

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
26 February 2004 (26.02.2004)

PCT

(10) International Publication Number
WO 2004/016811 A2

(51) International Patent Classification⁷: **C12Q 1/68**

(21) International Application Number:
PCT/DK2003/000545

(22) International Filing Date: 15 August 2003 (15.08.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/404,132 19 August 2002 (19.08.2002) US

(71) Applicant (for all designated States except US): **DAN-MARKS TEKNISKE UNIVERSITET (DTU)** [DK/DK]; Bygning 101A, Anker Engelunds Vej 1, DK-2800 Lyngby (DK).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **NIELSEN, Jens** [DK/DK]; Hans Bruuns Vej 8, DK-2920 Charlottenlund (DK). **MORTENSEN, Uffe** [DK/DK]; Sveasvej 8, 2. sal, lejl. 3, DK-1917 Frederiksberg (DK). **BORGES, Ana, Claudia, Cantisani** [BR/BR]; Rua Hilda Lucena 60, PB 58043 Joao Pessoa (BR). **NIELSEN, Michael, Lyngø** [DK/DK]; Kochsvej 27, 3.tv., DK-1917 Frederiksberg C (DK).

(74) Agent: **HØIBERG A/S**; Store Kongensgade 59A, DK-1264 Copenhagen K (DK).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to the identity of the inventor (Rule 4.17(i)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations
- of inventorship (Rule 4.17(iv)) for US only

Published:

- without international search report and to be republished upon receipt of that report

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

(54) Title: A METHOD AND A KIT FOR GENOME-WIDE GENE SHUFFLING USING TAGGED PRIMER TECHNOLOGY

(57) Abstract: The present invention relates to methods and kits for gene shuffling, as well as to shuffled genes. In particular, the methods comprise amplification of the sequences to be shuffled using affinity tagged primers, separation of the strands of the amplified sequences using the affinity tag and cleavage of the strands. The strands may be cleaved as single stranded nucleic acids or heteroduplexes may be formed from the strands and the double stranded heteroduplexes may be cleaved. Shuffled genes may be obtained by using the cleaved fragments as both primers and templates in a nucleotide polymerisation reaction, for example PCR. The kits of the present invention comprises primers and enzymes suitable for performing the methods.

WO 2004/016811 A2

A method and a kit for genome-wide gene shuffling using tagged primer technology

Field of invention

5

The present invention relates to a method and a kit for shuffling of genes, the method and the kit being based on shuffling of single-stranded DNA to reduce the occurrence of homoduplexes. All patent and non-patent references cited in the application are hereby incorporated by reference in their entirety.

10

Background of invention

Methods for gene shuffling are generally known in the art.

15

DNA shuffling has provided a paradigm shift in recombinant nucleic acid generation, manipulation and selection. Fast artificial evolution methodologies for generating improved industrial, agricultural, and therapeutic genes and encoded proteins have been developed. These methods, and related compositions and apparatus for practising these methods represent an important advance in the field of molecular biology.

20

25

A number of publications describe DNA shuffling. For example, Stemmer et al. (1994) "Rapid Evolution of a Protein", Nature 370: 389-391; Stemmer (1994) "DNA Shuffling by Random Fragmentation and Reassembly: in vitro Recombination for Molecular Evolution", Proc. Natl. Acad. USA 91: 10747-10751; Stemmer U. S. Patent No. 5,603,793 METHODS FOR IN VITRO RECOMBINATION; Stemmer et al. U. S. Pat. No. 5,830,721 DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY; Stemmer et al., U. S. Pat. No. 5,811,238 METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND RECOMBINATION describe, e. g., in vitro and in vivo nucleic acid, DNA and protein shuffling in a variety of formats, e. g., by repeated cycles of mutagenesis, shuffling and selection, as well as methods of generating libraries of displayed peptides and antibodies.

30

Applications of DNA shuffling technology have also been developed. In addition to the publications noted above, Minshull et al., U. S. Pat. No. 5,837,458 METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING provides, e. g., for the evolution of metabolic pathways and the enhancement of bioprocessing through recursive shuffling techniques. Crameri et al. (1996),
5 "Construction And Evolution Of Antibody-Phage Libraries By DNA Shuffling" Nature Medicine 2 (1): 100-103 describe, e. g., antibody shuffling for antibody phage libraries.

10 Additional details regarding DNA Shuffling can be found in WO95/22625, WO97/20078, WO96/33207, WO97/33957, WO98/27230, WO97/35966, WO98/31837, WO98/13487, WO98/13485 and WO98/42832, as well as a number of other publications.

15 A number of publications further describe techniques which facilitate DNA shuffling, e. g., by providing for reassembly of genes from small fragments, or even oligonucleotides. For example, in addition to the publications noted above, Stemmer et al. (1998) U. S. Pat. No. 5,834,252 END COMPLEMENTARY POLYMERASE REACTION describe processes for amplifying and detecting a target sequence
20 (e.g., in a mixture of nucleic acids), as well as for assembling large polynucleotides from nucleic acid fragments.

Review of the foregoing publications reveals that forced evolution by gene shuffling is an important new technique with many practical and powerful applications.
25 However, existing methods for shuffling based on double-stranded DNA fragments suffer the inherent disadvantage that the majority of the "shuffled" sequences are one of the two or more wild-type, because these have a high tendency to re-assemble during shuffling. Consequently there is a need for developing methods for improving the shuffling of genes.

30 JP 2000-245473 (Marine Biotechnol Inst) discloses methods of gene shuffling with single-stranded DNA. As a specific example the reference discloses three hybrid genes obtained by cutting the two strands with one, two and three different restrictionenzymes. The examples cover gene shuffling with nahH and xylE, which
35 are approximately 80 % homologous. It is not entirely clear from the disclosure how

the single-stranded DNA has been isolated from the source. Kikuchi et al, 2000 (Gene 243: 133-137) discloses the same method for gene shuffling with single-stranded DNA. The single-stranded DNA has been obtained by directional cloning the two genes in opposite directions. Plasmids carrying the two opposite strands have been transformed into E. coli. Using a helper-phage, particles with single-stranded DNA have been obtained. This is a very laborious method for isolating the opposite strings of two homologous genes. Furthermore, as evidenced by the occurrence of wild-type genes among the shuffled genes the method does not result in an efficient elimination of the strands not intended to take part in the shuffling.

WO 98/32845 (Bioinvent AB) discloses a method for site directed introduction of mutations, using PCR, where one of the primers is biotinylated. The amplified doublestranded DNA oligonucleotides are captured on avidin and one of the strands is removed using alkaline denaturing. The strands with introduced mutations are re-assembled into a complete doublestranded DNA using overlapping homologous regions and PCR. The method is especially adapted for introduction of mutations into domains, e.g. in the manipulation of genes coding for antibodies.

US 6,159,687 (Novo Nordisk AS) discloses a method for gene shuffling, relying on a template switch during replication of a template. A primer, for example a random primer is extended and said extended primer is used as a primer in a new replication process. The extended primer may be separated from the templates using a biotin label. According to one embodiment one strand is isolated from one gene and the opposite strand from another gene using biotinylated primers. The mixing of the gene sequences is carried out using random primers or primers that can hybridise to the end of the strands. After a number of rounds in a thermocycler, the nucleotides can be separated and full length hybrid polynucleotides are obtained.

Thus, new techniques, which improve gene shuffling are highly desirable especially new techniques which address the problem of wild-type background among the shuffled genes.

Summary of invention

Accordingly, there is provided a method for shuffling of genes, said method comprising the steps of

- 5 selecting at least two polynucleotides to be shuffled,
performing one round of PCR on the polynucleotides with primer pairs, wherein one of the primers incorporated into each polynucleotide is an affinity-tagged primer, the affinity-tagged primers being selected so that one affinity-tagged primer is incorporated into one strand of one polynucleotide and the other affinity-tagged
10 primer is incorporated into the corresponding complementary strand of another polynucleotide,
isolating the synthesised double-stranded DNA polynucleotides using the affinity-tags,
separating the tagged from non-tagged DNA strands,
15 cleaving isolated tagged DNA strands or non-tagged DNA strands,
shuffling the cleaved DNA strands.

Preferably, the non-tagged DNA strands are cleaved.

- 20 As the method is only workable when the first round of PCR is performed successfully, and as the efficiency of PCR is often variable, since primers for PCR cannot be chosen freely, it is often preferable to use affinity-tagged primers in a subsequent reamplification step.

- 25 Accordingly, in a further aspect of the invention, there is provided a method for shuffling of genes, said method comprising the following steps:

- a) selecting at least two double stranded DNA polynucleotides to be shuffled,
b) amplifying the at least two double-stranded DNA polynucleotides using PCR with primers, each primer pair having a sequence for templating the selected
30 polynucleotide and an oligonucleotide tag, the oligonucleotide tags of the two primers of a pair being different, thereby obtaining amplified tagged polynucleotides,
c) performing one further round of PCR on each of the amplified tagged polynucleotides with primer pairs for templating the oligonucleotide tags, wherein
35 one of the primers thereby incorporated into each polynucleotide is an affinity-

tagged primer, the affinity-tagged primers being selected so that one affinity-tagged primer is incorporated into one strand of one polynucleotide and one affinity-tagged primer is incorporated into the corresponding complementary strand of another polynucleotide,

- 5 d) isolating the synthesised double-stranded DNA polynucleotides using the affinity-tags,
e) separating the tagged from non-tagged DNA strands,
f) cleaving isolated non-tagged DNA strands, and
g) shuffling the cleaved DNA strands.

10

By the word "different" when it refers to the oligonucleotide tags is intended that the tags are sufficiently different that a primer, which hybridises to one tag does not hybridise to another under PCR conditions.

- 15 In one embodiment of the invention, step g) involves shuffling cleaved single-stranded DNA strands.

This method according to the invention combines several advantages. Firstly, shuffling is performed with single-stranded DNA thereby minimising the chance of
20 homoduplex formation. Furthermore, a simple and efficient method for separating the two strands is used, namely the use of affinity tags incorporated into the genes by using affinity-tagged primers. By carefully selecting the affinity tag, it is possible to perform the separation of the two strands very efficiently, so that contamination of the shuffling composition with non-wanted strands is avoided. Thirdly, a two step
25 PCR method is used for amplifying and tagging the genes to be shuffled. The advantage is that the amplification step is separated from the step of affinity tagging the genes to be shuffled. Thereby the oligonucleotide tags can be designed optimally for the second round of PCR and the use of the costly, affinity tagged primers can be reduced.

30

In another embodiment of the present invention the method further comprises the step of preparing heteroduplexes of DNA by incubating non-tagged DNA strands under hybridisation conditions, wherein said step preferably is performed between steps e) and f). Hence, in said embodiment, step f) preferably involves cleaving the

newly formed heteroduplexes of non-tagged DNA strands and step g) preferably involves shuffling the double stranded cleaved heteroduplexes.

5 The method according to this embodiment has in addition to the above mentioned advantages also further advantages. Certain endonucleases (and optionally chemicals) preferably cleaves double stranded nucleic acids. Hence, cleavage may be directed to regions of the heteroduplex, which are perfectly complementary. The ends of the individually cleaved nucleic acids will thus mainly contain perfectly complementary overlaps with other cleaved nucleic acids (example thereof shown in
10 fig 21). This facilitates the subsequent shuffling process (see herein below).

According to an especially preferred embodiment of the invention, the forward primer of each pair used for amplifying the selected polynucleotides have identical first oligonucleotide tags. Similarly, it is an advantage that the reverse primer of each
15 pair used for amplifying the selected polynucleotides have identical second oligonucleotide tags. By using these embodiments, it is possible to perform the affinity tagging with just two primer pairs. Each primer pair consists of a forward and a reverse primer of which only one contains an affinity tag. The two forward primers are identical, except that one contains an affinity tag. Likewise, the two reverse
20 primers are identical, except that one contains an affinity tag. Therefore, when the first of these primer pairs is used for the subsequent round of PCR of one polynucleotide, one of the strands of the resulting PCR fragment will be specifically affinity labelled. In addition, when the second primer pair is used for the subsequent round of PCR of another homologous polynucleotide, one strand, the one that
25 corresponds to the complementary strand of the one labeled in the previous reaction, will be specifically affinity tagged.

Of course, it is also possible in the amplification step to use primers pairs, which altogether have both a first, a second, a third, a fourth, a fifth, a sixth, a seventh, an eighth, a ninth, a tenth or more oligonucleotide tags, but this requires that more
30 primer sets with affinity tags are used for the second PCR step, thereby making the method more costly.

When more than two genes are to be shuffled together, e.g. a whole family of genes, it is still only necessary to have two sets of affinity tagged primers for
35 performing the affinity tagging in the second PCR step. The only requirement is that

in at least one case, a coding strand must be affinity tagged and in at least one case a corresponding complementary strand must be affinity tagged.

According to one preferred embodiment of the invention, the primer pair used for
5 amplifying the genes to be shuffled are adaptamers. Examples of adaptamers are disclosed in US 5,942,422 (Rothstein) and US 6,291,213 (Rothstein) One advantage of this embodiment is that adaptamers that amplify all genes in the yeast genome and that have been tested for functionality are commercially available at a low price. It is expected that adaptamers will be available for other organisms in the
10 future and become available at low price.

Affinity end-labeled primer-pairs that match the sequence of the constant parts of the adaptamers are used. This means that all amplified and tagged fragments can be affinity labelled by using the same two primer pairs. This allows purification of
15 single stranded DNA at a low cost and high purity.

Shuffling of complementary single stranded DNA molecules from two or more homologous genes reduce the chance of obtaining wild-type (un-shuffled) gene products.
20

Adaptamers can be designed so that the primer-pair above is optimal for PCR, e.g. good melting temperature, good GC content according to the commonly accepted guidelines for good primer design, see e.g. <http://genome-www.stanford.edu/Saccharomyces/help/pcrinfo.primer.html>. In the case where an
25 ORF is hard to amplify due to an unfavourable sequence environment, this primer pair provides the possibility to perform a robust re-amplification of the ORF and robust shuffling of homologous ORFs.

In a further preferred embodiment the method provides an additional PCR step that
30 eliminates wild-type background by amplifying only such chimeric genes that have one end from one gene and the other end from another gene.

The constant oligonucleotide tags incorporated into the genes to be shuffled provide a possibility of recombination cloning in yeast (no need for an E. coli based cloning
35 step).

The constant oligonucleotide tags may also include rare restriction enzyme cut-sites, thus providing a possibility of high efficiency cloning in *E. coli*.

- 5 The constant oligonucleotide tags provide an easy way to sequence the products - a fast way to address shuffling efficiency.

According to a further embodiment, the invention relates to a kit for shuffling of genes, said kit comprising

- 10 two sets of primers, each set having one primer to template a first oligonucleotide tag and one to template a second oligonucleotide tag, one set having the primer to template the first oligonucleotide tag linked to an affinity-tag and the other set having the primer to template the second oligonucleotide tag linked to an affinity-tag,
15 at least one endonuclease,
 suitable buffers and reagents.

The kit is adapted for use in the shuffling method of the invention.

- 20 In a further aspect the invention relates to a chimeric hexose transporter, said chimeric hexose transporter being obtained by gene shuffling and having a higher affinity for xylose and/or a lower affinity for glucose than the affinity of the unshuffled hexose transporter for said carbohydrates.

- 25 Preferably said chimeric hexose transporter has a K_M for xylose lower than 90 mM and/or a K_M for glucose higher than 30 mM.

- 30 The improved chimeric hexose transporter can be used for the production of fuel ethanol production from the cheap substrate is hydrolysed lignocellulose, found for example, in agricultural wastes. Several sugars are present in this raw material, including pentoses such as arabinose and xylose.

Description of Drawings

Figure 1. Amplification of members of a gene family or of homologous genes from different species by PCR using tagged primers.

5

Figure 2. Reamplification of fragments using a primer pair of which one is biotinylated.

Figure 3 a b c d. Purification of single stranded DNA fragments

10

Figure 4 a b c. Shuffling of single stranded DNA fragments.

Figure 5. Schematic illustration of possible interactions and non-interactions between two proteins, Prot1 and Prot2, for which two homologous proteins (Hs and Sc) exist. The Sc- proteins interact with each other but not with the Hs- proteins and vice versa.

15

Figure 6. Schematic illustration of shuffling of single strands of Sc-Prot1 with Hs-Prot1 and of Sc-Prot2 with Hs-Prot2.

20

Figure 7. Analysis of Prot1 chimeras.

Figure 8. Mapping of essential amino acids in a protein. Purification of single stranded DNA coding for gene X and synthesis of oligonucleotides that cover the complementary strand of the entire gene X. The oligos are designed so that they will change every 5 (this number is only a suggestion) codon for one that encodes alanine.

25

Figure 9. Mapping of essential amino acids in a protein. Shuffling of the one strand with the complementary codon-altered strand.

30

Figure 10. Mapping of essential amino acids in a protein. Further steps in the mapping.

Figure 11. Mapping of essential amino acids in a protein. Functionality map of gene X.

5 Figure 12. Consequences of the limited transport of xylose by low affinity glucose transporters on the ethanol production by xylose-utilizing *S. cerevisiae*. PPP: Pentose Phosphate Pathway, EMP: Glycolysis. Legend: Ribo5P: ribose-5-phosphate, D-Xylu: xylulose, Sedo7P: sedoheptulose-7-phosphate, F6P: fructose-6-phosphate, Glyceral3P: glyceraldehyde-3-phosphate, Acetal: acetaldehyde, 6GL: 6-phospho-gluconolactone, XK: xylulokinase, XR: xylose reductase, XDH: xylitol dehydrogenase. NADP⁺, NADPH, NAD⁺, NADH, ATP, cofactors. Solid lines
10 represent the flow of glucose whereas dotted lines the flow of xylose.

Figure 13.1 Growth of TMB 3001 in glucose, xylose or maltose plates in the absence and presence of DOG. Cells were pre-grown in 10 ml of YPD for 20 hours,
15 at 30°C, with agitation and collected by centrifugation and washed with 40 ml of sterile distilled water, followed by re-centrifugation. Cells were then re-suspended in 1 ml of dH₂O, diluted 1:10,000 and 100 ml of the suspension were plated in: A) Glucose, B) Xylose, C) Xylose +0.01% DOG, D) Xylose + 0.05% DOG, E) Xylose + 0.1% DOG, F) Xylose + 0.25% DOG, G) Maltose, H) Maltose + 0.1% DOG and I)
20 Maltose + 0.25% DOG. Plates were incubated at 30°C for 2 days.

Figure 13.2. Growth of TMB 3201 in maltose plates in the absence of and presence of DOG. Cells were pre-grown in 10 ml of YPMaltose for 20 h, at 30 °C, with agitation and harvested and washed as described in Figure 2.1. Cells were then
25 resuspended in 1 ml dH₂O, diluted 1:10,000 and 100 ml of the suspension were plated in : A) Maltose, B) maltose + 0.01% DOG and C) Maltose + 0.05% DOG. Plates were incubated at 30 °C for 2 days.

Figure 14. Features of the plasmid p426 TEF. It contains the 2 μ origin of replication
30 in *S. cerevisiae*, the URA3 marker, the ampR marker, the pMB1 origin of replication in *E. coli*, as well as the TEF promoter and the CYC1 terminator. The unique restriction enzyme sites are shown. (Vector described in Mumberg *et al.*, 1995 *Gene* 156:119-122).

Figure 15. Genes amplified in a second-round PCR, where one of the primers was biotinylated. Samples (2 µl) were loaded as replicas, from a 50-µl PCR reaction.

Figure 16. Typical experiment for ssDNA isolation. Relative amount of each samples was loaded.

Figure 17. Single strand DNA (ssDNA) before and after SephadexG-50 purification

Figure 18. Schematic view of the procedure for isolation of ssDNA. The star represents a biotin molecule linked to the 5' end of the primer.

Figure 19. Annealing experiment between *HXT4* and *HXT7* ssDNA. The mixture was submitted to the following program in the PCR machine: 94 °C , 2 min for denaturation followed by 30 cycles at 94 °C, 30 sec, setting the dT[°C] to -2.0 and the ramping rate to 0.01 °/sec.

Figure 20. DNase I digestion pattern. The 1 kb ladder ranges from 10,000 to 500 bp and the 100 bp ladder from 1,500 to 100bp. Thus, the size of fragments generated by digestion might be around 300 bp.

Figure 21 illustrates methods of gene shuffling using heteroduplex DNA. Isolated partly complementary single stranded DNA molecules are hybridised to form a heteroduplex (Fig. 21a), which is digested with DnaseI (fig. 21b). The fragments are denatured and single strands are extended in a PCR reaction without primers (fig. 21c).

