POLYMERASE

The present invention relates to DNA polymerases. In particular the invention relates to a method for the generation of DNA polymerases which exhibit a relaxed substrate specificity. Uses of engineered polymerases produced using the methods of the invention are also described.

Background:

Accurate DNA replication is of fundamental importance to all life ensuring the maintenance and transmission of the genome and limiting tumorigenesis in higher organisms. High-fidelity DNA polymerases perform an astonishing feat of molecular recognition, incorporating the correct nucleotide triphosphate (dNTP) substrate molecules as specified by the template base with minimal error rates. For example, even without exonucleolytic proofreading, the replicative DNA polymerase III from E.coli on average only makes one error in ~10^5 base pairs (Schaaper JBC 1993).

As energetic differences between correctly and mispaired nucleotides *per se* are much too small to give rise to a 10^5 fold discrimination, the structure of the polymerase active site in high-fidelity polymerases has evolved to enhance those differences. Recent structural studies of the A-family (Pol I-like) DNA polymerases from *Thermus aquaticus* (Taq) (Li 98), phage T7 (Ellenberger) and *B. stearothermophilus* (Bst) (Beese) in particular have revealed how conformational changes during the catalytic cycle may exclude non-cognate base-pairing geometries because of steric clashes within the closed active site. As a result of these tight steric constraints, not only are mismatched nucleotides excluded but catalysis becomes exquisitely sensitive to even slight distortions in the primer-template duplex. This precludes or greatly diminishes the replication of modified or damaged DNA templates, the incorporation of modified or unnatural

deoxinucleotide triphosphates (dNTP) and the extension of mismatched or unnatural 3' termini.

While desirable in nature, such stringent substrate discrimination is limiting for many applications in biotechnology. Specifically, it restricts the use of unnatural or modified nucleotide bases and the applications they enable. It also precludes the efficient PCR amplification of damaged DNA templates.

Some other naturally occurring polymerases are less stringent with regard to their substrate specificity. For example, viral reverse transcriptases like HIV-1 reverse transcriptase or AMV reverse transcriptase and polymerases capable of translesion synthesis such as polY-family polymerases, pol X (Vaisman et al, 2001, JBC) or pol X (Washington (2002), PNAS; or the unusual polB-family polymerase pol X (Johnson, Nature), all extend 3' mismatches with elevated efficiency compared to high fidelity polymerases. The disadvantage of the use of translesion synthesis polymerases for biotechnological uses is that they depend on cellular processivity factors for their activity, such as PCNA. Moreover such polymerases are not stable at the temperatures at which certain biotechnological techniques are performed, such as PCR. Furthermore most Translesion synthesis polymerases have a much reduced fidelity, which would severely compromise their utility for cloning.

Using another approach, the availability of high-resolution structures has guided efforts to rationally alter the substrate specificity of high fidelity DNA polymerases by site-directed mutagenesis e.g. to increase acceptance of dideoxi- (ddNTPs) (Li 99) or ribonucleotides (rNTPs) (Astatke 98). In vivo complementation followed by screening has also yielded polymerase variants with increased rNTP incorporation and limited bypass of template lesions (Patel 01). Recently, two different *in vitro* strategies for selection of polymerase activity have been described (Jestin 00, Ghadessy 01, Xia 02). One is based on the proximal attachent of polymerase and template-primer duplex on the same phage particle and has allowed the isolation mutants of Taq polymerase, which incorporate rNTPs and dNTPs with comparable efficiency (Xia 02). However, such methods are complex, prone to error and are laborious.

Recently, the technique of compartmentalized self-replication (CSR) (Ghadessy 01), which is based on the self-replication of polymerase genes by the encoded polymerases within discrete, non-communicating compartments has allowed the selection of mutants of Taq polymerase with increased thermostability and/or resistance to the potent inhibitor heparin (Ghadessy et al 01).

However, there still remains a need in the art for an efficient and simple method for relaxing the substrate specificity of high fidelity DNA polymerases whilst maintaining high catalytic turnover and processivity of DNA fragments up to several tens of kb. Such polymerases will be of particular use in applications such as PCR amplification and sequencing of damaged DNA templates, for the incorporation of unnatural base analogues into DNA (such as is required for sequencing or array labelling) and as a starting point for the creation of novel polymerase activities using compartmentalised self replication or other methods.

Summary of the invention

The present inventors modified the principles of directed evolution, (in particular compartmentalised self replication) described in GB97143002, 986063936 and GB 01275643 in the name of the present inventors, to relax the steric control of high fidelity DNA polymerases and consequently to expand the substrate range of such polymerases. All of the documents listed above are herein incorporated by reference.

They surprisingly found that by performing the technique of compartmentalised self replication referenced above, using repertoires of randomly mutated Taq genes, and flanking primers bearing the mismatches A*G and C*C at their 3' terminus/end, then mutants were generated which not only exhibited the ability to extend the A*G and C*C tranversion mismatches used in the CSR selection, but also surprisingly exhibited a generic ability to extend mispaired 3' termini. This finding is especially significant since Taq polymerase is not able to extend 3' mismatches (Kwok wt al, (1990), Huang (1992).

The mutant polymerases generated also exhibit high catalytic turnover, concomitant with other high fidelity polymerases and are capable of efficient amplification of DNA fragments up to 26kb.

Thus in a first aspect the present invention provides a method for the generation of an engineered DNA polymerase with an expanded substrate range which comprises the step of preparing and expressing nucleic acid encoding an engineered DNA polymerase utilising template nucleic acid and flanking primers which bear one or more distorting 3' termini/ends.

As herein defined 'flanking primers which bear a 3' distorting terminus/end' refer to those primers which possess at their 3' ends one or more group/s, preferably nucleotide group/s which deviate from cognate base-pairing geometry. Such deviations from cognate base-pairing geometry includes but is not limited to: nucleotide mismatches, base lesions (i.e. modified or damaged bases) or entirely unnatural, synthetic base substitutes.. According to the above aspects of the invention, advantageously, the flanking primer/s bear one or more nucleotide mismatches at their 3' end/terminus.

Advantageously, according to the above aspects of the invention the flanking primers may have one, two, three, four, or five or more nucleotide mismatches at the 3' primer end. More advantageously, the one or more nucleotide mismatches are consecutive mismatches. More advantageously, according to the above aspects of the invention, the flanking primers have one or two nucleotide mismatches at the 3' primer end. Most preferably according to the above aspects of the invention, the flanking primers have one nucleotide mismatch at their 3' primer end.

More specificallythe term 'distorting 3' termini/ends' includes within its scopethe phenomenon f whereby, for example, either the 3' terminal base (1-mismatch) or the 3' terminal and upstream base (2-mismatch, 3-mismatch, 4-mismatch and so on) are not complementary to the template base. Preferably mismatches are transversion mismatches i.e. apposing purines with purines and pyrimidines with pyrimidines. Preferably

transversion mismatches are G.A and C.C. This type of primer terminus distortion is referred to herein as 'primer mismatch distortion'.

In addition, and as eluded to above, the term 'flanking primers bearing distorting 3' termini/ends' includes within its scope flanking primers bearing one or more unatural base analogues at the 3' termini/end of the one or more flanking primers so that distortion of the cognate DNA duplex geometry is created.

The method of the invention may be used to expand the substrate range of any DNA polymerase which lacks an intrinsic 3-5' exonuclease proofreading activity or where a 3-5' exonuclease proofreading activity has been disabled, e.g. through mutation . Suitable DNA polymerases include polA, polB (see e.g. Patrel & Loeb, Nature Struc Biol 2001) polC, polD, polY, polX and reverse transcriptases (RT) but preferably are processive, high-fidelity polymerases.

Advantageously, an engineered DNA polymerase with an expanded substrate range according to the invention is generated from a pol A-family DNA polymerase. Advantageously, the DNA polymerase is generated from a repertoire of pol A DNA polymerase nucleic acid as template nucleic acid. Preferably the pol A polymerase is Taq polymerase and the flanking primers used in the generation of the polymerase are one or more of those primers selected from the group consisting of the following: 5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GA-3';**A•G** mismatch); 5'GTA AAA CGA CGG CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GCC-3' C*C mismatch.

More advantageously, according to the above aspect of the invention, the nucleic acid encoding the engineered polymerase according to the invention is generated using PCR using one or more flanking primers listed herein.

Advantageously, the method of the present invention involves the use of compartmentalised self replication, and consists of the steps listed below:

(a) preparing nucleic acid encoding a engineered DNA polymerase, wherein the polymerase is generated using a repertoire of nucleic acid molecules encoding one or more DNA polymerases and flanking primers which bears a 3'distorting end.

(b) compartmentalising the nucleic acid of step (a) into microcapsules;

(c) expressing the nucleic acid to produce their respective DNA polymerase within the microcapsules;

(d) sorting the nucleic acid encoding the engineered DNA polymerase which exhibits an expanded substrate range; and

(e) expressing the engineered DNA polymerase which exhibits an expanded substrate range.

Most advantageously, the method of the invention comprises the use of one or more DNA polymerases and flanking primers which bears one or more nucleotide mismatches at their 3'primer ends.

According to the above aspects of the invention, the term 'engineered DNA polymerase' refers to a DNA polymerase which has a nucleic acid sequence which is not 100% identical at the nucleic acid level to the one or more DNA polymerase/s or fragments thereof, from which it is derived, and which is synthetic. According to the invention, an engineered DNA polymerase may belong to any family of DNA polymerase. Advantageously, an engineered DNA polymerase according to the invention is a pol A DNA polymerase. As referred to above the term 'engineered DNA polymerase' also includes within its scope fragments, derivatives and homologues of an 'engineered DNA polymerase' as herein defined so long as it exhibits the requisite property of possessing an expanded substrate range as defined herein. In addition, it is an essential feature of the present invention that an engineered DNA polymerase according to the invention does not include a polymerase with a 3-5' exonuclease activity under the conditions used for the polymerisation reaction. (This definition includes polymerases in which the 3-5' exonuclease is not part of the polymerase polypeptide chain but is associated non-covalently with the active polymerase). Such a proofreading activity would remove any 3' mismatches incorporated according to the

method of the invention, and thus would prevent a polymerase according to the invention possessing an expanded substrate range as defined herein.

As defined herein the term 'expanded substrate range' (of an engineered DNA polymerase) means that substrate range of an engineered DNA polymerase according to the present invention is broader than that of the one or more DNA polymerases, or fragments thereof from which it is derived. The term 'a broader substrate range' refers to the ability of an engineered polymerase according to the present invention to extend one or more 3'distorting ends, advantageously transversion mismatches (purine*purine, pyrimidine*pyrimidine) for example A*A, C*C, G*G, T*T and G*A, which the one or more polymerase/s from which it is derived cannot extend. That is, essentially, a DNA polymerase which exhibits a relaxed substrate range as herein defined has the ability not only to extend the 3' distorting endsused in its generation, IE those of the flanking primers) but also exhibits a generic ability to extend 3' distorting ends (for example A*G, A*A, G*G mismatches). Preferably, 'expanded substrate range' (of an engineered DNA polymerase) includes a wider spectrum of unnatural nucleotide substrates including αS dNTPs, dye-labelled nucleotides, damaged DNA templates and so on. More details are given in the Examples.

According to the above aspect of the invention advantageously the DNA polymerase generated using CSR technology is a pol A polymerase and it is generated using flanking primers selected from the group consisting of the following: 5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GA-3';**A•G** mismatch); 5'GTA AAA CGA CGG CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GC**C**-3' C*C mismatch.

One skilled in the art will appreciate that in essence, any DNA polymerase flanking primer which incorporates a 3' mismatch will work with any suitable repertoire. The process of mismatch extension will vary in characteristics from polymerase to polymerase, and will also vary according to the experimental conditions. For example,

G*A and C*C are the most disfavoured mismatches for extension by Taq polymerase (Huang et al, 92). Other mismatches are favoured for extension by other polymerases and this can be routinely determined by the skilled person.

One skilled in the art will also appreciate that it is an essential feature of the present invention that the methods described herein will only work for polymerases which are devoid of 3-5' exonuclease activity proofreading under the conditions used for the polymerisation reaction, as such activity would result in the removal of the incorporated mismatches.

Using the method of the invention, the present inventors generated a number of pol A polymerase mutants. Two of the mutants named M1 and M4 not only exhibit the ability to extend the G*A and C*C transversion mismatches used in the CSR selection, but also surprisingly exhibit a generically enhanced ability to extend 3' mismatched termini.

Thus in a further aspect the present invention provides an engineered DNA polymerase which exhibits an expanded substrate range. Preferably such an engineered polymerase is obtainable using one or more method/s of the present invention.

