Title: MOLECULAR BIOLOGY METHOD

Fig. 2) Construction of a small VH Library. An assembly reaction experiment was done, using CDR3H single-stranded fragments from five antibody scFv clones with different sequences and lengths (6-22 aa) as a pool, while keeping the CDR1H and CDR2H single-stranded fragments constant. The template fragment used, had a 10 as CDR3H.

Abstract: The invention relates to a method for assembling a composite polynucleotide sequence and for producing a library of different composite polynucleotide sequences. In particular, the invention relates to a method for assembling polynucleotide sequences encoding immunoglobulin variable regions and for identifying immunoglobulin variable regions with desired characteristics.
Published:
— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
MOLECULAR BIOLOGY METHOD

The invention relates to a method for assembling a composite polynucleotide sequence and for producing a library of different composite polynucleotide sequences. In particular, the invention relates to a method for assembling polynucleotide sequences encoding immunoglobulin variant regions and for identifying immunoglobulin variant regions with desired characteristics.

Binding surfaces for ligands on proteins are often composed of two or more polypeptide segments. This is, for example, true for the antigen-binding surfaces on immunoglobulin molecules such as antibodies, B cell receptors and T cell receptors. How these different polypeptide segments are assembled provides the basis for a large variability in the binding properties of such proteins. This needs to be mimicked in biotech applications. There is, for example, a need to generate specific antibodies for analytical and therapeutic purposes. To generate antibodies optimal for such applications, ideally all natural sequences involved in forming the binding surface, in all combinations, should be investigated in order to find an antibody with the desired maximised properties. Theoretically, this can be achieved by combining fragments of antibody variable region genes in different ways, which then will be expressed to produce an array of proteins with different binding properties.

It is well known that the generation of libraries, such as antibody fragment libraries, with existing methods results in a biased representation of variabilities, with only a subset of naturally-occurring sequences being represented in the resulting library.

Many techniques are known in the art that can be used to recombine DNA fragments in order to assemble genes. The usefulness of these techniques
for the assembly of genes encoding antibodies and other proteins with alternating regions of homology and variability have limitations, however. Some of these techniques are based on DNA-shuffling (for example, US 6,132,970), where a typical reaction involves fragmentation of a gene segment to oligonucleotide sequences by randomly fragmenting the DNA using DNase I, polymerase chain reaction (PCR) extension is initiated using a polymerase, and the complete gene is assembled by random recombination and amplified. The DNA-shuffling approach relies on random recombination by polymerisation and is, therefore, not very controllable. Also, the frequency of errors made by the polymerase is quite high, making unwanted mutations a reality.

Methods for the generation of synthetic antibody gene libraries by which spiked oligonucleotides are assembled by a ligase chain reaction protocol (Deng S-J et al. (1993); Deng et al. (1995) Methods in Mol Biol, 51, 329-42) for in vitro affinity maturation purposes have been devised. Briefly, several oligonucleotides covering the antibody variable domain are synthesized. The sense oligonucleotides have partly randomized complementarity-determining region (CDR) sequences due to ‘spiking’ at their synthesis and the antisense oligonucleotides have the sequence positions corresponding to the randomized position on the sense strand filled with inosine, a universal base-pairing base, to facilitate annealing. The oligonucleotides are assembled on a template consisting of double-stranded vector DNA encoding the antibody variable domain in a ligase chain reaction, and the complete gene is then purified and cloned. The reaction may also be performed without template DNA, in that case overlapping oligonucleotides must be used to form the DNA duplexes. This method will be limited to the use of synthetically mutated CDRs, where only a limited amount of the residues will be mutated and most of the clones will be wild type. Also, there is no consideration of the length variations
occurring among natural antibody CDR sequences, which means that parts of the diversity in natural sequences will be lost.

L-shuffling is another technique for recombination of polynucleotide sequences using a template for assembly of gene fragment variants and ligation by a suitable DNA ligase of the ends of the fragments (WO 02/086121 A1). After the hybridization to template and ligation, the reaction is denatured again, and ligated and non-ligated fragments are allowed to hybridize again. The cycles of hybridization — ligation — denaturation are repeated until the full length gene is assembled. The template is then removed and the gene variants amplified by PCR and cloned to create a library. Neither data on length variability, nor on whether the method can handle looping out of the DNA, has been reported for the L-shuffling technique.

Thus there is a need for alternative approaches to assemble polynucleotide sequences encoding antibodies or other proteins that fully utilise the information present in natural gene sequences.

The present invention relates to a method for directing and controlling the assembly of gene fragments which provides an alternative protocol with potential to generate new and/or different combinations of gene fragments, and/or an improved representation of natural sequence segments. This allows for a more controlled selection and assembly of gene segments, which has utility, for example, in the assembly of immunoglobulin variable region sequences. In addition, the invention permits looping out of single-stranded polynucleotide sequences allowing gene segments of different length to be assembled into a composite polynucleotide.

The invention also provides a method useful for library construction, both with respect to primary libraries composed of natural immunoglobulin gene
segments and secondary libraries of, for example, mutated sequences starting from a template of a selected immunoglobulin gene with interesting albeit not optimal binding properties for a given ligand. In general, the method provides a means to construct libraries of any sequence but is particularly useful for the assembly of sequences that encode a protein comprising a ligand-binding site composed of more than one linear sequence segments.

In a first aspect, the invention provides a method for assembling a composite polynucleotide sequence comprising the steps:

a) providing a single-stranded template polynucleotide sequence comprising at least one invariant region and at least one variant region;

b) providing a plurality of single-stranded invariant polynucleotide sequences that are at least partially complementary to the at least one invariant region of the template polynucleotide sequence, and a plurality of single-stranded variant polynucleotide sequences that are at least partially complementary to the at least one variant region of the template polynucleotide sequence,

wherein the single-stranded invariant and variant polynucleotides, when ligated together, are capable of forming a polynucleotide which is at least partially complementary to the template polynucleotide;

c) annealing the single-stranded polynucleotide sequences provided in steps (a) and (b) to generate a double-stranded polynucleotide sequence wherein one strand is the template polynucleotide sequence, and one strand is composed of at least one invariant
polynucleotide sequence and at least one variant polynucleotide sequence that are not ligated together;

d) subjecting the double-stranded polynucleotide sequence generated in step (c) to conditions that ligate the at least one invariant polynucleotide sequence and the at least one variant polynucleotide sequence together to generate a double-stranded hybrid polynucleotide sequence composed of one strand that is the template polynucleotide sequence, and one strand that is a composite polynucleotide sequence comprising at least one invariant polynucleotide sequence and at least one variant polynucleotide sequence;

e) separating the single-stranded composite polynucleotide sequence generated in step (d) from the template polynucleotide sequence

Thus, in a first aspect the invention provides a method for assembling a composite polynucleotide sequence in a controlled and ordered manner. By providing a single-stranded template polynucleotide sequence, single-stranded gene segments can be assembled to generate a composite sequence which has, for example, a similar number and order of gene segments as the template sequence. The single-stranded polynucleotide sequences effectively act as "building blocks" which can be aligned by being annealed to the template polynucleotide sequence and subsequently joined to one another to create a composite polynucleotide sequence. The invention also permits looping out of single-stranded polynucleotide sequences allowing gene segments of different length to be assembled into a composite polynucleotide.

It will be understood that the composite molecule generated by the method of the invention may be single-stranded or partially double-stranded.
By “polynucleotide sequence” we include molecules of DNA (deoxyribonucleic acid) and/or RNA (ribonucleic acid) and derivatives thereof. Deoxyribonucleic acid and ribonucleic acid are known to exist as single-stranded or double-stranded molecules.

By “composite polynucleotide sequence” we include a polynucleotide sequence that has been assembled by the method of the invention by joining polynucleotide sequences together.

By “ligated together” we include the process whereby the free ends of annealed single-stranded polynucleotide sequences are enzymatically joined via a phosphodiester bond to form a single polynucleotide sequence composed of the joined sequences.

By “at least partially complementary” we include polynucleotide sequences that are capable of hybridising under low, medium or high stringency conditions to the template polynucleotide sequence.

It will be understood that the single-stranded invariant polynucleotide sequences need not be completely complementary to the at least one invariant region of the template polynucleotide sequence, and that the single-stranded variant polynucleotide sequences need not be completely complementary to the at least one variant region of the template polynucleotide sequence. These single-stranded sequences may be partially complementary with the appropriate regions of template polynucleotide sequence provided that they are sufficiently complementary to anneal to the template and become ligated to adjacent annealed polynucleotide sequences.
It will be understood that in order to anneal to the template polynucleotide sequence and become ligated to adjacent polynucleotide sequences, it is necessary for the 5' and 3' termini of the single-stranded invariant polynucleotide sequences and single-stranded variant polynucleotide sequences to anneal to the template polynucleotide such that adjacent polynucleotides can be ligated to each other. There is no requirement for the sequence of the single-stranded variant or invariant polynucleotide sequence between each of the termini to anneal to the template polynucleotide sequence so these regions may be more variable than the termini.

Preferably, approximately 15 to 20 bases at each of the 5' and 3' termini of the single-stranded variant polynucleotide sequence and the single-stranded invariant polynucleotide sequence are required to anneal to the template polynucleotide sequence. However, it will be appreciated that a greater or lesser number of bases may be required to anneal to the template polynucleotide sequence depending on the stringency of the hybridisation conditions and the nature of the sequences to be annealed.

It will also be appreciated that the termini of the single-stranded variant polynucleotide sequence and/or the single-stranded invariant polynucleotide sequence need not be fully complementary with the template polynucleotide sequence. Mismatches between bases at each of the termini of single-stranded variant polynucleotide sequence and the bases of the template sequence and/or between each of the termini of single-stranded invariant polynucleotide sequence and the bases of the template sequence may be present provided that the termini are sufficiently complementary to anneal to the template and become ligated to adjacent annealed polynucleotide sequences. Preferably, each of the termini of the single-stranded variant polynucleotide sequence and/or single-stranded invariant polynucleotide sequence will have a mismatch of 10% or less with the template
polynucleotide sequence. However, it will be appreciated that more or less mismatches may be required or sufficient to anneal to the template polynucleotide sequence depending on the stringency of the hybridisation conditions and the nature of the sequences to be annealed.

The single-stranded invariant polynucleotide sequences and/or single-stranded variant polynucleotide sequences may be longer or shorter than the region of the template polynucleotide sequence to which they anneal. If the single-stranded invariant polynucleotide sequences and/or single-stranded variant polynucleotide sequences are longer than the region of the template sequence to which they anneal, regions of the sequence may “loop out” so that the sequence is accommodated on the template polynucleotide sequence. If the single-stranded variant polynucleotide sequences and/or single-stranded invariant polynucleotide sequences are shorter than the region of the template sequence to which they anneal, regions of the template polynucleotide sequence may “loop out” to accommodate the variant and/or invariant polynucleotide sequence (see for example, Figure 1A).

