METHOD FOR LINKING SEQUENCES OF INTEREST

Technical Field of the Invention

[0001] The present invention relates to a multiplex molecular amplification procedure capable of linking nucleotide sequences of interest in connection with the amplification, in particular polymerase chain reaction (multiplex PCR). The method is particularly advantageous for the generation of cognate pair libraries as well as combinatorial libraries of variable region encoding sequences from immunoglobulins, T cell receptors or B cell receptors.

Background of the Invention

[0002] Antigen binding proteins involved in the immune response are present in mammals as large polyclonal repertoires representing a broad diversity of binding specificities. This diversity is generated by rearrangement of gene sequences encoding variable regions of these binding proteins. Such variable region binding proteins include soluble and membrane-bound forms of the B cell receptor (also known as immunoglobulins or antibodies) and the membrane-bound T cell receptors (TcR). With respect to immunoglobulins, their affinity is enhanced subsequent to recognition of an antigen by a B cell antigen receptor, through a process termed affinity maturation which involves cycles of somatic hypermutation of these variable genes.

[0003] Notably, immunoglobulins or fragments thereof, such as Fab fragments, Fv fragments and single chain Fv (scFv) molecules have been subject to cloning and recombinant expression. However, all other variable region binding proteins can in principle also be cloned and expressed using the same concepts as for antibodies.

[0004] Known approaches for isolating antibodies with a desired binding specificity most often involves generation of hybridomas from immunized hosts followed by screening for specific clones or involves the generation of combinatorial expression libraries in *E. coli* composed of immunoglobulin variable domains, which are subsequently enriched using techniques such as, for example, phage display.

[0005] The main restriction in the use of the hybridoma technology for making therapeutic antibodies is the absence of a human lymphoma suitable as fusion partners for human B lymphocytes. Heterohybridomas (i.e., fusion of human B cells with mouse lymphomas) are notoriously unstable and thus rarely lead to suitable cell lines for production purposes.

Human B cells immortalized through infection with Epstein-Barr virus exhibit similar challenges of instability. The lack of robust cellular methodology for making human antibodies for therapy can be compensated with more recent advantages in molecular biology.

[0006] The use of combinatorial libraries and phage display allows for generation of large repertoires of antibody clones with a potential diversity in excess of 10^{10}. From this repertoire selection for binding to a specific target can be performed thereby generating a sub-library. This sub-library can be used to generate either polyclonal or monoclonal antibodies. The variable region encoding sequences (for example immunoglobulin heavy chain variable region and light chain variable region encoding sequences) which constitute the library can be amplified from lymphocytes, plasma cells, hybridomas or any other immunoglobulin expressing population of cells. Current technologies for generating combinatorial libraries involve separate isolation of the variable region encoding sequences from a population of cells. Thus, the original pairing of for example immunoglobulin heavy chain variable region and light chain variable region encoding sequences will be lost. Rather, in a combinatorial library said sequences are randomly paired and the original combinations of these variable sequences will only occur by chance. Thus, in order to isolate variable region encoding sequences responsible for a desired binding specificity, a considerable amount of screening is necessary. This is typically performed in combination with methods for enrichment of clones exhibiting a desired specificity, such as ribosome display or phage display. Even then, the diversity achieved might not be sufficiently large to isolate variable region encoding sequence pairs giving rise to binding proteins of similar high affinity as those found in the original cells. Further, the enrichment procedures normally used to screen combinatorial libraries introduce a strong bias e.g. for polypeptides of particular low toxicity in *E. coli*, efficient folding, slow off-rates, or other system dependent parameters, that reduce the diversity of the library even further. In addition, clones derived from such combinatorial libraries will be more prone to produce binding proteins with cross-reactivity against self-antigens because they as pairs, in contrast to original pairs (hereafter called cognate pairs), never have been through *in vivo* negative selection against self-antigens, such as it is the case for B and T lymphocyte receptors during particular stages of their development. Therefore, the cloning of original pairs of variable region encoding sequences is a desirable approach. Moreover, the frequency of clones exhibiting a desired binding specificity is expected to be considerably higher within a library of cognate pairs, than in a conventional combinatorial

library, particularly if the starting material cells are derived from a donor with high frequency of cells encoding specific binding pairs e.g. immune or immunized donors. It follows that the size of a cognate pair library will not need to be as large as a combinatorial library: a cognate pair library size of 10^4 to 10^5 clones or even as small as 10^2 to 10^3 clones derived from a donor with a relevant ongoing immune response might very well suffice in order to obtain binding proteins representing a broad diversity of desired binding specificities.

[0007] In order to generate cognate pair libraries the linkage of the variable region encoding sequences derived from the same cell is required. At present, two different approaches which can achieve cognate pairing of variable region encoding sequences have been described.

[0008] In-cell PCR is an approach where a population of cells is fixed and permeabilized, followed by in-cell linkage of heavy chain variable region and light chain variable region encoding sequences from immunoglobulins. This linkage can be performed either by overlap-extension RT-PCR (WO 93/03151) or by recombination (Chapal, N. *et al.* 1997 BioTechniques 23, 518-524). The amplification process as described in these publications is a three or four step process consisting of i) reverse transcription utilizing constant region primers generating immunoglobulin cDNA, ii) PCR amplification of the heavy and light chain variable region encoding sequences utilizing primer sets containing either overlap-extension design or recombination sites, iii) linkage by recombination, if this approach is chosen, iv) nested PCR of the products generating restriction sites for cloning. Since the cells are permeabilized there is a considerable risk that amplification products might leak out of the cells, thereby introducing scrambling of the heavy chain variable region and light chain variable region encoding sequences, resulting in the loss of cognate pairing. Therefore, the procedure includes washing steps after each reaction which makes the process laborious and reduces the efficiency of the reactions.

[0009] More generally, the in-cell PCR is notoriously inefficient, resulting in low sensitivity. Accordingly, the in-cell PCR linkage technique has never found widespread usage, and the original study has in fact never been reliably repeated in a way which can be used to verify that the linkage actually occurs within the cell. This, however, is absolutely crucial to avoid scrambling of the heavy chain variable region and light chain variable region encoding sequences and thereby disrupting the cognate pairs.

[0010] A different in-cell approach is described in WO 01/92291. This approach is based on RNA trans-splicing, and achieves joining of V_H and V_L encoding mRNA within the cell. This approach requires the presence of a DNA construct driving the trans-splicing within the cells.

[0011] Single-cell PCR is a different approach to achieve cognate pairing of heavy chain variable region and light chain variable region encoding sequences (see, for example, Coronella, J.A. *et al.* 2000 Nucleic Acids Res. 28, E85; Wang, X., *et al.* 2000 J. Immunol. Methods 20, 217-225). In these publications a population of immunoglobulin expressing cells are distributed by diluting to a density of one cell per reaction, thereby eliminating scrambling of heavy chain variable region and light chain variable region encoding sequences during the cloning process. Basically, the process described is a three to four step procedure consisting of i) reverse transcription utilizing oligo-dT-, random hexamer- or constant region primers generating cDNA, ii) fractionating the cDNA product into several tubes and performing PCR amplification on the individual variable chain encoding sequences (in separate tubes) with primer sets containing restriction sites for cloning, iii) nested PCR of the products generating restriction sites for cloning (optionally) and iv) linking the heavy chain variable region and light chain variable region encoding sequences from the separate tubes by cloning them into an appropriate vector, which in itself is a multi-step process.

[0012] In humans there are two types of light chains: lambda (λ) and kappa (κ). This means that with the cDNA generated from every single cell at least three separate PCR reactions must be performed followed by analysis and cloning of the appropriate fragments into a single vector to achieve the cognate pairing. Thus, the single-cell PCR approach as described requires a large number of manipulations to generate a library of cognate pairs. Although, a cognate pair library does not need to be as large as a combinatorial library in order to obtain binding proteins representing a broad diversity of binding specificities it would still be a laborious task to generate a library of for example 10^4 to 10^5 clones by the described single-cell PCR approach. Further, the large number of manipulations highly increases the risk of contamination and human error.

[0013] In order to obtain high affinity binding proteins corresponding to the affinities normally observed during an immune response, cognate pairing of the variable region sequences in association with their amplification is highly advantageous. To generate a library of large diversity it is necessary to have a cloning technique that can be fitted to a high-throughput format and where the risk of contamination and scrambling is minimal.

[0014] A reduction of the number of cloning steps allowing the generation of combinatorial libraries to be fitted to a high-throughput format, is likewise desired.

Disclosure of Contribution

[0015] The present invention provides an efficient method of linking two or more nucleotide sequences of interest, e.g. variable region encoding sequences, utilizing a multiplex molecular amplification procedure, such as multiplex overlap-extension RT-PCR or multiplex RT-PCR followed by linkage by ligation or recombination. The method is applicable on a single cell, thereby enabling cloning of cognate pairs in a high-throughput format.

Description of the Figures

[0016] FIGURE 1 is a diagram illustrating the different types of overlap-extension tails. The bold line corresponds to a gene-specific part of the primer and the regular line corresponds to the overlapping tail. The vertical bars illustrate complementary regions. The primers facilitate the linkage of two nucleotide sequences of interest. FIGURE 1 (I) illustrates two varieties of the type I overlap-extension tails where only the extension tails overlap, either completely or partially; FIGURE 1 (II) illustrates the type II overlap-extension tails where some of the 5'-nucleotides of the first primer extension tail are complementary to the gene-specific part of the neighboring primer; FIGURE 1 (III) illustrates the type III overlap-extension tails where the entire overlap-extension tails are complementary to the gene-specific region of the neighboring primer.

[0017] FIGURE 2 is a diagram showing a schematic overview of a multiplex overlap-extension primer mix applicable in the linkage of immunoglobulin variable region encoding sequences. The cDNA encoding the light chain (LC) and the heavy chain variable region (V_H) to be linked are illustrated as tubes with indication of their sense strand 5' and 3'ends as well as the expected size of the amplified product. The multiplex overlap-extension primer sets used to amplify the encoding sequence are illustrated by the arrows. The bended arrow tails with dashed 5' overhangs illustrate the cloning tails. The overlap-extension tails are in bold. The restriction sites present in the tails are named in connection with the tail. The total number of primers in the multiplex overlap-extension primer mix is sixteen, distributed on the outer primers comprising one C_κ and one C_{H1} primer and the overlap extension primers comprising six V_L and eight V_H primers. The C_{H1} primer anneal at the 5' end of the heavy chain constant domain 1. The product resulting from the multiplex overlap-extension RT-PCR is expected to be approximately 1070 bp constituting the entire kappa light chain composed of the constant region, joining gene and variable gene ($C_\kappa + J_L + V_L$) and the

heavy chain variable region, composed of the variable gene, diversity segment and joining gene ($V_H + D + J_H$). 5' and 3' indicates the direction of the open reading frame. Only a small portion of the C_{H1} region encoding sequence is amplified by this multiplex overlap-extension primer mix, because the annealing position of the C_{H1} primer is close to the heavy chain J-region.

[0018] FIGURE 3 is a series of diagrams illustrating the different linking direction of the products that can be obtained depending on which primers are equipped with the linkage tail. Solid black illustrates the overlap region. 5' and 3' indicates the direction of the open reading frame. FIGURE 3A is a diagram illustrating a head-to-head orientation of the products. FIGURE 3B is a diagram illustrating a tail-to-tail orientation. FIGURE 3C is a diagram illustrating a head-to-tail orientation with the light chain encoding sequence first. FIGURE 3D is a diagram illustrating a head-to-tail orientation with the heavy chain encoding sequence first.

[0019] FIGURE 4 is a schematic diagram of the immunoglobulin expression vector pLL113, where the encoding sequences are in a head-to-tail orientation. The vector comprises the following elements: bla = promoter allowing expression of the ampicillin resistance gene. Amp = gene encoding for ampicillin resistance. pUC ori = pUC origin of replication. AdMLP= Adenovirus major late promoter. Human IgG1 = Sequence encoding for immunoglobulin isotype G1 heavy chain. hGH pA = Human growth hormone poly A signal sequence. bGH polyA= Bovine Growth Hormone poly A sequence. Human kappa LC = Sequence encoding immunoglobulin kappa light chain. FRT = A Flp recognition target site. Hygromycin = gene encoding hygromycin resistance. SV40 poly A = Simian virus 40 poly A signal sequence.

[0020] FIGURE 5A is an electrophoretic gel showing the results from a two-step multiplex overlap-extension RT-PCR followed by a semi-nested PCR. The amplification products are derived from cDNA isolated from single CHO Flp-In pLL113 cells. Lanes 1-12 are sample lanes and the arrows indicate correct nested multiplex-overlap-extension RT-PCR products of 1076 bp; M1 is a 100 bp ladder. W is water used as negative control template. C is a positive cDNA control template derived from cell-line HB-8501. In a separate panel the same lanes W, C and the 100bp ladder have been depicted with less contrast in order to resolve the individual DNA fragments. FIGURE 5B is a sketch of the gel shown in Figure 5A, illustrating the relevant fragments from the gel.

[0021] FIGURE 6 is a series of photographs and a graphical representation of electrophoretic gels showing the results from a single-step multiplex overlap-extension RT-PCR reaction without additional PCR amplification. In each panel, M1 is a 100 bp ladder and M2 is a 500 bp ladder. FIGURE 6A is an electrophoretic gel showing the amplification products derived from lysate corresponding to 100, 10, 1 or 0 cells. The arrow indicates the overlap-extension product. FIGURE 6B is a sketch of the gel in Figure 6A. FIGURE 6C is an electrophoretic gel verifying the presence of overlap-extension product in the 100 and 1 cell lanes in Figure 6A. FIGURE 6D is an electrophoretic gel showing restriction enzyme cleavage with NheI and NcoI, respectively, of the overlap-extension product from the 1 cell lane in Figure 6C.

[0022] FIGURE 7 is an electrophoretic gel showing the results from a single-step multiplex overlap-extension RT-PCR followed by a semi-nested PCR amplification. M1 is a 100 bp ladder and M2 is a 500 bp ladder. Results from multiplex overlap-extension primer mixtures containing either C_{H1}, C_{H2}, C_{H3}, C_{H4} or C_{H5} as the outer primer in the multiplex overlap-extension RT-PCR reaction. The reactions were performed on cell lysates corresponding to 100, 10, 1 or 0 cells. The size of the overlap-extension product is indicated with an arrow.

[0023] FIGURE 8 is an electrophoretic gel showing the results from a single-step multiplex overlap-extension RT-PCR followed by a semi-nested PCR amplification using enriched human B lymphocytes as template. M1 is a 100 bp ladder. Lane 5 and 6 show bands of the size expected for the overlap-extension product.

[0024] FIGURE 9A is a schematic diagram of the mammalian expression vectors (Em465/01P582/Em465/01P581) used for generation of IgG1-lambda expressing cell lines, where the encoding sequences are in a head-to-head orientation. The vectors comprise the following elements: Amp = gene encoding for ampicillin resistance. pUC ori = pUC origin of replication. AdMLP= Adenovirus major late promoter. EFP=Elongation factor promoter. AP leader= alkaline phosphatase leader sequence. VH= heavy chain variable region encoding sequence. IgG1 HC = Sequence encoding for immunoglobulin isotype G1 heavy chain constant region. rBG polyA =Rabbit beta-globin poly A signal sequence. bGH polyA= Bovine Growth Hormone poly A sequence. IgK leader= sequence encoding for murine kappa leader. IgL (1b or 1c) = Sequence encoding immunoglobulin lambda light chain family 1b or 1c. FRT = A Flp recognition target site. Hygromycin = gene encoding hygromycin resistance. SV40 poly A = Simian virus 40poly A signal sequence. FIGURES 9B and 9C are ethidium bromide stained agarose gels loaded with PCR products isolated form cell line CHO Flp-In/Em464/01P581 and CHO Flp-In/Em464/01P582, respectively. Lanes 1 to 4

correspond to total RNA template concentrations of 50 pg, 5 pg, 0.5 pg or 0 pg used for the multiplex overlap-extension RT-PCR reaction. M is a 100 bp ladder (New England Biolabs, New England, USA). The arrows indicate the overlap-extension PCR product.

[0025] FIGURE 10 is a flow chart showing the steps applied in order to generate a cognate antibody expression library from which polyclonal or monoclonal antibodies can be expressed.

[0026] FIGURE 11 is a schematic diagram of JSK301, an *E. coli* vector used to generate a library of Fab vectors by inserting the overlap-extension fragments comprising the cognate variable region encoding sequences into the vector at the indicated NotI/ XhoI restriction sites. The vector comprises the following elements: Amp and Amp pro =ampicillin resistance gene and its promoter. pUC19 Ori=origin of replication. Human CH1=sequence encoding human immunoglobulin gamma 1 heavy chain domain 1. Stuffer =an irrelevant sequence insert which is cut out upon insertion of the overlap-extension fragments. tac P and lac Z = bacterial promoters which can be excised at the NheI and AscI restriction sites.

[0027] FIGURE 12 is a schematic diagram illustrating the generation of a library of cognate Fab expression vectors. Step I illustrates the insertion of cognate pairs of variable region encoding sequences (VH_1-VL_1 to VHx-VLx) into *E. coli* vector JSK301 by XhoI-NotI digestion. Step II illustrates the insertion of a bacterial promoter and leader cassette (*pel*B leader-P *tac*-promoter driving the expression of VHx and P *lac* promoter- *pel*B leader driving the expression of VLx) by AscI-NheI digestion.

[0028] FIGURE 13 illustrate the linkage of α-, β- and γ-subunits constituting a G-protein, utilizing single-step multiplex overlap-extension RT-PCR followed by an additional PCR amplification. The sizes of the individual coding regions are given as well as the size of the linked product. Restriction sites introduced by the primer tails during amplification are indicated for the final product.

[0029] FIGURE 14 shows dot plots of an analytic FACS staining of (A) PBMC purified from donor blood; (B) the magnetically sorted non-labelled CD19 negative cell fraction and (C) magnetically sorted CD19+ cells fraction. A scatter plot, a CD19/CD38 plot, and a CD38/CD45 plot is shown for each fraction.

[0030] FIGURE 15 shows the CD19+ fraction from Figure 9 C, which had been stored in liquid nitrogen, thawed and stained with anti-CD19, anti-CD38 and anti-CD45. Dot plots corresponding to Figure 9C are shown.

[0031] FIGURE 16 shows gates used for sorting on the CD19+ cell fraction. A scatter gate and a fluorescence gate based on CD38 and CD45 were used for isolating the CD38high (CD38hi), CD45intermediate (CD45in) cells.

[0032] FIGURE 17 is an electrophoretic gel showing the successful multiplex overlap-extension RT-PCR reaction on donor TT03 (row A, well 1-12 from eight 96 well plates). The samples have been applied to the agarose gel in two rows (A and B) with 48 samples in each. The expected size of the overlap-extension fragment was approximately 1070bp. The putative overlap-extension fragments are marked with arrows.

[0033] FIGURE 18 shows ELISA analysis of periplasmic extracts from plate G060. The ELISA plate was coated with goat(gt)-anti-human Kappa, and captured Fab fragments were detected with a HRP-conjugated gt-anti-human Fab-specific antibody.

[0034] FIGURE 19 shows ELISA analysis of periplasmic extracts from plate G060. The ELISA plate was coated with 10 µg/ml Ovalbumin (Sigma A-5503), and captured Fab fragments were detected with a HRP-conjugated gt-anti-human Fab-specific antibody.

[0035] FIGURE 20 shows ELISA analysis of periplasmic extracts from plate G060. The ELISA plate was coated with Tetanus Toxoid, and captured Fab fragments were detected with a HRP-conjugated gt-anti-human Fab-specific antibody.

[0036] FIGURE 21 shows a one step competition ELISA analysis of periplasmic extracts from plate G060. The ELISA plate was coated with Tetanus Toxoid (TT), and soluble TT was added to each well at 10^{-7} M in order to compete for the binding of Fab fragments from the bacterial supernatants, to immobilized TT. Captured Fab fragments were detected with a HRP-conjugated gt-anti-human Fab-specific antibody.

[0037] FIGURE 22 shows alignment of variable heavy chain protein sequences from TT antigen-binding clones from plate G060. The degree of sequence homology was represented by different shadings; 100%, 80% and 60% were depicted with black, grey and light grey, respectively. CDR1 is located at alignment position 34 to 41. CDR2 is located at alignment position 55 to 73. CDR3 is located at alignment position 107 to 127. Premature stop codons were denoted by an asterisk. The alignment is divided into 8 separate Figures (a-h) distributed in two rows from left to right with Figure 22a to d in the top row and Figure 22e to h in the bottom row.

[0038] FIGURE 23 shows alignment of variable light chain protein sequence from TT antigen-binding clones from plate G060. The degree of sequence homology was represented by different shadings; 100%, 80% and 60% were depicted with black, grey and light grey,

respectively. CDR1 is located at alignment position 26 to 42. CDR2 is located at alignment position 58 to 64. CDR3 is located at alignment position 97 to 106. Premature stop codons were denoted by an asterisk. The alignment is divided into 8 separate Figures (a-h) distributed in two rows from left to right with Figure 23a to d in the top row and Figure 23e to h in the bottom row.

[0039] FIGURE 24 shows a competition ELISA assay for determination of apparent affinities of selected clones from plate G060. Soluble TT dilutions in concentrations from 100 nM to 25 pM (four-fold dilutions) were added to the Fab fragments, thereby competing out the binding of the Fab-fragments to immobilized TT. The extents of the reactions are given as the ratio of the observed binding at a given soluble TT concentration to the binding found when no soluble TT was added to the reactions.

[0040] FIGURE 25 shows a double phylogenetic dot matrix plot showing the intra- and intergenetic relationship between antibody heavy and light chain variable domain sequences. Phylogenetic threes of V_H and V_L sequences are paired in a dot matrix in order to indicate actual pairing of particular V genes. A) TT binding clones obtained from a combinatorial library using phage display. B) TT binding clones obtained from a library of cognate pairs using the present invention.

Description of the Invention

[0041] The present invention sets out to provide an amplification and linkage process of two or more non-contiguous nucleotide sequences of interest which enables the cloning of such sequences to be fitted to a high-throughput format. This is basically achieved by reducing the number of steps necessary to amplify and link sequences to be cloned.

[0042] One aspect of the invention is a method of linking a plurality of non-contiguous nucleotide sequences comprising amplifying, in a multiplex molecular amplification procedure, nucleotide sequences of interest using a template derived from an isolated single cell, a population of isogenic cells or a population of genetically diverse cells and effecting a subsequent linkage of the amplified sequences. If the template is an isolated single cell or a population of isogenic cells, the linkage results in a nucleic acid segment comprising nucleotide sequences of interest associated with each other in a cognate manner. If the template is a population of genetically diverse cells, the linkage results in a library of segments where each segment comprise nucleic acid sequences of interest associated in a random manner, this is also termed a combinatorial library.

[0043] In one embodiment of the present invention this multiplex molecular amplification procedure is a multiplex PCR amplification, preferably preceded by a reverse transcription step. In a preferred embodiment reverse transcription, amplification and linkage are performed in a single step, using multiplex overlap-extension RT-PCR or alternatively in two steps using multiplex RT-PCR followed by linkage by ligation or recombination.

[0044] A further embodiment of the present invention relates to the generation of libraries of cognate pairs comprising linked variable region encoding sequences, in particular heavy chain variable region and light chain encoding sequences or T cell receptor (TcR) alpha chain encoding sequences and beta chain encoding sequences. The process involves obtaining a lymphocyte-containing cell fraction from at least one suitable donor and optionally enriching for a particular lymphocyte population from this fraction, for example B lymphocytes or T lymphocytes, depending on whether variable region encoding sequences from immunoglobulins or TcRs are desired. The lymphocyte-containing cell fraction or the enriched cell fraction is distributed into an array of vessels, obtaining one cell in each vessel. The array of single cells are subjected to a reverse transcription (RT) step or an alternative cDNA generating procedure, using the nucleic acids derived from the population of single cells as template. The RT step is followed by multiplex molecular amplification procedure and linkage of pairs of variable region encoding sequences generated from each cell according to one of the methods of the present invention.

[0045] The cloning techniques disclosed in the present invention omit laborious and inefficient cloning approaches and additionally reduces the risk of contamination and loss of diversity during multiple cloning steps.

[0046] Another aspect of the invention relates to libraries of cognate pairs produced by the multiplex molecular amplification and linkage process. The initial library of cognate pairs (parent library) generated by the method of the present invention can be subjected to screening, thereby generating a sub-library of cognate pairs encoding target-specific binding protein variable domains or full-length binding proteins.

[0047] In a further embodiment of the invention, the libraries and sub-libraries of the present invention can be used in the expression of recombinant monoclonal or polyclonal proteins, where the original binding affinities and specificities present in the donor are preserved.

Definitions

[0048] The term "cognate pair" describes an original pair of non-contiguous nucleic acids of interest that are contained within or derived from a single cell. In preferred embodiments, a cognate pair comprises two variable region encoding sequences which together encode for a binding protein variable domain and which gene sequences are derived from the same cell. Thus, when expressed either as a complete binding protein or as a stable fragment thereof, they preserve the binding affinity and specificity of the binding protein originally expressed from this cell. A cognate pair can for example be comprised of an antibody variable heavy chain encoding sequence associated with a variable light chain encoding sequence from the same cell, or a T cell receptor α chain encoding sequence associated with a β chain encoding sequence from the same cell. A library of cognate pairs is a collection of such cognate pairs.

[0049] The term "hot-start polymerase" describes polymerases that are inactive or have very low activity at temperatures used for reverse transcription. Such polymerases need to be activated by high temperatures (90 to 95°C) to become functional. This is for example an advantage in single-step RT-PCR procedures, since this prohibits interference of the polymerase with the reverse transcriptase reaction.

[0050] The term "isogenic population of cells" describes a population of genetically identical cells. In particular, an isogenic population of cells derived by clonal expansion of an isolated single cell is of interest in the present invention.

[0051] The term "isolated single cell" describes a cell that has been physically separated from a population of cells corresponding to "a single cell in a single vessel". When distributing a population of cells individually among a plurality of vessels, a population of isolated single cells is obtained. As specified in the section entitled "*Template sources*" the proportion of vessels with a single cell is not necessarily a 100% in order to call it a population of single cells.

[0052] Terms derived from "link" or "linkage" in relation to amplification describes the association of the amplified nucleic acid sequences encoding the nucleic acid sequences of interest into a single segment. In relation to cognate pairs a segment comprises nucleic acid sequences encoding a variable domain, e.g. an antibody heavy chain variable region associated with an antibody light chain variable region encoding sequence, derived from the same cell. The linkage can either be achieved simultaneously with the amplification or as an immediate step following the amplification. There are no requirements to the form or functionality of the segment, it may be linear, circular, single stranded or double stranded.

Nor is the linkage necessarily permanent, one of the nucleic acid sequences of interest may be isolated from the segment if desired, one of the variable region encoding sequence may for example be isolated from a cognate pair segment. However, as long as the original variable regions constituting the cognate pair are not scrambled with other variable regions, they are still considered a cognate pair, although not linked together into a single segment. The linkage is preferably a nucleotide phosphodiester linkage. However, linkage can also be obtained by different chemical cross linking procedures.

[0053] The term "multiplex molecular amplification" describes the simultaneous amplification of two or more target sequences in the same reaction. Suitable amplification methods include the polymerase chain reaction (PCR) (U.S. 4,683,202), ligase chain reaction (LCR), (Wu and Wallace, 1989, Genomics 4, 560-9), strand displacement amplification (SDA) technique (Walker et al., 1992, Nucl. Acids Res. 20, 1691-6), self-sustained sequence replication (Guatelli et al., 1990, Proc. Nat. Acad. Sci. U.S.A., 87, 1874-8) and nucleic acid based sequence amplification (NASBA) (Compton J., 1991, Nature 350, 91-2). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA).

[0054] The term "multiplex PCR" describes a variant of PCR in which two or more target sequences are amplified simultaneously, by including more than one set of primers in the same reaction, e.g. one primer set adapted for amplification of the heavy chain variable region and one primer set adapted for amplification of the kappa chain variable region in the same PCR reaction. Additionally a primer set adapted for amplification of the lambda chain variable region may be combined with these primer sets.

[0055] The term "multiplex RT-PCR" describes a multiplex PCR reaction, which is preceded by a reverse transcription (RT) step. The multiplex RT-PCR, can either be performed as a two-step process with a separate RT step prior to the multiplex PCR, or as a single-step process where all components for both RT and multiplex PCR are combined in a single tube.

[0056] The terms "multiplex overlap-extension PCR" and "multiplex overlap-extension RT-PCR" implies that the multiplex PCR or multiplex RT-PCR is performed utilizing a multiplex overlap-extension primer mix to amplify the target sequences, thereby enabling simultaneous amplification and linkage of the target sequences.

[0057] The term "a plurality of vessels" describes any object (or collection of objects) which enables the physical separation of a single cell from a population of cells. This may be tubes, multiwell plates (e.g. 96-well, 384-well, microtitter plates or other multiwell plates), arrays,

microarrays, microchips, gels, or a gel matrix. Preferably the object is applicable for PCR amplification.

[0058] The term "polyclonal protein" or "polyclonality" as used herein, refers to a protein composition comprising different, but homologous protein molecules, preferably selected from the immunoglobulin superfamily. Thus, each protein molecule is homologous to the other molecules of the composition, but also contains one or more stretches of variable polypeptide sequence, which is/are characterized by differences in the amino acid sequence between the individual members of the polyclonal protein. Known examples of such polyclonal proteins include antibody or immunoglobulin molecules, T cell receptors and B cell receptors. A polyclonal protein may consist of a defined subset of protein molecules, which has been defined by a common feature such as the shared binding activity towards a desired target, e.g. a polyclonal antibody exhibiting binding specificity towards a desired target antigen.

[0059] The term "a population of genetically diverse cells" as used herein, refers to a cell population where the individual cells in the population differ among each other on the genomic level. Such a population of genetically diverse cell is for example a population of cells derived from a donor, or a fraction of such cells, e.g. a B lymphocyte or a T lymphocyte containing cell fraction.

[0060] The term "primer set" is used interchangeably with the term "primer pair" and describes two or more primers which together are capable of priming the amplification of a nucleotide sequence of interest (i.e., one member of a cognate pair). A primer set of the present invention might be designed to prime a family of nucleotide sequences containing variable region encoding sequences. Examples of different families are antibody kappa light chains, lambda light chains, heavy chain variable regions, and α, β, γ, or δ T cell receptor variable regions. A primer set for the amplification of a family of nucleotide sequences containing variable region encoding sequences often constitutes a plurality of primers where several primers can be degenerate primers.

[0061] The term "sequence identity" is expressed as a percentage which indicates the degree of identity between to nucleic acid sequences over the length of the shortest of the two sequences. It can be calculated as $(N_{ref} - N_{dif})\times 100/N_{ref}$, wherein N_{ref} is the number of residues in the shorter of the sequences, and wherein N_{dif} is the total number of non-identical residues in an N_{ref} long optimally aligned match between the two sequences. Hence, the DNA sequence AGTCAGTC will have a sequence identity of 75% with the sequence

TAATCAATCGG (Ndif=2 and Nref=8) (underlining shows the optimal alignment, and bold indicates the two non-identical residues out of 8).

[0062] The terms "randomly" or "random" with respect to linkage refers to linkage of nucleotide sequences which are not derived from the same cell but are linked transversely among a population of genetically diverse cells. If the nucleotide sequences of interest are variable region encoding sequences, this will result in a combinatorial library of linked sequences. If, on the other hand, the nucleotide sequences of interest encode for a non-diverse heteromeric protein the randomly linked sequences will appear similar to sequences linked from a single cell.

[0063] The term "template derived from an isolated single cell," with regard to reverse transcription, relates to the nucleic acids within such an isolated cell. The nucleic acids can for example be in the form of RNA, mRNA, DNA or genomic DNA. The nucleic acids can either be isolated from the cell or still be with the remaining contents of the cell, where the cell is in an intact form or a lysed form.