According to the above aspect of the invention, preferably the DNA polymerase is a pol A polymerase.

According to the above aspect of the invention, preferably the engineered DNA polymerase is obtained using the method of the invention.

In a further aspect still, the present invention provides a pol A DNA polymerase with an expanded substrate range, or the nucleic acid encoding it, wherein the DNA polymerase is designated M1 or M4 as shown in fig 1 and fig 2 respectively and depicted as SEQ No 1 and SEQ No 2 respectively.

According to the above aspect of the invention, preferably the engineered DNA polymerase as herein defined is that polymerase designated M1 in fig 1 and depicted SEQ No 1.

In yet a further aspect the invention provides a pol A DNA polymerase with an expanded substrate range, wherein the polymerase exhibits at least 95% identity to one or more of the amino acid sequences designated M1 and M4 as shown in fig 1 and fig 2 respectively and depicted SEQ No 1 and SEQ No 2 respectively and which comprises any one or more of the following mutations: E520G, D144G, L254P, E520G, E524G, N583S, 1.1-D144G, L254P, E520G, E524G, N583S, V113I, A129V, L245R, E315K, G364D, G403R, E432D, P481A, I614M, R704W, D144G, G370D, E742G, K56E, I63T, K127R, M317I, Q680R, R343G, G370D, E520G, G12A, A109T, D251E, P387L, A608V, R617K, D655E, T710N, E742G, A109T, D144G, V155I, P298L, G370D, I614M, E694K, R795G, E39K, R343G, G370D, E520G, T539A, M747V, K767R, G84A, D144G, K314R, E520G, F598L, A608V, E742G, D58G, R74P, A109T, L245R, R343G, G370D, E520G, N583S, E694K, A743P.

Advantageously, the invention provides a pol A DNA polymerase with an expanded substrate range, or the nucleic acid encoding it, wherein the polymerase exhibits at least 95% identity to one or more of the amino acid sequences designated M1 and M4 as shown in fig 1 and fig 2 respectively and depicted SEQ 1 and 2 respectively and which comprises any one or more of the following mutations: G84A, D144G, K314R, E520G, F598L, A608V, E742G, D58G, R74P, A109T, L245R, R343G, G370D, E520G, N583S, E694K, A743P.

Most advantageously, the invention provides a pol A DNA polymerase with an expanded substrate range, or the nucleic acid encoding it, wherein the polymerase exhibits at least 95% identity to one or more of the amino acid sequences designated M1 and M4 as shown in fig 1 and fig 2 respectively and depicted SEQ 1 and 2 respectively and which comprises any one or more of the following mutations: G84A, D144G, K314R, E520G, F598L, A608V, E742G.

According to the above aspect of the invention the mutation 'E520G' describes a DNA polymerase according to the invention in which glycine is present at position 520 of the amino acid sequence. The present inventors were surprised to find that E520, which is located at the tip of the thumb domain at a distance 20A from the 3'OH of the mismatched primer terminus, would be involved in mismatch recognition or extension . The mutation of E520 to G520 is clearly important in such roles however as the present inventors have demonstrated. This aspect of the invention is described further in the detailed description of the invention.

The present inventors consider that the method of the invention is applicable to the generation of 'blends' of engineered DNA polymerases with an expanded substrate range. According to the present invention the term a 'blend' of more than one polymerase refers to a mixture of 2 or more, 3 or more 4 or more, 5 or more enginerred polymerases. Preferably the term 'blends' refers to a mixture of 6, 7, 8, 9 or 10 or more 'engineered polymerases'.

It is important to note that the extension of mismatched 3' primer termini is a feature of naturally occurring polymerases. Viral reverse transcriptases (RT) like HIV-1 RT or AMV RT and polymerases capable of translesion synthesis (TLS) such as the polY-family polymerases pol ι (Vaisman 2001JBC) or pol κ (Washington 2002 PNAS) or the unusual polB-family polymerase polζ (Johnson Nature), all extend 3' mismatches with elevated efficiency compared to high-fidelity polymerases. Thus, the mutant polA polymerases according to the present invention share significant functional similarities with other polymerases found in nature but so far represent, the only known member of the polA-family polymerases that are proficient in mismatch extension (ME) and translesion synthesis (TLS).

In contrast to TLS polymerases, which are distributive and depend on cellular processivity factors such as PCNA, M1 and M4 combine mismatch extension (ME) and translesion synthesis (TLS) with high processivity and in the case of M1 are capable of efficient amplification of DNA fragments of up to 26kb.

In a further aspect still the present invention provides a nucleic acid construct which is capable of encoding a pol A DNA polymerase which exhibits an expanded substrate range, wherein said pol A DNA polymerase is depicted in fig 1 and fig 2 as SEQ No 1 or SEQ No 2 and is designated M1 and M4 respectively.

According to the above aspect of the invention, preferably the nucleic acid construct encodes the M1 pol A polymerase as described herein.

In a further aspects the invention provides a pol A DNA polymerase with an expanded substrate range, in particular which is capable of mismatch extension, wherein the DNA polymerase comprises, preferably consists of the amino acid sequence of any one or more of the clones designated herein as 3B5, 3B8, 3C12 and 3D1.

In yet a further aspect the invention provides a pol A DNA polymerase with an expanded substrate range, in particular which is capable of abasic site bypass, wherein the DNA polymerase comprises, preferably consists of the amino acid sequence of any one or more of the clones designated herein as 3A10, 3B6 and 3B11.

In a further aspect still the invention provides a pol A DNA polymerase with an expanded substrate range, in particular which is capable of DNA replication involving the incorporation of unatural base analogues into the newly replicated DNA, wherein the pol A DNA polymerase comprises, preferably consists of the amino acid sequence of any one or more of the clones designated herein as 4D11 and 5D4.

In a further aspect the present invention provides a pol A DNA polymerase with an expanded substrate range, wherein the polymerase exhibits at least **95%** identity to one or more of the amino acid sequences designated 3B5, 3B8, 3C12, 3D1, 3A10, 3B6, 3B11, 4D11 and 5D4. which comprises any one or more of the mutations (with respect to either of the three parent genes Taq, Tth, Tfl) or gene segments found in clones 3B5, 3B8, 3C12, 3D1, 3A10, 3B6, 3B11, 4D11 and 5D4.

In a further aspect still, the present invention provides a vector comprising a nucleic acid construct according to the present invention.

In a further aspect still the present invention provides the use of a DNA polymerase according to the present invention in any one or more of the following applications selected from the group consisting of the following: PCR amplification, sequencing of damaged DNA templates, the incorporation of unnatural base analogues into DNA and the creation of novel polymerase activities.

According to the above aspect of the invention, preferably the use is of a 'blend' of DNA polymerases according to the invention or selected according to the method of the invention. The use of blends of polymerases will be familiar to those skilled in the art and is described in Barnes, W. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 2216-2220 which is herein incorporated by reference.

According to the above aspect of the invention, preferably the DNA polymerase is a pol A DNA polymerase. Advantageously, it is generated using CSR technology using flanking primers bearing one or more 3' mismatch pairs of interest as described herein. Other suitable methods include screening after activity preselection (see Patel & Loeb 01) and phage display with proximity coupled template-primer duplex substrate (Jestin 01, Xue, 02. CST is also ideally suited as the present inventors have demonstrated.

According to the above aspect of the invention, preferably the use of a polymerase according to the invention is in PCR amplification and the polymerase is M1 as herein described.

According to the above aspect of the invention, advantageously, the creation of novel polymerase activities is produced using the technique of compartmentalised self replication as described herein.

Definitions

The term '**engineered DNA polymerase**' refers to a DNA polymerase which has a nucleic acid sequence which is not 100% identical at the nucleic acid level to the one or more DNA polymerase/s or fragments thereof, from which it is derived, and which has been generated using one or more biotechnological methods. Advantageously, an engineered DNA polymerase according to the invention is a pol-A family DNA polymerase or a pol-B family DNA polymerase. More advantageously, an engineered DNA polymerase according to the invention is a pol-A family DNA polymerase.As referred to above the term 'engineered DNA polymerase' also includes within its scope fragments, derivatives and homologues of an 'engineered DNA polymerase' as herein defined so long as it exhibits the requisite property of possessing an expanded substrate range as defined herein. In addition, it is an essential feature of the present invention that an engineered DNA polymerase according to the invention does not include a polymerase with a 3-5' exonuclease activity under the conditions used for the polymerisation reaction. Such a proofreading activity would remove any 3' mismatches incorporated according to the method of the invention, and thus would prevent a polymerase according to the invention possessing an expanded substrate range as defined herein.

As herein defined '**flanking primers which bear a 3' distorting terminus**' refer to those DNA polymerase primers which possess at their 3' ends one or more group/s, preferably nucleotide group/s which deviate from cognate base-pairing geometry. Such deviations from cognate base-pairing geometry includes but is not limited to: nucleotide mismatches, base lesions (i.e. modified or damaged bases) or entirely unnatural, synthetic base substitutes at the 3 end of a flanking primer used according to the methods of the invention. According to the above aspects of the invention, advantageously, the flanking primer/s bear one or more nucleotide mismatches at their 3' end. Advantageously, according to the above aspects of the invention the flanking primers may have one, two, three, four, or five or more nucleotide mismatches at the 3' primer end. Preferably according to the above aspects of the invention, the flanking primers have one or two nucleotide mismatches at the 3' primer end. Most preferably according to the above aspects of the invention, the flanking primers have one nucleotide mismatch at their 3' primer end.

As defined herein the term '**expanded substrate range**' (of an engineered DNA polymerase) means that substrate range of an engineered DNA polymerase according to the present invention is broader than that of the one or more DNA polymerases, or fragments thereof from which it is derived. The term 'a broader substrate range' refers to the ability of an engineered polymerase according to the present invention to extend one or more 3'distorting ends, advantageously transversion mismatches (purine*purine, pyrimidine*pyrimidine) for example A*A, C*C, G*G, T*T and G*A, which the one or more polymerase/s from which it is derived cannot extend. That is, essentially, a DNA polymerase which exhibits a relaxed substrate range as herein defined has the ability not only to extend the 3' distorting ends used in its generation, IE those of the flanking primers) but also exhibits a generic ability to extend 3' distorting ends(for example A*G, A*A, G*G mismatches).

Brief description of thefigures

Figure 1 shows the M1 nucleic acid (a) and amino acid sequence (b)

Figure 2 shows the M4 nucleic acid (a) amino acid sequence (b)

Figure 3 shows the general scheme of mismatch extension CSR selection. Self-replication of the *pol* gene by the encoded polymerase requires extension of flanking primers bearing G·A and C·C 3' mismatches. Polymerases capable of mismatch extension (Pol*) replicate their own encoding gene (*pol**), while Pol^{x} cannot extend mismatches and fails to self-replicate. Black bars denote incorporation of the mismatch into replication products.

Fig. 4. Mismatch extension properties of selected polymerases. (A) Polymerase activity in PCR for matched 3' ends and mismatches. Only mutant polymerases M4 and M1 (not shown) generate amplification products using primers with 3' transversion mismatches. (B) Mismatch extension PCR assay. Mismatch extension capability is expressed as arbitrary mismatch extension units (ratio of polymerase activity in PCR with matched vs. mismatched flanking primers). Different

polymerases (black diamonds) and derivatives (open squares, triangles) are shown in separate columns.

Fig. 5. Lesion bypass activity (A) wt*Taq*, (B) M1, (C) M4. Each polymerase was assayed over time for its ability to extend a radiolabeled primer annealed to either an undamaged template, or a template containing an abasic site or a *cis-syn* cyclobutane thymine-thymine dimer (CPD). Template sequence was identical except for three bases located immediately downstream of the primer (N1-3). The local sequence context in the N1-3 region is given on the right hand side of each respective panel. X= abasic site; T-T = CPD.

Fig. 6. Polymerase activity on unnatural substrates. (A) Polymerase activity in PCR using all αS dNTPs. αS DNA amplification products of 0.4kb, 0.8kb and 2kb, are obtained with M1 but not with wt*Taq* (wt). ϕX, HaeIII-digested phage ϕX174 DNA marker. λH, HindIII-digested phage λ DNA marker. (B) Polymerase activity in PCR with complete replacement of dATP with FITC-12-dATP (left) or dTTP with Biotin-16-dUTP (right). Only M1 yields amplification products. M, 1kb DNA ladder (Invitrogen). (C) Bypass of a 5-nitroindol template (5NI) base. Polymerase activity was assayed over time for its ability to extend a radiolabeled primer annealed to a template containing a 5NI template base.