It will be understood that it is therefore not necessary for every base of the template polynucleotide sequence to be annealed to a base of a variant polynucleotide sequence and/or an invariant polynucleotide sequence, or for every part of the template polynucleotide sequence to be completely “covered” by variant polynucleotide sequence and/or an invariant polynucleotide sequence.

“Stringency” of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon the length of the probe (i.e. sequence to be hybridised to a template sequence), washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter
probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to re-anneal when complementary strands are present in an environment below their melting temperatures. The higher the degree of desired homology between the probe and hybridisable sequence, the higher the relative temperature that can be used. For additional details and explanation of stringency of hybridization reactions, see Ausubel et al. (1995) or Protocols Online URL.

DNA-DNA, DNA-RNA and RNA-RNA hybridisation may be performed in aqueous solution containing between 0.1XSSC and 6XSSC and at temperatures of between 55°C and 70°C. It is well known in the art that the higher the temperature, or the lower the SSC (1XSSC is 0.15M NaCl/0.015M sodium citrate) concentration, the more stringent the hybridisation conditions. In addition, compounds such as dimethylsulfoxide (DMSO), formamide, glycerol and ammonium sulphate, which are capable of interacting with the DNA and/or RNA, may be added to the hybridisation reaction and will affect the conditions required to achieve hybridisation.

"Stringent conditions" or high-stringency may be identified by those that use low ionic strength and high temperature for washing, for example 0.1X SSC, 0.2% SDS @ 65-70°C. By "high stringency" we mean 2XSSC and 65°C.

"Moderately-stringent conditions" may be identified as described by Sambrook et al. (2001), and include the use of washing solution and hybridization conditions (e.g. temperature, ionic strength, and % SDS) less stringent than those described above. An example of moderately stringent conditions is 0.2X SSC, 0.1% SDS @ 58-65°C. The skilled artisan will recognize how to adjust temperature, ionic strength, etc. as necessary to accommodate factors such as probe length, degree of homology between probe and target site and the like. Therefore, in addition to the sequence of
interest, it is contemplated that additional or alternative probe sequences which vary from that of the sequence of interest will also be useful in screening for the sequence of interest.

By “annealing” we include the process of single-stranded polynucleotide sequences associating with one another via hydrogen bonding between specific complementary bases in the single-stranded sequences. For example, in DNA the bases cytosine and guanine bind with one another, and the bases adenine and thymine bind to one another.

By “template” we include a polynucleotide sequence or sequences that is/are used to orientate single-stranded polynucleotide sequences such that they are aligned and may be joined to form an assembled composite polynucleotide sequence. For example, the template may be a single-stranded polynucleotide sequence to which single-stranded “building blocks” of sequence may anneal, thus orientating and aligning the sequences so that they may be joined together.

By “invariant region” we include a region of sequence of a polynucleotide that is typically conserved within the sequence and/or a region of polynucleotide sequence which it is desirable to maintain. For instance, structural regions of the protein may be encoded by the sequence, such as framework regions in an immunoglobulin molecule.

By “variant regions” we include a region of a sequence of a polynucleotide that is typically varied within the sequence and/or may be a region of polynucleotide sequence which it is desirable to alter or vary. For instance, specified regions of the protein may be encoded by the sequence that have a specific functional role in the protein, such as CDR regions in an immunoglobulin molecule.
By “hybrid” we include a double-stranded polynucleotide sequence in which the two strands are sufficiently complementary to anneal to one another but which have sufficiently different sequence to one another that each encodes a different polypeptide and/or contains regulatory sequence elements that direct expression of the encoded polypeptide in a different way.

Preferably, the invention provides a method further comprising the step performed after step (e) of introducing the composite polynucleotide sequence in step (e) or its complement into a vector.

Methods for introducing (i.e. cloning) a polynucleotide sequence into a vector are well known to those skilled in the art of molecular biology. For example, double-stranded polynucleotide sequences may be introduced into a vector by blunt-end or cohesive-end ligation with the vector sequence.

If the single-stranded composite polynucleotide sequence is separated from the single-stranded template polynucleotide, it may be introduced/cloned into a vector by annealing with single-stranded overhangs in the vector. Alternatively, the single-stranded polynucleotide sequence may be manipulated to generate a double-stranded polynucleotide sequence which may be introduced/cloned into a vector by standard techniques using blunt-end or cohesive-end ligation. Methods for generating a double-stranded polynucleotide sequence from a single-stranded polynucleotide sequence are well known in the art of molecular biology and include using a single-stranded polynucleotide sequence as a template for PCR (polymerase chain reaction).

Preferably, the single-stranded composite polynucleotide sequence is separated from the template polynucleotide sequence and PCR-amplified to
generate a double-stranded composite polynucleotide sequence that may be cloned into a vector.

By "vector" we include a vehicle used in gene cloning and/or gene expression to introduce a nucleic acid of interest into a host cell, bacteriophage, virus or yeast. The nucleic acid of interest may be joined to a wide variety of vectors for introduction into an appropriate host. The vector will depend upon the nature of the host, the manner of the introduction of the vector into the host, and whether episomal maintenance or integration is desired.

For example, in bacterial hosts three different types of vector can be used: bacteriophage, cosmids, plasmids and their hybrid derivatives. In these vectors, the nucleic acid of interest can be spliced into the vector using specific restriction enzymes and ligases, or by using homologous recombinantion, although other methods will be well known by those in the art. In some phage vectors part of the viral genome may be removed and replaced with the nucleic acid of interest.

More preferably, the vector is selected from pUC18, pBluescript, pFab5c.His or pFAB60. The pUC18 and pBluescript vectors are cloning vectors that are well known in the art and can be purchased from BioRad Laboratories (Richmond, CA, USA) and Stratagene (La Jolla, USA), respectively.

More preferably, the vector is an expression vector.

By "expression vector" we include vectors that possess regions of nucleotide sequence that direct transcription and/or translation of the nucleic acid of interest such that the protein(s) encoded by the nucleic acid is expressed. If necessary, the nucleic acid of the invention may be linked
to the appropriate transcriptional and translational regulatory control nucleotide sequences recognised by the desired host, although such controls are generally available in expression vectors. Thus, the nucleic acid of the invention may be operatively linked to an appropriate promoter. Bacterial promoters include the *Escherichia coli* (*E. coli*) *lacI* and *lacZ* promoters, the T3 and T7 promoters, the *gpi* promoter, the phage λ PR and PL promoters, the *phoA* promoter and the *trp* promoter. Eukaryotic promoters include the CMV (cytomegalovirus) immediate early promoter, the HSV (herpes simplex virus) thymidine kinase promoter, the early and late SV40 promoters and the promoters of retroviral LTRs (long terminal repeats). Other suitable promoters will be known to the skilled artisan. Expression vectors will desirably also contain sites for transcription initiation and termination, and in the transcribed region, a ribosome binding site for translation (see, for example, WO 98/16643). In addition, expression vectors may contain regions of nucleic acid encoding polypeptide sequences that are fused to the nucleic acid of interest such that the nucleic acid of interest is expressed to form the polypeptide of interest fused to an additional region of polypeptide sequence. The additional region of polypeptide sequence may impart additional properties to the polypeptide of interest, or allow it to be detected using an antibody or bound to a molecule of interest for purification purposes. Such techniques are well known to those in the art and are described, for example, in Sambrook *et al.* (2001).

Many expression systems are known, including systems employing: bacteria (e.g. *E. coli* and *Bacillus subtilis*) transformed with, for example, recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeasts (e.g. *Saccharomyces cerevisiae*) transformed with, for example, yeast expression vectors; insect cell systems transformed with, for example, viral expression vectors (e.g. baculovirus); plant cell systems transfected with, for example viral or bacterial expression vectors; animal cell systems transfected with, for example, adenovirus expression vectors.
The vectors can include a prokaryotic replicon, such as the Col E1 ori, for propagation in a prokaryote, even if the vector is to be used for expression in other, non-prokaryotic cell types. The vectors can also include an appropriate promoter such as a prokaryotic promoter capable of directing the expression (transcription and translation) of the genes in a bacterial host cell, such as E. coli, transformed therewith.

A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with exemplary bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA segment of the present invention.

Typical prokaryotic vector plasmids are: pUC18, pUC19, pBR322 and pBR329 available from BioRad Laboratories (Richmond, CA, USA); pTrc99A, pKK223-3, pKK233-3, pDR540 and pRIT5 available from Pharmacia (Piscataway, NJ, USA); pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16A, pNH18A, pNH46A available from Stratagene Cloning Systems (La Jolla, CA, USA).

A typical mammalian cell vector plasmid is pSVL available from Pharmacia (Piscataway, NJ, USA). This vector uses the SV40 late promoter to drive expression of cloned genes, the highest level of expression being found in T antigen-producing cells, such as COS-1 cells. An example of an inducible mammalian expression vector is pMSG, also available from Pharmacia (Piscataway, NJ, USA). This vector uses the glucocorticoid-inducible promoter of the mouse mammary tumour virus long terminal repeat to drive expression of the cloned gene.
Useful yeast plasmid vectors are pRS403-406 and pRS413-416 and are generally available from Stratagene Cloning Systems (La Jolla, CA, USA). Plasmids pRS403, pRS404, pRS405 and pRS406 are Yeast Integrating plasmids (YIps) and incorporate the yeast selectable markers \( \text{HIS3}, \text{TRP}1, \text{LEU2} \) and \( \text{URA3} \). Plasmids pRS413-416 are Yeast Centromere plasmids (YCps).

Preferably, the expression vector is pFab51His or pFAB60 (described in Engberg \textit{et al.}, 1996 and Dziegieł \textit{et al.}, 1995).

Methods well known to those skilled in the art can be used to construct expression vectors containing the coding sequence and, for example appropriate transcriptional or translational controls. One such method involves ligation via homopolymer tails. Homopolymer polydA (or polydC) tails are added to exposed 3' OH groups on the DNA fragment to be cloned by terminal deoxynucleotidyl transferases. The fragment is then capable of annealing to the polydT (or polydG) tails added to the ends of a linearised plasmid vector. Gaps left following annealing can be filled by DNA polymerase and the free ends joined by DNA ligase.

Another method involves ligation via cohesive ends. Compatible cohesive ends can be generated on the DNA fragment and vector by the action of suitable restriction enzymes. These ends will rapidly anneal through complementary base pairing and remaining nicks can be closed by the action of DNA ligase.

A further method uses synthetic molecules called linkers and adaptors. DNA fragments with blunt ends are generated by bacteriophage T4 DNA polymerase or \( \text{E. coli} \) DNA polymerase I which remove protruding 3' termini and fill in recessed 3' ends. Synthetic linkers, pieces of blunt-ended double-stranded DNA which contain recognition sequences for defined restriction
enzymes, can be ligated to blunt-ended DNA fragments by T4 DNA ligase. They are subsequently digested with appropriate restriction enzymes to create cohesive ends and ligated to an expression vector with compatible termini. Adaptors are also chemically synthesised DNA fragments which contain one blunt end used for ligation but which also possess one preformed cohesive end.