Definitions

Oligonucleotide tag: a nucleotide sequence incorporated into a primer, preferably into one end of a primer. The sequence must be capable of hybridising to the second primer under conditions that will allow a specific PCR reaction to be performed efficiently. The sequences should be chosen based on the commonly accepted guidelines for good primer design, see e.g.:

<http://genome-www.stanford.edu/Saccharomyces/help/pcrinfo.primer.html> and herein below.

5 Affinity tag: by an affinity tag is intended a molecule covalently linked to the primers in question. The affinity tag must be capable of binding to another member and this binding must be strong enough to resist conditions, which will cause denaturing of double-stranded DNA into single stranded.

10 Adaptamer: Adaptamers are chimeric oligonucleotides that are used to amplify an allele and differentially tag it's 5' and 3' ends (Hudson JR Jr, Dawson EP, Rushing KL, Jackson DH, Lockshon D, Conover D, Lanciault C, Harris JR, Simmons SJ, Rothstein R, Fields S, Genome Res 1997, Dec, 7 (12): 1169-73; Erdeniz N, Mortensen UH, Rothstein R, Genome Res 1997 Dec 7(12): 1174-83), US 5,942,422 (Rothstein) and US 6,291,213 (Rothstein)

15 Complementary: Two single stranded oligonucleotide sequences are said to be "complementary", when they are capable of hybridising to each other under low stringency conditions. Accordingly, not all nucleotides of the sequences must participate in specific base pairing.

20 Perfectly complementary: Two nucleotide sequences are "perfectly complementary" to each other when all nucleobases of one sequence can specifically base-pair with all nucleobases of the other sequence.

25 Sense strand: The sense strand, is the strand of a DNA molecule, wherein at least a part of said strand encodes a polypeptide or a fragment thereof. Hence the sense strand is also designated the "coding strand".

30 Antisense strand: The antisense strand, is the strand of a DNA molecule, which is complementary, preferably perfectly complementary to the sense strand. The antisense strand is also designated the "non-coding strand".

35 Forward primer: A forward primer is capable of hybridising to the antisense strand of a DNA molecule and may thus prime the synthesis of a DNA sense strand.

Reverse primer: A reverse primer is capable of hybridising to the sense strand of a DNA molecule and may thus prime the synthesis of an antisense DNA strand.

5 **Detailed description**

The process of gene shuffling using tagged primers is now described with reference to the drawings.

10 **DNA sequences to be shuffled**

15 The DNA sequences to be shuffled comprise different nucleotide sequences. It is however preferred that the DNA sequences to be shuffled are homologous to each other. At least the polynucleotides to be shuffled should have sequences with sufficient identity to allow gene shuffling. Hence, it is preferred that the sense strand of one DNA sequence is capable of hybridising to the antisense strand of one or more preferably of all the other DNA sequences to be shuffled under low stringency conditions.

20 Low stringency conditions of hybridisation are well described in the prior art. A preferred condition for low stringency according to the present invention is incubation for at least 30 minutes in the presence of 10 mM MgCl₂, 50 mM NaCl and 20 mM Tris-HCL, pH 7.5, wherein the temperature gradually in changed from 96°C to 30°C over that time.

25 It is thus preferred that the DNA sequences to be shuffled shares at least some degree of sequence identity. Preferably any two nucleotide sequences to be shuffled are at least 40%, such as at least 50%, for example at least 60%, such as at least 70%, for example at least 80%, such as at least 90%, for example at least 95% identical over a stretch of at least 60 bp, such as at least 90 bp, for example at least 120 bp, such as at least 150 bp, for example at least 300 bp, such as at least 600 bp, for example at least 900 bp, such as at least 1500 bp, for example at least 2400 bp.

30

The degree of sequence identity may be calculated using any suitable matrix, for example CLUSTAL in the PC/Gene program by Intelligenetics or GAP, BESTFIT, BLAST, FASTA and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG).

5

In one embodiment of the invention the DNA sequences to be shuffled are two or more homologous genes obtained from different species. In another embodiment the sequences to be shuffled are one or more DNA sequences encoding proteins belonging to the same family. Hence, a family of genes may be selected for shuffling, such as a family of at least 3 genes, for example at least 4 genes, such as at least 5 genes, for example at least 6 genes, such as at least 7 genes, for example at least 8 genes, such as at least 9 genes, for example at least 10 genes, for example at least 15 genes, such as at least 20 genes or more.

10

In preferred embodiment of the present invention, the DNA sequences to be shuffled encodes a polypeptide or a fragment thereof. Hence, the DNA sequences may encode proteins which are homologous. Said proteins may have a desired functionality such as capability to interact with another molecule, such as another protein, peptides, nucleic acids, carbohydrates, lipids, small organic molecules, hormones or the like.

15

20

In one specific embodiment of the invention the two double stranded DNA polynucleotides to be shuffled comprise carbohydrate transporters such as HXT-genes and xylE.

25

Amplification of individual DNA sequences to be shuffled.

Specific primer pairs are designed that allow amplification of all the genes that need to be shuffled. The primer pairs may preferably be designed as described herein below. Thus, in general one primer pair, a forward- and a reverse primer, is required for each sequence to be amplified. The forward primer should preferably be situated upstream of the sequences desirable to amplify whereas the reverse primer should preferably be situated downstream of the sequence desirable to amplify.

30

35

The primer may comprise any suitable number of nucleotides, preferably the primer comprises in the range of 10 to 200, such as in the range of 10 to 100, for example in the range of 15 to 50 nucleotides. If the primer comprises an oligonucleotide tag, then preferably the primer comprises in the range of 20 to 200, such as in the range of 30 to 100, for example in the range of 30 to 70 nucleotides.

It is preferred that the forward primer comprises a specific sequence, which is at least 80%, preferably at least 90%, more preferably at least 95%, most preferably 100% identical to a sequence within the sense strand of the sequence desirable to amplify, preferably upstream of the sequence desirable to amplify. Preferably, said specific sequence is at least 10 bp, more preferably at least 15 bp, such as at least 20 bp, for example in the range of 20 to 30 bp in length. Said specific sequence is preferably situated in the 3' end of the primer.

Likewise, it is preferred that the reverse primer comprises a specific sequence, which is at least 80%, preferably at least 90%, more preferably at least 95%, most preferably 100% identical to a sequence within the antisense strand of the sequence desirable to amplify, preferably downstream of the sequence desirable to amplify. Preferably, said specific sequence is at least 10 bp, more preferably at least 15 bp, such as at least 20 bp, for example in the range of 20 to 30 bp in length. Said specific sequence is preferably situated in the 3' end of the primer.

In one embodiment of the invention, the primers comprising oligonucleotide tags are adaptamers.

In one preferred embodiment of the invention at least one primer comprises a sequence tag or an oligonucleotide tag. The terms "sequence tag" and "oligonucleotide tag" are used interchangeably herein. The primer may comprise any of the oligonucleotide tags described herein below. More preferably at least all forward primers comprise an oligonucleotide tag or all reverse primers comprise an oligonucleotide tag, even more preferably all primers, such as all forward and reverse primers comprises an oligonucleotide tag.

The oligonucleotide tags may be identical or different. Hence, one or more primers may comprise identical oligonucleotide tags different to the oligonucleotide tags of other primers.

5 In a preferred embodiment all forward primers comprises first oligonucleotide tags, wherein said first oligonucleotide tags all are capable of hybridising to the same primer under PCR conditions. This allows amplification of all PCR products comprising forward primers using the same primer. Hence, all first oligonucleotide tags may comprise a common sequence and optionally sequences, which are not
10 common. For example one first oligonucleotide tag may consists of a fragment of another oligonucleotide tag. In another example a first oligonucleotide tag may differ from another oligonucleotide tag by substitution of one or more nucleotides, such as substitution of 2, for example 3, such as 4, for example 5, such as more than 5 nucleotides. However, more preferably all first nucleotide tags are identical.

15 Similarly, in a preferred embodiment all reverse primers comprises second oligonucleotide tags, wherein said second oligonucleotide tags all are capable of hybridising to the same primer under PCR conditions. This allows amplification of all PCR products comprising reverse primers using the same primer. Hence, all second
20 oligonucleotide tags may comprise a common sequence and optionally sequences, which are not common. For example one second oligonucleotide tag may consists of a fragment of another oligonucleotide tag. In another example a second oligonucleotide tag may differ from another oligonucleotide tag by substitution of one or more nucleotides, such as substitution of 2, for example 3, such as 4, for example
25 5, such as more than 5 nucleotides. However, more preferably all second nucleotide tags are identical.

It is preferred that first and second oligonucleotide tags are different, i.e. that a primer which hybridises to first oligonucleotide tags does not hybridise to second
30 nucleotide tags under PCR conditions.

The oligonucleotide tag may be situated in any desirable position within the primer. It is preferred that the oligonucleotide tag is situated in a position within the primer allowing template specific elongation of the primer, hence it is generally preferred
35 that the oligonucleotide tag is situated at least 10 nucleotides, more preferably at

least 15 nucleotides, such as at least 20 nucleotides, for example at least 25 nucleotides away from the 3' end. It is most preferred that the oligonucleotide tag is situated in one end of primer, preferably the 5' end of the primer.

5 In figure 1 a preferred embodiment of the invention is disclosed. In figure 1, the horizontal lines to the left, indicate sequences to be amplified. The arrows above and below the lines indicate forward- and reverse primers, respectively. The arrowheads symbolize the 3-OH group that will act as substrate for DNA polymerases. The forward primers are all extended in their 5'-end with a sequence tag. This sequence tag is common for all forward primers. In figure 1, the sequence tag is called A. Similarly, all reverse primers are extended in their 5'-end by a specific sequence tag. This tag is common for all reverse primers, but different from that added to the forward primers. In figure 1, this sequence tag is called B. Accordingly, the primer pair required to amplify sequence 1, termed A1 and B1, the pair required to amplify sequence 2 is called A2 and B2 and so forth. As a result each unique PCR amplified DNA sequence will contain a constant region, A, in the upstream end and a constant region, B, in the downstream end as depicted in the right side of figure 1. This fact allows each unique sequence to be re-amplified with one primer pair where one primer will anneal to sequence A and the other will anneal to sequence B, see figure 2. If one of these primers, i.e. either A or B contain an affinity tag in its 5'-end, one DNA strand of the resulting PCR product will contain this tag in its 5'-end. In figure 2, a reaction is shown that incorporates the affinity tag in the "top" strand (the reaction to the left) and a reaction that incorporates the affinity tag in the "bottom" strand (the reaction to the right). In these examples, the affinity tag is biotin.

The PCR (Polymerase chain reaction) may be performed using any suitable protocol known to the person skilled in the art. Preferably, the PCR is a cyclic reaction involving several rounds of denaturation, annealing and elongation.

30 In general, the denaturation step is performed by heating the reaction to a denaturing temperature, for example in the range of 80 to 100°C, usually in the range of 90 to 98°C.

Annealing is usually performed by lowering the temperature to a temperature at which the primers used in the specific PCR reaction may hybridise with the DNA template. The temperature used for annealing (or hybridisation) thus depends on the melting temperature of the primers. The approximate melting temperature of a primer may be calculated using suitable algorithms (see for example herein below).

Elongation is preferably performed using a DNA polymerase, which is resistant to the denaturing conditions, i.e. heat resistant. Elongation should be performed at a temperature, under which the polymerase is functional.

Oligonucleotide tags

The oligonucleotide tags according to the invention are nucleotide sequences as defined herein above.

The oligonucleotide tags preferably comprise a sequence which is optimal for PCR, e.g. good melting temperature of dsDNA, good GC content according to the commonly accepted guidelines for good primer design (see details herein below).

Preferably oligonucleotide tags consists of in range of 5 to 100, more preferably in the range of 10 to 50, for example in the range of 15 to 30, such as in the range of 17 to 25 nucleotides. Hence, the length of the oligonucleotide tags may be at least 15 nucleotides, more preferably at least 18 nucleotides, such as at least 21 nucleotides, for example at least 24, such as at least 27, for example at least 30, such as at least 35, for example at least 40, such as at least 50.

In one embodiment the oligonucleotide tag comprises at least one restriction site. Restriction sites according to the invention are sequences, which may be specifically cleaved by an endonuclease in a sequence specific manner. Any suitable restriction site known to the person skilled in the art may be used with the invention.

In a preferred embodiment of the invention the restriction site is a rare restriction site having at least 6 nucleotides in the recognition sequence, preferably at least 8 nucleotides in the recognition sequence.

The presence of a restriction site within the oligonucleotide tag facilitates handling of the shuffled sequences and enables for example subsequent cloning into a suitable vector comprising compatible restriction sites.

5 The different oligonucleotide tags may comprise different restriction sites or they may comprise identical restriction sites. In one embodiment of the invention all first oligonucleotide tags comprises identical restriction sites and all second oligonucleotide tags comprises identical restriction sites, wherein the restriction sites of first oligonucleotide tags are different from the restriction sites of second
10 oligonucleotide tags. This allows directional cloning of shuffled sequences into a suitable vector comprising compatible restriction sites.

In one embodiment of the present invention all oligonucleotide tags comprises the same restriction site. Hence, first and second oligonucleotide tags may share the
15 sequence of the restriction site, but otherwise be different.

Affinity tags and primers comprising affinity tags

20 Some of the primers used with the present invention are preferably affinity tagged. Within any specific PCR reaction performed according to the invention it is preferred that the forward primer is differentially affinity tagged to the reverse primer, i.e. the forward primer comprises a different affinity tag than the reverse primer. It is more preferred that only one of the primers comprises an affinity tag, i.e. that only the for-
25 ward primer or only the reverse primer comprises an affinity tag. Because the primers are differentially affinity tagged, the individual strands of the PCR product may be separated.

It is in particularly preferred that one of the primers of a primer pair for templating
30 oligonucleotide tags is affinity tagged. In a preferred embodiment of the invention, DNA sequences to shuffled are amplified once using specific primers comprising oligonucleotide tags and subsequently once using primers specific for the oligonucleotide tags.

35 Primers for templating oligonucleotide tags are preferably capable of hybridising to a sequence perfectly complementary to the oligonucleotide tag under low stringency

conditions; more preferably under high stringency conditions. Hence, said primer may for example be identical to the oligonucleotide tag, it may comprise the sequence of the oligonucleotide tag or it may comprise or consist of a fragment of the sequence of the oligonucleotide tag, or it may comprise a sequence at least 5 80%, preferably at least 90%, such as at least 95% identical to the oligonucleotide tag. Said primers individually preferably comprises in the range of 10 to 50 nucleotides, such as in the range of 5 to 100, more preferably in the range of 10 to 50, for example in the range of 15 to 30, such as in the range of 17 to 25 nucleotides. Hence, the length of the primer may be at least 15 nucleotides, more 10 preferably at least 18 nucleotides, such as at least 21 nucleotides, for example at least 24, such as at least 27, for example at least 30, such as at least 35, for example at least 40, such as at least 50. The primers should be suitable for PCR amplification (see herein below).

15 The methods of the invention relates to shuffling of at least two DNA polynucleotide sequences. It is preferred, that the PCR amplification of said polynucleotide sequences results in that only the sense strand of one polynucleotide sequence is affinity tagged, whereas only the antisense strand of the other polynucleotide strand is affinity tagged.

20

By way of example, when 2 DNA polynucleotide sequences are to be shuffled, then it is preferred that an affinity tagged forward primer and a non-tagged reverse primer is used for PCR amplification of the first DNA polynucleotide sequence, whereas an non-tagged forward primer and an affinity tagged reverse primer is used for PCR 25 amplification of the second DNA polynucleotide sequence. Hence the sense strand of the first DNA polynucleotide sequence may be isolated or removed from a mixture using the affinity tag. Likewise, the antisense strand of the second DNA polynucleotide sequence may be isolated or removed from a mixture using the affinity tag.

30 Hence, in one embodiment of the invention the primer pair for the further round of PCR for one polynucleotide, i.e. a primer pair for templating oligonucleotide tags comprises an affinity-tagged primer for templating the first oligonucleotide tag, and the primer pair for the further round of PCR for another polynucleotide, i.e. a primer pair for templating oligonucleotide tag comprises an affinity-tagged primer for 35 templating the second oligonucleotide tag.

If more than 2 DNA polynucleotide sequences are to be shuffled, then some sequences are preferably amplified using an affinity tagged forward primer and a non-tagged reverse primer and others are amplified using an non-tagged forward primer and an affinity tagged reverse primer. Preferably, the affinity tagged primers are only used during the final PCR amplification.

The affinity tag according to the present invention may be any molecule specifically associating with another molecule herein designated "binding partner", wherein the binding between the affinity tag and the binding partner are strong enough to resist at least one condition, which cause denaturing of double-stranded DNA into single stranded. Conditions for denaturing DNA are described herein below. The affinity tag and the binding partner are collectively referred to as a species-specific pair. Hence, the affinity tagged primer may comprise one member of a species-specific pair. The association between the two members of the species-specific pair, should thus preferably be sufficient strong to survive conditions which cause denaturation of double-stranded DNA.

In one example the affinity tag may be a protein. Suitable proteins include for example antibodies. If the affinity tag is an antibody, the binding partner will in general be a molecule comprising or consisting of the epitope for that particular antibody.

In a preferred embodiment of the present invention the affinity tag is a biotin. Biotin is capable of binding specifically and with high affinity to avidin and streptavidin and is thus useful for the present invention. Hence, if the affinity tag is biotin, the binding partner may be avidin and/or streptavidin.

The primers may be affinity tagged using any conventional method known to the person skilled in the art. In one embodiment of the invention affinity tagged nucleotides are incorporated into the primer during synthesis. Affinity tagged nucleotides are commercially available, for example biotinylated nucleotides are available from MWG, The Genome Company, Germany.

Isolation of affinity tagged DNA PCR products.

5 Once the DNA polynucleotides to be shuffled have been amplified by one or more rounds of PCR, wherein at least one primer used for the amplification process is affinity tagged, the affinity tagged DNA PCR products may be purified.

10 Any suitable process for purification known to the person skilled in the art may be used. It is however preferred that the methods comprise the step of contacting a composition comprising the affinity tagged DNA PCR products with a binding partner specifically interacting with the affinity tag.

15 In a preferred embodiment the binding partner is immobilised on a solid support or a solid surface prior to contacting the said composition. Then the affinity tagged DNA PCR products may be separated from the compositions simply by isolating the solid support. Thus isolation of the syntheses double stranded DNA polynucleotides may be performed by capturing said polynucleotides on a member of species specific pair being bound to a solid surface.

20 Hence, it is preferred that the method for purification of affinity tagged DNA PCR products comprises the steps of

1. Providing a composition obtained by amplification by one or more rounds of PCR of at least one DNA polynucleotide to be shuffled, wherein at least one primer used for the amplification process is affinity tagged,
- 25 2. Providing a binding partner capable of specifically interacting with the affinity tag immobilised on a solid support,
3. Incubating said composition with said binding partner under conditions allowing for binding between said affinity tag and said binding partner
4. Isolating said solid support, thereby purifying the affinity tagged DNA PCR fragments from said composition.
- 30

The solid support may be any solid support suitable for immobilisation of the binding partner. Preferably, the solid support is compatible with at least one condition, which causes denaturing of double-stranded DNA into single stranded. Said conditions are

described herein below. For example the solid support may be the surface of a well or test tube, or it may be a bead, such as a magnetic bead or an agarose bead.

5 In one preferred embodiment of the invention the solid support is magnetic beads, such as stainless steel beads. These beads may be isolated using a magnetic field.

The methods may also comprise one or more additional steps, for example a step of crude purification of PCR products prior to the purification steps described above. This may be done by conventional methods known in the art, for example by gel
10 electrophoresis or chromatography, such as affinity chromatography.

A non-limiting example of purification of affinity tagged PCR fragments is described below:

15 Biotin labeled PCR fragments were gel-purified before further usage. The purified biotin labeled fragment was incubated with streptavidin coated stainless steel beads. In this way the DNA fragments will be attached to the beads via the strong streptavidin - biotin interaction, see figure 3. An easy way to do these first steps is to dissolve the agarose plug that contains the DNA fragment by sodium iodine and
20 then directly add then streptavidin beads to this mixture. After 30 minutes incubation at room temperature, the beads may be pelleted by placing the reaction tube in a magnetic field. The supernatant is discarded and the beads may washed in appropriate buffer.

25

Isolation of single stranded DNA from PCR products.

The methods according to the present invention involve separating tagged from non-tagged DNA strands or separating differentially tagged strands. For example, non-
30 tagged and tagged strands of isolated affinity tagged double stranded DNA may be separated. The double stranded DNA may be isolated as described herein above.