The Amplification and Linkage Process

[0064] One feature of the present invention reduces the number of tubes necessary to amplify the nucleotide sequences of interest, utilizing a variant of PCR in which two or more target sequences are amplified simultaneously in the same tube, by including more than one set of primers, for example all the primers necessary to amplify variable region encoding sequences, in the same reaction. Generally this approach is known as multiplex polymerase chain reaction (multiplex PCR).

[0065] Multiplex PCR amplification and multiplex PCR preceded by reverse transcription (multiplex RT-PCR) are well known techniques within the diagnostic field, for example in the analysis of mutations, deletions and polymorphisms of DNA, for quantitative assays of mRNA levels and for identification of viruses, bacteria and parasites (reviewed in Markoulatos, P. *et al.* 2002. J. Clin. Lab. Anal. 16, 47-51). However, there are only very few examples where immunoglobulin light chain variable region encoding sequences has been amplified in the same vessel as immunoglobulin heavy chain variable region encoding sequences using a multiplex primer mix composed of more than four primers constituting a $V\kappa$ and/or V_λ primer set together with a V_H primer set (Chapal, N. *et al.* 1997. BioTechniques 23, 518-524.; Liu, A.H. *et al.* 1992. Proc. Natl. Acad. Sci. U.S.A. 89, 7610-7614; Embleton, M.J. *et al.* 1992. Nucleic Acids Res. 20, 3831-3837). The reason for this,

might be that primer sets adapted for the amplification of sequences encoding the variable domains of antigen binding proteins generally are constituted of a plurality of degenerate primers in order to capture the diversity of these variable region encoding sequences. Thus, the complexity of the PCR reaction is highly increased when performing multiplex PCR amplification on variable region encoding sequences.

[0066] A further feature of the present invention is that two or more target sequences amplified by multiplex PCR are linked in close proximity to the amplification process. In particular cognate pairs of variable region encoding sequences are linked by this process.

[0067] One embodiment of the present invention exploits that a multiplex primer mix can be designed to work in an overlap-extension PCR procedure, resulting in a simultaneous amplification and linkage of nucleotide sequences of interest. This multiplex overlap-extension PCR technique serves to reduce the number of reactions necessary to isolate and link nucleotide sequences of interest, in particular cognate pairs of linked variable regions.

[0068] Other embodiments of the present invention apply linkage by ligation or by recombination as an alternative to linkage by multiplex overlap-extension PCR. In these procedures, the linkage is not performed simultaneously with the multiplex PCR amplification, but as an immediate step following the amplification. However, linkage can still be performed in the same tube as the multiplex PCR was performed in.

[0069] In order to perform multiplex overlap-extension PCR, the presence of two or more primer sets (a multiplex primer mix), where at least one primer of each set is equipped with an overlap-extension tail is needed. The overlap-extension tails enable the linkage of the products generated by each of the primer sets during amplification. Such a primer mix is called a multiplex overlap-extension primer mix. The multiplex overlap-extension PCR, differ from conventional overlap-extension PCR in that the sequences to be linked are generated simultaneously in the same tube, thereby providing immediate linkage of the target sequences during amplification, without any intermediate purification. Further, conventional overlap-extension PCR requires a separate linking PCR reaction either with an outer primer set or a nested primer set in order to generate the linked product (Horton, R.M. *et al.* 1989. Gene 77, 61-68). Such an additional amplification step is optional in the multiplex overlap-extension PCR of the present invention.

[0070] A further feature of the present invention is a reverse transcription (RT) step preceding the multiplex PCR or multiplex overlap-extension PCR amplification, utilizing a template derived from an isolated single cell or a population of isogenic cells.

[0071] A further feature of the present invention is the use of nucleotide sequences derived from an isolated single cell or a population of isogenic cells as template for the multiplex PCR amplification. Preferably, RNA from a single cell is reverse transcribed into cDNA prior to the multiplex PCR. For the amplification of some nucleic acid sequences of interest genomic DNA may be used as an alternative to mRNA. By using isolated single cells or a population of isogenic cells derived by clonal expansion of an isolated single cell as template source, it is possible to avoid scrambling of nucleotide sequences encoding a heteromeric protein of interest, with nucleotide sequences derived from different cells within a population of cells. This is of importance if one wishes to obtain the original composition of the sequences of interest. Especially for the generation of a cognate pair of variable region encoding sequences, is the use of an isolated single cell or a population of isogenic cells as template source an important feature.

[0072] Multiplex overlap-extension PCR is a rarely used technology. WO 99/16904 discloses the linkage of exons from a genomic sequence in a single reaction, thereby generating cDNA without utilizing reverse transcription. The process as described, utilized one primer set (constituted of two primers) per exon to be linked, thereby composing a multiplex overlap-extension primer mix. Each individual primer set was capable of overlapping with the adjacent primer set by complementary overlap-extension tails. The cDNA was generated from a template of genomic DNA by performing an overlap-extension PCR reaction utilizing the multiplex overlap-extension primer mix followed by a nested PCR, which was described as a necessary step.

[0073] The generation of cDNA from the exons of genomic DNA as described in WO 99/16904 is a different field than the cloning of sequences encoding for heteromeric proteins. First of all, heteromeric proteins are generally expressed from different genes, whereas the exon linkage as described in WO 99/16904 relates to the linkage of exons from a single gene. Additionally, the present invention facilitates the generation of libraries of linked nucleic acid sequences of interest, in particular combinatorial libraries and libraries of cognate pairs of variable regions; a completely different situation than linking a series of exons from a single gene, resulting in a single non-variable cDNA. Further, the present invention utilizes nucleic acids derived from single cells, preferably in the form of RNA that does not need to be isolated from the remaining cell contents before it can be utilized as template.

[0074] There are few publications, where multiplex overlap-extension RT-PCR has been described in relation to linkage of variable region encoding sequences.

[0075] The most simple form of multiplex overlap-extension RT-PCR was described for the isolation of a scFv encoding sequence from a hybridoma cell line (Thirion, S. *et al.* 1996. Eur. J. Cancer Prev. 5, 507-511 and Mullinax, R.L *et al.* 1992. BioTechniques 12, 864-869). The methods described by Thirion and Mullinax utilized reverse transcription of mRNA with oligo-dT primers on total RNA extracted from the hybridoma cell line followed by a separate linkage step. The linkage step was performed with a total of four primers, constituting two primer pairs for amplification of heavy chain variable region and light chain variable region encoding sequences, respectively. The V_L forward primer and the V_H or C_H reverse primer contained complementary overlap-extension tails, thereby enabling simultaneous amplification and linkage of the heavy chain variable region and light chain variable region encoding sequences. These methods did not use a nested PCR to increase the sensitivity of the linkage method.

[0076] The other example of multiplex overlap-extension RT-PCR in relation to linkage of variable region encoding sequences was described in the previously mentioned WO 93/03151, providing a method of cloning heavy chain variable region and light chain variable region encoding sequences originating from the same cell without having to isolate single cells prior to the cloning. The method described in WO 93/03151 required washing between the RT step and the multiplex overlap-extension PCR step. Further, one of the specific aims brought about by WO 93/03151, was to solve the problem of having to isolate single cells in order to obtain cognate pairs of variable region encoding sequences.

[0077] None of these known multiplex overlap-extension RT-PCR techniques were developed to function on template derived from an isolated single cell. Neither were any of the methods able to perform as single-step RT-PCR reactions.

[0078] One embodiment of the present invention encompasses the linkage of a plurality of non-contiguous nucleotide sequences of interest. The method comprises amplifying, in a multiplex PCR or multiplex RT-PCR amplification procedure, nucleotide sequences of interest using a template derived from an isolated single cell or a population of isogenic cells and effecting linkage of the amplified nucleotide sequences of interest. Further, the method comprises an optional step of performing an additional amplification of the linked products.

[0079] A further embodiment of the present invention encompasses a method of producing a library of cognate pairs comprising linked variable region encoding sequences. The method comprises providing a lymphocyte-containing cell fraction from a donor, which is optionally enriched for a particular lymphocyte population from said cell fraction. Further, a population

of isolated single cells is obtained by distributing cells from the lymphocyte-containing cell fraction, or the enriched cell fraction, individually among a plurality of vessels. Multiplex molecular amplification (multiplex RT-PCR amplification) of the variable region encoding sequences contained in the population of isolated single cells is performed and linkage of pairs of variable region encoding sequences, wherein an individual pair is derived from a single cell, within the population of isolated single cell, is effected. Further, the technique comprises two optional steps: in the first step the individual isolated single cell in the population of single cells is expanded to a population of isogenic cells prior to performing multiplex RT-PCR amplification. Thereby obtaining a plurality of vessels with a diverse population of isogenic cells (one population of isogenic cells in one vessel). The second optional step encompasses performing an additional amplification of the linked variable region encoding sequences.

[0080] In preferred embodiments of the present invention, an individual member of said library of cognate pairs comprised of an immunoglobulin light chain variable region encoding sequence is associated with an immunoglobulin heavy chain variable region encoding sequence, originating from the same cell or of sequences encoding a T cell receptor binding domain, constituted of an alpha chain variable region associated with a beta chain variable region or a gamma chain variable region associated with a delta chain variable region, where the associated variable regions originate from the same cell.

[0081] The multiplex RT-PCR amplification of the present invention can be performed either as a two-step process, where reverse transcription (RT) is preformed separate from the multiplex PCR amplification (or alternative multiplex molecular amplification), or as a single-step process, where the RT and multiplex PCR amplification steps are performed with the same primers in one tube.

[0082] The reverse transcription (RT) is performed with an enzyme containing reverse transcriptase activity resulting in the generation of cDNA from total RNA, mRNA or target specific RNA from an isolated single cell. Primers which can be utilized for the reverse transcription are for example oligo-dT primers, random hexamers, random decamers, other random primers, or primers that are specific for the nucleotide sequences of interest.

[0083] The two-step multiplex RT-PCR amplification procedure, allows for the cDNA generated in the RT step, to be distributed to more than one vessel permitting for the storage of a template fraction before proceeding with the amplification. Additionally, the distribution of cDNA to more than one tube, allows for the performance of more than one multiplex PCR

amplification of nucleic acid derived from the same template. Although, this results in an increased number of separate reactions, it opens for the possibility to decrease the complexity of the multiplex primer mix if this should be desired. This two-step approach can for example be applied to amplify and link heavy chain variable region and kappa light chain variable region encoding sequences in one tube and heavy chain variable region and lambda light chain variable region encoding sequences in a different tube utilizing the same template. A single cell usually only expresses one of the light chains. However, it will often be easier to perform the reactions simultaneously instead of awaiting the result of one of the reactions before performing the other. Further, the amplification of both kappa and lambda serves as an internal negative control, since it would be expected that only kappa or lambda amplify from a single cell.

[0084] In the single-step multiplex RT-PCR procedure, reverse transcription and multiplex PCR amplification is carried out in the same vessel. All the components necessary to perform both the reverse transcription and the multiplex PCR in a single step are initially added into the vessels and the reaction is performed. Generally, there is no need to add additional components once the reaction has been started. The advantage of single-step multiplex RT-PCR amplification is that it reduces the number of steps necessary to generate the linked nucleotide sequences of the present invention even further. This is particular useful when performing multiplex RT-PCR on an array of single cells, where the same reaction needs to be carried out in a plurality of vessels. Single-step multiplex RT-PCR is performed by utilizing the reverse primers present in the multiplex primer mix needed for the multiplex PCR amplification as primers for the reverse transcription as well. Generally, the composition needed for the single-step multiplex RT-PCR comprises a nucleic acid template, an enzyme with reverse transcriptase activity, an enzyme with DNA polymerase activity, deoxynucleoside triphosphate mix (dNTP mix comprising dATP, dCTP, dGTP and dTTP) and a multiplex primer mix. The nucleic acid template is preferably total RNA or mRNA derived from an isolated single cell either in a purified form, as a lysate of the cell or still within the intact cell. Generally, the exact composition of the reaction mixture requires some optimization for each multiplex primer mixture to be used with the present invention. This applies both for the two-step and the single-step multiplex RT-PCR procedures.

[0085] In alternative embodiments of the present invention it may be appropriate to use genomic DNA instead of RNA as template. In such cases the reverse transcription step is

omitted, and the remaining steps of the invention are performed as described throughout the application.

[0086] For some single-step multiplex RT-PCR reactions it may be an advantage to add additional components during the reaction. For example, addition of the polymerase following the RT step. Other components could for example be a dNTP mixture or a multiplex primer mix possibly with a different primer composition. This can then be considered as a one-tube multiplex RT-PCR, which generally has the same advantages as the single-step multiplex RT-PCR, since it also limits the number of tubes necessary to obtain the desired linked products.

[0087] The nucleotide sequences of interest, amplified by the multiplex RT-PCR, can be linked to one another by several methods, such as multiplex overlap-extension RT-PCR, ligation or recombination, using different multiplex primer mixes. Preferably the multiplex RT-PCR amplification and linkage process is a single step or a two step process. However, the linkage process may also be performed as a multi step process, using for example a stuffer fragment to link the nucleic acid sequences of interest, either with PCR, ligation or recombination. Such a stuffer fragment may contain cis-elements, promoter elements or a relevant coding sequence or recognition sequence. In a preferred embodiment the linkage process is performed in the same vessel as the multiplex RT-PCR amplification.

[0088] In one embodiment of the present invention the linkage of a plurality of non-contiguous nucleotide sequences of interest is performed in association with the multiplex PCR amplification, utilizing a multiplex overlap-extension primer mix. This results in the combined amplification and linkage of the target sequences. Generally, the composition needed for the multiplex overlap-extension PCR comprises, a nucleic acid template, an enzyme with DNA polymerase activity, deoxynucleoside triphospate mix (dNTP mix comprising dATP, dCTP, dGTP and dTTP) and a multiplex overlap-extension primer mix.

[0089] In a particular embodiment of the present invention, the linkage of a plurality of non-contiguous nucleotide sequences of interest is performed by multiplex overlap-extension RT-PCR using a template derived from an isolated single cell or a population of isogenic cells. Further, the method comprises an optional step of performing an additional molecular amplification of linked products. Preferably, the multiplex overlap-extension RT-PCR is performed as a single-step/one-tube reaction.

[0090] A multiplex overlap-extension primer mix of the present invention comprises at least two primer sets capable of priming the amplification and linkage of at least two variable

region encoding sequences, for example, amplification and linkage of sequences from immunoglobulin heavy chain variable region families with kappa or lambda light chain variable region families, or amplification and linkage of sequences from T cell receptor families α, β, γ, or δ.

[0091] In another embodiment of the present invention the plurality of nucleotide sequences of interest, amplified by multiplex RT-PCR, are linked by ligation. To achieve this, the multiplex primer mix used for the multiplex RT-PCR, is designed such that the amplified target sequences can be cleaved with appropriate restriction enzymes, and covalent linkage by DNA ligation can be performed (the primer design is described in the section *"Primer Mixtures and Design"*). Following multiplex RT-PCR amplification with such a multiplex primer mix, the restriction enzymes needed to form compatible ends of the target sequences, are added to the mixture together with the ligase. No purification of the PCR products is needed prior to this step, although purification may be performed. The reaction temperature for the combined restriction cleavage and ligation is approximately between 0 and 40°C. However, if the polymerase from the multiplex PCR reaction is still present in the mixture, an incubation temperature below room temperature is preferred, most preferred are temperatures between 4 and 16°C.

[0092] In yet another embodiment of the present invention, the plurality of nucleotide sequences of interest, amplified by multiplex RT-PCR, are linked by recombination. In this approach, the target sequences amplified can be joined using identical recombination sites. Linkage is then performed by adding the recombinases facilitating recombination. Some suitable recombinase systems are Flp recombinase with a variety of FRT sites, Cre recombinase with a variety of lox sites, integrase ΦC31 which carries out recombination between the attP site and the attB site, the β-recombinase-six system as well as the Gin-gix system. Linkage by recombination has been exemplified for two nucleotide sequences (V_H linked with V_L) (Chapal, N. *et al.* 1997 BioTechniques 23, 518-524), hereby incorporated by reference.

[0093] In a preferred embodiment of the present invention, the nucleotide sequences of interest comprise variable region encoding sequences and the linkage generates a cognate pair of variable region encoding sequences. Such a cognate pair may comprise one or more constant region encoding sequences in addition to the variable regions.

[0094] In an even more preferred embodiment of the present invention, the nucleotide sequences of interest comprise immunoglobulin variable region encoding sequences and the

linkage generates a cognate pair of light chain variable region and heavy chain variable region encoding sequences. Such a cognate pair may comprise one or more constant region encoding sequences in addition to the variable regions. Further, such a cognate pair may be isolated from template derived from cells of the B-lymphocyte lineage enriched from a lymphocyte-containing cell fraction, such as whole blood, mononuclear cells or white blood cells.

[0095] In a just as preferred embodiment of the present invention, the nucleotide sequences of interest comprise TcR variable region encoding sequences and the linkage generates a cognate pair of α chain variable region and β chain variable region encoding sequences or γ chain variable region and δ chain variable region encoding sequences. Such a cognate pair may comprise one or more constant region encoding sequences in addition to the variable regions. Further, such a cognate pair may be isolated from template derived from cells of the T-lymphocyte lineage enriched from a lymphocyte-containing cell fraction, such as whole blood, mononuclear cells or white blood cells.

[0096] Another aspect of the present invention, is to utilize the multiplex RT-PCR with a population of genetically diverse cells as template source. The majority of heteromeric protein encoding sequences do not vary from cell to cell as is the case with variable region encoding sequences from binding proteins. Thus, when utilizing the present invention for the cloning of such non-variable heteromeric protein encoding sequences there is no need to perform an initial isolation of single cells.

[0097] In this embodiment of the present invention, a plurality of non-contiguous nucleotide sequences of interest are linked randomly by a method comprising, performing multiplex RT-PCR amplification of nucleotide sequences of interest using a template derived from a population of genetically diverse cells and effecting linkage of the amplified nucleotide sequences of interest. Further, the method comprises an optional step of performing an additional amplification of the linked products. As with the single cell approach the linkage can either be performed utilizing a multiplex overlap-extension primer mix for the amplification or alternatively by ligation or recombination. Preferably the template derived from the population of cells is not strictly contained within the cells. The population of cells may for example be lysed.

[0098] Application of the process of random linkage on a population of cells expressing variant binding proteins, allows for a simplified generation of combinatorial libraries of variable region encoding sequences. Preferably, the population of cells constitutes cells that

express variable region binding proteins, such as B lymphocytes, T lymphocytes, hybridoma cells, plasma cells, or a mixture of these cells.

[0099] The population of cells in the above mentioned embodiment can for example be permeabilized or lysed, without additional purification, or the template nucleic acids can be isolated from the cells by standard procedures. The single-step multiplex RT-PCR procedure is preferred. However, the two-step procedure may also be used in the embodiment.

[00100] The invention also provides a combinatorial library comprised of linked pairs of immunoglobulin light chain variable region and heavy chain variable region encoding sequences.

[00101] An efficient way to increase the specificity, sensitivity, and yield of the multiplex RT-PCR-linkage process, is by performing an additional molecular amplification of the linked nucleotide sequences obtained from the multiplex RT-PCR followed by linkage by ligation or recombination or linkage using the multiplex overlap-extension RT-PCR. This additional amplification is preferably performed with PCR amplification, utilizing a primer mix adapted for amplifying the linked nucleic acid sequences of interest. The primer mix utilized may be the outer primers of the multiplex primer mix or multiplex overlap-extension primer mix, meaning the primers which anneal to the outermost 5' end and 3' end of the sense strand of the linked variable region encoding sequences, thereby enabling the amplification of the entire linked product. The outer primers can also be described as the primers of the multiplex overlap-extension primer mixture that do not contain overlap extension tails. Alternatively, a nested or semi-nested primer set can be used for the additional amplification of the linked nucleotide sequences. Such a nested PCR especially serves to increase the specificity of the method as well as to increase the amount of linked product. For the present invention, semi-nested PCR (as described in the section entitled *Primer Mixtures and Design)* is considered to function as well as the nested PCR. Thus, it is desired although not necessary for the present invention to perform an additional PCR amplification of the linked products from the multiplex overlap-extension RT-PCR or of the products linked by ligation or recombination, preferably using nested PCR or semi-nested PCR.

[00102] The additional amplification can either be performed directly using a fraction or the entire multiplex overlap-extension RT-PCR reaction product or ligation product or recombination product, or a fraction of any one of these products or using partially purified linked products from any one of these reactions, e.g. by performing an agarose gel electrophoresis of the linked products, and excising the fragment corresponding to the

expected size of the linked variable region encoding sequences. For products linked by multiplex overlap-extension RT-PCR, the additional amplification is preferably performed directly on a fraction from the multiplex overlap-extension RT-PCR reaction, since this would assist linkage of the individual target sequences that were not linked in the first reaction.

Sequences of interest

[00103] The nucleotide sequences of interest of the present invention can be selected from sequences that encode different subunits or domains, which when expressed, forms a protein or part of a protein. Such proteins that are composed of at least two non-identical subunits are known as heteromeric proteins. Heteromeric proteins are common in all kinds of species. Some of the classes to which such proteins belong are for example enzymes, inhibitors, structural proteins, toxins, channel proteins, G-proteins, receptor proteins, immunoglobulin superfamily proteins, transportation proteins etc. The nucleotide sequences encoding such heteromeric proteins are non-contiguous, meaning for example that they originate from different genes, or different mRNA molecules. However, non-contiguous as used in the present invention may also mean nucleotide sequences encoding domains of the same protein, where the domains are separated by nucleotide sequences which are not of interest.

[00104] In one embodiment of the present invention the nucleotide sequencess of interest contain variable region encoding sequences from the immunoglobulin superfamily, such as immunoglobulins (antibodies), B cell receptors and T cell receptors (TcR's). Especially variable region encoding sequences from immunoglobulins are of interest. Such variable region encoding sequences comprise full-length antibodies as well as Fab's, Fv's, scFv's and combinations of fragments of the variable region encoding sequences, e.g. complementarity determining regions (CDR's), joining genes or V-genes or combinations of these. Generally the present invention can be applied with any combinations of variable region encoding sequences and fragments thereof. The present application exemplifies the linkage of the entire light chain with the variable domain of the heavy chain. However, the present invention also allow for the linkage of only the variable domains of the heavy and light chains generating Fv or scFv encoding sequences, or the linkage of the entire light chain with the heavy chain variable region + constant region domain C_{H1} + parts of the hinge region, generating Fab, Fab' or $F(ab)_2$. Further, it is possible to add any region of the heavy chain

constant region domains to the variable heavy chain, thereby generating truncated antibody encoding sequences or full-length antibody encoding sequences.

[00105] In a further embodiment of the present invention variable region encoding sequences comprise one type of immunoglobulin light chain (kappa or lambda) encoding sequence and one immunoglobulin heavy chain variable region encoding sequence.

[00106] Variable region encoding sequences derived from T cell receptors (TcR's) are also of interest. Such TcR encoding sequences comprise encoding sequences for full-length alpha and beta chains or gamma and delta chains as well as soluble TcR's or only the variable domains of these chains or single chain fusion proteins thereof (e.g. single chain αβ or single chain γδ).

Template sources

[00107] One feature of the present invention is the ability to link nucleotide sequences derived from an isolated single cell, a population of isogenic cells, or a genetically diverse population of cells which have not been separated into single vessels. The cells utilized in the present invention can for example be bacteria, yeast, fungi, insect cells, plant cells or mammalian cells or fractions of such cells. Blood cells derived from mammals are one example of a fraction of cells that can be utilized in the present invention.

[00108] A preferred feature of the present invention is the use of isolated single cells or a population of isogenic cells as template source, since scrambling of the nucleic acid sequences of interest, in particular variable region encoding sequences is avoided. This is of importance if one wishes to obtain an original pair of for example variable region encoding sequences.

[00109] Another preferred feature of the present invention, is obtaining a single cell or population of single cells from a cell fraction comprising lymphocytes, such as B lymphocytes, T lymphocytes, plasma cells and/or various developmental stages of these cell lineages. Other populations of cells that express binding proteins from the immunoglobulin superfamily might also be used to obtain single cells. Cell lines such as hybridoma cells, cell lines of B lymphocyte or T lymphocyte lineage or virus immortalized cell lines or donor derived cells participating in the immune response are also applicable in the present invention. Donor derived lymphocyte-containing cell fractions may be obtained from natural tissue or fluid which is rich in such cells, e.g. blood, bone marrow, lymph nodes, spleen tissue, tonsil tissue or from infiltrations in and around tumors or inflammatory tissue

infiltrations. Suitable cell donors for the present invention can be selected from vertebrates that all contain an acquired immune system. Donors can either be naïve or hyperimmune with respect to a desired target. For the isolation of antigen binding proteins with binding specificities toward a desired target, hyperimmune donors are preferred. Such hyperimmune donors can either be donors immunized with the target, or fragments of the target, or it can be convalescent patients, or non-healthy individuals which are running a natural immune response towards the target e.g. autoimmune patients, cancer patients, patients with infectious diseases e.g. HIV patients, Hepetitis A, B or C patients, SARS patients etc., or patients with chronic diseases.

[00110] When utilizing recombinant proteins for treatment, it is preferable that they are derived from sequences that have species identity with the individual to be treated (e.g. human sequences for treatment of humans). Firstly, because recombinant proteins derived from a foreign sequence (i.e. non-human) will be recognized by the immune system leading to an immune response implicating polyclonal anti-protein antibodies. These anti-protein antibodies can block the drug action by occupying the active site, they will accelerate drug clearance and they could potentially induce adverse reactions such as hypersensitivity reactions upon repeated exposure.

[00111] Immunogenicity may however, also be seen in cases where the recombinant protein is derived from a sequence that have species identity. Such immunogenicity can for example be induced by post-translational modifications that might differ from those seen *in vivo*. Combinatorial libraries of variable region encoding sequences might also give rise to immunogenicity, since they consist of random pairs of variable region encoding sequences created in vitro. The rules that govern formation of antibody heavy and light chain encoding sequence pairs (or T cell receptors) *in vivo* are not completely understood. Hence, it follows that some of the *in vitro* formed pairs could be recognized as foreign by the immune system, even though both sequences constituting the pair are perfectly human. Binding proteins obtained from cognate libraries do on the other hand not create said abnormal combinations and they are consequently of lesser potential immunogenicity than binding proteins from combinatorial libraries. This does not mean that products from combinatorial libraries are unsuitable for treatment, they just require a larger degree of monitoring with respect to the above mentioned side effects.

[00112] For use in the present invention, cell donors should preferably be of the same species as the species to be treated with the products obtainable from the linked nucleotide sequences

of the present invention. Preferably, a cell donor is a domestic animal, a pet, a human or a transgenic animal. Transgenic animals carrying human immunoglobulin loci are described in U.S. Patent No. 6,111,166 and Kuroiwa, Y. *et al.* Nature Biotechnology; 2002; 20: 889-893. Such transgenic animals are capable of producing human immunoglobulins. Thus, fully human antibodies against a specific target can be raised by usual immunization techniques of such transgenic animals. This allows for generation of libraries encoding for binding proteins with specificities towards more difficult targets such as human antigens to which no or limited natural human antibody response exist. Such transgenic animals can likewise be developed to produce human T cell receptors.

[00113] In a further embodiment of the present invention, the lymphocyte-containing cell fraction is constituted of whole blood, bone marrow, mononuclear cells, or white blood cells obtained from a donor. Mononuclear cells can be isolated from blood, bone marrow, lymph nodes, spleen, infiltrations around cancer cells and inflammatory infiltrations. Mononuclear cells can be isolated by density centrifugation techniques, e.g. Ficoll gradients. If the mononuclear cells are isolated from samples composed of tissue, the tissue is disintegrated before the gradient centrifugation is performed. Disintegration can be performed, for example, by mechanical methods such as grinding, electroporation and/or by chemical methods such as enzymatic treatments. The isolation of white blood cells can be performed directly from donors using leukopheresis. Raw preparations of for example bone marrow or tissue, which contain lymphocytes, can also be used in the present invention. Such preparations will need to be disintegrated, for example as described above, in order to facilitate single cell distribution.

[00114] A further feature of the present invention is enrichment of the lymphocyte-containing cell fraction e.g. whole blood, mononuclear cells, white blood cells or bone marrow, with respect to a particular lymphocyte population, such as cells from the B lymphocyte or T lymphocyte lineage. Enrichment of B lymphocytes can for example be performed, using magnetic bead cell sorting or fluorescence activated cell sorting (FACS) taking advantage of lineage-specific cell surface marker proteins such as CD19 or other B cell lineage-specific markers. Enrichment of T lymphocytes can for example be performed, utilizing a cell surface marker such as CD3 or other T cell lineage-specific markers.

[00115] A preferred feature of the present invention is to sort the enriched B lymphocytes further in order to acquire plasma cells, before distributing the cells individually among a plurality of vessels. Isolation of plasma cells is generally performed by FACS sorting,

utilizing surface markers such as CD38 possibly in combination with CD45. Other plasma cell-specific surface markers or combinations thereof can be utilized as well, for example CD138, CD20, CD21, CD40, CD9, HLA-DR or CD62L, the exact choice of marker depends on the plasma cell source, e.g. tonsils, blood or bone marrow. Plasma cells can also be obtained from a non-enriched lymphocyte-containing cell population obtained from any of these sources. The plasma cells isolated from blood are sometimes called early plasma cells or plasmablasts. In the present invention these cells are also termed plasma cells although they are CD19 positive in contrast to plasma cells residing in the bone marrow. Plasma cells are desired for the isolation of cognate pairs of immunoglobulin encoding sequences because a higher frequency of these cells produces antigen-specific antibodies that reflect the acquired immunity toward the desired antigen and most of the cells have undergone somatic hypermutation and therefore encode for high-affinity antibodies. Further, the mRNA levels in plasma cells are elevated compared to the remaining B lymphocyte population, thus the reverse transcription procedure is more efficient when using single plasma cells. As an alternative to plasma cell isolation, memory B cells may be isolated from a lymphocyte containing cell fraction utilizing a cell surface marker such as CD22.

[00116] An alternative feature of the present invention, is selecting the enriched B lymphocytes for antigen specificity before distributing the cells among a plurality of vessels. Isolation of antigen-specific B lymphocytes is performed by contacting the enriched B lymphocytes with the desired antigen or antigens enabling binding of antigen to surface exposed immunoglobulin, followed by isolation of binders. This can be done, for example, by coating magnetic beads with the desired antigen or antigens followed by magnetic bead cell sorting, by FACS, by coating a column with the antigens followed by affinity chromatography, by filter screening assays or other methods known in the art. Plasma cells as well as B lymphocytes, non-enriched mononuclear cells, white blood cells, whole blood, bone marrow or tissue preparations can be subjected to isolation with respect to antigen specificity if this is desired.

[00117] Another feature of the present invention, is to sort enriched T lymphocytes (e.g. CD3 positive cells) using surface markers CD45R0 and/or CD27 to obtain a fraction of memory T cells. T lymphocytes can also be selected for MHC-antigen specificity using MHC-peptide complexes (e.g. Callan, M.F. *et al.* 1998. J. Exp. Med. 187, 1395-1402; Novak, E.J. *et al.* 1999. J. Clin. Invest 104, R63-R67).

[00118] A further feature of the present invention is immortalization of any of the isolated cell fractions described in the above (e.g. B lymphocytes, plasma cells, memory cells or T lymphocytes). Immortalization may for example be performed with Epstein-Barr virus (Traggiai, E., et al., 2004. Nat Med 10, 871-875) prior to cell distribution. Alternatively, isolated single cells may be immortalized and expanded prior to reverse transcription. Traggiai *et al.*, Nat Med. 2004 Aug;10(8):871-5.