More preferably, the invention provides a method further comprising the step of expressing the composite polynucleotide sequence generated in step (d), (e) or (f) or its complement to generate a corresponding polypeptide.

By “expression” or “expressing” we include the events that direct transcription and/or translation of a nucleic acid of interest such that the protein(s) encoded by the nucleic acid is produced. Many methods of expressing nucleic acids are known to those in skilled the relevant arts of molecular biology and biochemistry. For example, nucleic acids may be expressed by incubating the nucleic acid in vitro with components of the cellular machinery that direct transcription and translation. Alternatively, nucleic acids may be linked to the appropriate transcriptional and translational regulatory control nucleotide sequences recognised by a host cell, thereby enabling expression of the protein encoded by the nucleic acid in vivo (i.e. in that cell).

More preferably, the invention provides a method further comprising the step of testing the polypeptide for desired characteristics.

Polypeptides display many characteristics, including for example, binding affinity for a ligand, stability, cellular location, sensitivity to a drug, and the ability to function as an enzyme in a biological or chemical reaction. By “desired characteristics” we mean a characteristic displayed by a polypeptide that is desired by the user of the method of the invention – for
example, the user may use the method of the invention to increase or reduce an existing characteristic of the polypeptide or to impart a new characteristic to the polypeptide. Suitable methods for testing a polypeptide for a desired characteristic such as binding affinity, stability, enzymatic activity, cellular location, drug sensitivity and other properties of proteins will be well known to those in the relevant arts of biochemistry, molecular biology and chemistry.

For example, when the polypeptide comprises an immunoglobulin binding domain, for example is an antibody or antibody fragment, methods that may be used in determining the binding characteristics of the domain, antibody or antibody fragment include binding assays such as, for example, ELISA, and functional assays, as would be appreciated by a person skilled in the art. Further studies may include immunohistochemistry for assessment of target specificity and cross-reactivity. Such techniques are well known in the art and are described, for example, in Sambrook et al. (2001).

In a second aspect, the invention provides a method for producing a library of different composite polynucleotide sequences comprising the steps:

a) providing one or more single-stranded template polynucleotide sequence comprising at least one invariant region and at least one variant region;

b) providing a plurality of different single-stranded invariant polynucleotide sequences that are at least partially complementary to the at least one invariant region of the one or more template polynucleotide sequence, and a plurality of different single-stranded variant polynucleotide sequences that are at least partially complementary to the at least one variant region of the one or more template polynucleotide sequence,
wherein the single-stranded invariant and variant polynucleotides, when ligated together, are capable of forming a polynucleotide which is at least partially complementary to the one or more template polynucleotide;

c) annealing the single-stranded polynucleotide sequences provided in steps (a) and (b) to generate a plurality of double-stranded polynucleotide sequences wherein each of the plurality of double-stranded polynucleotide sequences comprises one strand that is a template polynucleotide sequence, and one strand that is composed of at least one invariant polynucleotide sequence and at least one variant polynucleotide sequence that are not ligated together;

d) subjecting the plurality of double-stranded polynucleotide sequences generated in step (c) to conditions that ligate the at least one invariant polynucleotide sequence and the at least one variant polynucleotide sequence together to generate a plurality of double-stranded hybrid polynucleotide sequences, wherein each of the plurality of double-stranded hybrid polynucleotide sequences is composed of one strand that is a template polynucleotide sequence, and one strand that is a different composite polynucleotide sequence comprising at least one invariant polynucleotide sequence and at least one variant polynucleotide sequence;

e) separating a plurality of different composite polynucleotide sequences generated in step (d) from the plurality of template polynucleotide sequences.
Thus, the second aspect of the invention provides a method of producing a collection of composite polynucleotide sequences, each of which has a different sequence. Such a collection of molecules is known in the art as a "library". Once produced, a library of molecules such as polynucleotides or polypeptides may be screened for individual members that have desirable characteristics and once identified can be isolated for further study and/or use.

As discussed above, it will be understood that the single-stranded invariant polynucleotide sequences need not be completely complementary to the at least one invariant region of the template polynucleotide sequence, and that the single-stranded variant polynucleotide sequences need not be completely complementary to the at least one variant region of the template polynucleotide sequence.

More preferably, the invention provides a method further comprising the step performed after step (e) of introducing the plurality of different polynucleotide sequences generated in step (e) or their complements into a plurality of vectors.

The step of introducing the plurality of different polynucleotide sequences (whether double-stranded or single-stranded) into a plurality of vectors may be performed using methods as described above. Preferably, the single-stranded composite polynucleotide sequences are separated from the template polynucleotide sequences and PCR-amplified to generate a double-stranded composite polynucleotide sequence that may be cloned into a vector.

Preferably, the invention provides a method further comprising the step of expressing the plurality of different composite polynucleotide sequences
generated in steps (d), (e) or (f) of the second aspect of the invention (and/or their complements) to generate a corresponding library of polypeptides.

More preferably, the invention provides a method further comprising the step of screening the library of polypeptides for desired characteristics and selecting a polypeptide with desired characteristics.

By “screening” we include the process and/or step of testing polypeptides in the library of polypeptides for desired characteristics. Only a relatively small number of the total number of polypeptides in a library may have the desired characteristics and screening is necessary to identify those with the characteristics of interest. Once the screening step has identified one or more polypeptide with the desired characteristics, they may be selected for further study and/or use.

The screening method preferably allows the sequence encoding the polypeptide with the desired characteristics to be identified or retrieved.

Screening may be performed could be based on any of the methods commonly used for library screening, such as phage display, ribosome display and chip-based technologies. Preferably, screening is performed using phage display methods based on the pFab5c.His vector mentioned above. Engberg et al. (1996) provides details on how to construct, express and select (by “panning”) antibodies from phage display libraries using the pFab5c.His vector.

The sequence may then be manipulated as well known to those skilled in the art, for example in order to express a fusion polypeptide comprising sequences useful for a particular application. Such sequences may include those encoding an antibody (such as IgG) or a fragment or fragments thereof (such as the constant region of an Fab molecule), and sequences for
detection and/or purification of the polynucleotide or polypeptide such as those encoding a poly-histidine tag, a FLAG tag, a V5 tag, and GST (glutathione S-transferase).

Preferably, the invention provides a method further comprising the step of selecting the polynucleotide sequence encoding the polypeptide with desired characteristics.

More preferably, the invention provides a method further comprising the step of introducing the polynucleotide or part thereof into a vector.

Preferably, the invention provides a method further comprising the step of expressing a polypeptide encoded at least in part by the selected polynucleotide or part thereof.

More preferably, the invention provides a method further comprising the step of formulating the expressed polypeptide into a pharmaceutical composition.

Pharmaceutical compositions and methods for formulating polypeptides into a pharmaceutical composition are well-known to those in the relevant art.

Preferably, the invention provides a method wherein the template polynucleotide sequence encodes (or is complementary to a polynucleotide sequence encoding) an immunoglobulin or a fragment or domain thereof.

More preferably, the invention provides a method wherein the immunoglobulin is an antibody.
More preferably, the invention provides a method wherein the template polynucleotide sequence (or its complement) encodes an immunoglobulin variable domain.

More preferably, the invention provides a method wherein the immunoglobulin variable domain is a light chain immunoglobulin variable domain ($V_L$) or a heavy chain immunoglobulin variable domain ($V_H$).

Even more preferably, the invention provides a method wherein the light chain immunoglobulin variable domain ($V_L$) is DPL3 or the heavy chain immunoglobulin variable domain ($V_H$) is DP-47. DPL3 and DP-47 are of human germline origin for light and heavy chain respectively ($VH3$ and $\lambda\lambda1$ gene families) and are well expressed in both in bacterial and mammalian expression systems.

By “immunoglobulins” we include polypeptides comprising one or more immunoglobulin complementarity-determining region (CDR), such as antibodies, B cell receptors or T cell receptors, or fragments thereof.

Antibodies comprise two identical polypeptides of $M_r$ 50,000-70,000 (termed “heavy chains”) that are linked together by a disulphide bond, each of which is linked to one of an identical pair of polypeptides of $M_r$ 25,000 (termed “light chains”). There is considerable sequence variability between individual N-termini of heavy chains of different antibody molecules and between individual light chains of different antibody molecules and these regions have hence been termed “variable domains”. Conversely, there is considerable sequence similarity between individual C-termini of heavy chains of different antibody molecules and between individual light chains of different antibody molecules and these regions have hence been termed “constant domains”.
The antigen-binding site is formed from hyper-variable regions in the variable domains of a pair of heavy and light chains. The hyper-variable regions are also known as complementarity-determining regions (CDRs) and determine the specificity of the antibody for a ligand. The variable domains of the heavy chain \( (V_H) \) and light chain \( (V_L) \) typically comprise three CDRs, each of which is flanked by sequence with less variation, which are known as framework regions (FRs).

The variable heavy \( (V_H) \) and variable light \( (V_L) \) domains of the antibody are involved in antigen recognition, a fact first recognised by early protease digestion experiments. Further confirmation was found by “humanisation” of rodent antibodies. Variable domains of rodent origin may be fused to constant domains of human origin such that the resultant antibody retains the antigenic specificity of the rodent parented antibody (Morrison et al. (1984).

That antigenic specificity is conferred by variable domains and is independent of the constant domains is known from experiments involving the bacterial expression of antibody fragments, all containing one or more variable domains. These molecules include Fab-like molecules (Better et al., 1988); Fv molecules (Skerra et al., 1988); single-chain Fv (ScFv) molecules where the \( V_H \) and \( V_L \) partner domains are linked via a flexible oligopeptide (Bird et al., 1988; Huston et al., 1988) and single domain antibodies (dAbs) comprising isolated V domains (Ward et al., 1989). A general review of the techniques involved in the synthesis of antibody fragments which retain their specific binding sites is to be found in Winter et al. (1991).

By “ScFv molecules” we mean molecules wherein the \( V_H \) and \( V_L \) partner domains are linked via a flexible oligopeptide.

The advantages of using antibody fragments, rather than whole antibodies, are several-fold. The smaller size of the fragments may lead to improved
pharmacological properties, such as better penetration of solid tissue. Effector functions of whole antibodies, such as complement binding, are removed. Fab, Fv, ScFv and dAb antibody fragments can all be expressed in and secreted from E. coli, thus allowing the facile production of large amounts of the said fragments.

Whole antibodies, and F(\text{ab}')_2 fragments are “bivalent”. By “bivalent” we mean that the said antibodies and F(\text{ab}')_2 fragments have two antigen combining sites. In contrast, Fab, Fv, ScFv and dAb fragments are monovalent, having only one antigen combining sites.