It is preferred that the separation comprises contacting an affinity tagged double stranded DNA with a binding partner capable of associating specifically and with
35 high affinity to the affinity tag of said particular DNA molecule under a condition,

which causes denaturing of double-stranded DNA into single stranded. Preferably, only one strand of said double stranded DNA is affinity tagged.

In one preferred embodiment of the invention the method preferably comprises the steps of

1. Providing a composition comprising at least one double stranded DNA molecule, wherein one strand of said DNA molecule is affinity tagged with an affinity tag;
2. Providing a binding partner capable of binding to the affinity tag immobilised on a solid support;
3. Incubating said composition with said binding partner under conditions allowing for binding between said affinity tag and said binding partner, wherein said conditions causes denaturation of said double stranded DNA
4. Isolating the solid support, thereby separating affinity tagged DNA strands from non-tagged DNA strand.

The solid support may be any of the solid supports described herein above.

Conditions, which causes denaturation of a double stranded DNA may be selected from the group consisting of high temperature, alkine treatment and incubation in the presence of a denaturing agent.

High temperature is defined as a temperature above the melting temperature of the double stranded DNA. Hence, depending on the sequence and length of said double stranded DNA, high temperature is a temperature above 55°C, such as above 60°C, for example above 65°C, such as above 70°C, for example above 75°C, such as above 80°C, for example above 85°C, such as above 90°C.

Alkaline treatment is defined as incubation under high pH, i.e. pH above 9, such as above 10, for example above 11. For example alkaline treatment may be incubation in the presence of a strong base, for example NaOH or KOH.

Denaturing agents may for example be selected from the group consisting of formamide, guanidinium and UREA.

Once the tagged strands are separated from the untagged strands, the single stranded DNA may be further processed by any desirable method. If the tagged strands are to be used, then the further processing may for example involve purification of the tagged strands from the solid support.

5

If the untagged strands are to be used, the further processing may involve further purification of the single stranded DNA by any conventional method, for example by gel electrophoresis and/or chromatography, such as affinity chromatography, size exclusion chromatography or gel-filtration. It is also contained within the present invention that the methods of separating tagged from untagged strands are repeated at least once, such as twice, for example 3 times, such as more than 3 times, in order to remove any residual tagged strands.

10

A non-limiting example of isolation of single stranded DNA is outlined below:

15

Double stranded DNA trapped on the streptavidin beads prepared as described in the section above, is treated by sodium hydroxide. This denatures the DNA without significantly affecting the streptavidin interaction, see figure 3C. Accordingly, the DNA single strand that does not contain the biotin label will be released from the bead. After pelleting the magnetic beads, the single stranded DNA will stay in the supernatant, See figure 3D. In a single step, this DNA preparation is purified and adjusted to neutral pH by using a gel-filtration spin-column, see figure 17. The whole process is summarized in figure 18. If any biotin labelled DNA material is present at this point this can be easily removed by adding additional streptavidin beads. Pelleting the beads will remove any biotin labelled fragments from the supernatant. This purification step can be repeated if necessary.

20

25

Heteroduplexes

30

In one preferred embodiment of the invention 2 isolated single stranded DNAs are hybridised to each other. In particular, it is preferred that the sense strand of some of the DNA polynucleotide sequences to be shuffled are hybridised with the antisense strand of other DNA polynucleotides sequences to be shuffled.

By way of example, when 2 DNA polynucleotide sequences are to be shuffled, then it is preferred that the sense strand of the first DNA polynucleotide sequence is hybridised to the antisense strand of the second DNA polynucleotide.

5 Because several endonucleases have a preference for cleaving double stranded DNA over single stranded DNA, it is possible to preferentially cleave a heteroduplex in regions wherein the strands of the heteroduplex are perfectly complementary. Because the endonuclease will cleave at different positions within the perfectly complementary region different cleaved polynucleotides will comprise perfectly
10 matching overlaps, which facilitates the shuffling process (see herein below). The principle of forming, digesting and shuffling heteroduplexes is illustrated in figure 21.

If random cutting is performed, many fragments will have 3'-OH ends that are not perfectly complementary to the 3'-end of any other fragment. Such fragments are
15 poor substrates, if they are useful at all, for priming a DNA polymerase elongation and the efficiency of the essential DNA polymerase step during shuffling that ensures DNA duplex formation will be reduced. Random cutting may be performed using an endonuclease that will cleave both single stranded and double stranded DNA approximately equally well or by digesting single stranded DNA. Preferred
20 methods of fragmenting heteroduplexes are described herein below.

The heteroduplex may be formed using any suitable protocol. Preferably the heteroduplex is formed by a method comprising the steps of:

- 25 1. Providing at least one single stranded sense strand of a DNA sequence to be shuffled
2. Providing at least one single stranded antisense strand of another DNA sequence to be shuffled
3. Incubating said single stranded sense and antisense strand under conditions of
30 low stringency hybridisation
4. Thereby forming a heteroduplex between said sense and said antisense strand

More than 2 different single stranded DNA sequences may be hybridised, such as 3, for example 4, such as in the range of 5 to 10, for example in the range of 10 to 20
35 different single stranded DNA sequences.

Conditions for low stringency hybridisation is well described in the art. The specific hybridisation conditions may be selected according to the nature of the single stranded DNA molecules to be hybridised. A preferred condition for low stringency hybridisation is described herein above.

5

Non-limiting examples of formation of a heteroduplex is given below:

A heteroduplex is formed by mixing and treating two single-stranded DNA molecules (for example isolated according to the methods described herein above) as follows:

10 0.5 µg of each of the two complementary DNA strands are combined in a total volume of 50 µl containing 10 mM MgCl₂, 50 mM NaCl and 20 mM Tris-HCL and pH 7.5 (final concentrations).

15 If it is desirable to mix strands from more than two homologous genes at the same time, it may be performed as follows. By way of example, in the case that four homologous genes (A, B, C and D) are to be shuffled in one experiment, then, of the four ssDNA strands that are going to be mixed together in the experiment, at least one needs to be the coding strand and at least one from another gene needs to be the non-coding strand. Example 1: coding strands from gene A, B and C are mixed with the non-coding strand from gene D. Example 2: coding strands from gene A and B are mixed with non-coding strands from genes C and D. In this type of experiments the total DNA content should be between 0.5 - 2 µg in a 50 µl reaction.

25 In all cases, the reaction mixture is incubated at 96°C for 5 min before being cooled down slowly and gradually to 30°C over a period of 30 minutes. At this point the heteroduplex formation is complete.

Gene shuffling of PCR fragments.

30

The shuffling is preferably performed by a method involving fragmentation of DNA. The DNA may be fragmented to obtain just two fragments or more than 2 fragments, such as in the range of 3 to 5, for example in the range of 5 to 10, such as in the range of 10 to 30, for example in the range of 30 to 100, such as more than 100 fragments of each DNA to be shuffled. The DNA is preferably either single stranded

35

DNA prepared as described in the section "Isolation of single stranded DNA from PCR products" herein above or heteroduplex DNA prepared as described in the section "Heteroduplexes" herein above.

5 Single stranded DNA may be fragmented using any suitable method known to the person skilled in the art. For example DNA may be fragmented using a chemical or an enzyme. For example the DNA may be fragmented using e.g. classical Maxam-Gilbert chemistry or using ultrasound.

10 Preferably, an endonuclease capable of cleaving single stranded DNA is used, for example DNase I or DNase II.

Heteroduplexes may also be fragmented by any suitable method, however it is preferred within the present invention that the heteroduplexes are fragmented using an
15 endonuclease which preferentially digest double stranded DNA rather than single stranded DNA. Even if the endonuclease used, is not entirely specific for dsDNA, but displays a pronounced preference for duplex DNA over ssDNA, then the likelihood of formation of fragments with the capacity to form chimeras with perfectly complementary 3'-OH ends is expected to be high. For example DNase I preferen-
20 tially cleaves double stranded DNA (Dietrick Suck, DNA recognition by structure-selective nucleases, *Biopolymers*, 1997, 44:405-21)

In a preferred embodiment of the invention heteroduplexes are fragmented using an
25 endonuclease selected from the group consisting of DNase I and DNase II.

Subsequent to fragmentation of heteroduplexes it is preferred that the double
stranded fragments of the heteroduplexes are denatured to obtain single stranded fragments. Denaturation may be performed by incubation under any condition,
30 which cause denaturation of double stranded DNA, for example any of the conditions mentioned herein above. Preferably, denaturation is performed by heating.

The shuffling methods according to the present invention furthermore comprises the step of mixing single stranded DNA fragments and hybridising them. The single
35 stranded DNA fragments may have been obtained either by cleaving single stranded DNA or by denaturation of heteroduplex DNA fragments. It is preferred that single

stranded DNA fragments derived from at least one sense strand of a DNA sequence to be shuffled and single stranded DNA fragments derived from at least one antisense strand of another DNA sequence to be shuffled are mixed. Preferably said fragments cover the entire sense/antisense strand.

5

Preferably the DNA fragments are mixed and incubated under hybridisation conditions. Hybridisation may be done by any suitable method. It is preferred that hybridisation results in formation of at least some hetero-DNA molecules with single stranded overhangs, more preferably with single stranded 5' overhangs. Preferably, the newly formed hetero-DNA molecules are hybridised to each other in a region wherein the molecules are perfectly complementary or at least 90% perfectly complementary, such as at least 95% perfectly complementary.

10

15

Hybridisation may be done using any suitable conditions allowing hybridisation. In one embodiment of the invention hybridisation is performed in a PCR reaction using a suitable annealing temperature (see herein above).

20

Hetero-DNA molecules with single stranded 5' overhangs may be subjected to elongation primed by the 3'-OH using the 5' overhangs as template using a DNA polymerase. This principle is for example illustrated in fig. 4 and fig. 21.

25

Accordingly, shuffled genes may be obtained by using the cleaved, single stranded fragments as both primers and templates in a nucleotide polymerisation reaction, for example PCR.

30

Hence, it is preferred that the shuffling method also comprises a step of incubating the hetero-DNA molecules in the presence of a DNA polymerase. In one embodiment of the invention said DNA polymerase is a thermostable polymerase, which allows the denaturation step and the elongation step to be performed in the same reaction tube.

35

The elongation step may be performed more than once, for example using a PCR reaction without addition of primers. Instead, the single stranded fragments may be used both as primers and templates in the PCR reaction.

Hence, the shuffling may be performed by a method comprising the steps of

- i) removing the endonuclease or chemical reaction mix,
- ii) reassembling double-stranded DNA using PCR without primers.

Said method may be performed using either single stranded or heteroduplex
5 fragments as starting material.

The methods of shuffling according to the present invention may also comprise additional steps. For example, the methods may involve the use of an exonuclease during shuffling. The methods may also further involve a step of introduction of
10 mutations during shuffling. Mutations may be introduced using any conventional method, such as random mutagenesis or site directed mutagenesis.

Subsequent to elongation, optionally several rounds of elongation, shuffled full length sequences may be obtained. A selection step for full length sequences may
15 also be performed. For example that may be done by subjecting the shuffled sequence to a PCR reaction using primers. Preferably, the primers recognise the ends of the full length molecule or alternatively, and more preferably the primers may prime the oligonucleotide tags incorporated into the DNA sequences to be shuffled during one of the amplification steps. Even more preferably, the primers
20 used for the PCR are selected to ensure that the amplified sequence comprises one end from one selected polynucleotide and the other end from another selected polynucleotide.

Full length sequences may also be isolated using other methods or in combination
25 with other methods such as gel electrophoresis

After shuffling, the newly formed shuffled sequences may be screened for a desired property or a pre-selected characteristic and sequences meeting at least one pre-determined selection criterion may be selected. For example, if the DNA sequences
30 to be shuffled encodes protein with a known functionality, then shuffled sequences encoding a protein with an improved functionality may be selected.

The methods for gene shuffling described herein may also be combined with any other shuffling method known in the art. It is also comprised within the present in-
35 vention that the shuffling methods according to the invention are repeated, for ex-

ample once, such as twice, for example in the range of 3 to 5 times, such as in the range of 5 to 10 times, for example more than 10 times.

Hence, the method may for example be repeated until at least one shuffled gene
5 having a pre-determined functionality is obtained.

An illustrative but non-limiting example of shuffling according to the invention is described below:

10 Single stranded material from homologous sequence may now be gene shuffled. In the simplest case, complementary single stranded DNA from two genes are fragmented, e.g. by DNaseI, the fragmenting agent removed and the DNA fragments from the two genes mixed. Alternatively, heteroduplexes are fragmented using DNase I.

15 By adding a heat-stable DNA polymerase, nucleotides and the appropriate buffer, the fragments may now assemble into a DNA double stranded molecule of the same length as the original genes by subjecting the mix to a PCR program, see figure 4B. If a heteroduplex is used, then the fragments of said heteroduplex will be denatured
20 during the first step of the PCR (see also fig. 21).

The resulting shuffled gene may be further amplified by adding primers A1 and B2, respectively. Since primers A1 and B2 only bind productively to gene 1 and gene 2, respectively, they will only amplify a chimera of the two genes, see figure 4C. Some
25 chimeras will be lost in this process. However, these can be regenerated by repeating the procedure and then selecting with primer pair (A1 + B1) and primer pair (A2 + B2) in the final step.

30 Primers

In the following conditions for the design of PCR primer sequences are described.

35

Location of Primers

When the DNA source is a Locus, the location of the primers will be determined relative to the start and stop codons of the gene. The default option will find "forward" primers of given length(s) that wholly reside somewhere within the first 35 basepairs upstream of the coding sequence, and likewise will find "reverse" primers that reside within the 35 basepairs immediately following the coding sequence. The user may alter the endpoints of either of these "primer selection regions" by changing the number in the "Distance from Start" and "Distance from Stop" fields (note that entering negative numbers will cause the primer selection region to be located inside the coding region of the gene). The user may also define exact 5' endpoints of the primers by selecting the button marked "YES" on the line which asks about exact endpoints. Thus while the default option will allow amplification of a region whose endpoints are somewhere within 35 basepairs upstream and 35 bp downstream of the gene, choosing an exact endpoint will cause each primer to share the same 5' end.

Location of primers (with respect to DNA entered)

Primer location is most influenced by selection of a gene. Possible primers are determined by their relationship to this gene. The user may choose where in relation to the start and stop codons the primers are located. The default option will find primers in the first and last 35 basepairs of the DNA sequence entered. The user may define exact endpoints to start and stop the primer, thus the default option will allow amplification of a region whose endpoints are in the first and last 35 basepairs, while choosing exact endpoints will cause all primers evaluated to start and end with exactly the same sequence.

Primer composition

Primers that contain a skewed AT/GC ratio can fail to give high specificity, or yield primers that are in other ways not well behaved. The user is allowed to enter minimum, optimal, and maximum values for the percentage of basepairs which are either G or C.

Primer melting temperature

Melting temperature heavily influence the results of PCR. This utility calculates the T_m of an oligonucleotide by the nearest neighbour method (Borer, et al. J. Mol., Biol. 86; 843-53, 1974), (Rychlik, et al. NAR, 18; 6409-12, 1990), and Breslauer, et al. PNAS, 83; 3746-50, 1986. Minimum, optimum, and maximum values are able to be user customised.

Primer Annealing

Primers also tend to dimerize and anneal to themselves, this can present significant problems in using PCR. One method for accounting for this problem was developed by Hillier and Green, PCR Method. Applic., 1; 124-8, 1991. Maximum values annealing between primers are able to be set by the user.

Ranking the pairs of primers

Primers are assigned a value based on their features and based on the user defined preferences. The best pair of primers is defined to be the pair of primers with the lowest score. This score is calculated in the following way: +1 per 10% difference from the optimal GC percentage, +1 per degree Celsius difference from the optimal T_m , +1 per 5 units of annealing at the end of primers, +1 per 10 units of annealing in the middle of primers, and +1 per 2 basepairs difference from the optimal length.

Kit for shuffling genes

In one embodiment the present invention relates to a kit useful for shuffling genes according to the methods described herein above.

The kit preferably comprises

two sets of primers, each set having one primer to template a first oligonucleotide tag and one to template a second oligonucleotide tag, one set having the primer to template the first oligonucleotide tag linked to an affinity-tag

and the other set having the primer to template the second oligonucleotide tag linked to an affinity-tag,
at least one endonuclease,
suitable buffers and reagents.

5

The primers to template oligonucleotide tags may be any of the primers described herein above. The affinity tags may also be any of the tags described above.

10

The endonuclease may be any suitable endonuclease as described herein above. In a preferred embodiment the endonuclease is DNase I (EC 3.1.21.1).

15

The kit may further comprise two or more sets of primers, each primer having a sequence for templating a polynucleotide to be shuffled and a first and/or a second oligonucleotide tag, each pair having the first tag in the forward primer and the second tag in the reverse primer. The primers may for example be any of the primers described herein above. In one embodiment, the two or more sets of primers comprise adaptamers.

20

The kit may also comprise means to capture the affinity tag, these means comprising a member of a species specific pair being linked to a solid surface as described above.

25

Hexose transporter

30

The invention also relates to a chimeric hexose transporter obtained through gene shuffling, said chimeric hexose transporter having a higher affinity for xylose and/or a lower affinity for glucose than the affinity of the unshuffled hexose transporter for said carbohydrates.

35

Preferably, the chimeric hexose transporter has a K_M for xylose lower than 90 mM and/or a K_M for glucose higher than 30 mM.

In a preferred embodiment the chimeric hexose transporter comprises sequences from at least one HXT-gene and at least one xylE-gene. The chimeric hexose

transporter may have been obtained by the shuffling method according to the present invention, for example by shuffling at least one HXT-gene and at least one xylE-gene.

5

Examples

Example 1. Mapping of protein-protein interaction domains.

10 Two proteins, Prot1 and Prot2, interact physically (in vivo)

Task: To map regions/domains within Prot1 And Prot2 that mediate the interaction between the two proteins.

15 Requirements:

- 1) Known, or partially known, DNA sequences for both proteins.
- 2) Existence of a pair of homologous proteins in another organism
- 3) The interaction between the two proteins is species specific as illustrated schematically in figure 5. In this figure, a double headed arrow symbolizes
20 a protein-protein interaction, Hs means Homo sapiens and Sc means Saccharomyces cerevisiae.
- 4) An assay to detect the protein-protein interaction of interest must exist , e.g. a two-hybrid assay

25 Experimental procedure. First, PCR products containing the entire open reading frames of the genes that encodes the two proteins, Prot1 and Prot2, of interest are produced. Each primer will contain an affinity tag, A or B, as described above. For each protein, individual PCR products are obtained by using at least two different homologues, albeit an orthologue or a paralogue, as templates. The homologues
30 are chosen so that the requirements listed above are fulfilled. In the simplest case, where only two homologous genes of each protein are employed in the mapping experiment, single stranded DNA corresponding to the coding strand of one of the homologues and single stranded DNA corresponding to the non-coding strand of the other homologue are isolated. This can be done by reamplifying the individual PCR
35 products by primers A and B of which the appropriate primer is biotinylated.

Subsequently, the desired single stranded DNA can be purified by the streptavidin based method described above. Accordingly, for each protein, chimeras of its corresponding homologues can be generated by shuffling (and reshuffling) the two homologous genes by our method. For each protein, a library of different chimeras is produced. See figure 6. Each chimera in the libraries can now be tested for its ability to interact with its original un-shuffled protein partners. For example chimeras of Prot1 can be tested for their abilities to interact with Prot2 encoded by one of the original homologues as well as for their abilities to interact with Prot2 encoded by the other original homologue. Sequencing of a number of the Prot1 chimeras followed by a comparison of their individual abilities of forming a protein-protein interaction to the two original Prot2 homologues reveal the region in Prot1, which is important for forming an interaction with Prot2, see figure 7.

Example 2. Mapping of essential amino acids in a protein.

Task: To map regions/domains within a protein, which is important for its function.

Requirements:

- 1) Known sequence of the gene encoding the protein.
- 2) An assay to detect the functionality of the protein, e.g. a phenotype or an enzymatic reaction.

In the first step, the gene of interest is amplified by PCR. The primers will contain affinity tags, A and B, respectively, as described above. If primer A is biotinylated it will allow one strand of the gene of interest to be purified. In this case, the purified single stranded DNA will contain: affinity tag B followed by the gene of interest, then tag a, where a is a sequence complementary to the sequence tag A, see figure 8.

Next, a set of primers are constructed that can basepair to the single stranded DNA that was isolated in the previous step. The primer set may cover the entire gene or part of the gene. The primers are designed in a way so that they will change given codons in the original gene of interest to alanine codons (or another mutation) in a system of choice. For example, every fifth codon could be an alanine codon, see figure 8

The single stranded DNA obtained in step one is now degraded by, e.g. DNase I, and mixed with the synthetic oligonucleotides described in step two. Shuffling of the sequences are now performed in the same way as described previously. This will yield a library where different sections of the gene of interest will contain mutations.

