFv n. 6 (B) to wheat (lanes 1 and 4), rye (lanes 2 and 5) and barley (lanes 3 and 6) protein extracts. Molecular weight standards are shown on the left.

4.2.4 VH family usage and somatic mutations

The ten anti- α-gliadin clones were sequenced and the VH family used assessed by screening against the VBASE (Tomlinson, Williams et al. 1996) database. The results are reported in table 2. VH utilisation was restricted to four (VH 2, VH 3, VH 4, VH 6) of the seven human antibody VH families, with most of the VH genes belonging to the VH 4 family (7 cases) with a preferential use of segments DP-78 (2 cases) and DP-79 (4 cases). The V_L sequences were found to be Vk and Vλ chains belonging to various of the ten Vλ and six Vk families. One scFv was found to lack a VL sequence (clone n. 16).

The level of somatic mutation was assessed by comparing the sequences of the isolated clones with the closest corresponding germline sequence. In general, the ratio of replacement over silent mutations for the CDRs of the V_H domains was significantly higher than for the frameworks 1-3 (Tab. 2) suggesting that the mutations are the consequence of antigen-driven selections. The VL sequences showed a scattered number of differences from germline not allowing an interpretation as above. Also the lengths of the CDR3 regions were reported. For the V_H eight clones out of ten had a CDR length ranging between 11 and 16 residues, two clones were shorter: with 9 (clone 8) and 6 residues (clone 10).

a) VH

Clone	VH famiy	VH segment	FR		CDR		VH CDR3	CDR 3 lenght	JH
			R	S	R	S			
1	VH-6	DP-74	5	1	8	1	DGSIGLDALDI	11	3b
2	VH-4	DP-78	0	0	3	1	VSSNRYYYGMDV	12	6b
3	VH-4	DP-79	3	2	5	0	LGRTVATGHYYGMDV	15	6b
4	VH-4	DP-78	1	0	3	0	VSSNRYYYGMDV	12	6b
5	VH-4	DP-79	1	0	2	0	IPRRSVEYSSSSGHY	15	4b
7	VH-3	DP-42	7	3	4	2	TGSSGPYGDAFDI	13	3b
8	VH-4	DP-79	0	0	3	0	LAMDDTFDI	9	3b
10	VH-4	DP-63	5	7	7	4	KRAWDY	6	4b
16	VH-4	DP-79	1	0	6	0	LWRSWSDSGSYYGFDY	16	6b
17	VH-2	DP-76	8	3	5	1	ITLGYGDYYYGMDV	14	6b

b) VL

Clone	VL family	VL Segment	FR R	FR S	CDR R	CDR S	VL CDR 3	JL
1	VK-I	DPK-9	3	1	2	2	QQSYSTPVT	Jκ5
2	VK-III	vg/38K	6	2	4	1	QQRTNWPWT	Jκ1
3	VL-1	DPL-5	4	6	7	2	GVWDEGLSEEV	Jλ2
4	VK-II	DPK-15	0	2	3	0	MQGLQSPYT	Jκ2
5	VK-II	DPK-15	1	0	1	0	MQALQTPRT	Jκ1
7	VL-3	IGLV3S2	1	1	2	1	GVWDVSLSEEV	Jλ2/3a
8	VL-2	DPL-11	9	7	7	3	TSESRTSDVL	Jλ2/3a
10	VK-I	L12a	0	0	1	1	QQYNSYPLT	Jκ4
16	N/A							
17	VK-III	DPK-22	1	4	11	1	QQYRDSHT	Jκ2

Tab 4.2 a) V_H chains sequence analysis. The sequence of isolated V_H genes was screened against VBASE. The family, V_H segment, mutations, CDR3 length and correspondent sequences are reported. **b)** As above except CDR3 length for V_L chains.

4.2.5 Construction and selection of a library from intestinal lymphocytes.

Having found the lack of anti tTG from a PBL library, a second library was constructed from lymphocytes obtained from a biopsy of an active celiac patient. As it is well known that the response agains tTG is mainly of the IgA class an IgA specific primer was designed and a library of $5x10^6$ clones was constructed. As for the PBL library also this one was selected against α-gliadin and tTG. Interestingly anti tTG could be easily isolated from the library after 1 round of selection. The scFvs were found to be specific for the tTG, with some of them having a fine specificity able to specifically recognize the human form but not the guinea pig one, even though the two are 93% homologous (see figure 4.5).

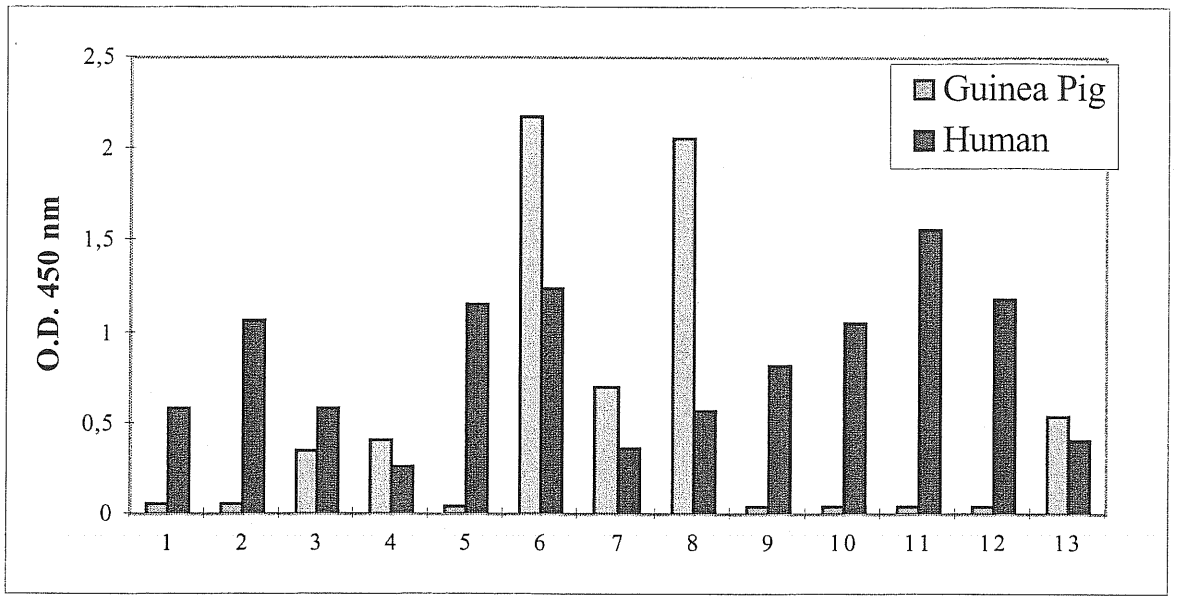


Fig 4.5 Reactivity of isolated clones against Human tTG and Guinea Pig tTG. Both antigens were coated at 10 µg/ml.

4.2.6 Anti tTG scFv as immunochemical reagents

During the active phase, a characteristic IgA auto-antibody response recognising the transglutaminase is produced, these antibodies can be detected in an immunofluorescent assay using either monkey oesophagus or human umbilical cord. In order to prove that the selected scFvs could be used as reagents in immunochemistry, we compared the results obtained with the patient's serum with that obtained with the scFvs anti Hu-tTG isolated from the libray made from his lymphocytes.

Previous results showed that scFvs purified using the His tag are able to work in Western blot. To avoid the tedious purification protocol we first verified if bacteria supernatant can be used directly without further purification. In the first analysis bacteria supernatant was dialyzed against PBS and used as a primary antibody in immunofluorescence on human umbilical cord sections. IFA was also performed with the serum of the patient on adjacent sections. The two reactions were revealed with anti SV5 and fluoresceinated anti mouse for the scFv and with anti human IgA conjugated with texas red for the patient serum. The results clearly shows that the scFv give an identical pattern compared with the serum and with far less background staining. In particular when the two flourescence patterns were revealed simultaneously they give a perfectly superimposable staining as can be seen in fig. 4.6

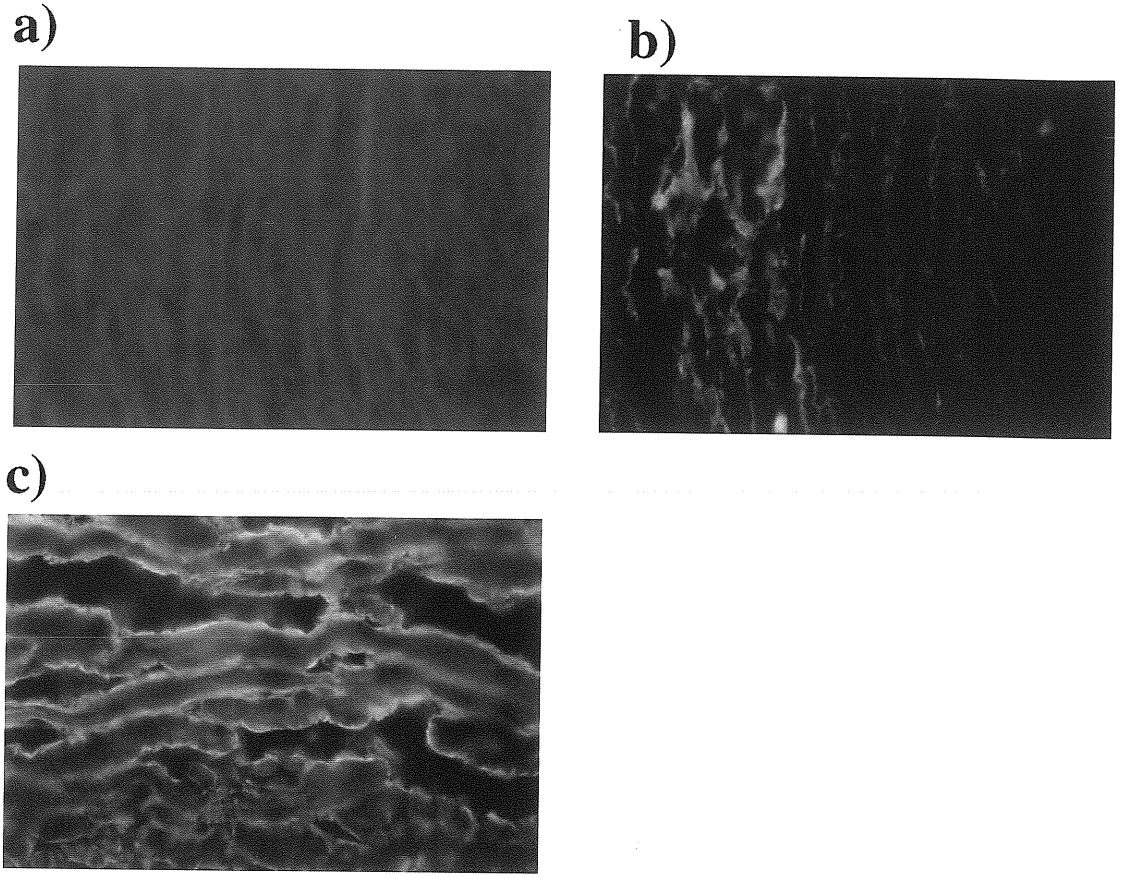


Fig 4.6 Immunofluorecence staining of humbilical cord vein cryosection. **a)** Anti endomysium (transglutaminase) staining for serum sample was revealed with an anti IgA texas red conjugated secondary antibody. The staining revealed the typical honeycomb-like flourescence pattern. **b)** Dialyzed bacteria supernatant of a scFv anti Hu-tTG was used as primary antibody, the staining was revealed with the anti SV5 and anti mouse foureseinated. The fluorescence pattern was similar to that observed using the serum. **c)** a section was incubated with the serum sample and the scFv, the fluorescence was revealed simoultaneously. The almost perfectly superiposable staining is indicated by the yellow (red+green) fluorescence.

The same isolated anti tTG scFv were used also in a dot blot assay (Not, Sblattero et al. 1999). The signal obtained using bacterial supernatant was comparable with that obtained with a strong positive serum sample.

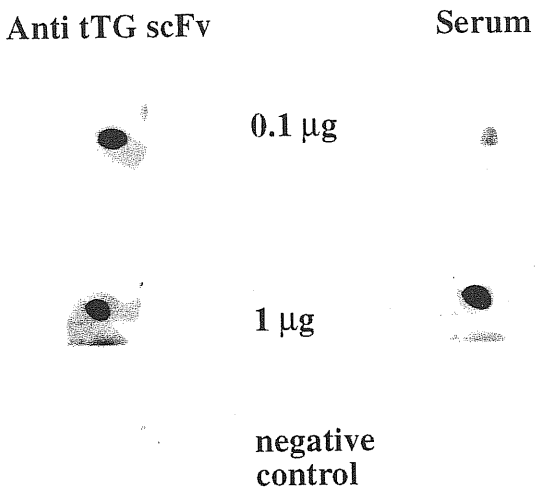


Fig 4.7 Dot Blot assay. Hu-tTG was spotted at various concentration on an nitrocellulose membrane and incubated with the patient serum or bacteria supernatant of a scFv anti Hu-tTG. On the right the results with the scFv and on the left with the serum. The sensitivity and also the background staining on the membrane was comparable

As in all the previous tests the scFvs were found to be extremely stable. In order to quantify this, we performed an accelerated degradation test. Bacteria supernatant was dialyzed and then incubated at 37°C for a week with periodic measurements of the ELISA signal. After one week the signal decrease was limited to 50% of the original one indicating a very stable structure and limited proteolytic degradation of the scFv.

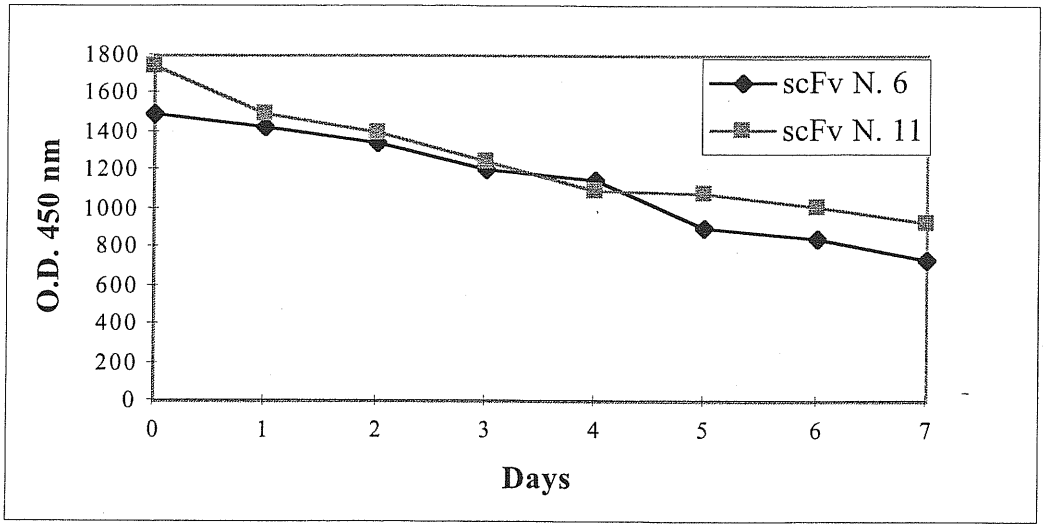


Fig 4.8 Stability of the scFv in a degradation test. The ELISA signal was measure at various time. ELISA signal is given as a % of the signal obtained with freshly prepared scFv.

4.3 Discussion

4.3.1 Immune libraries as a model

In the present study we created two antibody libraries from a patient with coeliac disease. Although the main aim was to test the vector system described, it is clear that interesting information on the humoral response in coeliac disease was also obtained. This approach has a major criticism based on the fact that the VH-VL pairings occurring *in vivo* are not usually recreated *in vitro*, as VH and VL genes are paired at random. However, a number of findings suggest that phage antibody libraries can act as surrogates for the humoral response, or at least that part of it originating in the tissue used to create the library.

4.3.2 CD specific antibodies

Using the PBL library after two cycles of selection on α-gliadin and β-lactoglobulin using low stringency washing nearly 85% of clones recognised the antigens, but all of these scFvs proved to be polyreactive, able to recognise unrelated antigens as well as that used for selection. This is characteristic of the serum response in patients with coeliac disease. At higher stringency, a lower number of positive clones were selected, but all were shown to be monoreactive. Subsequent fingerprinting revealed the presence of at least 10 individual clones to α-gliadin and 8 to β-lactoglobulin in the library. We further investigated the reactivity pattern of ten monoreactive anti α-gliadin scFvs. In ELISA all the individual clones were inhibited by soluble α-gliadin and showed different degrees of reactivity to other flours known to be toxic in CD and which share conserved sequences with wheat gliadins.

The sequences of the ten scFvs isolated revealed that most (7/10) belonged to the V_H 4 family, with DP-78 or 79 being the most likely genes used. The sequence of the V_H genes isolated revealed that all the clones showed amino-acid substitutions with respect to the germ line sequences, suggesting that these V_H genes arise from an immune response involving somatic mutation and selection.

The reactivity of the selected clones was further investigated. In particular the purified scFvs were used for western blot. The performance was comparable with that of the monoclonal Ab GL30. Infact the same

extracts used to test polyreactivity in ELISA challenged in Western blotting revealed a complex recognition pattern.

Interestingly, no positive clones could be selected against guinea pig or human tTG despite a number of different attempts using different conditions. These results suggest that the B lymphocytes synthesising antibodies against tTG are not present in peripheral blood and support the hypothesis that the primary response to tTG in CD occurs at the intestinal level. To test this hypotesis we constructed a second library from intestinal lymphocytes. This library was selected with the same procedure as the first and allowed the isolation of several specific anti tTG scFvs after 1selection cycle.

All the screening tests for CD are based on the detection of anti transglutaminase antibodies (Not, Sblattero et al. 1999; Sblattero, Berti et al. 1999) and this was an ideal situation to compare serum from patients. By both dot blot and immunofluorescence there was no difference between specific monoclonal scFvs and serum.

In conclusion, by creating two immune libraries from lymphocytes derived from patients with coeliac disease, the pDAN3 vector designed was shown to be functional by a number of different criteria: 30 of 30 unselected scFvs were found to be full length, indicating that scFvs constructed using this vector are stable; scFvs recognising specific antigens could be selected from the libraries, and the vast majority of these were full length and finally, selected scFvs were functional in a number of different immunological tests (western blot, ELISA, dot blot). Furthermore, the heat stability of scFvs selected from this library (50% decrease in ELISA signal after 7 days at 37°C) was remarkable. These results together confirmed that pDAN3 was functional and that a large phage antibody library could be made using this vector.

The results also provided an interesting insight into the humoral response found in coeliac disease. In particular, it is clear that the anti-tTG response found in CD is one generated at the mucosal level, while that against gliadin appears to be also generated by blood lymphocytes. The scFvs selected from the libraries had the same properties as the serum from the patient used to construct the libraries, and the sequences of the VH genes isolated appear to indicate that specific genes are involved in the response to gliadin.

Chapter Five:
Single bacteria making many different antibodies: exploiting recombination to make large phage antibody libraries

5.1 Introduction

5.1.1 Making libraries by recombination.

As already discussed in chapter 1 it is possible in theory to create a diversity in a population of polypeptides by *in vivo* recombination of DNA. Several methods have been proposed and used ((Waterhouse, Griffiths et al. 1993; Geoffroy, Sodoyer et al. 1994; Griffiths, Williams et al. 1994; Tsurushita, Fu et al. 1996), and all have the common feature of using a "two plasmid" system. However, this system has several drawbacks. In particular, the use of reversible recombination can lead to heavy contamination of the library with the starting phage; there are byproducts of the recombination which can also be packaged into phage particles, so reducing the diversity and as the system is presently designed only a single round of recombination is possible. What is especially limiting is that only one recombination event is possible in a single cell. This limits the diversity of the library to the number of bacteria used in the recombination step.