Fig. 7. Long range PCR. PCR amplification of fragments of increasing length from a phage λ DNA template. Wt*Taq* (wt) fails to generate amplification products larger than 8.8kb while M1 is able to amplify fragments of > 25kb. λH, HindIII-digested phage λ DNA marker.

Fig. 8. Hairpin-ELISAs to test nucleotide analogue incorporation by mismatch extension clones.

Fig 9.. Clones 3B5. 3B8, 3C12 and 3D1 (where 3 indicates that these are third round clones) were able to extend primers containing four mismatches.

Fig 10. A list of polymerases selected to extend four mismatches were assayed for their ability to extend abasic sites in PCR.

Figure 11. Seven polymerases were assayed for their ability to bypass abasic sites in a primer extension assay.

Figure 12.. Several samples of cave hyena (*Crocuta spelaea)* were extracted and analysed.

Figure 13. Appropriate primers for use in the method of the invention. See example 15 for details.

Figure 14. Polymerases selected for replication of 5NI were tested for activity with a range of substrates using the hairpin ELISA assay described in example 8. See example 16 for details.

Figure 15. Polymerases selected for replication of 5NI were tested for activity with a range of substrates . Polymerase 4D11. P is primer, Ch is the chase reaction. Reaction times in minutes. See example 16 for details.

Figure 16. Polymerases selected for replication of 5NI were tested for activity with a range of substrates Polymerase 5D4. P is primer, Ch is the chase reaction. Reaction times in minutes. See example 16 for details.

Figure 17. Polymerases selected for replication of 5NI were tested for activity with a range of substrates Polymerase 4D11. P is primer, Ch is the chase reaction. Reaction times in minutes. See example 16 for details.

Figure 18.. Polymerases selected for replication of 5NI were tested for activity with a range of substrates Polymerase 5D4. P is primer, Ch is the chase reaction. Reaction times in minutes. See example 16 for details.

Figure 19. Microarray hybridisations of FITC-labelled probes. Microarrays contained 5 replicate features of serial dilutions of *Taq*, RT and genomic salmon sperm DNA target sequences, as indicated. Labelled randomers were used to visualise the microarray and assess the availability of target sequences for hybridisation. Array co-hybridisations were performed with a Cy5-labelled Taq probe ($Cy5_{Taq}$), as a reference, and equivalent unlabelled or FITC-labelled probes ($FITC10_{Taq}$, $FITC10_{M1}$, $FITC100_{M1}$). Single examples from 3 replicate experiments are displayed for each co-hybridisation.

Figure 20, Figure 21. Microarray signals from FITC-labelled probes. Mean FITC fluorescence signal of FITC-labelled probes ($FITC10_{Taq}$, $FITC10_{M1}$, $FITC100_{M1}$) for each co-hybridisation is plotted against the Cy5 fluorescence signal of the reference probe ($Cy5_{Taq}$) for A) Taq, B) RT and C) genomic salmon sperm DNA target sequences, as indicated. D) Microarray background signals from FITC-labelled probes are determined using 3 replicate microarrays for each co-hybridisation experiment of a Cy5-labelled Taq probe ($Cy5_{Taq}$), as a reference, and unlabeled or FITC-labelled probes ($FITC10_{Taq}$, $FITC10_{M1}$, $FITC100_{M1}$). Background information was generated by measuring fluorescence signal from 12 non-feature areas of each microarray. Mean pixel intensities were generated and used to derive a ratiometric value for each non-feature area. A mean of the mean ratio +/- 1 standard deviation is displayed for each co-hybridisation experiment.

Figure 22. Fidelity. (A) MutS ELISA. Relative replication fidelity of wt*Taq*, *M1* and *M4* was determined using mutS ELISA of two different DNA fragments (either a 0.4kb or 2.5kb region of the cloned *Taq* gene) obtained by PCR and probed at two different concentrations. (B) Spectra of nucleotide substitutions observed in PCR fragments amplified with either wt*Taq* or *M1*. Types of substitutions are given as % of total substitutions (wt*Taq*: 48, *M1*: 74). Equivalent substitutions on either strand (e.g. G->A, C->T) were added together (GC->AT). Observed –1 delections (wt*Taq*: 3, *M1*: 1) are not shown.

Figure 23. Processivity of wt*Taq*, *M1* and *M4* was measured at three different polymerase concentrations in the absence (A) or presence (B) of trap DNA. The processivity for nucleotide incorporation at each position was variable but essentially identical for all three polymerases. For example, the probability of enzyme dissociation is higher at positions 2-5 compared to positions 6 and 7 for all three polymerases. In the presence of trap DNA (to ensure all primer extension is the result of a single DNA binding event) 13% of bound wtTaq, 28% of M1 and 15% of M4 extended primers to the end of the template. The termination probabilities for positions 2 through 5 varied from 15-25% for wt*Taq* and *M1* and from 13-35% for *M4*, while at positions 6 and 7 the termination proabability was 5% for wt*Taq*, 1% for *M1*, and 2-4% for *M4*. DNA replication has been characterized as low processive when the termination probability reaches 40-80%[15]. Our results suggest that *M1* and *M4* are both processive polymerases, with processivity equal or higher than wt*Taq*,

arguing against a mechanistic interdependence of low processivity and translesion synthesis.

Detailed description of the invention

(A) Principles underlying CST technology according to the invention.

In a preferred embodiment the present invention provides a method for the generation of an engineered DNA polymerase with an expanded substrate range which comprises the steps of:

(a) preparing nucleic acid encoding a mutant DNA polymerase, wherein the polymerase is generated using flanking primers which bear a 3' distorting end

(b) compartmentalising the nucleic acid of step (a) into microcapsules;

(c) expressing the nucleic acid to produce their respective DNA polymerase within the microcapsules;

(d) sorting the nucleic acid encoding the mutant DNA polymerase which exhibits an expanded substrate range; and

(e) expressing the mutant DNA polymerase which exhibits an expanded substrate range.

The techniques of directed evolution and compartmentalised self replication are detailed in GB 97143002 and GB 98063936 and GB 01275643, in the name of the present inventors. These documents are herein incorporated by reference.

The inventors modified the methods of compartmentalised self replication and surprisingly generated DNA polymerases which exhibited an expanded substrate range as herein defined.

In particular, the inventors realised that for self-replication of Taq polymerase, compartments must remain stable at the high temperatures of PCR thermocycling. Encapsulation of PCRs has been described previously for lipid vesicles (Oberholzer, T.,

Albrizio, M. & Luisi, P. L. (1995) *Chem. Biol.* **2,** 677-82 and fixed cells and tissues (Haase, A. T., Retzel, E. F. & Staskus, K. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4971-5; Embleton, M. J., Gorochov, G., Jones, P. T. & Winter, G. (1992) *Nucleic Acids*) but with low efficiencies.

The present inventors used recently developed oil in water emulsions but modified the composition of the surfactant as well as the oil to water ratio. Details are given in Example 1. These modifications greatly increased the heat stability of the compartments and allowed PCR yields in the emulsion to approach those of PCR in solution. Further details of the method of compartmentalised self replication are given below.

Microcapsules

The microcapsules used according to the method of the invention require appropriate physical properties to allow the working of the invention.

First, to ensure that the nucleic acids and gene products may not diffuse between microcapsules, the contents of each microcapsule must be isolated from the contents of the surrounding microcapsules, so that there is no or little exchange of the nucleic acids and gene products between the microcapsules over the timescale of the experiment.

Second, the method of the present invention requires that there are only a limited number of nucleic acids per microcapsule. This ensures that the gene product of an individual nucleic acid will be isolated from other nucleic acids. Thus, coupling between nucleic acid and gene product will be highly specific. The enrichment factor is greatest with on average one or fewer nucleic acids per microcapsule, the linkage between nucleic acid and the activity of the encoded gene product being as tight as is possible, since the gene product of an individual nucleic acid will be isolated from the products of all other nucleic acids. However, even if the theoretically optimal situation of, on average, a single nucleic acid or less per microcapsule is not used, a ratio of 5, 10, 50, 100 or 1000 or more nucleic acids per microcapsule may prove beneficial in sorting a large library. Subsequent rounds of sorting, including renewed encapsulation with

differing nucleic acid distribution, will permit more stringent sorting of the nucleic acids. Preferably, there is a single nucleic acid, or fewer, per microcapsule.

Third, the formation and the composition of the microcapsules must not abolish the function of the machinery the expression of the nucleic acids and the activity of the gene products.

Consequently, any microencapsulation system used must fulfil these three requirements. The appropriate system(s) may vary depending on the precise nature of the requirements in each application of the invention, as will be apparent to the skilled person.

A wide variety of microencapsulation procedures are available (see Benita, 1996) and may be used to create the microcapsules used in accordance with the present invention. Indeed, more than 200 microencapsulation methods have been identified in the literature (Finch, 1993).

These include membrane enveloped aqueous vesicles such as lipid vesicles (liposomes) (New, 1990) and non-ionic surfactant vesicles (van Hal *et al.*, 1996). These are closed-membranous capsules of single or multiple bilayers of non-covalently assembled molecules, with each bilayer separated from its neighbour by an aqueous compartment. In the case of liposomes the membrane is composed of lipid molecules; these are usually phospholipids but sterols such as cholesterol may also be incorporated into the membranes (New, 1990). A variety of enzyme-catalysed biochemical reactions, including RNA and DNA polymerisation, can be performed within liposomes (Chakrabarti *et al.*, 1994; Oberholzer *et al.*, 1995a; Oberholzer *et al.*, 1995b; Walde *et al.*, 1994; Wick & Luisi, 1996).

With a membrane-enveloped vesicle system much of the aqueous phase is outside the vesicles and is therefore non-compartmentalised. This continuous, aqueous phase should be removed or the biological systems in it inhibited or destroyed (for example, by digestion of nucleic acids with DNase or RNase) in order that the reactions are limited to the microcapsules (Luisi *et al.*, 1987).

Enzyme-catalysed biochemical reactions have also been demonstrated in microcapsules generated by a variety of other methods. Many enzymes are active in reverse micellar solutions (Bru & Walde, 1991; Bru & Walde, 1993; Creagh *et al.*, 1993; Haber *et al.*, 1993; Kumar *et al.*, 1989; Luisi & B., 1987; Mao & Walde, 1991; Mao *et al.*, 1992; Perez *et al.*, 1992; Walde et al., 1994; Walde *et al.*, 1993; Walde *et al.*, 1988) such as the AOT-isooctane-water system (Menger & Yamada, 1979).

Microcapsules can also be generated by interfacial polymerisation and interfacial complexation (Whateley, 1996). Microcapsules of this sort can have rigid, nonpermeable membranes, or semipermeable membranes. Semipermeable microcapsules bordered by cellulose nitrate membranes, polyamide membranes and lipid-polyamide membranes can all support biochemical reactions, including multienzyme systems (Chang, 1987; Chang, 1992; Lim, 1984). Alginate/polylysine microcapsules (Lim & Sun, 1980), which can be formed under very mild conditions, have also proven to be very biocompatible, providing, for example, an effective method of encapsulating living cells and tissues (Chang, 1992; Sun *et al.*, 1992).

Non-membranous microencapsulation systems based on phase partitioning of an aqueous environment in a colloidal system, such as an emulsion, may also be used.

Preferably, the microcapsules of the present invention are formed from emulsions; heterogeneous systems of two immiscible liquid phases with one of the phases dispersed in the other as droplets of microscopic or colloidal size (Becher, 1957; Sherman, 1968; Lissant, 1974; Lissant, 1984).

Emulsions

Emulsions may be produced from any suitable combination of immiscible liquids. Preferably the emulsion of the present invention has water (containing the biochemical components) as the phase present in the form of finely divided droplets (the disperse, internal or discontinuous phase) and a hydrophobic, immiscible liquid (an 'oil') as the

matrix in which these droplets are suspended (the nondisperse, continuous or external phase). Such emulsions are termed 'water-in-oil' (W/O). This has the advantage that the entire aqueous phase containing the biochemical components is compartmentalised in discreet droplets (the internal phase). The external phase, being a hydrophobic oil, generally contains none of the biochemical components and hence is inert.

The emulsion may be stabilised by addition of one or more surface-active agents (surfactants). These surfactants are termed emulsifying agents and act at the water/oil interface to prevent (or at least delay) separation of the phases. Many oils and many emulsifiers can be used for the generation of water-in-oil emulsions; a recent compilation listed over 16,000 surfactants, many of which are used as emulsifying agents (Ash and Ash, 1993). Suitable oils include light white mineral oil and non-ionic surfactants (Schick, 1966) such as sorbitan monooleate (Span™80; ICI) and polyoxyethylenesorbitan monooleate (Tween™ 80; ICI) and Triton-X-100..