B cell receptors comprise a membrane-bound antibody molecule which differs from secreted antibody only in the C-terminal domain.

The structure of T cell receptors is similar to that of antibodies. The T cell receptor consists of two polypeptide chains joined by a disulphide bond, each of which has an N-terminal variable domain and a C-terminal constant domain. The variable domains possess three CDRs which appear to be equivalent to CDRs in antibody heavy and light chains.

By “immunoglobulin variable domain” we include the region of a heavy chain or light chain of an immunoglobulin located at its N-terminal end, which may comprise three complementarity-determining regions (termed CDRs or CDR1, CDR2 and CDR3) and four framework regions (termed FRs or FR1, FR2, FR3 and FR4), as distinct from the constant region located at the C-terminal end of the heavy or light chain.

Preferably, the invention provides a method wherein the template sequence comprises three variant regions and four invariant regions.
Preferably, the invention provides a method wherein the variant regions encode (or are complementary to polynucleotides encoding) immunoglobulin complementarity-determining regions (CDRs) and the invariant regions encode (or are complementary to polynucleotides encoding) immunoglobulin framework regions (FRs).

More preferably, the invention provides a method wherein the immunoglobulin framework regions are from the DP-47 heavy chain immunoglobulin variable domain (V_H) or the DPL-3 light chain immunoglobulin variable domain (V_L).

By “complementarity-determining regions” or “CDRs” we include regions comprising the amino acid residues located within the hyper-variable regions of the immunoglobulin variable domain that form an antigen binding site, and hence determine the specificity of an antibody or antigen-binding fragment thereof.

By “framework regions” or “FRs” we include regions of the immunoglobulin domain other than the CDRs which form the structural framework within which the CDRs are positioned to form the antigen-binding site. FR1 is typically the region at the N-terminus of the variable domain located on the N-terminal side of CDR1; FR2 and FR3 are located between CDR1 and CDR2 and between CDR2 and CDR3, respectively; FR4 is located at the C-terminus of the variable domain on the C-terminal side of CDR3.

Preferably, the invention provides a method wherein the single-stranded invariant polynucleotide sequences are obtained from sequences encoding (or complementary to sequences encoding) immunoglobulin framework regions (FRs) and the single-stranded variant polynucleotide sequences are
obtained from sequences encoding (or complementary to sequences encoding) immunoglobulin complementarity-determining regions (CDRs).

 Preferably, the invention provides a method wherein the single-stranded invariant polynucleotide sequences are obtained from sequences encoding (or complementary to sequences encoding) immunoglobulin framework regions (FRs) and the single-stranded variant polynucleotide sequences are oligonucleotides of random sequence.

 An oligonucleotide is typically a single-stranded polynucleotide sequence that is synthesised in vitro and may comprise the nucleotides adenine, thymine, cytosine and guanine which are capable of forming hydrogen bonds with complementary nucleotides. Alternative nucleotides may be included in oligonucleotides to facilitate hydrogen bonding to a larger number of nucleotides – for example, the nucleotide inosine, a universal base-pairing base. Techniques for synthesising oligonucleotide sequences are well known in the art and oligonucleotides can be synthesised that have a known sequence defined by the user or that have a random sequence.

 Preferably, the invention provides a method wherein the sequences encoding (or complementary to sequences encoding) immunoglobulin framework regions (FRs) and/or immunoglobulin complementarity-determining regions (CDRs) are obtained from a library of different polynucleotide sequences encoding (or complementary to sequences encoding) immunoglobulins with different binding specificities.

 Sequences encoding different immunoglobulin framework regions (FRs) and different immunoglobulin complementarity-determining regions (CDRs) for use in the method of the invention may be obtained from a pre-existing library of sequences that encode antibodies with different binding specificities. Such libraries are well known to those in the art and standard
molecular biology techniques (such as polymerase chain reaction (PCR) using oligonucleotide primers specific for variable domain regions, or restriction enzyme digestion) may be used to obtain sequences encoding FRs and CDRs from such libraries. It will be understood that libraries containing nucleotide sequences encoding full-length immunoglobulin polypeptides or fragments thereof (for example, encoding immunoglobulin variable domains or regions thereof) may be used to obtain sequences for use in the method of the invention.

Kits for generating phage display libraries are available commercially. Kit of use in the method of the invention include the Recombinant Phage Antibody System (Pharmacia, Piscataway, NJ, USA) and the Surf.ZAP™ phage display kit (Stratagene Europe, Amsterdam, The Netherlands).

One library that may be used in the method of the invention is the “Tomlinson I + J scFv” library that can be obtained through the Medical Research Council gene service (MRC, Cambridge, UK - http://www.hgmp.mrc.ac.uk/geneservice/index.shtml)

Preferably, the invention provides a method wherein the sequences encoding (or complementary to sequences encoding) immunoglobulin framework regions (FRs) and/or immunoglobulin complementarity-determining regions (CDRs) are obtained by mutagenising a polynucleotide encoding (or complementary to a polynucleotide encoding) an immunoglobulin.

In order to generate sequences encoding immunoglobulin framework regions (FRs) and immunoglobulin complementarity-determining regions (CDRs) for use in the method of the invention, a polynucleotide encoding an immunoglobulin may be subjected to mutagenesis to create a plurality of differently mutated derivatives thereof. Random mutagenesis can be
accomplished by a variety of methods known in the art, such as site-directed mutagenesis (Alber et al., 1987), combinatorial cloning (Huse et al., 1989; Marks et al., 1992) and random mutagenesis combined with appropriate selection systems (Barbas et al, 1992).

Preferably, the invention provides a method wherein mutagenising the polynucleotide encoding an immunoglobulin involves a polymerase, for example, error-prone polymerase chain reaction (PCR) amplification.

A particularly preferred method is error-prone polymerase chain reaction (PCR) wherein a polynucleotide sequence of interest is amplified using a polymerase under conditions which incorporate errors into the amplification products to generate a plurality of differently mutated derivatives of the polynucleotide sequence.

Preferably, the invention provides a method wherein ligating the at least one invariant single-stranded polynucleotide sequence and the at least one variant single-stranded polynucleotide sequence together in step (d) is performed using DNA ligase.

It is well known that DNA ligase is an enzyme that catalyses the formation of a phosphodiester bond between the 3'-hydroxyl (3'-OH) group at the end of a first DNA chain and the 5'-phosphate group at the end of a second DNA chain, thereby joining the first and second DNA chains. Various forms of DNA ligase have been described and are used by those in the art such as, for example, T4 DNA ligase which is encoded by the T4 bacteriophage and is available commercially from various manufacturers such as Invitrogen (Invitrogen Ltd, Paisley, UK).

Preferably, the invention provides a method according to any preceding claim further comprising step (d2) performed after step (d) but before step
(e) of separating the double-stranded hybrid polynucleotide sequence generated in step (d) from single-stranded invariant polynucleotide sequences and/or single-stranded variant polynucleotide sequences that are not annealed to the template polynucleotide sequence.

More preferably, the invention provides a method wherein the single-stranded template polynucleotide sequence is linked to a member of a specific binding pair (MSBP) and step (d2) comprises binding the MSBP to its specific binding partner.

By “a specific binding pair” we include two molecules that bind specifically to one another. Such pairs of molecules may include a receptor and its specific ligand or an enzyme and its specific substrate.

By “member of a specific binding pair (MSBP)” we include either one of the two molecules comprising that binding pair. Many binding pairs are known to those skilled in the art and are of use in the method of the invention. For example, a binding pair may consist of an antibody and its specific target (such as digoxigenin and anti-digoxigenin antibodies, available from Roche Diagnostics GmbH, Germany) or nucleic acid sequences with regions of homology (such as polyA and polydT sequences).

Conveniently, the MSBP is biotin and its specific binding partner is streptavidin. The use of these molecules is well known to those skilled in the art and biotin and streptavidin reagents are widely available, for example from Sigma-Aldrich Company Ltd, Fancy Road, Poole, Dorset, BH12 4QH, UK, and Dynal ASA, Oslo, Norway.

Preferably, the invention provides a method wherein the template polynucleotide sequence is more sensitive to degradation than the composite polynucleotide sequence and step (e) comprises the step of
subjecting the double-stranded hybrid polynucleotide sequence generated in step (d) to conditions capable of degrading the template polynucleotide sequence but not the composite polynucleotide sequence.

It will be understood that the composite polynucleotide sequence could be separated from the template polynucleotide sequence by selectively degrading the template polynucleotide sequence. Methods for doing so are well known to those skilled in the art. For example, the template polynucleotide sequence for use in the method of the invention may be transformed into and replicated in a bacterial strain that has been modified so that replicated nucleic acids are less stable than those replicated in a bacterial strain that has not been modified.

Preferably, the invention provides a method wherein thymine in the template polynucleotide sequence has been replaced with uracil. Conveniently, this is achieved by generating the template polynucleotide sequence using a bacterial strain that is dut and ung. One such bacterial strain that is dut and ung and may be used in the method of the invention is the E. coli strain K12 CJ236 which can be purchased from New England Biolabs (Beverly, MA, USA).

The dut gene encodes the dUTPase enzyme that is involved in converting dUTP to dUMP. Replicating DNA in bacterial strains lacking a functional dut gene (i.e. dut) results in an increased frequency of uracil incorporation into positions in the replicated DNA that should be occupied by thymine.

The ung gene encodes a uracil-DNA glycosylase enzyme that is involved in removing uracil that has been incorporated into DNA. Replicating DNA in bacterial strains lacking a functional ung gene (i.e. ung) results in replicated DNA containing uracil instead of thymine.
Preferably, the invention provides a method wherein the conditions capable of degrading the template polynucleotide sequence but not the composite polynucleotide sequence comprise transforming the double-stranded hybrid polynucleotide sequence into a bacterial cell capable of replicating the composite polynucleotide sequence but not the template polynucleotide sequence.

Introducing DNA that contains uracil instead of thymine (i.e. DNA that has been generated in a $dut^- ung^-$ bacterial strain) into a bacterial strain which is $dut^+$ and $ung^+$ results in the degradation of DNA containing uracil instead of thymine.

The method of using a bacterial strain that is $dut^-$ and $ung^-$ to generate an unstable template strand and selectively degrade it is described in Kunkel, (1985) and reviewed in Kunkel et al. (1991).

Preferably, the invention provides a method wherein step (e) comprises the steps:

i) ligating the double-stranded hybrid polynucleotide sequence generated in step (d) or (d2) into a vector;

ii) introducing the vector containing the double-stranded hybrid polynucleotide sequence into a cell capable of replicating the vector containing the double-stranded hybrid polynucleotide sequence;

iii) isolating a plurality of vectors from the cell progeny comprising one or more vectors containing the template polynucleotide sequence and one or more vectors containing the composite polynucleotide sequence;
iv) subjecting the plurality of vectors to conditions that fragment the one or more vectors containing the template polynucleotide sequence but that do not fragment the one or more vectors containing the composite polynucleotide sequence;

v) separating the one or more vectors containing the composite polynucleotide sequence from the fragmented one or more vectors containing the template polynucleotide sequence.