5 The library may be reamplified by a PCR reaction that includes primers A and B for further analysis. Alternatively the reamplification may be performed in two different reactions: one that employs the primer pair (A and biotinylated-B) and one that employs the primer pair (biotinylated-A and B), see figure 9. In this case, complementary single stranded DNA molecules can be generated that can be used
10 for additional rounds of shuffling, see figure 10. Shuffled DNA molecules can be cloned into an expression system and the functionality of the gene evaluated in the appropriate system. Sequencing of mutated genes that encode functional- as well as non-functional proteins will provide the desired information, i.e., which regions of the protein that are important for its function(s), see figure 10 and 11.

15 **Example 3. Development of a xylose-metabolizing *S. cerevisiae* strains with enhanced xylose transport by *in vitro* homologous recombination (DNA shuffling)**

20 **Introduction**

The yeast *S. cerevisiae* is the most efficient microorganism for the industrial production of ethanol. This trait is given by its well-developed sugar transport system and glycolytic pathway. For the production of fuel ethanol production, a potentially cheap substrate is hydrolysed lignocellulose, found for example, in
25 agricultural wastes. Several sugars are present in this raw material, including pentoses such as arabinose and xylose.

S. cerevisiae does not ferment xylose, so genes from the xylose-utilizing yeasts have been inserted into it, to enable the fermentation of this pentose in addition to
30 hexoses.

A. Sugar transport in *Saccharomyces cerevisiae*

S. cerevisiae possesses a family of highly homologous sugar transporter genes. Glucose uptake is carried out by hexose transporters encoded by the *HXT* genes. In
35 *S. cerevisiae* there are 20 genes encoding proteins similar to hexose transporters -

HXT1 to *HXT17*, *GAL2*, *SNF3* and *RGT2* genes (Özcan and Johnston, 1999). None of these transporters are essential for growth on glucose, indicating their functional redundancy. Seven members of the *HXT* family are known to encode functional glucose transporters (*HXT1* through *HXT7*). A *hxt1Δ-hxt7Δ* mutant is unable to grow on glucose, fructose or mannose and has no glycolytic flux and the introduction of any of the 7 *HXT* genes into the null mutant is sufficient to recover its grow on glucose. To block uptake of hexoses however, the knock out of at least 20 transporter genes, including all the *HXT* genes (*HXT1* to *HXT17*), plus the galactose transporter gene *GAL2* and three members of maltose transporter (*AGT1*, *YDL247* AND *YJR160*) is required, as occurred with strain EBY.VW4000 (Wieczorke et al. 1999) (see Table 1).

Xylose can also be taken up by glucose transporters, but its affinity is about two orders of magnitude lower than that for glucose and the efficiency of this uptake decreases even more in the presence of glucose. This minimal uptake of xylose when glucose is present in the growth medium is schematically shown in Fig. 12. Since *Saccharomyces* spp. is not able to ferment xylose, the xylose reductase (*XR*) and xylitol dehydrogenase (*XDH*) genes from *P. stipitis* were introduced into the chromosome of the *S. cerevisiae* strain EBY.VW4000, generating strain TMB 3201 (collection of Bärbel Hahn-Hägerdahl from Lund Universtiy, Sweden; personal communication) (see Table 1), to study xylose utilization in a system that dos not utilize hexose transporters.

On the other hand, having in mind the production of fuel ethanol, the conversion ion of xylose to this alcohol has limitations being the transport of xylose a critical one (Fig. 12). To improve the xylose uptake by ethanol-producing *S. cerevisiae* strains, directed evolution of the glucose transporter genes based on DNA shuffling technique was approached.

30

35

Table 1. Yeast strains		
Strain	Genotype	Source
EBY.VW4000	<i>MATα leu2-3,112 ura3-52 trp1-289 his3-Δ1 MAL2-8^c SUC2 hxt17Δ hxt13Δ::loxP hxt15Δ::loxP hxt16Δ::loxP hxt14Δ::loxP hxt12Δ::loxP hxt9Δ::loxP hxt11Δ::loxP hxt10Δ::loxP hxt8Δ::loxP hxt514Δ::loxP hxt2Δ::loxP hxt67Δ::loxP gal2Δ stl1Δ::loxP agt1Δ::loxP ydl247Δ::loxP yjr160cΔ::loxP</i>	Wieczorke et al. 1999
TMB3201	EBY.VW4000 containing the plasmid YlpXR/XDH/XK integrated into HIS3 locus	Bärbel Hahn-Hägerdahl collection, Lund, Sweden
TMB3001	<i>MATα MAL2-8^c SUC2</i> containing the plasmid YipXR/XDH/XK integrated into HIS3 locus	Eliasson et al., 2000

B. Directed evolution

- 5 Improvements in biological functions have occurred by evolutionary changes, including homologous recombination (Stemmer, 1994). In eukaryotic genes the coding sequence is interrupted by introns. The introns are spliced out after transcription, thus only the coding sequence (the exons) are left behind. This process, called exon shuffling, can result in rapid protein evolution, and occurs when
- 10 recombination within introns assembles independent exons into genes for novel proteins, as in the blood-clotting proteins. So the gene shuffling technique mimics the natural evolutionary process, to accelerate such a process *in vitro*. This has been named molecular breeding or direct evolution, and is based in gene shuffling techniques (homologous recombination *in vitro*; Jermutus et al. 2000; Gibbs et al
- 15 2001). It is the systematisation and acceleration of the evolutive process by breaking apart and then mixing genes from (the same or) different organisms and testing the hybrid genes to see if they work better than the original ones (Harayama 2001, Arnold's Webpage).
- 20 The method consists in fragmenting the genes and subsequently recombining the fragments *in vitro* by using Polymerase Chain Reaction (PCR). The process utilises naturally occurring nucleotide substitutions among family genes as the driving force for the *in vitro* evolution (Judo et al. 1998). There are two options for cutting DNA: first, restriction enzyme in which only double stranded DNA (dsDNA) can be used as
- 25 substrate and second, DNase I treatment, where either single or double stranded DNA, ssDNA and dsDNA, respectively, can be used.

A potential problem of the family shuffling is the low yield of recombinants, especially if the identity of nucleotide sequences between parental genes is relatively low. By consequence, the efficiency of hybrid formation of shuffled products will be very low due to a low heteroduplex formation. To overcome this problem, improved techniques in terms of chimeric gene formation have been developed (Kikuchi et al 1999a).

In the case of dsDNA, fragments are generated by restriction enzymes and since fragments from the same gene have no overlaps with each other, no homoduplex formation occurs. Nonetheless, the variety of shuffled products might be limited to a certain degree, since the variation of gene fragments generated by restriction enzyme is limited, compared to gene fragments generated by DNaseI digestion. Moreover, the use of related genes as complementary ssDNA is interesting, since a reduction in the homoduplex formation is less likely to occur during the shuffling process, because complementary fragments derived from the same gene are missing (Kikuchi et al 1999b).

This new technique (gene shuffling) has a large range of potential application, i.e., production of pharmaceuticals and vaccines (Chartrain et al. 2000), therapeutic proteins (Kurtzman et al. 2001), plant biotechnology (Lassner and Bedbrook, 2001), enzymes (Arnold's Webpage, Arnold and Volkov, 1999; Hyun Joo et al. 1999; Shibuya et al 2000; Bornscheur and Pohl, 2001; Morawski et al, 2001, Hayashi et al. 2001). Arnold (webpage) gives a long list of enzymes with increased activity and higher stability resulting from the application of this technique.

The aim was to construct a new family of sugar transporters able to uptake xylose more specific and efficiently. Directed evolution (or molecular breeding or gene shuffling) of *S. cerevisiae* and *E. coli* sugar transporters was approached. The host strain for the screening of the enhanced-xylose-transport trait is *S. cerevisiae* TMB 3201 (kindly provided by Barbel Hahn-Hägerdahl), in which the entire hexose transport system has been deleted.

Work strategy

The strategy encompassed the following steps:

1. Test for analyzing new phenotype
2. Preparation of clean DNA of sugar transporters genes suitable for DNA shuffling
3. DNase I cleavage of several different genes
- 5 4. Mix of short-length ssDNA fragments for reassembling a new full-length dsDNA and PCR of the fragments for amplification
5. Restriction enzyme digestion of the amplified products
6. Ligation of the newly constructed fragments into a transformation vector
7. Transformation of the engineered strain and selection in xylose plate
- 10 8. Assay for enhanced xylose uptake

1. Test for analyzing the new phenotype (improved xylose transport)

Any Molecular Biology work requires a good system for assessing the effect of changes introduced in the system, i.e., to test the phenotype given by the modified
15 genotype. Sugar analogs are useful compounds to test sugar transport capability. Non metabolizing sugar analogs, such as 6-deoxy-D-glucose and 2-deoxy-D-glucose (the latter herein referred to as DOG), have been used to address an accurate measurement of sugar uptake. At high-concentrations, DOG is considered
20 an energy poison that rapidly depletes the cells of ATP, resulting in cell death, probably by accumulation of significant amount of 2,2'-dideoxy- α , α' -trehalose in the non-ionic sugar pool.

We successfully tested the effect of this compound on the growth of two different *S. cerevisiae* strains a procedure to be used as a tool for screening for new strains
25 displaying improved xylose transporter system. A dose-response test was established to identify the concentration range (0.01 to 0.25% was used) affecting the growth of *S. cerevisiae* TMB 3001 in either xylose or maltose. This reference strain contains the xylose-metabolizing genes integrated into the chromosome (Eliasson et al. 2000) (see Table 1) and is isogenic to TMB 3201. For the screening
30 the ideal concentration is the one that yields colonies not as large as in the positive control plate (without the compound) due to impaired metabolism, rather than a concentration where the cells die.

With the improvement in xylose uptake one expects the cells to grow better on
35 plates containing that amount of DOG when compared to the "negative" control

(reference strain growing in the same condition). Thus, Fig. 13.1 shows that at concentrations of 0.1 and 0.25% the reference strain did not survive, regardless of the type of sugar used as carbon source, so, such concentrations were considered not appropriate. However, at 0.05%, the difference in colonies size is readily noticed and we considered that sufficient for the analysis (plating on maltose 0.05% DOG presented similar result but is not shown).

As the strain TMB 3201 cannot grow on glucose or xylose because of the deletion of its hexose transport system, we tested the effect of DOG only in media containing maltose as carbon source to prove that the compound is related to the glucose system and not maltose. The compound (Fig. 13.2) did not affect the growth, corroborating the lack of glucose transport in this modified strain and further indicating that DOG may be a good tool for the proposed screening.

2. Preparation of pure ssDNA

Although a few shuffling procedures employ double stranded DNA, we concentrated efforts into the isolation of single stranded DNA to minimize the chances that after applying the shuffling method, the original gene is formed again. Although the initial difficulty was the isolation of pure DNA, this was solved by using a commercial kit especially designed for removal of small molecules that worked properly to get rid of traces of NaOH. The addition of beads to the biotinylated PCR product prior to complete isolation (see item d) shortened the time frame of the entire procedure even more.

The steps followed to prepare clean ssDNA were:

a) The genes of interest were first amplified by PCR by using genomic DNA from *S. cerevisiae* strain W2148-7A, a derivative of the common laboratory yeast strain W303, (for the *HXT* genes). or from *E. coli* DH5 α (*xyIE* gene) as templates and by using primer pairs (BWG Biotech) that match the sequences of the individual genes to be amplified. Additionally, every primer's pair has non-homologous common sequences at the 5' ends providing adequate restriction enzyme sites that are useful to clone the genes in a directly oriented fashion into the plasmid shown in Fig 14. The sequence of the primers is shown in Table 2 and table 3 shows general volumes and conditions used for each reaction.

b) Amplified genes were isolated from agarose gel using the GFX DNA purification Kit (Pharmacia) and used as template for a second round of PCR.

- c) In the second reaction, one of the primers (AB1 or AB2) is biotinylated at its 5' end and is complementary to the common sequence of each other primers, therefore being suitable for using with any of the amplified genes (see Fig. 15 for examples of some genes amplified using this procedure)
- 5 d) The gel slice containing the biotinylated PCR product was dissolved as mentioned in item b, cooled down to room temperature and an appropriate volume of streptavidin-coated beads was added (Beads are previously equilibrated in TE).
- e) Incubated for 30' at room temperature
- 10 f) Placed the tube on the magnet for the separation process
- g) Three washing steps in TE followed
- h) Non-biotinylated ssDNA was released by a freshly prepared solution of 0.1 M NaOH /1mM EDTA followed by incubation for 15 minutes at room temperature.
- 15 i) After separation using the magnet, the ssDNA was rescued with a pipette, with special care to avoid disturbing the beads (see Fig. 16)
- j) Residual NaOH was removed from the DNA suspension by applying the recovered volume directly to the top of the mini Quick Spin DNA Column (Boehringer Mannheim), following instructions of the manufacturer (see Fig. 17).

20

Fig. 18 shows an overview of the procedure. Genes amplified were *HXT1*, *HXT2*, *HXT4*, *HXT7* and *xyIE* that present approximately 80% homology. Before shuffling *HXT4* + *HXT7* ssDNA were re-annealed (see Fig. 19). A G50 column was also tried and worked fine.

25

2. DNase I cleavage

Each ssDNA was fragmented separately by adding DNase I in 100 mM Tris HCl pH 7.6 in the presence of 10 mM MnCl₂ at 20 °C. After 2, 5, 10 and 20 min, adding 6x gel-loading buffer to obtain 1x final concentration, followed by boiling for 10 minutes

30 stopped reactions. Aliquots were applied to a 2% agarose gel to assess the efficiency of digestion and reactions (see Fig. 20) to isolate fragments ranging from 50 to 200 bases long.

35

4. Mix of DNA fragments and amplification

- 5 Fragments isolated conventionally (*HXT4* or *HXT7*, 200 bp, and *xylE*, 300 bp) with the GFX kit were submitted to the first PCR (w/o primers) for the shuffling, using the PCR program below:

94 °C, 5 min
94 °C, 45 sec, 24 cycles
37 °C, 1 min, 24 cycles
72 °C, 1 min 30 sec, 24 cycles
72 °C, 5 min

- 10 In the next round (for amplification) the primers AB1 and AB2 and more units of Taq Polymerase were added to an aliquot of the reaction above and submitted to the same program, except that the annealing temperature was raised to 54°C.

Another programs was also tried with the mixture above and it was as follow:

94 °C, 5 min
94 °C, 1 min, 24 cycles
37 °C, 1 min, 24 cycles
72 °C, 1 min + 5 sec/cycle, 24 cycles

- 15 Use was made of the Mermaid kit, especially designed for the isolation of small fragments and/or ssDNA (BIO 101).

- 20 The use of biotinylated primers is a simple and effective procedure for generating clean ssDNA suitable for the shuffling step, without time consuming and exhaustive cloning and screening steps. Here is an scheme that illustrates that the time-frame of the method can be significantly shortened by using biotinylated material.

Non-biotinylated

PCR (2 h)



Cloning (1 day or more)

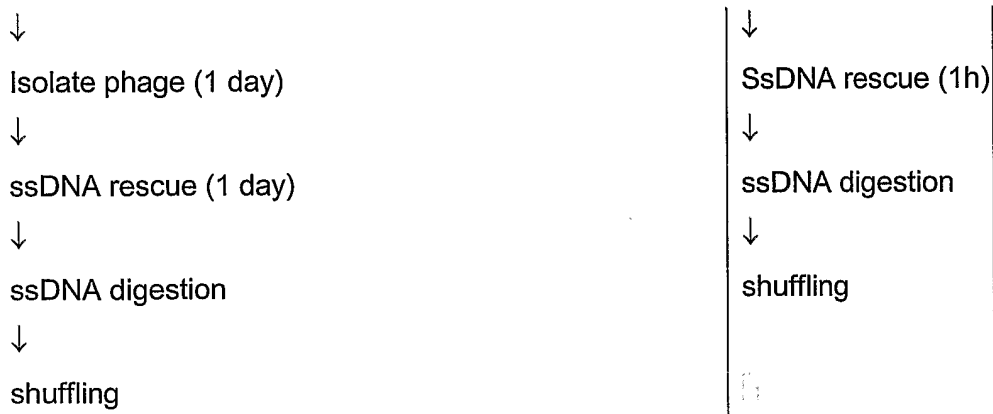
Biotinylated

PCR (2 h)



Isolate dsDNA (1h)

45



6. Transformation of the engineered strains and plating in the presence of DOG

- 5 The vector chosen for cloning the shuffled genes is p426 TEF (Mumberg et al. 1995), depicted in Fig 14, due to the presence of a sugar non-dependent promoter (TEF-Translation Elongation Factor 1 α) that enables the expression of any open reading frame introduced. So, each shuffling reaction should be cloned into this vector, previously digested with SmaI and XhoI, and after ligation introduced into the strain *S. cerevisiae* TMB 3201. Growth of this strain on plates containing xylose (as the carbon source) and DOG should be assessed. Clones displaying modified sugar transport should uptake xylose preferentially, not glucose.

Notes:

1. Selection for the plasmid is *URA3*!
- 15 2. All genes are cloned into pYES2.1 TOPO vector (Invitrogen) and can be used to put these genes back into the strain, each one at a time.

7. Assay for enhanced xylose uptake

- 20 Several alternatives are available, such as uptake of ^{13}C or ^{14}C labeled xylose for uptake experiments. However, such procedures that handle radioactive, besides validation by corresponding authorities still need implementation. Assay of intracellular xylose after metabolic arrest is another possibility, but it may be complicated. In summary, such assays are not practical.

25

The easiest procedure would be to compare the growth of new strains (with enhanced xylose uptake) to the un-modified strain. Fermentations in 50-ml flask

fermentation, employing a growth medium containing xylose as the sole carbon source could be easily performed in-house, since equipments and reagents are available.