[00119] A further feature of the present invention, is the distribution of a population of desired cells (e.g. hybridoma cells, cell lines of B lymphocyte or T lymphocyte lineage, whole blood cells, bone marrow cells, mononuclear cells, white blood cells, B lymphocytes, plasma cells, antigen-specific B lymphocytes, memory B cells, T lymphocytes, antigen/MHC-specific T lymphocytes, or memory T cells) individually, into a plurality of vessels, in order to obtain a population of isolated single cells. This isolation of single cells refers to the physical separation of cells from a population of cells in such a way that a single vessel contains a single cell, or a micro array, chip or gelmatrix is loaded in a manner that produce single cells. The cells may be distributed directly into multitudes of vessels such as arrays of single vessels by limiting dilution. The single vessels utilized in the present invention are preferably those applicable in PCR (e.g. PCR tubes and 96 well or 384 well PCR plates or larger arrays of vessels). However other vessels may also be used. When distributing single cells into a large number of single vessels (e.g. 384 well plates), a population of single cells is obtained. Such a distribution may be performed, for example, by dispensing a volume into a single vessel that on average encompasses a cell concentration of one, 0.5 or 0.3 cell, thereby obtaining vessels that on average contain a single cell or less. Since distribution of cells by limiting dilution is a statistical event, a fraction of the vessels will be empty, a major fraction will contain a single cell, and a minor fraction will contain two or more cells. Where two or more cells are present in a vessel some scrambling of the variable region encoding sequences may occur among the cells present in the vessel. However, since it is a minor event it will not affect the overall utility of the present invention. Additionally, combinations of variable region encoding sequences which do not posses the desired binding affinity and specificity will most likely not be selected and hence eliminated during a screening process. Therefore, minor events of scrambling will not significantly affect the final library of the present invention.

[00120] There are alternatives to cell distribution by limiting dilution using, for example, cell sorters such as FACS machines or robots that can be programmed to accurately dispense

single cells into single vessels. These alternatives are preferable, since they are less laborious and are more efficient in uniformly obtaining a distribution of single cells into single vessels.

[00121] The enrichment, sorting and isolation procedures described in the above, are performed such that the majority of the cells are kept intact. Rupture of cells during enrichment and sorting might result in scrambling of the variable region encoding sequences. However, this is not expected to be a problem since the frequency of rupture is expected to be low. Washing and possible RNAse treatment of the cells prior to distribution into single vessels will remove any RNA that has leaked during the process.

[00122] Further, when considering the above descriptions of how to distribute cells in order to obtain a population of single cells in a population of single vessels, it is not to be interpreted as an absolutely required feature that every vessel must contain a single cell. Rather, it indicates that a majority of the vessels contain single cells, e.g. the number of vessels with two or more cells is below 25% of the total amount of cells distributed, or even better it is below 10%.

[00123] A further feature of the present invention is the performance of a reverse transcription using template derived from cells distributed individually among a plurality of vessels.

[00124] For the purpose of reverse transcription (RT), in accordance with the present invention, the nucleic acids within a single cell that is to serve as template source for the RT, are considered to be derived from a single cell although they have not necessarily been separated from the remaining contents of that single cell.

[00125] When the final distribution of the single cells to their single vessels has been performed, the single cells may be expanded in order to obtain a population of isogenic cells prior to reverse transcription. This process yields more mRNA to be used as template, which might be important if a rare target is to be amplified and linked. However, the cells should remain genetically identical with respect to the target gene during the expansion. The isolated cells or the population of isogenic cells can either be kept intact or lysed, as long as the template for the reverse transcription is not degraded. Preferentially, the cells are lysed in order to ease the following reverse transcription and PCR amplification.

[00126] In a different embodiment of the present invention, the disclosed multiplex overlap-extension RT-PCR method or multiplex RT-PCR followed by linkage by ligation or recombination may also be utilized on template derived from a genetically diverse population of cells which have not been separated into single vessels, but all remain together as a pool of cells. This method may be used for the generation of combinatorial libraries. Such an

approach will not require the distribution of single cells. However, the cells which may be used in this approach are the same as those described for the single cell approach, for example a population (pool) of sorted B lymphocytes or T lymphocytes. When performing the single-step multiplex overlap-extension RT-PCR or single-step multiplex RT-PCR followed by linkage by ligation or recombination on such a population of cells, it is preferable to lyse the cells prior to the reaction and if desired total RNA or mRNA may be isolated from the lysate.

[00127] The sensitivity of the single-step multiplex overlap-extension RT-PCR of the present invention enables the use of a very low amount of template. As shown in Example 2 and 3, single-step multiplex overlap-extension RT-PCR may be carried out on an amount of template corresponding to the lysate of a single cell.

Primer Mixtures and Design

[00128] The primer mixtures of the present invention comprise at least four primers that form primer sets two by two, which are capable of amplifying at least two different target sequences of interest. Mixtures of two or more of such primer sets constitute a multiplex primer mix. Preferably, a multiplex mix comprise at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20, 30, 40, 50, 60, 70, 80, 90 100, 110, 120, 130, 140 or 150 primer sets (primer pairs). In particular for the amplification of variable region encoding sequences, may an individual primer set within the multiplex primer mix constitute several more than two primers. Preferably, an individual primer set comprise at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 280 or 300 primers. Preferably the total number of primers in a multiplex primer mix is at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 45, 50, 60, 70, 80, 90, 100, 125, 150 or 200 and at the most 225, 250, 275, 300, 325, 350, 375 or 400 primers.

[00129] All the primers of the present invention comprise a gene-specific region, and preferably all primers are additionally equipped with a primer tail at the 5' end of the primer, i.e. 5' non-coding sequences which are fused to the 3' end of the gene-specific primer part. Such a primer tail is approximately from 6 to 50 nucleotides long, but it may also be longer if desired. Upon amplification the primer tails are added to the target sequences.

[00130] Primer tails of the present invention are for example, cloning tails and linkage tails such as, tails adapted for linkage by ligation, tails adapted for linkage by recombination or overlap-extension tails.

[00131] Cloning tails may be from 6 to 20 nucleotides long or longer and comprise restriction sites and/or recombination sites, which are useful for the insertion of the linked product into an appropriate vector.

[00132] To enable linkage by ligation, the primer sets of the multiplex primer mix are designed such that one part (forward or reverse primer(s)) of the first primer set is equipped with a linkage-tail containing a restriction site that upon cleavage will be compatible with a restriction site located in the linkage tail of one part of the second primer set. For linkage of more than two target sequences, the second part of the second primer set is equipped with a restriction site that upon cleavage will be compatible with a restriction site located in one part of the third primer set. This second restriction site located in the second primer set should be non-compatible with that of the first primer set. A considerable number of target sequences can be linked by designing primer sets this way. Restriction sites with a low frequency or no occurrence, in the target sequences should be chosen. Further, it is preferable that compatible restriction sites are not identical, such that the site of ligation becomes cleavage-resistant for the particular restriction enzymes used. This will drive the reaction towards linkage of target sequence one with target sequence two, since linkage between identical target sequences will be cleavable by the restriction enzymes. Suitable pairs of restriction sites are for example, SpeI with XbaI (alternatively NheI or AvrII can substitute one or both of these), NcoI with BspHI , EcoRI with MfeI or PstI with NsiI. For linkage, SpeI can for example be located in target sequence one, XbaI can be located in target sequence two, NcoI can be located at the other end of target sequence two and BspHI in target sequence three and so forth. To simplify the process further, it is an advantage if the restriction enzymes function in the same buffer.

[00133] To enable linkage by recombination the primer sets of the multiplex primer mix can for example be designed as exemplified in the article by Chapal (1997 BioTechniques 23, 518-524), which is hereby incorporated by reference.

[00134] To enable the linkage of the nucleotide sequences of interest in the same step as the multiplex PCR amplification, tails adapted for overlap-extension PCR are added to at least one primer of each primer set of the multiplex primer mix, thereby generating a multiplex overlap-extension primer mixture.

[00135] The overlap-extension tails are typically longer, ranging from 8 to 75 nucleotides in length and may contain restriction sites or recombination sites which allow for subsequent insertion of regulatory elements such as promoters, ribosomal binding sites, termination sequences, or linker sequences such as in a scFv. The overlap-extension tail may also contain a stop codon if that is desired. Generally there are three types of overlap-extension tails, as illustrated in Figure 1. In type I the overlap-extension tails of two primer sets solely overlap with each other. Not necessarily all of the nucleotides of two overlap-extension tails are complementary to each other. In one aspect of the present invention the complementary nucleotides represent between 60 to 85% of the overlap-extension tail. In type II overlap-extension tails, 4 to 6 of the 5' nucleotides are complementary to the gene-specific region of the adjacent target sequence. In type III overlap-extension tails, the entire overlap is complementary to the adjacent target sequence. The type I and II overlap-extension tails are preferred when regulatory elements and the like are later to be inserted between the linked target sequences. Type II overlap-extension tails are preferred if the target sequences are to be linked by a defined linker as seen with scFv. Type III overlap-extension tails are preferred if the target sequences are to be linked in-frame to each other.

[00136] Design of overlap-extension tails is dependent on sequence features such as length, relative GC content (GC%), presence of restriction sites, palindromes, melting temperature, the gene-specific part to which they are coupled etc. The length of the overlap-extension tails should be between 8 and 75 nucleotides long, preferably they are from 15 to 40 nucleotides long. Even more preferred they are from 22 to 28 nucleotides long. The use of very long overlap-extension tails (50 to 75 nucleotides) could favor the linkage of the products produced by each primer set. However, the proportion between the length of the overlap-extension tail and the gene-specific region probably will need to be adjusted when using very long overlap-extension tails. The GC% preference is dependent on the length of the overlap-extension tail. Since shorter tails have a shorter area where they are complementary they need a higher GC% to strengthen the interaction than longer tails. Other principles of primer design should likewise be observed, e.g. primer dimerization and hairpin formation should be minimized. Neither shall they engage in false priming. Further, it is known that Taq DNA polymerase often adds an adenosine (A) at the 3' end of the newly synthesized DNA strand, and this can be accommodated for in overlap-extension tail design by enabling overlap-extension tails to accommodate 3' non-template A addition.

[00137] The choice of primers that carry the linkage tail, e.g. the overlap-extension tail, tail adapted for linkage by ligation or tail adapted for linkage by recombination, defines the order and direction of linkage of the target sequences. It is not essential to the invention whether it is the forward primer(s) or reverse primer(s) of a primer set or possibly both forward and reverse primers that are equipped with the linkage tail. However, some consideration should be given to this anyway since the order and direction of the target sequences in the final product might be of relevance e.g. for the insertion of regulatory elements such as promoters and termination sequences or for the in-frame linkage of the individual target sequences.

[00138] For the linkage of two nucleotide sequences of interest the linkage tail may be added either to the reverse primer(s) or forward primer(s) of each primer set used for the PCR amplification of each target sequence. The present invention exemplifies addition of overlap-extension tails and tails adapted for linkage by ligation, to the V_H and V_L forward primers of each set (e.g. Figure 2 and example 9, respectively). This results in a linking direction of the products that is 5' to 5' (head-to-head and bi-directional). However, linkage tails might as well be added to the reverse primer(s) of each set (e.g. C_κ and/or C_λ in the first set and C_H or J_H in the second set). This results in a linking direction of the product that is 3' to 3' (tail-to-tail and bi-directional). A third option is adding the linkage tails to the reverse primer(s) of the first primer set (e.g. C_κ and/or C_λ primers) and the forward primer(s) of the second primer set (e.g. V_H primer(s)) or visa versa. This results in a 3' to 5' orientation (head-to-tail and uni-directional). Figure 3 illustrates the possible directions that can be generated depending on which primer of each primer set that is equipped with the linkage tail.

[00139] When linking more than two nucleotide sequences of interest some of the primer sets need to have linkage tails on both the forward and reverse primers, such that one tail is complementary to a tail of the preceding primer set and the other tail is complementary to one of the primers of the subsequent primer set. This principle holds for all the primer sets that amplify target sequences that are to be linked between two other target sequences.

[00140] The design of the gene-specific primer part generally should observe known primer design rules such as minimizing primer dimerization, hairpin formation and non-specific annealing. Further, multiple G or C nucleotides as the 3' bases are to be avoided when possible. The melting temperature (Tm) of the gene-specific regions in a primer set should preferably be equal to each other plus/minus 5°C. In the present invention Tm values between 45°C and 75°C are desirable and Tm values of about 60°C are optimal for most applications. Advantageously, the initial primer design can be aided by computer programs developed for

this task. However, primer designs generally need laboratory testing and routine optimization. This may be done, for example, by analyzing size, restriction fragment length polymorphism (RFLP) and sequencing of the amplification products obtained using the primer sets. The use of degenerate positions within primers is a useful approach when amplifying sequences with variable regions or when searching for new family members belonging to a specified class of proteins. The numbers of degenerate positions may also require optimization.

[00141] The present invention encompasses improved primer sets that can be used together in a highly multiplexed format. The primer set described by de Haard (de Haard, H.J. *et al.* 1999. J. Biol. Chem. 274, 18218-18230) were used as starting point, and was modified by trimming the 3' ends of the primers to reduce non-specific interactions and adding overlap-extension tails or tails adapted for linkage by ligation.

[00142] One feature of the present invention, are primer mixes composed of at least two primer sets that are able to prime amplification and promote linkage of at least two nucleotide sequences of interest. The primer mixes of the present invention are capable of priming the amplification of at least two subunits or domains from heteromeric proteins, e.g. belonging to the class of enzymes, inhibitors, structural proteins, toxins, channel proteins, G-proteins, receptor proteins, immunoglobulin superfamily proteins, transportation proteins etc.

[00143] A further feature of the present invention is a multiplex overlap-extension primer mix comprising primer sets wherein at least one primer set member of each primer set comprises an overlap-extension tail capable of hybridizing to the overlap-extension tail of a primer set member of a second primer set.

[00144] The overlap-extension tails enables the immediate linkage of the nucleotides of interest during the multiplex overlap-extension PCR amplification by equipping each individual product arising from the primer sets with a tail that is complementary to an adjoining product. This however does not mean that the linkage necessarily occur during this first PCR amplification. Depending on the reaction setup, the majority of the actual linkage may be performed during an additional amplification with the outer primers of the first PCR amplification (multiplex PCR amplification).

[00145] A further feature of the present invention, is a primer set designed to amplify a family of nucleotide sequences containing variable region encoding sequences. Examples of such families are kappa light chains (e.g. VKI-VI in humans), lambda light chains (e.g. VL1-10 in humans) and variable heavy chains (e.g. VH1-7 in humans and VH1-15 in mice) from

immunoglobulins, and α, β, γ or δ TcR variable regions. A primer set for the amplification of a family of nucleotide sequences containing variable region encoding sequences often comprise a plurality of primers where several primers can be degenerate primers. Amplification of families of immunoglobulin light chain variable region encoding sequences is for example performed using a primer set comprised of a plurality of primers complementary to the 5' end of the variable region of the kappa chain ($V_{L\kappa}$ primer(s)) or the kappa leader sequence ($V_{L\kappa L}$ primer(s)) and/or the lambda chain ($V_{L\lambda}$ primer(s)) or the lambda leader sequence ($V_{L\lambda L}$ primer(s)) (forward primers) together with constant region kappa (C_κ primer(s)) and/or lambda primers (C_λ primer(s)) (reverse primers) or a plurality of such primers. Alternatively, light chain joining region primers ($J_{L\kappa}$ and/ or $J_{L\lambda}$ primer(s)) may be used as reverse primers instead of the constant region primers. Alternatively, forward primers anneal in the UTR region preceding the leader sequence of the variable light chain. Equally, families of immunoglobulin heavy chain variable region encoding sequences can be amplified with one primer set utilizing various primer combinations. For example, a plurality of primers complementary to the 5' end of the heavy chain variable region (V_H primer(s)) or the leader sequence of this region (V_{HL} primer(s)) (forward primers) together with a plurality of heavy chain joining region primers (J_H primer(s)) or heavy chain constant region primer(s) (reverse primers). The C_H primer may be isotype-specific and in principle any C_H primer can be utilized (e.g. C_{H1}, C_{H2}, C_{H3}, or C_{H4}), also one that would result in a full-length heavy chain. Alternatively, forward primers anneal in the UTR region preceding the leader sequence of the variable heavy chain.

[00146] The use of forward primers annealing in the leader sequence instead of the 5' end of the variable region is particular useful if cross-hybridization is observed for the variable region primers. Since mutations due to cross-hybridization will be eliminated from the final protein because leader sequences are cleaved off during protein processing within the cell. Example 9 describes the design of antibody variable heavy chain and kappa light chain leader primers. The aspect that the site of priming is located in the 3' end of the leader encoding sequence (C-terminal) is an advantage over previous antibody leader primers, since this allows for shuttling of the amplified sequences between bacterial and eukaryotic expression vectors in a way that allow functional leader sequences in both system. The design system described in Example 9 can easily be applied to antibody lambda light chains, TcR α, β, γ or δ chains as well.

[00147] One feature of the present invention are primers which anneal in the 3' end of the leader encoding sequence preceding a variable region encoding sequence, and their use for amplification of variable region encoding sequences.

[00148] A preferred feature of the present invention is the application of primers with at least 90% sequence identity (preferably at least 95% identity) with the gene-specific region of SEQ ID NO 86 to 92, which correspond to primers annealing in the C-terminal of heavy chain leader encoding sequences (V_{HL} primers). The gene-specific sequence of these SEQ ID NO's correspond to base number 18 to the 3' end of the sequences (see also table 11).

[00149] Another preferred feature of the present invention is the application of primers with at least 90% sequence identity (preferably at least 95% identity) with the gene-specific region of SEQ ID NO 93 to 98, which correspond to primers annealing in the C-terminal of kappa light chain leader encoding sequences ($V_{L\kappa L}$ primers). The gene-specific sequence of these SEQ ID NO's correspond to base number 25 to the 3' end of the sequences (see also table 11).

[00150] In one embodiment of the present invention, the multiplex overlap-extension primer mix utilized for the multiplex overlap-extension PCR and possibly for the reverse transcription step as well comprises a) at least one C_L or J_L primer complementary to the sense strand of an immunoglobulin light chain region encoding sequence; b) at least one V_L 5' primer or V_L leader primer complementary to the antisense strand of an immunoglobulin light chain variable region encoding sequence and capable of forming a primer set with the primer(s) in a); c) at least one C_H or J_H primer complementary to the sense strand of an immunoglobulin constant heavy chain domain encoding sequence or the heavy chain joining region and d) at least one V_H 5' primer or V_H leader primer complementary to the antisense strand of an immunoglobulin heavy chain variable region encoding sequence, and capable of forming a primer set with the primer(s) in c).

[00151] Primer sets of the present invention can for example be $V_{L\kappa} + C_\kappa$, $V_{L\lambda} + C_\lambda$, $V_{L\kappa} + J_{L\kappa}$, $V_{L\lambda} + J_{L\lambda}$, $V_{L\kappa L} + C_\kappa$, $V_{L\lambda L} + C_\lambda$, $V_{L\kappa L} + J_{L\kappa}$, $V_{L\lambda L} + J_{L\lambda}$, $V_H + J_H$, $V_H + C_H$, $V_{HL} + J_H$ or $V_{HL} + C_H$ or combinations of these, capable of amplifying a variable region encoding target sequence.

[00152] In a further embodiment are the C_L (J_L) primer(s) and V_L (V_{LL}) primer(s) adapted for amplifying sequences comprising kappa light chain variable regions or lambda light chain variable regions.

[00153] In a preferred embodiment of the present invention the are the C_L (J_L) primer(s) and V_L (V_{LL}) primer(s) adapted for amplifying both kappa and lambda light chain variable region encoding sequences.

[00154] In an even more preferred embodiment of the present invention, are the forward primers for light chain amplification V_{LL} primers with at least 90% sequence identity (preferably at least 95% identity) with the gene-specific region of SEQ ID 93 to 98, and the forward primers for heavy chain amplification are V_{HL} primers with at least 90% identity (preferably at least 95%) with the gene-specific region of SEQ ID 86 to 92.

[00155] In a further embodiment of the present invention, carries the immunoglobulin V_L/ V_{LL} and V_H/V_{HL} primers linkage tails, preferably in the form of complementary overlap-extension tails. This generates variable region encoding sequences that are linked in a head-to-head fashion. For the linkage of variable region encoding sequences in a head-to-tail fashion, either the C_L/J_L and V_H/V_{HL} primers contain linkage tails or the V_L/V_{LL} and C_H/ J_H primers contain linkage tails, preferably in the form of complementary overlap-extension tails. For the linkage of variable region encoding sequences in a tail-to-tail fashion, the C_L/J_L and C_H/J_H primers contain linkage tails, preferably in the form of complementary overlap-extension tails (Figure 3).

[00156] Preferentially, the multiplex primer mixes, including multiplex overlap-extension primer mixes, of the present invention comprise two primer sets. Thus, a multiplex primer mix comprises at least four different primers. In a further aspect of the present invention a multiplex primer mix comprises more than four different primers. A multiplex primer mix of the present invention is used for the amplification of target sequences in a single vessel. For example are kappa, lambda and heavy chain variable regions all amplified in the same vessel. One multiplex primer mix of the present invention was comprised of 16 different degenerate primers distributed as follows: eight V_H primers, one C_{H1} primer, six $V_{L\kappa}$ primers, and one C_κ primer (Figure 2). Another set was comprised of 19 degenerate primers distributed as follows: eight V_H primers, four J_H primers, six $V_{L\kappa}$ primers, and one C_κ primer. A third set was comprised of 22 degenerate primers distributed as follows: eight V_H primers, one C_{H1} primer, eleven $V_{L\lambda}$ primers, and two C_λ primers. A fourth set was comprised of 27 degenerate primers distributed as follows: eight V_H primers, one C_{H1} primer, six $V_{L\kappa}$ primers, one C_κ primer, eleven $V_{L\lambda}$ primers, and two C_λ primers.

[00157] The present invention also encompasses primers for an additional PCR amplification of the linked products obtained by multiplex RT-PCR followed by linkage by ligation or

recombination or by multiplex overlap-extension RT-PCR. This additional PCR amplification can be performed using a primer mix adapted for amplifying the linked target sequences. Such a primer mix may comprise the outer primers of the multiplex primer mix or multiplex overlap-extension primer mix, meaning the primers that anneals to the outermost 5' end and 3' end of the sense strand of the linked nucleotide sequences, thereby enabling the amplification of the entire linked product. One example of primers that can be utilized as outer in the present invention are the C_κ/J_κ and/or C_λ/J_λ primers forming a primer set with the J_H or C_H primers. This process generally serves to increase the amount of linked product obtained from the multiplex RT-PCR followed by linkage by ligation or recombination or from the multiplex overlap-extension RT-PCR.

[00158] Alternatively, a primer set which is nested compared to the outer primers used in the primary multiplex RT-PCR or multiplex overlap-extension RT-PCR reaction can be used for the additional amplification of the linked nucleotide sequences. In the present invention such a primer set is termed a nested primer set. The design of nested primers generally observes the same design rules as for the gene-specific primers previously described, except that they prime 3' to the annealing position of the outer primers used in the multiplex RT-PCR or multiplex overlap-extension RT-PCR. The product resulting from a nested PCR is therefore shorter than the linked product obtained by the multiplex RT-PCR followed by linkage by ligation or recombination or by multiplex overlap-extension RT-PCR. In addition to increase the amount of linked product, the nested PCR further serves to increase the overall specificity, especially of the multiplex overlap-extension RT-PCR technology. However, it should be noted that not all multiplex primer mixes/multiplex overlap-extension primer mixes that have been described previously are suitable for combination with a nested primer set when performing the additional amplification. In such cases the outer primers of the multiplex primer mix/multiplex overlap-extension primer mix can be used for the additional amplification or a semi-nested PCR can be applied as described later.

[00159] In one embodiment of the present invention, a mixture of J_L and J_H primers is used as nested primers for the additional amplification of the linked immunoglobulin variable region encoding sequences.

[00160] Nested primer sets of the present invention can also be comprised of a reverse (or forward) outer primer(s) from the first multiplex primer mix/multiplex overlap-extension primer mix and a second nested primer that prime 3' to the annealing position of the forward (or reverse) outer primer(s) of the first multiplex primer mix/multiplex overlap-extension

primer mix. The use of such a primer set for an additional PCR amplification is generally known as a semi-nested PCR. Semi-nested PCR can for example be applied if it is difficult to design a nested primer in one specific region e.g. for the variable region sequences (V and J primers), because such a primer would have to anneal in the complementarity determining regions (CDRs). Further, semi-nested PCR can be used when it is desirable to keep one end of the linked sequences intact e.g. for cloning purposes.

[00161] One feature of the present invention keeps the constant light chain encoding sequence kept intact during the additional amplification reaction. The C_L (reverse) primer(s) used for the additional amplification is only slightly modified compared to the outer primer(s) used for the primary reaction (the multiplex RT-PCR or multiplex overlap-extension RT-PCR reaction). The modification comprises the addition of a few bases to the 3' end of the C_L primer(s) used for the primary amplification. Further, a different cloning tail might be added to the nested C_L primer(s). The forward primer(s) used in the additional amplification is completely nested compared to the constant heavy chain-specific outer primer(s) used in the primary reaction. The combined use of the outer primer(s) in the primary reaction and the slightly modified nested C_L primer(s) together with the completely nested forward primer(s) in the additional amplification results in an increase in specificity that is comparable to what is achievable with a nested PCR using a completely nested primer set.

[00162] In a preferred embodiment of the present invention, the nested PCR is performed with J_H primer(s) and modified C_L primer(s) where between 2 and 10 gene-specific base pairs have been added at the 3'end compared to the first C_L primer(s) of the multiplex primer mix/multiplex overlap-extension primer mix.

Optimization of Multiplex Overlap-extension PCR

[00163] The parameters of the multiplex overlap-extension PCR step of both the two-step and the single-step procedure can be optimized on several parameters (see, for example, Henegariu, O. *et al.* 1997. BioTechniques 23, 504-511; Markoulatos, P. *et al.* 2002. J. Clin. Lab. Anal. 16, 47-51). Generally the same optimization parameters apply for multiplex RT-PCR, although the ratio between outer and inner primers is less important for such a reaction.

a. Primer Concentration

[00164] The concentration of the primers carrying the overlap-extension tail (for example the V_H and V_L primers) is preferably lower than the concentration of the outer primers without overlap-extension tail (for example J_H and kappa primers).

[00165] If one of the target sequences amplifies with a lower efficiency than the others, for example, as a result of a higher GC%, it may be possible to equalize the amplification efficacy. This may be done by using a higher concentration of the primer set which mediates amplification with low efficiency, or lowering the concentration of the other primer set. For example, sequences encoding for heavy chain variable regions tend to have a higher GC% and hence lower amplification efficiency than light chain variable regions. This points towards using V_L primers at a lower concentration than the V_H primers.

[00166] Further, when using a large number of primers the total primer concentration might be an issue. The upper limit is determined experimentally by titration experiments. For the "AmpliTaq Gold" PCR system from Applied Biosystems the upper limit was found to be 1.1 μM total oligonucleotide concentration, for other systems it may however be as high as 2.4 μM. Such an upper limit of total oligonucleotide concentration influences the maximal concentration of individual primers. If the individual primer concentration is too low it is likely to cause a poor PCR sensitivity.

[00167] The quality of the oligonucleotide primers have also been found to be important for the multiplex overlap-extension PCR. HPLC-purified oligonucleotides, have produced the best results.

b. PCR Cycling Conditions:

[00168] Preferentially the cycling conditions are as follows:

Denaturation:	10-30 s	94°C	
Annealing:	30-60 s	50-70°C	Approximately 5°C below Tm of primers.
Extension:	1 min x EPL	65-72°C	EPL is Expected Product Length in kb.
Cycle number:	30-80		
Final extension:	10 min	65-72°C	

[00169] For the single-step multiplex overlap-extension RT-PCR the following steps were built into the cycling program prior to the amplification cycling outlined above:

Reverse transcription:	30 min	42-60°C	These conditions are also used where separate reverse transcription is performed.
Polymerase activation:	10-15 min	95°C	Hot-start polymerases are favorable in single-step RT-

PCR. Activation according to manufacturer.

[00170] It is possible to optimize on all these parameters. Especially the annealing temperature is important. Thus, initially all the individual primer sets that are to constitute the final primer mix should be tested separately in order to identify optimal annealing temperature and time, as well as elongation and denaturing times. This will give a good idea about the window within which these parameters can be optimized for the multiplex overlap-extension primer mix.

[00171] Problems with poor PCR sensitivity, for example due to low primer concentration or template concentration can be overcome by using a high number of thermal cycles. A high number of thermal circles constitute between 35 and 80 cycles, preferably around 40 cycles.

[00172] Further, longer extension times can improve the multiplex overlap-extension PCR process. Long extension times constitute 1.5 – 4 min x EPL compared to the normal 1 min extension.

c. Use of Adjuvants

[00173] Multiplex PCR reactions can be significantly improved by using a PCR additive, such as DMSO, glycerol, formamide, or betaine, which relax DNA, thus making template denaturation easier.

d. dNTP and $MgCl_2$

[00174] Deoxynucleoside triphosphate (dNTP) quality and concentration is important for the multiplex overlap-extension PCR. The best dNTP concentration is between 200 and 400 μM of each dNTP (dATP, dCTP, dGTP and dTTP), above which the amplification is rapidly inhibited. Lower dNTP concentrations (100 μM of each dNTP) suffice to achieve PCR amplification. dNTP stocks are sensitive to thawing/freezing cycles. After three to five such cycles, multiplex PCR often do not work well. To avoid such problems, small aliquots of dNTP can be made and kept frozen at -20°C.

[00175] Optimization of Mg^{2+} concentration is critical since most DNA polymerases are magnesium-dependent enzymes. In addition to the DNA polymerase, the template DNA primers and dNTP's bind Mg^{2+}. Therefore, the optimal Mg^{2+} concentration will depend on the dNTP concentration, template DNA, and sample buffer composition. If primers and/or template DNA buffers contain chelators such as EDTA or EGTA, the apparent Mg^{2+} optimum may be altered. Excessive Mg^{2+} concentration stabilizes the DNA double strand and

prevents complete denaturation of DNA, which reduces yield. Excessive Mg^{2+} can also stabilize spurious annealing of primer to incorrect template sites, thereby decreasing specificity. On the other hand, an inadequate Mg^{2+} concentration reduces the amount of product.

[00176] A good balance between dNTP and $MgCl_2$ is approximately 200 to 400 μM dNTP (of each) to 1.5 to 3 mM $MgCl_2$.

e. PCR Buffer Concentration

[00177] Generally KCl based buffers suffice for multiplex overlap-extension PCR; however, buffers based on other components such as $(NH_4)_2SO_4$, $MgSO_4$, Tris-HCl, or combinations thereof may also be optimized to function with the multiplex overlap-extension PCR. Primer pairs involved in the amplification of longer products work better at lower salt concentrations (e.g. 20 to 50 mM KCl), whereas primer pairs involved in the amplification of short products work better at higher salt concentrations (e.g. 80 to 100 mM KCl). Raising the buffer concentration to 2X instead of 1X may improve the efficiency of the multiplex reaction.

f. DNA Polymerase

[00178] The present invention is exemplified with Taq polymerase. Alternatively, other types of heat-resistant DNA polymerases including, for example, Pfu, Phusion, Pwo, Tgo, Tth, Vent, Deep-vent may be used. Polymerases without or with 3´to 5´exonuclease activity may either be used alone or in combination with each other.

Vectors and Libraries

[00179] The linkage of nucleotide sequences of interest according to the present invention produces a nucleotide segment comprising the linked nucleotide sequences of interest. Further, libraries of such linked nucleic acid sequences of interest are produced by the methods of the present invention, in particular libraries of variable region encoding sequences.