The use of anionic surfactants may also be beneficial. Suitable surfactants include sodium cholate and sodium taurocholate. Particularly preferred is sodium deoxycholate, preferably at a concentration of 0.5% w/v, or below. Inclusion of such surfactants can in some cases increase the expression of the nucleic acids and/or the activity of the gene products. Addition of some anionic surfactants to a non-emulsified reaction mixture completely abolishes translation. During emulsification, however, the surfactant is transferred from the aqueous phase into the interface and activity is restored. Addition of an anionic surfactant to the mixtures to be emulsified ensures that reactions proceed only after compartmentalisation.

Creation of an emulsion generally requires the application of mechanical energy to force the phases together. There are a variety of ways of doing this which utilise a variety of mechanical devices, including stirrers (such as magnetic stir-bars, propeller and turbine stirrers, paddle devices and whisks), homogenisers (including rotor-stator homogenisers, high-pressure valve homogenisers and jet homogenisers), colloid mills, ultrasound and 'membrane emulsification' devices (Becher, 1957; Dickinson, 1994).

Aqueous microcapsules formed in water-in-oil emulsions are generally stable with little if any exchange of nucleic acids or gene products between microcapsules. Additionally, we have demonstrated that several biochemical reactions proceed in emulsion microcapsules. Moreover, complicated biochemical processes, notably gene transcription and translation are also active in emulsion microcapsules. The technology exists to create emulsions with volumes all the way up to industrial scales of thousands of litres (Becher, 1957; Sherman, 1968; Lissant, 1974; Lissant, 1984).

The preferred microcapsule size will vary depending upon the precise requirements of any individual selection process that is to be performed according to the present invention. In all cases, there will be an optimal balance between gene library size, the required enrichment and the required concentration of components in the individual microcapsules to achieve efficient expression and reactivity of the gene products.

Details of one example of an emulsion used when performing the method of the present invention are given in Example 1.

Expression within microcapsules

The processes of expression must occur within each individual microcapsule provided by the present invention. Both *in vitro* transcription and coupled transcription-translation become less efficient at sub-nanomolar DNA concentrations. Because of the requirement for only a limited number of DNA molecules to be present in each microcapsule, this therefore sets a practical upper limit on the possible microcapsule size. Preferably, the mean volume of the microcapsules is less that 5.2×10^{-16} m^3, (corresponding to a spherical microcapsule of diameter less than 10μm, more preferably less than 6.5×10^{-17} m^3 (5μm), more preferably about 4.2×10^{-18} m^3 (2μm) and ideally about 9×10^{-18} m^3 (2.6μm).

The effective DNA or RNA concentration in the microcapsules may be artificially increased by various methods that will be well-known to those versed in the art. These include, for example, the addition of volume excluding chemicals such as polyethylene

glycols (PEG) and a variety of gene amplification techniques, including transcription using RNA polymerases including those from bacteria such as *E. coli* (Roberts, 1969; Blattner and Dahlberg, 1972; Roberts *et al.*, 1975; Rosenberg *et al.* , 1975) , eukaryotes e. g. (Weil *et al.* , 1979; Manley *et al.*, 1983) and bacteriophage such as T7, T3 and SP6 (Melton et al., 1984); the polymerase chain reaction (PCR) (Saiki *et al.*, 1988); Qβ replicase amplification (Miele *et al.*, 1983; Cahill *et al.*, 1991; Chetverin and Spirin, 1995; Katanaev *et al.*, 1995); the ligase chain reaction (LCR) (Landegren *et al.*, 1988; Barany, 1991); and self-sustained sequence replication system (Fahy *et al.*, 1991) and strand displacement amplification (Walker *et al.*, 1992). Even gene amplification techniques requiring thermal cycling such as PCR and LCR could be used if the emulsions and the *in vitro* transcription or coupled transcription-translation systems are thermostable (for example, the coupled transcription-translation systems could be made from a thermostable organism such as *Thermus aquaticus).*

Increasing the effective local nucleic acid concentration enables larger microcapsules to be used effectively. This allows a preferred practical upper limit to the microcapsule volume of about 5.2×10^{-16} m^3 (corresponding to a sphere of diameter 10um).

The microcapsule size must be sufficiently large to accommodate all of the required components of the biochemical reactions that are needed to occur within the microcapsule. For example, *in vitro*, both transcription reactions and coupled transcription-translation reactions require a total nucleoside triphosphate concentration of about 2mM.

For example, in order to transcribe a gene to a single short RNA molecule of 500 bases in length, this would require a minimum of 500 molecules of nucleoside triphosphate per microcapsule (8.33×10^{-22} moles). In order to constitute a 2mM solution, this number of molecules must be contained within a microcapsule of volume 4.17×10^{-19} litres (4.17×10^{-22} m^3 which if spherical would have a diameter of 93nm.

Furthermore, particularly in the case of reactions involving translation, it is to be noted that the ribosomes necessary for the translation to occur are themselves approximately

20nm in diameter. Hence, the preferred lower limit for microcapsules is a diameter of approximately 100nm.

Therefore, the microcapsule volume is preferably of the order of between 5.2 x 10^{-22} m^3 and 5.2 x 10^{-16} m^3 corresponding to a sphere of diameter between 0.1um and 10um, more preferably of between about 5.2 x 10^{-19} m^3 and 6.5 x 10^{-17} m^3 (1um and 5um). Sphere diameters of about 2.6um are most advantageous.

It is no coincidence that the preferred dimensions of the compartments (droplets of 2.6um mean diameter) closely resemble those of bacteria, for example, *Escherichia* are 1.1-1.5 x 2.0-6.0 um rods and *Azotobacter* are 1.5-2.0 um diameter ovoid cells. In its simplest form, Darwinian evolution is based on a 'one genotype one phenotype' mechanism. The concentration of a single compartmentalised gene, or genome, drops from 0.4 nM in a compartment of 2 um diameter, to 25 pM in a compartment of 5 um diameter. The prokaryotic transcription/translation machinery has evolved to operate in compartments of ~1-2 um diameter, where single genes are at approximately nanomolar concentrations. A single gene, in a compartment of 2.6 um diameter is at a concentration of 0.2 nM. This gene concentration is high enough for efficient translation. Compartmentalisation in such a volume also ensures that even if only a single molecule of the gene product is formed it is present at about 0.2 nM, which is important if the gene product is to have a modifying activity of the nucleic acid itself. The volume of the microcapsule should thus be selected bearing in mind not only the requirements for transcription and translation of the nucleic acid/nucleic acid, but also the modifying activity required of the gene product in the method of the invention.

The size of emulsion microcapsules may be varied simply by tailoring the emulsion conditions used to form the emulsion according to requirements of the selection system. The larger the microcapsule size, the larger is the volume that will be required to encapsulate a given nucleic acid/nucleic acid library, since the ultimately limiting factor will be the size of the microcapsule and thus the number of microcapsules possible per unit volume.

The size of the microcapsules is selected not only having regard to the requirements of the transcription/translation system, but also those of the selection system employed for the nucleic acid/nucleic acid construct. Thus, the components of the selection system, such as a chemical modification system, may require reaction volumes and/or reagent concentrations which are not optimal for transcription/translation. As set forth herein, such requirements may be accommodated by a secondary re-encapsulation step; moreover, they may be accommodated by selecting the microcapsule size in order to maximise transcription/translation and selection as a whole. Empirical determination of optimal microcapsule volume and reagent concentration, for example as set forth herein, is preferred.

A "nucleic acid/nucleic acid" in accordance with the present invention is as described above. Preferably, a nucleic acid is a molecule or construct selected from the group consisting of a DNA molecule, an RNA molecule, a partially or wholly artificial nucleic acid molecule consisting of exclusively synthetic or a mixture of naturally-occurring and synthetic bases, any one of the foregoing linked to a polypeptide, and any one of the foregoing linked to any other molecular group or construct. Advantageously, the other molecular group or construct may be selected from the group consisting of nucleic acids, polymeric substances, particularly beads, for example polystyrene beads, magnetic substances such as magnetic beads, labels, such as fluorophores or isotopic labels, chemical reagents, binding agents such as macrocycles and the like.

The nucleic acid portion of the nucleic acid may comprise suitable regulatory sequences, such as those required for efficient expression of the gene product, for example promoters, enhancers, translational initiation sequences, polyadenylation sequences, splice sites and the like.

Product selection

Details of a preferred method of performing the method of the invention are given in Example 1. However, those skilled in the art will appreciate that the examples given are

non-limiting and methods for product selection are discussed in more general terms below.

A ligand or substrate can be connected to the nucleic acid by a variety of means that will be apparent to those skilled in the art (see, for example, Hermanson, 1996). Any tag will suffice that allows for the subsequent selection of the nucleic acid. Sorting can be by any method which allows the preferential separation, amplification or survival of the tagged nucleic acid. Examples include selection by binding (including techniques based on magnetic separation, for example using Dynabeads™), and by resistance to degradation (for example by nucleases, including restriction endonucleases).

One way in which the nucleic acid molecule may be linked to a ligand or substrate is through biotinylation. This can be done by PCR amplification with a 5'-biotinylation primer such that the biotin and nucleic acid are covalently linked.

The ligand or substrate to be selected can be attached to the modified nucleic acid by a variety of means that will be apparent to those of skill in the art. A biotinylated nucleic acid may be coupled to a polystyrene microbead (0.035 to 0.2um in diameter) that is coated with avidin or streptavidin, that will therefore bind the nucleic acid with very high affinity. This bead can be derivatised with substrate or ligand by any suitable method such as by adding biotinylated substrate or by covalent coupling.

Alternatively, a biotinylated nucleic acid may be coupled to avidin or streptavidin complexed to a large protein molecule such as thyroglobulin (669 Kd) or ferritin (440 Kd). This complex can be derivatised with substrate or ligand, for example by covalent coupling to the alpha-amino group of lysines or through a non-covalent interaction such as biotin-avidin. The substrate may be present in a form unlinked to the nucleic acid but containing an inactive "tag" that requires a further step to activate it such as photoactivation (e.g. of a "caged" biotin analogue, (Sundberg *et al.*, 1995; Pirrung and Huang, 1996)). The catalyst to be selected then converts the substrate to product. The "tag" could then be activated and the "tagged" substrate and/or product bound by a tag-binding molecule (e.g. avidin or streptavidin) complexed with the nucleic acid. The

ratio of substrate to product attached to the nucleic acid via the "tag" will therefore reflect the ratio of the substrate and product in solution.

When all reactions are stopped and the microcapsules are combined, the nucleic acids encoding active enzymes can be enriched using an antibody or other molecule which binds, or reacts specifically with the "tag". Although both substrates and product have the molecular tag, only the nucleic acids encoding active gene product will co-purify.

The terms "isolating", "sorting" and "selecting", as well as variations thereof, are used herein. Isolation, according to the present invention, refers to the process of separating an entity from a heterogeneous population, for example a mixture, such that it is free of at least one substance with which it was associated before the isolation process. In a preferred embodiment, isolation refers to purification of an entity essentially to homogeneity. Sorting of an entity refers to the process of preferentially isolating desired entities over undesired entities. In as far as this relates to isolation of the desired entities, the terms "isolating" and "sorting" are equivalent. The method of the present invention permits the sorting of desired nucleic acids from pools (libraries or repertoires) of nucleic acids which contain the desired nucleic acid. Selecting is used to refer to the process (including the sorting process) of isolating an entity according to a particular property thereof.

Initial selection of a nucleic acid/nucleic acid from a nucleic acid library (for example a mutant taq library) using the present invention will in most cases require the screening of a large number of variant nucleic acids. Libraries of nucleic acids can be created in a variety of different ways, including the following.

Pools of naturally occurring nucleic acids can be cloned from genomic DNA or cDNA (Sambrook *et al.*, 1989); for example, mutant Taq libraries or other DNA polymerase libraries, made by PCR amplification repertoires of taq or other DNA polymerase genes have proved very effective sources of DNA polymerase fragments. Further details are given in the examples.

Libraries of genes can also be made by encoding all (see for example Smith, 1985; Parmley and Smith, 1988) or part of genes (see for example Lowman *et al.*, 1991) or pools of genes (see for example Nissim *et al.*, 1994) by a randomised or doped synthetic oligonucleotide. Libraries can also be made by introducing mutations into a nucleic acid or pool of nucleic acids 'randomly' by a variety of techniques *in vivo,* including; using 'mutator strains', of bacteria such as *E. coli mutD5* (Liao *et al.*, 1986; Yamagishi *et al.*, 1990; Low *et al.*, 1996). Random mutations can also be introduced both *in vivo* and *in vitro* by chemical mutagens, and ionising or UV irradiation (see Friedberg *et al.*, 1995), or incorporation of mutagenic base analogues (Freese, 1959; Zaccolo *et al.*, 1996). 'Random' mutations can also be introduced into genes in vitro during polymerisation for example by using error-prone polymerases (Leung *et al.*, 1989). In a preferred embodiment of the method of the invention, the repertoire of nucleic fragments used is a mutant Taq repertoire which has been mutated using error prone PCR. Details are given in Examples 1. According to the method of the invention, the term 'random' may be in terms of random positions with random repertoire of amino acids at those positions or it may be selected (predetermined) positions with random repertoire of amino acids at those selected positions.