Separating the double-stranded hybrid polynucleotide sequence generated in step (d) from single-stranded template polynucleotide sequence and/or single-stranded invariant and variant polynucleotide sequences that have not been ligated together may be accomplished using standard molecular biology techniques known in the art. For example, the double-stranded hybrid polynucleotide sequence may be separated from single-stranded polynucleotide sequences by using agarose gel electrophoresis and/or ethanol precipitation and/or membrane filtration and/or size-charge separating matrix.

Alternatively, the template strand may be linked to a member of a specific binding pair (MSBP) and separating the double-stranded hybrid polynucleotide sequence generated in step (d) from single-stranded invariant and variant polynucleotide sequences that have not been ligated together may be accomplished by binding the MSBP to its specific binding partner. Preferably, the MSBP is biotin and its specific binding partner is streptavidin. The use of these molecules is well known to those skilled in the art and biotin and streptavidin reagents are widely available, for example from Sigma-Aldrich Company Ltd, Fancy Road, Poole, Dorset, BH12 4QH, UK, and Dynal ASA, Oslo, Norway.
It will be understood that it may be necessary to generate suitable fragments of the double-stranded hybrid polynucleotide sequence generated in step (d) or (d2) for ligation into a vector (in step (i), above). Suitable fragments for ligation into a vector may be generated by digesting the termini of the double-stranded hybrid polynucleotide sequence with one or more restriction enzyme(s), thereby providing suitable termini for ligation with the termini of the vector. Suitable termini may be blunt-ends and/or cohesive-ends.

Alternatively, suitable fragments for ligation into a vector may be generated by designing the outermost invariable polynucleotide sequences to be ligated on the template strand such that they have blunt or cohesive ends.

Alternatively, the double-stranded hybrid polynucleotide sequence may be denatured to obtain single stranded polynucleotide sequence which may be annealed to complementary single-stranded ends in a denatured vector or to single-stranded overhangs at the end of a non-denatured vector.

Introducing vectors and nucleic acids into appropriate cell hosts can be accomplished by well known methods that typically depend on the type of vector used. With regard to transformation of prokaryotic host cells, see, for example, Cohen et al (1972) and Sambrook et al (2001). Transformation of yeast cells is described in Sherman et al (1986). The method of Beggs (1978) is also useful. With regard to vertebrate cells, reagents useful in transfecting such cells, for example calcium phosphate and DEAE-dextran or liposome formulations, are available from Stratagene Cloning Systems, or Life Technologies Inc., Gaithersburg, MD, USA.

Physical methods may be used for introducing DNA into animal and plant cells. For example, microinjection uses a very fine pipette to inject DNA molecules directly into the nucleus of the cells to be transformed. Another
example involves bombardment of the cells with high-velocity micro-
projectiles, usually particles of gold or tungsten that have been coated with
DNA.

Successfully transformed cells, i.e. cells that contain a vector of the invention,
can be identified by well known techniques. For example, one selection
technique involves incorporating into the expression vector a DNA sequence
(marker) that codes for a selectable trait in the transformed cell. These
markers include dihydrofolate reductase, G418 or neomycin resistance for
eukaryotic cell culture, and tetracyclin, kanamycin or ampicillin resistance
genes for culturing in E. coli and other bacteria. Alternatively, the gene for
such selectable trait can be on another vector, which is used to co-transform
the desired host cell.

The marker gene can be used to identify transformants but it is desirable to
determine which of the cells contain recombinant DNA molecules and which
contain self-ligated vector molecules. This can be achieved by using a cloning
vector where insertion of a DNA fragment destroys the integrity of one of the
genes present on the molecule. Recombinants can therefore be identified
because of loss of function of that gene.

Another method of identifying successfully transformed cells involves
growing the cells resulting from the introduction of an expression construct of
the invention to produce a polypeptide of the invention. Cells can be
harvested and lysed and their DNA content examined for the presence of the
DNA using a method such as that described by Southern (1975) or Berent et
al. (1985). Alternatively, the presence of the protein in the supernatant can be
detected using antibodies as described below.

In addition to directly assaying for the presence of vector, successful
transformation can be confirmed by well known immunological methods
when the vector is capable of directing the expression of the protein. For example, cells successfully transformed with an expression vector produce proteins displaying appropriate antigenicity. Samples of cells suspected of being transformed are harvested and assayed for the protein using suitable antibodies.

Thus, in addition to the transformed host cells themselves, the present invention also contemplates a culture of those cells, preferably a monoclonal (clonally homogeneous) culture, or a culture derived from a monoclonal culture, in a nutrient medium.

It will be understood by those skilled in the art that as vector are replicated semi-conservatively, introducing a vector containing a double-stranded hybrid molecule into a cell and allowing the cell to replicate the vector will result in the production of a plurality of vectors comprising one or more vectors containing the template polynucleotide sequence and one or more vectors containing the composite polynucleotide sequence. Purifying vector from these cells will therefore result in a plurality of vectors comprising one or more vectors containing the template polynucleotide sequence and one or more vectors containing the composite polynucleotide sequence.

By “a cell capable of replicating the vector” we include any cell capable of generating copies of the vector during growth and/or division of the cell such that one or more copy of the vector is transferred to one or more daughter cell(s) generated by growth and/or division of the parent cell. Suitable cells for use in the method of the invention may be prokaryotic or eukaryotic cells, such as those described in the preceding paragraphs.

The step of isolating/purifying vector from the cell progeny may be performed using well known methods. For example, plasmid vector DNA can be prepared on a large scale from cleaved lysates by banding in a caesium
chloride (CsCl) gradient according to the methods of Clewell et al. (1970) and Clewell (1972). Plasmid DNA extracted in this way can be freed from CsCl by dialysis against sterile, pyrogen-free buffer through dialysis tubing or by size-exclusion chromatography. Alternatively, plasmid vector DNA may be purified from cleared lysates using ion-exchange chromatography, for example those supplied by Qiagen. Hydroxyapatite column chromatography may also be used.

Following the isolation/purification step, the isolated/purified vector is subjected to conditions that fragment the one or more vectors containing the template polynucleotide sequence. By “fragment” we include the step of cutting/cleaving the strands of the vector containing the template polynucleotide sequence so that two or more pieces (i.e. fragments) of vector and/or polynucleotide sequence are formed. Methods for selectively fragmenting a vector include digestion with a restriction enzyme that selectively recognises and cleaves that vector.

Preferably, the invention provides a method wherein the one or more vectors containing the template polynucleotide sequence comprise a recognition site for a restriction endonuclease that is not present in the one or more vectors containing the composite polynucleotide sequence and step (v) comprises treating the plurality of vectors with a restriction endonuclease specific for said recognition site.

Restriction enzymes (also known as restriction endonucleases) recognise specific sequences of nucleotides in double-helical DNA and cleave both strands of the double-stranded molecule at specific locations in the sequence. Many restriction enzymes are known, and each recognises a specific (and often unique) sequence of nucleotides in double-stranded DNA.
Following fragmentation of the one or more vector containing the template polynucleotide sequence, the one or more vector containing the composite polynucleotide sequence is isolated from other sequences and reagents.

Preferably, the invention provides a method wherein step (vi) comprises introducing the one or more vectors containing the composite polynucleotide sequence and the fragmented one or more vectors containing the template polynucleotide sequence into a cell capable of replicating the one or more vector containing the composite polynucleotide sequence but not capable of replicating the fragmented one or more vectors containing the template polynucleotide sequence.

It is well known that vectors may contain regulatory control nucleotide sequences that are recognised by the desired host cell and direct replication of the vector in the cell. For example, prokaryotic vectors can include a prokaryotic replicon, such as the ColEl ori, for propagation in a prokaryote. Therefore, fragmenting a vector may result in the cleaving/cutting of regulatory control nucleotide sequences required for replication resulting in the fragments of vector not being replicated when they are introduced into a cell. By “a cell capable of replicating the one or more vector containing the composite polynucleotide sequence but not capable of replicating the fragmented one or more vectors containing the template polynucleotide sequence” we include a cell that is capable of replicating a complete vector but not capable of replicating fragments of a vector.

More preferably, the invention provides a method wherein the cell is a bacterial cell. Conveniently, the bacterial cell is Escherichia coli (E. coli). Whilst it will be understood that many strains of E. coli may be used in the method of the invention, particularly preferred strains are TOP10F’ (available from Invitrogen Ltd, Paisley, UK) and HB101F’ (NCCB4285,

In a further aspect, the invention provides a polynucleotide sequence obtained according to the methods of the invention.

In a further aspect, the invention provides a polypeptide obtained according to the methods of the invention. In a further aspect, the invention provides a library of different composite polynucleotide sequences obtained according to the methods of the invention.

In a further aspect, the invention provides a library of different composite polypeptides obtained according to the methods of the invention. In a further aspect, the invention provides a pharmaceutical preparation obtained by the methods of the invention.

Whilst it is possible for a polypeptide of the invention to be administered alone, it is preferable to present it as a pharmaceutical formulation, together with one or more acceptable carriers. The carrier(s) must be “acceptable” in the sense of being compatible with the polypeptide of the invention and not deleterious to the recipients thereof. Typically, the carrier(s) will be water or saline which will be sterile and pyrogen free. The method and route of administration of an effective amount of the polypeptide or pharmaceutical composition of the invention will depend on the reason for administration (such as, for example, the treatment of a medical condition) and will be readily apparent to those skilled in the relevant art.

Preferred, non-limiting examples which embody certain aspects of the invention will now be described, with reference to the following figures:
Figure 1A: *Annealing of single-stranded oligonucleotide fragments onto single-stranded template polynucleotide molecules.* Pools of single-stranded oligonucleotide fragments containing variability (multi-coloured) or constituting framework regions with no variability (black) are allowed to hybridize to single-stranded template DNA (red) encoding a specific gene or part of a gene. The fragments anneal through complementary base-pairing and differences in length will be accommodated by looping out of either the template DNA or the oligonucleotide fragments.

Figure 1B: *Principle of the method of the invention.* The steps included in the method of the invention up to the creation of a new library are shown.

Figure 2: *Construction of a small V\textsubscript{H} library.* An assembly reaction experiment was done, using CDRH3 single-stranded fragments from five antibody scFv clones with different lengths (6 to 22 amino acids in length) and sequences as a pool, while keeping the CDRH1 and CDRH2 single-stranded fragments constant. The length of the CDRH3 sequence used as the template was 10 amino acids in length.

Figure 3: *Distribution of CDRH3 length in a V\textsubscript{H} library with CDRs from lymphoid tissue.* An assembly reaction was done with single-stranded DNA fragments encoding CDRH1, CDRH2 and CDRH3 regions isolated from immunoglobulin heavy chain mRNA from lymphoid tissue. The length of the CDRH3 sequence used as the template was 10 amino acids in length.