Many of these problems could be overcome if a "one vector" system could be developed. This would involve the creation of a library of VH and VL genes separated by two non-homologous recombination lox signals. In practice, one such signal would have to be found in the linker in the case of scFvs and between the VH-CH and VL-CL genes in the case of Fabs, and the other anywhere else within the vector backbone. If infection of cre recombinase expressing bacteria by such a vector at a high multiplicity of infection, could result in the ingress of many different phagemid particles, VH and VL genes could be exchanged between different phagemids, creating many new VH/VL combinations. The use of one vector would be ideal for several reasons, in particular all the new combinations created would be functional and the recombination procedure could be repeated indefinitely

because the products maintain their original structure and so could be substrates for further rounds of recombination.
To achieve this goal several points were taken into account in order to overcome the limits of the two vector system.

The first and most important issue concerned the possibility of introducing two or more plasmids (of the same kind but with different recombination substrates) simultaneously into a bacterial cell.
Probably the most efficient process of introducing DNA molecules into a bacterial cell is by phage infection. Some early papers (Marvin and Hohn 1969) demonstrated that a bacteria can be infected by 3 phages simultaneously. However, when co-infection happen this involves three identical phage which do not differ in nucleic acid sequence. This is because it is generally believed in the field that "plasmids that utilize the same replication system (origin of replication) cannot co-exist stably and are said to be incompatible" (see pages 1.3-4 of (Sambrook, Fritsch et al. 1989) and cited references).
One case in which the same origin of replication is present on two different vectors present simultaneously within a cell is when producing phagemid particles. In this procedure bacteria are infected or transfected with phagemid DNA, and infectious (phagemid) particles are rescued by subsequent infection with a helper phage. Both phagemid and helper phage contain the Ff origin of replication. The infection of bacteria containg a phagemid with the helper phage indicate that a superinfection event is possible and that the plasmids can-coexist. However, it should be pointed out that in most cases, the helper phage and phagemid to be packaged also contain additional origins of replication.
These results open the way to test the possibility of multiple infection as a way of introducing several recombination substrate into a bacterium. To test this we obtained a series of phages with different antibiotic resistances that would allow the determination of how many different phages can infect a single cell on the bases of the resistances it can express.

If multiple infection can be shown to occur, the second point which needs to be analysed regards the possibility of the plasmids being able to stay in the cell long enough to undergo recombination. It is well known that plasmids belonging to the same incompatibility group are not stable in the same cell. To ascertain how strict this point is, we designed a new vector based on pDAN3 that could be used for recombination and tested its functionality with a model experiment using the well characterized mAb, D1.3. This model will allow both the determination of the stability of the

system as well as the extent of recombination tested at both the genotypic and phenotypic level.

The last point was to exploit the possibility of creating a library by this "single vector" *in vivo* recombination procedure. The main problem is how to measure the diversity of the library made in this way. We designed a series of experiments that allowed a definitive determination of the stability of the system, the extent of recombination, and so the diversity of the final library.

5.2 Results

5.2.1 Bacteria can be infected by more than one phagemid particle

The *in vivo* recombination procedure has as an essential requisite that at least two different plasmids (which can recombine with one other) should be introduced inside the same cell. There are several ways that allow the introduction of DNA within a cell, with phage infection being the most efficient. As a first step towards the development of an *in vivo* recombination system the number of different phagemids which could infect a single bacteria was assessed. Five phagemids expressing different antibiotic resistances (see tab 5.1) were obtained. These phagemids do not contain gene 3 and so overcome the problems related to the inhibition of pilus synthesis by p3 (Boeke, Model et al. 1982). These phagemid particles were prepared with M13KO7 as helper phage.

The origins of replication of the vectors were PMB1 (15-20 copies bacteria) and ColE1 that confer a very high copy number (500-700 copies per bacteria). The two origin of replication are closely related and plasmids carrying these are incompatible with one another. This means that when two such plasmids are introduced into the same cell, they compete with one another both during replication and the subsequent step of partition into daughter cells. Over the course of few generations of bacterial growth one of the plasmid is lost (Sambrook, Fritsch et al. 1989).

Plasmid	Origin of replication	Resistances	Source
pMPM-A1	ColE1	ampicillin	Stratagene, La Jolla
pMPM-K1	ColE1	kanamycin	(Mayer 1995)
pMPM-T1	ColE1	tetracycline	(Mayer 1995)
pBSL121	pMB1	chloramphenicol	(Alexeyev, Shokolenko et al. 1995)
pBSL141	pMB1	gentamycin	(Alexeyev, Shokolenko et al. 1995)

Tab 5.1 List of plasmids used for multiple phage infection experiment. For each plasmid used origin of replication, resistance and source are indicated.

DH5αF' bacteria were infected simultaneously with equal titres (measured by cfu) of the 5 phagemids particles. To increase the possibility of multiple infection the multiplicity of infection (MOI) was kept very high with a ratio of 20:1 (phage:bacteria). The possible contamination of helper phage genome was considered not relevant because all the 5 preparation have a phagemid/phage DNA ratio equal or higher than 100:1. After 1h of infection bacteria were plated out on single or double antibiotics. The total number of bacteria present in the culture was calculated plating on plates without antibiotics and the number of colonies obtained was considered as the reference for 100%.

Antibiotic resistance	**% of resistent bacteria**
Tetracycline	100 %
Tet + ampicillin	94 %
Ampicillin	92 %
Tet + Kanamycin	85 %
Kan + Amp	81 %
Kanamycin	74 %
Chloramphenicol	64 %
Gentamycin	52 %

Table 5.2 Entry and survival of phagemid carrying different antibiotic resistances into single cells. Percentages of bacteria showing the resistances is indicated. The number of bacteria surviving in the absence of antibiotics was considered to be 100%. Resistance to some antibiotic pairs is not indicated as the plasmids used did not permit us to test these combinations.The data are the mean of 6 different measurements.

As can be seen in table 5.2, bacteria infected with the mix of phagemid particles show resistance to all 5 antibiotics. Considering the percentages of single resistance to all 5 antibiotics a minimum estimate of 21% of bacteria (1x0.92x0.74x0.64x0.52) appear to have had the potential to show resistance to all five antibiotics simultaneously. This means that at least a population of bacteria in the culture has been infected in 1h by 5 phagemid particles. The real percentage of bacteria with multiple infection is probably higher, in fact, while Ampicillin, Kanamycin and Tetracycline all give values of resistances over 74%, and near 100%, resistance to Gentamycin and Chloramphenicol was found most infrequently (64% and 52%). This is probably due to the characteristic of the plasmids used. Infact the pBSL plasmids have the pMB1

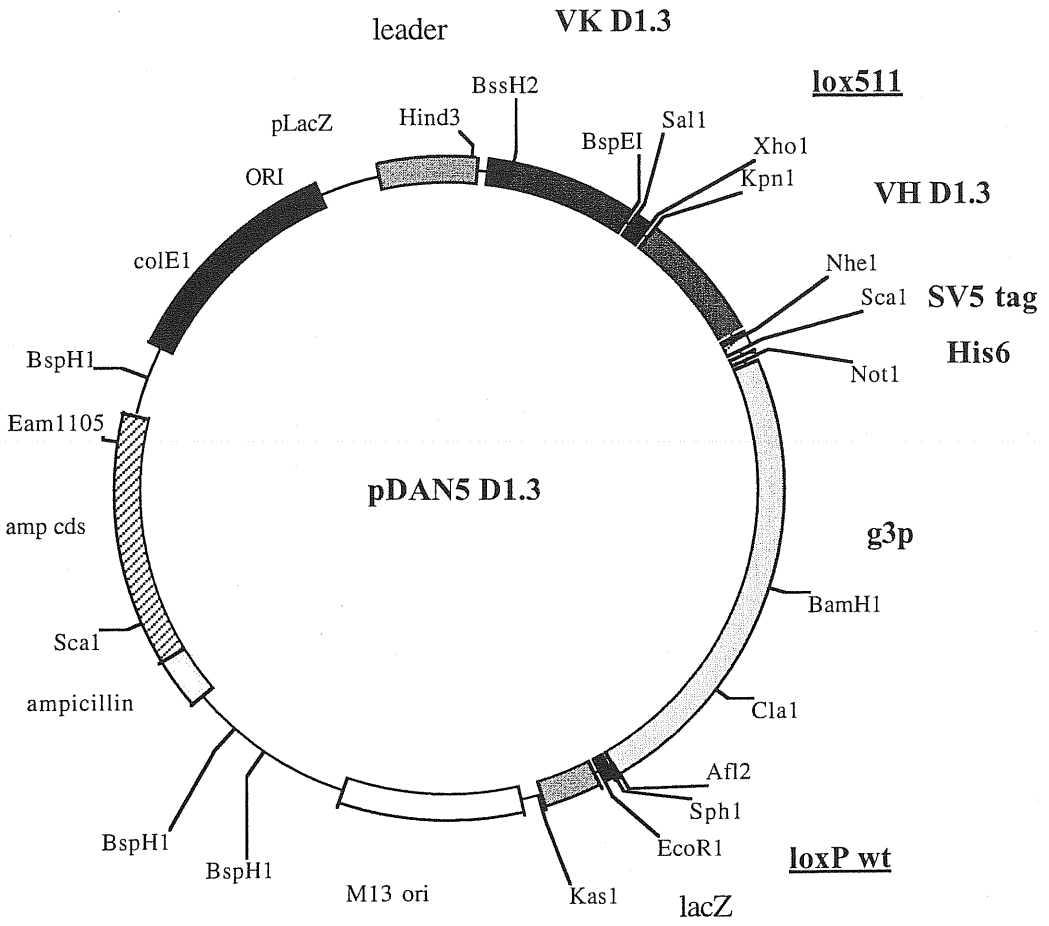
origin of replication that have a much lower copy number (15-20) compared to the colE1 (500-700). As the two origins are incompatible one possible explanation could be that the pBLS vectors are much more easily lost during replication compared to the pMPM vectors. The possibility of an higher percentage of multiple infection is supported also by the results obtained with double resistance. For example, 94% of all bacteria are resistant to ampicillin and Tetracycline, and 85% are resistant to Tetracycline and Kanamycin, 81% of the total bacteria are resistent to ampicillin and Kanamycin. These data clearly confirm that almost 90% of bacteria are infected by more than 1 phage during the 1h incubation time. The results show also that in the presence of a selective pressure (antibiotic resistence) the plasmids with the same origin of replication are stably maintained within the cell after an overnight growth.

5.2.2 Construction of pDAN5: a vector for *in vivo* recombination

The previous results demonstrated that more that one plasmid with the same origin of replication can be introduced by infection into a single cell. This opens the way for the construction of a phage display vector for *in vivo* recombination. In order to construct such a vector for recombination of heavy and light V-genes in the scFv format the presence of another *loxP* recombination site was needed. Two possible positions for the site were available: upstream of the leader sequence or downstream of the gene 3 sequence. The second position was selected, and as the *loxP*511 was present in the linker the wild type *loxP* sequence (*loxP*WT) was used. gene 3 was amplified with the primers gene 3 back and gene3lox. As can be seen in Tab 2.1 this primer introduce the *loxP* WT sites at the end of the gene 3. The PCR fragment was subsequently cloned replacing the original gene 3 in pDAN3 vector, this created the pDAN5 display vector as shown in fig 5.1. a) The insertion of the *loxPWT* sequence was confirmed by DNA fingerprinting and sequencing.

If two pDAN5 plasmids containing different V genes can be introduced into a cell expressing Cre recombinase recombination should occur, and the fragment comprising VH-tag-gene3 should be shuffled between the two plasmids (see fig 5.1b).

a) Map of the plasmid pDAN5



b) Scheme of the recombination with one vector system

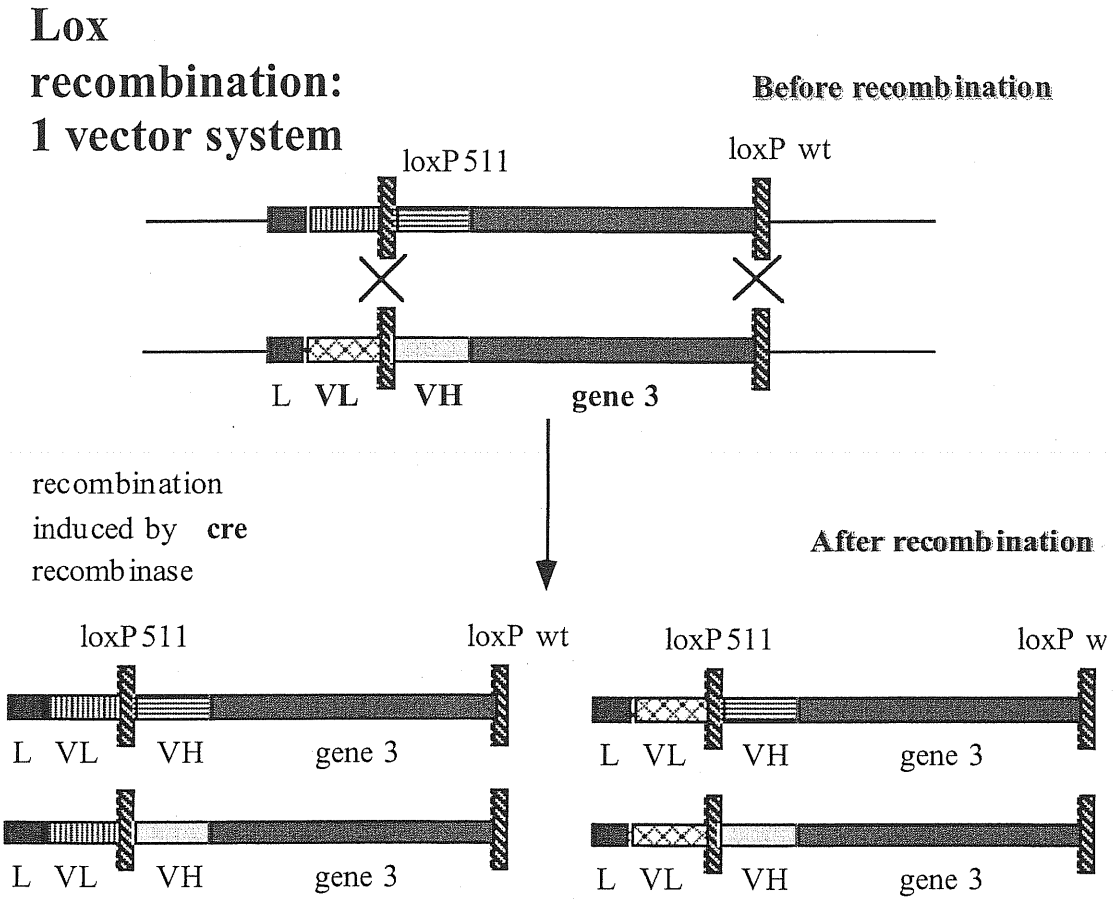
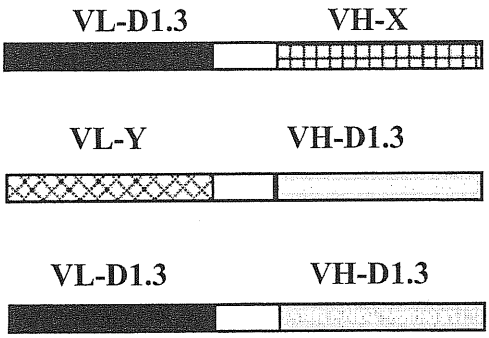


Fig 5.1 a) Map of the plasmid pDAN5. The two *loxP* sites are underlined. The *loxP*511 is positioned in the linker sequence, the *loxP*WT is found after the gene 3 sequence. All the other features are identical to the vector pDAN described in chapter 3. **b)** Schematic representation of the recombination event mediated by cre recombinase and two lox sites.

5.2.3 Recreating a functional scFv by shuffling the V-genes

The ability of the new vector to recombine *in vivo* V-genes to create functional new antibody specificities was tested in a model experiment. The scFv derived from the anti-lysozyme mAb D1.3. (Mariuzza, Jankovic et al. 1983) was selected as reference. Two scFvs which contained either D1.3 VH (VL/Y-VH/D1.3) or D1.3 VL (VL/D1.3-VH/X) with irrelevant partner chains were created (see fig 5.2 a)) Recognition of lysozyme by D1.3 scFv was shown by ELISA to require the presence of both D1.3 heavy and light chains; single D1.3 chains associated with irrelevent partner chains were non-functional (see Fig 5.2 b)).

a)



b)

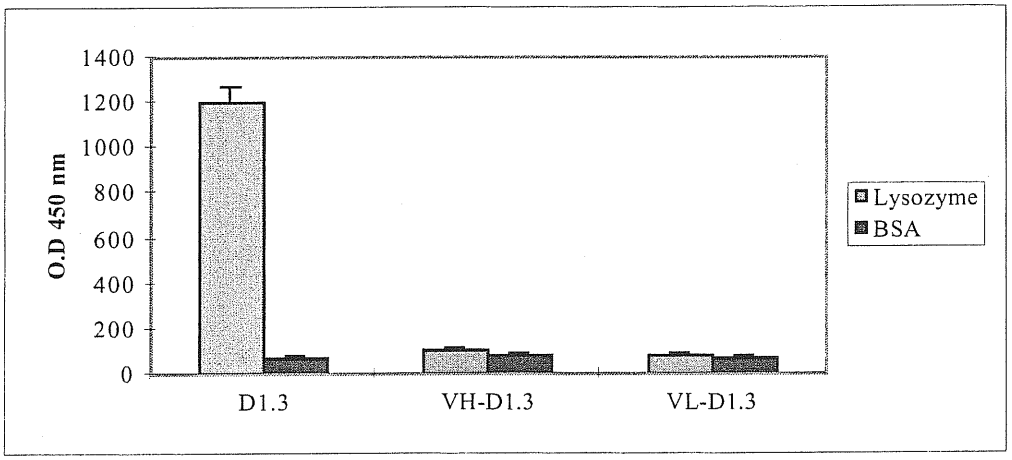


Fig 5.2 a) Scheme of the scFv constructs containing either the VH or the VL of D1.3 and the complete scFv **b)** ELISA with the two construct and with the complete D1.3 scFv as reference. The results shows that only the complete scFv is functional, plasmids containg either VH or VL of D1.3 alone are not functional.

Phagemid particles were prepared from the two plasmids, titrated and mixed
in a 1:1 ratio. On the bases of previous multiple infection results, bacteria were infected for 1h at an MOI of 20:1. The recombination experiment was performed in parallel infecting either normal DH5α E. coli cells (as negative control for background recombination efficiency) or BS1365 cells constitutively expressing cre recombinase. Infected bacteria were then grown overnight at 30°C to allow recombination to take place. If recombination was successful, each cre expressing bacteria should contain four different scFv genes (VL/D1.3-VH/D1.3; VL/D1.3-VH/X; VL/Y-VH/D1.3 and VL/Y-VH/X - see figure 5.3) and if the stability of the plasmids is comparable, each gene should be (in theory) present in the same

percentage. To analyse the phagemid composition of the cells after recombination phagemid particles were rescued from such bacteria and coupled to the appropriate scFv protein by passage through normal E. coli infected at MOI <1 (see figure 5.3).