Further diversification can be introduced by using homologous recombination either *in vivo* (see Kowalczykowski *et al.*, 1994 or *in vitro* (Stemmer, 1994a; Stemmer, 1994b)).

Microcapsules/sorting

In addition to the nucleic acids described above, the microcapsules according to the invention will comprise further components required for the sorting process to take place. Other components of the system will for example comprise those necessary for transcription and/or translation of the nucleic acid. These are selected for the requirements of a specific system from the following; a suitable buffer, an *in vitro* transcription/replication system and/or an *in vitro* translation system containing all the necessary ingredients, enzymes and cofactors, RNA polymerase, nucleotides, nucleic acids (natural or synthetic), transfer RNAs, ribosomes and amino acids, and the

substrates of the reaction of interest in order to allow selection of the modified gene product.

A suitable buffer will be one in which all of the desired components of the biological system are active and will therefore depend upon the requirements of each specific reaction system. Buffers suitable for biological and/or chemical reactions are known in the art and recipes provided in various laboratory texts, such as Sambrook *et al.*, 1989.

The *in vitro* translation system will usually comprise a cell extract, typically from bacteria (Zubay, 1973; Zubay, 1980; Lesley *et al.*, 1991; Lesley, 1995), rabbit reticulocytes (Pelham and Jackson, 1976), or wheat germ (Anderson *et al.*, 1983). Many suitable systems are commercially available (for example from Promega) including some which will allow coupled transcription/translation (all the bacterial systems and the reticulocyte and wheat germ TNT™ extract systems from Promega). The mixture of amino acids used may include synthetic amino acids if desired, to increase the possible number or variety of proteins produced in the library. This can be accomplished by charging tRNAs with artificial amino acids and using these tRNAs for the *in vitro* translation of the proteins to be selected (Ellman *et al.*, 1991; Benner, 1994; Mendel *et al.*, 1995).

After each round of selection the enrichment of the pool of nucleic acids for those encoding the molecules of interest can be assayed by non-compartmentalised *in vitro* transcription/replication or coupled transcription-translation reactions. The selected pool is cloned into a suitable plasmid vector and RNA or recombinant protein is produced from the individual clones for further purification and assay.

Microcapsule identification

Microcapsules may be identified by virtue of a change induced by the desired gene product which either occurs or manifests itself at the surface of the microcapsule or is detectable from the outside as described in section iii (Microcapsule Sorting). This change, when identified, is used to trigger the modification of the gene within the

compartment. In a preferred aspect of the invention, microcapsule identification relies on a change in the optical properties of the microcapsule resulting from a reaction leading to luminescence, phosphorescence or fluorescence within the microcapsule. Modification of the gene within the microcapsules would be triggered by identification of luminescence, phosphorescence or fluorescence. For example, identification of luminescence, phosphorescence or fluorescence can trigger bombardment of the compartment with photons (or other particles or waves) which leads to modification of the nucleic acid. A similar procedure has been described previously for the rapid sorting of cells (Keij *et al.*, 1994). Modification of the nucleic acid may result, for example, from coupling a molecular "tag", caged by a photolabile protecting group to the nucleic acids: bombardment with photons of an appropriate wavelength leads to the removal of the cage. Afterwards, all microcapsules are combined and the nucleic acids pooled together in one environment. Nucleic acids encoding gene products exhibiting the desired activity can be selected by affinity purification using a molecule that specifically binds to, or reacts specifically with, the "tag".

Multi step procedure

It will be also be appreciated that according to the present invention, it is not necessary for all the processes of transcription/replication and/or translation, and selection to proceed in one single step, with all reactions taking place in one microcapsule. The selection procedure may comprise two or more steps. First, transcription/replication and/or translation of each nucleic acid of a nucleic acid library may take place in a first microcapsule. Each gene product is then linked to the nucleic acid which encoded it (which resides in the same microcapsule). The microcapsules are then broken, and the nucleic acids attached to their respective gene products optionally purified. Alternatively, nucleic acids can be attached to their respective gene products using methods which do not rely on encapsulation. For example phage display (Smith, G.P.,1985), polysome display (Mattheakkis *et al.*, 1994), RNA-peptide fusion (Roberts and Szostak, 1997) or lac repressor peptide fusion (Cull, *et al.*, 1992).

In the second step of the procedure, each purified nucleic acid attached to its gene product is put into a second microcapsule containing components of the reaction to be selected. This reaction is then initiated. After completion of the reactions, the microcapsules are again broken and the modified nucleic acids are selected. In the case of complicated multistep reactions in which many individual components and reaction steps are involved, one or more intervening steps may be performed between the initial step of creation and linking of gene product to nucleic acid, and the final step of generating the selectable change in the nucleic acid.

Amplification

In all the above configurations, genetic material comprised in the nucleic acids may be amplified and the process repeated in iterative steps. Amplification may be by the polymerase chain reaction (Saiki *et al.*, 1988) or by using one of a variety of other gene amplification techniques including; Qβ replicase amplification (Cahill, Foster and Mahan, 1991; Chetverin and Spirin, 1995; Katanaev, Kurnasov and Spirin, 1995); the ligase chain reaction (LCR) (Landegren *et al.*, 1988; Barany, 1991); the self-sustained sequence replication system (Fahy, Kwoh and Gingeras, 1991) and strand displacement amplification (Walker *et al.*, 1992).

(B) DNA polymerases according to the invention.

(i) General

High fidelity DNA polymerases such as Pol A(like Taq polymerase) and Pol-B family polymerases which lack a 3'-5' exonuclease proofreading capability show a strict blockage to the extension of distorted or mismatched 3' primer termini to avoid propagation of misincorporations. While the degree of blockage varies considerably depending on the nature of the mismatch, some transversion (purine·purine / pyrimidine·pyrimidine) mismatches are extended up to 10^6-fold less efficiently than

matched termini (Huang 92). Likewise, many unnatural base analogues, while incorporated efficiently, act as strong terminators (Kool, Loakes).

The present inventors have modified the principles described in Ghadessy, F. G et al (2001) Proc. Nat. Acad. Sci, USA, 93, 4552-4557 (compartmentalised self replication) and Ghadessy 2003, and outlined above. Both these documents are herein incorporated by reference. The present inventors have used these modified techniques to develop a method by which the substrates specificity of high fidelity DNA polymerases may be expanded in a generic way.

The inventors have exemplified the technique by expanding the substrate specificity of the high-fidelity pol-A family polymerases. In particular, the present inventors created two repertoires of randomly mutated Taq genes, as described in Ghadessy, F. G et al (2001) referred to above. Three cycles of mismatch extension CSR was performed using flanking primers bearing the mismatches A*G and C*C at their 3' ends. Selected clones were ranked using a PCR extension assay described herein.

Selected mutants exhibited the ability to extend the G*A and C*C tranversion mismatches used in the CSR selection, but also exhibited a generic ability to extend mispaired 3' termini. These results are surprising, especially since Taq polymerase is unable to extend such mismatches (Kwok et al, (1990); Huang (1992).

Thus, using this approach, the inventors have generated DNA polymerases which exhibit a relaxed substrate specificity/expanded substrate range.

According to the present invention, the term '**expanded substrate range**' (of an engineered DNA polymerase) means that substrate range of an engineered DNA polymerase according to the present invention is broader than that of the one or more DNA polymerases, or fragments thereof from which it is derived. The term 'a broader substrate range' refers to the ability of an engineered polymerase according to the present invention to extend one or more 3' mismatches, for example A*A, G*A, G*G, T*T, C*C, which the one or more polymerase/s from which it is derived cannot extend. That

is, essentially, a DNA polymerase which exhibits a relaxed substrate range as herein defined has the ability not only to extend the 3' mismatches used in its generation, (IE those of the flanking primers), but also exhibits a generic ability to extend 3' mismatches (for example A*G, A*A, G*G).

The two best mutants M1 (G84A, D144G, K314R, E520G, F598L, A608V, E742G) and M4 (D58G, R74P, A109T, L245R, R343G, G370D, E520G, N583S, E694K, A743P) were chosen for further investigation.

M1 and M4 not only had greatly increased ability to extend the G•A and C•C transversion mismatches used in the CSR selection, but appeared to have acquired a more generic ability to extend 3' mispaired termini, including other strongly disfavoured transversion mismatches (such as A•G, A•A, G•G) (Fig. 1B), which wtTaq polymerase was unable to extend, as previously reported (Kwok et al 1990, Huang 92).

(ii) M1 and M4 mutants according to the invention.

Nucleic acid sequences encoding M1 and M4 pol A DNA polymerase mutants are depicted SEQ No 1 and SEQ No 2 respectively and are shown in Fig 1 and fig 2 respectively.

Despite very similar properties, M1 and M4 (and indeed other selected clones) have few mutations in common, suggesting there are multiple molecular solutions to the mismatch extension phenotype. One exception was E520G, a mutation that is shared by all but one of the four best clones of the final selection. Curiously, E520 is located at the very tip of the thumb domain at a distance of 20Å from the 3' OH of the mismatched primer terminus and its involvement in mismatch recognition or extension is unclear. However, E520G is clearly important for mismatch extension as backmutation reduces mismatch extension in both M1 and M4 to near wt levels (Fig. 2).

The only other feature clearly shared by both M1 and M4 are mutations targeting residues, which may be involved in flipping out the +1 template base. Residue E742

mutated in M1 (E742G) forms a direct contact with the flipped out +1 base on the template strand (Li et al), while in M4 the adjacent residue A743 is mutated to proline (A743P), which may disrupt interactions by distorting local backbone conformation. Back mutation of E742G in M1 reduced mismatch extension, but only by ca. 20% indicating that it does not contribute decisively to mismatch extension.

Surprisingly, mutations in the N-terminal 5'-3' exonuclease domain (53exoD) also appear to be contributing to mismatch extension as suggested by the 2-4 fold increased mismatch extension ability of chimeras of the 53exoD of M1, M4 and polD of wtTaq (Fig. 4). How they promote mismatch extension is unclear but given the apparent distance of the 53exoD from the active site (Utz 99, Eom 96) is unlikely to involve direct effects on extension catalysis. Increased affinity for primer-template duplex could also increase mismatch extension (Huang 92) but dissociation constants of wtTaq, M1 and M4 for matched and mismatched primer-template duplex were indistinguishable as judged by an equilibrium binding assay (Huang 92) (not shown).

The relationship of M1 and M4 with other naturally occurring DNA polymerases

Extension of mismatched 3' primer termini is a feature of naturally occurring polymerases. Viral reverse transcriptases (RT) like HIV-1 RT or AMV RT and polymerases capable of translesion synthesis (TLS) such as the polY-family polymerases pol ι (Vaisman 2001JBC) or pol κ (Washington 2002 PNAS) or the unusual polB-family polymerase polζ (Johnson Nature), all extend 3' mismatches with elevated efficiency compared to high-fidelity polymerases. Thus, the selected polymerases share significant functional similarities with preexisting polymerases but represent, to our knowledge, the only known polA-family polymerases that are proficient in mismatch extension (ME) and translesion synthesis (TLS). In contrast to TLS polymerases, which are distributive and depend on cellular processivity factors such as PCNA (Prakash refs for eta/kappa and iota), M1 and M4 combine ME and TLS with high processivity and in the case of M1 are capable of efficient amplification of DNA fragments of up to 26kb.

In the case of viral RTs, ME may play a crucial role in allowing error-prone yet processive replication of a multi-kb viral genome. For TLS polymerases, proficient mismatch extension is also a necessary prerequisite for their biological function as unpaired and distorted primer termini necessarily occur opposite lesions in the DNA template strand. The ability of TLS polymerases to traverse replication blocking lesions in DNA is thought to arise from a relaxed geometric selection in the active site (Goodman 02). The ability of M1 and M4 to process both bulky mispairs and a distorting CPD (cys-syn thymidine-thymidine dimer) dimer makes it plausible that, in analogy to TLS polymerases, they also have acquired a more open active site. Indeed, modelling showed that a CPD dimer can not be accommodated in the wtTaq polymerase active site without mayor steric clashes (Trincao01).