[00180] One feature of the present invention is the insertion of a segment containing linked nucleotide sequences of interest or a library of linked nucleotide sequences of interest, generated by a method of the present invention, into suitable vectors. The libraries may be combinatorial libraries or libraries of cognate pairs of variable region encoding sequences. The restriction sites generated by the outer primers, nested primers or semi-nested primers are preferably designed to match appropriate restriction sites of the vector of choice. The linked nucleic acid sequences of interest can also be inserted into vectors by recombination,

if one of the semi-nested, nested primers or outer primers were equipped with a suitable recombination site and the vector of choice contains one as well.

[00181] Basically there are no limitations to the vectors that can be used as carriers of the products generated by one of the multiplex RT-PCR-linkage methods of the present invention. Vectors of choice may be those suitable for amplification and expression in cells including, for example, bacteria, yeast, other fungi, insect cells, plant cells, or mammalian cells. Such vectors may be used to facilitate further cloning steps, shuttling between vector systems, display of the product inserted into the vector, expression of the inserted product and/or integrate into the genome of a host cell.

[00182] Cloning and shuttle vectors are preferably bacterial vectors. However, the other types of vectors may also be applied in cloning and shuttle procedures.

[00183] Display vectors can for example be phage vectors or phagemid vectors originating from the class of fd, M13, or f1 filamentous bacteriophages. Such vectors are capable of facilitating the display of a protein including, for example, a binding protein or a fragment thereof, on the surface of a filamentous bacteriophage. Display vectors suitable for display on ribosomes, DNA, yeast cells or mammalian cells are also known in the art. These comprise for example viral vectors or vectors encoding for chimeric proteins.

[00184] Expression vectors exist for all the mentioned species and the one to be chosen completely depend on the protein to be expressed. Some expression vectors are additionally capable of integrating into the genome of a host cell either by random integration, or by site-specific integration, utilizing appropriate recombination sites. Expression vectors may be designed to provide additional encoding sequences that, when the linked product is inserted in-frame to these sequences, enable the expression of a larger protein, e.g. a full-length monoclonal antibody, when introduced into an appropriate host cell. This in-frame insertion may also facilitate the expression of chimeric proteins that facilitate the display on the surface of a filamentous bacteriophage or cell. In a bacteriophage display system, the linked nucleotide sequences of interest may be inserted in-frame to a sequence encoding a coat protein such as pIII or pVIII (Barbas, C.F. *et al.* 1991. Proc. Natl. Acad. Sci. USA 88, 7978-7982; Kang, A.S. *et al.* 1991. Proc. Natl. Acad. Sci. USA 88, 4363-4366).

[00185] In one embodiment of the present invention, the individual segments of linked nucleotide sequences of interest is comprised of an immunoglobulin heavy chain variable region encoding sequence associated with a light chain variable region encoding sequence. Preferably, these linked sequences are inserted into a vector that contains sequences encoding

one or more immunoglobulin constant domains. The insertion is engineered such that the linked heavy chain variable region and/or light chain variable region encoding sequences are inserted in-frame to the constant region encoding sequences. Such an insertion can for example generate a Fab expression vector, a full-length antibody expression vector or an expression vector encoding a fragment of a full-length antibody. Preferentially such a vector is an expression vector suitable for screening (e.g. *E. coli*, phagemid, or mammalian vectors) and the constant region heavy chain encoding sequences are chosen from the human immunoglobulin classes IgGl, IgG2, IgG3, IgG4, IgM, IgA1, IgA2, IgD, or IgE, thereby enabling the expression of a Fab or full-length recombinant antibody. In addition to the constant heavy chain encoding sequences the vector may also contain a constant light chain encoding sequence chosen from human lambda or kappa chains. This is appropriate when the linked nucleotide sequences only encode the immunoglobulin variable region encoding sequences (Fv's).

[00186] In another embodiment of the present invention, the individual segments of the linked nucleotide sequences is comprised of a TcR α chain variable region encoding sequence associated with a β chain variable region encoding sequence or a γ chain variable region encoding sequence associated with a δ chain variable region encoding sequence. Preferably, these linked sequences are inserted into a vector that contains sequences encoding one or more TcR constant domains. The insertion is engineered such that the inserted linked variable region encoding sequences are in-frame to the corresponding TcR constant region encoding sequences. In a further embodiment, such a vector is a chimeric expression vector comprising sequences that encode a leucine zipper in-frame to the TcR constant regions. It has been shown that such constructs increase stability of soluble TcR's (Willcox, B.E. et al. 1999. Protein Sci 8, 2418-2423).

[00187] Libraries of cognate pairs of the present invention may be introduced into vectors by two different approaches. In the first approach, the single cognate pairs are inserted individually into a suitable vector. This library of vectors may then either be kept separate or be pooled. In the second approach, all the cognate pairs are pooled prior to vector insertion, followed by in-mass insertion into suitable vectors generating a pooled library of vectors (illustrated in Figure 12). Such a library of vectors comprises a large diversity of pairs of variable region encoding sequences.

[00188] One aspect of the present invention is a library of cognate pairs of linked variable region encoding sequences. Preferably the individual cognate pairs of the library comprise an

immunoglobulin light chain variable region encoding sequence associated with a heavy chain variable region encoding sequence.

[00189] Another preferred library of cognate pairs comprise linked TcR region encoding sequences, where each individual TcR region encoding sequences comprise an alpha chain variable region encoding sequence associated with a beta chain variable region encoding sequence and/or a TcR gamma chain variable region encoding sequence associated with a delta chain variable region encoding sequence.

[00190] An embodiment of the present invention is a sub-library of cognate pairs of linked variable region encoding sequences which encode for desired binding specificities directed against a particular target. Preferably these cognate pairs comprise linked immunoglobulin light chain variable region and heavy chain variable region encoding sequences, TcR alpha chain variable region and beta chain variable region encoding sequences and/or TcR gamma chain variable region and delta chain variable region encoding sequences.

[00191] A further embodiment is a sub-library selected from a parent library of cognate pairs of variable region encoding sequences as described throughout the invention.

[00192] A preferred embodiment of the present invention is a library or sub-library encoding for cognate pairs of full-length immunoglobulins selected from human immunoglobulin classes IgA1, IgA2, IgD, IgE, IgG1, IgG2, IgG3, IgG4, or IgM.

[00193] Another preferred feature of the present invention is a library or sub-library encoding for soluble and stable cognate pairs of TcRs.

[00194] A feature of the present invention is the diversity of said libraries, which are comprised of at least 5, 10, 20, 50, 100, 1000, 10^4, 10^5 or 10^6 different cognate pairs.

[00195] In a further embodiment of the present invention, said libraries of cognate pairs of linked variable region encoding sequences are obtainable by a method comprising the steps described herein. This library is also termed the parent library.

Screening and Selection

[00196] The parent library of pairs of linked variable region encoding sequences isolated from a donor, utilizing one of the methods of the present invention, is expected to represent a diversity of binding proteins of which some will be irrelevant, i.e. not binding to a desired target, in particular for combinatorial libraries. Therefore, the present invention encompasses enrichment and screening, for a sub-library encoding a subset of diversities of binding specificities directed against a particular target.

[00197] For libraries of cognate pairs the diversity of the library is expected to represent the diversity present in the donor material, with only a minor number of randomly linked variable regions. Thus, an enrichment step may not be necessary prior to the screening for target-specific binding affinities in a library composed of cognate pairs.

[00198] In a further embodiment of the present invention, the method of generating a library of pairs of linked variable region encoding sequences, further comprises creating a sub-library by selecting a subset of pairs of linked variable region sequences that encode binding proteins with a desired target specificity. Such a selection of linked variable region encoding sequences, is also termed a library of target-specific cognate pairs.

[00199] In a preferred embodiment of the present invention, is the library of target-specific cognate pairs of variable region encoding sequences transferred to a mammalian expression vector.

[00200] Immunological assays are generally suitable for the selection of target-specific immunoglobulin variable region encoding sequences. Such assays are well know in the art and constitute for example ELISPOTS, ELISA, membrane assays (e.g. Western blots), arrays on filters, or FACS. The assays can either be performed in a direct manner, utilizing the polypeptides produced from the immunoglobulin variable region encoding sequences. Alternatively, the immunoassays can be performed in combination with or following enrichment methods such as phage display, ribosome display, bacterial surface display, yeast display, eukaryotic virus display, RNA display or covalent display (reviewed in FitzGerald, K., 2000. Drug Discov. Today 5, 253-258). As illustrated in Figure 10, both cognate Fab expression libraries and cognate full-length antibody expression libraries can be subjected to screening, thereby generating a sub-library of positive clones. Such screening assays and enrichment procedures are also suitable for Fv or scFv fragments or combinatorial libraries of linked variable regions.

[00201] In a preferred embodiment of the present invention, the selection of a sub-library of target-specific cognate pairs or combinatorial pairs of variable region encoding sequences is performed by using a high-throughput screening assay. High-throughput screening assays could be, but are not restricted to, ELISA assays performed with semi-automated or fully automated equipment. It could also be a membrane assays in which bacteria are robotically picked and gridded onto an appropriate membrane on top of agar plates generating arrays of colonies expressing antigen-binding molecules. The molecules are secreted through the membrane onto a second underlying antigen-coated membrane which can be developed

separately and used to identify clones that secrete antigen binding molecules towards the desired target (de Wildt, R.M., *et al.* 2000. Nat. Biotechnol. 18, 989-994).

[00202] When a sub-library of cognate pairs or combinatorial pairs of antigen-binding clones has been selected by an appropriate technology it is possible to perform an additional analysis by DNA sequencing of the linked immunoglobulin light chain variable region and heavy chain variable region encoding sequences. First of all such a DNA sequencing will provide information about the library diversity such as germline origin, family distribution and maturation within the CDR regions. Such an analysis will enable the selection of clones which represent a broad diversity, and leaving out repeated clones. Secondly, DNA sequencing will reveal mutations introduced during the isolation process.

[00203] When analyzing variable region encoding sequences there are three types of mutations to consider when assessing whether a mutation is acceptable: i) The most frequent type of mutations result from intra-family cross-priming, where V gene primers prime to the wrong subset within one particular V gene family. The changes introduced are mainly substitutions of naturally occurring codons at one particular position. Due to the high degree of sequence homology within a V gene family these changes usually can be regarded as conservative and acceptable changes; ii) Less frequent mutations are induced by inter-family cross-priming (e.g., a VH3 family primer primes a VH1 family encoding sequence) and induces more significant structural changes, sometimes with no natural counterpart. Such changes could potentially affect the immunogenicity of the variable region by creating new epitopes. Such changes can easily be identified and subsequently repaired using standard molecular biological techniques or the clones can be excluded from the library; iii) Errors created by the Taq DNA polymerase are most easily identified in the constant region encoding sequences and can easily be eliminated. However, Taq induced mutations will of course also be present in the variable region encoding sequences where they are indistinguishable from the naturally occurring somatic mutations, which are also the result of random mutations in the variable region encoding sequences. Considering that the mutations are non-systematic and only affect particular pairs in distinct ways, it appears reasonable to disregard such changes.

[00204] Further the sequence analysis can be used to identify the degree of scrambling in a cognate pair library, as illustrated in table 20 for VH group H4.

[00205] As described in Example 9 the presence of the mutations described in i) and ii) can be circumvented in the expression library, when utilizing primers that anneal in the leader

sequence of the variable region encoding sequences instead of primers that anneal in the 5' region of the variable regions.

[00206] In a further embodiment of the present invention, the sub-library of target-specific and possibly sequence analysed pairs of linked immunoglobulin light chain variable region and heavy chain variable region encoding sequences are transferred to a mammalian expression vector. Such a transfer can be performed into any of the vectors described in the previous section, enabling the expression of a full-length recombinant antibody. If the screening is performed with a mammalian cognate full-length antibody expression library such a transfer may not be needed.

[00207] In another embodiment of the present invention, the parent library is generated from a lymphocyte-containing cell fraction which is enriched for T lymphocytes. The pairs of linked variable region encoding sequences constituting the parent library, may be selected for encoding a subset of pairs of linked variable region sequences, composed of alpha and beta and/or gamma and delta chains that encode binding proteins with a desired target specificity, generating a sub-library of cognate pairs or combinatorial pairs. Antigen-specific T cell receptors can subsequently be identified from a pool of transfected cells using standard methodology such as staining with tetrameric MHC-peptide complexes (e.g., Callan, M.F. *et al.* 1998. J. Exp. Med. 187, 1395-1402; Novak, E.J. *et al.* 1999. J. Clin. Invest 104, R63-R67), by measuring cellular responses in the form of IL-2 release or by more sophisticated means such as yeast or retroviral display techniques.

Host cells and Expression

[00208] The libraries of the present invention can be transferred to vectors suitable for expression and production of proteins encoded from the linked nucleic acid sequences of interest, in particular variable region containing binding proteins or fragments thereof. Such vectors are described in the *Vectors and Libraries* section, and provide for the expression of for example full-length antibodies, Fab fragments, Fv fragments, scFv, membrane bound or soluble TcRs or TcR fragments of a species of choice.

[00209] One feature of the present invention is the introduction into a host cell of a library or a sub-library of vectors of cognate pairs of linked variable region encoding sequences or a single clone encoding a cognate pair of linked variable region encoding sequences, for amplification and/or expression. Host cells can be chosen from bacteria, yeast, other fungi, insect cells, plant cells, or mammalian cells. For expression purposes mammalian cells, such

as Chinese hamster ovary (CHO) cells, COS cells, BHK cells, myeloma cells (e.g., Sp2/0 cells, NS0), NIH 3T3, fibroblast or immortalized human cells such as HeLa cells, HEK 293 cells, or PER.C6 are preferred.

[00210] The introduction of vectors into host cells may be accomplished by a number of transformation or transfection methods known to those skilled in the art, including calcium phosphate precipitation, electroporation, microinjection, liposome fusion, RBC ghost fusion, protoplast fusion, viral infection and the like. The production of monoclonal full-length antibodies, Fab fragments, Fv fragments and scFv fragments is well known.

[00211] The production of recombinant polyclonal antibodies to be used for treatment is a quite new area. A recombinant polyclonal manufacturing technology has been described in PCT application WO 2004/061104. In brief, this technology involves the generation of a collection of cells, suitable as a manufacturing cell line. The following description of the technique is made for a library of cognate pairs, it is however just as applicable for a combinatorial library. The individual cells in the collection of cells are capable of expressing a distinct member of the recombinant polyclonal binding protein for example from a library of cognate pairs. In order to ensure that the individual cells express a single cognate pair and not several cognate pairs of the polyclonal binding protein, the nucleic acid sequences encoding the cognate pairs are introduced into a single site-specific site in the genome of each individual cell. This is an important feature of the collection of cells, since this prevent scrambling of the heavy and light chains expressed from each cell, but also because it generates cells that are virtually identical to one another, except for the small differences in the variable regions of the individual cognate pairs. This trait will enable an unbiased growth of the collection of cells over the period of time necessary for the production. To ensure single site-specific integration, a host cell line with only one integration site should be used, these are commercially available e.g. Invitrogen's CHO Flp-In cells containing a single FRT site. Appropriate vectors for this cell line contain a corresponding FRT site and are introduced into the genome using the Flp recombinase. There are several other known recombinases e.g. Cre, beta-recombinase, Gin, Pin, PinB, PinD, R/RS, lambda integrase, or phage ΦC31 integrase that can be used in combination with their corresponding recombination sites. Further, appropriate vectors contain a selection marker that enables the selection of site-specific integrants.

[00212] The generation of a polyclonal manufacturing cell line and the production of a recombinant polyclonal protein from such a cell line can be obtained by several different transfection and manufacturing strategies.

[00213] One way, is to use a library of vectors mixed together into a single composition, for the transfection of a host cell line with a single integration site per cell. This method is termed bulk transfection or transfection in bulk. Generally, the vector and host cell design previously described will ensure that a polyclonal cell line capable of unbiased growth will be obtained upon appropriate selection. A frozen stock of the polyclonal cell line will be generated before initiation of the recombinant polyclonal protein manufacturing.

[00214] Another way, is to use a library of vectors split into fractions, containing approximately 5 to 50 individual vectors of the library in a composition, for transfection. Preferably, a fraction of the library constitutes 10 to 20 individual vectors. Each composition is then transfected into an aliquot of host cells. This method is termed semi-bulk transfection. The number of aliquots transfected will depend on the size of the library and the number of individual vectors in each fraction. If the library for example constitutes 100 distinct cognate pairs, which are split into fractions containing 20 distinct members in a composition, 5 aliquots of host cells would need to be transfected with a library composition constituting a distinct fraction of the original library. The aliquots of host cells are selected for site-specific integration. Preferably, the distinct aliquots are selected separately. However, they can also be pooled before selection. The aliquots can be analyzed for their clonal diversity and only those with sufficient diversity will be used to generate a polyclonal cognate pair library stock. To obtain the desired polyclonal cell line for manufacturing, the aliquots can be mixed before generating the freezing stock, immediately after they have been retrieved from the stock or after a short proliferation and adaptation time. Optionally, the aliquots of cells are kept separate throughout production, and the polyclonal protein composition is assembled by combining the products of each aliquot rather than the aliquots of cells before production.

[00215] A third way, is a high throughput method in which host cells are transfected separately using the individual vectors constituting the library of cognate pairs. This method is termed individual transfection. The individually transfected host cells are preferably selected for site specific integration separately. The individual cell clones generated upon selection may be analyzed with respect to proliferation time and preferably, those with similar growth rates are used to generate a polyclonal cognate pair library stock. The individual cell clones can be mixed to obtain the desired polyclonal cell line before

generating the stock, immediately after they have been retrieved from the stock, or after a short proliferation and adaptation time. This approach may eliminate any possible residual sequence bias during transfection, integration and selection. Alternatively the individually transfected host cells are mixed before selection is performed, this will enable control of sequence bias due to transfection.

[00216] A shared feature in the manufacturing strategies outlined in the above is that all the individual cognate pairs constituting the recombinant polyclonal protein can be produced in one, or a limited number of bioreactors. The only difference is the stage at which one chooses to generate the collection of cells that constitutes the polyclonal manufacturing cell line.

[00217] One embodiment of the present invention, is a population of host cells comprising a cognate library or sub-library of linked pairs of variable region encoding sequences.

[00218] In a further embodiment, a population of host cells comprises a library obtained from a population of isolated single cells constituting lymphocytes, utilizing the multiplex RT-PCR amplification followed by linkage by ligation or recombination or the multiplex overlap-extension RT-PCR technology of the present invention, to link the cognate pairs.

[00219] Another embodiment of the present invention, is a population of host cells comprising a combinatorial library or sub-library of linked pairs of variable region encoding sequences.

[00220] A population of host cells according to the present invention, will encompass a diverse population of cells corresponding to the diversity of the library the cells have been transformed/transfected with. Preferably, each cell of the population of cells only constitutes one cognate pair of the entire library of cognate pairs, and no individual member of the library of cognate pairs exceeds more than 50%, more preferred 25%, or most preferred 10%, of the total number of individual members expressed from the population of host cells.

[00221] In a preferred embodiment of the present invention, the population of host cells is mammalian cells.

[00222] A population of host cells as described in the above can be utilized for the expression of a recombinant polyclonal binding protein, since individual cells of the population constitute variable region encoding sequences of different diversity.

[00223] One embodiment of the present invention, is a recombinant polyclonal protein expressed from a population of host cells comprising a library of vectors encoding diverse cognate pairs of linked variable region encoding sequences, where such a library is obtainable by the method of the present invention. Typically, a recombinant polyclonal

protein of the present invention is comprised of at least 2, 5, 10, 20, 50, 100, 1000, 10^4, 10^5 or 10^6 proteins composed of different cognate pairs.

[00224] A preferred embodiment of the present invention, is a recombinant polyclonal immunoglobulin expressed from a population of host cells comprising a library of vectors encoding diverse cognate pairs of heavy chain variable region and light chain variable region encoding sequences.

[00225] Another preferred embodiment of the present invention, is a recombinant polyclonal TcR expressed from a population of host cells comprising a library of vectors encoding diverse cognate pairs of TcR alpha chain variable region linked with beta chain variable region encoding sequences and/or TcR gamma chain variable region linked with delta chain variable region encoding sequences.

[00226] Another embodiment of the present invention is a host cell suitable for production of a monoclonal protein. In particular a monoclonal antibody comprised of a cognate pair of a light chain variable region with a heavy chain variable region or a monoclonal TcR comprised of a cognate pair of an alpha variable region with a beta variable region or a delta variable region with a gamma variable region. Preferably such a monoclonal production cell line is not a hybridoma cell line.

[00227] Such a monoclonal antibody or TcR can be generated by adding the following steps to the method of linking a plurality of non-contiguous nucleotide sequences of interest a) inserting said linked nucleic acid sequences into a vector; b) introducing said vector into a host cell; c) cultivating said host cells under conditions suitable for expression; and d) obtaining the protein product expressed from the vector inserted into said host cell. Preferably, the vector introduced into the host cell encodes an individual cognate pair of variable region encoding sequences.

Applications of the Invention

[00228] One of the major applications of the present invention is the linkage of cognate pairs of variable region encoding sequences, especially immunoglobulin heavy and light chain variable region encoding sequences or TcR alpha and beta chain or gamma and delta chain variable region encoding sequences, by a high-throughput method for the generation of libraries of cognate pairs. In addition to the generation of cognate pair libraries, the multiplex RT-PCR followed by linkage by ligation or recombination or the multiplex overlap-extension RT-PCR techniques of the present invention may be utilized in the generation of

combinatorial libraries by performing the technique on a population of genetically diverse cells, cell lysates from such a population of cells, or on RNA purified from such a population of cells. The libraries, sub-libraries, or single clones from one of these libraries facilitate the expression of polyclonal or monoclonal proteins. Especially monoclonal or polyclonal antibodies may be obtained from the libraries of the present invention.

[00229] The use of recombinant monoclonal antibodies in diagnostics, treatment, and prophylaxis is well known. Recombinant monoclonal and polyclonal antibodies generated by the present invention will have the same applications as antibody products generated by existing technologies. In particular, a pharmaceutical composition comprising a polyclonal recombinant immunoglobulin as active ingredient, combined with at least one pharmaceutically acceptable excipient, can be produced by means of the present invention. More preferred are pharmaceutical compositions where the polyclonal recombinant immunoglobulin is comprised of cognate pairs of variable region encoding sequences. Such pharmaceutical compositions of polyclonal recombinant immunoglobulins can be used as medicaments. The polyclonal recombinant immunoglobulin of the composition can be specific for or reactive against a predetermined disease target, and the composition can thus be used for the treatment, amelioration or prevention of diseases such as cancer, infections, inflammatory diseases, allergy, asthma and other respiratory diseases, autoimmune diseases, immunological malfunctions, cardiovascular diseases, diseases in the central nervous system, metabolic and endocrine diseases, transplant rejection, or undesired pregnancy, in a mammal such as a human, a domestic animal, or a pet.

[00230] The present invention has a further application which cannot be obtained with conventional monoclonal combinatorial antibody techniques. In situations where protective antigens are either poorly characterized or completely unknown, such as in emerging infectious diseases, it is impossible to screen for a monoclonal antibody that will provide protection against the disease.

[00231] However, with the present invention it will be possible to obtain antibody-expressing cells directly from donors with an established protective antibody response, e.g. convalescent patients, and use the starting material from these individuals to generate a library of cognate pairs of immunoglobulin heavy and light chain variable region encoding sequences.

[00232] In situations where for example the virus is known, but the protective antigens are unknown, it will be possible to generate a sub-library of cognate antibody gene pairs with

broad reactivity towards antigenic structures on the virus. If a recombinant polyclonal antibody is produced from such a sub-library, it will likely contain protective antibodies.

[00233] In situations where antigens are completely unknown, a recombinant polyclonal antibody generated from a cognate pair library from e.g. convalescent patients can be used in the same way as hyperimmune immunoglobulins are being used today. The reason for this is the cognate pairing which ensures that the recombinant polyclonal antibody produced closely resemble the antibody immune response of the convalescent patient.

[00234] Another application of the techniques for linking cognate pairs of variable regions, described in the present invention, is for diagnostic and analytical purposes. When administering a medicament to a patient, the possibility of an immune response directed towards the medicament always exists. This immunogenecity can be assessed by conventional techniques, such as medicament-specific binding assays with serum or plasma derived from the individual treated with the medicament. Alternatively, the methods of the present invention can be used to mirror the patients' immune response shortly after administration of the medicament, by isolating cognate pairs of variable heavy chain and light chain encoding sequences. Antibodies expressed from such a library can then be screened for reactivity towards the medicament or components of the medicament. This method is particular useful if the presence of the medicament in plasma or serum would interfere with the conventional method. Such medicaments are for example antibodies. With conventional methods, identification of an anti-medicament immune response (e.g. an anti-idiotypic, anti-framework or anti-Fc immune response) in relation to treatment with an antibody cannot be performed until the antibody used for treatment has been cleared completely from the blood. With the present invention sequences encoding such anti-medicament antibodies can be isolated and analyzed from the individual treated with the medicament within a couple of weeks. Thus, it would be an alternative method for assessing whether drugs in general and in particular antibody-based drugs are immunogenic.

[00235] A further application of the present invention is in the validation and comparison of vaccines and immunization programmes. This is particular useful during vaccine development, since it will be possible to assess and compare the sequence diversity of antibody responses generated in response to novel candidate vaccines, in addition to the current comparisons of antibody binding affinities and serum titer. Further, the present invention can be used to analyze, monitor and compare antibody responses in surveillance of vaccine efficacy in populations.

[00236] The present invention also finds applications outside the field of variable region containing binding proteins. It is merely a matter of optimization for one skilled in the art to adapt the technique disclosed in the present invention to link two or more transcribed nucleotide sequences encoding a heteromeric protein other that a binding protein. Such a linkage of sequences encoding for domains or subunits from a heteromeric protein may be an advantage in the isolation of these protein encoding sequences since it would reduce the number of steps considerably. Further, it is possible to isolate for example splice variants, mutations or new family members of such proteins by using only one multiplex primer mix/multiplex overlap-extension primer mix.

[00237] The linkage of sequences which encode distant domains of a single protein is also a possibility. Such a technique would ease the research in relation to the importance of certain domains in a multidomain protein, since it would ease the deletion of intermediate domains.

[00238] The generation of chimeric or coupled proteins is also an area where the present invention may be applied. Even if the proteins to be linked are of different origin, e.g. a chimera between a human and a mouse protein, the present invention may be utilized by mixing the cells prior to reverse transcription. Such a cell mixture may constitute either a population of cells from each species or a single cell from each species.

Examples

Example 1: Two-step Multiplex Overlap-extension RT-PCR

[00239] In this example, reverse transcription (RT) was performed using a template derived from an isolated single cell and the cDNA produced was used as template for more than one multiplex overlap-extension PCR.

a. Cells

[00240] An IgG1-kappa expressing Chinese hamster ovary (CHO) cell line was generated using the Flp-In technology (Invitrogen, Carlsbad, CA, USA).

[00241] An IgG-kappa mammalian expressing plasmid vector pLL113 (Figure 4) was constructed based on the Flp-In expression vector, pcDNA5/FRT. CHO-Flp-In cells were co-transfected, using Lipofectamin2000 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions, with pLL113 containing the antibody encoding genes and pOG44 conferring transient expression of the Flp recombinase. Transformants were selected and verified for production of IgG-kappa by immunoassays. The selected cell line was named

CHO Flp-In pLL113, and maintained in Ham's F-12 medium, supplemented with 2 mM L-glutamine, 10% FCS and 900 Hygromycin B (maintenance media).

[00242] CHO Flp-In pLL113 cells were harvested using trypsin and washed 3x in maintenance media. After the final wash the cells were re-suspended in the original volume of maintenance media. The cell concentration was determined using a Casy-1 System (Schärfe System GmbH, Reutlingen, Germany) and diluted in maintenance media to a concentration of 1 cell/5µl.

b. Reverse transcription

[00243] CHO Flp-In pLL113 cell suspension, on average containing one cell, was dispensed into single wells of a 96-well PCR plate (Thermo-Fast 96, skirted AB-0800, ABgene, Epsom, Surrey, UK).

[00244] The cDNA was synthesized from the distributed cells by utilizing the reverse transcriptase (RT) step in the Qiagen One-Step RT-PCR protocol (Qiagen OneStep RT-PCR Kit, Cat# 210210, Hilden, Germany).

[00245] Each well contained the following reagents in a total volume of 20 µl:

1 x One Step RT-PCR buffer,
dNTP's in a final concentration of 1 mM of each,
5 pmol Oligo poly-dT (18),
26 U RNase inhibitor (RNasin, Promega, Madison USA, cat. No N2111),
2 µl One Step RT-PCR enzyme mix, and
5 µl CHO Flp-In pLL113 cell suspension.

[00246] The RT reaction was performed by incubating the reaction mixes at 55°C for 30 min. Subsequently the reverse transcriptase was inactivated by incubating the reaction mixture at 94°C for 10 min.

c. Multiplex Overlap-extension PCR

[00247] A fraction of the cDNA products generated in step b) was used as template for multiplex overlap-extension PCR. The reactions were performed in 96-well plates.

[00248] Each well contained, in a total volume of 40 µl, the following reagents:

1 x One Step RT-PCR buffer,
dNTP's in a final concentration of 500 µM of each,
Multiplex overlap-extension primer mix in the concentrations as indicated in Table 1,

26 U RNase inhibitor (RNasin, Promega, Madison USA, cat. No N2111),

2 µl One Step RT-PCR enzyme mix, and

1 µl cDNA template (derived from a single cell (step b)).

[00249] The multiplex overlap-extension primer mix used comprised the primers shown in table 1.

Table 1

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
V_H	20 nM of each	1	agcctatactattgattaggcgcgccCAGRTGCAGCTGGTGCART
		2	agcctatactattgattaggcgcgccSAGGTCCAGCTGGTRCAGT
		3	agcctatactattgattaggcgcgccCAGRTCACCTTGAAGGAGT
		4	agcctatactattgattaggcgcgccSAGGTGCAGCTGGTGGAG
		5	agcctatactattgattaggcgcgccCAGGTGCAGCTACAGCAGT
		6	agcctatactattgattaggcgcgccCAGSTGCAGCTGCAGGAGT
		7	agcctatactattgattaggcgcgccGARGTGCAGCTGGTGCAGT
		8	agcctatactattgattaggcgcgccCAGGTACAGCTGCAGCAGTC
C_H	100 nM	9	GACSGATGGGCCCTTGGTGG
V_L	20 nM of each	10	cgcctaatcaatagtataggctagccGACATCCAGWTGACCCAGTCT
		11	cgcctaatcaatagtataggctagccGATGTTGTGATGACTCAGTCT
		12	cgcctaatcaatagtataggctagccGAAATTGTGWTGACRCAGTCT
		13	cgcctaatcaatagtataggctagccGATATTGTGATGACCCACACT
		14	cgcctaatcaatagtataggctagccGAAACGACACTCACGCAGT
		15	cgcctaatcaatagtataggctagccGAAATTGTGCTGACTCAGTCT
$C_{L\kappa}$	100 nM	16	atatatatgcggccgcttaTTAACACTCTCCCCTGTTG

W=A/T, S=G/C, R=A/G. Capitalized sequences correspond to the gene-specific region.

[00250] The reactions were performed with a 96-well MWG Primus HT thermocycler (MWG Biotech AG, Ebersberg, Germany) with the following cycling conditions:

Denature	30 sec	95°C	50 cycles
Anneal	30 sec	50°C	
Extend	5 min	72°C	
Final extension	10 min	72°C	

d. Nested PCR

[00251] Nested PCR was performed using the multiplex overlap-extension products as template. The reactions were performed in 96-well plates.