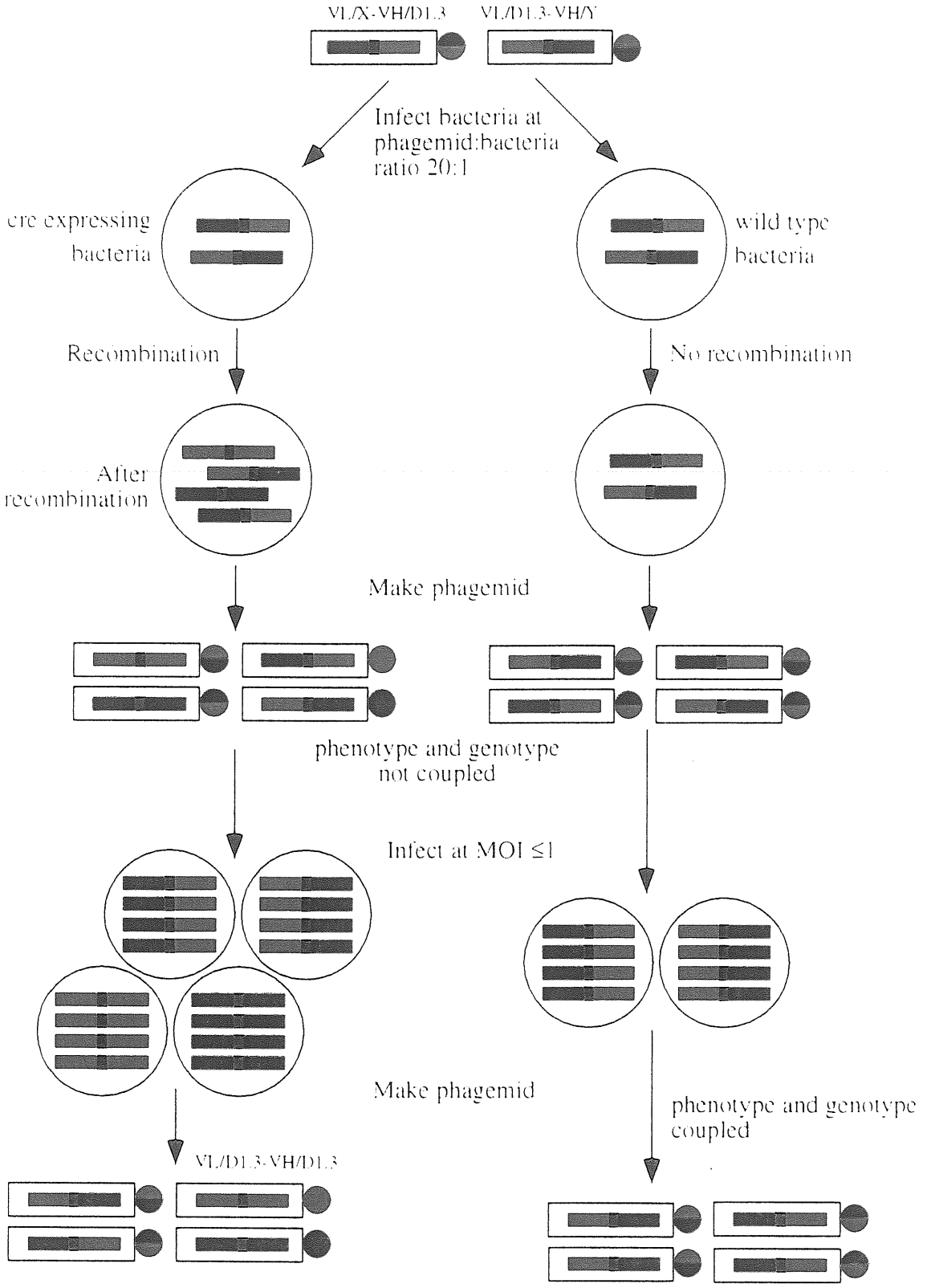


Figure 5.3 Scheme of D1.3 recombination experiment
Two phagemid containing VL/X-VH/D1.3 and VL/D1.3-VH/Y (where X and Y represent irrelevant antibodies) are added to either cre expressing bacteria or wild type DH5αF at high multiplicity of infection. After overnight growth, phagemid are made, reinfected into DH5αF to couple genotype and phenotype, and tested by PCR and ELISA for the presence of functional VL/D1.3-VH/D1.3.

The extent of recombination was tested both at the genotypic and phenotypic level. For the DNA level two specific primers for the VH and VL CDR3 sequences were designed and used in PCR, such that only the recombined phagemids carrying D1.3 VL and VH genes in cis could give rise to an amplification product. The phenotypic level was tested by ELISA. This confirmed both recombination as well as the maintenance of scFv function (the absence of aberrant recombination forms). The results obtained were in accordance with the theoretical analysis. After coupling of genotype and phenotype 48 or 96 colonies were tested. In the presence of Cre recombinase, the recombination was present in 25% of phagemid if assesed by PCR and 17% by ELISA (see table 5.3), whereas in bacteria not expressing cre, no recombination was found and no functional D1.3 identified.

These results indicate that recombination induced by cre recombinase can be used to shuffle heavy and light chain genes between different phagemids and so give rise to new specificities, even when those phagemid carry the same antibiotic resistance and origin of replication. Furthermore, the recombination reaction appears to go to equilibrium, since VL/D1.3-VH/D1.3 comprises approximately 25% of the total.

Analysis of D1.3 recombination by PCR and ELISA

ELISA testing	**Positive colonies**	**% of positivity**
Cre recombinase bacteria	16/96	17%
Wild type bacteria (DH5aF)	0/96	0%
PCR testing using D1.3 specific primers		
Cre recombinase bacteria	12/48	25%
Wild type bacteria (DH5aF)	0/48	0%

Table 5.3 ELISA and PCR analysis of D1.3 recombined phagemid
The ELISA signals observed for 96 different colonies, and PCR signals for 48 different colonies, derived after growth in cre expressing bacteria or wild type bacteria are shown. ELISAs were performed using phagemid binding to lysozyme adsorbed to 96 well plates. PCR was performed using a 5' primer specific for CDR3 of VL and a 3' primer specific for the CDR3 of VH. This combination is only able to give a PCR product when D1.3 VH and VL are present in the same scFv

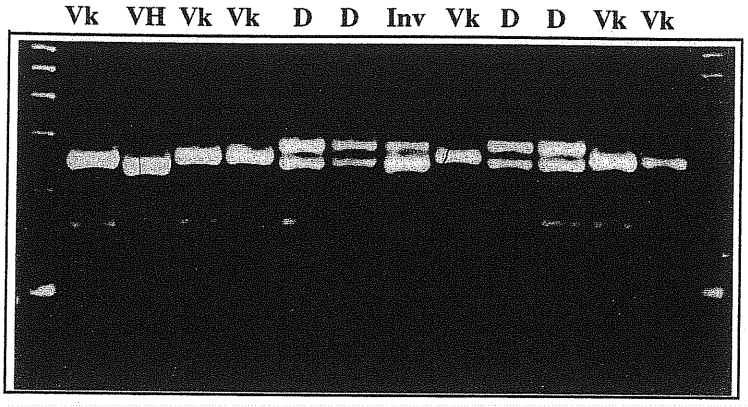


Fig 5.4 Fingerprinting analysis of recombined colonies. Single colonies after recombination and coupling of genotype and phenotype (see fig 5.3) were selected and scFv PCR amplified. BstNI fingerprinting analysis shows the presence of 4 different clones corresponding to the 4 possible recombination results. The percentage of each form was near 25% confirming PCR results.

5.2.4 Creation of a small primary antibody library in pDAN5.

In order to create an antibody library as a source of V-genes to use for *in vitro* recombination more than 40 samples of peripheral blood lymphocytes were collected from healthy blood donors. The library was created as described in fig 5.5. Total RNA and cDNA was prepared from each sample and then cDNA was pooled and V-genes amplified using the new set of primers (Sblattero and Bradbury 1998) described in chapter 3.
One of the major concern during the library construction was to maintain the diversity of the original V-genes repertoire through all the amplification and cloning steps. The ratio between the V-genes as represented in the PBL repertoire was maintained by amplifying the heavy and light chain genes from cDNA with each Back primer and the pool of the For primers. After the primary amplification the purified V-genes amplification products were pooled, so that the relative abundance of each amplification was maintained, and reamplified with external primers. On the bases of both theoretical consideration (see introduction) and practical results obtained in previous works (Marks, Hoogenboom et al. 1991) the VH repertoire was limited to the IgM genes. An IgM CH1 specific primer was designed and used in the primary amplification. The library was created by assembling VH and VL in

the scFv format. Two long pull through PTL primers (see tab 2.1) were designed in order to cover the whole linker length. The presence of such a long tail in the VH and VL fragments facilitate the assembly reducing the probability of incorrect annealing. Moreover the use of the PT1 and PT2 primers added long tails at the 5' and 3' ends of the PCR fragments in the reamplification step. This was designed in order to facilitate the cloning procedure. In fact, the scFv assembled PCR fragment is much more efficiently cut and the extent of completely digested fragment could be easily determined thanks to the change in size that could be seen in a normal agarose gel.

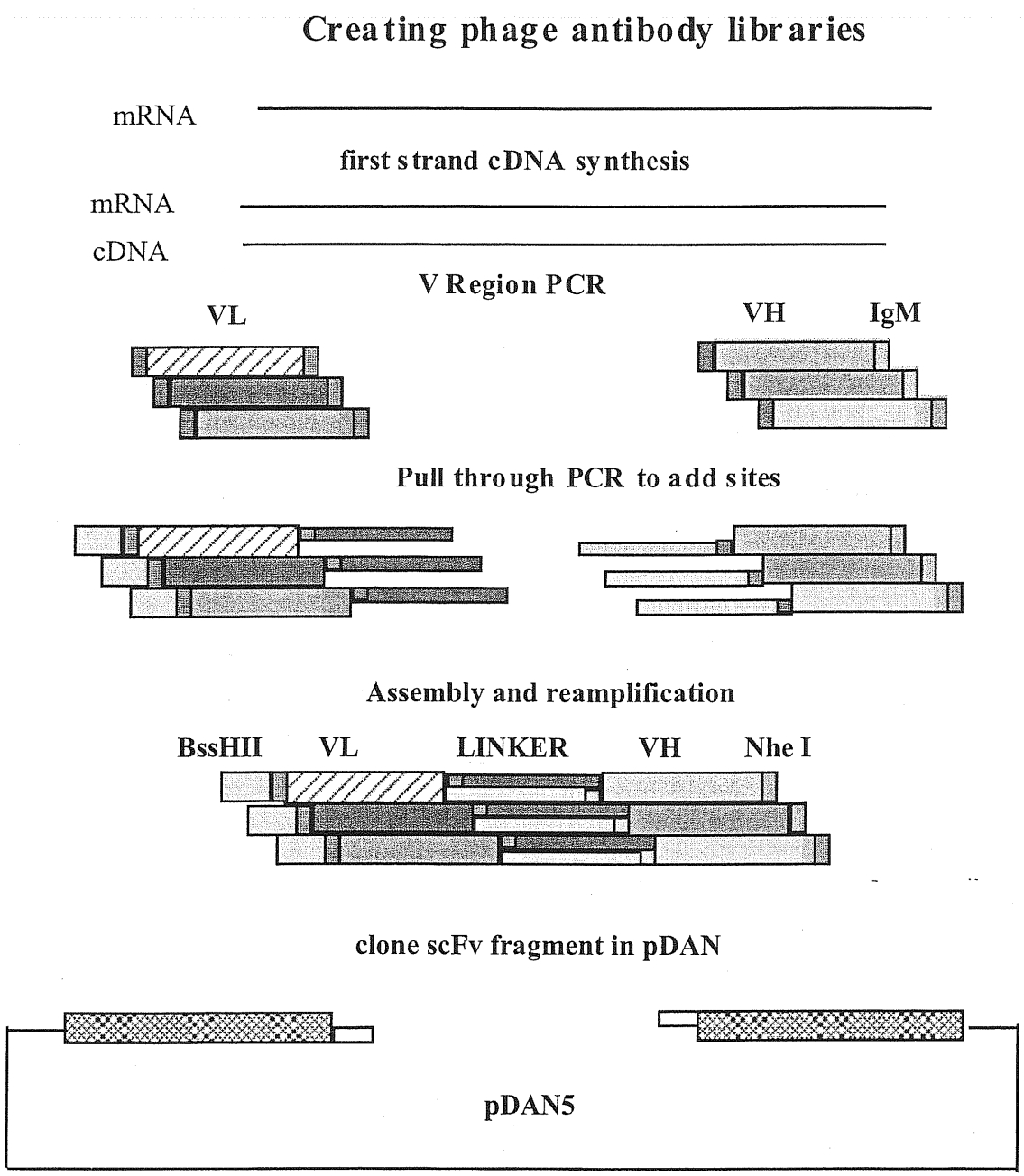


Fig 5.5 Scheme used in the creation of the primary library in the vector pDAN 5. cDNA obtained from total RNA is amplified with specific human primers. After primary amplification DNA is pooled and reamplified in order to add a complementary sequence to allow the assembly of scFv. Finally the scFv is amplified with external primers to add long tail to facilitate the cut and cloning procedures.

A small scFv phagemid library consisting of $7x10^7$ independent clones was created in pDAN5 by cloning PCR assembled scFv in this manner.

5.2.5 Assessing diversity and quality of the library

The diversity and integrity of this primary library was tested in several ways: first of all at the DNA level by DNA fingerprinting and sequencing and second by Western blot analysis. More than 50 scFvs were PCR amplified from single randomly selected clones. This showed that 96% of clones were full length. Furthermore, BstNI fingerprints were all different, confirming the diversity of the library (see fig 5.6).

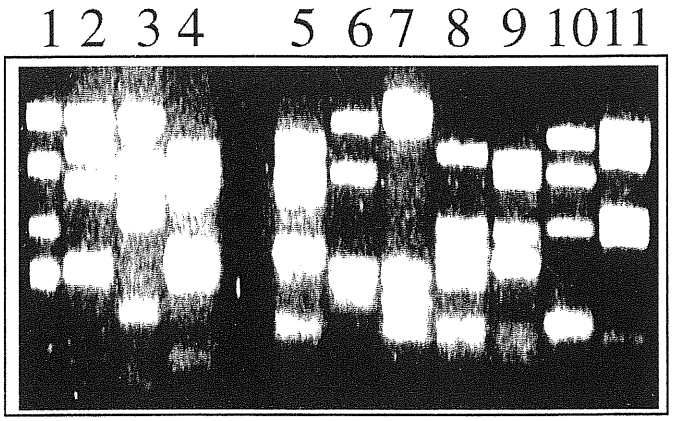


Fig 5.6 DNA fingerprinting analysis of the clones of the primary library. The scFv fragment from random selected clones was amplified with the primers PT2 and digested with BstNI. All the clone have a different pattern confirming the diversity of the library.

The quality of the primary library at the phenotypic level was also assessed by Western blot of the whole library and phagemid particles made from randomly isolated single clones. The library showed a good expression level of the full length recombinant protein, ranging between 10-20% of total p3 present. 12 out of 15 of the selected single clones showed a full length p3-scFv protein band, with variable intensity in the levels of expression. This

variability could be partially explained by the different titers of phagemid particles loaded on the gel.

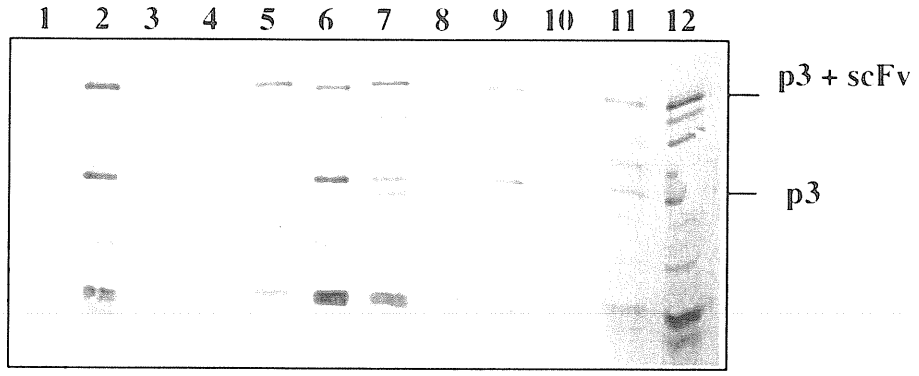


Fig 5.7 12 clones of the primary library ($7x10^7$) were selected randomly. Phagemid particles were prepared O/N at 30°C and concentrated by Peg precipitation. 10 µl of concentrated phages were loaded. Recombinant p3 protein was revealed with SV5 (Hanke, Szawlowski et al. 1992) antibody.

5.2.6 Creation of a large library by *in vivo* recombination

The primary library obtained above was used to infect bacteria constitutively expressing cre recombinase. The infection was performed at a high MOI, as for the D1.3 experiment, following the scheme illustrated in figure 5.6, the library of $7x10^7$ phagemids was used. 10^{10} cre bacteria were infected with $3x10^{12}$ phages, resulting in a MOI of 300:1. This leads to a population of bacteria containing multiple phagemids encoding different VH and VL genes which can be recombined by the cre recombinase. Following overnight recombination, phagemid particles were derived from these bacteria and used to infect bacteria not expressing cre (DH5αF') at a phagemid particles:bacteria ratio <1. This is important to couple phenotype and genotype. In this step 1 liter of DH5αF' was infected with $5x10^{11}$ phages. This culture was used for the production of a large quantity of phages that were cesium cloride purified and constitute the final library.

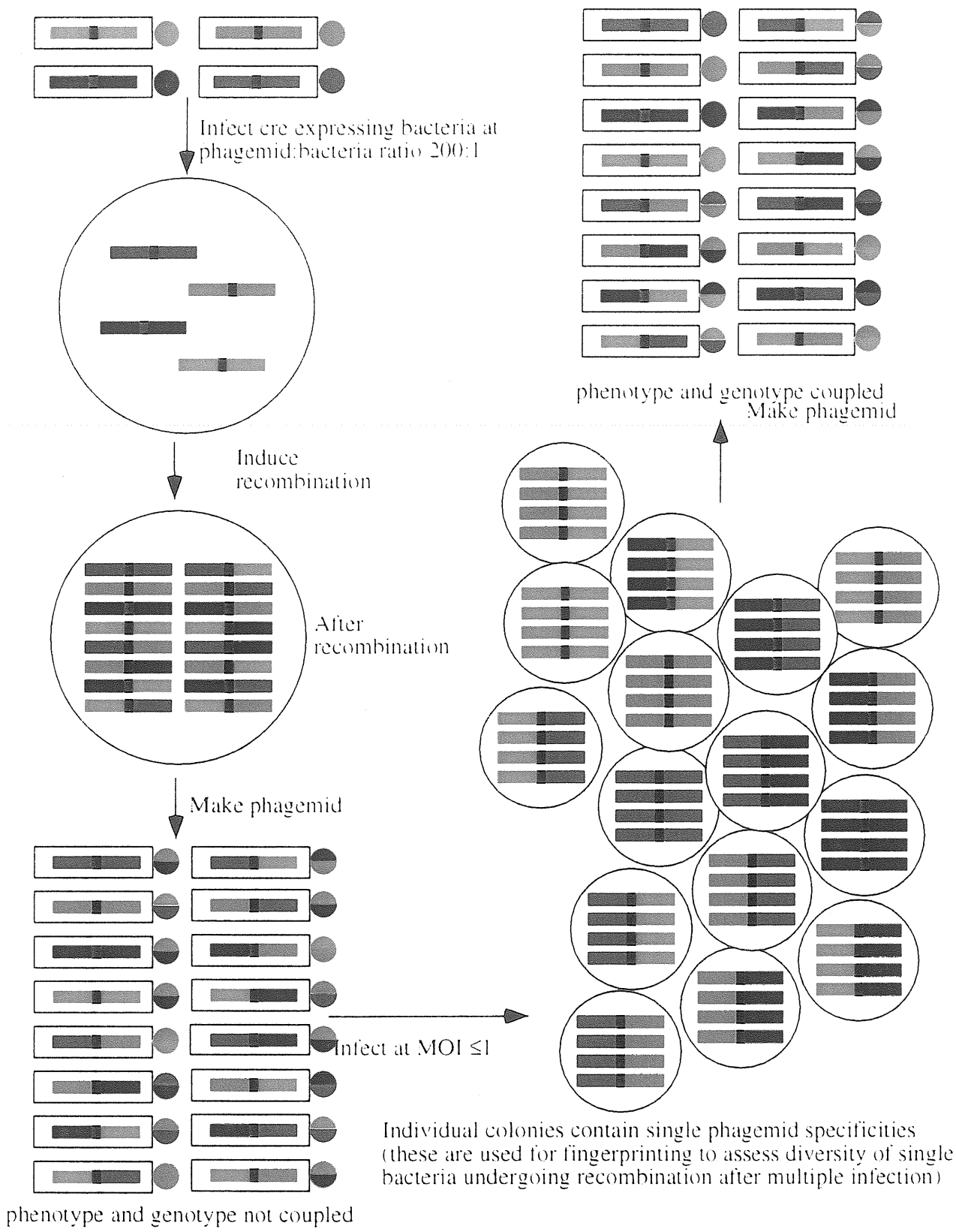


Fig 5.8 Scheme of the steps required to recombine the primary library and obtain a genotype phenotype coupled final library.