Figure 4: *Distribution of CDRH3 length in a V\textsubscript{H} library with CDRs from lymphoid tissue.* An assembly reaction was done as in Fig. 3 with single-stranded DNA fragments encoding CDR regions, isolated from immunoglobulin heavy chain mRNA from lymphoid tissue. The length of the CDRH3 sequence used as the template was 6 amino acids in length.
EXAMPLE – Experimental data

Material and Methods

Construction of naïve V\textsubscript{H} and V\textsubscript{L} sub-libraries

Two naïve sub-libraries were constructed, one containing the immunoglobulin (Ig) variable light chain (V\textsubscript{L}) and the other containing the variable heavy chain (V\textsubscript{H}).

Starting with mRNA originating from human lymphoid material, a first strand cDNA synthesis was made using Ig-specific primers complementary to the \(\lambda\)-constant region or the \(\gamma_1\)-constant region and SuperScript Choice System for cDNA Synthesis (Invitrogen Ltd, Paisley, UK).

Briefly, 5\(\mu\)g mRNA and 2\(\mu\)M primer in DEPC H\textsubscript{2}O were heated to 80\(^\circ\)C for 5 min and slowly allowed to cool to a temperature suitable for the primer. The rest of the reagents were mixed according to the manufacturer’s protocol, heated to the same temperature as for the mRNA and added to the mRNA tube. The reaction was allowed to proceed for 1h and then heat inactivated at 85\(^\circ\)C for 5 min.

A standard second strand synthesis using Rnase H, DNA polymerase and DNA ligase was then made using the first strand synthesis as template.

The Ig-specific primer is biotinylated, this enables the product to be bound to streptavidin coated surfaces and thereby facilitates downstream applications. The double stranded (ds) cDNA product was coupled to streptavidin coated magnetic beads (Dynal ASA, Norway; Dynabeads M-280) and incubated on a Dynal sample mixer in room temperature for 2-3h and then washed using 1x digestion buffer of choice. Restriction enzymes cutting at naturally occurring restriction sites upstream the gene were used
to give the fragment correct sticky ends. The digestion was incubated on the sample mixer at 37°C for 2h. The digestion is then washed a second time using 1x T4 DNA ligation buffer (1x ligase buffer: 50 mM Tris-HCl pH 7.6, 10 mM MgCl2, 1 mM ATP, 1 mM DTT, 12.5% (w/v) polyethylene glycol-8000). To isolate the double-stranded (ds) complementary DNA (cDNA) from the biotin:streptavidin complex another restriction site either introduced by the specific primer or a natural restriction site downstream the gene was used (incubated 2h, 37°C). This digestion was made in 1x ligation buffer. The supernatant, now containing the isolated ds cDNA and the second restriction enzyme in ligation buffer, was heat inactivated at 65°C for 15 min.

The isolated fragments were ligated into either the pUL (for light chains) or the pUH (for heavy chains) vectors by incubation together with T4 DNA ligase in room temperature for 1h. The ligation mix was diluted 10 times and then electroporated into E. coli TOP10 or HB101. To determine the library size a small part of the electroporation was spread on agar plates. The larger part was grown over night in TB medium, separated by mild centrifugation and resolved in 50% sterile glycerol and kept in -70°C. A smaller part was used to isolate the DNA that is to be used in the single stranded CDR amplification.

**Sequencing**

To be able to determine the variability of the naive sub-libraries, a large number of clones were sequenced. Sequencing reactions were performed using the BigDye terminator kit v3 (Applied Biosystems) and the purified reactions were run on a ABI Prism Genetic Analyzer 3100.

**Single stranded CDR fragments**

The single-stranded (ss) CDR fragments from light (V_L) and heavy (V_H) variable region of the antibody were amplified from the sub-libraries in
pUL- or pUH-vectors. The three CDRs from the immunoglobulin light chain are referred to as CDRL1, L2 and L3 respectively, and the three CDRs from the heavy chain are referred to as CDRH1, H2 and H3 respectively. The amplification was made using a single primer PCR (linear PCR) on the antibody sequence-containing vectors digested with a restriction enzyme specific for a sequence upstream of each CDR and that is naturally occurring. In this way the ss CDR fragments produced by the linear PCR are defined by a site at 5′end and a specific primer at the 3′end. The 3′ region of the CDR fragments are complementary to the template sequence onto which the fragment are going to be annealed. The ss CDR fragments can, after gel purification, be annealed to the template sequence chosen for the template based ligation assembly. For a ligation assembly twice as much of each ss CDR fragment compared to the amount of template is needed.

*Single-stranded framework fragments*

The framework fragments have to be complementary to the template sequence. Twice as much of each ss framework fragments compared to the amount of template is needed for a template based ligation assembly. The framework fragments, HFR3 (i.e. heavy framework region 3), LFR1 (i.e. light framework region 1) and LFR3 (i.e. light framework region 3), can be ordered as oligonucleotides or amplified with PCR and strand separated using magnetic streptavidin beads (Dynal ASA, Norway; Dynabeads M-280). If the framework fragments are amplified by PCR they have to be bound to washed streptavidin beads according to the manufacturer’s protocol. Incubate for 30 min on a roller at room temperature (r.t. – i.e. 20-25°C) and wash away non-coupled framework fragments. To strand separate the coupled framework fragments add 40 μl 0.1 M newly made sodium hydroxide and incubate on a roller for 10 min at room temperature (r.t. – i.e. 20-25°C). Save the supernatant, that includes the wanted ss framework fragments, and ethanol precipitate by adding 2.5 times 99%
ethanol and 10% 3 M sodium acetate (pH 5.2). Store at -20°C overnight and pellet the precipitations by centrifuging twice, including a wash step with 70% ethanol. The pellets are dissolved in an appropriate volume of water. The framework fragments are now ready for the ligation assembly.

Annealing of CDR fragments onto a template sequence

The template has a sequence similar to the light or heavy variable region of an antibody, in our case the DPL3 and DP-47 sequences. These two template sequences, have to be made with one identical restriction site located in each CDR. The template sequence also have to have two different sites for restriction enzymes in the constant framework at the 5′ and 3′ end, so it can be ligated into a vector.

Strand separation of template sequence

The double-stranded (ds) template DNA sequence is coupled onto streptavidin coated magnetic beads (Dynabeads M-280 from Dynal), washed and strand separated with sodium hydroxide (see Dynal instructions). To 10 pmole ds template, 100µl streptavidin coated magnetic beads (Dynabeads M-280 from Dynal) were used. Wash the beads according to vendor instructions and bind the template sequence onto the beads in 1x wash buffer (see Dynal instructions) for 30 min on a roller at room temperature (r.t. – i.e. 20-25°C). Non-bound template was washed away with wash buffer and 40µl 0.1 M freshly made sodium hydroxide was added to the bound ds template. Let the strand separation proceed for 10 min on a roller at r.t (i.e. 20-25°C). The unbound strand can now be washed away leaving the bound single stranded template sequence ready for annealing of CDR and framework fragments, i.e. the template based ligation assembly.

Template based ligation assembly with T4 DNA ligase
Wash the ss template sequence coupled to streptavidin coated magnetic beads with 1x T4 DNA ligase buffer (Invitrogen Ltd, Paisley, UK) and heat to 65°C. Heat the ss CDR and framework fragments, 20 pmole from each (see previous sections) to 95°C.

Mix heated template sequence with heated ss framework and ss CDR fragments. Allow to cool down slowly from 65°C to room temperature (r.t. – i.e. 20-25°C) allowing the fragments to anneal onto the coupled ss template sequence. The CDR and framework fragments are designed to anneal back to back on the template sequence. To the hybrid DNA formed add 0.1 U/µl T4 DNA ligase (Invitrogen Ltd, Paisley, UK) and incubate at 24°C for 60 min on a roller. Non-ligated fragments are washed off. The hybrid DNA is now ready to be digested with the appropriate restriction enzymes at the 5’ and the 3’ end. The first digest is made on the non-coupled constant end of the hybrid DNA with 50U enzyme and in 1x T4 DNA ligase buffer. Incubate the 100µl digest at 1.5 hrs on a roller at suitable temperature for the restriction enzyme. Wash away the cut-off end fragments with 1x T4 DNA ligase buffer and add the second restriction enzyme to digest at the coupled constant end of the hybrid DNA. Incubate the 100µl digest in 1x T4 DNA ligase buffer and with 50U enzyme at 1.5 hrs on a roller at, a for the restriction site suitable, temperature. The digested hybrid DNA fragments are located in the supernatant and can, after heat inactivation at 65°C 20 min be ligated into a suitable vector and transformed into chemically or electro competent E. coli cells. Since plasmids replicate semi-conservatively the amplified DNA can be digested with the identical restriction sites located in each CDR, to remove plasmids harbouring the template sequence. The remaining circular DNA, which is the annealed fragments from the ligation assembly, is transformed into chemically or electro competent E. coli cells. The above procedure is made
for both $V_H$ and $V_L$. These two libraries will then be joined into one complete ScFv library.

*Template based ligation assembly with Pfu DNA ligase*

Wash the ss template sequence coupled to streptavidin coated magnetic beads with 1x Pfu DNA ligase buffer (Stratagene Europe, Amsterdam, the Netherlands) and heat to 65°C. Heat the ss CDR and framework fragments, 20 pmole from each (see previous sections) to 95°C. Mix heated template sequence with heated ss framework and ss CDR fragments. Also add 0.1U/μl of thermo-stable Pfu DNA ligase (Stratagene Europe, Amsterdam, the Netherlands). Allow to cool down slowly from 65°C to 50°C allowing the fragments to anneal onto the coupled ss template sequence. The CDR and framework fragments are designed to anneal back to back on the template sequence. Incubate the hybrid DNA formed for 60 min at 50°C. The thermo-stable Pfu DNA ligase will ligate at 50°C. After incubation let the hybrid DNA cool down to room temperature (r.t. – i.e. 20-25°C). Non-ligated fragments are washed off. The hybrid DNA is now ready to be digested with the appropriate restriction enzymes at the 5’ and the 3’ end. The first digest is made on the non-coupled constant end of the hybrid DNA with 50U enzyme and in 1x T4 DNA ligase buffer. Incubate the 100μl digest at 1.5 hrs on a roller at suitable temperature for the restriction enzyme. Wash away the cut-off end fragments with 1x T4 DNA ligase buffer and add the second restriction enzyme to digest at the coupled constant end of the hybrid DNA. Incubate the 100μl digest in 1x T4 DNA ligase buffer and with 50U enzyme at 1.5 hrs on a roller at, a for the restriction site suitable, temperature. The digested hybrid DNA fragments are located in the supernatant and can, after chloroform extraction be ligated into a suitable vector and transformed into chemically or electro competent *E. coli* cells. Since plasmids replicate semi-conservatively the amplified
DNA can be digested with the identical restriction sites located in each CDR, to remove plasmids harbouring the template sequence. The remaining circular DNA, which is the annealed fragments from the ligation assembly, is transformed into chemically or electro competent *E. coli* cells. The above procedure is made for both *V*<sub>H</sub> and *V*<sub>L</sub>. These two libraries will then be joined into one complete ScFv library.