5 References

- Arnold FH A discussion on enzyme directed evolution approaches
(<http://www.che.caltech.edu/groups/fha/Enzyme/directed.html>)
- Arnold FH (1999) Directed evolution of biocatalysts. *Curr Opin Chemical Biol* 3:
54-59
- 10 Chartrain M, Salmon PS, Robinson DK, Buckland BC (2000) Metabolic engineering
and directed evolution for the production of pharmaceuticals. *Curr Opin
Biotechnol* 11: 209-214
- Eliasson A, Christensson C, Wahlbom CF and Hahn-Hägerdal B (2000). Anaerobic
fermentation by recombinant *S. cerevisiae* carrying *XYL1*, *XYL2* and *XKS1* in
15 mineral medium chemostat cultures. *Applied and Environmental Microbiology*
66:3381-3386.
- Gibbs MD, Nevaleinen HKH, Bergquist PL (2001) Degenerate oligonucleotide gene
shuffling (DOGS): a method for enhancing the frequency of recombination with
family shuffling. *Gene*: 271: 13-20
- 20 Harayama S (1998) Artificial evolution by DNA shuffling. *TIBTECH* 16: 76-82
- Jermutus L, Honegger A, Schweslinger F, Hanes J, Plückthun A (2001) Tailoring in
vitro evolution for protein affinity or stability. *PNAS* 98: 75-80
- Joo HJ, Srisawa A, Lin Z, Arnold FH (1999) A high-throughput digital image screen
for the discovery and directed evolution of oxygenases. *Chemistry & Biology* 6:
25 699-706
- Judo, MSB, Wedel AB, Wilson C (1998) Stimulation and suppression of PCR-
mediated recombination *Nucleic Acids Res* 26: 1819-1825
- Kaneko S, Hayashi S (2000) Enhancement of the thermostability and hydrolytic
activity of xylanase by random gene shuffling. *Biochem J* 349: 651-
- 30 Kikuchi M, Ohnishi K, Harayama S (1999) Novel family shuffling methods for the in
vitro evolution of enzymes. *Gene*: 236: 159-167
- Kikuchi M, Ohnishi K, Harayama S (2000) An effective family shuffling method using
single stranded DNA. *Gene* 243: 133-137

- Kurtzman AL, Govindarajan S, Vahle K, Jons JT, Heinrichs V, Patten PA (2001) Advances in directed protein evolution by recursive genetic recombination: application to therapeutic proteins. *Curr Opin in Biotechnol* 12: 361-370
- Lassner M, Bedbrook J (2001) Directed molecular evolution in plant improvement. *Curr Opin Plant Biol* 4: 152-156
- 5 Mumberg D, Funk M (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 156:11-122.
- Özcan S, Johnston M (1999). Function and regulation of the yeast hexose transporters *Microbiol. Molec. Biol. Rev.* 63:554-569.
- 10 Soga T, Serwe, M (2000) Determination of carbohydrates in food samples by capillary electrophoresis with indirect UV detection. *Food Chemistry*, 69: 339-344
- Stemmer W (1994) DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *PNAS* 91: 10747-10751
- 15 Wieczorke R, Krampe S, Weirstall T, Friedel K, Hollenberg CP, Boles E (1999). Concurrent knock out of at least 20 transporter genes is required to block uptake of hexoses in *S. cerevisiae*. *FEBS Lett.* 464: 123-128.
- Yano T, Kagamiyama H (2001) Directed evolution of ampicillin-resistant activity from a functionally unrelated DNA fragment: a laboratory model of molecular evolution. *PNAS* 98: 903-907
- 20 Zhao H, Arnold F (1997) Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res* 25: 1307-1308

Table 2 – Nucleotide sequence of the primers (5' → 3')

FORWARD	REVERSE	GENE
GTTCACTAGTTACCCGGGATGAATTCAACTCCCGATCTAATATCTCC (100 pmol/μl) (SEQ ID 1)	TCTTGTAGTCGACCTCGAGCGTTTATTCTCTGCTAAACAACTCTTG (100 pmol/μl) (SEQ ID 8)	HXT1
GTTCACTAGTTACCCGGGATGTCTGAATTCGGCTACTAGCCGC (100 pmol/μl) (SEQ ID 2)	TCTTGTAGTCGACCTCGAGCGCTTATTCTCTCGGAAACTC (100 pmol/μl) (SEQ ID 9)	HXT2
GTTCACTAGTTACCCGGGATGTCTGAAGAAGCTGCCTATCAAG (100 pmol/μl) (SEQ ID 3)	TCTTGTAGTCGACCTCGAGCGCATTAGATCATCAGCGTTGTAGTC (100 pmol/μl) (SEQ ID 10)	HXT4
GTTCACTAGTTACCCGGGATGTCACAAAGACGCTGCTATTGCAG (100 pmol/μl) (SEQ ID 4)	TCTTGTAGTCGACCTCGAGCGTGTTCGCAAAATTATTGGTGCTG (100 pmol/μl) (SEQ ID 11)	HXT7
GTTCACTAGTTACCCGGGATGAATACCCAGTATAATTCAGTTA (100 pmol/μl) (SEQ ID 5)	TCTTGTAGTCGACCTCGAGCGTTACAGCGTAGCAGTTTGTGT (100 pmol/μl) (SEQ ID 12)	xyIE
Biotin-GTTCACTAGTTACCCGGGATG (primer A) (265 pmol/μl) (SEQ ID 6)	Biotin-TCTTGTAGTCGACCTCGAGC (primer Biotin_B) (265 pmol/μl) (SEQ ID 13)	Remplification primers
GTTCACTAGTTACCCGGGATG (primer A) (265 pmol/μl) (SEQ ID 7)	TCTTGTAGTCGACCTCGAGC (primer_B) (265 pmol/μl) (SEQ ID 14)	Remplification primers

- Primers were re-suspended in Nano-pure water and final concentrations are given in parenthesis.

Table 3

- PCR reaction to pull out HXT genes

5	Template	0.1-0.5 µg
	Forward Primer	10 pmol
	Reverse Primer	10 pmol
	MgCl ₂	1.5 mM
	dNTPs	200 µM
10	H ₂ O to 50 µl	

* Usually the yield was 10-20 µg PCR product in 200 µl Rxn

- | | |
|----|---|
| 15 | <ul style="list-style-type: none"> • Isolation from gel using GFX Kit from Pharmacia |
| | <ul style="list-style-type: none"> - 4 lanes containing 50 µl Rxn each - elution in 100 µl TE |

- | | |
|----|--|
| 20 | <ul style="list-style-type: none"> • Biotinylated PCR |
|----|--|

	Template 1:10	0.5 µl
	AB1/AB2 Primer	10 pmol
25	Forward/Reverse Primer	10 pmol
	MgCl ₂	1.5 mM
	dNTPs	200 µM
	H ₂ O to 50 µl	

- | | |
|----|---|
| 30 | <ul style="list-style-type: none"> • ssDNA isolation |
|----|---|

Beads were in PBS at 10.8 mg/ml.

- | | |
|----|---|
| 35 | <ul style="list-style-type: none"> - Equilibration of the beads: |
| | 1 ml beads into an Eppendorf |
| | Separation using the magnet |
| | Removal of the supernatant |
| | 1 ml TE and separation (3x) |

- | | | |
|----|--|--------|
| 40 | <ul style="list-style-type: none"> - Isolation: | |
| | Beads | 100 µg |
| | PCR | 5 µg |
| | TE to 250 µl | |

- | | |
|----|---|
| 45 | Incubation for 30 minutes at room temperature |
| | Separation |
| | 3x washing in 200 µl TE |
| | Elution in 100 µl 100 mM NaOH/1 mM EDTA |

- | | |
|----|---|
| 50 | <ul style="list-style-type: none"> * about 4-4.5 µg ssDNA were rescued |
|----|---|

- ssDNA cleaning

Volume recovered as above directly applied to the Pre-packed Mini column

5

- * about 4 µg ssDNA NaOH-free were recovered

- PCR program for amplification of genes

10

1- 94°C, 3 min for initial denaturation

2- 94°C, 30 sec

3- 54°C, 30 sec

4- 72°C, 1 min 30 sec

15

5 – 24 repetitions from step 2

6- 72°C, 5 min

7- 4°C

- DNaseI treatment

20

4-5 µg ssDNA

0.005 U DNaseI

100 mM TrisHCl pH 7.6

25

10 mM MnCl₂

H₂O to 100 µl

Incubation at 20°C, 5-20 minutes and reaction stopped by adding loading buffer and boiling. Samples applied to 2% agarose-TAE gel and fragments isolated.

30

Claims

1. A method for shuffling of genes, said method comprising the following steps:
 - a) selecting at least two double stranded DNA polynucleotides to be shuffled,
 - 5 b) amplifying the at least two double-stranded DNA polynucleotides using PCR with primers, each primer pair having a sequence for templating the selected polynucleotide and an oligonucleotide tag, the oligonucleotide tags of the two primers of a pair being different, thereby obtaining amplified tagged polynucleotides,
 - 10 c) performing one further round of PCR on each of the amplified tagged polynucleotides with primer pairs for templating the oligonucleotide tags, wherein one of the primers thereby incorporated into each polynucleotide is an affinity-tagged primer, the affinity-tagged primers being selected so that one affinity-tagged primer is incorporated into one strand of one polynucleotide and one
 - 15 affinity-tagged primer is incorporated into the corresponding complementary strand of another polynucleotide,
 - d) isolating the synthesised double-stranded DNA polynucleotides using the affinity-tags,
 - e) separating the tagged from non-tagged DNA strands,
 - 20 f) cleaving isolated non-tagged DNA strands, and
 - g) shuffling the cleaved DNA strands.
2. The method according to claim 1, wherein step g) involves shuffling the cleaved single stranded DNA strands.
- 25 3. The method according to claim 1, wherein the method further comprises the step of preparing heteroduplexes of DNA by incubating non-tagged DNA strands under hybridisation conditions, wherein said step is performed between steps e) and f).
- 30 4. The method according to claim 3, wherein step f) involves cleaving the heteroduplexes.
5. The method according to claim 4, wherein step g) involves shuffling the double
- 35 stranded cleaved heteroduplexes.

6. The method according to claim 1, wherein the forward primer of each pair used for amplifying the selected polynucleotides have identical first oligonucleotide tags.
- 5
7. The method according to claim 6, wherein the reverse primer of each pair used for amplifying the selected polynucleotides have identical second oligonucleotide tags.
- 10
8. The method according to claim 7, wherein the primer pair for the further round of PCR for one polynucleotide comprises an affinity-tagged primer for templating the first oligonucleotide tag, and the primer pair for the further round of PCR for another polynucleotide comprises an affinity-tagged primer for templating the second oligonucleotide tag.
- 15
9. The method according to claim 1, wherein the oligonucleotide tags comprise a restriction site for inserting the shuffled sequences into a vector.
10. The method according to claim 9, wherein the restriction site is a rare restriction site having at least 6 nucleotides in the recognition sequence, preferably at least 8 nucleotides in the recognition sequence.
- 20
11. The method according to claim 1, wherein the length of the oligonucleotide tags at least 15 nucleotides, more preferably at least 18 nucleotides, such as at least 21 nucleotides, for example at least 24, such as at least 27, for example at least 30, such as at least 35, for example at least 40, such as at least 50.
- 25
12. The method according to claim 1, wherein the oligonucleotide tags comprise a sequence which is optimal for PCR, e.g. good melting temperature of dsDNA, good GC content according to the commonly accepted guidelines for good primer design.
- 30
13. The method according to claim 1, wherein the primers with oligonucleotide tags comprise an adaptamer.
- 35

14. The method according to claim 1, wherein the isolated single-stranded DNA is cleaved using an endonuclease.
15. The method according to claim 14, wherein the endonuclease is DNase I.
- 5 16. The method according to claim 1, wherein the cleavage of the isolated single-stranded DNA is performed by using e.g. classical Maxam-Gilbert chemistry.
- 10 17. The method according to claim 1, wherein the cleavage of the isolated single-stranded DNA is performed using ultrasound.
- 15 18. The method according to claim 4, wherein the heteroduplexes are cleaved using an endonuclease, which preferentially cleaves double stranded DNA over single stranded DNA.
19. The method according to claim 18, wherein the endonuclease is DNase I.
- 20 20. The method according to claim 1, further comprising use of an exonuclease during shuffling.
21. The method according to claim 1, further comprising introduction of mutations during shuffling.
- 25 22. The method according to claim 1, wherein shuffling is performed by
iii) removing the endonuclease or chemical reaction mix,
iv) reassembling double-stranded DNA using PCR without primers.
- 30 23. The method according to claim 1, further comprising amplification of the shuffled double stranded DNA using PCR with primers.
24. The method according to claim 23, wherein the primers used for the PCR are selected to ensure that the amplified sequence comprises one end from one selected polynucleotide and the other end from another selected polynucleotide.
- 35

25. The method according to claim 1, wherein separation of the tagged from non-tagged strands is performed by denaturing the double-stranded DNA.
- 5 26. The method according to claim 25, wherein denaturing comprises use of a base, preferably NaOH and/or KOH.
27. The method according to claim 1, wherein the affinity-tagged primer comprises one member of a species-specific pair.
- 10 28. The method according to claim 27, wherein association between the two members of the species-specific pair is sufficiently strong to survive conditions which cause denaturation of double-stranded DNA.
- 15 29. The method according to claim 27, wherein the member of the species specific pair on the affinity-tagged primer is biotin
- 20 30. The method according to claim 1, wherein isolation of the synthesised doublestranded DNA polynucleotides is performed by capturing said polynucleotides on a member of species specific pair being bound to a solid surface.
31. The method according to claim 30, wherein the member of the species specific pair bound to the solid support is avidin or streptavidin.
- 25 32. The method according to claim 30, wherein the member of the species specific pair is linked to the surface of a well, a test tube..
- 30 33. The method according to claim 30, wherein the member of the species specific pair is linked to a bead, such as a magnetic bead, agarose bead.
34. The method according to claim 1, further comprising screening and selecting shuffled genes for a pre-selected characteristic.
- 35 35. The method according to claim 1, further comprising repeating the process a number of times.

36. The method according to claim 34, comprising repeating the steps until at least one shuffled gene having a pre-determined functionality is obtained.
- 5 37. The method according to claim 1, wherein the polynucleotides to be shuffled have homologous sequences.
38. The method according to claim 1, wherein the polynucleotides to be shuffled have sequences with sufficiently identity to allow gene shuffling.
- 10 39. The method according to claim 1, wherein more than two polynucleotide sequences are selected for shuffling, the last PCR amplification being performed to ensure that at least one coding strand and at least one corresponding complementary strand is affinity-tagged.
- 15 40. The method according to claim 39, wherein a family of genes are selected for shuffling, such as a family of at least 3 genes, for example at least 4 genes, such as at least 5 genes, for example at least 6 genes, such as at least 7 genes, for example at least 8 genes, such as at least 9 genes, for example at least 10 genes, for example at least 15 genes, such as at least 20 genes or more.
- 20 41. The method according to claim 1, wherein the at least two double stranded DNA polynucleotides to be shuffled comprise carbohydrate transporters such as HXT-genes and xylE.
- 25 42. The method according to claim 1, wherein the at least two double-stranded DNA polynucleotides to be shuffled encode two homologous proteins, one of which interact with another protein.
- 30 43. A method for shuffling of genes, said method comprising the steps of
- i) selecting at least two polynucleotides to be shuffled,
 - ii) performing one round of PCR on the polynucleotides with primer pairs, wherein one of the primers incorporated into each polynucleotide is an affinity-tagged primer, the affinity-tagged primers
- 35 being selected so that one affinity-tagged primer is incorporated into

- one strand of one polynucleotide and the other affinity-tagged primer is incorporated into the corresponding complementary strand of another polynucleotide,
- iii) isolating the synthesised double-stranded DNA polynucleotides using the affinity-tags,
- iv) separating the tagged from non-tagged DNA strands,
- v) cleaving either the isolated tagged DNA strands or the non-tagged DNA strands,
- vi) shuffling the cleaved DNA strands.

10

44. A kit for shuffling of genes, said kit comprising

- two sets of primers, each set having one primer to template a first oligonucleotide tag and one to template a second oligonucleotide tag, one set having the primer to template the first oligonucleotide tag linked to an affinity-tag and the other set having the primer to template the second oligonucleotide tag linked to an affinity-tag,
- at least one endonuclease,
- suitable buffers and reagents.

15

45. The kit according to claim 43, wherein the endonuclease is DNase I (EC 3.1.21.1).

20

46. The kit according to claim 43, further comprising two or more sets of primers, each primer having a sequence for templating a polynucleotide to be shuffled and a first or a second oligonucleotide tag, each pair having the first tag in the forward primer and the second tag in the reverse primer.

25

47. The kit according to claim 46, wherein the two or more sets of primers comprise adaptamers.

30

48. The kit according to claim 43, wherein the affinity-tags comprise part of a species specific pair.

49. The kit according to claim 46, wherein the species specific pair is biotin.

35

50. The kit according to claim 43, further comprising means to capture the affinity tag, these means comprising a member of a species specific pair being linked to a solid surface.
- 5 51. The kit according to claim 50, wherein the solid surface comprises a test tube, a multi-well dish, beads including magnetic beads, or other useful solid supports.
52. The kit according to claim 50, wherein the member of the species specific pair bound to a solid surface comprises avidin, streptavidin
- 10 53. A chimeric hexose transporter obtained through gene shuffling, said chimeric hexose transporter having a higher affinity for xylose and/or a lower affinity for glucose than the affinity of the unshuffled hexose transporter for said carbohydrates.
- 15 54. The chimeric hexose transporter of claim 53, said chimeric hexose transporter having a K_M for xylose lower than 90 mM and/or a K_M for glucose higher than 30 mM.
- 20 55. The chimeric hexose transporter of claim 53, containing sequences from at least one HXT-gene and at least one xylE-gene.
56. The chimeric hexose transporter of claim 53, being obtained by shuffling according to any of the claims 1 to 42.

25

Amplification of members of a gene family or of homologous genes from different species by PCR using generically tagged primers

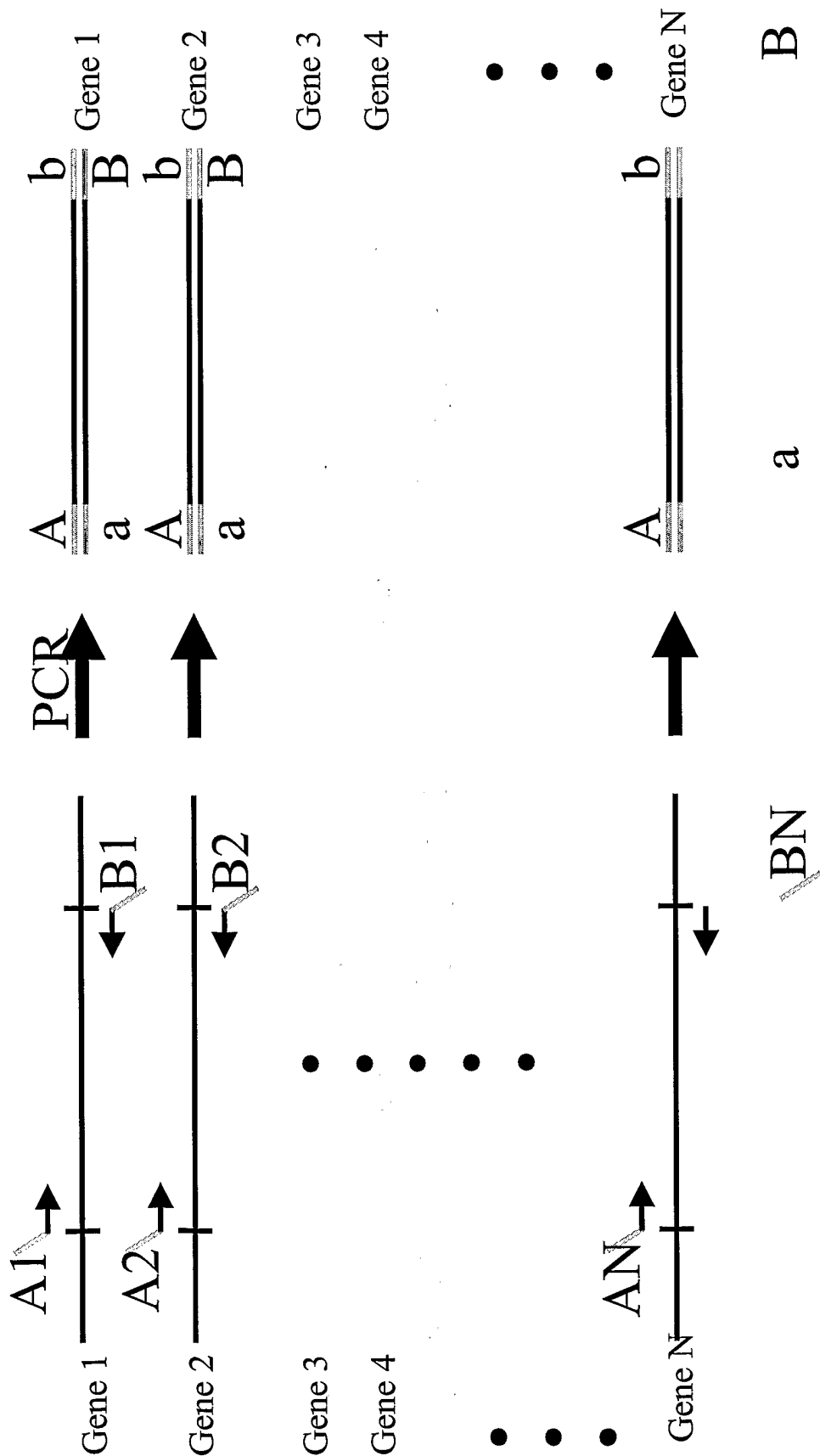
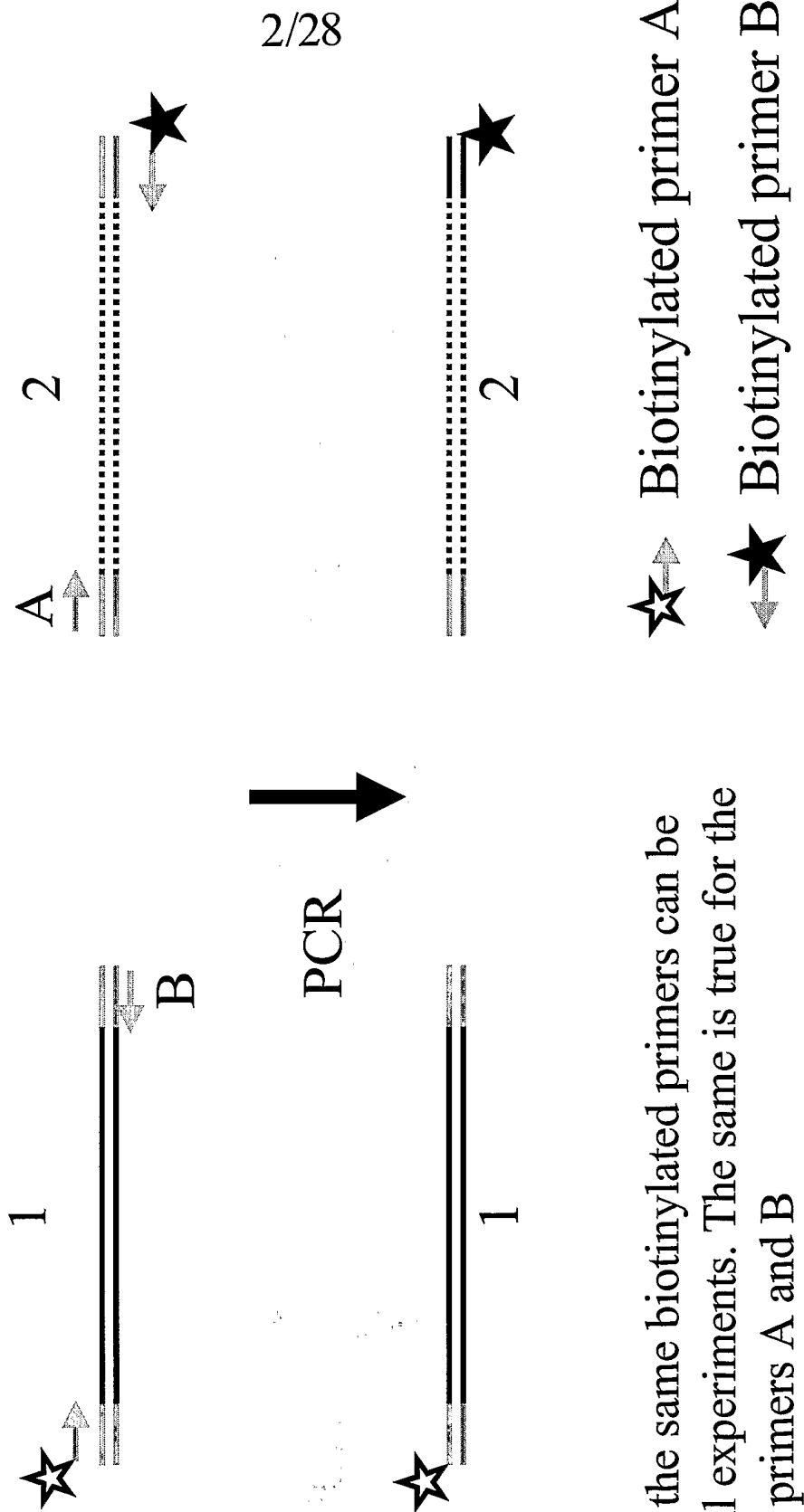