5.2.7 Assessing diversity and quality of the library

The diversity of this library was confirmed by DNA fingerprinting. scFvs were amplified from more than 100 colonies, 94% contain full length clones and after BstNI fingerprinting all were found to be different. This result confirmed that the system is stable as the percentage of deleted clones even after 2 additional overnight growth of the library is essentially identical

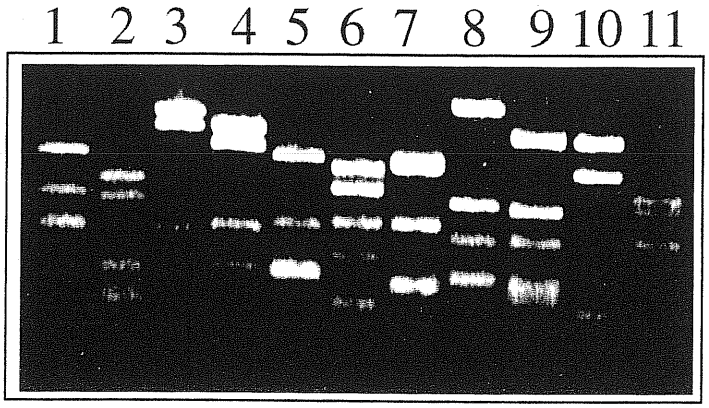


Fig. 5.9 DNA fingerprinting analysis of the clones of the recombined library. The scFv fragment from randomly selected clones was amplified with the primers PT2 and digested with BstNI. All the clone have a different pattern confirming the diversity of the library.

Also for the recombined library, 14 random clones were selected and the expression of recombinant p3 was assessed by western blot. The result show that more than 65% of the clones express a detectable level of p3-scFv fusion protein. As for the primary library the level of expression is variable from cell to cell.

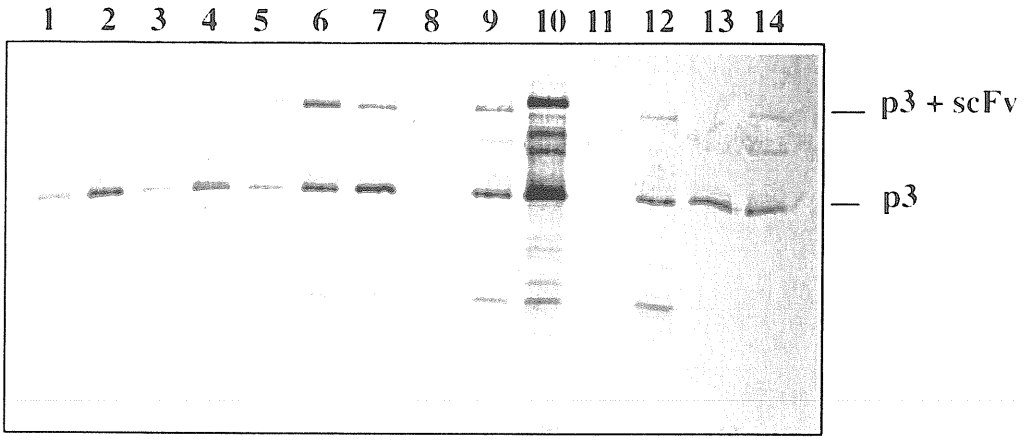


Fig 5.10 14 clones of the recombined final library ($3x10^{11}$) were selected randomly. Phagemid particles were prepared O/N at 30°C and concentrated by Peg precipitation. 10 µl of concentrated phages were loaded. Recombinant p3 protein was revealed with SV5 (Hanke, Szawlowski et al. 1992) antibody.

5.2.8 Testing the extent of *in vivo* recombination

The experiments described in the paragraphs above, indicated firstly that at least five different phagemid particles could infect a single bacteria, and second that two phagemid that entered a single bacteria were able to recombine with one another reaching an equilibrium state, resulting in four different phagemids inside a single cell. However, they did not give an indication of how long such phagemid would survive after growth in either liquid or solid culture. In order to determine what the real number of phages that could infect a single cell and also what the extent of recombination between these plasmids was, two different analysis were performed.

The first analysis was called CDR3 fingerprinting, this method is based on the detection of the VH CDR3 length which varies from 3 to 22 codons. A system to measure the length of all the VH CDR3s that are present simultaneously within a single cell that has been multiple infected was designed. The system works as described in fig 5.11. Briefly the V-genes from a single colony are amplified with primers annealing in the vector, the PCR product is then used as template for a single cycle of extension using a labelled primer annealing within the framework 3 region. Each V-gene with different CDR length will give a different extension product, all these products are then separated and visualized on a sequencing gel.

Infected cre bacteria after overnight recombination were plated and several colonies were selected and analysed following the method described. The results obtained shows the number of different VH CDR3s present in a single bacteria, after the multiple infection and after overnight recombination. As can be seen in tab 5.4 the average number of different VH CDR3s present in a single cell was 7. This is another confirmation that at least 7 different phages are able to simultaneously infect a single cell and be stable after overnight growth.

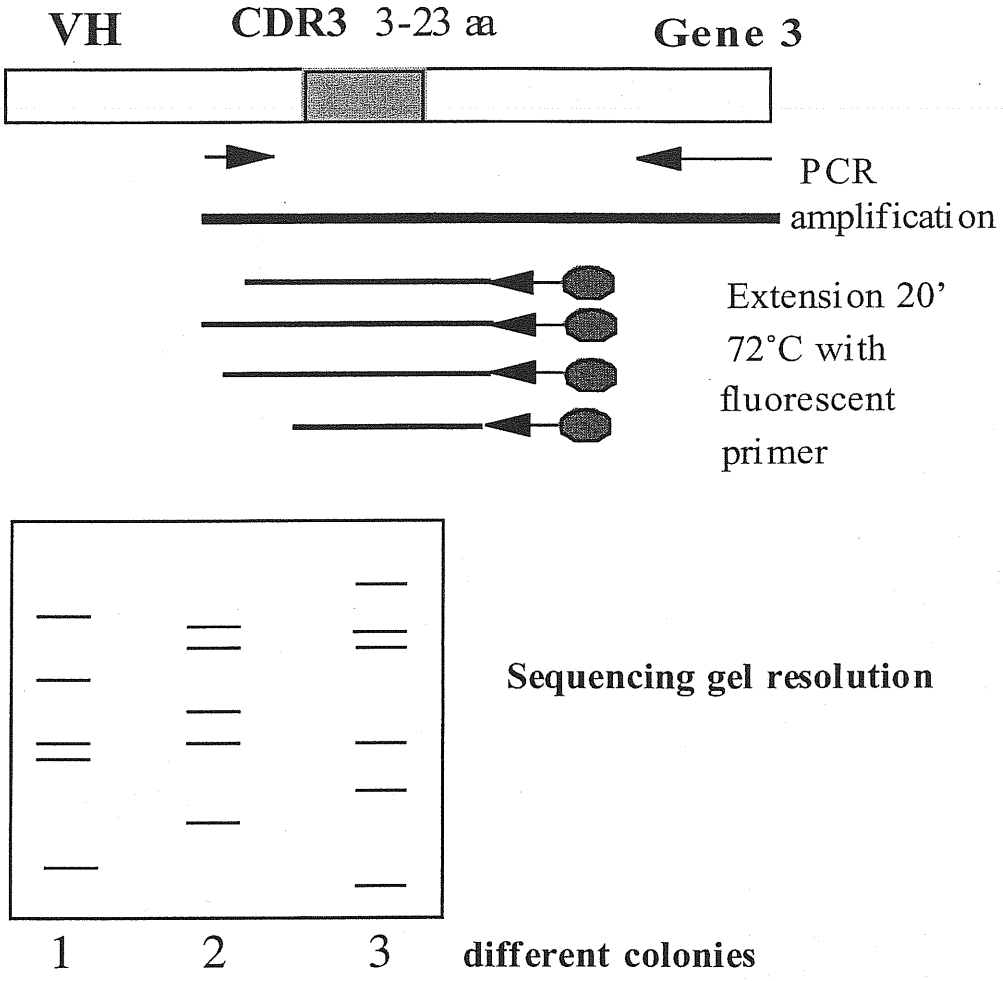


Fig 5.11 Scheme of VH CDR3 fingerprinting analysis of single colonies after recombination. The V-gene is first amplified with external primers and then PCR fragment is used as a template for a cycle of extension using a fluorescent labeled primer. The products of extension are then run on a sequencing gel and the number of different bands (corresponding of different VHCDR3 lengths) determined. A sequence reaction is run simultaneously to confirm that all the CDR3 bands differ in length by three or multiples of three nucleotides (indicating an in frame productive rearrangement).

	fingerprinting bands
colony 1	6
colony 2	8
colony 3	6
colony 4	7
colony 5	8
average 5 colonies	7

Tab 5.4 Results of VH CDR3 fingerprinting. 5 single colonies were analysed and the nnumber of different VH CDR3 present in each one was determined. The average of results is also reported.

This simple analysis revealed that the number of phages that can enter in a single cell is high, but it is clear that there are some limits to this technique. In fact, although the theoretical distribution of CDR3 length covers all the range between 3 to 22 aminoacids it is well known that the length of many VH CDR3s is concentrated in a reduced range of residues (9-15 see paragraph 5.2.12) and it is also clear that there are many VH genes having the same CDR3 length but different sequences. All these points lead to an underestimation of the real number of phages present in a cell and also does not indicate whether there has been recombination within the cell.

A more complete analysis was consequently designed (see figure 5.8): after overnight recombination, an aliquote of the culture of cre expressing bacteria was plated out to isolate individual (and multiple infected) colonies. If recombination has been successful, and all phagemid are still present, each of these colonies should contain many phagemids with multiple recombined V-genes. Phages were produced from these single colonies (these phages should consist of all the different scFv combinations which have arisen within the original starting cell) and used to infect DH5α cells in order to obtain single infected cells. This step is needed to isolate the many single scFv found in the starting colony. The scFvs present in a number of these colonies were then characterised by either sequencing or by separately amplifying VH and VL genes and fingerprinting them with BstNI. This analysis gave a more precise indication of the real number of different VH and VL present within a single bacterium expressing cre recombinase, and

furthermore, also gave an indication of the degree, if any, of recombination occurring.

Results obtained with the two methods (sequence and fingerprinting) were essentially identical, with different fingerprint patterns always representing different VH and VL genes, indicating that fingerprinting could substitute for sequencing in the assessment of diversity. In fact, if anything, fingerprinting tended to slightly underestimate the diversity, since in 12.5% of cases, (three cases out of twenty four) V-genes with identical BstNI fingerprints had different sequences.

On the basis of either sequencing or fingerprint patterns, the different VH and VL genes arising from a single cell were numbered. A total of five different individual colonies were analysed, with forty to eighty V genes analysed for each single colony. Results for all of these (3 colonies are shown in Tab 5.5 and 5.7) were very similar with 12-18 different VH and 12-15 different VL genes identified for each cell. These were present in 19-30 different scFv combinations, with the vast majority of scFvs containing shared V-genes. In the case of cell 1 (that analysed by sequencing) 11 examples of all four possible combinations of a pair of VH and VL genes can be identified (shown in bold in table 5.5), indicating that recombination has been extensive. Only one VH/VL pair does not appear to have participated in recombination in this cell.

In both cells, over 50% of the VH/VL pairs identified are present in single copies, suggesting that the small sample (37-41 scFv) analysed has not identified the full complement of VH and VL genes or their combinations, and that the true diversity is higher than that identified.

scFvs produced by single bacteria

Cell 1 (analysed by sequencing)

VL genes	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1									1	1		
2		1	1	1	1	3						5	
3			1	1									
4			1						1				1
5	1			2		1			1			1	
6							1						
7			1						2				
8									1				
9	3			2				1					
10			1	1									
11				1									
12						1							

13 different VH, 12 different VL out of a total of 35 different scFvs analysed
30 different VL-VH cominations, of which only 1 has not participated in recombination

Table 5.5: Single scFv genes from individual colonies of cre-expressing bacteria containing multiple recombined V genes were isolated by phagemid preparation and reinfection at low multiplicity of infection into DH5αF'. The individual VH and VL chains present in each phagemid were identified by sequencing and PCR amplification using V region specific primers and digestion with BstNI. The different sequences or fingerprint patterns found for VH and VL were numbered, with the VH genes listed horizontally and the VL fingerprints vertically. The number of scFvs found with a particular VH/VL combination is indicated in the corresponding cell. V genes which are derived from two VH/VL combinations in which all four possible recombination products have been identified are indicated in bold..

Colony 1

a) VH genes

Clone n.	Family	VH gene	CDR-3	BstNI fingerprint	sequence fingerprint
35	VH-1	DP-5	15	2	1
38	VH-1	DP-24	15	2	1
1	VH-1	DP5	15	2	1
28	VH-1	DP25	11	1	9
29	VH-1	DP-25	11	1	9
32	VH-1	DP-25	11	1	9
33	VH-1	DP-25	11	1	9
34	VH-1	DP-25	11	1	9
12	VH-2	V-II	9	10	5
42	VH-3	DP77	11	6	8
36	VH-3	DP-31	12	3	11
26	VH-3	DP-77	11	6	13
5	VH-4	DP-71	6	9	3
6	VH-4	DP-71	6	9	3
18	VH-4	DP-71	6	9	3
48	VH-4	DP71	6	9	3
22	VH-4	DP-71	6	9	3
43	VH-4	DP65	16	7	4
7	VH-4	DP-65	16	7	4
8	VH-4	DP-65	16	7	4
16	VH-4	DP65	16	7	4
24	VH-4	DP65	16	7	4
25	VH-4	DP-65	16	7	4
19	VH-4	DP-65	16	7	4
11	VH-4	DP-65	16	7	4
15	VH-4	VIV-DP71	15	5	6
39	VH-4	VIV/DP71	15	5	6
45	VH-4	VIV-DP71	15	5	6
21	VH-4	DP-65	18	11	7
40	VH-4	DP63	13	4	12
41	VH-4	DP63	13	4	12
47	VH-4	DP63	13	4	12
37	VH-4	DP-63	13	4	12
3	VH-5	DP-73	18	4	2
31	VH-6	DP-74	12	10	10

b) VL-genes

Clone n.	Family	VL gene	CDR3	BstNI fingerprint	sequence fingerprint
26	VK-1	DPK-6	9	3	4
34	VK-1	DPK6	9	3	4
6	VK-1	DPK6	9	3	4
21	VK-1	DPK8	9	11	6
42	VK-1	DPK1	9	7	9
24	VK-1	DPK1	9	7	9
38	VK-1	DPK1	9	7	9
11	VK-1	DPK1	9	7	9
19	VK-2	DPK15	8	10	11
31	VK-3	DPK21	10	1	1
36	VK-3	DPK21	10	1	1
1	VK-3	DPK21	10	1	1
3	VK-3	DPK22	7	3	2
18	VK-3	DPK22	7	3	2
47	VK-3	DPK22	7	3	2
37	VK-3	DPK22	7	3	2
39	VK-3	DPK22	7	3	2
25	VK-3	DPK22	7	3	2
40	VK-3	DPK22	7	3	2
12	VK-3	DPK22	7	3	2
45	VK-3	DPK22	8	8	12
22	VK-3	L16	9	2	7
28	VK-3	L16	9	2	7
29	VK-3	L16	9	2	7
16	VL-1	DPL7	12	9	10
48	VL-1	DPL7	12	9	10
5	VL-2	2b2	11	5	3
43	VL-2	2b2	11	5	3
33	VL-3	IGGLL	11	12	8
32	VL-9	DPL22	13	4	5
35	VL-9	DPL22	13	4	5
41	VL-9	DPL22	13	4	5
7	VL-9	DPL22	13	4	5
15	VL-9	DPL22	13	4	5
8	VL-9	DPL22	13	4	5

Tab 5.6 Sequences of the clones of cell N. 1. The family, the germline V-gene and the CDR3 length is indicated. In these tables the fingerprinting classifications made on the bases of BstNI digestion or sequence information are also indicated. This double analysis was performed to confirm that BstNI fingerprinting is almost as sensitive as sequencing for the identification of different V-regions and could be used without losing information about diversity.

cell 2 (analysed by BstNI fingerprinting)

VL genes \ VH genes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1							1									
2								1									
3			1														
4				1							1						
5					1												
6												1					
7													1				
8															1		
9	2		2														
10	3			1						1							
11	2	1			1				1								
12	2		1		1	1	1					2		1		1	3

17 different VH out of a total of 37 different scFvs
12 different VL
28 different VL-VH combination, of which 2 have not participated in recombination

cell 3 (analysed by BstNI fingerprinting)

VL genes \ VH genes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1																	
2		1							1									
3			1															1
4			5	1	1	5					1			1				
5	1					1	1	1		1							1	
6	2																	
7							1											
8			3				1								1			
9																1		
10								1										
11	1																	
12												1						
13							1											
14			1										1					

18 different VH out of a total of 41 different scFvs
14 different VL
29 different VL-VH combination, of which 2 have not participated in recombination

Table 5.7: Assessing the diversity of cre bacteria infected by multiple phagemid particles Two additional individual colonies were isolated after recombination and VH and VL genes were fingerprinted using BstNI. The different fingerprint patterns found for VH and VL were numbered, with the VH genes listed horizontally and the VL fingerprints vertically.

5.2.9 Selecting specific Abs from the library

The library produced after recombination was tested by selection on a number of different protein antigens (see table 5.8). The selection was performed using either immunotubes or selection in a 96 pin format. In the second case antigens were fixed on 96 pins and the selection was performed simultaneously for all the antigens in the first cycle and individually for each antigen in subsequent rounds. This new type of selection revealed to be very useful because it could be performed with very low quantity of antigen and 96 samples could be processed simultaneously working with 1 plate and using the same quantity of library as used for a single selection on immunotubes.