**Results**

*Light chain*

A small library was constructed using CDRs from 5 scFv clones. The length of all CDRL1 and CDRL2 were 14 respectively 7 amino acids. The length of CDRL3 varied between 12-14 amino acids. The template used had 12 amino acids in CDRL3.

**ScFv clones used in the V<sub>L</sub> library**

<table>
<thead>
<tr>
<th>Clone</th>
<th>Amino acids in CDRL3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT17</td>
<td>14</td>
</tr>
<tr>
<td>PSA21</td>
<td>13</td>
</tr>
<tr>
<td>LeY7</td>
<td>12</td>
</tr>
<tr>
<td>Muc46</td>
<td>12</td>
</tr>
<tr>
<td>Muc34</td>
<td>12</td>
</tr>
</tbody>
</table>

The resulting library hosted a mix of CDRs from all 5 scFv clones. Clones with the CDRL3 length of 12-13 amino acids were represented.

**Example of clones from the V<sub>L</sub> library**
<table>
<thead>
<tr>
<th>Clone</th>
<th>CDRL1</th>
<th>CDRL2</th>
<th>CDRL3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PSA21</td>
<td>Muc34/LeY7/CT17</td>
<td>LeY7</td>
</tr>
<tr>
<td>11</td>
<td>Muc46</td>
<td>Muc46</td>
<td>LeY7</td>
</tr>
<tr>
<td>12</td>
<td>PSA21</td>
<td>Muc34/LeY7/CT17</td>
<td>Muc46</td>
</tr>
<tr>
<td>30</td>
<td>CT17</td>
<td>PSA21</td>
<td>Muc34</td>
</tr>
<tr>
<td>37</td>
<td>PSA21</td>
<td>Muc34/LeY7/CT17</td>
<td>LeY7</td>
</tr>
<tr>
<td>41</td>
<td>Muc46</td>
<td>Muc34/LeY7/CT17</td>
<td>PSA21</td>
</tr>
<tr>
<td>43</td>
<td>Muc46</td>
<td>Muc34/LeY7/CT17</td>
<td>LeY7</td>
</tr>
</tbody>
</table>

CDRL2 is identical from the clones Muc34, LeY7 and CT17

*Heavy chain library*

A small library was constructed using CDRH3 from 5 scFv clones. CDRH1 and CDRH2 were constant and originated from Muc46. The length of CDRH3 varied between 6-22 amino acids. The template used had 10 amino acids in CDRH3.

In the resulting library clones with the CDRH3 length of 6-22 amino acids were represented.

Figure 2 summarizes 23 clones from the V_H library experiment.

*Heavy chain mRNA from human lymphoid tissue*

To further evaluate the system a library with CDRs prepared from lymphoid tissue was also constructed. The resulting library contained CDRs from different germline families with somatic mutations. The length of all CDRH1 were 7 amino acids, CDRH2 varied between 19-20 amino acids and CDRH3 varied between 8-20 amino acids, compare to template DNA CDRH1 7 amino acids, CDRH2 20 amino acids and CDRH3 10 amino
acids. 58 clones were evaluated and the length of CDRH3 is presented in Figure 3. Out of 58 investigated sequences of CDRH3 57 were unique.

When using a template with only 6 amino acids in CDRH3 a similar pattern can be seen in Figure 4. We investigated CDRH3 from 42 clones and CDRH3 varied between 9-20 amino acids, originated from different germ line families with somatic mutations — that is, the template allows a great variation to be implemented into the library.

REFERENCES

Ausubel et al. (1995) or Protocols Online URL: www.protocol-online.net/molbio/index.html.
Berent et al., 1985, Biotech, 3:208.
Engberg et al., 1996 Molecular Biotechnology 6, 287-310.
Marks et al., (1992), Biotechnology, 10:779-783.
Sherman et al., (1986), *Methods In Yeast Genetics, A Laboratory Manual,*
Cold Spring Harbor, NY.
US 6,132,970
WO 02/086121 A1
WO 98/16643
CLAIMS

1. A method for assembling a composite polynucleotide sequence comprising the steps:

   a) providing a single-stranded template polynucleotide sequence comprising at least one invariant region and at least one variant region;

   b) providing a plurality of single-stranded invariant polynucleotide sequences that are at least partially complementary to the at least one invariant region of the template polynucleotide sequence, and a plurality of single-stranded variant polynucleotide sequences that are at least partially complementary to the at least one variant region of the template polynucleotide sequence,

   wherein the single-stranded invariant and variant polynucleotides, when ligated together, are capable of forming a polynucleotide which is at least partially complementary to the template polynucleotide;

   c) annealing the single-stranded polynucleotide sequences provided in steps (a) and (b) to generate a double-stranded polynucleotide sequence wherein one strand is the template polynucleotide sequence, and one strand is composed of at least one invariant polynucleotide sequence and at least one variant polynucleotide sequence that are not ligated together;

   d) subjecting the double-stranded polynucleotide sequence generated in step (c) to conditions that ligate the at least one invariant polynucleotide sequence and the at least one variant
polynucleotide sequence together to generate a double-stranded hybrid polynucleotide sequence composed of one strand that is the template polynucleotide sequence, and one strand that is a composite polynucleotide sequence comprising at least one invariant polynucleotide sequence and at least one variant polynucleotide sequence;

e) separating the composite polynucleotide sequence generated in step (d) from the template polynucleotide sequence.

2. A method according to Claim 1 further comprising the following step performed after step (e):

f) introducing the composite polynucleotide sequence in step (e) or its complement into a vector.

3. A method according to Claim 1 or 2 further comprising the step of expressing the composite polynucleotide sequence generated in step (d), (e) or (f) or its complement to generate a corresponding polypeptide.

4. A method according to Claim 3 further comprising the step of testing the polypeptide for desired characteristics.

5. A method for producing a library of different composite polynucleotide sequences comprising the steps:

a) providing one or more single-stranded template polynucleotide sequence comprising at least one invariant region and at least one variant region;
b) providing a plurality of different single-stranded invariant polynucleotide sequences that are at least partially complementary to the at least one invariant region of the one or more template polynucleotide sequence, and a plurality of different single-stranded variant polynucleotide sequences that are at least partially complementary to the at least one variant region of the one or more template polynucleotide sequence,

wherein the invariant and variant polynucleotides, when ligated together, are capable of forming a single-stranded polynucleotide which is at least partially complementary to the one or more template polynucleotide;

c) annealing the single-stranded polynucleotide sequences provided in steps (a) and (b) to generate a plurality of double-stranded polynucleotide sequences wherein each of the plurality of double-stranded polynucleotide sequences comprises one strand that is a template polynucleotide sequence, and one strand that is composed of at least one invariant polynucleotide sequence and at least one variant polynucleotide sequence that are not ligated together;

d) subjecting the plurality of double-stranded polynucleotide sequences generated in step (c) to conditions that ligate the at least one invariant polynucleotide sequence and the at least one variant polynucleotide sequence together to generate a plurality of double-stranded hybrid polynucleotide sequences, wherein each of the plurality of double-stranded hybrid polynucleotide sequences is composed of one strand that is a template polynucleotide sequence, and one strand that is a different composite polynucleotide sequence comprising at least one
invariant polynucleotide sequence and at least one variant polynucleotide sequence;

e) separating a plurality of different composite polynucleotide sequences generated in step (d) from the plurality of template polynucleotide sequences.

6. A method according to Claim 5 further comprising the following step performed after step (e):

f) introducing the plurality of different composite polynucleotide sequences generated in step (e) or their complements into a plurality of vectors.

7. A method according to Claim 5 or 6 further comprising the step of expressing the plurality of different composite polynucleotide sequences generated in steps (d), (e) or (f) or their complements to generate a corresponding library of polypeptides.

8. A method according to Claim 7 further comprising the step of screening the library of polypeptides for desired characteristics and selecting a polypeptide with desired characteristics.

9. A method according to Claim 8 further comprising the step of selecting the polynucleotide sequence encoding the polypeptide with desired characteristics.

10. A method according to Claim 9 further comprising the step of introducing the polynucleotide or part thereof into a vector.
11. A method according any one of Claims 2, 6 or 11 wherein the vector is an expression vector.

12. A method according to Claim 11 wherein the expression vector is pFab51.His or pFAB60.

13. A method according to any of Claims 9 to 12 further comprising the step of expressing a polypeptide encoded at least in part by the selected polynucleotide or part thereof.

14. A method according to any one of Claims 3, 4, 7, 8 or 11 further comprising the step of formulating the polypeptide into a pharmaceutical composition.

15. A method according to any preceding claim wherein the template polynucleotide sequence encodes (or is complementary to a polynucleotide encoding) an immunoglobulin or a fragment or domain thereof.

16. A method according to Claim 15 wherein the immunoglobulin is an antibody.

17. A method according to Claim 15 or 16 wherein the template polynucleotide sequence or its complement encodes an immunoglobulin variable domain.

18. A method according to any one of Claims 15 to 17 wherein the immunoglobulin variable domain is a light chain immunoglobulin variable domain (V_L).
19. A method according to Claim 18 wherein the light chain immunoglobulin variable domain (\(V_L\)) is DPL3.

20. A method according to any one of Claims 15 to 17 wherein the immunoglobulin variable domain is a heavy chain immunoglobulin variable domain (\(V_H\)).

21. A method according to Claim 20 wherein the heavy chain immunoglobulin variable domain (\(V_H\)) is DP-47.

22. A method according to any preceding claim wherein the template sequence comprises three variant regions and four invariant regions.

23. A method according to any preceding claim wherein the variant regions encode (or are complementary to sequences encoding) immunoglobulin complementarity-determining regions (CDRs) and the invariant regions encode (or are complementary to polypeptides encoding) immunoglobulin framework regions (FRs).

24. A method according to Claim 23 wherein the immunoglobulin framework regions are from the DP-47 heavy chain immunoglobulin variable domain (\(V_H\)) or the DPL-3 light chain immunoglobulin variable domain (\(V_L\)).

25. A method according to any one of Claims 1 to 24 wherein the single-stranded invariant polynucleotide sequences are obtained from sequences encoding (or complementary to sequences encoding) immunoglobulin framework regions (FRs) and the single-stranded variant polynucleotide sequences are obtained from sequences encoding (or complementary sequences encoding) immunoglobulin complementarity-determining regions (CDRs).
26. A method according to any one of Claims 1 to 24 wherein the single-stranded invariant polynucleotide sequences are obtained from sequences encoding (or complementary to sequences encoding) immunoglobulin framework regions (FRs) and the single-stranded variant polynucleotide sequences are oligonucleotides of random sequence.