Fig. 1 Note that grey Aa/Bb tags are identical in all fragments

Reamplification of fragments using a primer pair of which one is biotinylated



Note that the same biotinylated primers can be used in all experiments. The same is true for the "normal" primers A and B

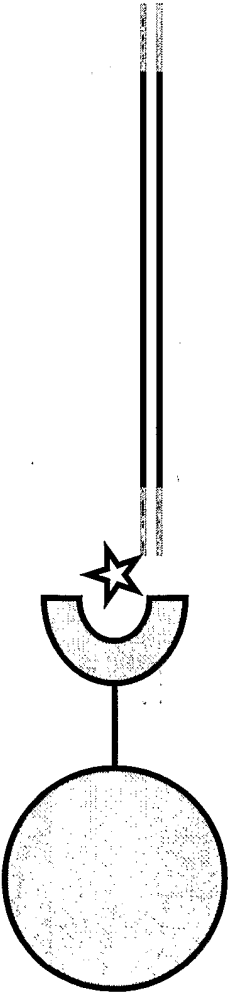
Fig. 2

Purification of single stranded DNA fragments - Part I



Bind to Streptavidin Coated Beads

Reaction 1



Reaction 2

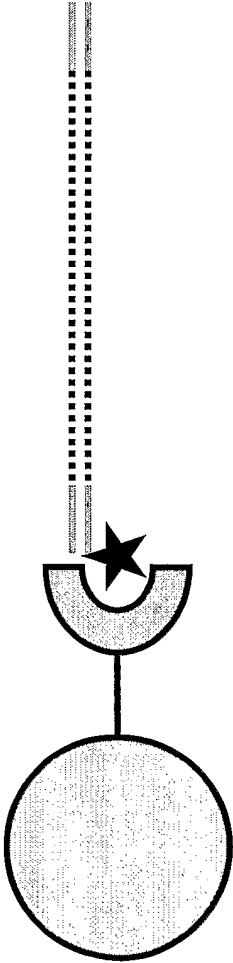


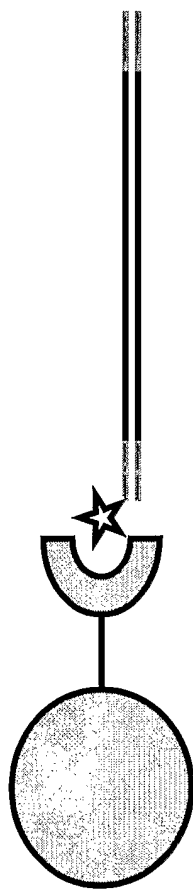
Fig. 3a

4/28

Purification of single stranded DNA fragments - Part II

Capture DNA via biotin-streptavidin interaction

Reaction 1



Reaction 2

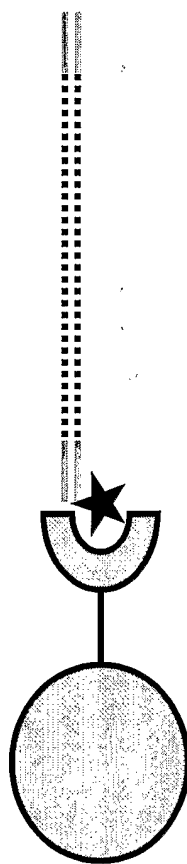
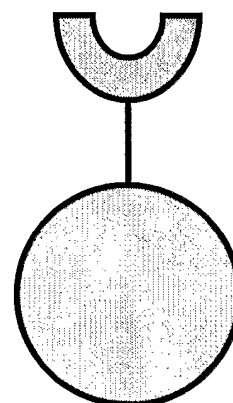


Fig. 3b



Streptavidin coupled to a magnetic bead

Purification of single stranded DNA fragments - Part III

WO 2004/016811

PCT/DK2003/000545

Denature DNA

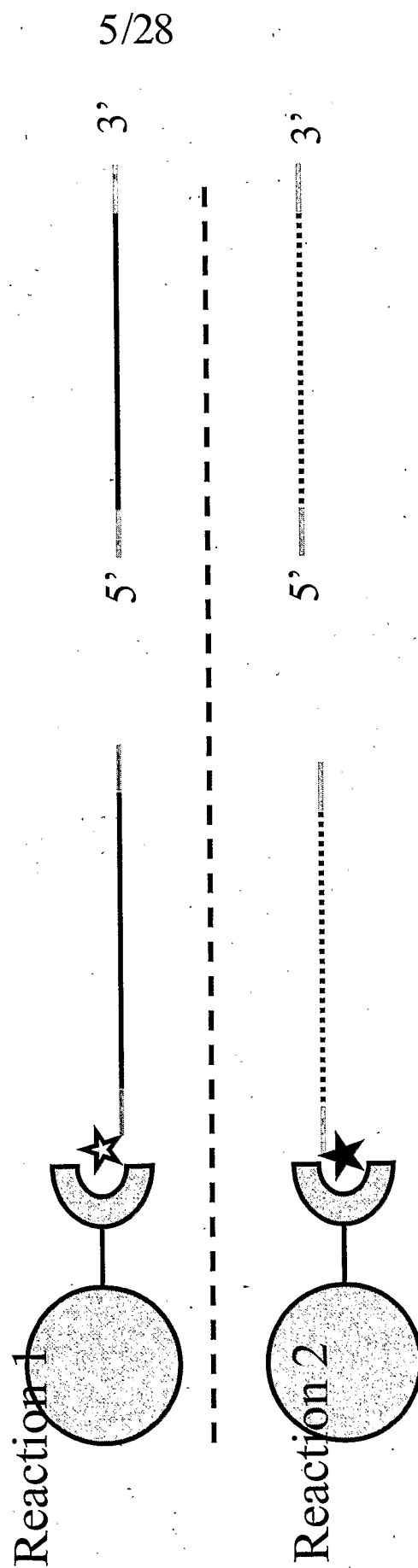
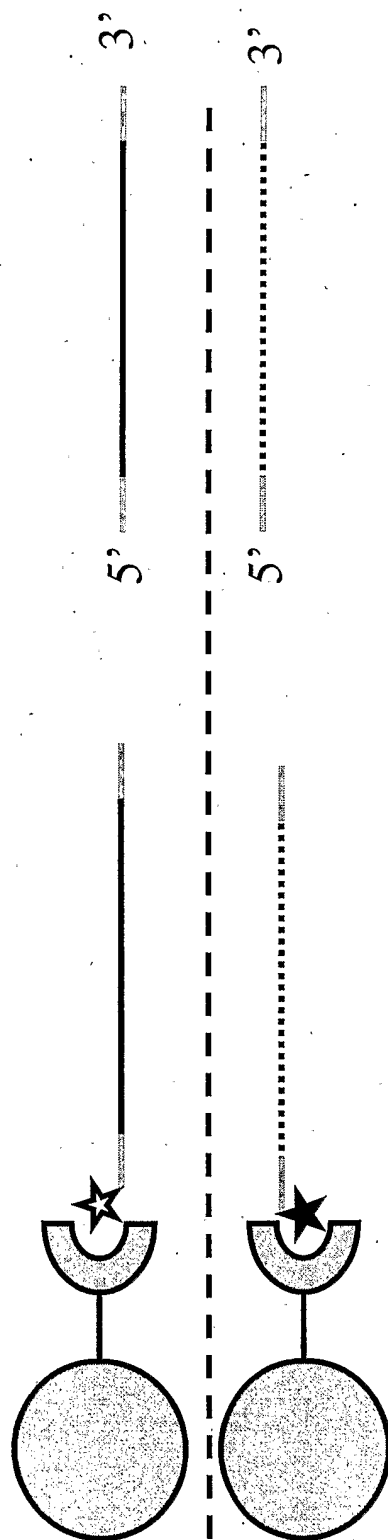


Fig. 3c

6/28

Purification of single stranded DNA fragments - Part IV

Reaction 1



Reaction 2

Remove Beads - Save Supernatants

Reaction 1



Reaction 2



Fig. 3d

Shuffling of single stranded DNA fragments - Part I

WO 2004/016811

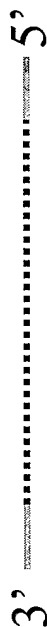
PCT/DK2003/000545

7/28

Reaction 1



Reaction 2



DNAse I (or other nuclease) or chemical treatment

Reaction 1



Reaction 2



Fig. 4a

Shuffling of single stranded DNA fragments - Part II

WO 2004/016811

PCT/DK2003/000545

8/28

Remove DNase I, other nuclease, or chemical and mix reactions

From Reaction 1 5' ————— 3'

From Reaction 2 3' 5'

PCR, without primers



Fig. 4b

5' ===== 3'

Shuffling of single stranded DNA fragments - Part III

Amplification of the shuffled fragment by PCR

(And a method to avoid remaining background)

Chimeras of this type can be selected by using primers A1 and B2 in new PCR reaction

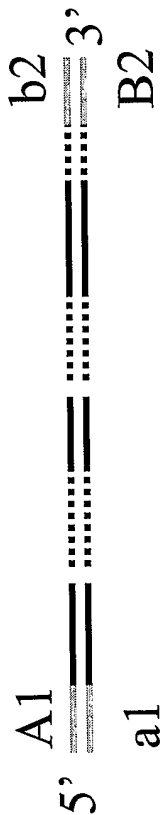
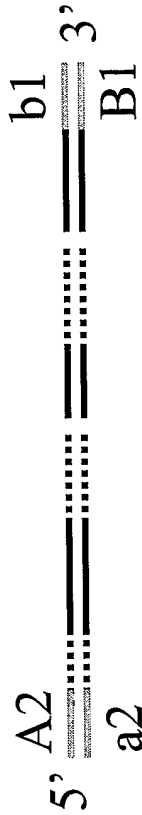


Fig. 4c

Chimeras of this type can be selected by using primers A2 and B1 in a new PCR reaction



Note: Some chimeras will be lost in this process. However, these can be regenerated by repeating the procedure (produce and isolate ssDNA from the reactions above followed by reshuffling) and then selecting with primer pair (A1 + B1) and primer pair (A2 + B2) in the final step.

10/28

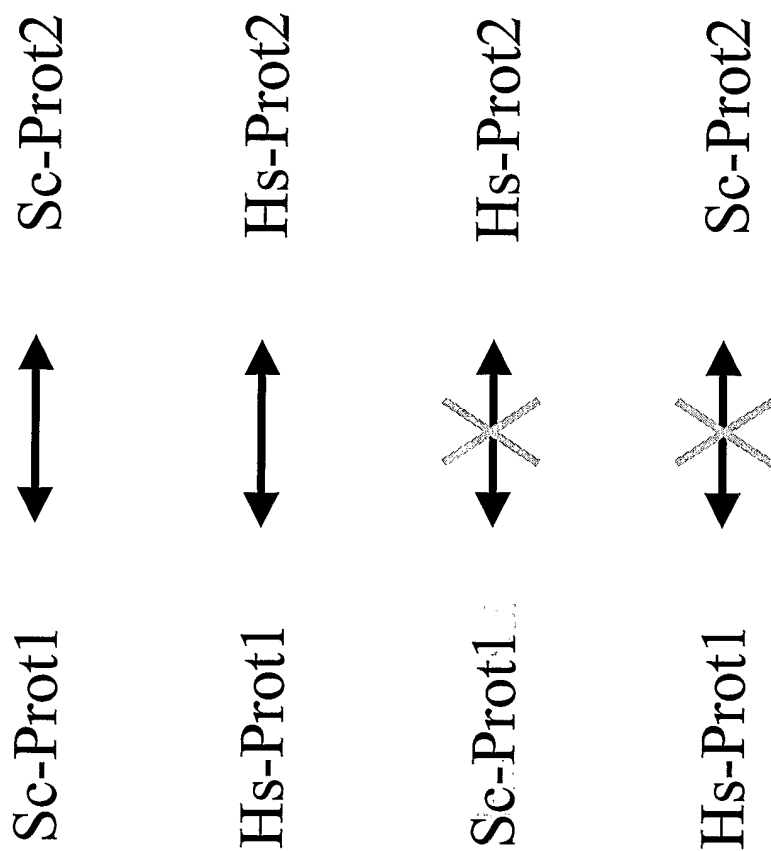


Fig. 5

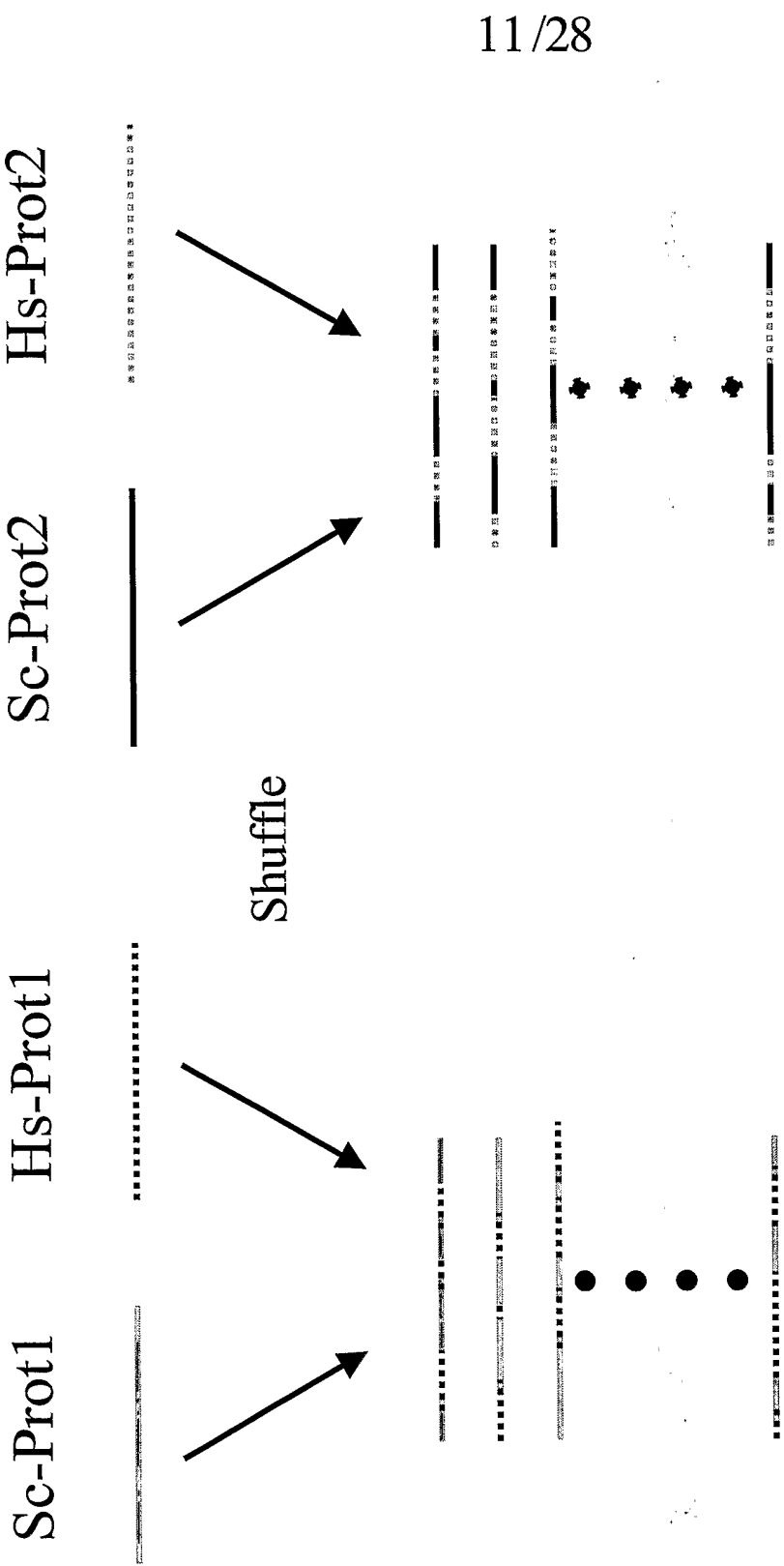


Fig. 6

Chimeras to be tested for their abilities to interact with either Sc-Prot2 And Hs-Prot2

Chimeras to be tested for their abilities to interact with either Sc-Prot1 And Hs-Prot1

(Exemplified on next page)

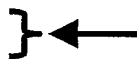
Analysis of Prot1 Chimeras

Prot1 Chimeras	Interaction to Sc-Prot2	Interaction to Hs-Prot2
.....	+	-
.....	-	+
.....	-	+
•	•	•
•	•	•
•	•	•
•	•	•
.....	+	-

12/28

Fig. 7

Putative interaction region



In this case the likely Prot2 interaction domain in Prot1 is situated in its C-terminus.

It would be advantageous to perform the experiment in two sections. One that shuffles big DNA pieces together (to get a rough idea about the location of the domain of interest), and one that shuffles small pieces of DNA together (to map the location of the domain more accurately).