Antibodies were obtained against all fourteen antigens attempted. 32 or 48 individual colonies after 2 round of selection were tested in ELISA and positive clones fingerprinted. Even using such a small number of colonies a range of 3 to 11 different antibodies per antigen and a mean of over 6 antibodies per antigen were detected. The antibodies isolated were tested for crossrectivity against a number of unrelated antigens. All the clones showed high specificity and no polyreactivity as can be seen in fig 5.12

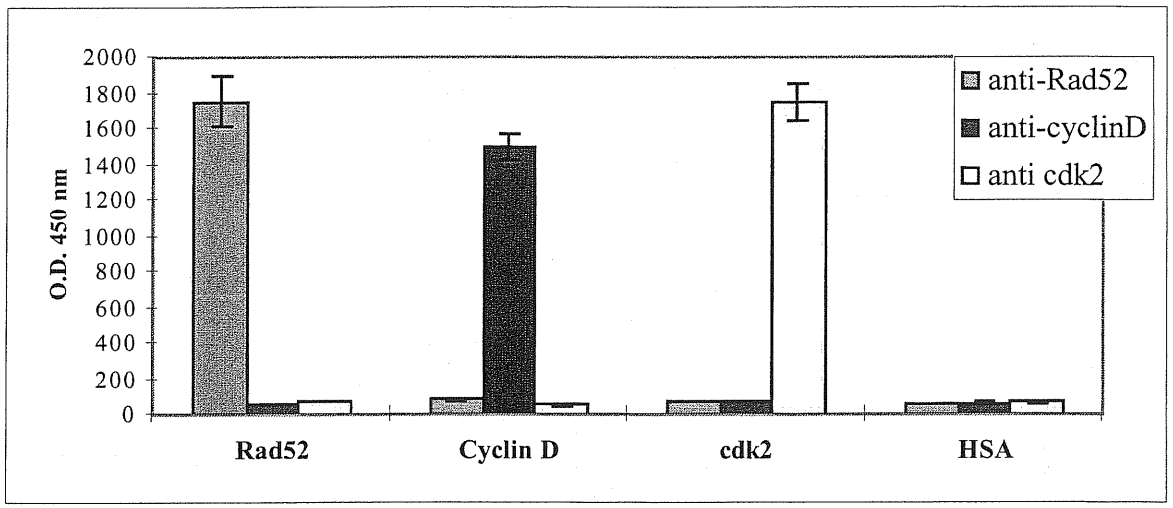


Fig 5.12 Analysis of the cross reactivity of the selected clones. All the clones were tested against the protein used for selection and positive clones further characterized with a series of irrelevant proteins

Antigens against which antibodies were selected

Immunotube selection:

antigen	source	tag	different scFv
Rad52	Human	his	11
PIGS12B	Pyrobaculum	his	4
cyclin D	Human	GST	4
cdk2	Human	GST	9
cdc25A	Human	GST	3
HSA	Human		7
loop2 gp120	HIV		6

Selection in microtitre plates

antigen	source	tag	positive colonies/8
FLAP	Human	his	4
PIGS 10	Pyrobaculum	his	3
cdc25C	Humam	GST	1
PARP DBD	Human	his	3
PARP 85K	Human	his	5
Tau	Human		1
Ku 70/80	Human		2

Table 5.8 Antiges against which antibodies have been selected
A list of all the antigens against which the library has been tested is given. For those selected in immunotubes (Nunc), this includes the number of different scFvs derived. For those selected in microtitre plates, the number of positives identified in 8 tested scFvs are given.

5.2.10 Affinity of the selected Abs

Eight of these scFv were purified by affinity chromatography using the His tag and characterised by gel filtration. All scFv were almost exclusively in the monomeric format, not showing the 'diabody' peak associated with scFv containing shorter linkers (Holliger, Prospero et al. 1993). A tipical gel filtration profile is shown in fig 5.13

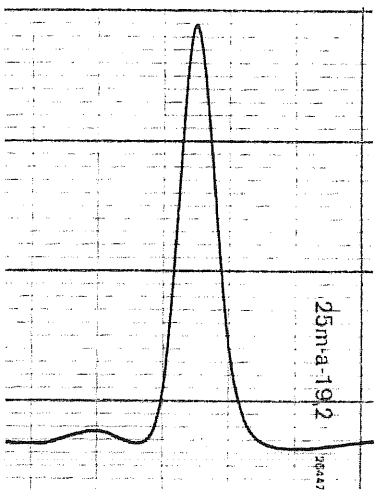


FIG 5.13 His tag purified scFv were gel filtrated to isolate the monomeric form. All the scFv purified showed a single peak elution profile corresponding to the monomeric form (Mon) of the scFv, the dimer was almost not visible (Dim).

Four of the monomeric scFv peaks were purified, and their dissociation equilibrium constant (K_d) were calculated from association and dissociation rate constants measured by surface plasmon resonance (BIACORE 2000). The results are summarised in tab 5.9 All scFvs had affinities less than 90nM, with the best having an affinity of 15nM.

Antigen	Clone	K_{on} $M^{-1}s^{-1}$	K_{off} s^{-1}	Kd
rad52	rad52-1	1.26e5	1.96e-3	1.56e-8
cdk2	cdk2-1	5.19e4	4.30e-3	8.29e-8
	cdk2-9	1.47e5	4.38e-3	2.97e-8
	cdk2-3	8.90e4	5.33e-3	5.98e-8

Tab 5.9 Summary of the affinity data determined by Biacore analysis. For each scFv the affinity, the Kon and Koff are reported. These data are obtained from at least 3 measure with different scFv concentration.

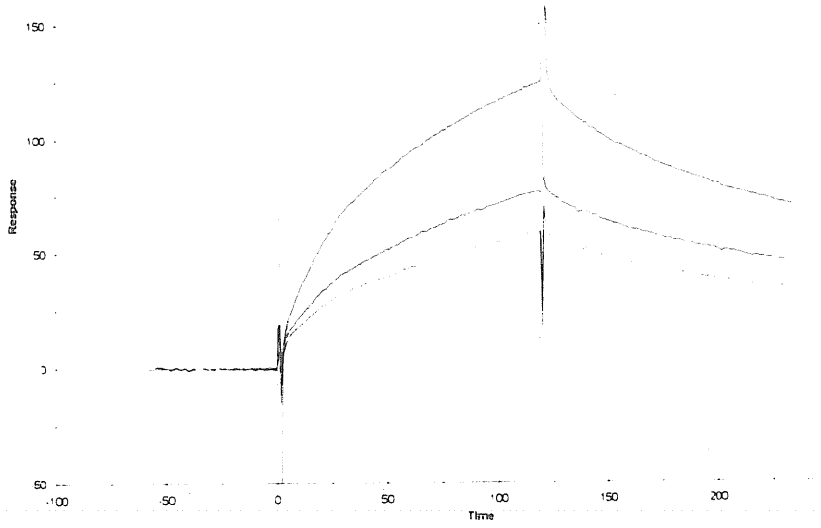


Fig 5.14 Biacore sensogram from an anti CDK2 scFv tested at differen concentration

5.2.11 V-gene sequence analysis

The distribution of V genes present in the primary library, the recombined library, a single cell and after selection was analysed by sequencing 22-26 V-genes for each category. After sequencing, the origin of the V gene and CDR3 length were determined using VBASE (Tomlinson, Williams et al. 1996) (see table 5.10)

a) Primary library sequences

	VH FAMILY	VH SEGMENT	CDR3-VH		VL FAMILY	VL SEGMENT	CDR3-VL
17	VH-1	DP-75	14	**15**	VK-1	DPK9	9
4	VH-1	DP-14	12	**28**	VK-1	DPK4	9
1	VH-1	DP-10	17	**32**	VK-1	DPK9	9
12	VH-1	DP-10	9	**30**	VK-1	DPK5	8
30	VH-1	DP-88	16	**27**	VK-1	DPK5	8
27	VH-1	DP-88	16	**14**	VK-1	L12	9
9	VH-2	DP-76	15	**25**	VK-3	DPK22	10
10	VH-2	DP-76	11	**20**	VK-3	DPK20	9
25	VH-2	DP-27	10	**17**	VK-3	DPK21	10
20	VH-3	DP-49	16	**19**	VK-4	DPK24	10
11	VH-3	DP-46	12	**31**	VK-4	DPK24	9
22	VH-4	DP78	10	**26**	VK-4	DPK24	10
15	VH-4	DP-63	7	**29**	VL-1	DPL7	11
29	VH-4	DP-63	17	**12**	VL-1	DPL5	11
18	VH-4	DP-63	17	**1**	VL-2	DPL11	11
19	VH-4	DP-63	10	**4**	VL-2	DPL11	11
31	VH-4	DP-70	10	**18**	VL-2	DPL11	10
32	VH-4	DP-64	14	**10**	VL-2	DPL11	11
14	VH-4	DP-71	10	**3**	VL-3	DPL23	9
28	VH-5	DP-73	8	**11**	VL-3	3h	11
3	VH-6	DP-74	16	**9**	ND	ND	ND
26	VH-6	DP-74	10	**22**	ND	ND	ND

b) Recombined library sequences

	VH FAMILY	VH SEGMENT	CDR3-VH		VL FAMILY	VL SEGMENT	CDR3-VL
7	VH-1	DP-10	22	17N	VK-1	DPK9	9
13N	VH-1	DP-10	18	10	VK-1	DPK9	9
13	VH-1	DP-10	13	11N	VK-1	DPK9	9
17N	VH-1	DP-15	13	22N	VK-1	DPK9	8
10	VH-1	DP-88	10	1	VK-1	L12	8
27	VH-2	DP-76	17	27	VK-1	L12	9
12	VH-2	DP-26	17	6	VK-1	L12	9
6	VH-2	DP-76	13	11	VK-2	DPK18	9
24N	VH-3	DP-37	13	24	VK-3	DPK21	9
15N	VH-4	DP-63	18	12	VK-3	DPK-21	9
1N	VH-4	DP-63	ND	7N	VK-3	DPK21	10
3N	VH-4	DP-63	14	6N	VK-3	DPK21	11
11N	VH-4	DP-65	19	14N	VK-3	DPK22	9
23N	VH-4	DP-65	15	21N	VK-3	DPK22	9
12N	VH-4	DP-65	13	23N	VK-3	DPK23	10
5N	VH-4	DP-66	17	1N	VK-4	DPK24	9
1	VH-4	DP-71	22	8N	VL-1	DPL7	9
15*	VH-4	DP-71	14	10N	VL-1	DPL8	10
24*	VH-4	DP-71	14	3N	VL-1	DPL5	11
10N	VH-4	DP-78	ND	24N	VL-2	DPL10	9
7N	VH-4	DP-79	8	15*	VL-2	DPL10	10
8N	VH-5	5-a	11	7*	VL-2	DPL10	10
22N	VH-5	DP-73	7	12N	VL-2	DPL12	9
11	VH-5	DP-73	13	13N	VL-2	DPL11	10
6N	VH-6	DP-63	16	15N	VL-3	DPL16	10
14N	VH-6	DP-74	11	13	ND	ND	ND
21N	VH-6	DP-74	8	5	ND	ND	ND

c) Selected clones sequences

	VH FAMILY	VH SEGMENT	CDR3-VH		VL FAMILY	VL SEGMENT	CDR3-VL
14.2	VH-1	DP-8	12	**LAST**	VK-1	L12	9
16.2	VH-1	DP25	14	**1.2**	VK-1	DPK1	9
12.1	VH-1	DP-14	8	**13.2**	VK-1	L12	9
4.2	VH-1	DP-7	9	**14.1**	VK-2	DPK18	8
C-TER	VH-1	DP-75	12	**16.2**	VK-3	DPK21	8
N-TER	VH-1	DP-88	9	**10.2**	VK-3	L6	9
LAST	VH-1	DP-75	13	**9.1**	VL-1	DPL3	10
1.2	VH-1	DP-88	11	**7.2**	VL-1	DPL1	11
13.2	VH-2	DP-27	16	**12.1**	VL-1	DPL2	12
13.1	VH-3	DP-35	16	**C-TER**	VL-1	DPL5	13
2.1	VH-3	DP-47	13	**14.2**	VL-1	DPL5	11
10.2	VH-3	DP-47	14	**11.1**	VL-2	DPL11	11
7.2	VH-3	DP-47	9	**2.1**	VL-3	DPL23	11
5.1	VH-4	DP-70	11	**7.1**	VL-3	DPL23	10
3.2	VH-4	DP-71	11	**N-TER**	VL-3	3h	12
6.1	VH-5	DP-73	12	**4.2**	VL-3	3h	10
7.1	VH-5	DP-73	12	**5.1**	VL-3	3h	11
15.1	VH-5	DP-73	17	**5.2**	VL-3	3h	8
8.1	VH-6	DP74	11	**6.1**	VL-3	3h	11
9.1	VH-6	DP74	8	**15.1**	VL-3	3h	11
11.1	VH-6	DP-74	13	**3.2**	VL-6	6a	10
14.1	VH-6	DP-74	14	**13.1**	ND	ND	ND
5.2	VH-7	DP21	16	**8.1**	ND	ND	ND

Tab 5.10 Sequence analysis of the clones from **a)** primary library; **b)** recombined library **c)** selected clones. The family, V-genes germline and CDR lenght is indicated.

As can be seen in figure 5.15, the primary and recombined library are diverse, with almost all families represented. VH genes for both libraries were derived from 6 of the 7 gene families and from 14 different germline genes for the primary and 15 for the recombined. VL genes were derived from 3 (4 for recombined) of the 6 VK families and 3 of the 9 Vλ families with 8 VK germline genes and 5 Vλ (7 for recombined) being used. These V-genes are not evenly distributed, with VH genes having a predominance of VH4 genes, and VL genes having more VK1 and VK3 genes. There is no great difference between the primary and recombined libraries, indicating that diversity does not appear to be compromised by recombination.

The distribution of V-genes found in a single cell is also similar to that found in both the primary and recombined library, with the identification of an extra Vλ gene, indicating that the recombination identified within a single cell mirrors that found in the library as a whole.

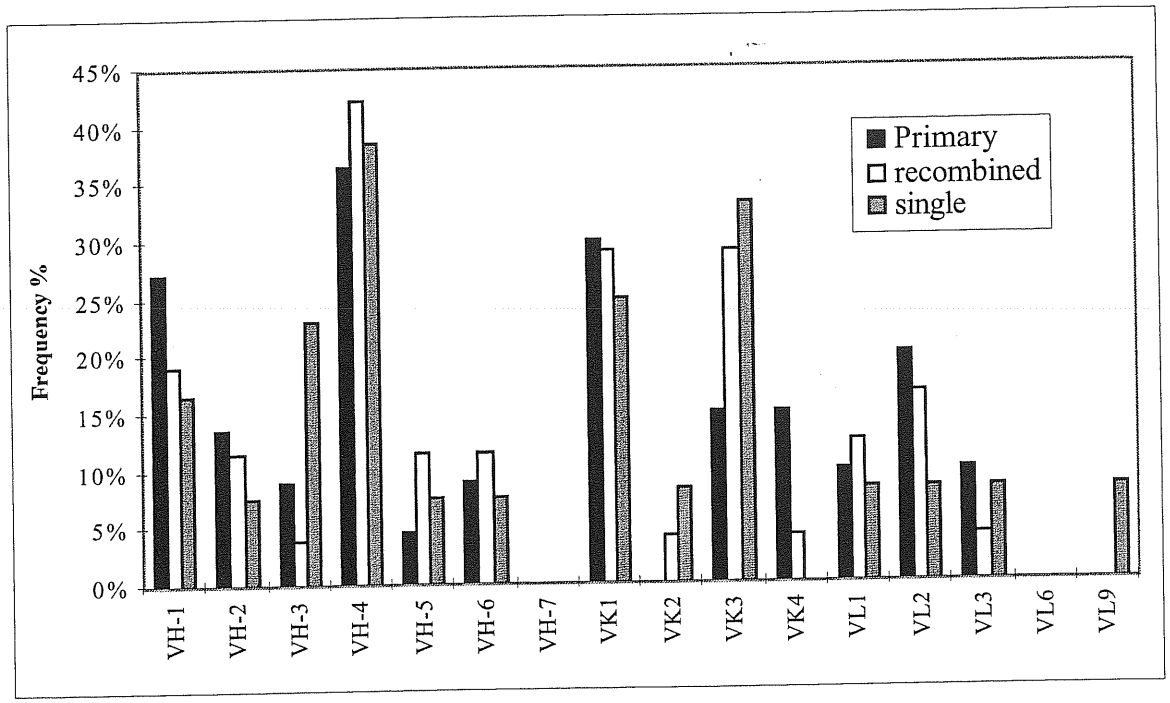
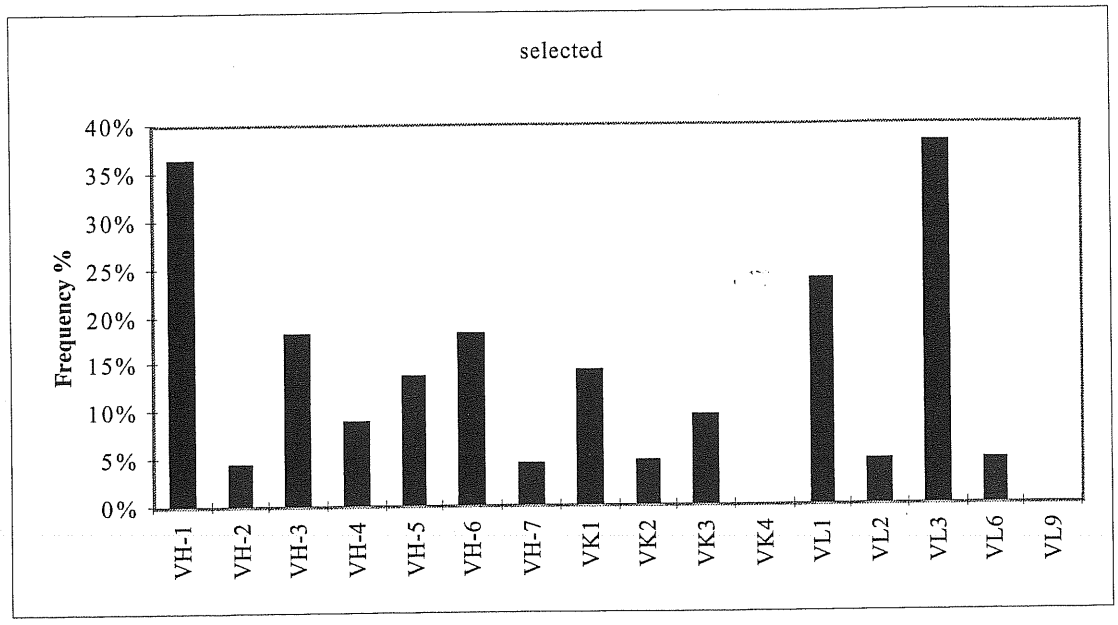


Fig 5.15 Frequency of use of human VH, VK and Vλ segments from the primary, recombined and single clone library. Frequency is plotted as % of total. VH, VK and Vλ are listed by family name.

Upon selection, the V-gene distribution changes somewhat, as has been previously shown (Vaughan, Williams et al. 1996; Sheets, Amersdorfer et al. 1998). V-genes were derived from all 7 VH families and use 14 different VH germline genes. The light chains derive from 3 VK families and 4 Vλ families with 5 K and 9 λ different germline genes used. While the primary and recombined libraries have an excess of VH4 genes, the V-genes found in antibodies recognising selected antigens appear to be more widely distributed, with VH1, VH3, VH5 and VH6 being the commonest VH gene classes found. In the VL genes, 38% of all selected antibodies contain a VL gene of the VL3 family, even though this accounts for less than 10% of the primary and recombined libraries.

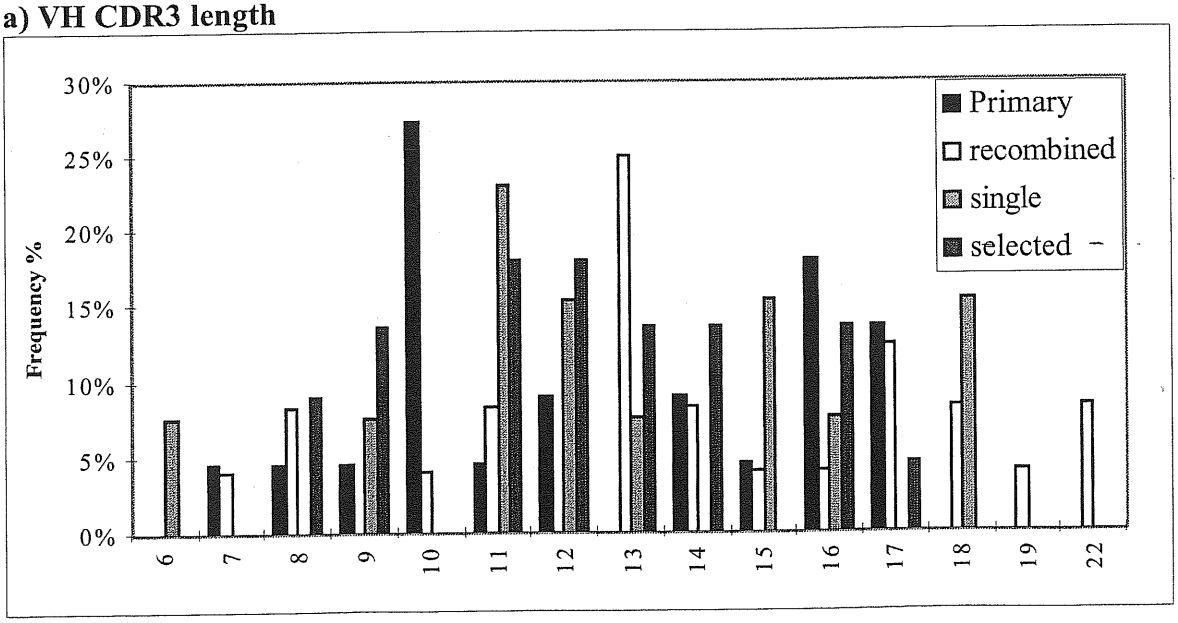


5.16 Frequency of use of human VH, VK and Vλ segments from the selected clones. Frequency is plotted as % of total. VH, VK and Vλ are listed by family name.