[00252] Each well contained, in a total volume of 50 µl, the following reagents:

1 x BioTaq buffer,

dNTP's in a final concentration of 400 µM of each,

2 mM $MgCl_2$,

Nested PCR primer mix,

1.25 U BIOTAQ DNA Polymerase (Cat. No BIO-21040, Bioline, UK.), and

1 µl Multiplex overlap-extension PCR product (step c).

[00253] The primers used are shown in table 2.

Table 2

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
J_H	100 nM of each	17	atattctcGAGACGGTGACCAGGGTG
		18	atattctcGAGACGGTGACCATTGT
		19	atattctcGAGACGGTGACCAGGGTTC
		20	atattctcGAGACGGTGACCGTGGT
$C_{L\kappa}$	100 nM	21	accgcctccaccggcggccgcttaTTAACACTCTCCCCTGTTGAAGCTCTT

Capitalized sequences correspond to the gene-specific region.

[00254] The reactions were performed with a 96-well MWG Primus HT thermocycler (MWG Biotech AG, Ebersberg, Germany) with the following cycling conditions:

Denature	30 sec	95°C	25 cycles
Anneal	30 sec	50°C	
Extend	90 sec	72°C	
Final extension	10 min	72°C	

[00255] The nested PCR products were analyzed by 1% agarose gel electrophoresis using ethidium bromide for detection (Figure 5). The expected size of the overlap-extension product was 1076 bp. Such a product can be observed in lanes 1, 5, 6, 7, 8 and 12 indicated by the arrows (Figure 5A). In the experiment shown, the negative control is contaminated and a similar contamination is also present in the samples. Figure 5B illustrates the fragments of Figure 5A, which are relevant to the present experiment.

[00256] The present experiment illustrates one way to perform two-step multiplex overlap-extension PCR, utilizing template derived from a single cell. Further, it was shown that it is

possible to perform approximately twenty multiplex overlap-extension PCR's using cDNA generated from an isolated single cell.

Example 2: Single-step Multiplex Overlap-extension RT-PCR

[00257] In this example reverse transcription and multiplex overlap-extension PCR was performed in a single step using template from lysed cells, in concentrations corresponding to 100, 10, or 1 cells.

a. Cells

[00258] The human hybridoma cell line HB-8501 producing an anti-tetanus IgG1-kappa antibody was acquired from American Type Culture Collection and cultured in Iscove's Modified Dulbecco's Medium (Vitacell, Kiev, Ukraine, cat. No 30-2005) containing 10% fetal bovine serum. Before multiplex overlap-extension RT-PCR was performed, the cells were harvested, counted and frozen at -80°C in culture medium in a concentration of 200 cells/µl.

b. Single-step Multiplex Overlap-extension RT-PCR

[00259] The Qiagen One-Step RT-PCR kit (Qiagen cat. No 210212, Hilden, Germany) was used for the multiplex overlap-extension RT-PCR essentially according to the manufacturer's recommendation. Before addition to the PCR tubes cell lysates were thawed and diluted in H_2O to yield a lysate concentration corresponding to 100, 10, and 1 cells per 5 µl.

[00260] Each PCR tube contained the following reagents in a total volume of 50 µl:

1 x One-Step RT-PCR buffer,
dNTP's in a final concentration of 400 µM of each,
Multiplex overlap-extension primer mix in the concentrations as indicated in table 3,
2 µl One-step RT-PCR enzyme mix,
50 U RNase inhibitor (RNasin, Promega, Madison USA, cat. No N2515), and
5 µl diluted cell lysate

[00261] The multiplex overlap-extension primer mix used comprised the primers shown in table 3.

Table 3

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
V_H	40 nM of each	22	gctagcattattattaccatggccCAGRTGCAGCTGGTGCART
		23	gctagcattattattaccatggccSAGGTCCAGCTGGTRCAGT
		24	gctagcattattattaccatggccCAGRTCACCTTGAAGGAGT
		25	gctagcattattattaccatggccSAGGTGCAGCTGGTGGAG
		26	tattggcgcgccatggccsaggtgCAGCTGGTGGAG
		27	gctagcattattattaccatggccCAGGTGCAGCTACAGCAGT
		28	gctagcattattattaccatggccCAGSTGCAGCTGCAGGAGT
		29	gctagcattattattaccatggccGARGTGCAGCTGGTGCAGT
		30	gctagcattattattaccatggccCAGGTACAGCTGCAGCAGTC
J_H	200 nM of each	31	atattctcGAGACGGTGACCAGGGTG
		32	atattctcGAGACGGTGACCATTGTCC
		33	atattctcGAGACGGTGACCAGGGTTC
		34	atattctcGAGACGGTGACCGTGGTCC
V_L	40 nM of each	35	ccatggtaataataatgctagccGACATCCAGWTGACCCAGTCT
		36	ccatggtaataataatgctagccGATGTTGTGATGACTCAGTCT
		37	ccatggtaataataatgctagccGAAATTGTGWTGACRCAGTCT
		38	ccatggtaataataatgctagccGATATTGTGATGACCCACACT
		39	ccatggtaataataatgctagccGAAACGACACTCACGCAGT
		40	ccatggtaataataatgctagccGAAATTGTGCTGACTCAGTCT
$C_{L\kappa}$	200 nM	16	atatatatgcggccgcttaTTAACACTCTCCCCTGTTG

W=A/T, S=G/C, R=A/G. Capitalized sequences correspond to the gene-specific region

[00262] The reactions were performed using the following cycling conditions:

Reverse transcription:	30 min	55°C	
Polymerase activation:	15 min	95°C	inactivating reverse transcriptase and activating Taq polymerase.

[00263] PCR reaction:

Denature	30 sec	94°C	50 cycles
Anneal	30 sec	44°C	
Extend	3 min	72°C	
Final extension	10 min	72°C	

[00264] Ten microliters of the reaction products were analyzed by 1.5 % agarose gel electrophoresis using ethidium bromide for detection (Figure 6A).

[00265] Expected size of the fragments (the exact size depends upon the lengths of the variable regions):

V_H: 410 bp

LC: 680 bp

Overlap-extension fragment: 1070 bp

[00266] Discrete DNA fragments with mobilities corresponding to the heavy chain variable region (V_H) and the light chain variable and constant region (LC) lengths are seen for all dilutions of cell lysate. A less intense fragment with mobility corresponding to the putative overlap-extension fragment is seen in lysates from 100 and 10 cells. The overlap-extension fragment of the one cell lysate sample, is difficult to recognize, although it can be seen on the original gel (see arrow in photo shown in Figure 6A and sketch shown in Figure 6B).

c. Identification of the Overlap-extension Fragment

[00267] From an experiment reproduced as the one described above the presence of the overlap-extension band was verified. Regions in the agarose gel around 1070 bp from a lane corresponding to 1 cell and from a lane corresponding to 100 cells were excised and the DNA purified using Qiaex II (Qiagen cat. No 20051, Hilden, Germany) and eluted in 20 µl of water.

[00268] One microliter of the eluate was subjected to PCR (Biotaq kit, Bioline, UK cat. No BIO-21040) according to the manufacturers instructions with primers flanking the putative overlap-extension fragment (J_H primers corresponding to SEQ ID NO 32 to 35 and $C_{L\kappa}$ primer corresponding to SEQ ID NO: 17, each primer at a concentration of 0.5 µM). Cycling parameters were:

Denature	30 sec	95°C	30 cycles
Anneal	30 sec	55°C	
Extend	1 min	72°C	

[00269] Ten microliters of each reaction product was analyzed by 1% agarose gel electrophoresis using ethidium bromide for detection. Several fragments can be seen including, both from 1 and 100 cells, a fragment with mobility of the expected overlap-extension fragment (arrow in Figure 6C). The 1 kb fragment from the gel lane corresponding to one cell was cut out of the gel and purified using Qiaex II as described above. The purified fragment was digested with the restriction enzymes NheI and NcoI (separately) and the reaction products were analyzed by 1% agarose gel electrophoresis using ethidium bromide

for detection (Figure 6D). These restriction sites are present in the overlap between VH and LC and the expected sizes after digestion are approximately 410 and 680 bp long, respectively (Figure 2). The NheI digestion was partial since a large fraction is still present at the original size.

Example 3: Combined Single-step Multiplex Overlap-extension RT-PCR and Nested PCR

[00270] In this example, reverse transcription and multiplex overlap-extension PCR reactions were performed in a single step followed by a semi-nested PCR amplification, using template from lysed cells, in concentrations corresponding to 100, 10 or 1 cells.

a. Single-step Multiplex Overlap-extension RT- PCR

[00271] Multiplex overlap-extension RT-PCR using HB-8501 cell lysate was performed as described in Example 2 utilizing a multiplex overlap-extension primer mix comprising the primers shown in table 4.

Table 4

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
V_H	40 nM of each	41	tattggcgcgccatggccCAGRTGCAGCTGGTGCART
		42	tattggcgcgccatggccSAGGTCCAGCTGGTRCAGT
		43	tattggcgcgccatggccCAGRTCACCTTGAAGGAGT
		44	tattggcgcgccatggccSAGGTGCAGCTGGTGGAG
		45	tattggcgcgccatggccCAGGTGCAGCTACAGCAG
		46	tattggcgcgccatggccCAGSTGCAGCTGCAGGAGT
		47	tattggcgcgccatggccGARGTGCAGCTGGTGCAGT
		48	tattggcgcgccatggccCAGGTACAGCTGCAGCAGTC
C_{H1-5}	200 nM of one of these primers	49	AAGTAGTCCTTGACCAGGCAGCC
		50	TGAGTTCCACGACACCGTCA
		51	AGAGGTGCTCTTGGAGGAGG
		52	AGTTTTGTCACAAGATTTGGG
		53	GTTGCAGATGTAGGTCTGGGTGC

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
V_L	20 nM of each	54	gccatggcgcgccaatagctagccGACATCCAGWTGACCCAGTCT
		55	gccatggcgcgccaatagctagccGATGTTGTGATGACTCAGTCT
		56	gccatggcgcgccaatagctagccGAAATTGTGWTGACRCAGTCT
		57	gccatggcgcgccaatagctagccGATATTGTGATGACCCACACT
		58	gccatggcgcgccaatagctagccGAAACGACACTCACGCAGT
		59	gccatggcgcgccaatagctagccGAAATTGTGCTGACTCAGTCT
$C_{L\kappa}$	200 nM	60	atatatatgcggccgcttaTTAACACTCTCCCCTGTTGAA

W=A/T, S=G/C, R=A/G. Capitalized sequences correspond to the gene-specific region.

[00272] It should however be noted that for each multiplex overlap-extension primer mix only one of the C_{H1-5} primers were used, resulting in five different multiplex overlap-extension primer mixes.

[00273] Remaining parameters were as described for the multiplex overlap-extension RT-PCR reaction in Example 2, only with a change in annealing temperature to 50°C. A reaction was performed for each heavy chain constant region reverse primer (C_{H1}, C_{H2}, C_{H3}, C_{H4}, C_{H5} corresponding to SEQ ID NOs: 49 to 53, respectively) utilizing lysates corresponding to 100, 10, 1 and 0 cells.

b. Semi-nested PCR

[00274] One microliter of the multiplex overlap-extension RT-PCR reaction product was subjected to semi-nested PCR (Biotaq kit, Bioline, UK cat. No BIO-21040), essentially as proposed by the manufacturer. The total volume of the reactions were 50 µl. The primers used are shown in table 5.

Table 5

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
J_H	200 nM of each	61	ggaggcgctcGAGACGGTGACCAGGGTGCC
		62	ggaggcgctcGAGACGGTGACCATTGTCCC
		63	ggaggcgctcGAGACGGTGACCAGGGTTCC
		64	ggaggcgctcGAGACGGTGACCGTGGTCCC
$C_{L\kappa}$	200 nM	21	accgcctccaccggcggccgcttaTTAACACTCTCCCCTGTTGAAGCTCTT

Capitalized sequences correspond to the gene-specific region.

[00275] Cycling conditions were as follows:

Denature	30 sec	95°C	}
		65	}

Anneal	30 sec	50°C	25 cycles
Extend	1.5 min	72°C	
Final extension	5 min	72°C	

[00276] Ten microliters of each reaction was analyzed by 1.5% agarose gel electrophoresis using ethidium bromide for detection (Figure 7).

[00277] The putative overlap-extension fragment from the semi-nested PCR corresponding to lysate from one cell and where the C_{H4} primer was used in the first reaction (see arrow in Figure 7), was excised from the agarose gel, purified using a Qiaex II kit (Qiagen cat. No. 20051, Hilden, Germany) and inserted into pCR2.1-TOPO using an Invitrogen TOPO TA cloning kit (Invitrogen cat. No 45-0641, Carlsbad, CA, USA). Inserts of eight clones were sequenced and seven of these appeared to consist of a heavy chain variable region linked to a light chain variable and constant region, by the expected overlap region.

[00278] In summary, the sensitivity of the combined multiplex overlap-extension RT-PCR and the semi-nested PCR reaction was very satisfactory, with 4 out of 5 constant region primers capable of amplifying significant amounts of overlap-extension products from lysate corresponding to a single cell.

Example 4: Combined Single-step Multiplex Overlap-extension RT-PCR and Nested PCR Using Enriched Human B Lymphocytes as Template Source

[00279] In this example, reverse transcription and multiplex overlap-extension PCR reactions were performed in a single step followed by a semi-nested amplification PCR, using isolated single human B lymphocytes as template source.

a. B-cell isolation

[00280] A human male donor was immunized with Tetanus toxoid. A blood sample of 120 ml was collected from the donor 6 days post immunization and peripheral blood mononuclear cells (PBMC's) were isolated using Lymphoprep (Axis-Shield, Oslo Norway, prod. No 1001967) according to the manufacturer's instructions. The CD19-positive cell population was enriched utilizing magnetic bead cell sorting. The PBMC's were stained with FITC-conjugated anti-CD19 antibody (Becton Dickinson, NJ, USA, cat. No 345776). Magnetic bead cell sorting using anti-FITC-conjugated magnetic microbeads and column purification was performed according to manufactures instructions (Miltenyi Biotec, Gladbach, Germany, cat. No 130-042-401). The cells were diluted to a concentration of 200 cells per ml in PBS containing 2 nM EDTA and 0.5% BSA. Five microliters of the diluted cells was distributed

to PCR tubes obtaining approximately a single cell pr. tube. The tubes were stored at -80°C until use.

b. Multiplex Overlap-extension RT- PCR and Semi-nested PCR

[00281] Conditions for multiplex overlap-extension RT-PCR and semi-nested PCR were as described in Example 3. However, the reactions were only performed with the multiplex overlap-extension primer mix constituting primer C_{H3} corresponding to SEQ ID NO: 51. Sixteen samples from the combined multiplex overlap-extension RT-PCR and semi-nested PCR reactions were analyzed by subjecting 10 µl from each semi-nested PCR reaction to 1% agarose gel electrophoresis using ethidium bromide for detection (Figure 8).

[00282] As can be seen in Figure 8, 2 out of 16 lanes (lanes 5 and 6) contained fragments of the expected mobility (around 1 kb). Further, lane 2 contained a less intense fragment at the expected mobility. The 1 kb fragments of lane 5 and 6 were excised from the agarose gel, purified using a Qiaex II kit (Qiagen cat. No 20051) and inserted into pCR2.1-TOPO using the Invitrogen TOPO TA cloning kit (Invitrogen cat. No 45-0641, Carlsbad, CA, USA). Two clones from each isolated fragment had inserts of the correct size. Restriction enzyme digestion (NcoI and NheI, separately) showed fragments of the expected sizes (410 and 680 bp) indicating a correct linkage between the heavy chain variable region and the light chain variable and constant region encoding sequences.

[00283] The two clones originating from the fragment in lane 5 were sequenced and shown to be identical, indicating that the linked V_H and LC were cognate pairs.

Example 5: Combined Single-step Multiplex Overlap-extension RT-PCR and Nested PCR Using V_λ-specific Primers

[00284] In this example, reverse transcription and multiplex overlap-extension PCR reactions were performed in a single step using V_λ-specific primers, followed by a semi nested PCR reaction. Total RNA purified from two different cell lines expressing lambda gene families 1b and 1e in combination with the same heavy chain variable region were used as templates.

a. Cells

[00285] Two IgG1-lambda expressing Chinese hamster ovary (CHO) cell lines were generated using the Flp-In technology (Invitrogen, Carlsbad, CA, USA).

[00286] IgG1-lamda mammalian expressing vectors pEm465/01P581 and pEm465/01P582 (Figure 9A) representing two lambda gene families were constructed based on the Flp-In

expression vector, pcDNA5/FRT. CHO-Flp-In cells were co-transfected, using Fugene 6 (Roche, Mannheim, Germany) according to the manufacturer's instructions, with the above mentioned plasmids containing the antibody encoding genes and pOG44 conferring transient expression of the Flp recombinase. Transformants were selected for insertions. The selected cell lines were named CHO Flp-In/Em464/01P581 & CHO Flp-In/Em464/01P582, and maintained in Ham's F-12 medium, supplemented with 2mM L-glutamine, 10% FCS and 900 µg/ml Hygromycin B (maintenance media).

[00287] The cell lines were harvested using trypsin and washed 3x in maintenance media. Approximately 10^7 cells were used for Total RNA purification using Nucleo Spin L kit (Macherey- Nagel, Düren, Germany) according to the manufacturer's description. The final RNA concentrations were determined by OD_{260} measurements.

b. Single-step Multiplex Overlap-extension RT-PCR

[00288] Multiplex overlap-extension RT-PCR using 50 pg, 5 pg, or 0.5 pg total RNA as template was essentially performed as described in Example 3 utilizing a multiplex overlap-extension primer mix comprising the primers shown in Table 6 and the cycling conditions specified below.

Reverse transcription:	30 min	55°C		
Polymerase activation:	15 min	95°C	inactivating reverse transcriptase and activating Taq polymerase.	
PCR reaction:				
Denature	30 sec	95°C	}	35 cycles
Anneal	30 sec	50°C	}	
Extend	5 min	72°C	}	
Final extension	10 min	72°C		

Table 6

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
V_H	40 nM of each	1	agcctatactattgattaggcgcgccCAGRTGCAGCTGGTGCART
		2	agcctatactattgattaggcgcgccSAGGTCCAGCTGGTRCAGT
		3	agcctatactattgattaggcgcgccCAGRTCACCTTGAAGGAGT
		4	agcctatactattgattaggcgcgccSAGGTGCAGCTGGTGGAG
		5	agcctatactattgattaggcgcgccCAGGTGCAGCTACAGCAGT
		6	agcctatactattgattaggcgcgccCAGSTGCAGCTGCAGGAGT
		7	agcctatactattgattaggcgcgccGARGTGCAGCTGGTGCAGT
		8	agcctatactattgattaggcgcgccCAGGTACAGCTGCAGCAGTC
C_H	100 nM	9	GACSGATGGGCCCTTGGTGG
V_λ	100 nM of each	65	cgcctaatcaatagtataggctagccCAGTCTGTGCTGACTCAGCCA
		66	cgcctaatcaatagtataggctagccCAGTCTGTGYTGACGCAGC
		67	cgcctaatcaatagtataggctagccCAGTCTGTCGTGACGCAGC
		68	cgcctaatcaatagtataggctagccCARTCTGCCCTGACTCAGCCT
		69	cgcctaatcaatagtataggctagccTCCTATGWGCTGACTCAGCCA
		70	cgcctaatcaatagtataggctagccTCTTCTGAGCTGACTCAGGAC
		71	cgcctaatcaatagtataggctagccCACGTTATACTGACTCAACCG
		72	cgcctaatcaatagtataggctagccCAGGCTGTGCTGACTCAGC
		73	cgcctaatcaatagtataggctagccAATTTTATGCTGACTCAGCCCCA
		74	cgcctaatcaatagtataggctagccCAGRCTGTGGTGACYCAGGA
		75	cgcctaatcaatagtataggctagccCWGCCTGTGCTGACTCAGC
$C_{\lambda 2/C\lambda 7}$	100 nM	76	gccgcttattaTGAACATTCTGTAGGGGCCA
		77	gccgcttattaAGAGCATTCTGCAGGGGCCA

Y=C/T, W=A/T, S=G/C, R=A/G. Capitalized sequences correspond to the gene-specific region.

c. Semi-nested PCR

[00289] One microliter of the multiplex overlap-extension RT-PCR reaction product was subjected to semi-nested PCR (Biotaq kit, Bioline, UK cat. No BIO-21040), essentially as proposed by the manufacturer. The total volumes of the reactions were 20 µl. The primers used are shown in table 7.

Table 7

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
J_H	200 nM of each	61	ggaggcgctcGAGACGGTGACCAGGGTGCC
		62	ggaggcgctcGAGACGGTGACCATTGTCCC
		63	ggaggcgctcGAGACGGTGACCAGGGTTCC

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
		64	ggaggcgctcGAGACGGTGACCGTGGTCCC
$C_{L\lambda}$	200 nM	78	atatatatgcggccgcttattaTGAACATTCTGTAGGGGCCACTG
		79	atatatatgcggccgcttattaAGAGCATTCTGCAGGGGCCACTG

Capitalized sequences correspond to the gene-specific region.

[00290] Cycling conditions were as follows:

Denature	30 sec	95°C	}	25 cycles
Anneal	30 sec	50°C		
Extend	1.5 min	72°C		
Final extension	5 min	72°C		

[00291] Ten microliters of the nested products were analyzed by 1.5% agarose gel electrophoresis using ethidium bromide for detection (Figures 9B and 9C). The arrows indicate overlap-extension products with the expected migration properties of approximately 1 kb.

[00292] The present experiment illustrates that the lambda multiplex overlap-extension primer mix shown in Table 6 is applicable in single-step multiplex overlap-extension RT-PCR independent on the template used. Further, specific overlap-extension PCR products were produced at 0.5 pg total RNA. This sensitivity suggest that cognate linked heavy chain variable region and light chain variable region encoding sequences can be amplified from a single cell.

Example 6: Generation of a Sub-library of Tetanus-specific Cognate Pairs of Antibody Encoding Sequences

[00293] In the present example the steps outlined in the flow chart illustrated in Figure 10, are exemplified using Tetanus Toxoid (TT) as target antigen.

a. Donors

[00294] Donors which previously have been immunized with the Tetanus vaccine are boosted with Tetanus vaccine (Statens Serum Institut, Denmark). Six days post the Tetanus vaccine boost a blood sample of approximately 200 ml is drawn from the donors into a tube containing anticoagulant.

[00295] It is required that the donors generally are healthy with no silent or chronic infections. They should not suffer from autoimmune diseases or receive any immunosuppressive

medication, and they should not have had any vaccinations within the last 3 months. Also, at the time of the TT-vaccine boost, the donors must not have had any serious infections within the last month.

b. Preparation of Peripheral Blood Mononuclear Cells (PBMC)

[00296] PBMC are isolated from the blood samples using Lymphoprep (Axis-Shield PoC AS, Norway, prod. No 1001967) according to the manufactures recommendations. Briefly, the blood is diluted 1:1 in PBS and this suspension is layered over Lymphoprep in a 2:1 ratio. Vials are centrifuged for 20 min, 25°C at 800g and the white inter phase band is collected. Cells are washed in PBS containing 2 mM EDTA.

c. Enrichment of B cells

[00297] The B-cell lineage (CD19+ cells) of the PBMC's is enriched by magnetic bead cell sorting using the following procedure.

[00298] The isolated PBMC are stained with anti-CD19-FITC (Becton Dickenson, NJ, USA, cat. No 345776). All steps are performed at 4°C in the dark. Staining is performed with 10 µl anti-CD19-FITC pr. $1x10^6$ cells in a volume of 100 µl per $1x10^6$ cells using M-buffer (PBS, pH 7.2, 0.5% BSA, 2 mM EDTA). This will stain the B-cell lineage of the PBMCs. Cells are incubated for 20 min followed by two washing steps with M-buffer. The anti-CD19-FITC stained cells are magnetically labeled with anti-FITC conjugated microbeads, using 10 µl of anti-FITC magnetic beads (Miltenyi Biotec, Gladbach, Germany, cat. No 130-042-401) per $1x10^6$ cells in a volume of 100 µl M-buffer per $1x10^6$ cells. Incubation is performed for 15 min followed by a single washing step with M-buffer. The cells are resuspended in degassed M-buffer.

[00299] A MACS LS column (Miltenyi Biotec, Gladbach, Germany, cat. No 130-042-401) is pretreated with degassed M-buffer according to the manufactures prescriptions. The suspension of cells stained with anti-CD19-FITC and labeled with anti-FITC-magnetic beads are applied to the column and allowed to run through. Stained and labeled cells (CD19+) will be retained in the magnetic field surrounding the column while unstained cells (CD19-) will pass through the column. The column is washed with degassed M-buffer. The magnetic field is removed and the CD19+ cells are collected.

d. Sorting of plasma cells

[00300] The eluate from the MACS column is centrifuged and resuspended in FACS buffer (PBS, pH 7.2, 2% BSA) in a concentration of $1x10^6$ cells/60 µl FACS buffer. Anti-CD19-FITC (Becton Dickenson, NJ, USA, cat. No 345776) (10 µl/10^6 cells), anti-CD38-PE (Becton Dickenson, NJ, USA, cat. No 555460) (10 µl/10^6 cells) and anti-CD45-PerCP (Becton Dickenson, NJ, USA, cat. No 345809) (20 µl/10^6 cells) are added. Incubation is performed at 4°C for 20 min in the dark, followed by two times washing and resuspension in FACS buffer.

The cells are sorted by fluorescence activated cell sorting (FACS) using the following gating parameters:

1. Forward scatter and side scatter in order to retain lymphocytes and monocytes including plasma blasters and plasma cells and to avoid dead cells and cells with very high side scatter which may be aggregates or granulocytes.
2. Cells that are CD19 positive and express increased levels of CD38 ($CD38^{hi}$). This is basically only a gate on CD38 since the PBMCs have been enriched on a MACS column for CD19 expression, but this will disclose any contaminants.
3. CD45 positive cells. All lymphocytes express CD45. However, plasma cells down regulate their CD45 expression compared to earlier lymphocyte differentiation stages. Therefore a discrete population of cells corresponding to plasma cells can be obtained when gating on CD45.

[00301] FACS sorted cells are collected as single cells directly in single wells of 96 well plates containing 5µl PBS buffer supplemented with 5 U RNase inhibitor (RNasin, Promega, Madison USA, cat. No N2515) per well. At this point the cells may be frozen for later RT-PCR or the cells may proceed immediately for RT-PCR.

e. Linkage of Cognate Pairs of Immunoglobulin Variable Region Encoding Sequences

[00302] The multiplex overlap-extension RT-PCR technology is applied to the single cells, thereby achieving cognate linkage of transcribed anti-Tetanus Toxoid heavy chain variable region and light chain variable region encoding sequences.

e-1. Single-step Multiplex Overlap-extension RT-PCR

[00303] The Qiagen One-Step RT-PCR kit (Qiagen cat. No 210212, Hilden, Germany) is used for the multiplex overlap-extension RT-PCR essentially according to the manufacturer's recommendation. Frozen 96-well plates containing a single cell per well are removed from

the freezer and when the wells are free of ice crystals 15 µl RT-PCR reaction mixture is added immediately to each sample (single cell).

[00304] The RT-PCR reaction mixture contains, in a total volume of 20 µl, the following reagents:

1 x One-Step RT-PCR buffer,
dNTP's in a final concentration of 400 µM of each,
Multiplex overlap-extension primer mix in the concentrations as indicated in table 8,
0.8 µl One-step RT-PCR enzyme mix, and
20 U RNase inhibitor (RNasin, Promega, Madison USA, cat. No N2515).

[00305] The composition of the multiplex overlap-extension primer mix is shown in table 8.

Table 8

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
V_H	40 nM of each	41	tattggcgcgccatggccCAGRTGCAGCTGGTGCART
		42	tattggcgcgccatggccSAGGTCCAGCTGGTRCAGT
		43	tattggcgcgccatggccCAGRTCACCTTGAAGGAGT
		44	tattggcgcgccatggccSAGGTGCAGCTGGTGGAG
		45	tattggcgcgccatggccCAGGTGCAGCTACAGCAG
		46	tattggcgcgccatggccCAGSTGCAGCTGCAGGAGT
		47	tattggcgcgccatggccGARGTGCAGCTGGTGCAGT
		48	tattggcgcgccatggccCAGGTACAGCTGCAGCAGTC
$C_{H(IgG)}$	200 nM	9	GACSGATGGGCCCTTGGTGG
V_L	40 nM of each	54	gccatggcgcgccaatagctagccGACATCCAGWTGACCCAGTCT
		55	gccatggcgcgccaatagctagccGATGTTGTGATGACTCAGTCT
		56	gccatggcgcgccaatagctagccGAAATTGTGWTGACRCAGTCT
		57	gccatggcgcgccaatagctagccGATATTGTGATGACCCACACT
		58	gccatggcgcgccaatagctagccGAAACGACACTCACGCAGT
		59	gccatggcgcgccaatagctagccGAAATTGTGCTGACTCAGTCT
$C_{L\kappa}$	200 nM	60	atatatatgcggccgcttaTTAACACTCTCCCCTGTTGAA

W=A/T, S=G/C, R=A/G. Capitalized sequences correspond to the gene specific region.

[00306] Cycling conditions are as follows:

Reverse transcription:	30 min	55°C	
Polymerase activation:	15 min	95°C	inactivates reverse transcriptase and activates Taq polymerase.

[00307] PCR reaction:

Denature	30 sec	94°C	35 cycles
Anneal	30 sec	52°C	
Extend	5 min	72°C	
Final extension	10 min	72°C	

e-2. Additional Amplification

[00308] One microliter of the multiplex overlap-extension RT-PCR reaction product, from each sample, is subjected to semi-nested PCR (Biotaq kit, Bioline, UK cat. No BIO-21040), essentially as proposed by the manufacturer utilizing 96 well plates. The total volume of each reaction is 50 µl, containing a final concentration of 1x Biotaq buffer, 200 µM dNTP's (of each), 2 mM $MgCl_2$, 1.25 U Bio Taq polymerase and the primers shown in Table 9.

[00309] Cycling conditions are as follows:

Denature	30 sec	95°C	30 cycles
Anneal	30 sec	55°C	
Extend	1.5 min	72°C	
Final extension	5 min	72°C	

Table 9

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
J_H	200 nM of each	61	ggaggcgctcGAGACGGTGACCAGGGTGCC
		62	ggaggcgctcGAGACGGTGACCATTGTCCC
		63	ggaggcgctcGAGACGGTGACCAGGGTTCC
		64	ggaggcgctcGAGACGGTGACCGTGGTCCC
$C_{L\kappa}$	200 nM	21	accgcctccaccggcggccgcttaTTAACACTCTCCCCTGTTGAAGCTCTT

Capitalized sequences correspond to the gene specific region.

[00310] Ten microliters of a limited set of samples are analyzed by 1.5% agarose gel electrophoresis using ethidium bromide for visualization, in order to verify that the multiplex overlap-extension RT-PCR has been successful.