Example 2: Mapping of essential amino acids in a Protein

Part II

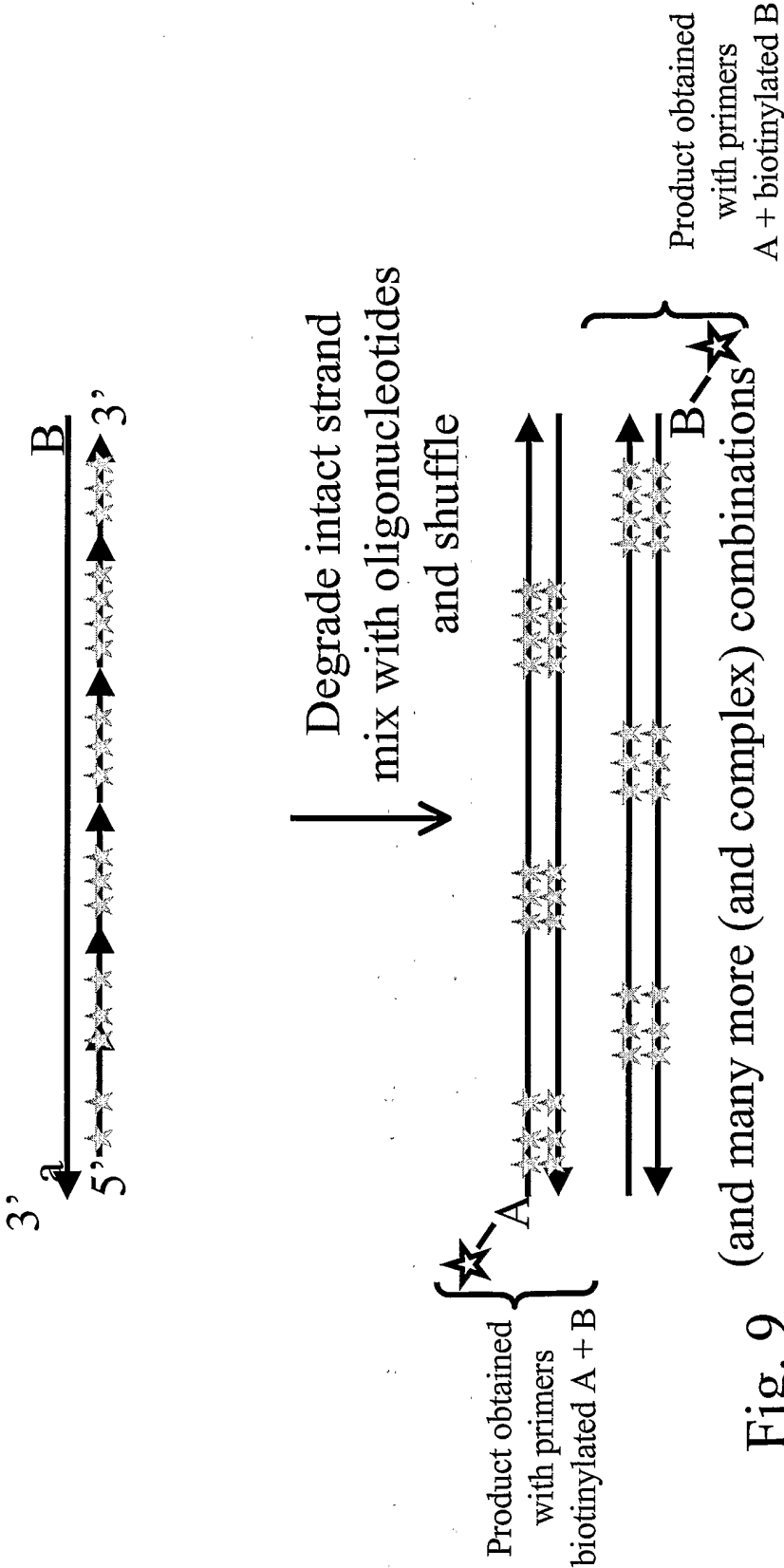


Fig. 9

Note - it may also be useful to prepare a double stranded DNA fragment that contain all mutations and use that for shuffling according to the present protocol.

Example 2: Mapping of essential amino acids in a Protein

Part III

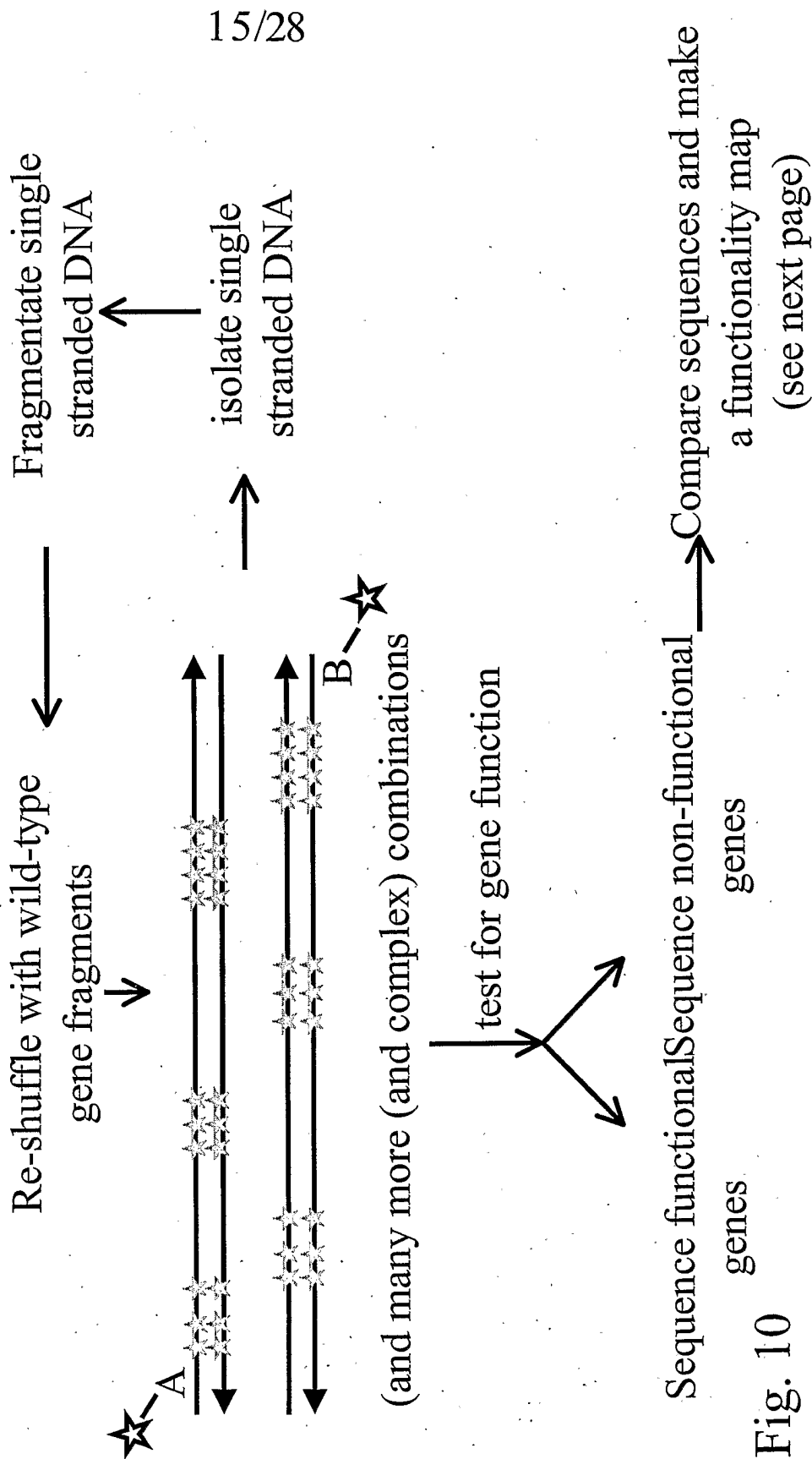


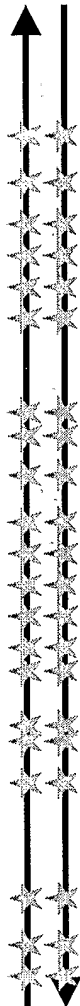
Fig. 10

Example 2: Mapping of essential amino acids in a Protein

Part III

Functionality map of Gene X

Location of mutations that does not influence gene function:



Location of mutations that does influence gene function:

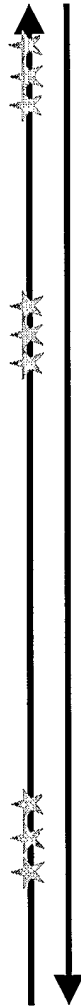


Fig. 11

17/28

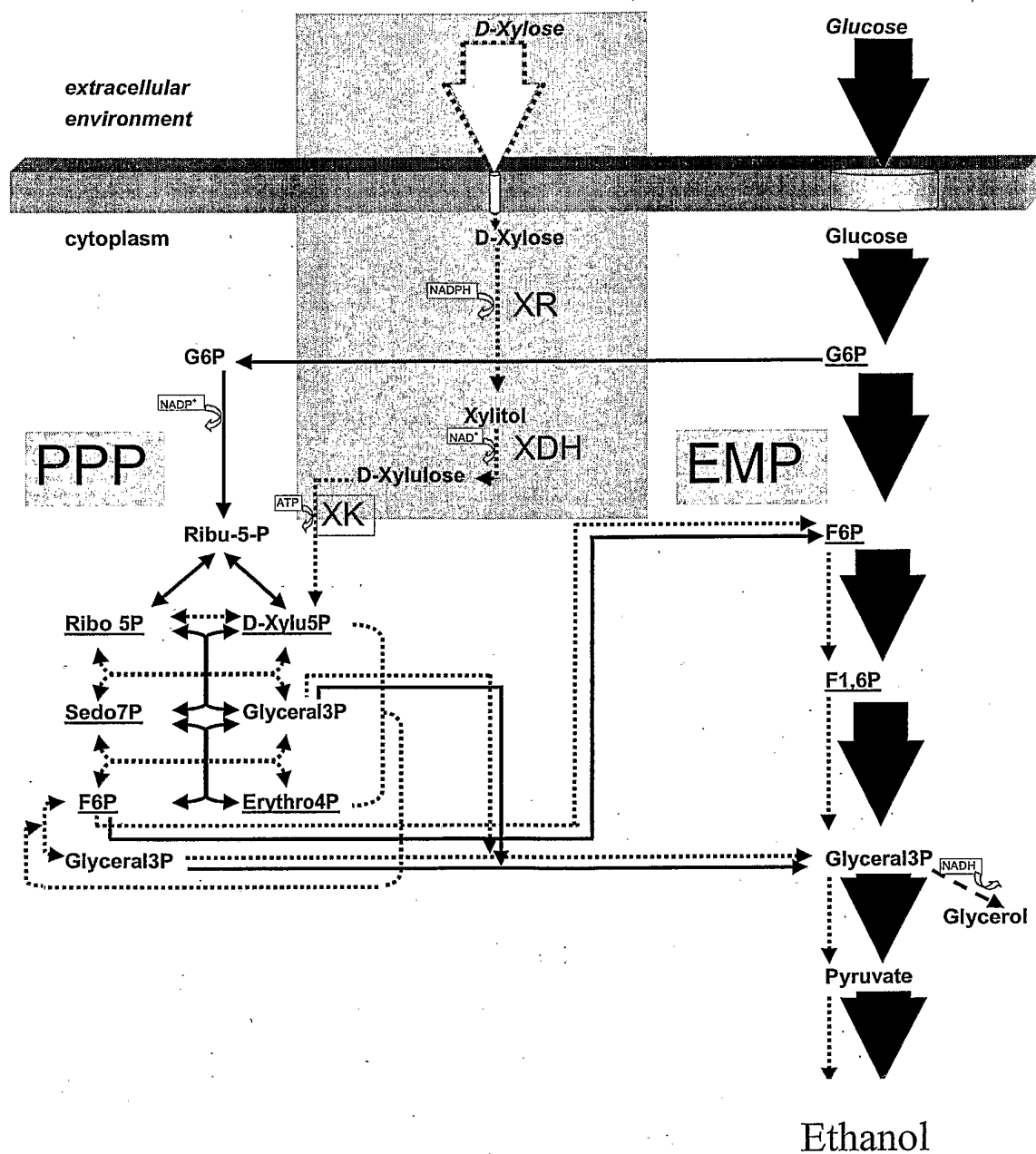
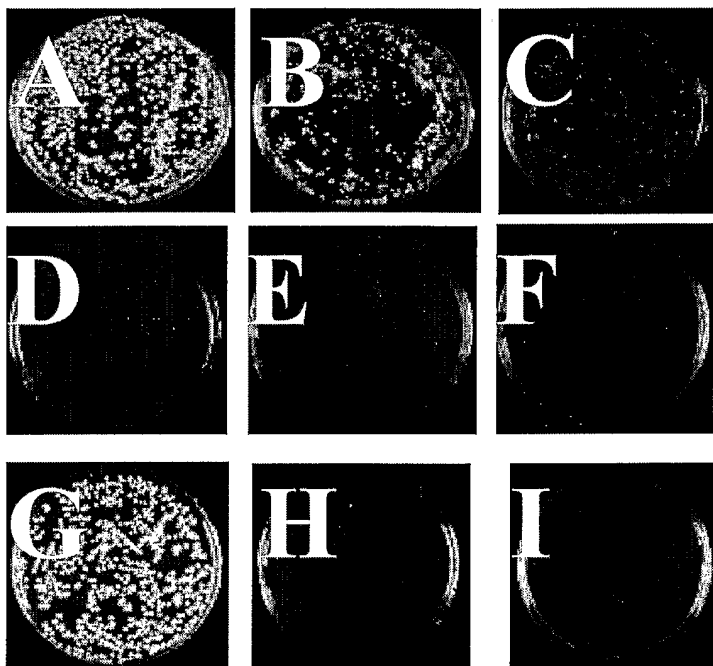
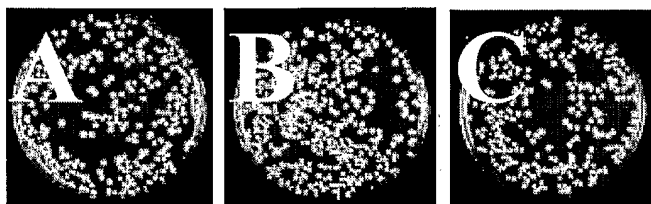


Figure 12

18/28

Figure 13.1.**Figure 13.2.**

19/28

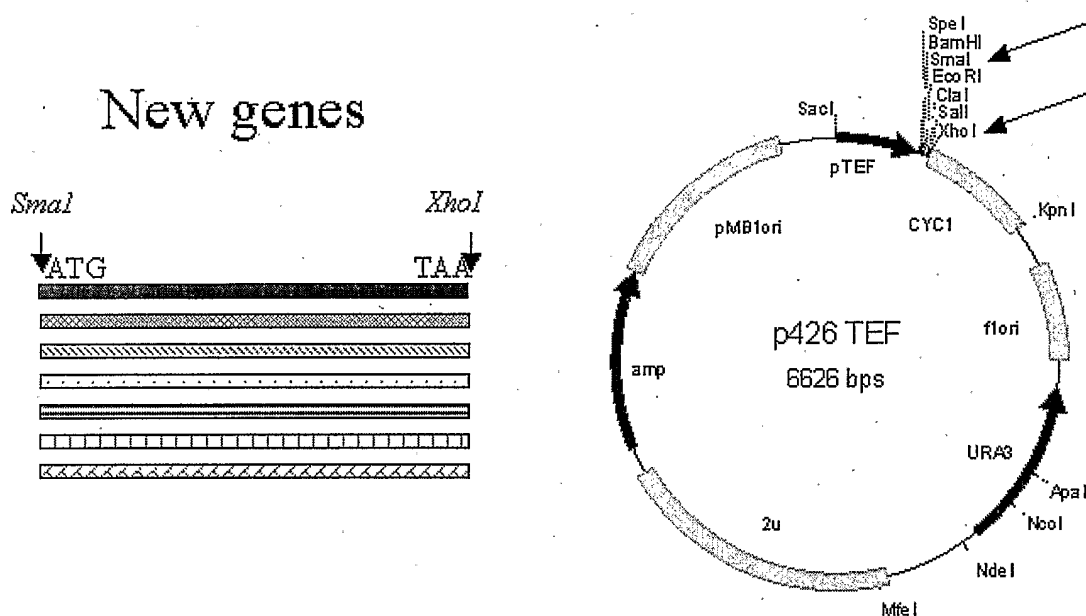


Figure 14.

20/28

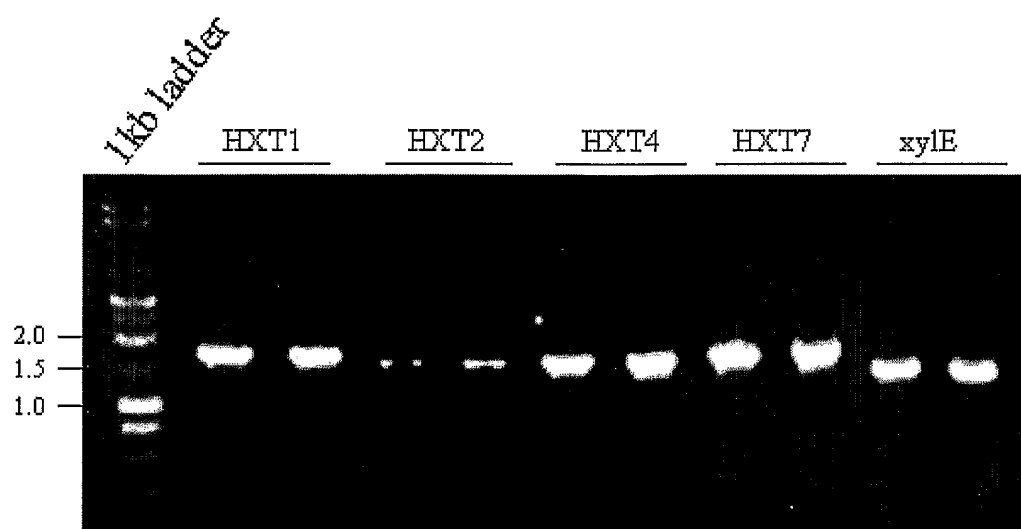
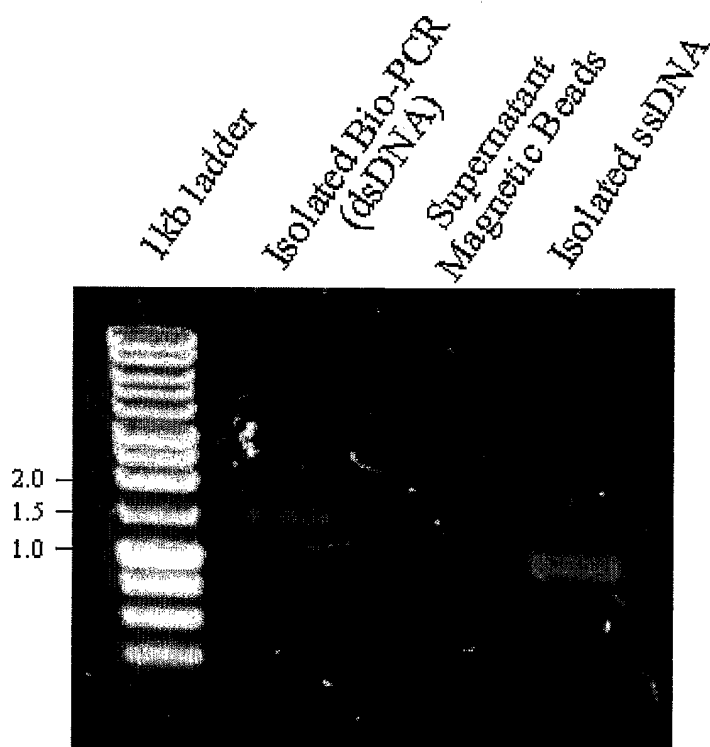


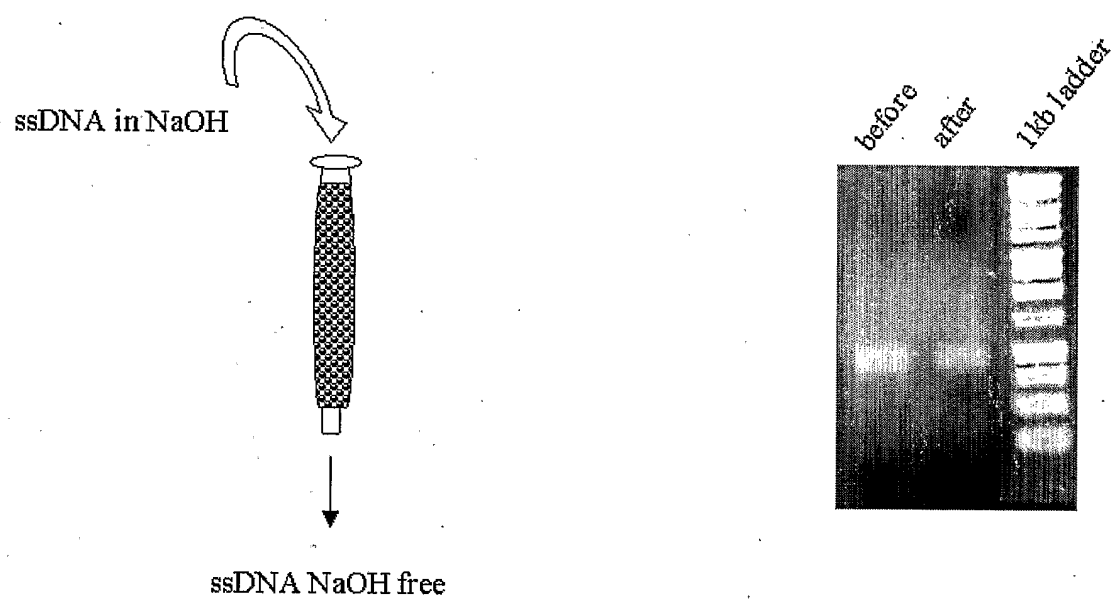
Figure 15.