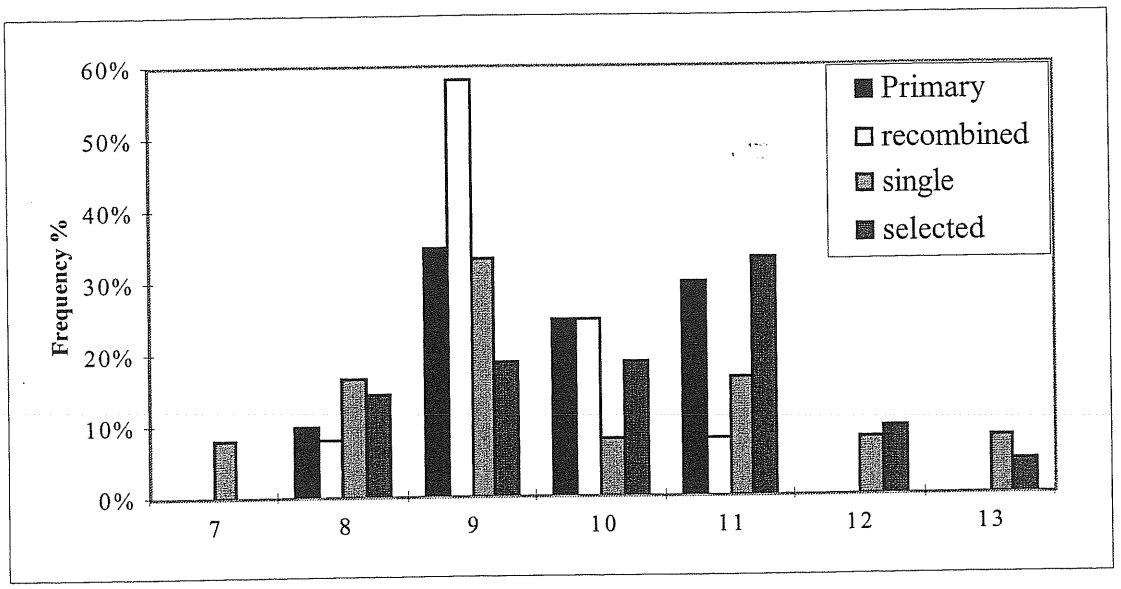
5.2.12 CDR3 length distribution

The length and the sequence of the CDR3 loops of the sequenced clones was also analysed. The length does not appear to change very much between the four populations of antibodies, with a wide distribution of all lengths found, ranging between 6-22 aminoacids for the VH genes with 55% to 72% of the clones concentrated in the range of 10-16 residues. The VL genes have a CDR3 distribution that cover the range of 7-13 residues.

a) VH CDR3 length



b) VL CDR3 length



5.17 CDR3 length distribution of the **a)** VH genes and **b)** VL genes from all the sequenced clones.

5.3 Discussion

5.3.1 Phage antibody libraries

Even though many results have been obtained with phage-displayed peptides and antibodies (Marks, Hoogenboom et al. 1991; Griffiths, Williams et al. 1994; Nissim, Hoogenboom et al. 1994; de Kruif, Boel et al. 1995; Vaughan, Williams et al. 1996; Sheets, Amersdorfer et al. 1998), there is still a need to improve this technology as the size of a useful library requires at least 10^9 independent clones. This represent only a very small fraction of the natural antibody repertoire. The construction of very large repertoires is an indispensable step to obtain "high affinity binders", antibodies directed against non classical antigens such as small organic molecules or antibodies with peculiar properties such as catalytic activity. Currently, the bottleneck is the efficiency of the transformation (electroporation) step, furthermore the library once made become a limited resource, as attempts to amplify such libraries result in a reduction of diversity each time the library is amplified.

Making larger and larger libraries, is not, however, an end in itself. The encoded sequence space of a library cannot be fully exploited unless the expressed shape space can be adequately sampled. If, for example, a library with 10^{14} elements was created, this would represent more clones than the number of phage which can be introduced into a typical selection procedure, which is 10^{13}. This problem is exacerbated when one considers the display levels, which with the best phagemid libraries, is probably not greater than 10% (Clackson, Hoogenboom et al. 1991) indicating, that in a standard selection, using 10^{13} phagemid, diversity greater than 10^{12} remains untapped.

5.3.2 *In vivo* recombination as a solution

Some recombination methods used or proposed to make antibody libraries (Waterhouse, Griffiths et al. 1993; Geoffroy, Sodoyer et al. 1994; Griffiths, Williams et al. 1994; Tsurushita, Fu et al. 1996) use two plasmids to generate the library. The heavy chains are cloned on one plasmid, and the light chains on another. After recombination four plasmid populations are generated, only one of which is correct (containing heavy and light library chains). The others contain copies of either single library chains in combination with dummy chains, or dummy chains alone. Although these

extra products should not package into phage, we have found that plasmids which lack an f1 origin of replication can still do so, and as a result will contaminate the library and so reduce the real diversity of the library, moreover these systems suffer severe problems of stability of the vector system.

In this chapter we have described the use of a "single vector" to generate diversity. This method have the following advantages: large (10^{12} -10^{14}) phagemid and phage libraries can be made quickly; all recombination products are functional; recombination after each cycle accesses even greater diversity (square of starting library size); affinity maturation can be simultaneous for both VH and VL; can recreate true VH-VL pairings (among others) and many more. All these points will be discussed in the next paragraphs.

5.3.3 Multiplicity of infection, *in vivo* recombination and stability of the vector.

The first requirement to develop a single vector recombination system is that infection can be used as a way to introduce multiple copies of the same vector (carrying different recombination products) into a bacteria and that these are sufficiently stable once introduced, to allow recombination to occur.

The experiments made with the phagemid particles carrying different antibiotic resistances shows immediately that at least 5 phages can infect one cell simultaneously. This is the first demostration that a cell could be infected by more than 3 phages, overcoming the limit described by (Marvin and Hohn 1969), although it should be noted that we have used phagemid particles in our experiments and Marvin et al., used native phage. The stability (compatibility) of the plasmids with the same origin of replication inside an infected cell was not directly proved at this stage. The only indirect evidence came from the fact that different antibiotic resistences could be retrived in the same cell after overnight growth in the presence of a selective pressure. Data confirming the plasmid stability will be discussed later. These results open the way to the creation of a new phagemid vector for *in vivo* recombination. The vector pDAN3 that in the previous chapter we have shown to work extremely well as a phagemid vector for phage display was modified to become also a recombination vector. In particular, we have used the *loxP* 511 sequence as part of the linker coding sequence and wild type

loxP as the second sequence used for the recombination of the scFv. This construct was used in a bacteria expressing cre constitutively.

The recombination system was tested with the D1.3 mAb and this confirmed, firstly that multiple infection can occur, and more importantly that the recombination process was very efficient and stable. This was shown by the fact that no deletion occurs and the reaction goes to equilibrium with the 4 possible recombination products present at similar percentages after recombination.

5.3.4 Experimental support and theoretical arguments to evaluate the degree of diversity

Following the results obtained in the model experiments a library of $7x10^7$ different colonies was constructed. This was used to infect cre expressing bacteria to induce recombination. Many different analyses were designed to test the stability and the extent of recombination.

One worry with the use of the cre/lox is the perceived instability of the system. An analysis of the number of full length scFvs by PCR showed that 4/96 (4%) of the primary library and 6/96 (6%) of the recombined library contained genes which had either the VH or VL gene deleted. These are likely to have occurred during library construction, probably as a result of spurious priming during the PCR assembly procedure, and not in any way linked to the presence of the lox site, which was found to be preserved in all sequenced scFvs, including that which did not appear to have participated in recombination. The limited increase in the percentage of deletion probably represent the growth advantage of preexinting deleted clones (de Bruin, Spelt et al. 1999) during the growth cycle required for the recombination process. These data indicate that scFvs containing lox sites are no less stable than others created using other systems, and suggest that the presence of a lox site, does not in itself, confer any tendency to instabilty or deletion.

The degree of diversity created in a single cell is difficult to assess accurately. The simple CDR3 finger printing shows that 7 phagemids can enter a cell and remain stable overnight but do not give any indication about whether recombination occurs. The definitive results come from the single cell sequence and fingerprinting analysis. The five colonies analysed give similar results, and those obtained for colony one (which was assessed by sequencing) will be discussed. 13 different VH genes and 12 different VL

genes were identified. These were recombined in a total of 30 different combinations, with all except one scFv containing VH or VL genes also found in other scFvs. One light chain was found with six different heavy chains, one heavy chain with six different light chains, and eleven overlapping cases of all four combinations of a VH/VL pair can be identified (shown in bold in Table 5.5), showing how extensive recombination has been. Interestingly, in all five cells analysed, no single scFv dominates the analysis, with the most abundant being present in only five copies (14%), indicating that all scFv appear to have similar probabilities of remaining within the cell. The single sequenced scFv which did not appear to have participated in recombination was normal with a functional lox site present. Consequently, its non-participation is probably a result of not having analysed enough clones, rather than an intrinsic problem of the scFv itself.

The minimum diversity generated by a single cell can be estimated to be the number of recombined scFv actually observed: 26-30. However, if the 13-18 different VH genes rescued from a single colony are matched by 13-18 different VL genes (which is certain if the diversity of the primary library is as high as observed), the potential diversity identified in this small sample is 169-324 (13^2-18^2). It is likely that not all the different V genes present were identified, and we have shown that V genes can have identical fingerprints but different sequences, suggesting that the diversity created in a single cell is likely to exceed this, giving a maximum estimate which approaches the 500-700 copy number of pUC based plasmids (Sambrook, Fritsch et al. 1989).

In the secondary library created here, 10^{10} bacteria (20ml) expressing the cre recombinase were used. On the basis of the diversity actually observed (each bacteria produces 26-30 recombined phage), the final library size will be approximately $3x10^{11}$ (30×10^{10}). While if all possible V gene combinations are present, the library will be up to ten times larger ($3.2x10^{12}$). The size of the library which can be used practically, however, is limited by the reinfection step at low MOI when phenotype and genotype are coupled. By using one litre to perform this step, it cannot exceed $5x10^{11}$, the number of bacteria in this volume. On the basis of the volumes used this gives an estimate of the final diversity of $3\text{-}5x10^{11}$.

These figures are clearly theoretical and must be tempered by upstream limiting factors. The number of B lymphocytes used to create the library can be estimated to be $5x10^8$. These are derived from forty different donors, and although they may represent $5x10^8$ different VH gene specificities, it is unlikely that the number of VL genes is this high. Another

step in which diversity can be compromised is in the PCR assembly. This is known to be difficult and can be overtaken by a small number of V genes. Although fingerprint analysis of the assembled band indicated diversity, the degree is impossible to assess and could be lower than the $7x10^7$ clones in the primary library.

5.3.5 V-genes usage in the primary library and selected clones

The possibility that the recombination process itself may induce bias in the library was examined by sequencing V-genes from the primary and recombined library. An increase in the representation of VK3 and VK2 genes, and a reduction in VH1 and VK4 genes in the recombined library compared to the primary was found, but in general the ratio of the different V-genes is remarkably well preserved. Interestingly, neither the primary nor the recombined library was dominated by single V gene families, as has been found in other naïve libraries, and representative genes from almost all VH and VL gene families were found. This may be due to several reasons, first we used a set of primers (Sblattero and Bradbury 1998), which was specifically designed to be able to amplify all known germline V-genes. This guaranteed that all the V families are amplified with the same probability. Second amplification of V-genes was carried out with individual primers, rather than mixtures. This protocol allow the amplification of all the V-gens and doesn't give the problems of biased amplification towards some families or specific genes that are much more represented in the natural repertoire. Third the reamplification and assembly steps carried out with external primers were designed to mantain the original diversity.
The diversity of V-gene families in the primary and recombined libraries is also reflected in the V-genes of scFvs selected for antigen binding. Although VL3 genes are frequently found, as has been shown for scFvs selected from other published naïve libraries (Vaughan, Williams et al. 1996; Sheets, Amersdorfer et al. 1998), there is no predominance of other V gene families, and members of almost all V gene families can be found. Of 22 different selected scFvs sequenced, 14 different VH genes, and 13 different VL genes were found, with no example of identical V genes being found in two scFvs recognising different antigens.

5.3.6 Selection procedures and affinities isolated

The affinities of the antibodies isolated were all better than 90nM, with the best having an affinity of 15nM. This is lower than the best affinities reported for the larger libraries (de Haard, van Neer et al. 1999), which in some cases (Vaughan, Williams et al. 1996; Sheets, Amersdorfer et al. 1998), were subnanomolar. However, several consideration about these results could be made. First the highest affinities reported in these papers were all obtained by selection in the soluble phase using biotinylated antigen and magnetic streptavidin beads, or alternatively by selecting on haptens. When selections were performed as described here (protein antigens coupled to immunotubes), the affinities obtained were similar (see Sheets et al.(Sheets, Amersdorfer et al. 1998), for a full discussion of this point). We selected only on proteins and perform the classical immunotube selection that is known (due to the high density of the antigen) to retrive also relatively low affinity binders. Furthermore, the bacterial elution method used here has recently been shown to be far less efficient than more stringent methods, such as 100mM HCl pH1.1 or 100mM Triethylamine, at eluting high affinity antibodies (Schier and Marks 1996), suggesting that the antibodies with the highest affinities may well have been left on the immunotube. Furthermore, we only measured the affinities of four scFvs and these have all been presented. ScFvs with better K_d's could probably have been found by increasing the number of scFvs analyzed.

5.3.7 Further exploitation of the *in vivo* recombination system

The search for larger and larger libraries reaches a practical limit in the volume required to perform a selection. By including cycles of recombination between rounds of selection, the effective diversity accessed would be expected to approach the theoretical maximum diversity of the library. Binding V-regions could either be shuffled against one another, or against the starting library. This would be the equivalent of shuffling the chains of all binding antibodies in parallel, in a fashion similar to that which has been carried out in series for single antibodies with notable increases in affinity (Marks, Griffiths et al. 1992; Schier and Marks 1996). We expect the affinities of antibodies obtained in this way to be higher than those reported here.

Affinity maturation, with mutations introduced simultaneously in both heavy and light chains, could also be carried out in a similar way, with the advantage that all potential combinations of mutations can be sampled simultaneously.

The ease with which libraries can be made using this method should also allow the testing of presently uncontrolled variables (such as V gene source, leader sequence, promoter, effect of different tags) in library construction, as well as allow the creation of very large libraries from any source of V genes, including synthetic V genes or limiting quantities of material derived from patients.

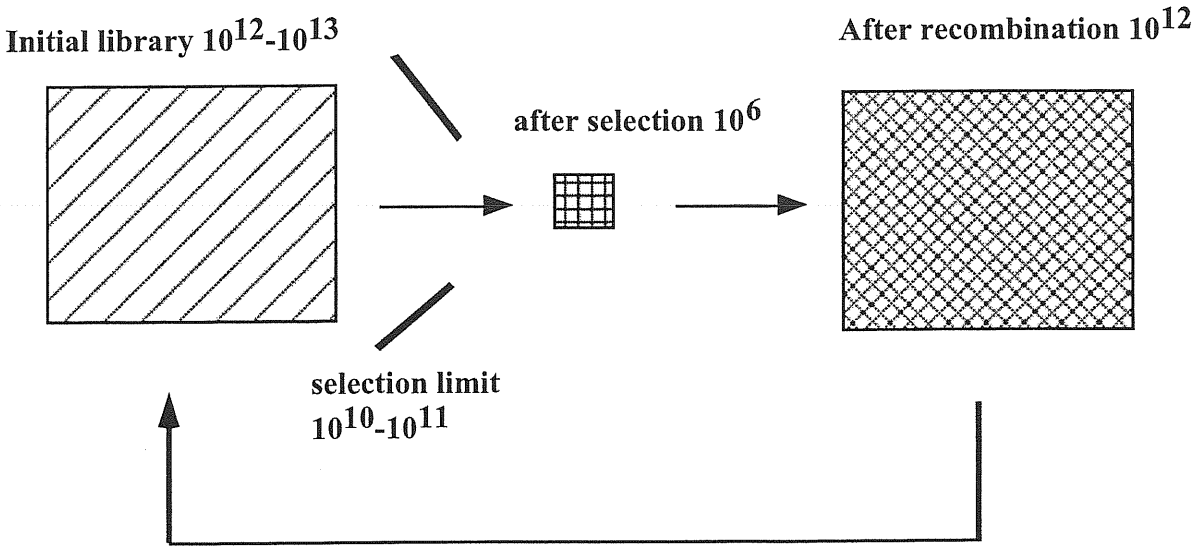


Fig. 5.18 Accessing the diversity. The full diversity of a very library remain untapped, by including cycles of recombination between selection the effective diversity could be accessed. This approach is equivalent of the V chain shuffling.

This method can also be used to create large libraries of any other protein (natural or artificial) displayed on phage (or not) in which a site-specific recombination sequence (e.g. *loxP*, or derivative) can be inserted, such that diversity can be created in the two parts of the protein separated by the linker, or in two polypeptide chains which can be covalently joined by such a linker. The use of the combination of the *loxP* or derivative sites and cre recombinase is also not necessary. Any other combination of specific recombinase recognition site and specific recombinase may also be used (e.g. Flp recombinae and Flp recombinase target sites). Furthermore, such site specific recombination sequences may also be inserted as non-coding sequences between two polypeptide chains which are separately synthesised and become associated, covalently or not, following synthesis (e.g. Fab chains, T cell receptors, cell surface receptors).

5.3.8 Conclusions

Here we describe a method which is capable of making libraries so large that the complete diversity cannot be accessed using traditional phage technology. The method involves the creation of a primary phage scFv library in a phagemid vector containing two lox sites which are unable to recombine with one another. Infection of cre recombinase expressing bacteria by such a primary library at a high multiplicity of infection, results in the ingress of many different phagemid and the exchange of VH and VL genes between different phagemids, creating many new VH/VL combinations, all of which are functional. On the basis of the observed recombination, the library is calculated to have a potential diversity between 10^{11} and 10^{12}. The library size is not limited by transfection efficiency, because it is created within bacteria, and as we have shown, the library size is limited by the number of bacteria or phage which can be grown. The fact that recombination is used to generate enormous diversity from small starting libraries means that diversity is regenerated each time recombination is used to create each new libraries. As a result sufficient library for 10^7 selections can be made from an initial primary library, whereas the number of selections which can be made from a standard ligation and transfection library is far lower.

A library created using this method was validated by the selection of antibodies, with high affinity, against a large number of different protein antigens.

Bibliography

Adams, D. E., J. B. Bliska, et al. (1992). "Cre-lox recombination in Escherichia coli cells. Mechanistic differences from the in vitro reaction." J Mol Biol **226**(3): 661-73.