[00311] Expected size of the fragments (the exact size depends upon the lengths of the variable regions):

V_H: ~ 410 bp

LC: ~ 680 bp

Overlap-extension fragment: ~ 1070 bp

[00312] Two microliters from all the samples, run in the 96 well plates, originating from the same donor, are combined in a single tube. The pooled samples are digested with XhoI and NotI. The digested overlap-extension fragments are purified by preparative 1% agarose gel electrophoresis; the overlap-extension fragments are excised from the agarose gel and purified using a Qiaex II kit (Qiagen cat. No 20051, Hilden, Germany). It is not necessary to pool the cognate pairs at this stage if this is not desired. In that case each individual reaction will be subjected to restriction cleavage and the products are cloned individually into the vector described below.

f. Cognate Fab Expression Library

[00313] A library of Fab vectors is generated by inserting the pool of XhoI/NotI digested overlap-extension fragments into E. coli vector JSK301 (Figure 11) by ligation. *E. coli* (TOP10) is transformed by electroporation with the library of Fab vectors and the transformants are selected on 2xYT agar containing 100µg/ml carbenicillin. Plasmid DNA is prepared from colonies directly off the agar plates. The plasmid preparation is derived from a minimum number of colonies per donor, corresponding to 3x the total number of single plasma cells originally sorted, in order to maintain diversity. An Fab expression library is generated by inserting a prokaryotic promoter and leader cassette derived from phh3 (Den, W. *et al.* 1999. J. Immunol. Methods 222, 45-57) into the resulting plasmid preparation by digesting the plasmid preparation with AscI and NheI and subsequent ligation. The library generation process is outlined in Figure 12. *E. coli* (TG1) is transformed by electroporation with the resulting Fab expression library and the transformants are selected on 2xYT agar containing 100µg/ml carbenicillin. The individual donors are kept separate during the described procedure in Den *et al*, J Immunol Methods. 1999 Jan 1;222(1-2):45-57. They can, however, be combined at any stage where this might be desired.

g. Screening of Clones

[00314] Fab expressing clones are screened for TT antigen-binding by antigen-specific ELISA assays.

g-1. Master Plate Generation and Fab Expression

[00315] Individual selected colonies of the TG1 cells, each harbouring a cognate Fab expression vector from the library generated in step (f) are picked into single wells of 96 well plates containing 2 x YT/100 µg/mL Amp/1 % glucose. The colonies are grown overnight at

37°C with gentle shaking. These plates are referred to as master plates, which are stored at -80°C upon addition of glycerol to a final concentration of 15%.

[00316] Before the master plates are stored, they are used for inoculation of one or more 384 well plates containing 2 x YT/100 μg/mL Amp 0.1 % glucose using a 96 pin replicator. The plates are sealed and shaken for 2 – 3 h at 37°C.

[00317] Fab expression is induced by adding an equal volume of 2 x YT/100 μg/mL Amp/ 0.2 mM IPTG, obtaining a final IPTG concentration of 0.1 mM. The plates are sealed and shaken overnight at 30°C. The following day, the Fab containing supernatants are analyzed for TT antigen-binding specificity by ELISA.

g-2. ELISA Analysis

[00318] Three hundred and eighty four (384) well ELISA plates (Nunc, Roskilde, Denmark, cat. No 265196) are coated overnight at 4°C with tetanus toxoid (TT) antigen, diluted to a final concentration of 1 μg/mL in PBS in a volume of 25 μl per well. Excess binding sites of the wells are blocked for 1 h at RT by adding 2 % M-PBS-T (2 % skim milk powder in PBS, 0.05 % Tween 20). The wells are washed 2 times with PBS-T (PBS, 0.05 % Tween 20).

[00319] The Fab containing bacterial supernatants from g-1. are diluted 1:2 in 2 % M-PBS-T, and transferred to the ELISA wells in duplicate. Incubation is performed for 1 h at RT. The wells are washed 4 times with PBS-T. Goat-anti-human Fab/HRP (Sigma, St. Louis, MO, USA, cat. No A0293) 1:10.000 dilution in 2 % M-PBS-T is added to the wells. Incubation is performed for 1 h at RT. The wells are washed 4 times with PBS-T. TMB Plus (KemEnTec, Copenhagen, Denmark, cat. No 4390L) substrate is added and incubation is performed for 5 – 15 minutes. The reactions are stopped by adding an equal volume of 1 M H_2SO_4. The extents of the reactions are read at 450 nm in an ELISA reader (Multiscan Ascent, Labsystems, Franklin, USA).

[00320] The original bacterial clones corresponding to TT antigen-binding clones are subsequently retrieved from the original master plates. Plasmid DNA is prepared from the isolated antigen positive Fab clones, generating a cognate Fab expression sub-library of TT antigen-binding clones.

[00321] The clones may also be analyzed by an anti-light chain ELISA assay, in order to get a correlation between the number of antigen-binding clones and the number of Fab expressing clones. Further, such an analysis provides information of kappa and lambda light chain representation in the clones.

h. Cognate Antibody Expression Library

[00322] In order to facilitate the expression of full-length human antibodies the cognate variable regions from the bacterial Fab expression vector must be transferred to a mammalian expression vector containing the constant domains from one of the human immunoglobulin isotypes. Such a transfer can either be performed clone by clone or in mass. Below it is described how to perform the mass transfer.

[00323] The mass transfer is performed in two steps. First, the plasmid preparations of the individually isolated antigen-binding Fab clones are pooled. The prokaryotic promoter and leader containing cassette is exchanged with a mammalian promoter and leader cassette consisting of human alkaline phosphatase leader sequence (AP Leader), human elongation factor 1-alpha promoter (EFP), adenovirus major late promoter (AdMLP), IgK leader sequence by digesting the pooled Fab expression sub-library with AscI and NheI followed by insertion of a mammalian promoter and leader cassette by a subsequent ligation. *E. coli* (TOP10) is transformed by electroporation with the resulting promoter exchanged Fab expression library and the transformants are selected on 2xYT agar containing 100µg/ml carbenicillin. Plasmid DNA is prepared from colonies directly off the agar plates. The plasmid preparation is derived from a minimum number of colonies per donor, corresponding to approximately 3x the total number of clones in the original library, in order to maintain diversity.

[00324] Second, the cognate variable regions from the promoter exchanged Fab expression vector are isolated by digesting the plasmid preparation with XhoI and NotI. The isolated fragment is inserted by ligation into a mammalian expression vector resulting in a cognate antibody expression library. The vector used is essentially the same as in Figure 9A, except that IgL in this case corresponds to the kappa light chain encoding sequence. The mammalian expression vector used above is based on the site-specific Flp-In system (Invitrogen Corporation, Carlsbad, CA, USA, Cat No K6010-01). E. coli (TOP10) is transformed by electroporation with the resulting cognate antibody expression library and the transformants are selected on 2xYT agar containing 100µg/ml carbenicillin. Plasmid DNA is prepared from colonies directly off the agar plates. The plasmid preparation is again derived from a minimum number of colonies per donor, corresponding to approximately 3x the total number of clones in the original library, in order to maintain diversity. The resulting plasmid preparation of the cognate antibody expression library can be used to transfect a mammalian host cell , e.g. the CHO cells from Invitrogen's Flp-In system (Invitrogen Corporation,

Carlsbad, CA, USA, Cat No K6010-01) in order to generate a stable mammalian expression cell lines by Flp recombinase-mediated integration. Such a cell line can be used in the production of recombinant polyclonal antibodies.

Example 7: Screening by High Density Membrane Assay

[00325] The Fab expressing clones from Example 6 f. are screened by a high density membrane assay.

a. Master Plate Generation

[00326] Master plates are generated as described in step (g-1) of Example 6.

b. PVDF Membrane Preparation

[00327] PVDF membranes (Amersham, Uppsala, Sweden, cat. NoRPN2020F) are coated according to the manufacturer's instructions overnight at 4°C with TT antigen diluted to a final concentration of 1 μg/ml in PBS. Excess binding sites on the membranes are blocked for 1 h in 2 % M-PBS (2 % skim milk powder in PBS). The membranes are rinsed 3 times in PBS. The membranes are soaked in 2 x YT for a few minutes.

c. Clone Consolidation

[00328] Consolidation of clones into 384 well plates is performed by transferring clones from the master plates into one or more 384 well plates containing 20 μl 2 x YT/well, using a 96 pin replicator.

[00329] Using a 384 pin replicator the bacteria are transferred to one or more nylon membranes in duplicate (Amersham, Uppsala, Sweden, cat. No RPN2020B) placed on 2 x YT/Carb/1 % glucose agar plates. The plates are incubated 4 – 6 h at 37°C.

d. Fab Induction

[00330] The PVDF membranes prepared in step b, are placed on a 2 x YT/Carb/0,1 mM IPTG agar plates. The nylon membranes with the growing bacterial colonies are placed on top of the PVDF membranes (one nylon membrane per PVDF membrane) on the IPTG-containing agar plate. The IPTG-containing agar plate is incubated at 30°C overnight to facilitate IPTG induced Fab expression from the colonies. The Fab molecules diffuse through the nylon membrane onto the PVDF membrane, where TT-antigen binding Fab's will be retained on the membrane.

e. Detection

[00331] The nylon membranes are removed and the PVDF membranes are washed 2 x 5 minutes with PBS-T (PBS, 0.05 % Tween 20). The membranes are incubated for 30 minutes with 2 % M-PBS. The PVDF membranes are washed 2 x 5 minutes with PBS-T, followed by incubation for 1 h with goat-anti-human IgG/HRP (Sigma, St. Louis, MO, USA, cat. No A0293), 1:10.000 diluted in 2 % M-PBS. The PVDF membranes are washed 3 x 5 minutes with PBS-T. Finally, the PVDF membranes are incubated for 5 minutes with SuperSignal West Femto chemiluminiscent substrate (Pierce, Rockford, Il, USA, cat. No 34095) according to the manufactures instructions. Excess substrate is removed and the PVDF membranes are placed under a CCD camera for detection of the chemiluminescent signal generated at positions on the PVDF membrane where a Fab fragment binds the TT-antigen.

[00332] Positive TT-antigen binding clones will appear as dots on the image. The original bacterial clone is subsequently retrieved from the original master plates. Plasmid DNA is prepared from the isolated antigen positive Fab clones, generating a cognate Fab expression sub-library of TT-antigen binding clones.

f. Correlation Between Antigen Binding Clones and Fab Expressing Clones

[00333] The nylon membranes removed in step e. are placed on a second series of PVDF membranes coated with anti-light chain antibody or anti-Fab antibody according to the procedure in step b. Step d. and e. are then repeated.

[00334] Fab positive clones will then appear as dots on the image and a correlation between the number of TT-antigen binding clones and the number of Fab expressing clones can be made.

Example 8: Illustrative Example Describing the Isolation of Sequences Encoding a Heteromeric Protein

[00335] The present example illustrates how a trimeric guanine nucleotide binding protein (G-protein) can be isolated in a single step utilizing the multiplex overlap-extension RT-PCR technique of the present invention. The family of G-protein subunits is large, alone in humans there are approximately 16 different alpha subunits, 5 beta subunits and 12 gamma subunits. For the present example subunit GαS (GenBank accession No X04408), subunit Gβ1 (GenBank accession No AF501822) and subunit Gγ1 (GenBank accession No BC029367) were chosen. The linkage procedure is illustrated in Figure 13, where step 1 is the multiplex overlap-extension RT-PCR step and step 2 is the additional amplification of the linked product.

a. Cells

[00336] A population of human liver cells is obtained from surgical discharge samples. The liver cells are disintegrated and lysed, total RNA is isolated from the lysate using RNA L kit (Macherey – Nagel, Düren, Germany, cat. No 740 962.20) according to the manufactures instructions.

b. Multiplex Overlap-extension RT-PCR

[00337] The Qiagen One-Step RT-PCR kit (Qiagen cat. No 210212, Hilden, Germany) is used for the multiplex overlap-extension RT-PCR essentially according to the manufacturer's recommendation.

[00338] Each PCR tube contains, in a total volume of 50 µl, the following reagents:

1 x One Step RT-PCR buffer,

dNTP's in a final concentration of 400 µM of each,

Multiplex overlap-extension primer mix in the concentrations as indicated in table 10,

2 µl One step RT-PCR enzyme mix,

1 U/µl RNase inhibitor (RNasin, Promega, Madison USA, cat. No N2515), and

50 ng total RNA.

[00339] The multiplex overlap-extension primer mix used comprises the primers shown in table 10.

Table 10

Primer Name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
α1	400 nM	80	atatatatgcggccgcttaATGGGCTGCCTCGGGAA
α2	80 nM	81	gccatggcgcgccaatagctagccTTAGAGCAGCTCGTACTGACGAA
β1	80 nM	82	tattggcgcgccatggccATGAGTGAGCTTGACCAGTTACG
β2	80 nM	83	ggtacctaataataatcgatcgcTTAGTTCCAGATCTTGAGGAAGCT
γ1	80 nM	84	cgatcgattattattaggtaccccATGCCAGTAATCAATATTGAGGAC
γ2	400 nM	85	ggaggcgctcgagTTATGAAATCACACAGCCTCCTTTG

Capitalized sequences correspond to the gene-specific region.

[00340] The cycling conditions used are:

Reverse transcription:	30 min	42°C	
Polymerase activation:	15 min	95°C	inactivating reverse transcriptase and activating Taq polymerase.

[00341] PCR reaction:

Denature	30 sec	94°C	45 cycles
Anneal	30 sec	49°C	
Extend	3 min	72°C	
Final extension	10 min	72°C	

c. Amplification PCR

[00342] One microliter of the multiplex overlap-extension RT-PCR reaction product is subjected to an additional PCR amplification step (Biotaq kit, Bioline, UK cat. No BIO-21040) essentially as proposed by the manufacturer. The primers used are α1 and γ2 as shown in table 10 corresponding to SEQ ID NO 80 and 85. The total volume of the reaction is 50 µl.

[00343] Cycling conditions are as follows:

Denature	30 sec	95°C	25 cycles
Anneal	30 sec	53°C	
Extend	3 min	72°C	
Final extension	10 min	72°C	

[00344] Five microliters of the reaction product is analyzed by 1 % agarose gel electrophoresis using ethidium bromide for detection. The size of the product is expected to be 2215 bp. However, additional unlinked fragments of the subunits might be present. The expected sizes of these fragments are Gα: 1228 bp, Gβ: 1064 bp and Gγ: 262 bp.

d. Cloning

[00345] The remaining reaction product is subjected to 1 % agarose gel electrophoresis using ethidium bromide for detection and the overlap-extension fragment of 2215 bp is excised from the agarose gel, purified using a Qiaex II kit (Qiagen, Hilden, Germany, cat. No 20051) and inserted into pCR2.1-TOPO using Invitrogen's TOPO TA cloning kit (Invitrogen cat. No 45-0641, Carlsbad, CA, USA).

[00346] Furthermore, promoter sequences or ribosome binding sites might be inserted upstream of each subunit encoding sequence in order to facilitate their expression from the vector. The restriction sites present in the overlap-extension sequences are applicable for such insertion.

e. General Considerations

[00347] The linkage of sequences encoding subunits constituting a heteromeric protein, as described above, can be performed for most heteromeric proteins where the sequence of the individual subunits is known. Moreover, when utilizing a multiplex overlap-extension primer mixture capable of amplifying several members of a particular family, e.g. all the G-protein α-, β-, and γ-subunits, it is possible to identify and link any combination of subunits from a cell type without first having to analyze which family members are expressed by what cell type. For example, the above example can be performed utilizing template derived from isolated single liver cells employing a multiplex overlap-extension primer mix capable of amplifying and linking all 16 Gα subunits with the 5 Gβ subunits and 12 Gγ subunits, thereby identifying which combination of subunits are expressed in each cell and simultaneously isolating the encoding sequences responsible for that combination of subunits. In addition, the use of degenerate primers might even serve to identify new members of a particular family.

[00348] When linking subunits of a heteromeric protein using the multiplex overlap-extension RT-PCR technique of the present invention, the overall size of the linked product should be considered, since there are limits to the number of base pairs a DNA polymerase can amplify. The Deep Vent polymerase from New England Biolabs, MA, USA is one DNA polymerase capable of generating very long primer extensions, of up to 14,000 bp in length. Theoretically, 14,000 bp can encode a protein with an average weight of 510 kDa. Thus, it should be possible to isolate encoding sequences for very large heteromeric proteins utilizing the method of the present invention. The extension capability could possibly be increased even further by utilizing mixtures of DNA polymerases.

Example 9: Design of leader primers:

[00349] In this example, a multiplex set of primers in which the site of priming is kept in the sequence encoding the leader region of the human antibody heavy chain and kappa light chain families was designed.

[00350] When using primers annealing to the N-terminal encoding sequences of the variable region families, some degree of cross-hybridization may be observed. Primers of one particular gene family sequence can hybridize to the cDNA of another gene family, which in many cases cause creation of novel sequences. Proteins produced from such sequences are potentially immunogenic when used in treatment. Generally, these novel sequences can be corrected by PCR, or eliminated from the library.

[00351] An alternative solution, is to position the site of priming in the sequence encoding the leader peptide region of the antibody variable region. Hybrid sequences, generated as a result of cross-priming, will be eliminated during intra-cellular processing of the antibody, in which the peptide leader containing the potential immunogenic sequences will be cleaved off, and hence will not be present in the secreted antibody product. Previous sets of leader primers for antibody cloning, have been located in the 5' end of the leader encoding sequence, allowing for direct eukaryotic expression (Campbell M.J. et al. 1992 Mol Immunol. 29, 193-203). Unfortunately, eukaryotic leader peptides are poorly suited for prokaryotic processing.

[00352] The ease by which nucleic acid sequences are manipulated in bacteria makes it attractive to develop cloning systems in which sequence shuttling between bacterial vectors and vectors for other host organisms is possible. Therefore, a multiplex primer set, in which the site of priming is kept at a position in the leader region encoding sequence, that allow for functional expression of antibodies or antibody fragments in both eukaryotic and prokaryotic expression systems, was designed in the present example.

[00353] The sequence requirement for signal peptidase cleavage at the C-terminal region of any leader peptide seems very similar for prokaryote and eukaryote systems (Nielsen H. et al. 1997 Protein Eng. 10, 1-6.). Further, it follows that mutations in the C-terminal region of the leader peptide is likely to have little effect on the conformation of the N-terminal region and vice versa. Thus, it is expected that C-terminal sequences could be transferable among species without loss of function.

[00354] The basic concept of the primer design in the present example, was to place the site of priming at the last six codons of the C-terminal region of the leader peptide. The remaining part of the leader encoding sequence, up-stream of the site of priming (approximately 50 nucleotides) should be supplied by the appropriate expression vector matching the host species. For example, vectors suitable for expression in gram negative bacteria can be designed with a partial or full-length PelB leader sequence. For expression of antibodies in eukaryotes, a vector with a partial native antibody heavy chain leader encoding sequence and a partial variable kappa chain leader encoding sequence will be suitable. Regardless of expression system, the last six amino acids of the leader, will originate from the endogenous antibody leader encoding sequence contained in the nucleic acid segment inserted into the vector.

[00355] The following primer sequences were designed for use in the multiplex overlap-extension RT-PCR (table 11). However, the overlapping tails (lowercase letters) can easily be

redesigned to function in the linkage by ligation or recombination procedures.

Table 11

Primer name	SEQ ID NO	primer sequence in 5' to 3' direction
V_{HL}	86	tattcccagcggccgccGCCACAGGTGCCCACTC
	87	tattcccagcggccgccCCTTCMTGGGTCTTGTC
	88	tattcccagcggccgccTTARAAGGTGTCCAGTGT
	89	tattcccagcggccgccCCCAGATGGGTCCTG
	90	tattcccagcggccgccCTCCAAGGAGTCTGTTC
	91	tattcccagcggccgccCCATGGGGTGTCCTGTCA
	92	tattcccagcggccgccGCAACAGGTGCCCACT
$C_{H(IgG)}$	9	GACSGATGGGCCCTTGGTGG
$V_{LκL}$	93	ggcggccgctgggaatagctagcgYTCCCAGGTGCCAGATG
	94	ggcggccgctgggaatagctagcgGTCCCTGGATCCAGTG
	95	ggcggccgctgggaatagctagcgCTCCCAGATACCACCGGA
	96	ggcggccgctgggaatagctagcgATCTCTGGTGCCTAC
	97	ggcggccgctgggaatagctagcgATCTCTGATACCAGGGCA
	98	ggcggccgctgggaatagctagcgGTTCCAGCCTCCAGGGGT
$C_{Lκ}$	99	atatatatggcgcgccTTAACACTCTCCCCTGTTGAA

M=A/T, R=A/G, Y=C/T, S=G/C. Capitalized sequences correspond to the gene-specific region.

[00356] It should be noted that the restriction sites in this primer design were slightly modified compared to the primers described in the previous examples. Thus, the overlap sequence between the heavy chain and light chain variable region encoding sequences comprised a NotI and a NheI restriction site, and in the constant kappa primer ($C_{Lκ}$), the previous NotI site has been exchanged with an AscI site. The later modification should of course be observed when designing the primers for the additional PCR amplification. Further, the NotI site in the cloning vector, shown in Figure 11, should be changed to a AscI site, and the AscI site in the promoter unit should be changed to a NotI site.

[00357] The functionality with respect to leader cleavage has been tested *in silico* using SignalP provided by CBS at the Danish Technical University.

[00358] The chimeric leader peptides tested for the variable heavy chain, were encoded by nucleic acid sequences composed of a truncated PelB sequence, a C-terminal NotI site (SEQ ID NO 100: ATG AAA TAT CTT CTA CCA ACA GCG GCA GCT GGA TTA TTG GCG GCC GCC), and in association with the NotI site the gene specific part of one of SEQ ID NO 86 to 92, encoding the six C-terminal amino acids from native V_{HL} family members. All

seven sequences showed signal peptide cleavage in gram negative bacteria, using the SignalP program.

[00359] The chimeric leader peptides tested for the variable light chain, were encoded by nucleic acid sequences composed of a truncated PelB sequence, a C-terminal NheI site (SEQ ID NO 101: ATG AAA TAT TTG CTA CCA ACA GCG GCA GCT GGA TTA TTG TTA CTA GCG), and in association with the NheI site the gene specific part of one of SEQ ID NO 93 to 98, encoding the six C-terminal amino acids from native V_{LkL} family members. All six sequences showed signal peptide cleavage in gram negative bacteria, using the SignalP program.

[00360] The sequences of SEQ ID NO 100 and 101 are suitable for the construction of a bacterial expression vector similar to the one illustrated in Figure 11, in that the AscI site illustrated in figure 11 is substituted with SEQ ID NO 100 and the NheI site is substituted with SEQ ID NO 101. Further, the NotI site illustrated in figure 11 needs to be substituted for a AscI site.

[00361] The mammalian expression vector (Figure 9) can likewise be re-designed to enable expression in mammalian cells following transfer of the variable region encoding segments from the bacterial vector just described.

Example 10: One tube multiplex RT-PCR and linkage by ligation

[00362] The present example illustrates how a cognate pair comprising linked heavy chain variable region and light chain variable region encoding sequences can be generated in a single tube reaction using ligation instead of overlap-extension PCR, to obtain linkage.

a. multiplex RT-PCR

[00363] An in-house produced cell line expressing Fab molecules with specificity towards Tetanus Toxoid are distributed by limiting dilution to obtain a single cell in a single PCR tube.

[00364] In addition to the single cell, each PCR tube contains the following reagents in a total volume of 20 µl:

1 x Phusion HF buffer

dNTP's in a final concentration of 200µM each

Multiplex primer mix in the concentrations as indicated in table 12

0.8 U Phusion polymerase (FinnZymes Cat. No F-530-L)

1 µl Sensiscript reverse transcriptase (Qiagen Cat. No 205213

20 U Rnase inhibitor

Table 12

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
V_H	200 nM of each	102	tgttgttctagatgaaggcgcgccCAGRTGCAGCTGGTGCART
		103	tgttgttctagatgaaggcgcgccSAGGTCCAGCTGGTRCAGT
		104	tgttgttctagatgaaggcgcgccCAGRTCACCTTGAAGGAGT
		105	tgttgttctagatgaaggcgcgccSAGGTGCAGCTGGTGGAG
		106	tgttgttctagatgaaggcgcgccCAGGTGCAGCTACAGCAGT
		107	tgttgttctagatgaaggcgcgccCAGSTGCAGCTGCAGGAGT
		108	tgttgttctagatgaaggcgcgccGARGTGCAGCTGGTGCAGT
		109	tgttgttctagatgaaggcgcgccCAGGTACAGCTGCAGCAGTC
J_H	200 nM of each	61	ggaggcgctcGAGACGGTGACCAGGGTGCC
		62	ggaggcgctcGAGACGGTGACCATTGTCCC
		63	ggaggcgctcGAGACGGTGACCAGGGTTCC
		64	ggaggcgctcGAGACGGTGACCGTGGTCCC
V_L	200 nM of each	110	aacaacactagtgataggctagccGACATCCAGWTGACCCAGTCT
		111	aacaacactagtgataggctagccGATGTTGTGATGACTCAGTCT
		112	aacaacactagtgataggctagccGAAATTGTGWTGACRCAGTCT
		113	aacaacactagtgataggctagccGATATTGTGATGACCCACACT
		114	aacaacactagtgataggctagccGAAACGACACTCACGCAGT
		115	aacaacactagtgataggctagccGAAATTGTGCTGACTCAGTCT
$C_{L\kappa}$	200 nM	16	atatatatgcggccgcttaTTAACACTCTCCCCTGTTG

W=A/T, R=A/G, S=G/C. Capitalized sequences correspond to the gene-specific region

[00365] The cycling conditions are:

Reverse transcription	30 min	37°C	
Denature	30 sec	98°C	
Denature	10 sec	98°C	40 cycles
Anneal	30 sec	55°C	
Extend	30 sec	72°C	
Final extension	5 min	72°C	

b. linkage by ligation

[00366] To perform the linkage by ligation, restriction enzymes and ligase are added directly to the multiplex RT-PCR reaction products. Alternatively, the multiplex RT-PCR can be purified prior to this addition in order to free the mixture from the polymerase.

[00367] The following reagents are added to each tube in a total volume of 40 µl:

1 x NEBII buffer

1 mM ATP

50 U XbaI (New England Biolabs Cat. No R0145L)

25 U SpeI (New England Biolabs Cat. No R0133S)

100 µg/ml BSA

400U T4 DNA ligase (New England Biolabs Cat. No M0202S)

[00368] The reaction is incubated at 16°C over night.

[00369] The linked heavy chain and light chain variable region encoding sequences are purified from the reaction by gel electrophoresis and excision of the band of approximately 1000bp.

[00370] In an alternatively version of this method, the multiplex RT-PCR reaction is performed with a C_H primer instead of the J_H primers in the multiplex primer set. This primer may either be equipped with a cloning tail that allows in-frame insertion of the heavy chain variable region into an expression vector, or a semi nested PCR may be performed with the primers of table 5.

Example 11: Generation of a recombinant polyclonal immunoglobulin with specificity towards Tetanus Toxoid

[00371] In the present example results from a single donor (TT03) immunized with Tetanus Toxoid (TT) is used to illustrate the steps outlined in the flow chart in Figure 10.

a. Donors

[00372] Eight Donors which previously have been immunized with the Tetanus vaccine were boosted with Tetanus vaccine (Statens Serum Institut, Denmark). The donors were assigned with numbers TT01 to TT08. Six days after the Tetanus vaccine boost a blood sample of approximately 200 ml was drawn from the donors into a tube containing anticoagulant.

[00373] The donors were healthy without chronic infections, autoimmune disease, or immunosuppressive medication, and they had not had any vaccinations within the last 3 months.

a-1. Monitoring Donor Quality

[00374] Pre-bleedings of the donors were taken at the time of immunization. Fourteen days later additional bleedings were taken to determine the serum titer. All donors responded with an increase in the TT-titer. An ELISPOT assay was also set up to measure the frequency of

TT-specific plasma cells. Different cell fractions can be used for the ELISPOT, e.g. the major bleed from day 6, the PBMC fraction, the magnetic sorted fraction, or the FACS sorted fraction. ELISPOT can be used to evaluate the donor material and to identify the best responders, which then may proceed through the sorting step and multiplex overlap-extension RT-PCR. For donor TT03 an ELISPOT was performed using the CD19+ cell fraction (see section c). The ELISPOT assay was performed as described by Lakew (Lakew, M. et al. 1997. J. Immunol. Methods 203, 193-198) with minor modifications. The frequency of TT specific plasma cells in the PBMC fraction was calculated to 0.021%

b. Preparation of Peripheral Blood Mononuclear Cells (PBMC)

[00375] PBMC's were isolated from the blood sample using Lymphoprep (Axis-Shield PoC AS, Norway, prod. No 1001967) according to the manufactures recommendations. Briefly, the blood was diluted 1:1 in PBS and this suspension was layered over Lymphoprep in a 2:1 ratio. Vials were centrifuged for 20 min, 25°C at 800g and the white inter phase band was collected. Cells were washed in PBS containing 2 mM EDTA.

[00376] From the 200 ml full blood drawn from donor TT03 approximately $2x10^8$ PBMC's were obtained.

c. Enrichment of B cells

[00377] The B-cell lineage (CD19+ cells) of the PBMC's was enriched by magnetic bead cell sorting using the following procedure.

[00378] The isolated PBMC's were stained with anti-CD19-FITC (Becton Dickenson, NJ, USA, cat. No 345776). All steps were performed at 4°C in the dark. Staining was performed with 10 µl anti-CD19-FITC pr. $1x10^6$ cells in a volume of 100 µl per $1x10^6$ cells using M-buffer (PBS, pH 7.2, 0.5% BSA, 2 mM EDTA). This will stain the B-cell lineage of the PBMC's. Cells were incubated for 20 min followed by two washing steps with M-buffer. The anti-CD19-FITC stained cells were magnetically labeled with anti-FITC-conjugated microbeads, using 10 µl of anti-FITC magnetic beads (Miltenyi Biotec, Gladbach, Germany, cat. No 130-042-401) per $1x10^6$ cells in a volume of 100 µl M-buffer per $1x10^6$ cells. Incubation was performed for 15 min followed by a single washing step with M-buffer. The cells were resuspended in degassed M-buffer.

[00379] A MACS LS column (Miltenyi Biotec, Gladbach, Germany, cat. No 130-042-401) was pretreated with degassed M-buffer according to the manufacturers descriptions. The suspension of cells stained with anti-CD19-FITC and labeled with anti-FITC-magnetic beads

were applied to the column and allowed to run through. Stained and labeled cells (CD19+) were retained in the magnetic field surrounding the column while unstained cells (CD19÷) passed through the column. The column was washed with degassed M-buffer. The magnetic field was removed and the CD19+ cells were collected.

[00380] An analytic staining of the starting material (PBMC), the unlabelled fraction and the CD19 labelled fraction from TT03, using anti-CD19-FITC, anti-CD38-PE, and anti-CD45-PerCP was performed (Figure 14). This shows that the Magnetic cell sorting results in two distinct fractions compared to the PBMC fraction. The CD19 negative cells shown in panel B and the CD19 positive cells shown in panel C. Of the PBMC fraction 11% of the cells were CD19+, the extent of CD19 positive cell purification was 99.5% of R1 (the scatter gate in Figure 14C) for TT03.

[00381] As seen in Figure 14C, the anti-CD38/anti-CD45 plot showed a distinct population of CD38hi, CD45in cells (R2) corresponding to 1.1% of R1. This population contained the plasma cells and was collected during the later sorting step. This indicated that in the PBMC fraction 0.12% of the cells corresponded to plasma cells.

[00382] The CD19 positive cell fraction was frozen in FCS (Invitrogen, Cat. No. 16000-044) +10% DMSO (Sigma, Cat. No. D2650) for later sorting. An analysis like the one seen in Figure 14 was made on the MACS purified cells that had been frozen to make sure that the cells were intact (Figure 15). The staining patterns seen in Figure 15 were the same as seen in Figure 14C although a slightly broader staining intensity were observed. The cells that were collected by sorting is the subset of R1 and R2. This corresponds to approx. 1.1 % of the MACS purified cells.

d. Sorting of plasma cells

[00383] The frozen eluate from the MACS column was thawed, centrifuged and resuspended in FACS buffer (PBS, pH 7.2, 2% BSA) in a concentration of $1x10^6$ cells/60 µl FACS buffer. Anti-CD19-FITC (Becton Dickenson, NJ, USA, cat. No 345776) (10 µl/10^6 cells), anti-CD38-PE (Becton Dickenson, NJ, USA, cat. No 555460) (10 µl/10^6 cells) and anti-CD45-PerCP (Becton Dickenson, NJ, USA, cat. No 345809) (20 µl/10^6 cells) were added, and the mixture was incubated at 4°C for 20 min in the dark, followed by two times washing and resuspension in FACS buffer.