21/28

**Figure 16.**

22/28

Figure 17.



23/28

Procedure for isolation of ssDNA

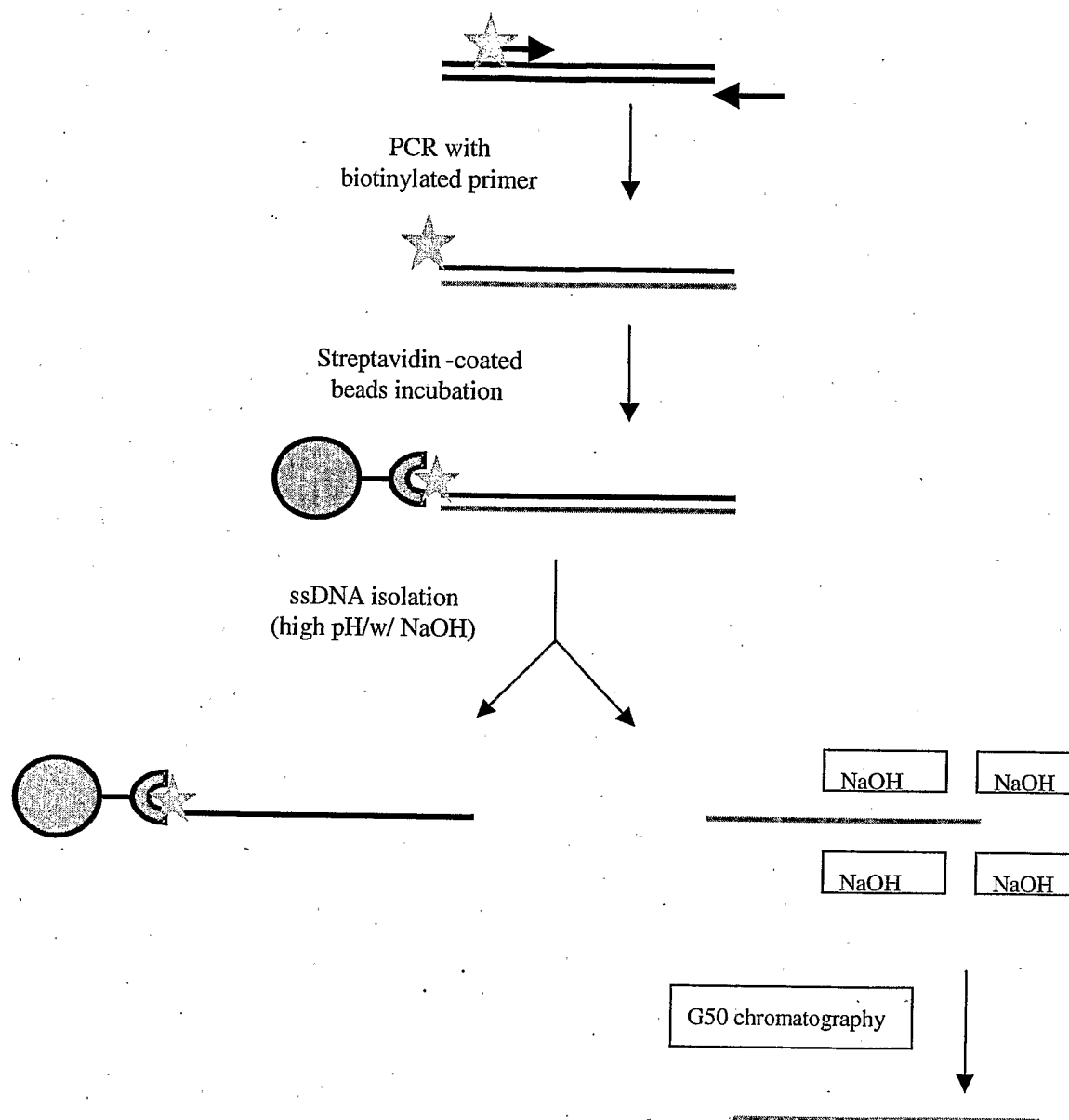
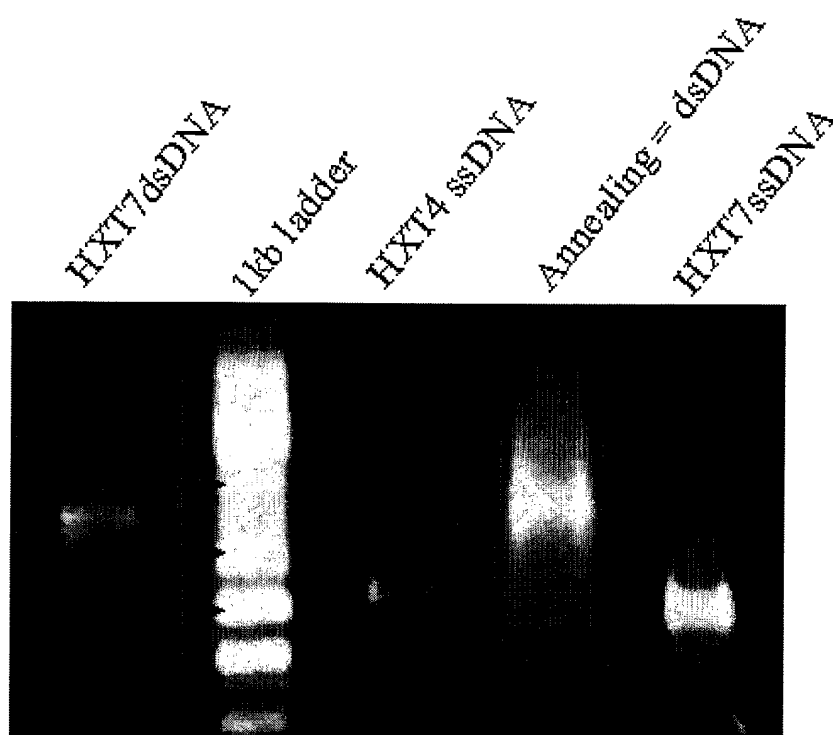


Figure 18.

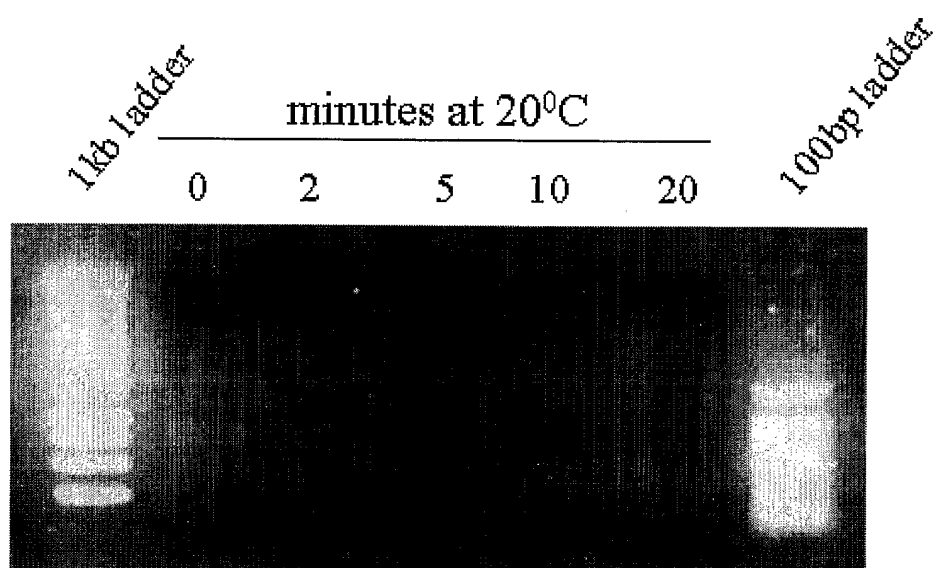
24/28

Figure 19.



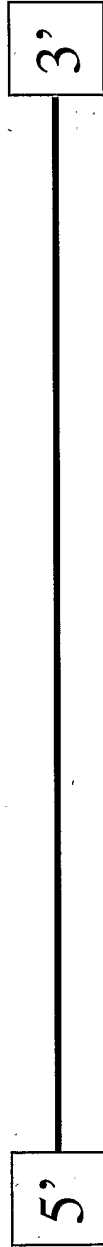
25/28

Figure 20.



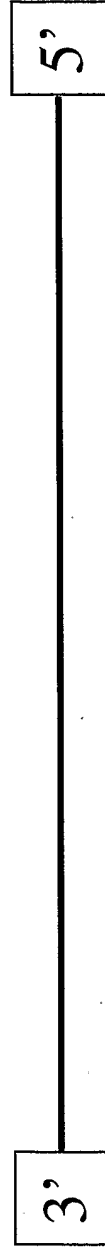
Example of shuffling using heteroduplex DNA - Step 1

ssDNA Orf-1:



+

ssDNA Orf-2:



Heteroduplex:

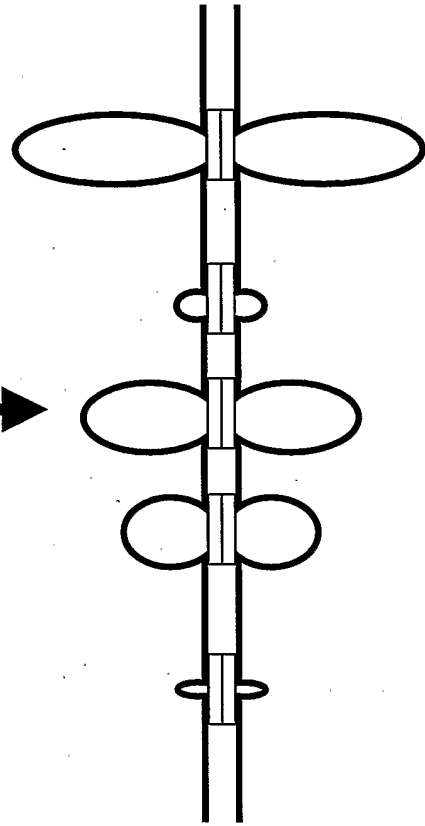


Fig. 21a

27/28

Example of shuffling using heteroduplex DNA - Step 2

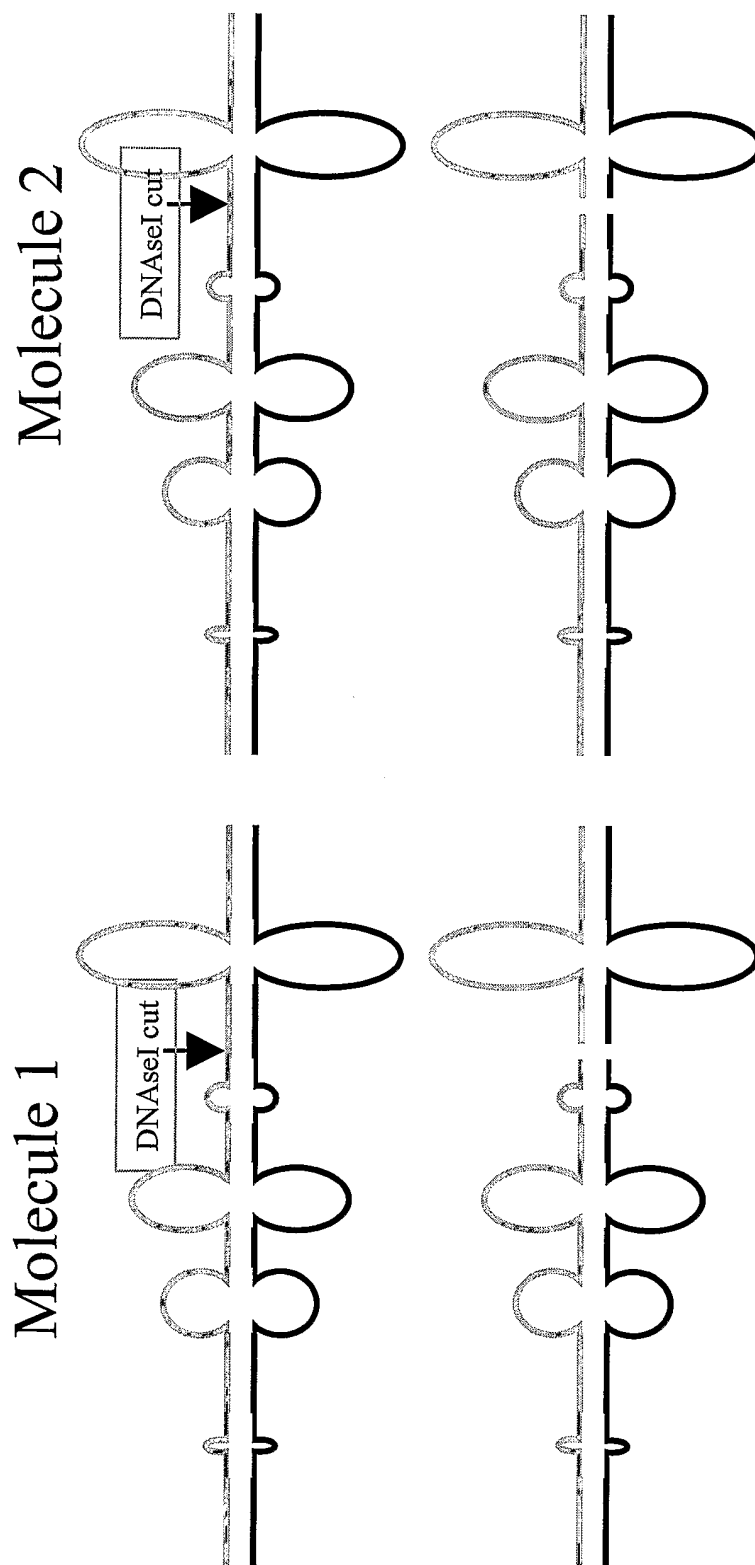


Fig. 21b

In this step, the heteroduplex is cut in the regions, which form dsDNA. In this example DNaseI is used. This enzyme prefers to cut dsDNA. For simplicity only cleavage of two molecules in a large population of molecules is shown. Each of the two molecules are cut once and in the same segment of dsDNA. However, not at the same position.

Example of shuffling using heteroduplex DNA - Step 3

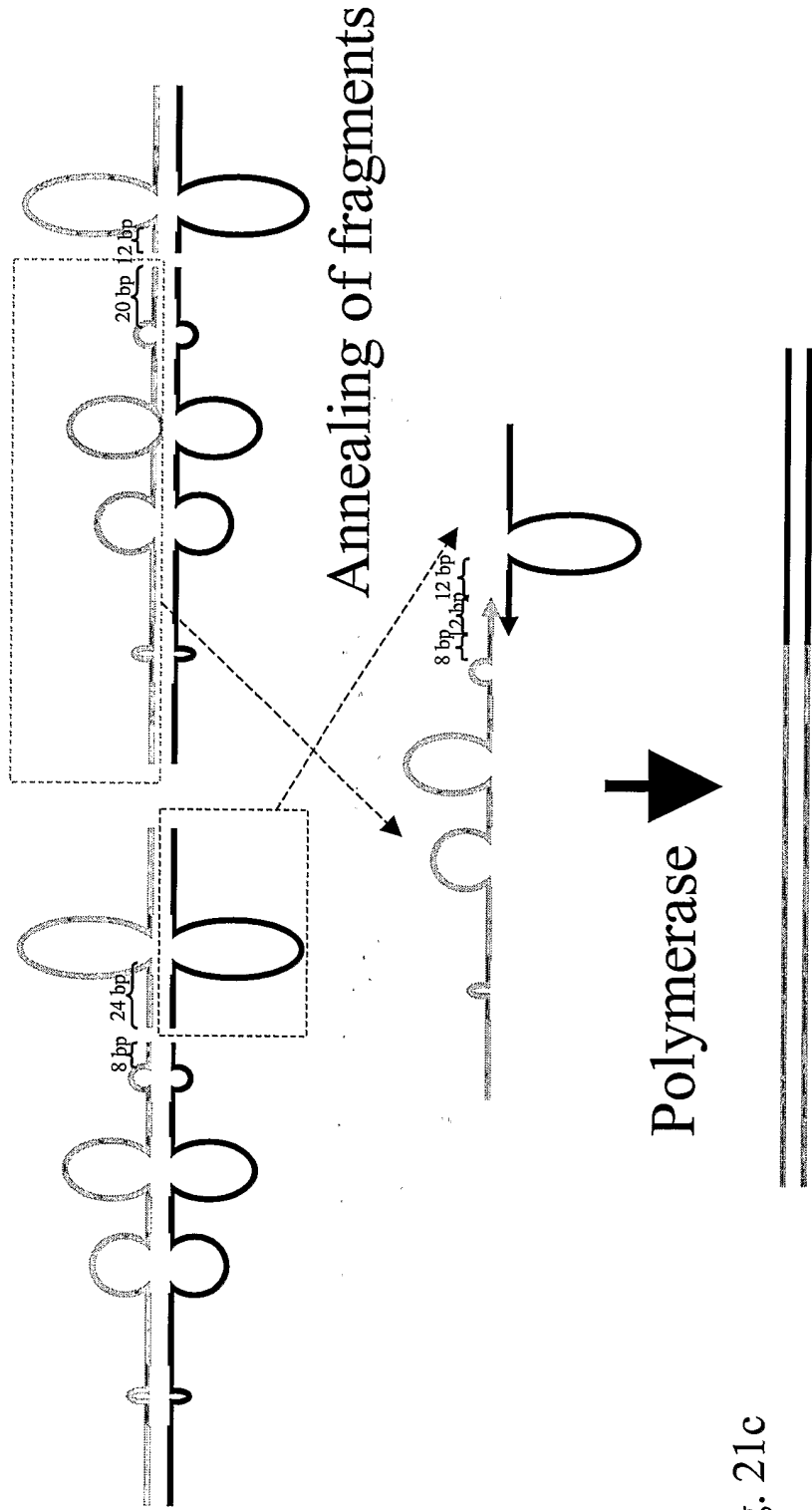


Fig. 21c

The DNaseI treated heteroduplex DNA is now allowed to reassemble in a PCR reaction without primers. Note that perfect base-pairing in the overlap is ensured since, DNaseI has cut double stranded regions of the heteroduplex.

SEQUENCE LISTING

<110> Danmarks Tekniske Universitet

<120> A method and a kit for genome-wide gene shuffling using tagged primer technology

<130> P 668 PC00

<160> 14

<170> PatentIn version 3.1

<210> 1

<211> 47

<212> DNA

<213> Synthetic sequence

<400> 1
gttcactagt taccCGggat gaattcaact cccgatctaa tatctcc

47

<210> 2

<211> 42

<212> DNA

<213> Synthetic sequence

<400> 2
gttcactagt taccCGggat gtctgaattc gctactagcc gc

42

<210> 3

<211> 43

<212> DNA

<213> Synthetic sequence

<400> 3
gttcactagt taccCGggat gtctgaagaa gctgcctatc aag

43

<210> 4

<211> 43

<212> DNA

<213> Synthetic sequence

<400> 4

gttcactagt tacccgggat gtcacaagac gctgctattg cag

43

<210> 5

<211> 44

<212> DNA

<213> Synthetic sequence

<400> 5

gttcactagt tacccgggat gaatacccag tataattcca gtta

44

<210> 6

<211> 21

<212> DNA

<213> Synthetic sequence

<400> 6

gttcactagt tacccgggat g

21

<210> 7

<211> 21

<212> DNA

<213> Synthetic sequence

<400> 7

gttcactagt tacccgggat g

21

<210> 8

<211> 47

<212> DNA

<213> Synthetic sequence

<400> 8

tcttgtagtc gacctcgagc gtttatttcc tgctaaacaa actcttg

47

<210> 9

<211> 40

<212> DNA

<213> Synthetic sequence

<400> 9

tcttgtagtc gacctcgagc ctcttattcc tcggaaactc

40

<210> 10

<211> 45

<212> DNA

<213> Synthetic sequence

<400> 10

tcttgtagtc gacctcgagc gcattagatc atcagcggtg tagtc

45

<210> 11

<211> 44

<212> DNA

<213> Synthetic sequence

<400> 11

tcttgtagtc gacctcgagc gtgttcgcaa attatttggt gctg

44

<210> 12

<211> 42

<212> DNA

<213> Synthetic sequence

<400> 12

tcttgtagtc gacctcgagc ttacagcgta gcagtttggt gt

42

<210> 13

<211> 20

<212> DNA

<213> Synthetic sequence

<400> 13
tcttgtagtc gacctcgagc

20

<210> 14

<211> 20

<212> DNA

<213> Synthetic sequence

<400> 14
tcttgtagtc gacctcgagc

20