Al-Lazikani, B., A. M. Lesk, et al. (1997). "Standard conformations for the canonical structures of immunoglobulins." J Mol Biol **273**(4): 927-48.

Alexeyev, M., I. Shokolenko, et al. (1995). "Improved antibiotic-resistance gene cassettes and omega elements for Eschericia coli vector construction and in vitro deletion/insertion mutagenesis." Gene **160**: 63-67.

Alt, F. W. and D. Baltimore (1982). "Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions." Proc Natl Acad Sci U S A **79**(13): 4118-22.

Andersen, P. S., A. Stryhn, et al. (1996). "A recombinant antibody with the antigen-specific, major histocompatibility complex-restricted specificity of T cells." Proc. Natl. Acad. Sci. U.S.A. **93**(5): 1820-1824.

Ayala, M., R. F. Balint, et al. (1995). "Variable region sequence modulates periplasmic export of a single-chain Fv antibody fragment in Escherichia coli." Biotechniques **18**(5): 832, 835-8, 840-2.

Balass, M., Y. Heldman, et al. (1993). "Identification of a hexapeptide that mimics a conformation-dependent binding site of acetylcholine receptor by use of a phage-epitope library." Proc. Natl. Acad. Sci. U.S.A. **90**: 10638-10642.

Barbas, C. F., A. S. Kang, et al. (1991). "Assembly of combinatorial antibody libraries on phage surfaces: The gene III site." Proc. natl. Acad. Sci. USA **88**: 7978-7982.

Barbas, S. M., H. J. Ditzel, et al. (1995). "Human autoantibody recognition of DNA." Proc Natl Acad Sci U S A **92**(7): 2529-2533.

Bauer, T. R. J. and B. Blomberg (1991). "The human λ L chain Ig locus. Recharacterization of JCλ6 and identification of a functional JCλ7." J. Immunol. **146**: 2813-2820.

Better, M., C. P. Chang, et al. (1988). "Escherichia coli secretion of an active chimeric antibody fragment." Science **240**(4855): 1041-3.

Bird, R. E., K. D. Hardman, et al. (1988). "Single-chain antigen-binding proteins [published erratum appears in Science 1989 Apr. 28;244(4903):409]." Science **242**(4877): 423-426.

Bird, R. E. and B. W. Walker (1991). "Single chain antibody variable regions." Trends Biotech **9**: 132-138.

Blond-Elguindi, S., S. E. Cwiria, et al. (1993). "Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP." Cell **75**: 717-728.

Boeke, J. D., P. Model, et al. (1982). "Effects of bacteriophage f1 gene III protein on the host cell membrane." Mol Gen Genet **186**(2): 185-92.

Bonnycastle, L., J. Mehroke, et al. (1996). "Probing the basis of antibody reactivity with a panel of constrained peptide libraries displayed by filamentous phage." J Mol Biol **258**: 747-762.

Boyd, D. and J. Beckwith (1990). "The role of charged amino acids in the localization of secreted and membrane proteins." Cell **62**(6): 1031-3.

Bradbury, A., L. Persic, et al. (1993). "From gene to antibody: the use of living columns to select specific phage antibodies." BioTech. **11**: 1565-1569.

Brown, S. (1992). "Engineered iron oxide-adhesion mutants of the Escherichia coli phage lambda receptor." Proc Natl Acad Sci U S A **89**(18): 8651-5.

Charbit, A., G.-C. Boulain, et al. (1986). "Probing the topology of a bacterial membrane protein by genetic insertion of a foreign epitope. Expression a the cell surface." EMBO J. **5**: 3029-3037.

Chester, K. A., R. H. Begent, et al. (1994). "Phage libraries for generation of clinically useful antibodies." Lancet **343**(8895): 455-6.

Chomczynski, P. and N. Sacchi (1987). "Single step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction." Anal. Biochem. **162**: 156-159.

Clackson, T., H. R. Hoogenboom, et al. (1991). "Making antibody fragments using phage display libraries." Nature **352**: 624-628.

Clackson, T. and J. A. Wells (1994). "In vitro selection from protein and peptide libraries." TIBTECH **12**: 173-184.

Cook, G. P. and I. M. Tomlinson (1995). "The human immunoglobulin VH repertoire." Immunol Today **16**(5): 237-242.

Cook, G. P., I. M. Tomlinson, et al. (1994). "A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q." Nat Genet **7**(2): 162-168.

Corbett, S. J., I. M. Tomlinson, et al. (1997). "Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, "minor" D segments or D-D recombination." J Mol Biol **270**(4): 587-97.

Cortese, R., F. Felici, et al. (1994). "Epitope discovery using peptide libraries displayed on phage." TIBTECH **12**: 262-267.

Courtney, B. C., K. C. Williams, et al. (1995). "A phage display vector with improved stability, applicability and ease of manipulation." Gene **165**(1): 139-140.

Cox, J. P., I. M. Tomlinson, et al. (1994). "A directory of human germ-line V kappa segments reveals a strong bias in their usage." Eur J Immunol **24**(4): 827-836.

Craig, N. L. (1988). "The mechanism of conservative site-specific recombination." Annu Rev Genet **22**: 77-105.

Cwirla, S. E., P. Balasubramanian, et al. (1997). "Peptide agonist of the thrombopoietin receptor as potent as the natural cytokine." Science **276**: 1696-1699.

Cwirla, S. E., E. A. Peters, et al. (1990). "Peptides on phage: a vast library of peptides for identifying ligands." Proc. Natl. Acad. Sci. U.S.A. **87**(16): 6378-82.

Dariavich, P., G. Lefranc, et al. (1987). "Human immunoglobulin Cλ6 gene encodes the $Kern^{+}Oz^{-}$ λ chain and Cλ4 and Cλ5 are pseudogenes." Proc. Natl. Acad. Sci. USA **84**: 9074-9078.

Davidkova, G., S. Pettersson, et al. (1997). "Selective usage of VH genes in adult human B lymphocyte repertoires." Scand. J. Immunol. **45**: 62-73.

Davies, D. R. and G. H. Cohen (1996). "Interactions of protein antigens with antibodies." Proc Natl Acad Sci U S A **93**(1): 7-12.

Davies, J. and L. Riechmann (1995). "An antibody VH domain with a lox-Cre site integrated into its coding region: bacterial recombination within a single polypeptide chain." FEBS Lett **377**(1): 92-96.

De Bellis, D. and I. Schwartz (1990). "Regulated expression of foreign genes fused to lac: control by glucose levels in growth medium." Nucleic Acids Res **18**: 1311.

De Boer, M., S.-Y. Chang, et al. (1994). "Desing and analysis of PCR primers for the amplification and cloning of human immunoglobulin Fab fragments." Hum. Antibod. Hybridodomas **5**(§): 57-64.

de Bruin, R., K. Spelt, et al. (1999). "Selection of high-affinity phage antibodies from phage display libraries." Nat Biotechnol **17**(4): 397-9.

de Haard, H. J., N. van Neer, et al. (1999). "A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies." J Biol Chem **274**(26): 18218-30.

de Kruif, J., E. Boel, et al. (1995). "Selection and application of human single chain Fv antibody fragments from a semi-synthetic phage antibody display library with designed CDR3 regions." J Mol Biol **248**(1): 97-105.

Deng, L. W., P. Malik, et al. (1999). "Interaction of the globular domains of pIII protein of filamentous bacteriophage fd with the F-pilus of Escherichia coli." Virology **253**(2): 271-7.

Dieterich, W., T. Ehnis, et al. (1997). "Identification of tissue transglutaminase as the autoantigen of celiac disease [see comments]." Nat Med **3**(7): 797-801.

Dunn, I. S. (1996). "Total modification of the bacteriophage lambda tail tube major subunit protein with foreign peptides." Gene **183**(1-2): 15-21.

Dziegiel, M., L. K. Nielsen, et al. (1995). "Phage display used for gene cloning of human recombinant antibody against the erythrocyte surface antigen, rhesus D." J Immunol Methods **182**(1): 7-19.

Endemann, H., V. Gailus, et al. (1993). "Interchangeability of the adsorption proteins of bacteriophages Ff and IKe." J Virol **67**(6): 3332-7.

Endemann, H. and P. Model (1995). "Location of filamentous phage minor coat proteins in phage and in infected cells." J Mol Biol **250**(4): 496-506.

Engels, W. R. (1993). "Contributing software to the Internet: the Amplify program." Trends in Biochem. Sci. **18**: 448-450.

Fan, H., C. Villegas, et al. (1998). "Myc-epitope tagged proteins detected with the 9E10 antibody in immunofluorescence and immunoprecipitation assays but not in western blot analysis." Biochem Cell Biol **76**(1): 125-8.

Felici, F., A. Luzzago, et al. (1993). "Mimicking of discontinuous epitopes by phage-displayed peptides, II. Selection of clones recognized by a protective monoclonal antibody against the Bordetella pertussis toxin from phage peptide libraries." Gene **128**(1): 21-7.

Fisch, I., R. E. Kontermann, et al. (1996). "A strategy of exon shuffling for making large peptide repertoires displayed on filamentous bacteriophage." Proc Natl Acad Sci U S A **93**(15): 7761-6.

Francisco, J. A., R. Campbell, et al. (1993). "Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface." Proc.Natl.Acad.Sci. **90**: 10444-10448.

Francisco, J. A., C. F. Earhart, et al. (1992). "Transport and anchoring of ß-lactamase to the external surfaceof Escherichia coli." Proc.Natl.Acad.Sci., U.S.A. **89**: 2713-2717.

Francisco, J. A., C. Stathopoulos, et al. (1993). "Specific Adhesion and Hydrolysis of Cellulose by Intact Escherichia coli Expressing Surface Anchored Cellulase or Cellulose Binding Domains." Bio/Technology **11**: 491-495.

Frippiat, J. P., S. C. Williams, et al. (1995). "Organization of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2." Hum Mol Genet **4**(6): 983-991.

Furth, M. E., L. J. Davis, et al. (1982). "Monoclonal antibodies to the p21 products of the transforming gene of Harvey murine sarcoma virus and of the cellular ras gene family." J Virol **43**(1): 294-304.

Gailus, V. and I. Rasched (1994). "The adsorption protein of bacteriophage fd and its neighbour minor coat protein build a structural entity." Eur J Biochem **222**(3): 927-31.

Geoffroy, F., R. Sodoyer, et al. (1994). "A new phage display system to construct multicombinatorial libraries of very large antibody repertoires." Gene **151**(1-2): 109-113.

Georgiou, G., H. L. Poetschke, et al. (1993). "Practical applications of engineering gram negative bacterial cell surfaces." TIBTECH **11**: 6-10.

Georgiou, G., C. Stathopoulos, et al. (1997). "Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines." Nat. Biotechnol **15**: 29-34.

Glaser-Wuttke, G., J. Keppner, et al. (1989). "Pore-forming properties of the adsorption protein of filamentous phage fd." Biochim. Biophys. Acta **985**: 239-247.

Goggins, M. and D. Kelleher (1994). "Celiac disease and other nutrient related injuries to the gastrointestinal tract." Am J Gastroenterol **89**(8 Suppl): S2-17.

Grant, R. A., T. C. Lin, et al. (1981). "Structure of filamentous bacteriophage: isolation, characterization, and localization of the minor coat proteins and orientation of the DNA." Prog Clin Biol Res **64**: 413-28.

Graus, Y. F., M. H. de Baets, et al. (1997). "Human anti-nicotinic acetylcholine receptor recombinant Fab fragments isolated from thymus-derived phage display libraries from myasthenia gravis patients reflect predominant specificities in serum and block the action of pathogenic serum antibodies." J Immunol **158**(4): 1919-29.

Graus, Y. F., M. H. de Baets, et al. (1997). "Anti-acetylcholine receptor Fab fragments isolated from thymus-derived phage display libraries from myasthenia gravis patients reflect predominant specificities in serum and block the action of pathogenic serum antibodies." Immunol Lett **57**(1-3): 59-62.

Graus, Y. F., J. J. Verschuuren, et al. (1998). "Selection of recombinant anti-HuD Fab fragments from a phage display antibody library of a lung cancer patient with paraneoplastic encephalomyelitis." J Neuroimmunol **82**(2): 200-9.

Griffiths, A. D., M. Malmqvist, et al. (1993). "Human anti-self antibodies with high specificity from phage display libraries." EMBO. J. **12**(2): 725-734.

Griffiths, A. D., S. C. Williams, et al. (1994). "Isolation of high affinity human antibodies directly from large synthetic repertoires." EMBO J. **13**(14): 3245-3260.

Guigou, V., A. M. Cuisinier, et al. (1990). "Human immunoglobulin VH and VK repertoire revealed by in situ hybridization." Mol Immunol **27**(9): 935-40.

Guihard, G., P. Boulanger, et al. (1994). "Colicin A and the Tol proteins involved in its translocation are preferentially located in the contact sites between the inner and outer membranes of Escherichia coli cells." J Biol Chem **269**(8): 5874-80.

Guo, F., D. N. Gopaul, et al. (1997). "Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse." Nature **389**(6646): 40-6.

Hanes, J., L. Jermutus, et al. (1998). "Ribosome display efficent selects and evolves high-affinity antibodies in vitro from immune libraries." Proc. Natl. Acad. Sci **95**: 14130-14135.

Hanes, J. and A. Plückthun (1997). "In vitro selection and evolution of functional proteins using ribosome display." Proc. Natl. Acad. Sci., U.S.A. **94**: 4937-4942.

Hanke, T., P. Szawlowski, et al. (1992). "Construction of solid matrix-antibody-antigen complexes containing simian immunodeficiency virus p27 using tag-specific monoclonal antibody and tag-linked antigen." J Gen Virol **73**(Pt 3): 653-60.

Hansson, M., S. Stahl, et al. (1992). "Expression of recombinant proteins on the surface of the coagulase- negative bacterium Staphylococcus xylosus." J Bacteriol **174**(13): 4239-45.

He, M. and M. J. Taussig (1997). "Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites." Nucleic Acids Res **25**: 5132-5134.

Hexham, J. M., L. J. Partridge, et al. (1994). "Cloning and characterisation of TPO autoantibodies using combinatorial phage display libraries." Autoimmunity **17**(3): 167-79.

Hirst, K., F. Fisher, et al. (1994). "The transcription factor, the Cdk, its cyclin and their regulator: directing the transcriptional response to a nutritional signal." Embo J **13**(22): 5410-20.

Hoess, R., A. Wierzbicki, et al. (1986). "The role of the loxP spacer region in P1 site-specific recombination." Nucleic Acids Res **14**: 2287-2300.

Hoess, R. H. and K. Abremski (1984). "Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP." Proc Natl Acad Sci U S A **81**(4): 1026-9.

Hoess, R. H., A. J. Mack, et al. (1994). "Identification of a structural epitope by using a peptide library displayed on filamentous bacteriophage." J Immunol **153**(2): 724-9.

Hofnung, M. (1991). "Expression of foreign polypeptides at the Escherichia coli cell surface." Methods Cell Biol. **34**: 77-105.

Holliger, P., T. Prospero, et al. (1993). ""Diabodies": small bivalent and bispecific antibody fragments." Proc Natl Acad Sci U S A **90**(14): 6444-8.

Hollinger, P. and L. Riechmann (1997). "A conserved infection patway for filamentous bacteriophages is suggested by the structure of the membrane penetration domain of the minor coat protein g3p from phage fd." Structure **5**: 265-275.

Hoogenboom, H. R., A. D. Griffiths, et al. (1991). "Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains." Nucl. Acids Res. **19**: 4133-4137.

Hoogenboom, H. R. and G. Winter (1992). "By-passing immunisation. Human antibodies from synthetic repertoires of germline VH gene segments rearranged in vitro." J. Mol. Biol **227**: 381-388.

Hoogenboom, H. R. and G. Winter (1992). "Bypassing immunisation: human antibodies from synthetic repertoires of germ line VH-gene segments rearranged *in -vitro*." J. Mol. Biol **227**: 381-388.

Iannolo, G., O. Minenkova, et al. (1995). "Modifying filamentous phage capsid: limits in the size of the major capsid protein." J Mol Biol **248**(4): 835-844.

Ignatovich, O., I. M. Tomlinson, et al. (1997). "The creation of diversity in the human immunoglobulin V(lambda) repertoire." J Mol Biol **268**(1): 69-77.

Irvine, D., C. Tuerk, et al. (1991). "Selexion: systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis." J. Mol. Biol. **222**: 739-761.

Jellis, C. L., T. J. Cradick, et al. (1993). "Defining critical residues in the epitope for a HIV-neutralizing monoclonal antibody using phage display and peptide array technologies." Gene **137**(1): 63-8.

Jespers, L., S. Jenne, et al. (1997). "Epitope mapping by negative selection of randomized antigen libraries displayed on filamentous phage." J Mol Biol **269**(5): 704-18.

Jiang, J., L. Abu-Shilbayeh, et al. (1997). "Display of a PorA peptide from Neisseria meningitidis on the bacteriophage T4 capsid surface." Infect Immun **65**(11): 4770-7.

Johansen, L. K., B. Albrechtsen, et al. (1995). "pFab60: a new, efficient vector for expression of antibody Fab fragments displayed on phage." Protein Eng **8**(10): 1063-7.

Johansson, M., I. Nilsson, et al. (1993). "Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in Escherichia coli than in mammalian microsomes." Mol Gen Genet **239**(1-2): 251-6.

Kabat, E. A., T. T. Wu, et al. (1987). Sequences of proteins of immunological interest, US Department of Health and Human Services, US Government Printing Office.

Khatoon, H., R. V. Iyer, et al. (1972). "A new filamentous bacteriophage with sex-factor specificity." Virology **48**: 145-155.

Kilby, N. J., M. R. Snaith, et al. (1993). "Site-specific recombinases: tools for genome engineering." Trends Genet **9**(12): 413-21.

Knappik, A. and A. Plückthun (1994). "An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody fragments." Biotechniques **17**: 754-761.

Krebber, A., S. Bornhauser, et al. (1997). "Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system." J. Immunol. Methods. **201**: 35-55.

Kretzschmar, T. and M. Geiser (1995). "Evaluation of antibodies fused to minor coat protein III and major coat protein VIII of bacteriophage M13." Gene **155**(1): 61-65.

Kuwabara, I., H. Maruyama, et al. (1997). "Efficient epitope mapping by bacteriophage lambda surface display." Nat Biotechnol **15**(1): 74-8.

Larrick, J. W., L. Danielsson, et al. (1989). "Polymerase chain reaction using mixed primers: Cloning of human monoclonal antibody variable region genes from single hybridoma cells." Bio/Technology **7**(9): 934-938.

Lei, S.-P., H.-C. Lin, et al. (1987). "Characterization of the Erwinia carotovora pelB gene and its product pectase lyase." J. Bacteriol. **169**: 4379.

Li, P., J. Beckwith, et al. (1988). "Alteration of the amino terminus of the mature sequence of a periplasmic protein can severely affect protein export in Escherichia coli." Proc Natl Acad Sci U S A **85**(20): 7685-9.

Livnah, O., E. Stura, et al. (1996). "Functional mimicry of a protein hormone by a peptide agonist: The EPO receptor complex at 2.8 Å." Science **273**: 464-471.

Lubkowski, J., F. Hennecke, et al. (1998). "The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p." Nat Struct Biol **5**(2): 140-7.

Lundin, K. E., H. A. Gjertsen, et al. (1994). "Function of DQ2 and DQ8 as HLA susceptibility molecules in celiac disease." Hum Immunol **41**(1): 24-7.

Lundin, K. E., H. Scott, et al. (1994). "T cells from the small intestinal mucosa of a DR4, DQ7/DR4, DQ8 celiac disease patient preferentially recognize gliadin when presented by DQ8." Hum Immunol **41**(4): 285-91.