27. A method according to any preceding claim wherein the sequences encoding (or complementary to sequences encoding) immunoglobulin framework regions (FRs) and/or immunoglobulin complementarity-determining regions (CDRs) are obtained from a library of different polynucleotide sequences encoding (or complementary to sequences encoding) immunoglobulins with different binding specificities.

28. A method according to any preceding claim wherein the sequences encoding (or complementary to sequences encoding) immunoglobulin framework regions (FRs) and/or immunoglobulin complementarity-determining regions (CDRs) are obtained by mutagenising a polynucleotide encoding (or complementary to a polynucleotide encoding) an immunoglobulin.

29. A method according to Claim 28 wherein mutagenising the polynucleotide encoding an immunoglobulin involves a polymerase.

30. A method according to any preceding claim wherein ligating the at least one invariant single-stranded polynucleotide sequence and the at least one variant single-stranded polynucleotide sequence together in step (d) is performed using DNA ligase.
31. A method according to any preceding claim further comprising step (d2) performed after step (d) but before step (e) of separating the double-stranded hybrid polynucleotide sequence generated in step (d) from single-stranded invariant polynucleotide sequences and/or single-stranded variant polynucleotide sequences that are not annealed to the template polynucleotide sequence.

32. A method according to Claim 31 wherein the single-stranded template polynucleotide sequence is linked to a member of a specific binding pair (MSBP) and step (d2) comprises binding the MSBP to its specific binding partner.

33. A method according to Claim 32 wherein the MSBP is biotin and its specific binding partner is streptavidin.

34. A method according to any preceding claim wherein the template polynucleotide sequence is more sensitive to degradation than the composite polynucleotide sequence and step (e) comprises the step of subjecting the double-stranded hybrid polynucleotide sequence generated in step (d) to conditions capable of degrading the template polynucleotide sequence but not the composite polynucleotide sequence.

35. A method according to Claim 34 wherein thymidine in the template polynucleotide sequence has been replaced with uracil.

36. A method according to Claim 35 wherein the template polynucleotide is generated using a bacterial strain that is $duf^-$ and $ung^-$. 
37. A method according to any one of Claims 34 to 36 wherein the conditions capable of degrading the template polynucleotide sequence but not the composite polynucleotide sequence comprise transforming the double-stranded hybrid polynucleotide sequence into a bacterial cell capable of replicating the composite polynucleotide sequence but not the template polynucleotide sequence.

38. A method according to any one of Claims 1 to 33 wherein step (e) comprises the steps:

i) ligating the double-stranded hybrid polynucleotide sequence generated in step (d) or (d2) into a vector;

ii) introducing the vector containing the double-stranded hybrid polynucleotide sequence into a cell capable of replicating the vector containing the double-stranded hybrid polynucleotide sequence;

iii) isolating a plurality of vectors from the cell progeny comprising one or more vectors containing the template polynucleotide sequence and one or more vectors containing the composite polynucleotide sequence;

iv) subjecting the plurality of vectors to conditions that fragment the one or more vectors containing the template polynucleotide sequence but that do not fragment the one or more vectors containing the composite polynucleotide sequence;
v) separating the one or more vectors containing the composite
polynucleotide sequence from the fragmented one or more
vectors containing the template polynucleotide sequence.

39. A method according to Claim 38 wherein the one or more vectors
containing the template polynucleotide sequence comprise a
recognition site for a restriction endonuclease that is not present in
the one or more vectors containing the composite polynucleotide
sequence and step (v) comprises treating the plurality of vectors with
a restriction endonuclease specific for said recognition site.

40. A method according to Claim 38 or 39 wherein step (vi) comprises
introducing the one or more vectors containing the composite
polynucleotide sequence and the fragmented one or more vectors
containing the template polynucleotide sequence into a cell capable
of replicating the one or more vector containing the composite
polynucleotide sequence but not capable of replicating the
fragmented one or more vectors containing the template
polynucleotide sequence.

41. A method according to Claim 38 or 40 wherein the cell is a bacterial
cell.

42. A method according to Claim 41 wherein the bacterial cell is
*Escherichia coli*.

43. A polynucleotide sequence obtained by the method of any one of
Claims 1, 2, 9, 10 to 12, or 15 to 42.

44. A polypeptide obtained by the method of any one of Claims 3, 8 or
13.
45. A library of different composite polynucleotide sequences obtained by the method of Claim 5 or 6.

46. A library of different composite polypeptides obtained by the method of Claim 7.

47. A pharmaceutical preparation obtained by the method of Claim 14.
Fig. 1A) Annealing of single-stranded oligonucleotide fragments onto single-stranded template polynucleotide molecules. Pools of single-stranded oligonucleotide fragments containing CDR variability (multi-coloured) or constituting framework regions with no variability (black) are allowed to hybridise to single-stranded template DNA (red) encoding a specific gene or part of a gene. The fragments will base pair with complementary bases and differences in length will be accommodated by looping out of either the template DNA or the oligonucleotide fragments.
Principle of Template-Based Ligation Assembly

1. Anneal fragments on template coupled to paramagnetic beads.
2. Ligate fragments with T4 DNA ligase. Trim ends with restriction enzyme and ligate hybrid DNA into vector.
3. Transform into E.coli.
4. Plasmid purification. Digest with template-unique restriction enzyme.
5. Transform again into E.coli → New Library.
Fig. 2) Construction of a small VH library. An assembly reaction experiment was done, using CDRH3 single-stranded fragments from five antibody scFv clones with different sequences and lengths (6-22 aa) as a pool, while keeping the CDRH1 and CDRH2 single-stranded fragments constant. The template fragment used, had a 10 aa CDRH3.
Fig 3) Distribution of CDRH3 length in a VH library with CDRs from lymphoid tissue. An assembly reaction was done with single-stranded DNA fragments encoding CDRH1, CDRH2 and CDRH3 regions respectively, isolated from immunoglobulin heavy chain mRNA from lymphoid tissue. The length of the template CDRH3 is 10 aa.
Fig. 4) Distribution of CDRH3 length in a VH library with CDRs from lymphoid tissue. An assembly reaction was done as in Fig. 3 with single-stranded DNA fragments encoding CDR regions, isolated from immunoglobulin heavy chain mRNA from lymphoid tissue. The length of the template CDRH3 is only 6 aa.
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N15/10 C40B50/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N C40B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the International search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, CAB Data, Sequence Search, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>WO 01/29211 A (ENCHIRA BIOTECHNOLOGY CORPORATION; COCO, WAYNE, M; ARLNSDORF, JOSEPH,) 26 April 2001 (2001-04-26)</td>
<td>1-14, 30-47</td>
</tr>
<tr>
<td>Y</td>
<td>the whole document</td>
<td>15-29</td>
</tr>
<tr>
<td>Y</td>
<td>the whole document</td>
<td>15-29</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C.

Special categories of cited documents:

*A* document defining the general state of the art which is not considered to be of particular relevance

*E* earlier document but published on or after the international filing date

*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

*O* document referring to an oral disclosure, use, exhibition or other means

*P* document published prior to the international filing date but later than the priority date claimed

**F** later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

**Y** document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, each combination being obvious to a person skilled in the art

"A" document member of the same patent family

Date of the actual completion of the international search: 13 March 2006

Date of mailing of the international search report: 22/03/2006

Name and mailing address of the ISA:

European Patent Office, P.B. 5513 Patentlaan 2 NL – 2280 HV Glossip, Tel.: (+31-70) 340-2040, Fax: (+31-70) 340-2016

Authorized officer: Hornig, H.
## INTERNATIONAL SEARCH REPORT

### DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category*</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>WO 01/75091 A (BIOINVENT INTERNATIONAL AB; OHLIN, MATS; SODERLING, ESKIL; CARLSSON, R) 11 October 2001 (2001-10-11) the whole document</td>
<td>15-29</td>
</tr>
<tr>
<td>Y</td>
<td>WO 02/086121 A (PROTEUS; DUPRET, DANIEL; LEFEVRE, FABRICE; MASSON, JEAN-MICHEL) 31 October 2002 (2002-10-31) the whole document</td>
<td>1-14, 15-29</td>
</tr>
<tr>
<td>X</td>
<td>WO 00/09679 A (PROTEUS; DUPRET, DANIEL; MASSON, JEAN-MICHEL; LEFEVRE, FABRICE) 24 February 2000 (2000-02-24) the whole document</td>
<td>1-14, 15-29</td>
</tr>
<tr>
<td>X</td>
<td>WO 98/15567 A (VITALITY BIOTECHNOLOGIES LTD; STRIZHOV, NICOLAI; KONCZ, CSABA; SCHELL,) 16 April 1998 (1998-04-16) the whole document</td>
<td>1, 3, 4</td>
</tr>
<tr>
<td>X</td>
<td>EP 1 108 783 A (MAXYGEN, INC) 20 June 2001 (2001-06-20) the whole document</td>
<td>1-14, 15-29</td>
</tr>
<tr>
<td>X</td>
<td>WO 01/64864 A (MAXYGEN, INC; AFFHOLTER, JOSEPH, A; COX, ANTHONY; NESS, JON, E; CARR,) 7 September 2001 (2001-09-07) the whole document</td>
<td>1-14, 43-47</td>
</tr>
<tr>
<td>A</td>
<td>WO 02/06469 A (ENCHIRA BIOTECHNOLOGY CORPORATION; COCO, WAYNE, M; ENCELL, LANCE, P; A) 24 January 2002 (2002-01-24) the whole document</td>
<td>1-14</td>
</tr>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 5225301 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2402240 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DE 60103988 D1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DE 60103988 T2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1268801 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2003529368 T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NO 20024723 A</td>
</tr>
<tr>
<td>US 2004191772</td>
<td>A1 30-09-2004</td>
<td>NONE</td>
</tr>
<tr>
<td>WO 02086121</td>
<td>A 31-10-2002</td>
<td>CA 2445258 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1381680 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2004531258 T</td>
</tr>
<tr>
<td>WO 0009679</td>
<td>A 24-02-2000</td>
<td>AT 309342 T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 769516 B2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 5171799 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2339901 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DE 69928263 D1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1104457 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FR 2782323 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2002538763 T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NO 20015787 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DE 69711667 D1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 0948626 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 0935603 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 9815630 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NZ 335358 A</td>
</tr>
<tr>
<td>EP 1352959</td>
<td>A 15-10-2003</td>
<td>NONE</td>
</tr>
<tr>
<td>WO 0164864</td>
<td>A 07-09-2001</td>
<td>AU 4193901 A</td>
</tr>
<tr>
<td>WO 0206469</td>
<td>A 24-01-2002</td>
<td>AU 7355901 A</td>
</tr>
</tbody>
</table>