M1 (and to a lesser degree M4) also display a much increased ability to incorporate extend and replicate different types of unnatural nucleotide substrates that deviate to varying degrees from the canonical nucleobase structure. Of these the αS substitution is the most conservative. However, the sulfur anion is significantly larger than oxygen anion and coordinates cations poorly, which may be among the reasons why the wt enzyme will not tolerate full αS substitution. Fluorescently-labelled nucleotides like aS nucleotides retain base-pairing potential but include a bulky and hydrophobic substituent that must be accomodated by the polymerase active site. Steric clashes in the active site are allievated by the presence of a long, flexible linker. Indeed, we find biotin-16-dUTP amuch better substrate for M1 than biotin-11-dUTP, while wtTaq cannot utilize either. The hydrophobic analogue 5NI represents the most drastic departure from standard nucleotide chemistry we investigated. Comparable in size to a purine base, 5NI competely lacks any hydrogen bonding potential but like the natural bases, favours the anti-position with respect to the ribose sugar as judged by NMR (J. Gallego, D.L. and P.H., unpublished results). Therefore, a 5NI•A or 5NI•G basepair would closely resemble a purine-purine transversion mismatch and may cause similar distortions to the canonical DNA duplex geometry. Elegant experiments using isosteric non-hydrogen bonding base analogues have shown that Watson-Crick hydrogen bonding *per se* is not required for efficient insertion or replication (reviewed by Kool 02). However, while many non-hydrogen-bonding hydrophobic base analogues are efficiently incorporated, they

subsequently lead to termination, both at the 3' end and as a template base (Kool , Romesberg).

Structural and biochemical studies have previously identified regions of the polymerase structure that are important for mismatch discrimination such as motif A (involved in binding the incoming dNTP), the O-helix (motif B) and residues involved in minor groove hydrogen bonding (24, 25). Inspection of the sequence of M1 and M4 reveals a conspicuous absence of mutations in these regions. Rather mutations in M1 and M4 implicate regions of the polymerase not previously associated with substrate recognition such as the tip of the thumb subdomain (E520), the +1 template base-flipping function (E742, A743) in the finger subdomain and the 5-3' exonuclease domain (53exoD).

The 53exoD is too distant from the active site to have direct effects on mismatch extension. It is, however, thought to be crucial for polymerase processivity and may thus influence mismatch extension (24). Indeed, the Stoffel fragment of *Taq* polymerase (26), which lacks the 53exoD, displays both reduced processivity and more stringent mismatch discrimination (27). Mutations in the 53exoD of M1 and M4 may therefore contribute to mismatch extension by enhancing polymerase processivity. Together with the ability to bypass abasic sites (generated in large DNA fragments during thermocycling) this may also contribute to the proficiency of M1 at long PCR (Fig. 5). E520 is located at the very tip of the thumb domain at the end of the H2 helix at a distance of 20Å from the 3' OH of the mismatched primer terminal base (P1) (2). Mechanistic aspects of the involvement of the E520G mutation in mismatch recognition or extension are therefore not obvious either. It is worth noting though that adjacent regions, especially the preceding loop connecting helices H1 and H2 and parts of helix I, make extensive contacts with the template-primer duplex between P3-P7 (2). It has previously been observed that mismatch incorporation affects extension kinetics up to the P4 position (24). E520G may modify the structure of these regions to ease passage of mismatches and increase elongation efficiency post incorporation. Base flipping, i.e. rotation of the designated base out of the DNA helix axis is a common mechanism among DNA modifying enzymes (e.g. glycosylases) but its precise role for polymerase function is less clear. It

has been speculated that flipping out of the +1 template base may contribute to polymerase fidelity by preventing out-of-register base-pairing (25) of the 3' nucleotide to cognate upstream template bases. Interference with this mechanism therefore might promote apparent mismatch extension but would produce -1 base deletions. However, neither primer extensions nor sequencing of PCR products generated with M1 or M4 using primers with 3' G•A and C•C mismatches revealed any template slippage but on the contrary, confirmed in-register extension of the mismatches (not shown). The utility of reduced base-flipping in the context of the TLS capability of M1 and M4 is easier to understand, especially on the CPD dimer, as the two covalently linked thymine template bases would be refractory to flipping out. Indeed, TLS polymerases, which are naturally able to bypass CPD dimers, appear to lack a base-flipping function (28).

Extension and incorporation kinetics of polymerases according to the invention.

Examination of the extension and incorporation kinetics of the mutant polymerases suggests that they have a significantly increased propensity to not only extend but also incorporate transversion mispairs and consequently should have a significantly increased mutation rate compared to the wt enzyme. More relaxed geometric selection in the active site might also be expected to come at the price of significantly reduced fidelity as indeed is the case for TLS polymerases (23). However, measurement of the overall mutation rate using the MutS assay (not shown) and sequencing of PCR products generated by M1 indicated only a modest (< 2-fold) increase in the mutation rate (Table 1) mostly due to an increased propensity for transversions. As discussed previously (10), CSR should select for optimal self-mutation rates within the error threshold (31). A change in the mutation spectrum towards a more even distribution of transition and transversion mutations may be an effective solution to accelerate adaptation, while maintaining a healthy distance from the error threshold. This may also make M1 a useful tool for protein engineering as the bias of Taq (and other DNA polymerases) for transition mutations limits the regions of sequence space that can be accessed effectively using PCR mutagenesis

Table 1: Mutation spectrum of wtTaq and M1 in PCR

	Transitions		Transversions				Deletions
	AT->GC	GC->AT	AT->TA	AT->CG	GC->TA	GC->CG	
WtTaq*	25	9	8	2	3	1	3
M1*	25	16	15	4	5	10	1

*Mutations derived from sequencing of 40 clones (800bp) each.

In summary DNA polymerases according to the present invention, in particular M1 and M4 respectively as depicted in SEQ No 1 and SEQ No 2 possess the following properties:

(1) DNA Translesion synthesis

(2) A generic ability to incorporate unnatural base analogues into DNA.

(3) M1 has the ability to efficiently amplify DNA targets up to 26kb.

Uses of DNA polymerases according to the invention.

Directed evolution towards extension of distorting transversion mismatches like G·A or C·C by CSR yields novel, "unfussy" polymerases with an ability to perform not only efficient mismatch extension and TLS but also accept a range of unnatural nucleotide substrates. The present inventors have shown that the evolution of TLS from a high-fidelity, polA-family, pol B family or other polymerases requires but few mutations, suggesting that TLS and relaxed substrate recognition are functionally connected and may represent a default state of polymerase function rather than a specialization.

The unusual properties of the DNA polymerases according to the present invention, in particular M1 and M4 may have immediate uses for example for the improved incorporation of dye-modified nucleotides in sequencing and array labelling and/or the amplification of ultra-long DNA targets. They may prove useful in the amplification of damaged DNA templates in forensics or paelobiology, may permit an expansion of the chemical repertoire of aptamers or deoxi-ribozymes (Benner, Barbas, ribozyme review)

and may aid efforts to expand the genetic alphabet (Benner, Schultz). The altered mutation spectrum of M1 may make a useful tool in random mutagenesis experiments as the strong bias of Taq and other polymerases towards (A->G, T->C) transitions limits the combinatorial diversity accessible through PCR mutagenesis. Furthermore, the ability of M1 & M4 to extend 3' ends in which the last base is mismatched with the template strand and the ability of H10 (see example 6) to extend 3' ends in which the last two bases are mismatched with the template strand may extend the scope of DNA shuffling methods (Stemmer) by allowing to recombine more distantly related sequences.

In addition, DNA polymerases according to the invention, in particular pol A polymerases, for example M1 and M4 pol A polymerases as herein described may serve as a useful framework for mutagenesis and evolution towards polymerases capable of utilizing an ever wider array of modified nucleotide substrates. The inventors anticipate that directed evolution may ultimately permit modification of polymerase chemistry itself, allowing the creation of amplifiable DNA-like polymers of defined sequence thus extending molecular evolution to material science.

The invention will now be described by the following examples which are in no way limiting of the invention claimed herein.

Example 1

General Methods

List of primers:

1: 5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GA-3'; A•G mismatch

2: 5'-GTA AAA CGA CGG CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GCC-3'; C•C mismatch

3: 5'-AAA AAT CTA GAT AAC GAG GGC AA-3'

4: 5'-ACC ACC GAA CTG CGG GTG ACG CCA AGC G-3'

5: 5'-GAA CTG CGG GTG ACG CCA AGC GCA 3'; A•A mismatch

6: 5'-CC GAA CTG CGG GTG ACG CCA AGC G**G** 3'; **G**•G mismatch

7: 5'-GAA CTG CGG GTG ACG CCA AGC GC**G**-3'; **G**•A mismatch

8: 5'-AAA AAT CTA GAT AAC GAG GGC AA-3'

9: 5'-CCG ACT GGC CAA GAT TAG AGA GTA TGG-3'

10: 5'-GAT TTC CAC GGA TAA GAC TCC GCA TCC-3'

11: 5'-GGC AGA CGA TGA TGC AGA TAA CCA GAG C-3'

12: 5'-GCC GAT AGA TAG CCA CGG ACT TCG TAG-3'

13: 5'-GGA GTA GAT GCT TGC TTT TCT GAG CC-3'

14: 5'-GAG TTC GTG CTT ACC GCA GAA TGC AG-3'

15: 5'-ACC GAA CTG CGG GTG ACG CCA AGC G 3'

16: 5'-ACC GAA CTG CGG GTG ACG CCA AGC C 3'

17: 5'-ACC GAA CTG CGG GTG ACG CCA AGC A 3'

18: 5'-AAA CAG CGC TTG GCG TCA CCC GCA GTT CGG T-3'

19: 5'-AAA CAG GGC TTG GCG TCA CCC GCA GTT CGG T-3'

20: 5'-AAA CAG AGC TTG GCG TCA CCC GCA GTT CGG T-3'

21: 5'-AAA CAC CGC TTG GCG TCA CCC GCA GTT CGG T-3'

22: 5'-AGC TAC CAT GCC TGC ACG AAT TCG GCA TCC GTC GCG ACC ACG GTC GCA GCG-3' (undamaged)

23: 5'-AGC TAC CAT GCC TGC ACG ACA **X**CG GCA TCC GTC GCG ACC ACG GTC GCA GCG-3'; **X**= abasic site

24: 5'-AGC TAC CAT GCC TGC ACG AA**X** **X**CG GCA TCC GTC GCG ACC ACG GTC GCA GCG-3, **XX**= CPD dimer

25: 5'-CGT GGT CGC GAC GGA TGC CG-3'

26: 5'-TAA TAC GAC TCA CTA TAG GGA GA-3'

27: 5'-ACT G**X**T CTC CCT ATA GTG AGT CGT ATT A-3'; **X**=5NI

Materials and Methods

DNA manipulation and protein expression. Expression of Taq clones for screening and CSR selection was as described (10). For kinetic measurements and gel extension assays, polymerases were purified as described (32) using a Biorex70 ion exchange resin (BioRad). All PCR and primer extensions were performed in 1x *Taq* buffer (50mM KCl/10mM Tris•HCl (pH 9.0) / 0.1% Triton X-100 / 1.5mM $MgCl_2$), with

dNTPs (0.25mM (Amersham Pharmacia Biotech, NJ)) and appropriate primers unless specified otherwise. Primer sequences are provided in Supplementary information. Primer extension reactions were terminated by addition of 95% formamide/ 10 mM EDTA and analysed on 20 % polyacrylamide / 7 M Urea gels.

CSR selection. Activity preselected libraries L1* and L2* (10) were combined and 3 rounds of CSR selection carried out as described (10) except using primers 1: (A•G mismatch) and 2: (C•C mismatch) and 15 cycles of (94°C 1 min, 55°C 1 min, 72°C 8 min). Round 2 clones were recombined by staggered extension process (StEP) PCR shuffling (33) as described . For round 3, CSR cycles were reduced to 10 and annealing times to 30 sec.

PCR. A PCR assay was used to screen and rank clones. Briefly, clones were normalized for activity in PCR with matched primers 3, 4 and activity with mismatched primers 1 and 2 (1μM each) determined at minimal cycle number (15-25 cycles). Extension capability for different mismatches was determined by the same assay using mismatch primers 2 (C•C mismatch), 5 (A•A mismatch), 6 (G•G mismatch), 7 (G•A mismatch) with matched primer 3 or primer 1 (A•G mismatch) with matched primer 4. Incorporation of unnatural substrates in 50 cycle PCR was carried out using standard conditions and 50μM αS dNTPs (Promega) or 50μM FITC-12-dATP (Perkin-Elmer), Rhodamine-5-dUTP (Perkin-Elmer) or Biotin-16-dUTP (Roche) with equivalent amounts of the other 3 dNTPs (all 50μM). Long PCR was carried out using a two-step cycling protocol as described (22) 94°C for 2 minutes, followed by 20 cycles of (94°C 15 sec, 68°C 30 min) using 5ng of phage λ DNA (New England Biolabs) template and either primers 9, 10, 11 with primer 12 or primer 13 with primers 10, 14.