Luo, D., N. Mah, et al. (1995). "Vl-linker-Vh orientation-dependent expression of single chain Fv- containing an engineered disulfide-stabilized bond in the framework regions." J Biochem (Tokyo) **118**(4): 825-31.

Luzzago, A., F. Felici, et al. (1993). "Mimicking of discontinuous epitopes by phage-displayed peptides, I. Epitope mapping of human H ferritin using a phage library of constrained peptides." Gene **128**: 51-56.

Makeyev, E. V., V. A. Kolb, et al. (1999). "Cell-free immunology: construction and in vitro expression of a PCR- based library encoding a single-chain antibody repertoire." FEBS Lett **444**(2-3): 177-80.

Malik, P., T. D. Terry, et al. (1998). "Factors limiting display of foreign peptides on the major coat protein of filamentous bacteriophage capsids and a potential role for leader peptidase." FEBS Lett **436**(2): 263-6.

Malik, P., T. D. Terry, et al. (1996). "Role of capsid structure and membrane protein processing in determining the size and copy number of peptides displayed on the major coat protein of filamentous bacteriophage." J Mol Biol **260**(1): 9-21.

Marchbank, M. T. and S. L. Deutscher (1995). "Autoimmune derived combinatorial phage display libraries: methods in construction of and affinity selection for anti-RNA Fabs." Nucleic Acids Symp Ser(33): 120-2.

Mariuzza, R. A., D. L. Jankovic, et al. (1983). "Preliminary crystallographic study of the complex between the Fab fragment of a monoclonal anti-lysozyme antibody and its antigen." J Mol Biol **170**(4): 1055-8.

Marks, J. D., A. D. Griffiths, et al. (1992). "By-passing immunization: building high affinity human antibodies by chain shuffling." BioTechnology **10**: 779-783.

Marks, J. D., H. R. Hoogenboom, et al. (1991). "By-passing immunization- human antibodies from V-gene libraries displayed on phage." J. Mol. Biol. **222**: 581-597.

Marks, J. D., M. Tristem, et al. (1991). "Oligonucleotide primers for polymerase chain reaction amplification of human immunoglobulin variable genes and design of family-specific oligonucleotide probes." Eur J Immunol **21**(4): 985-91.

Marullo, S., C. Delavier-Klutchko, et al. (1989). "Expression of human ß1 and ß2 adrenergic receptors in E. coli as a new tool for ligand screening." BioTech. **7**: 923-927.

Maruyama, I. N., H. I. Maruyama, et al. (1994). "lambda-foo: a lambda vector for the expression of foreign proteins." Proc. Natl. Acad. Sci. USA **91**: 8273-8277.

Marvin, D. A., R. D. Hale, et al. (1994). "Molecular models and structural comparisons of native and mutant class I filamentous bacteriophages Ff (fd, f1, M13), If1 and IKe." J Mol Biol **235**(1): 260-86.

Marvin, D. A. and B. Hohn (1969). "Filamentous bacterial viruses." Bacteriol Rev **33**(2): 172-209.

Matthews, D. J., L. J. Goodman, et al. (1994). "A survey of furin substrate specificity using substrate phage display." Protein Sci **3**(8): 1197-1205.

Matthews, D. J. and J. A. Wells (1993). "Substrate phage: selection of protease substrates by monovalent phage display." Science **260**: 1113-1117.

Mattila, P. S., J. Schugk, et al. (1995). "Extensive allelic sequence variation in the J region of the human immunoglobulin heavy chain gene locus." Eur. J. Immunol. **25**: 2578-2582.

Max, E. E., J. G. Seidman, et al. (1980). "Variation in the crossover point of kappa immunoglobulin gene V-J recombination: evidence from a cryptic gene." Cell **21**(3): 793-9.

Mayer, M. P. (1995). "A new set of useful cloning and expression vectors derived from pBlueScript." Gene **163**(1): 41-6.

McCafferty, J., A. D. Griffiths, et al. (1990). "Phage antibodies: filamentous phage displaying antibody variable domains." Nature **348**(6301): 552-4.

McGuinness, B. T., G. Walter, et al. (1996). "Phage diabody repertoires for selection of large numbers of bispecific antibody fragments." Nat. Biotech. **14**: 1149-1154.

McIntosh, R. S., M. S. Asghar, et al. (1997). "Analysis of immunoglobulin G kappa antithyroid peroxidase antibodies from different tissues in Hashimoto's thyroiditis." J Clin Endocrinol Metab **82**(11): 3818-25.

McIntosh, R. S., M. S. Asghar, et al. (1996). "Cloning and analysis of IgG kappa and IgG lambda anti-thyroglobulin autoantibodies from a patient with Hashimoto's thyroiditis: evidence for in vivo antigen-driven repertoire selection." J Immunol **157**(2): 927-35.

McIntosh, R. S., M. S. Asghar, et al. (1997). "The antibody response in human autoimmune thyroid disease [editorial]." Clin Sci (Colch) **92**(6): 529-41.

Mead, D. A. and B. Kemper (1988). "Chimeric single-stranded DNA phage-plasmid cloning vectors." Biotechnology **10**: 85-102.

Medaglini, D., G. Pozzi, et al. (1995). "Mucosal and systemic immune responses to a recombinant protein expressed on the surface of the oral commensal bacterium Streptococcus gordonii after oral colonization." Proc Natl Acad Sci U S A **92**(15): 6868-6872.

Meola, A., P. Delmastro, et al. (1995). "Derivation of vaccines from mimotopes. Immunologic properties of human hepatitis B virus surface antigen mimotopes displayed on filamentous phage." J Immunol **154**(7): 3162-3172.

Mikawa, Y. G., I. N. Maruyama, et al. (1996). "Surface display of proteins on bacteriophage lambda heads." J Mol Biol **262**(1): 21-30.

Miller, G. and M. Feiss (1988). "The bacteriophage lambda cohesive end site: isolation of spacing/substitution mutations that result in dependence on Escherichia coli integration host factor." Mol Gen Genet **212**(1): 157-65.

Milstein, C. and M. S. Neuberger (1996). "Maturation of the immune response." Adv Protein Chem **49**: 451-85.

Model, P. and M. Russel (1988). Filamentous bacteriophage. The bacteriophages. R. Calendar. New York, Plenum Press. **2:** 375-456.

Molberg, O., S. N. McAdam, et al. (1998). "Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease [see comments] [published erratum appears in Nat Med 1998 Aug;4(8):974]." Nat Med **4**(6): 713-7.

Morea, V., A. Tramontano, et al. (1998). "Conformations of the third hypervariable region in the VH domain of immunoglobulins." J Mol Biol **275**(2): 269-94.

Mortari, F., W. J.Y., et al. (1993). "Human cord blood antibody repertoire. Mixed population of VH gene segments and CDR3 distribution in the expressed Cα and Cγ repertoires." J. Immunol. **150**(1348-1357).

Mortari, F., J. A. Newton, et al. (1992). "The human cord blood antibody repertoire. Frequent usage of the VH7 gene family." Eur. J. Immunol. **22**: 241.

Nemoto, N., E. Miyamoto-Sato, et al. (1997). "In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro." FEBS Lett **414**(2): 405-8.

Nissim, A., H. R. Hoogenboom, et al. (1994). "Antibody fragments from a 'single pot' phage display library as immunochemical reagents." Embo J. **13**: 692-698.

Not, T., D. Sblattero, et al. (1999). "Rapid and simple dot immunobinding assay to detect anti human transglutaminase antibodies in coeliac disease." J. Pediatr. Gastroenterol. Nutr. **26**(5): 563.

Orlandi, R., D. H. Gussow, et al. (1989). "Cloning immunoglobulin variable domains for expression by the polymerase chain reaction." Proc. Natl. Acad. Sci. U.S.A **86**(10): 3833-3837.

Overman, S. A. and G. J. Thomas, Jr. (1995). "Raman spectroscopy of the filamentous virus Ff (fd, fl, M13): structural interpretation for coat protein aromatics [published erratum appears in Biochemistry 1995 Jun 13;34(23):7780]." Biochemistry **34**(16): 5440-51.

Parmley, S. F. and G. P. Smith (1988). "Antibody-selectable filamentous fd phage vectors: affinity purification of target genes." Gene **73**(2): 305-18.

Parmley, S. F. and G. P. Smith (1989). "Filamentous fusion phage cloning vectors for the study of epitopes and design of vaccines." Adv. Exp. Med. Biol. **251**(215): 215-8.

Peeters, B. P. H., R. M. Peters, et al. (1985). "Nucleotide sequence and genetic organization of the genome of the N-specific filamentous bacteriophage IKe. Comparison with the genome of the F-specific filamentous phages M13, fd and f1." J. Mol. Biol. **181**: 27-39.

Perelson, A. S. and G. F. Oster (1979). "Theoretical studies of clonal selection: Minimial antibody repertoire size and reliability of self non-self discrimination." J. Theor. Biol. **81**: 645-670.

Persic, L., M. Righi, et al. (1997). "Targeting vectors for intracellular immunisation." Gene **187**: 1-8.

Persic, L., A. Roberts, et al. (1997). "An integrated vector system for the eukaryotic expression of antibodies or their fragments after selection from phage display libraries." Gene **187**: 9-18.

Picarelli, A., L. Maiuri, et al. (1996). "Production of antiendomysial antibodies after in-vitro gliadin challenge of small intestine biopsy samples from patients with coeliac disease [see comments]." Lancet **348**(9034): 1065-7.

Portolano, S., S. M. McLachlan, et al. (1993). "High affinity, thyroid-specific human autoantibodies displayed on the surface of filamentous phage use V genes similar to other autoantibodies." J Immunol **151**(5): 2839-51.

Portolano, S., M. F. Prummel, et al. (1995). "Molecular cloning and characterization of human thyroid peroxidase autoantibodies of lambda light chain type." Mol Immunol **32**(14-15): 1157-69.

Ren, Z. J., G. K. Lewis, et al. (1996). "Phage display of intact domains at high copy number: a system based on SOC, the small outer capsid protein of bacteriophage T4." Protein Sci **5**(9): 1833-43.

Riechmann, L. and P. Holliger (1997). "The C-terminal domain of TolA is the coreceptor for filamentous phage infection of E. coli." Cell **90**: 351-360.

Roben, P., S. M. Barbas, et al. (1996). "Repertoire cloning of lupus anti-DNA autoantibodies." J Clin Invest **98**(12): 2827-37.

Roberts, B. L., W. Markland, et al. (1992). "Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage." Proc Natl Acad Sci U S A **89**(6): 2429-33.

Russel, M. (1991). "Filamentous phage assembly." Mol. Microbiol. **5**: 1607-1613.

Russel, M. (1994). "Phage assembly: a paradigm for bacterial virulence factor export?" Science **265**(5172): 612-4.

Saggio, I. and R. Laufer (1993). "Biotin binders selected from a random peptide library expressed on phage [published erratum appears in Biochem J 1993 Nov 1;295(Pt 3):903]." Biochem J **293**(Pt 3): 613-6.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular cloning: a laboratory manual, Cold Spring Harbour Laboratory Press.

Samuelson, P., M. Hansson, et al. (1995). "Cell surface display of recombinant proteins on Staphylococcus carnosus." J Bacteriol **177**(6): 1470-1476.

Santini, C., D. Brennan, et al. (1998). "Efficient display of an HCV cDNA expression library as C-terminal fusion to the capsid protein D of bacteriophage lambda." J Mol Biol **282**(1): 125-35.

Sanz, I. (1991). "Multiple mechanisms participate in the generation of diversity of human H chain CDR3 regions." J. Immunol. **147**(5): 1720-1729.

Sauer, B. (1992). "Identification of cryptic lox sites in the yeast genome by selection for Cre-mediated chromosome translocations that confer multiple drug resistance." J Mol Biol **223**(4): 911-28.

Sauer, B. and N. Henderson (1988). "The cyclization of linear DNA in Escherichia coli by site-specific recombination." Gene **70**: 331-341.

Sblattero, D., I. Berti, et al. (1999). "Human tissue transglutaminase ELISA: a powerfull mass screening diagnostic assay for Coeliac Disease." J. Pediatr. Gastroenterol. Nutr. **28**(5): 568.

Sblattero, D. and A. Bradbury (1998). "A definitive set of oligonucleotide primers for amplifying human V regions." Immunotechnology **3**: 271-8.

Sblattero, D., T. Not, et al. (1999). "Phage antibody library from celiac disease patient." J. Pediatr. Gastroenterol. Nutr. **26**(5): 568.

Schatz, D. G., M. A. Oettinger, et al. (1989). "The V(D)J recombination activating gene, RAG-1." Cell **59**(6): 1035-48.

Schier, R. and J. D. Marks (1996). "Efficient in vitro affinity maturation of phage antibodies using BIAcore guided selections." Hum Antibodies Hybridomas **7**(3): 97-105.

Schroeder, H. J., J. L. Hillson, et al. (1987). "Early restriction of the human antibody repertoire." Science **238**(4828): 791-3.

Seidman, J. G., E. E. Max, et al. (1979). "A kappa-immunoglobulin gene is formed by site-specific recombination without further somatic mutation." Nature **280**(5721): 370-5.

Sheets, M. D., P. Amersdorfer, et al. (1998). "Efficient construction of a large nonimmune phage antibody library; the production of panels of high affinity human single-chain antibodies to protein antigens." Proc Natl Acad Sci USA **95**: 6157-6162.

Skerra, A. and A. Pluckthun (1988). "Assembly of a functional immunoglobulin Fv fragment in Escherichia coli." Science **240**(4855): 1038-41.

Smith, G. P. (1985). "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface." Science **228**: 1315-1317.

Smith, G. P. (1988). "Filamentous phages as cloning vectors." Biotechnology **10**: 61-83.

Smith, G. P., D. A. Schultz, et al. (1993). "A ribonuclease S-peptide antagonist discovered with a bacteriophage display library." Gene **128**(1): 37-42.

Smith, G. P. and J. K. Scott (1993). "Libraries of peptides and proteins displayed on filamentous phage." Methods Enzymol **217**(57): 228-57.

Sparks, A. B., L. A. Quilliam, et al. (1994). "Identification and characterization of Src SH3 ligands from phage- displayed random peptide libraries." J Biol Chem **269**(39): 23853-6.

Stark, W. M., M. R. Boocock, et al. (1992). "Catalysis by site-specific recombinases [published erratum appears in Trends Genet 1993 Feb;9(2):45]." Trends Genet **8**(12): 432-9.

Stephen, C. W., P. Helminen, et al. (1995). "Characterisation of epitopes on human p53 using phage-displayed peptide libraries: insights into antibody-peptide interactions." J Mol Biol **248**(1): 58-78.

Stephen, C. W. and D. P. Lane (1992). "Mutant conformations of p53. Precise epitope mapping using a filamentous phage epitope library." J. Mol. Biol. **225**: 577-583.

Sternberg, N. (1981). "Bacteriophage P1 site-specific recombination. III. Strand exchange during recombination at lox sites." J Mol Biol **150**(4): 603-8.

Sternberg, N. and R. H. Hoess (1995). "Display of peptides and proteins on the surface of bacteriophage lambda." Proc Natl Acad Sci U S A **92**(5): 1609-1613.

Stewart, A. K., C. Huang, et al. (1993). "High frequency representation of a single VH gene in the expressed human B cell repertoire." J. Exp. Med. **177**: 409.

Stewart, A. K. and R. S. Schwartz (1994). "Immunoglobulin V regions and the B cell." Blood **83**(7): 1717-30.

Suzuki, H., H. Takemura, et al. (1997). "Molecular cloning of anti-SS-A/Ro 60-kDa peptide Fab fragments from infiltrating salivary gland lymphocytes of a patient with Sjogren's syndrome." Biochem Biophys Res Commun **232**(1): 101-6.

Suzuki, I., L. Pfister, et al. (1995). "Representation of rearranged VH gene segments in the human adult antibody repertoire." J Immunol **154**(8): 3902-11.

Tesar, M., C. Beckmann, et al. (1995). "Monoclonal antibody against pIII of filamentous phage: an immunological tool to study pIII fusion protein expression in phage display systems." Immunotechnology **1**(1): 53-64.

Tomlinson, I. M., G. P. Cook, et al. (1995). "A complete map of the human immunoglobulin VH locus." Ann N Y Acad Sci **764**: 43-6.

Tomlinson, I. M., J. P. Cox, et al. (1995). "The structural repertoire of the human V kappa domain." Embo J **14**(18): 4628-38.

Tomlinson, I. M., G. Walter, et al. (1996). "The imprint of somatic hypermutation on the repertoire of human germline V genes." J Mol Biol **256**(5): 813-17.

Tomlinson, I. M., G. Walter, et al. (1992). "The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops." J. Mol. Biol. **227**: 776-798.

Tomlinson, I. M., S. C. Williams, et al. (1996). "V BASE sequence directory." MRC Centre for Protein Engineering, Cambridge, UK.

Tsurushita, N., H. Fu, et al. (1996). "Phage display vectors for in vivo recombination if immunoglobulin heavy and light chain genes to make large combinatorial libraries." Gene **172**: 59-63.

Van Es, J. H., F. M. Raaphorst, et al. (1993). "Expression pattern of the most JH-proximal human VH gene segment (VH6) in the B cell and antibody repertoire suggests a role of VH6-encoded IgM antibodies in early ontogeny." J Immunol **150**(1): 161-8.

Vaughan, T. J., A. J. Williams, et al. (1996). "Human antibodies with sub-nanomolar affinities isolated from a large non-immunised phage display library." Nat. Biotechnol. **14**: 309-314.

von Heijne, G. (1986). "A new method for predicting signal sequence cleavage sites." Nuc. Acids Res. **14**: 4683-4690.

Wang, C. I., Q. Yang, et al. (1995). "Isolation of a high affinity inhibitor of urokinase-type plasminogen activator by phage display of ecotin." J Biol Chem **270**(20): 12250-12256.

Waterhouse, P., A. D. Griffiths, et al. (1993). "Combinatorial infection and in vivo recombination: a strategy for making large phage antibody repertoires." Nucleic Acids Res **21**(9): 2265-6.

Watkins, B. A., A. E. Davis, et al. (1995). "V-region and class specific RT-PCR amplification of human immunoglobulin heavy and light chain genes from B-cell lines." Scand J Immunol **42**(4): 442-448.

Welschof, M., P. Terness, et al. (1995). "Amino acid sequence based PCR primers for amplification of rearranged human heavy and light chain immunoglobulin variable region genes." J Immunol Methods **179**(2): 203-214.

Werge, T. M., A. Bradbury, et al. (1992). "A recombinant cell line, expressing a form of the anti-p21ras antibody, Y13-259, which binds protein A and may be produced as ascites." Oncogene **7**: 1033-1035.

Whitlow, M., B. A. Bell, et al. (1993). "An improved linker for single-chain Fv with reduced aggregation and enhanced proteolytic stability." Protein Eng **6**(8): 989-95.

Williams, K. A., M. Glibowicka, et al. (1995). "Packing of coat protein amphipathic and transmembrane helices in filamentous bacteriophage M13: role of small residues in protein oligomerization." J Mol Biol **252**(1): 6-14.

Williams, S. C., J.-P. Frippiat, et al. (1996). "Sequence and evolution of the human germiline Vλ repertoire." J. Mol. Biol. **264**: 220-232.

Williams, S. C. and G. Winter (1993). "Cloning and sequencing of human immunoglobulin V lambda gene segments." Eur J Immunol **23**(7): 1456-61.

Winter, G., A. D. Griffiths, et al. (1994). "Making antibodies by phage display technology." Annu Rev Immunol **12**: 433-55.

Yamada, M., R. Wasserman, et al. (1991). "Preferential utilization of specific immunoglobulin heavy chain diversity and joining segments in adult human peripheral blood

B lymphocytes." J Exp Med **173**(2): 395-407.

Zacher, A. N., C. A. Stock, et al. (1980). "A new filamentous phage cloning vector: fd-tet." Gene **9**: 127-140.