[00384] The cells were sorted by fluorescence activated cell sorting (FACS) using the following gating parameters:

1. Forward scatter and side scatter in order to retain lymphocytes and monocytes including plasma blasters and plasma cells and to avoid dead cells and cells with very high side scatter which may be aggregates or granulocytes.
2. Cells that are CD19 positive and express increased levels of CD38 (CD38hi) This is basically only a gate on CD38 since the PBMC' s have been enriched on a MACS column for CD19 expression, but this will disclose any contaminants.
3. CD45 intermediate positive cells. All lymphocytes express CD45. However, plasma cells down regulate their CD45 expression compared to earlier lymphocyte differentiation stages. Therefore, a discrete population of cells corresponding to plasma cells can be obtained when gating on CD45.

[00385] FACS sorted cells from donor TT03 were collected as the subset of the two gates P2 and P1 (Figure 16A and 16B, respectively). The cells were CD38hi (FL2-A), CD45in (FL3-A) and CD19+, the latter due to the MACS purification. Briefly, the cells were collected in bulk, counted and diluted in RPMI (Invitrogen, Cat No. 21875-034) containing 10% FCS (Invitrogen Cat. No.16000-044), 100 units/ml Penicillin-streptomycin (Invitrogen, cat. No. 15140-122), 2mM L-glutamine (cat No. 25030-024). Cells were then dispensed into fifty 96 well PCR plates (ABgene, cat. No. AB-0800) with one cell per well in 5 µl medium. The plates were sealed and frozen immediately, and stored for later RT-PCR analysis.

e. Linkage of Cognate Pairs of Immunoglobulin Variable Region Encoding Sequences

[00386] The multiplex overlap-extension RT-PCR technology was applied to the single cells obtained from donor TT03, thereby achieving cognate linkage of transcribed anti-Tetanus Toxoid heavy chain variable region and light chain variable region encoding sequences.

e-1. Single-step Multiplex Overlap-extension RT-PCR

[00387] The Qiagen One-Step RT-PCR kit (Qiagen cat. No 210212, Hilden, Germany) was used for the multiplex overlap-extension RT-PCR essentially according to the manufacturer's recommendation. Fifty frozen 96-well PCR plates containing approximately a single cell per well were removed from the freezer and when the wells were free of ice crystals, 15 µl RT-PCR reaction mixture was added immediately to each well.

[00388] The RT-PCR reaction mixture contained, in a total volume of 20 µl, the following reagents:

1 x One Step RT-PCR buffer

dNTP's in a final concentration of 400 µM of each

Multiplex overlap-extension primer mix in the concentrations as indicated in table 13

0.8 µl One-step RT-PCR enzyme mix

20 U RNase inhibitor (RNasin, Promega, Madison USA, cat. No N2515)

[00389] The composition of the multiplex overlap-extension primer mix is shown in table 13.

Table 13

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
V_H	40 nM of each	116	tattcccatggcgcgccCAGRTGCAGCTGGTGCART
		117	tattcccatggcgcgccSAGGTCCAGCTGGTRCAGT
		118	tattcccatggcgcgccCAGRTCACCTTGAAGGAGT
		119	tattcccatggcgcgccSAGGTGCAGCTGGTGGAG
		120	tattcccatggcgcgccCAGGTGCAGCTACAGCAGT
		121	tattcccatggcgcgccCAGSTGCAGCTGCAGGAGT
		122	tattcccatggcgcgccGARGTGCAGCTGGTGCAGT
		123	tattcccatggcgcgccCAGGTACAGCTGCAGCAGTC
$C_{H(IgG)}$	200 nM	9	GACSGATGGGCCCTTGGTGG
V_L	40 nM of each	124	ggcgcgccatgggaatagctagccGACATCCAGWTGACCCAGTCT
		125	ggcgcgccatgggaatagctagccGATGTTGTGATGACTCAGTCT
		126	ggcgcgccatgggaatagctagccGAAATTGTGWTGACRCAGTCT
		127	ggcgcgccatgggaatagctagccGATATTGTGATGACCCACACT
		128	ggcgcgccatgggaatagctagccGAAACGACACTCACGCAGT
		129	ggcgcgccatgggaatagctagccGAAATTGTGCTGACTCAGTCT
$C_{L\kappa}$	200 nM	130	atatatatgcggccgcTTAACACTCTCCCCTGTTGAA

W=A/T, S=G/C, R=A/G. Capitalized sequences correspond to the gene-specific region

[00390] Cycling conditions were as follows:

Reverse transcription:	30 min	55°C	
Polymerase activation:	15 min	95°C	inactivates reverse transcriptase and activates Taq polymerase.
PCR reaction:			
Denature	30 sec	94°C	} 35 cycles
Anneal	30 sec	52°C	
Extend	5 min	72°C	
Final extension	10 min	72°C	

e-2. Additional Amplification

[00391] One microliter of the multiplex overlap-extension RT-PCR reaction product, from each sample, was subjected to semi-nested PCR (Biotaq kit, Bioline, UK cat. No BIO-21040), essentially as proposed by the manufacturer utilizing 96 well PCR plates (ABgene, cat. No. AB-0800). The total volume of each reaction was 50 µl, containing a final concentration of 1x Biotaq buffer, 200 µM dNTP's (of each), 2 mM $MgCl_2$, 1.25 U Bio Taq polymerase and the primers shown in table 14.

Table 14

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
J_H	200 nM of each	61	ggaggcgctcGAGACGGTGACCAGGGTGCC
		62	ggaggcgctcGAGACGGTGACCATTGTCCC
		63	ggaggcgctcGAGACGGTGACCAGGGTTCC
		64	ggaggcgctcGAGACGGTGACCGTGGTCCC
$C_{L\kappa}$	200 nM	21	accgcctccaccggcggccgcttaTTAACACTCTCCCCTGTTGAAGCTCTT

Capitalized sequences correspond to the gene-specific region.

[00392] Cycling conditions were as follows:

Denature	30 sec	95°C	30 cycles
Anneal	30 sec	55°C	
Extend	1.5 min	72°C	
Final extension	5 min	72°C	

[00393] Ten microliters of samples from row A, well 1-12 from each plate were analyzed by 1.5% agarose gel electrophoresis using ethidium bromide for visualization, in order to verify that the multiplex overlap-extension RT-PCR was successful. The expected size of the overlap-extension fragment was approximately 1070bp (the exact size depends upon the lengths of the variable regions). Figure 17 shows samples from eight 96 well plates. The average number of successful overlap-extension RT-PCR fragments from fifty 96 well plates, was estimated to eight per plate. Thus, the total number of successful overlap-extension RT-PCR fragments were estimated to approximately 400.

[00394] Ten microliters from all the performed reactions, run in the 96 well plates, originating from the same donor, were consolidated in a single tube. An aliquot consisting of 200 µl of the pooled PCR products were subsequently purified using QIAquick PCR purification kit according to the manufactures procedure (Qiagen cat. No. 28106, Hilden, Germany) using 60 microliters buffer EB for elution. The purified pool of overlap PCR products was digested

with XhoI and NotI and subsequently purified by preparative 1% agarose gel electrophoresis; the overlap-extension fragments are excised from the agarose gel and purified using a Qiaex II kit (Qiagen cat. No 20051, Hilden, Germany).

f. Cognate Fab Expression Library

[00395] A Fab expression library was generated by a two-step ligation procedure. Initially the pool of digested overlap-extension fragments described above was ligated into the XhoI/NotI digested E. coli vector JSK301 (Figure 11). The ligation reaction was subsequently transformed into electrocompetent E. coli cells (XL1-Blue electroporation competent, Stratagen, cat. No. 200228, La Jolla, USA), according to manufactures instructions. The transformed E. coli cells were plated onto 2xYT agar containing 100µg/ml carbenicillin. Biomass corresponding to approximately 10^{10}-10^{11} cells originating from a number of independent colonies that exceeded at least 5 times the total number of overlap PCR products was used as starting material for plasmid preparation using Qiagen Plasmid preparation Maxi kit (Qiagen cat. No. 12163, Hilden, Germany). To enable Fab expression of the cloned cognate linked VH and VL encoding sequences, a prokaryotic promoter and leader cassette was inserted in a second ligation step. The used bi-directional promoter fragment (SEQ ID NO 321) was extracted from the parental JSK301 by AscI/NheI digestion. The purified pool of plasmids was likewise digested with AscI /NheI restriction endonucleases and gel purified as previously described. The purified fragments were subsequently ligated and transformed into electrocompetent E. coli cells (TG1, Stratagene, cat. No. 200123, La Jolla, USA) and plated on 2xYT agar containing 100µg/ml carbenicillin, and 1% glycose. The Fab expression library generation process is outlined in Figure 12.

g. Screening of Clones

[00396] Fab expressing clones were screened for TT antigen-binding by antigen-specific ELISA assays.

g-1. Master Plate Generation and Fab Expression

[00397] Individual selected colonies of the TG1 cells, each harbouring a cognate Fab expression vector from the library generated as described in section (f) were picked into single wells of 96 well plates containing 2 x YT/100 µg/ml Carb/1 % glucose. The plates were incubated overnight at 37°C with gentle shaking. Four 96 well plates were consolidated into the wells of one 384 well plate, containing 2 x YT/100 µg/ml Carb/1 % glucose, using

96 pin replicators. The 384 well plates were incubated overday at 37°C. These plates are referred to as master plates, and they were stored at -80°C after addition of glycerol to a final concentration of 15%.

[00398] The masterplates were used for inoculation of 384 DeepWell plates containing 2 x YT/100 μg/ml Carb 0.1 % glucose using a 96 pin replicator. The plates were sealed and incubated for 2 – 3 h at 37°C with shaking.

[00399] Fab expression was induced by adding an equal volume of 2 x YT/100 μg/ml Carb/ 0.2 mM IPTG, obtaining a final IPTG concentration of 0.1 mM. The plates were sealed and incubated overnight at 30°C with shaking. The following day, the Fab-containing supernatants were analyzed for TT antigen-binding specificity by ELISA.

g-2. ELISA Analysis

[00400] Three hundred and eighty four (384) well ELISA plates (Corning Inc., Corning, NY, USA, cat. No 3700) were coated overnight at 4°C with Tetanus Toxoid (TT) antigen, diluted to a final concentration of 1 μg/ml in PBS in a volume of 25 μl per well. Excess binding sites of the wells were blocked for 1 h at room temperature (RT) by adding 2 % M-PBS-T (2 % skim milk powder in PBS, 0.05 % Tween 20). The wells were washed 2 times with PBS-T (PBS, 0.05 % Tween 20).

[00401] The Fab-containing bacterial supernatants from section (g-1) were diluted 1:2 in 2 % M-PBS-T, and transferred to the ELISA wells in duplicate. Incubation was performed for 1 h at RT. The wells were washed 4 times with PBS-T. Goat-anti-human Fab/HRP (Sigma, St. Louis, MO, USA, cat. No A0293) 1:10.000 dilution in 2 % M-PBS-T was added to the wells. Incubation was performed for 1 h at RT. The wells were washed 4 times with PBS-T. TMB Plus (KemEnTec, Copenhagen, Denmark, cat. No 4390L) substrate was added and incubation was performed for 5 – 15 minutes. The reactions were stopped by adding an equal volume of 1 M H_2SO_4, and analyzed using a spectrophotometer at 450 nm (Multiscan Ascent, Labsystems, Franklin, USA).

[00402] The original bacterial clones corresponding to TT antigen-binding clones may subsequently be retrieved from the original master plates. Plasmid DNA can be prepared from the isolated antigen positive Fab clones, generating a cognate Fab expression sub-library of TT antigen-binding clones.

[00403] A subset of the clones were further analyzed by an anti-kappa ELISA assay, in order to get a correlation between the number of antigen-binding clones and the number of Fab

expressing clones. The anti-kappa assay was generally performed as the TT-assay, except that the wells were coated with a 1:1000 dilution of goat anti-human kappa antibody (Caltag, California, USA, Cat. No H16000), using a carbonate buffer, pH 9.6.

g-3. Screening results

[00404] Clones from four 384 well plates were screened for reactivity with anti-kappa Ab and TT using ELISA assays according to the procedure described in g-2. The results obtained results are summarized in the tables 15 through 19. The anti-kappa ELISA results inform about the expression of Fab fragments in a given clone, the TT ELISA results inform about the functionality of the Fab fragments.

Table 15 ELISA screening, anti-kappa coat (1440 clones in total)

Plate #	2 x background	3 x background	4 x background
G050	105	79	73
G051	93	79	68
G052	120	96	73
G053	164	141	125
Sum	**482**	**395**	**339**
% Fab pos. clones	**34 %**	**27 %**	**24 %**

[00405] Of 1440 single clones analyzed, 482 clones or 34 % exhibited anti-Kappa reactivity at a level exceeding 2 x background reactivity. 395 clones or 27 % of the clones showed reactivity above 3 x background, etc.

[00406] The same clones were analyzed for TT-antigen reactivity by ELISA, and results are given in table 16 below:

Table 16 ELISA screening, TT coat (1440 clones in total)

Plate #	2 x background	3 x background	4 x background
G050	26	19	17
G051	24	18	15
G052	34	31	24
G053	46	36	30
Sum	**130**	**104**	**86**
% TT pos. clones	**9.0 %**	**7.2 %**	**6.1 %**

[00407] From this table it is seen that 9.0 % clones showed reactivity with TT at 2 x background (defined as the signal obtained by the reactivity of a Fab fragment with a irrelevant specificity). 104 clones reacted at 3 x background (7,2 %) etc.

[00408] Of the Fab-positive clones (482 clones in total), approximately 27 % of the clones (130/482) exhibited reactivity with TT at the 2x background level. This level does not change significantly at the other background levels.

[00409] Six 384 well plates were screened for clones with TT reactivity. The number of clones giving rise to reactivity with the antigen are shown in the table 17. Anti-kappa ELISA was not carried out with these plates.

Table 17 ELISA screening, TT coat (2160 clones in total)

Plate #	2 x background	3 x background	4 x background
G054	82	65	52
G055	30	24	21
G056	25	23	21
G057	30	28	26
G058	24	21	19
G059	18	14	13
Sum	**209**	**175**	**151**
% TT pos. clones	**9.7 %**	**8.1 %**	**7,0 %**

[00410] The percentage of TT positive clones in plates G054 – G059 was comparable to what was found in plates G050 – G053 (table 16).

[00411] The results from all the clones screened against TT (table 16 and 17) are summarized in table 18.

Table 18 All clones screened, TT reactive clones

Plates	2 x background	3 x background	4 x background
G050 – G053	130	104	86
G054 – G059	209	175	151
Sum	**339**	**279**	**237**
% of total (3600)	**9.4 %**	**7.8 %**	**6.6 %**

[00412] In summary, a total of 339 clones showed reactivity with TT at least at 2x background levels (table 18). This corresponds to 9.4 % of all clones screened.

[00413] All positive clones exhibiting reactivity with TT at 2x background was inoculated into 96 well plates as privously described, from the master plate. The next day, the bacteria was collected by centrifugation at 4000 rpm for 15 minutes, and the pellet was resuspended in 0,8 mM EDTA, 0,4 x PBS, 0,8 M NaCl, and incubated on ice for 15 minutes. The periplasmic extract was collected by centrifugation, and the reactivity of the clones was further analyzed. Here we show results from one such plate (G060) for reactivity with anti-Kappa (Figure 18), Ovalbumin (unrelated antigen) (Figure 19), TT (Figure 20), and a one-step competitive assay using 10^{-7} M concentration of TT-antigen in solution (Figure 21). These ELISA assays were performed on the same periplasmic extract, at the same dilution.

[00414] Most of the clones expressed Fab fragments (90/96) (Figure 13), and no clones reacted with Ovalbumin (Figure 19). The reactivity with immobilized TT was reduced or completely inhibited by TT in solution (Figure 21), indicating that the clones react specifically with TT. The reactivity of the clones in plate G060 is summarized in table 19.

Table 19 Summary of G060 plate (96 clones in total)

Background level	Anti-Kappa positive	TT positive
2 x	94 % (90/96)	74 % (71/96)
3 x	91 % (87/96)	72 % (69/96)
4 x	84 % (81/96)	69 % (66/96)

h. Diversity analysis and clone approval

[00415] Plasmid isolated from 47 clones (from plate G060) from the cognate TT antigen-binding Fab expression sub-library were subjected to sequence analysis. The variable heavy chain encoding sequences were sequenced using primer LSN-HCP: AGGAAACAGGAGATATACAT (SEQ ID NO131), annealing to the P *tac* promoter, and the light chain encoding sequences were sequenced using primer LSN-LCP: TCGCCAAGGAGACAGTCATA (SEQ ID NO132), annealing to the P *lac* promoter. The sequence data was analysed using the Vector NTI software (Informax, Frederick, MD, USA). The resulting sequence data of the heavy chain was trimmed to one base pair 5' of the upstream AscI restriction site and immediately 3' of the downstream XhoI restriction site. The light chain sequence data was trimmed to the second 5' base pair of the upstream NheI restriction site and 3' of the last codon "AAA" encoding the variable chain C-terminal lysine.

The trimming was performed in order to facilitate further analysis such as translation of each DNA sequence.

[00416] The variable heavy and light chain encoding sequences were analyzed for germline gene usage by comparing the sequences to the V-base germline variable region sequence database (MRC Centre for Protein Engineering, Cambridge, UK). The closest related germline allele was thus determined for each sequence showing a V-gene repertoire originating from 12 different variable heavy chain germline alleles belonging to VH-family VH1 through VH5 and 8 different variable kappa light chain germline alleles belonging to VKI, VKIII and VKIV (Table 20).

[00417] Furthermore, the variable gene sequences were translated into protein sequences which were aligned using the AlignX software (Informax, Frederick, MD, USA) as depicted in Figure 22 and Figure 23. Based on the protein sequence alignments, the variable chain sequences could be categorized into groups according to V(D)J rearrangement events, designated with H for the variable heavy chain sequence and L for the variable kappa chain sequence followed by a unique number. In each rearrangement group, the sequences were categorized according to the maturation genotype (M-type in table 20) within the CDR1, 2 and 3 regions, represented by a lower case letter in alphabetical order. Seven of the clones had premature stop codons, Whereof 6 clones were an amber mutation (TAG) (clone IDs g060:b12, d08, f06, c12, f03, c04) and 1 clone was an opal mutation (TGA) (clone ID g060: h12). These codons were most likely suppressed by *E. coli* resulting in functional Fab fragments. Four of these clones were members of groups containing similar clones making them redundant. Additional clones would have to be analyzed in order to replace the remaining 3 clones with premature stop codons as they were single members of their groups. Alternatively, the sequences could be corrected by standard molecular biology techniques such as PCR, replacing the stop codon with an appropriate codon.

[00418] The 47 analyzed V-region sequences could be divided into 20 unique V(D)J rearrangement groups (designated "groups" in table 20) both for the variable heavy and variable kappa gene. Four of the rearrangement groups could further be divided into 1 to 4 maturation types (a through d), resulting in 27 unique antibody encoding sequences (Table12). Generally, specific heavy chain rearrangement groups combine with specific light chain rearrangement groups (e.g. H1 with L1, H12 with L24 and so forth). Further, the maturation type match among pairs of variable heavy chain and the variable light encoding sequences (e.g. H4c match with L13c). Such stringency in paring among rearrangement

groups and maturation types indicating cognate paring of the variable region encoding sequences.

[00419] However, heavy chain rearrangement group H4 is an exception. H4 pairs with two different light chains L28 and L13. L13 was unique to H4, comprising maturation types a, b and c that matches with L13 maturations types. L28, on the other hand, was also found paired with the single member in heavy chain group H2. Two out of seven H4a heavy chain sequences pair with heavy chain L13a and five out of seven H4a heavy chain sequences pair with L28a. In summary, these observations suggested a multiplex overlap-extension RT-PCR event where two TT-specific antibody producing plasma cells, with the genotype combinations H2-L28 and H4a-L13a, were present in a single well. Rare scrambling events of the light chain and heavy chain gene pairing, such as for H2 and H4a with L28, was expected as the experiment was based on limiting dilution of plasma cells.

[00420] The sequence identity among individual cognate pairs of the variable heavy chain and variable light chain encoding sequences from the same group and maturation type is at least 90 % and preferably at least 95 %. Take for example g060g03 and g060a01 from group H1 maturation type a. Clone g060g03 correspond to nucl. SEQ ID pair 168:215, where the variable heavy chain encoding sequence correspond to SEQ ID NO 168 and the variable light chain encoding sequences correspond SEQ ID NO 215. Clone g060a01 correspond to nucl. SEQ ID pair 133:180. When SEQ ID NO 168 is aligned with SEQ ID NO 133 (the variable heavy chains) 4/369 bases are not identical and for the variable lights chains (SEQ ID NOs 215 and 180) 8/327 bases are not identical, this correspond to a sequence identity of 98.3 % between these two cognate pairs (g060g03 and g060a01). However, when looking at sequence identity among different groups it is not expected that the sequence identity will be high, since it is a polyclonal sub-library and diversity is desired. In this particular example the lowest identity among the cognate pairs is approximately 40 % (e.g. g060b11 and g060h11 have a sequence identity of 39.5%).

[00421] One embodiment of the present invention is a sub-library of cognate pairs of immunoglobulin heavy chain variable region and light chain variable region encoding sequences, where the immunoglobulins obtainable from said library are capable of reacting with or binding to Tetanus Toxin.

[00422] A further embodiment of the present invention is such a sub-library of cognate pairs of immunoglobulin heavy chain variable region and light chain variable region encoding sequences, comprising individual cognate pairs with at least 90 % sequence identity with one

individual SEQ ID pair, selected from the group consisting of SEQ ID pairs 135:182, 168:215, 146:193, 151:198, 173:220, 152:199, 164:211, 148:195, 137:184, 169:216, 138:185, 143:190, 161:208, 166:213, 157:204, 139:186, 134:181, 150:197, 156:203, 158:205, 170:217, 178:225, 141:188 or 144:191.

5

15881YY00

Table 20

	VH						Vkappa				
Clone ID	Group	M-type	Family	Germ-line	Seq nucl.	Seq prot.	Group	M-type	Family	Germ-line	Seq nucl.	Seq prot.
g060a03	H1	b	VH1	DP-14	42135	136229	L1	b	VKI	DPK9	89182	183276
g060g03	H1	a	VH1	DP-14	75168	169262	L1	a	VKI	DPK9	122215	216309
g060a01	H1	a	VH1	DP-14	13340	134227	L1	a	VKI	DPK9	87180	181274
g060b12	H11	a	VH3	DP-77	52145	146239	L10	a	VKIII	DPK22	99192	193286
g060c03	H11	a	VH3	DP-77	53146	147240	L10	a	VKIII	DPK22	100193	194287
g060d08	H11	a	VH3	DP-77	61154	155248	L10	a	VKIII	DPK22	108201	202295
g060f06	H11	a	VH3	DP-77	69162	163256	L10	a	VKIII	DPK22	116209	210303
g060d04	H12	a	VH1	DP-10	58151	152245	L24	a	VKI	DPK4	105198	199292
g060h02	H12	a	VH1	DP-10	82175	176269	L24	a	VKI	DPK4	129222	223316
g060h10	H12	a	VH1	DP-10	84177	178271	L24	a	VKI	DPK4	131224	225318
g060g11	H12	b	VH1	DP-10	80173	174267	L24	b	VKI	DPK4	127220	221314
g060d05	H12	c	VH1	DP-10	59152	153246	L24	c	VKI	DPK4	106199	200293
g060f09	H12	d	VH1	DP-10	71164	165258	L24	d	VKI	DPK4	118211	212305
g060c07	H15	a	VH5	DP-73	55148	149242	L16	a	VKIII	DPK22	102195	196289
g060d11	H15	a	VH5	DP-73	62155	156249	L16	a	VKIII	DPK22	109202	203296
g060g02	H15	a	VH5	DP-73	74167	168261	L16	a	VKIII	DPK22	121214	215308
g060a09	H16	-	VH1	DP-73	44137	138231	L4	-	VKI	DPK9	91184	185278
g060g04	H17	a	VH4	DP-66	76169	170263	L3	a	VKI	DPK9	123216	217310
g060h04	H17	a	VH4	DP-66	83176	177270	L3	a	VKI	DPK9	130223	224317
g060a10	H18	a	VH2	VII-5	45138	139232	L21	a	VKI	L12a	92185	186279
g060h12	H18	a	VH2	VII-5	86179	180273	L21	a	VKI	L12a	133226	227320
g060b09	H19	a	VH2	VII-5	50143	144237	L23	a	VKI	DPK8	97190	191284
g060f10	H19	a	VH2	VII-5	72165	166259	L23	a	VKI	DPK8	119212	213306
g060f05	H2*	-	VH1	DP-88	68161	162255	L28*	a	VKIV	DPK24	115208	209302
g060f11	H21	a	VH3	DP-32	73166	167260	L31	a	VKI	DPK9	120213	214307
g060g12	H21	a	VH3	DP-32	81174	175268	L31	a	VKI	DPK9	128221	222315

15881YY00

	VH						Vkappa					
Clone ID	Group	M-type	Family	Germ-line	Seq nucl.	Seq prot.	Group	M-type	Family	Germ-line	Seq nucl.	Seq prot.
g060f01	H10	-	VH3	COS-8	64157	158251	L7	-	VKI	DPK8	111204	205298
g060a11	H23	a	VH3	DP-77	46139	140233	L5	a	VKI	DPK9	93186	187280
g060c12	H23	b	VH3	DP-77	56149	150243	L5	b	VKI	DPK9	103196	197290
g060f03	H24	-	VH3	DP-38	66159	160253	L20	-	VKI	DPK8	113206	207300
g060a02	H3	-	VH3	DP-77	41134	135228	L9	-	VKI	O16	88181	182275
g060d03	H4	a	VH1	DP-88	57150	151244	L13	a	VKIII	DPK22	104197	198291
g060f08	H4	a	VH1	DP-88	70163	164257	L13	a	VKIII	DPK22	117210	211304
g060b05	H4*	a	VH1	DP-88	47140	141234	L28*	a	VKIV	DPK24	94187	188281
g060b07	H4*	a	VH1	DP-88	49142	143236	L28*	a	VKIV	DPK24	96189	190283
g060f04	H4*	a	VH1	DP-88	67160	161254	L28*	a	VKIV	DPK24	114207	208301
g060g10	H4*	a	VH1	DP-88	79172	173266	L28*	a	VKIV	DPK24	126219	220313
g060d07	H4*	a	VH1	DP-88	60153	154247	L28*	a	VKIV	DPK24	107200	201294
g060e05	H4	c	VH1	DP-88	63156	157250	L13	c	VKIII	DPK22	110203	204297
g060f02	H4	b	VH1	DP-88	65158	159252	L13	b	VKIII	DPK22	112205	206299
g060g09	H4	c	VH1	DP-88	78171	172265	L13	c	VKIII	DPK22	125218	219312
g060c04	H6	-	VH1	DP-88	54147	148241	L12	-	VKIII	DPK22	101194	195288
g060g07	H20	-	VH4	DP-71	77170	171264	L8	a	VKI	DPK9	124217	218311
g060h11	H7	-	VH2	VII-5	85178	179272	L18	-	VKI	L12a	132225	226319
g060b06	H8	a	VH3	DP-46	48141	142235	L17	a	VKI	DPK1	95188	189282
g060a04	H8	a	VH3	DP-46	43136	137230	L17	a	VKI	DPK1	90183	184277
g060b10	H9	-	VH1	DP-14	51144	145238	L30	-	VKI	DPK1	98191	192285

Asterisk indicates V encoding sequences implicated in scrambling. Seq prot. indicates SEQ ID numbers from the sequence list corresponding to the protein sequences and Seq nucl. indicate SEQ ID numbers corresponding to the nucleic acid sequence in the sequence list

i. Apparent affinities

[00423] A competition assay was set up in order to determine the apparent affinity or IC_{50} of selected clones from plate G060.

[00424] Briefly, Fab fragments were expressed in 50 ml cultures as follows: 50 ml 2 x YT/100 µg/ml Carb/ 0.1 % glucose was added 0.5 ml overnight culture, and shaken for approximately 2 h at 37°C. IPTG was added to a final concentration of 0,1 mM, and shaking was continued overnight at 30°C. The next day, the bacteria were collected by centrifugation at 4000 rpm for 15 minutes, and the pellet was resuspended in 1 ml 0.8 mM EDTA, 0.4 x PBS, 0.8 M NaCl, and incubated on ice for 15 minutes. The periplasmic extract was collected by centrifugation, and stored at -20°C.

[00425] The competition assay was performed as follows: the periplasmic extracts in appropriate dilutions estimated by titration were added to a series of tubes. As a positive control the phage display derived Fab fragment mp584, derived from the human hybridoma cell line HB8501 expressing an anti-TT antibody was used. Soluble TT was added to the first tube at 100 nM concentration, and subsequently diluted at four-fold steps in the following tubes, making a total of seven dilutions of TT (from 100 nM to 25 pM). The reactions were incubated for approximately 45 minutes at room temperature. The samples were transferred to ELISA plates coated with TT at 1 µg/ml, and blocked as previously described. The plates were incubated for 1 h at room temperature, followed by 4x wash with PBS-T, goat-anti-human Fab/HRP was added at a 1:10.000 dilution, and incubated for 1 h. TMB Plus substrate (KemEnTech, Denmark, cat. No. 4390A) was added, incubation was performed for approximately 10 minutes, and the reactions were stopped with 1 M H_2SO_4. The plates were read at 450 nm. The data are plotted in Figure 24.

[00426] The apparent affinities of the analyzed clones are given in the table 21:

Table 21

Clone	Apparent affinity
g060a10	0.50 nM
g060b06	0.60 nM
g060b12	1.2 nM
g060c04	0.55 nM
g060d04	1.1 nM
g060f01	1.2 nM
g060g04	0.9 nM
g060g07	0.35 nM
g060h11	3.0 nM
mp584 (pos. cont.)	6.0 nM

[00427] As seen in table 21, the Fab fragments have apparent affinities in the lower nano-molar or upper pico-molar range. Furthermore, all cognate paired Fab fragments exhibit a higher apparent affinity than that of the phage display derived Fab fragment.

j. Summary

[00428] In donor TT03 the frequency of TT specific plasma cells in the peripheral blood monocyte cell fraction was calculated to 0.022% . Approximately 400 cognate pairs were generated from TT03. 3600 clones from the cells transformed with the cognate pair library were screened using ELISA, of these 339 clones showed TT reactivity in the ELISA screening. 47 of these clones have been analysed with respect to their clonal diversity. Of these 47 clones, 27 proved to resemble unique non-scrambled variable region encoding sequences. Three of these clones contained a premature stop codon that will need correction before transfer to a mammalian expression vector. The transfer to mammalian expression vectors was described in example 1 section h. The apparent affinities were measured on selected clones, ranging from the lower nano-molar to upper pico-molar range.

k. Prospects

[00429] The Tetanus Toxin is one of the most toxic substances known with a lethal dose of a few nanogram. The toxin is produced by Clostridium tetani, a soil bacteria also present in the digestive tract of up to 25% of humans. The tetanus immunization program has effectively abolished the disease in the western world, although 100-200 cases are still observed in the major western countries annually, with a case-fatality ratio of 50%. In the developing world the number of cases is significantly greater. Bacterial growth in contaminated penetrating wounds may lead to toxin release, ultimately leading to rigidity, spasms, respiratory arrest and death.