Single nucleotide incorporation/extension kinetics. Kinetic parameters were determined using a gel-based assay essentially as described (16). Primers 15, 16, 17 (3' base = G, C, A respectively) were ^{32}P-labeled and annealed to one of template strands 18, 19, 20 (template base = C, G, A respectively) or 21 (template base C

different context). Duplex substrates were used at 50nM final concentration in 1 X Taq buffer with various concentrations of enzyme and dNTP. Reactions were carried out at 60°C for times whereby <20% of primer-template was utilized at the highest concentration of dNTP.

Template affinity assays. An equilibrium binding assay (12) was used to determine relative affinity of polymerases for the mismatched primer-templates used in the kinetics assays. Polymerases were preincubated at 60°C in 1xTaq buffer with 50nM ^{32}P-labeled matched primer-template and 50nM unlabeled mismatched competitor primer-templates. Reactions were initiated by simultaneous addition of dCTP (200μM) and trap DNA (*Xba*I / *Sal*I-restricted sheared salmon sperm DNA, 4.5 mg/ml). Prior experiments demonstrated trap-effectiveness over the time period used (15 seconds).

Translesion Replication Assay. Template primers 22 (undamaged) or 23 (containing a synthetic abasic site) were synthesized by Lofstrand Laboratories (Gaithersburg, MD). Template primer 24 (containing a single *cis-syn* thymine dimer), was synthesized as described (34). Primer 25 was ^{32}P-labeled and annealed to one of the three templates 22, 23, 24 (at a primer template ratio of molar 1:1.5) and extended in 40mM Tris•HCl at pH 8.0, 5mM $MgCl_2$, 100μM of each dNTP, 10mM DTT, 250 μg/ml BSA, 2.5% glycerol, 10 nM primer-template DNA and 0.1 Unit of polymerase at 60°C for various times.

5NI replication assay. Primer 26 was ^{32}P-labeled and annealed to template primer 27 (containing a single 5-nitroindole) in 1xTaq buffer, 0.1 or 0.5U of the polymerase was added and reactions incubated at 60°C for 15 mins, after which 40μM of each dNTP were added and incubation at 60°C continued for various times.

Fidelity assays. Mutation rates were determined using the mutS ELISA assay (Genecheck, Ft. Collins, CO) or by performing 2x50 cycles of PCR on three different templates and sequencing the cloned products.

Example 2

Kinetic analysis. Extension and incorporation kinetics of M1 and M4 for a selection of mismatches were measured using a gel-based steady-state kinetic assay (Goodman) (Tables 1 & 2). M1 and M4 respectively extend a C•C mispair 390 and 75-fold more efficiently than wtTaq. Examination of the other most disfavored mismatches (G•A, A•G, A•A, G•G) reveals generic, although less pronounced, increases of extension efficiencies, as suggested by the PCR assay (Fig. 4, fig 5). The gain in extension efficiency derives predominantly from increased relative Vmax values for the mutant polymerases, while Km for nucleotide substrates remains largely unchanged. For most DNA polymerases the relative efficiency of extending a given mispair (f0ext) is similar to the relative efficiency of forming it (finc) (Goodman 1993, Goodman 1990, Washington 2001).Indeed, M1 and M4 respectively incorporate dCTP opposite template base C 206- and 29-fold more efficiently than wtTaq (Table 2).

Table 2: Steady-state kinetic parameters for extension kinetics by wtTaq and mutant polymerases.

3'–Terminal Base pair *	Polymerase	V_{max} (%Min^{-1})	K_m (μM)	f †	f_{ext} ‡	Ratio of f_{ext} §
C·G	WtTaq	1477.0	0.016	92312.5	-	-
	M1	308.0	0.02	15400	-	-
	M4	817.0	0.012	68083	-	-
C·C	WtTaq	0.2	39.9	0.00546	5.9×10^{-8}	1.0
	M1	9.2	25.8	0.356	2.3×10^{-5}	390.0
	M4	11.1	36.6	0.303	4.5×10^{-6}	75.3
G·A	WtTaq	1.6	32.8	0.05	5.4×10^{-7}	1.0
	M1	2.4	22.0	0.111	7.2×10^{-6}	13.3
	M4	7.5	29.0	0.26	3.8×10^{-6}	7.0
A·G	WtTaq	28.0	45.2	0.02	2.1×10^{-7}	1.0

	M1	44.6	280.2	0.02	1.3×10^{-6}	6.2
	M4	50.0	259.0	0.1	1.5×10^{-6}	7.0
A:A	WtTaq	1.7	27.3	0.062	6.7×10^{-7}	1.0
	M1	1.5	40.9	0.037	2.4×10^{-6}	3.6
	M4	8.5	32.9	0.259	3.8×10^{-6}	5.7
G:G	WtTaq	20.4	174.0	0.117	1.3×10^{-6}	1.0
	M1	29.6	67.0	0.44	2.9×10^{-5}	22.5
	M4	70.6	107.0	0.66	9.7×10^{-6}	7.6

* Template base : 3' primer base; Incorporated base is dCTP

† f, enzyme efficiency = V_{max} / K_m

‡ f_{ext}, f (mismatched 3'terminus) / f (matched terminus)

§ f_{ext} (mutant polymerase) / f_{ext} (wtTaq)

Table 2: Steady-state kinetic parameters for incorporation kinetics by wtTaq and mutant polymerases.

Base pair *	Polymerase	V_{max} (%Min^{-1})	K_m (μM)	f †	f_{inc} ‡	Ratio of f_{inc} §
G: dCTP	WtTaq	1477	0.016	92312.5	-	-
	M1	308	0.02	15400	-	-
	M4	817	0.012	68083	-	-
G: dGTP	WtTaq	57.47	365.27	0.157	1.7×10^{-6}	1
	M1	215.98	377.1	0.573	3.72×10^{-5}	21.88
	M4	656.46	82.34	7.97	1.17×10^{-4}	68.82
G:dATP	WtTaq	1973.68	258.53	7.63	8.27×10^{-5}	1
	M1	681.82	257.2	2.65	1.72×10^{-4}	2.08
	M4	1935.48	157.77	12.27	1.80×10^{-4}	2.18
G:dTTP	WtTaq	25.08	1.64	15.29	1.65×10^{-4}	1
	M1	10.19	1.65	6.18	4.01×10^{-4}	2.43
	M4	63.20	5.10	12.39	1.82×10^{-4}	1.1

C: dGTP	WtTaq	2356.02	0.0366	64285.69	-	
	M1	111.66	0.0387	2884.55	-	
	M4	335.42	0.01	33542	-	
C: dCTP	WtTaq	3.3	1330.89	0.0025	3.86×10^{-8}	1
	M1	6.08	264.14	0.023	7.97×10^{-6}	206.74
	M4	52.63	1390.63	0.0378	1.13×10^{-6}	29.22

* Template base: incoming nucleotide

† f, enzyme efficiency = V_{max} / K_m

‡ f_{inc}, f (incorrect dNTP) / f (correct dNTP)

§ f_{inc} (mutant polymerase) / f_{inc} (wtTaq)

Example 3

Translesion synthesis. Transversion mispairs represent distorting deviations from the cognate duplex structure. We therefore investigated if M1 and M4 were capable of processing other deviations of the DNA structure such as lesions in the template strand. Using a gel-extension assay we investigated their ability to traverse an abasic site and a *cis-syn* thymine pyrmidine dimer (CPD) template strand lesion. In control assays using an undamaged template, wtTaq, M1 and M4 efficiently and rapidly extended primers to the end of the template (Fig. 5). On the template containing an abasic site, wtTaq efficiently inserts a base opposite the lesion but, further extension is largely abolished. In contrast, both M1 and M4 are able to extend past the lesion and to the end of the template. The size of the final product is similar to that observed on the undamaged template indicating that bypass occurred without deletions. Perhaps the most striking example of the proficiency of M1 and M4 to bypass template lesions is observed on the CPD-containing template (Fig. 5). Under the assay conditions, wtTaq utilizes a fraction of the available template and is only able to insert a base opposite the 3'T of the dimer after prolonged reaction conditions. In contrast, both M1 and M4 are able to readily extend all of the primer to the 3'T of the dimer. In addition, there is also considerable extension of these primers to the 5'T of the CPD. As with the abasic template, a significant fraction of these primers are subsequently fully extended to the end of the

template in an error-free manner without deletions. We estimate that trans-lesion synthesis (TLS) by M1 and M4 may only be 2-5 fold less efficient than that observed with a naturally occurring TLS polymerase, Dpo4 from *S. solfataricus* (Boudsocq et al (2001), Nucleic Acid Res, 29, 46072001) on the same template.

Example 4

Unnatural substrates. We reasoned that relaxed geometric selection might also aid the incorporation of unnatural base analogues, some of which inhibit or arrest polymerase activity due to poor geometric fit or lack of interaction with either polymerase or template strand. A first, conservative example are phosphothioate nucleotide triphosphates (αS dNTPs), in which one of the oxygen atoms in the α phosphate group is replaced by sulfur. As part of a dNTP mixture, αS dNTPs are generally well accepted as substrates by DNA polymerases but when we replaced all four dNTPs with their αS counterparts in PCR wtTaq failed to generate any amplification products, while M1 (and to lesser extent M4) were able to generate PCR products of up to 2kbp, indicating that they could utilize αS dNTPs with much increased efficiency compared to the wt enzyme (Fig. 6). As expected, the resulting αS DNA was completely resistant to cleavage by DNA endonucleases (not shown). Nucleotides bearing bulky adducts such as fluorescent dyes are widely used in applications such as dye terminator sequencing or array labelling. Although generally well tolerated they are incorporated considerably less efficiently than the natural dNTP substrates and can cause permature termination in homopolymeric runs, presumably because of steric crowding due to the bulky dye molecules. In PCR the effect is potentiated because both template and product strands are labelled. When we replaced dUTP with Biotin-16-dUTP or dATP with FITC-12-dATP in PCR, wtTaq was unable to generate any amplification products, while M1 was able to generate 2.7 kb amplification products fully labelled with Biotin-16-dUTP or a 0.4kb fully labelled with FITC-12-dATP (Fig. 6). Recently, there has been significant interest in hydrophobic, non-hydrogen bonding base-analogues and the applications they may enable. One of these is the candiate "universal base" 5-nitroindole (5NI) (Loakes & Brown 96), which, like other hydrophobic, strongly stacking base analogues , is readily accepted as a substrate, but

once incorporated, acts as a strong terminator both at the 3' end and as a template base. In contrast, M4 and in particular M1 efficiently bypass template strand 5NI (Fig. 6) and to a lesser degree, extend 5NI at the 3' end (not shown).

Example 5

Long PCR. Amplification product size with wtTaq is generally limited to fragments a few kb long but can be extended to much longer targets by inclusion of a proofreading polymerase (Barnes 92). We found that the selected polymerases, in particular M1 was able to efficiently amplify of targets up to 26kb (Fig. 7), using standard PCR conditions in the absence of auxiliary polymerases or other processivity factors. Under the same conditions wtTaq enzyme failed to amplify targets >9kb. The molecular basis for the product size limitation in the wt enzyme is thought to be premature termination due to an inability to extend mismatches following nucleotide misincorporation. These are thought to be removed by the proofreading polymerase allowing extension to resume. Our results that a generic mismatch extension ability to results in a similarly extended amplification range supports this concept.

Example 6: Libraries of polymerase chimeras

Libraries of chimeric polymerase gene variants were constructed using a gene shuffling technique called Staggered extension protocol (StEP, (Zhao, Giver et al. 1998)). This technique allows two or more genes of interest from different species to be randomly recombined to produce chimeras, the sequence of which contains parts of the original input parent genes. .

Thermus aquaticus (Taq) wild type and T8 (a previously selected 11 fold more thermostable Taq variant (Ghadessy, Ong et al. 2001)), *Thermus thermophilus (Tth)* and *Thermus flavus (Tfl)* polymerases had previously been amplified from genomic DNA and cloned into pASK75 (Skerra 1994) and tested for activity. These genes were then shuffled using the staggered extension protocol (StEP) as described (Zhao, Giver

et al. 1998) with (CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GCA A and GTA AAA CGA CG G CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GCG), recloned into pASK75 and transformed into *E. coli* TG1. The library size was scored by dilution assays and determining the ratio of clones containing insert using PCR screening and was approximately 10^8. A diagnostic restriction digest of 20 clones produced 20 unique restriction patterns, indicating that the library was diverse. Subsequent sequencing of selected chimeras showed an average of 4 to 6 crossovers per gene.