[00430] Hyperimmune immunoglobulin products isolated from human blood donors with a high titer of antibody response against Tetanus Toxoid may be used to prevent tetanus or if instituted early to treat established tetanus, also in conjunction with active immunization. However, due to shortage of human product, equine hyperimmune anti-Tetanus Toxoid is used in the developing world. Recombinant monoclonal or polyclonal antibodies against Tetanus Toxoid have potential to substitute hyperimmune globulin products for therapeutic and/or prophylactic use. Recombinant monoclonal antibodies originating from the conventional hybridoma technology has been described to be effective against TT (Chin, J. et al. 2003. Biologicals 31, 45-53). Interestingly, a synergistic effect was observed when mixing

two monoclonal antibodies. Thus, a recombinant polyclonal anti-Tetanus Toxoid antibody capable of reacting with or binding to Tetanus Toxin could potentially be very effective in treatment or prophylaxis protection of patients at risk of developing tetanus.

[00431] One embodiment of the present invention is a recombinant polyclonal immunoglobulin or fragments thereof capable of reacting with or binding to Tetanus Toxin.

[00432] A preferred embodiment of the present invention is a recombinant polyclonal immunoglobulin capable of reacting with or binding to Tetanus Toxin comprised of cognate pairs of immunoglobulin heavy chain variable region and light chain variable region.

[00433] A further embodiment of the present invention is a recombinant polyclonal immunoglobulin capable of reacting with or binding to Tetanus Toxin obtained by the method according to the present invention.

[00434] A further embodiment of the present invention is a recombinant polyclonal immunoglobulin capable of reacting with or binding to Tetanus Toxin comprising individual cognate pairs of immunoglobulin heavy chain variable region and light chain variable region with at least 90 % sequence identity with one individual SEQ ID pair, selected among SEQ ID pairs 229:276, 262:309, 240:287, 245:292, 267:314, 246:293, 258:305, 242:289, 231:278, 263:310, 232:279, 237:284, 255:302, 260:307, 251:298, 233:280, 228:275, 244:291, 250:297, 252:299, 264:311, 272:319, 235:282 or 238:285.

[00435] Another embodiment of the invention, is a pharmaceutical composition comprising a recombinant polyclonal antibody capable of reacting with or binding to Tetanus Toxin as active ingredient intended for the treatment or prevention of a tetanus. Preferably, the recombinant polyclonal antibody is combined with a pharmaceutically acceptable excipient.

[00436] A further embodiment of the present invention is the use of a recombinant polyclonal immunoglobulin capable of reacting with or binding to Tetanus Toxin as a medicament for the treatment or prophylactic protection of a patient at risk of developing tetanus.

[00437] An additional embodiment is a method of preventing or treating a patient at risk of developing tetanus by administering to a patient in need thereof a composition comprising a recombinant polyclonal antibody capable of reacting with or binding to Tetanus Toxin.

Example 12: Comparing results obtained from two donors

[00438] In the following example results obtained from donor TT08 are compared with the results obtained from donor TT03 in Example 11. The results are summarized in table 22.

Table 22

	TT03	TT08
TT specific plasma cells in PBMC fraction (ELISPOT)	0.021%	0.011%
CD19+ cells in PBMC fraction (FACS)	11%	5%
Plasma cells in PBMC fraction (FACS CD38hi, CD45in)	0.12%	0.08%
Estimated number of linked V_H and V_L sequences	400	400
Number of clones screened for Fab expression	1400	2000
Number of Fab positive clones (2 x background)	482	558
Number of clones screened for TT-specificity	3600	3400
Number of TT-specific clones (2 x background)	339	188
Number of sequences analyzed from TT-specific clones	47	102
VH germline alleles	12 (VH1-5)	NA
VL germline alleles	8 (VKI, III, IV)	NA
Number of unique cognate pairs	27	40
Apparent affinity range	0.35-3.0 nM	NA

NA=not analyzed

[00439] This clearly illustrated that libraries of similar quality could be isolated from two different donors immunized with TT.

[00440] The reason for the higher number of unique cognate pairs in the library obtained from donor TT08 is most likely due to the larger number of sequences analyzed.

Example 13: Comparing the library of cognate pairs from Example 11 with a combinatorial phage display library generated from the same donor

[00441] In the present example a combinatorial phage display library was prepared from the same donor previously used to prepare the library of cognate pairs, in order to compare library diversity, affinity and specificity between libraries of cognate pairs and combinatorial libraries.

a. Combinatorial phage library construction

[00442] A phage display library was generated from the $CD19^+$ fraction of cells from donor TT03 (identical to the cell fraction obtained in Example 11c).

[00443] Total RNA was prepared from app. $5x10^6$ $CD19^+$ cells using NucleoSpin RNA L kit (Machery-Nagel cat. no. 740 962.20). cDNA was subsequently synthesized in an oligo(dT)-primed reaction using ThermoScript reverse transcriptase (Invitrogen cat. no. 11146-016).

[00444] The V_H and Kappa chains were PCR amplified using HotMasterTaq DNA polymerase (Eppendorf cat. no. 0032 002.692) and primers essentially as described by de Haard *et al.* (J. Biol. Chem. 274, 18218-18230; 1999), only modified with respect to restriction enzyme recognition sequences in the 5'-end.

[00445] The combinatorial phage display library was generated by successive insertion of Kappa and V_H PCR products into Em351 phage display vector (modified from phh3 described in Den,W. et al. 1999 J. Immunol. Methods 222, 45-57).

[00446] The final library was electroporated into TG1 *E.coli* strain (Stratagene). The size of the combinatorial library contained 3 x 10^6 independent clones with a high insert frequency.

b. Panning of the combinatorial library

[00447] Fab displaying phage particles were prepared according to standard procedures (e.g. Antibody Engineering, A Practical Approach 1996, ed. McCafferty, Hoogenboom and Chiswell).

[00448] Panning was performed on tetanus toxoid (TT; SSI batch no. 89-2) diluted to 1 μg/mL in PBS, and immobilized in MaxiSorp immunotubes (Nunc cat. no. 444202). Following a one hour incubation period and several washing steps, bound phage particles were eluted using 100 mM TEA (Triethylamine).

[00449] The phage particle eluate was neutralized, and used to infect exponentially growing TG1 cells, from which Fab displaying phage particles enriched for TT specificity were obtained. A second round of panning using the eluted phages was performed following the general procedure outlined above.

[00450] In parallel, three panning rounds were performed on the C-fragment of the tetanus toxin molecule (Sigma cat. no. T3694), following the procedure outlined above.

[00451] Single colonies were screened for binding to tetanus toxoid and the C-fragment from the unselected library, and after each panning round, from both sets of panning described above.

c. Comparing specificity and affinity of combinatorial phage particle clones and cognate pair clones

[00452] Single colonies were picked from the unselected library and after each round of panning (on intact tetanus toxoid (TT) and the tetanus toxoid C-fragment, respectively). Fab displaying phage particles were analyzed for reactivity by ELISA assays. The number of Fab-positive clones exhibiting TT and/or C-fragment reactivity of at least 2 x background

reactivity and at least 4 x background reactivity are given in table 23 to 25 below. All the results are shown as number of specific clones/number of Fab positive clones.

Table 23: Clones selected on TT and analyzed by TT-specific ELISA

Enrichment	2 x background	4 x background
Unselected clones	-	13/app. 5000
After 1 round of panning	61/99	59/93
After 2 rounds of panning	165/169	163/166

Table 24: Clones selected on TT and analyzed by C-fragment-specific ELISA

Enrichment	2 x background	4 x background
Unselected clones	0/13	0/13
After 1 round of panning	0/85	0/85
After 2 rounds of panning	0/85	0/85

Table 25: Clones selected on C-fragment, analyzed by C-fragment-specific ELISA

Enrichment	2 x background	4 x background
After 1 round of panning	1/35	0/35
After 2 rounds of panning	12/47	9/47
After 3 rounds of panning	64/126	26/126

[00453] Panning of the phage display library against TT revealed an increasing number of TT specific clones when increasing the number of panning rounds, and only a few clones were identified in the unselected library. No clones were found to be reactive with the teatanus toxoid C-fragment from the unselected library or after two rounds of panning of the library against TT. To obtain Fab fragments with C-fragment specificity from the combinatorial phage display library, this library had to be panned against the C-fragment specifically.

[00454] For comparison, thirteen TT-specific clones from Example 11 were subjected to C-fragment-specific ELISA, of these seven showed reactivity above 2 x background. Thus, Fab fragments with specificity towards the tetanus toxin C-fragment could be expressed from a library of cognate pairs obtained from a TT-immunized individual.

[00455] This clearly illustrates the disadvantage of using panning to identify clones with specificity towards a particular antigen. If important antigen fragments or epitopes are unknown, they may be discharged during panning, thereby resulting in a less efficient product in the end.

[00456] Further, apparent affinities of the combinatorial clones were measured as described in the assay of Example 11i. Ten of the combinatorial clones obtained after two rounds of panning on TT were analyzed, revealing apparent binding affinities between 1 and 15 nM.

[00457] For comparison four of the nine clones analyzed from the library of cognate pairs (table 21), showed affinities in the pico-molar range.

[00458] This indicates that the pairing of variable regions as originally selected for by the donors immune system, in combination with the somatic hyper-mutations the pairs have been subjected to as a pair, potentially results in higher apparent binding affinities than random combinations of such variable regions.

d. Comparing sequences of TT specific combinatorial phage particle clones and cognate pair clones

[00459] The large amount of sequence data generated from the two libraries precludes direct comparisons of raw sequences. In order to visualize the difference between V_H and V_L sequence pairs in the phage display library an in the library of cognate pairs, phylogenetic threes were generated for the V_H and V_L sequences, and pairs were illustrated in a dot matrix. The pairing of phylogenetic information in a dot matrix (Figure 25) revealed very different distribution profiles of V_H and V_L sequence pairs in the two libraries. The scattered appearance of the V_H and V_L sequence pairs in the phage display library (Figure 25A) indicated little phylogenetic relation between V_H and V_L genes in agreement with the random paring of the V genes in this library. In contrast, V_H and V_L sequence pairs from the library of cognate pairs (Figure 25B) show a clustered appearance indicating co-evolution of V genes as expected for cognate pairs. Also, the genetic diversity is much greater for the V-genes in the library of cognate pairs compared to the V-genes from the combinatorial library, indicating that the method of isolating cognate paired V-genes is less biased.

Example 14: Combined Single-step Multiplex Overlap-extension RT-PCR and Nested PCR Using T-cells as Template Source

[00460] In this example, it is described how a single-step multiplex overlap-extension RT-PCR can be performed on T lymphocytes derived from a human donor.

a. Obtaining a lymphocyte containing cell fraction

[00461] A blood sample is obtained from human donors who have been subjected to the desired antigen, for example by immunization, natural infection, malignancy, through an autoimmune reaction, or other diseases. The peripheral blood mononuclear cells (PBMC's) are isolated using Lymphoprep (Axis-Shield, Oslo Norway, prod. No 1001967) according to the manufacturer's instructions.

[00462] In the present example antigen-specific T cells are generated by further stimulation of the PBMC fraction. However, the multiplex overlap-extension RT-PCR can also be performed directly on single cells from the PBMC fraction or on a cell fraction enriched for T cells (e.g. by FACS-sorting for CD3-positive cells).

b. Generating antigen-specific T cells

[00463] The PBMC cell fraction is re-suspended in an appropriate culture medium containing relevant cytokines such as IL2. Further, the desired antigen is added to the culture, where it will be presented to the T lymphocytes by antigen presenting cells (APC) present in the PBMC fraction. Alternatively, APC feeder cells treated in beforehand such that they are presenting the desired antigen may be added to the PBMC fraction. Such APC feeder cells can be APC exposed to the desired antigen in the form of for example peptide, protein or other molecular form, or microbially infected APC or cells transfected to express and present the antigen, or APC co-cultured with other antigenic cells, e.g. such as cancer cell in the form of primary tissue or cell lines. Many different types of APC feeder cell are known from the literature, including transformed cell lines, B cell lines, dendritic cells, etc.

[00464] The PBMC fraction is cultivated for approximately 3 to 5 weeks, during which fresh antigen presenting cells are added together with cytokines, e.g. on a weekly basis. This results in T lymphocyte proliferation, activation and maturation. At the end of the cultivation period the cell culture will be dominated by antigen-specific T cells. Whether these cells are CD4+ or CD8+ depends on the disease, the antigen, the APC cells, and the cytokine mixtures used during the stimulation period.

[00465] The antigen specificity can for example be tested with a CTL assay, proliferation assays and MHC tetramers loaded with the desired antigen (Altman, J.D., et al. 1996. Science 274, 94-96).

[00466] The antigen-specific T cells are distributed to PCR tubes, either by limiting dilution or using a FACS in order to obtain a single cell pr. vessel. The vessel can be stored at -80°C until use.

b. Single-step Multiplex Overlap-extension RT-PCR

[00467] The Qiagen One-Step RT-PCR kit (Qiagen cat. No 210212, Hilden, Germany) is used for the multiplex overlap-extension RT-PCR essentially according to the manufacturer's recommendation. Before addition of PCR reaction mixture to the PCR tubes, the cells are thawed.

[00468] PCR reaction mixtures and cycling conditions are initially set up as described in Example 11 e-1. However, a certain amount of optimization may be expected for a new set of primers.

[00469] The multiplex overlap-extension primer mix used comprise the primers shown in table 26.

Table 26

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
Vα	10 nM of each	322	tattggcgcgccatggccGCCCAGTCTGTGACCCAGC
		323	tattggcgcgccatggccGCCCAGAAGATAACTCAA
		324	tattggcgcgccatggccGATGCTAAGACCACCCAG
		325	tattggcgcgccatggccGCCCAGACAGTCACTCAG
		326	tattggcgcgccatggccAAACAGGAGGTGACACAGA
		327	tattggcgcgccatggccGGAGACTCGGTTACCCAGA
		328	tattggcgcgccatggccGGAGATTCAGTGACCCAGA
		329	tattggcgcgccatggccAGCAATTCAGTCAAGCAGA
		330	tattggcgcgccatggccGGACAAAACATTGACCAG
		331	tattggcgcgccatggccAAAAATGAAGTGGAGCAGA
		332	tattggcgcgccatggccGAAGACCAGGTGACGCAGA
		333	tattggcgcgccatggccGGAATACAAGTGGAGCAGA
		334	tattggcgcgccatggccCTACATACACTGGAGCAGA
		335	tattggcgcgccatggccGAAGACAAGGTGGTACAAA
		336	tattggcgcgccatggccGAAAACCAGGTGGAGCACA
		337	tattggcgcgccatggccCAGGAGAATGTGGAGCAG
		338	tattggcgcgccatggccGGAGAGAGTGTGGGGCTG
		339	tattggcgcgccatggccGGAGAGGATGTGGAGCAGA
		340	tattggcgcgccatggccAGCCAAAAGATAGAACAGA
		341	tattggcgcgccatggccAGTCAACAGGGAGAAGAG
		342	tattggcgcgccatggccCAACAACCAGTGCAGAGT
		343	tattggcgcgccatggccAGCCAAGAACTGGAGCAGA
		344	tattggcgcgccatggccGGTCAACAGCTGAATCAGA
		345	tattggcgcgccatggccACCCAGCTGCTGGAGCAGA
		346	tattggcgcgccatggccCAGCAGCAGGTGAAACAA
		347	tattggcgcgccatggccATACTGAACGTGGAACAA
		348	tattggcgcgccatggccGACCAGCAAGTTAAGCAA
		349	tattggcgcgccatggccGGACAACAGGTAATGCAA
		350	tattggcgcgccatggccAAGGACCAAGTGTTTCAG

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
Cα	200 nM	351	TAGGCAGACAGACTTGTCACT
Vβ	10 nM of each	352	gccatggcgcgccaatagctagccGGTGAAGAAGTCGCCCAGA
		353	gccatggcgcgccaatagctagccGATGCCATGGTCATCCAGA
		354	gccatggcgcgccaatagctagccGAAGCCCAAGTGACCCAGA
		355	gccatggcgcgccaatagctagccCATGCCAAAGTCACACAGA
		356	gccatggcgcgccaatagctagccGACACAGCCGTTTCCCAGA
		357	gccatggcgcgccaatagctagccGAAACGGGAGTTACGCAGA
		358	gccatggcgcgccaatagctagccGAACCTGAAGTCACCCAGA
		359	gccatggcgcgccaatagctagccGAAGCTGACATCTACCAGA
		360	gccatggcgcgccaatagctagccATATCTGGAGTCTCCCACA
		361	gccatggcgcgccaatagctagccGATGCTGGAATCACCCAGA
		362	gccatggcgcgccaatagctagccAATGCTGGTGTCACTCAGA
		363	gccatggcgcgccaatagctagccAGTGCTGTCGTCTCTCAA
		364	gccatggcgcgccaatagctagccGACGCTGGAGTCACACAA
		365	gccatggcgcgccaatagctagccGATGGTGGAATCACTCAGT
		366	gccatggcgcgccaatagctagccACTGCTGGGATCACCCAG
		367	gccatggcgcgccaatagctagccGAAGCTGGAGTTGCCCAGT
		368	gccatggcgcgccaatagctagccGAAGCTGGAGTGGTTCAGT
		369	gccatggcgcgccaatagctagccGAAGCTGGAGTTACTCAGT
		370	gccatggcgcgccaatagctagccGATGCTGTAGTTACACAA
		371	gccatggcgcgccaatagctagccGATGCTGGAGTTATCCAGT
		372	gccatggcgcgccaatagctagccGGTGCTGGAGTCTCCCAG
		373	gccatggcgcgccaatagctagccGATGCTGATGTTACCCAGA
		374	gccatggcgcgccaatagctagccTCTCAGACTATTCATCAAT
Cβ	200 nM	375	CAGGCACACCAGTGTGGCCTT

c. Semi-Nested PCR

[00470] Semi-nested PCR is likewise proposed to be performed as described in Example 11 e-2, and some optimization may be expected.

[00471] The primers used are shown in table 27.

Table 27

Primer name	Conc.	SEQ ID NO	primer sequence in 5' to 3' direction
Cα	200 nM	376	cacttagagctcTTAGAGTCTCTCAGCTGGTACAC
Cβ	200 nM	377	gacattatgCAtGATCTCTGCTTCTGATGGCTC

Capitalized sequences correspond to the gene-specific region.

[00472] In order to verify that the multiplex overlap-extension RT-PCR is successful a proportion of the samples from the semi-nested PCR reactions are analyzed by subjecting 10 µl from each semi-nested PCR reaction to 1% agarose gel electrophoresis using ethidium bromide for detection. The expected size of the overlap-extension fragment is approximately 850bp (the exact size depends upon the lengths of the variable regions).

[00473] Approximately ten microliters from all the performed reactions originating from the same donor, are consolidated in a single tube. An aliquot of the pooled PCR products is subsequently purified using QIAquick PCR purification kit according to the manufactures procedure (Qiagen cat. No. 28106, Hilden, Germany. The purified pool of overlap PCR products can be digested with appropriate restriction enzymes and subsequently purified and inserted into a suitable vector.

[00474] In the present experiment constant region primers (SEQ ID NO 376 and 377) used in the semi-nested PCR are designed for sub-cloning of semi-nested PCR product into a suitable vector. The design of the Cβ primer relies on changing a SER to Met at pos. 21 in the β chain constant region peptide, whereby a NsiI site can be introduced into the nucleic acid sequence:

```
                NsiI
tcc cac  -> atg cat
SER HIS     MET HIS
```

[00475] This transition should be relative safe since SER 21 is exposed on the β chain constant region in a loop structure at the membrane proximal end of the domain. This relative conservative change is therefore not likely to disturb the overall domain structure.

[00476] The design of the Cα primer relies on changing nucleotides corresponding to pos. 15-17 in the α chain constant region peptide, whereby a SacI site can be introduced into the nucleic acid sequence:

```
                    SacI
aaa tcc agt -> aag agc tct
LYS SER SER    LYS SER SER
```

[00477] Suitable vectors will therefore contain the remaining parts of the constant regions of the TcR α and β chains and will contain appropriate modification in terms of restriction sites. These changes can be performed using standard PCR and sub-cloning techniques.

e. Additional Considerations

[00478] The large amount of variable region primers may potentially interact in an inhibitory manner during the multiplex overlap-extension PCR amplification. To avoid this, variable region primers may be divided into sub-sets and used separately in proper combinations.

Other Embodiments

[00479] All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

[00480] What is claimed is:

Claims

1. A method of linking a plurality of non-contiguous nucleotide sequences of interest said method comprising:

a) amplifying, in a multiplex molecular amplification procedure, nucleotide sequences of interest using a template derived from an isolated single cell or a population of isogenic cells; and

b) effecting linkage of the nucleotide sequences of interest amplified in step a).

2. The method according to claim 1, wherein the nucleotide sequences of interest comprise variable region encoding sequences and the linkage generates a cognate pair of variable region encoding sequences.

3. The method according to claim 1 or 2, wherein the nucleotide sequences of interest comprise immunoglobulin variable region encoding sequences and the linkage generates a cognate pair of a light chain variable region encoding sequence associated with a heavy chain variable region encoding sequence.

4. The method according to claim 1 or 2 , wherein the nucleotide sequences of interest comprise T cell receptor variable region encoding sequences and the linkage generates a cognate pair constituted of an alpha chain variable region encoding sequence associated with a beta chain variable region encoding sequence or a gamma chain variable region encoding sequence associated with a delta chain variable region encoding sequence.

5. A method of linking a plurality non-contiguous nucleotide sequences of interest randomly said method comprising:

a) amplifying, in a multiplex molecular amplification procedure, nucleotide sequences of interest using a template derived from a population of genetically diverse cells; and

b) effecting linkage of the nucleotide sequences of interest amplified in step a).

6. The method according to claim 5, wherein said population of cells is lysed.

7. The method according to claim 5 or 6, wherein said nucleotide sequences of interest comprise variable region encoding sequences and the linkage generates a combinatorial library of pairs of variable region encoding sequences.

8. The method according to claim 7, wherein said nucleotide sequences of interest comprise immunoglobulin variable region encoding sequences and the linkage generates a combinatorial library of pairs of light chain variable region and heavy chain variable region encoding sequences.

9. The method according to claim 7, wherein the nucleotide sequences of interest comprise T cell receptor variable region encoding sequences and the linkage generates a combinatorial library of pairs of α chain variable region and β chain variable region encoding sequences.

10. The method according to any one of the preceding claims, wherein said multiplex molecular amplification procedure is a multiplex RT-PCR amplification.

11. The method according to claim 10, wherein said multiplex RT-PCR amplification is a two step process comprising a separate reverse transcription (RT) step prior to the multiplex PCR amplification.

12. The method according to claim 10, wherein said multiplex RT-PCR amplification is performed in a single step comprising initially adding all the components necessary to perform both reverse transcription (RT) and multiplex PCR amplification into a single vessel.

13. The method according to any one of the preceding claims, wherein said linkage of the nucleotide sequences of interest is performed in the same vessel as the multiplex molecular amplification.

14. The method according to any one of claims 10 to 13, wherein said linkage of the nucleotide sequences of interest is effected in association with the multiplex PCR amplification, utilizing a multiplex overlap-extension primer mix.

15. The method according to any one of the preceding claims, wherein said linkage of the nucleotide sequences of interest is effected by ligation.

16. The method according to any one of the preceding claims, wherein an additional molecular amplification, utilizing a primer mix adapted for amplifying the linked nucleic acid sequences of interest, is performed.

17. The method according to claim 14, wherein the multiplex overlap-extension primer mix comprises primer sets wherein at least one primer set member of each primer set comprises an overlap-extension tail capable of hybridizing to the overlap-extension tail of a primer set member of a second primer set.

18. The method according to claim 14 or 17, wherein the multiplex overlap-extension primer mix comprise:

a) at least one C_L or J_L primer complementary to the sense strand of an immunoglobulin light chain region encoding sequence;
b) at least one V_L 5' primer or V_{LL} primer complementary to the antisense strand of an immunoglobulin light chain variable region encoding sequence or light chain variable region leader sequence, and capable of forming a primer set with the primer(s) in step a);
c) at least one C_H or J_H primer complementary to the sense strand of an immunoglobulin constant heavy chain domain encoding sequence or a heavy chain joining region encoding sequence; and
d) at least one V_H 5' primer or V_{HL} primer complementary to the antisense strand of an immunoglobulin heavy chain variable region encoding sequence or heavy chain variable region leader sequence, and capable of forming a primer set with the primer(s) in step c)

19. The method according to claim 18, wherein the primers of step b) are V_{LL} primers with at least 90% sequence identity with the gene-specific region of SEQ ID 93 to 98, and the primers of step d) are V_{HL} primers with at least 90% sequence identity with the gene-specific region of SEQ ID 86 to 92.

20. The method according to any one of the preceding claims, wherein the single cell, the population of isogenic cells or the population of genetically different cells is obtained from a lymphocyte containing cell fraction.

21. A method of producing a library of cognate pairs comprising linked variable region encoding sequences, said method comprising:

a) providing a lymphocyte-containing cell fraction from a donor;

b) optionally enriching for a particular lymphocyte population from said cell fraction;

c) obtaining a population of isolated single cells, comprising distributing cells from said cell fraction individually into a plurality of vessels; and

d) amplifying and effecting linkage of the variable region encoding sequences contained in said population of isolated single cells, according to a method of any one of claims 1 to 4 or 10 to 20, insofar as these are dependent on any one of claims 1 to 4.

22. The method according to claim 21 wherein the individual isolated single cell in the population of single cells is expanded to a population of isogenic cells, prior to performing amplification and linkage (step d).

23. The method according to any one of claims 20 to 22, wherein the lymphocyte-containing cell fraction constitutes whole blood, bone marrow, mononuclear cells, or white blood cells.

24. The method according to any one of claims 20 to 23, wherein the lymphocyte-containing cell fraction is enriched for cells of the B lymphocyte lineage.

25. The method according to any one of claims 20 to 24, wherein the lymphocyte-containing cell fraction or B lymphocyte linage is enriched for plasma cells.

26. The method according to any one of claims 20 to 25, wherein cells from the lymphocyte-containing cell fraction, B lymphocyte linage or plasma cells are enriched for antigen-specificity.

27. The method according any one of claims 20 to 23 , wherein the lymphocyte-containing cell fraction is enriched for cells of the T lymphocyte lineage.

28. The method according any one of claims 20 to 23 , wherein antigen-specific T cells are generated by stimulation of the lymphocyte-containing cell fraction.

29. The method according to any one of the preceding claims, further comprising inserting the linked nucleotide sequences or a library of cognate pairs into a vector.

30. The method according to claim 29, wherein said vector is selected among cloning vectors, shuttle vectors, display vectors or expression vectors.

31. The method according to claim 29 or 30, wherein the linked nucleotide sequences or the individual members of the library of cognate pairs comprise an immunoglobulin heavy chain variable region encoding sequence associated with light chain variable region encoding sequence and said sequences are inserted in-frame into a vector already containing sequences encoding one or more immunoglobulin constant domains or fragments thereof.

32. The method according to claim 29 or 30, wherein the linked nucleotide sequences or the individual members of the library of cognate pairs comprise a T cell receptor α chain variable region encoding sequence associated with a β chain variable region encoding sequences and said sequences are inserted in-frame into a vector already containing sequences encoding one or more T cell receptor constant domains or fragments thereof.

33. The method according to any one of claims 21 to 32, further comprising creating a sub-library by selecting a subset of cognate pairs of linked variable region sequences that encode binding proteins with a desired target specificity, generating a library of target-specific cognate pairs of variable region encoding sequences.

34. The method according to any one of claims 31 to 33, further comprising transferring said cognate pair or library of target-specific cognate pairs of variable region encoding sequences to a mammalian expression vector.

35. The method according to claim 34, wherein the mammalian expression vector encodes one or more constant region domains selected from human immunoglobulin classes IgA1, IgA2, IgD, IgE, IgG1, IgG2, IgG3, IgG4, IgM, kappa light chain or lambda light chain or from human T cell receptor alpha, beta, delta and/or gamma chains.

36. The method according to one of the claims 29 to 35, further comprising the steps:

a) introducing a vector encoding a segment of linked nucleotide sequences into a host cell;

b) cultivating said host cells under conditions adapted for expression; and

c) obtaining the protein product expressed from the vector inserted into said host cell.

37. The method according to claim 36 wherein said protein product is a monoclonal antibody comprising a cognate pair of a light chain variable region associated with a heavy chain variable region.

38. The method according to claim 36, wherein said polypeptide product is a monoclonal T cell receptor comprising a cognate pair of an alpha chain variable region associated with a beta chain variable region.

39. A library of cognate pairs consisting of of linked variable region encoding sequences.

40. The library according to claim 39, wherein said cognate pairs of variable region encoding sequences are obtained by the method according to claim 21 or any claim dependent thereon.

41. The library according to claims 39 or 40, wherein an individual member of said cognate pairs comprise an immunoglobulin light chain variable region encoding sequence associated with a heavy chain variable region encoding sequence.

42. The library according to claim 41, wherein said individual member encode a full-length antibody selected from human immunoglobulin classes IgA1, IgA2, IgD, IgE, IgG1, IgG2, IgG3, IgG4 or IgM.

43. The library according to claims 39 or 40, wherein an individual member of said cognate pairs comprise a TcR alpha chain variable region encoding sequence associated with a beta chain variable region encoding sequence or a gamma chain variable region encoding sequence associated with delta chain variable region encoding sequence.

44. The library according claim 40, wherein said individual member encode a full-length TcR.

45. A sub-library of cognate pairs of variable region encoding sequences which encode for proteins exhibiting desired binding specificities directed against a particular target.

46. A sub-library which encode for proteins exhibiting desired binding specificities directed against a particular target, selected from a library according to one of claims 39 to 44.

47. A sub-library of cognate pairs of linked immunoglobulin heavy chain variable region and light chain variable region encoding sequences, where the immunoglobulins expressible from said library are capable of reacting with, or binding to Tetanus Toxin.

48. A population of host cells comprising a library or sub-library according to any one of the claims 39 to 47.

49. The population of host cells according to claim 48, wherein the cells are mammalian cells.

50. A recombinant polyclonal protein expressed from host cells according to claim 48 or 49.

51. A recombinant polyclonal immunoglobulin or fragments thereof capable of reacting with or binding to Tetanus Toxin.

52. The recombinant polyclonal immunoglobulin according to claim 51, wherein the individual members are cognate pairs.

53. The recombinant polyclonal immunoglobulin according to claim 51 or 52, comprising at least two individual cognate pairs comprising an immunoglobulin heavy chain variable region encoding sequence associated with a light chain variable region encoding sequence with at least 90 % identity to one individual SEQ ID pair, selected from the group consisting of SEQ ID pairs 229:276, 262:309, 240:287, 245:292, 267:314, 246:293, 258:305, 242:289, 231:278, 263:310, 232:279, 237:284, 255:302, 260:307, 251:298, 233:280, 228:275, 244:291, 250:297, 252:299, 264:311, 272:319, 235:282 or 238:285.

54. A recombinant immunoglobulin or fragments thereof capable of reacting with or binding to Tetanus Toxin, wherein the complementarity determining regions (CDR1,CDR2 and CDR3) are derived from one or more of the SEQ ID pairs of claim 53.

55. A pharmaceutical composition comprising a recombinant polyclonal immunoglobulin according to claim 50 as active ingredient, combined with at least one pharmaceutically acceptable excipient.

56. A pharmaceutical composition comprising a recombinant polyclonal antibody capable of reacting with or binding to Tetanus Toxin as active ingredient, optionally combined with a pharmaceutically acceptable excipient.

57. A pharmaceutical composition according to claim 55 or 56 for use as a medicament.

58. A method of preventing or treating a patient at risk of developing tetanus by administering to a patient in need thereof a composition comprising a recombinant polyclonal antibody capable of reacting with or binding to Tetanus Toxin.