Example 7: Selection of two mismatch extension polymerase.

CSR emulsification and selection was performed on the StEP *Taq, Tth* and *Tfl* library essentially as described (Ghadessy, Ong et al. 2001). Mismatch primers with two mismatches at their 3' end (5'-GTA AAA CGA CGG CCA GTT TAT TAA CCA CCG AAC T**GC**-3', 5'-CAG GAA ACA GCT ATG ACT CGA CAA AAA TCT AGA TAA CGA **CC**-3') were in the emulsion as the source of selective pressure. The aqueous phase was ether extracted, PCR purified (Qiagen, Chatsworth, CA) with an additional 35% GnHCl, digested with *Dpn*I to remove methylated plasmid DNA, treated with ExoSAP (USB) to remove residual primers, reamplified with outnested primers and recloned and transformed into *E. coli* as above.

The resultant clones were screened and ranked by PCR assay. Briefly, 2 µl of induced cells were added to 20 µl of PCR mix with the relevant mismatch primers. Clones that produced a band were then subjected to further analysis and the most active clones were sequenced.

In particular, clone H10 has significant activity on the primers with two mismatches. H10 is a chimera of *T. aquaticus* wild type (residues 4 to 20 and 221 to 640), T8 (residues 1 to 3 and 641 to 834) and *T. thermophilus* (residues 21 to 220). H10 has five detectable crossover sites and 13 point mutations, of which 4 are silent (F74→I, F280→L, P300→S, T387→A, A441→V, A519→V Q536→R, R679→G, F699→L).

Example 8: Selecting for a 4 mismatch extension polymerase.

CSR emulsification and selection was performed on the StEP *Taq, Tth* and *Tfl* library essentially as described (Ghadessy, Ong et al. 2001). The library had previously been cloned into pASK75 (see example 6). The aqueous phase was ether extracted and replication products were purified using a PCR purification kit (Qiagen, Chatsworth, CA) including a wash with an 35% GnHCl. 7 µl of purified replication products (from 48) were digested with 1µl *Dpn*I (20 Units) to remove plasmid DNA and treated with 2 µl ExoSAP (USB) to remove residual primers for 1h at 37°C and reamplified with outnested primers (GTAAAACGACGGCCAGT and CAGGAAACAGCTATGAC, 94 °C 2 minutes, and then 30cycles of 94 °C 30 seconds, 50 °C for 30 seconds and 72 °C for 5 minutes with a final 65 °C for 10 minutes). Reamplification products were digested with XbaI and SalI, recloned into pASK75 and transformed into *E. coli* as above.

In parallel an alternative selection approach was used: the induced library was emulsified as above with the additional presence of biotinylated dUTP and incubated at 94°C 5 minutes, 50 °C 1 minute and 72 °C 1 minute. The aqueous phase was ether extracted, the DNA in the aqueous phase was precipitated by addition of 1/10 volume of 3M NaAc, 1 µl glycogen and 2.5 volumes of 100% ethanol. This was then incubated for 1 hour at -20 °C, spun for at 13000rpm for 30 minutes in a benchtop microcentrifuge, washed with 70% ethanol and resuspended in 50 µl buffer EB (Qiagen). 20ul of Dynabeads (DynaL Biotech) were washed twice and resuspended in 20µl of bead buffer (10mM Tris pH 7.5, 1mM EDTA, 0.2M NaCl) The washed beads were then mixed with the selection in a total volume of 0.5 ml bead buffer and then incubated overnight under constant agitation at room temperature to capture biotinylated products. Beads were washed twice in bead buffer, twice in buffer EB and finally resuspended in 50 µl bead buffer. The resuspended beads were reamplified with outnested primers (sequences and programme as above) and recloned and transformed into *E. coli* as above.

Two sets of mismatch primers with four mismatches at their 3' end (underlined) (5'-CAG GAA ACA GCT ATG ACA AAA GTG AAA TGA ATA GTT CGA CTTTT-3' and 5'-GTA AAA CGA CGG CCA GTC TTC ACA GGT CAA GCT TAT TAA GGTG-3' as the first set and 5'-CAG GAA ACA GCT ATG ACC ATT GAT AGA GTT ATT TTA CCA CAGGG-3' and 5'-GTA AAA CGA CGG CCA GTC TTC ACA GGT CAA GCT TAT TAA GGTG-3' as the second set) were used in the emulsion as two separate sources source of selective pressure.

The resultant clones from both CSR and CST were screened and ranked by PCR assay. Briefly, 2 µl of induced cells were added to 20 µl of PCR mix with the relevant 4 mismatch primers. Clones that produced a band were then subjected to further analysis and their activity on single, double and quadruple mismatch primers (single mismatch primers: 5'–CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GA-3' and 5'-GTA AAA CGA CGG CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GCC 3'; double mismatch primers: CAG GAA ACA GCT ATG ACT CGA CAA AAA TCT AGA TAA CGA CC and GTA AAA CGA CGG CCA GTT TAT TAA CCA CCG AAC TGC; four mismatch primers above.) was investigated. Polymerases that could extend all of these mismatches were found, though many polymerases could do only one of the mismatches and none could do all.

The plasmid DNA of the ten best clones was then purified and shuffled as described above (StEP, (Zhao, Giver et al. 1998)). This was then purified, cut and cloned and the resultant library was subjected to another round of CSR as described (Ghadessy, Ong et al. 2001). The same two sets of mismatch primers with four mismatches at their 3' end were used in the emulsion as two separate sources source of selective pressure. This was then dealt with as above and the resultant clones were screened and ranked by PCR assay (as above). Once again, polymerases that could extend all of these mismatches were found (see Table), though many polymerases could do only one of the mismatches and none could do all. There was a notable increase in clones displaying mismatch activity over the first round.

The best clones from the second round were combined with the best clones from the first round on a 96 well plate and were subjected to further screening.

The following table is a summary of the results.

	1	2	3	4	5	6	7	8	9	10	11	12
	-,-,-	-,-,-	-,-,-	+,-,-	+,-,-	+,+,-	+,-,-	+,-,-	+,-,-	-,-,-	-,-,-	-,-,-
	+,-,-	-,-,-	-,-,-	-,-,-	-,-,-	+,-,-	+,-,-	-,-,-	+,-,-	-,-,-	-,-,-	-,-,-
	-,-,-	+,-,-	+,-,-	-,-,-	+,-,-	+,-,-	-,-,-	+,-,-	+,-,-	+,-,-	-,-,-	+,-,+
	+,-,-	-,-,-	-,-,-	-,+,-	+,-,-	+,-,-	+,-,-	+,-,-	-,-,+	-,-,-	-,-,-	-,-,-
	+,-,-	+,+,-	+,+,-	+,+,-	+,+,-	+,+,-	+,+,-	+,-,-	+,+,-	+,-,-	+,-,-	+,-,-
	-,-,-	+,-,-	+,+,-	+,-,-	+,+,-	+,-,-	+,-,-	+,-,-	+,-,-	+,+,-	+,-,-	+,+,-
	+,-,-	+,-,-	+,-,-	+,-,-	+,-,-	+,+,-	+,-,-	+,-,-	+,-,-	+,-,-	+,-,-	+,-,-
	+,-,-	+,-,-	-,-,-	+,-,-	+,-,-	+,-,-	-,-,-	+,-,-	+,-,-	+,-,-	+,-,-	+,-,-

A1 is Tth polymerase; A2 Tfl; A3 Taq; A4 M1; A5 M4; A6 H10 (see previous example. 1A7 to 1D12 are first round clones (where 1 indicates that these are first round clones), 2E1 to 2H12 are second round clones (where 2 indicates that these are second round clones)

The best first and second round clones were shuffled as described above and subjected to another round of CSR. The same two sets of mismatch primers with four mismatches at their 3' end were used in the emulsion as two separate sources of selective pressure. This was then dealt with as above and the resultant clones were screened and ranked by PCR assay (as above). Once again, polymerases that could extend all of these mismatches were found . In particular, clones 3B5. 3B8, 3C12 and 3D1 (where 3 indicates that these are third round clones) were able to extend primers containing four mismatches. See figure 9

Some promising clones were sequenced. All of the polymerases displayed a similar composition: the first part of the protein, roughly corresponding to the 5-3 exonuclease domain of the polymerase, was derived from *Tth,* whilst the remaining part of the protein was derived from *Taq*. Four point mutations (L33→P, E78→K, D145→G and E822→K) re-occurred in the majority of sequenced mutants and one (B10) had acquired an extra 16 amino acids at its C terminus through a frame shift at position 2499. *Tfl* was highly underrepresented, although some of its sequence was present.

Example 9: Hairpin ELISA to measure polymerase activity.

The below protocol is a sensitive method to measure polymerase activity both for the incorporation of unnatural nucleotide substrates (added to the reaction mixture) or the extension or replication of unnatural nucleotide substrates (incorporated as part of the hairpin oligo).

The assay comprises a hairpin oligonucleotide which constitutes both primer and template in one. In contains as part of the hairpin a biotinylated dU residue, which allows capture of the hairpin oligonucleotide on streptavidin-coated surfaces.

The oligonucleotide folds up into a hairpin with a 5' overhang, which serves as the template strand for the polymerase (typical sequence: 5'- AGC TAC CAT GCC TGC ACG CAG TCG GCA TCC GTC GCG ACC ACG TT5 TTC GTG GTC GCG ACG GAT GCC **G**-3', bases involved in hairpin formation are underlined, 3' base is in bold, 5= dU-biotin) .

Extension reactions are carried out in the presence of small amounts of a labelled nucleotide typically DIG-16-dUTP. Product is captured (for example on a streptavidin coated ELISA plate) and incorporation of labelled nucleotide into the product strand

is measured (using for example an anti-DIG antibody) and taken as a measure of polymerase activity.

Method:

Extension reactions are carried out in 1x Taq buffer including 1-100nM of hairpin primer and 100µM dNTP mixture (comprising 0.3-30% dUTP-DIG), typically incubated at 94°C for 1-5min, followed by incubation at 50°C for 1-5min, followed by incubation at 72°C for 1-5min. (1-10µl) Reaction products are added to Streptavidin coated ELISA plates (Streptawell, Roche) in 200µl PBS, 0.2% Tween20 (PBST) and incubated at room temperature for 10min to 1h. ELISA plates are washed 3x in PBST and 200µl of anti-DIG-POD Fab2 fragment (Roche) diluted 1/2000 in PBST is added and the plate is incubated at room temperature for 10min to 1h. The plate is washed 3-4x in PBST and developed with an appropriate POD substrate.

Example 10: Hairpin-ELISAs to test nucleotide analogue incorporation by mismatch extension clones

Clones previously selected for their ability to extend from a 4 basepair mismatch were assayed for their ability to incorporate a variety of nucleotide analogues.

Clones were grown at 30°C overnight in 200µl 2XTY + ampicillin (100µg/ml).
A 150µl (2xTY + ampicillin 100µg/ml) overday culture was started from the overnight and grown for 3 hours at 37°C. After 3 hours protein expression was induced by the addition of 50µl of 2XTY + anhydrous tetracycline (8ng/ml) to the culture which was then allowed to grow for a further 3h at 37°C. The cells were pelleted at 2254xg for 5 minutes and the growth medium removed by aspiration after which the cell pellet was resuspended in 100µl 1xTaq buffer (10mM Tris-HCl, pH 9.0, 1.5mM $MgCl_2$, 50mM KCl, 0.1% Triton X-100, 0.01% (w/v) stabiliser; HT Biotechnology Ltd). Resuspended cells were lysed by incubation at 85°C for 10 minutes and the cell debris was pelleted at 2254xg for 5 minutes.

ELISA protocol:

Extension reaction.

Reactions were performed in a final volume of 12.5µl comprising:

1x Taq buffer (10mM Tris-HCl, pH 9.0, 1.5mM $MgCl_2$, 50mM KCl, 0.1% Triton X-100, 0.01% (w/v) stabiliser; HT Biotechnology Ltd).

50 pmoles of primer.

25 µM of each dNTP (minus the nucleotide analogue) of which 10% (2.5µM) of the dTTP is digoxigenin-11-dUTP and 90% (22.5µM) is dTTP.

25 µM the nucleotide analogue.

2.5 µl of cell lysate.

The reaction conditions were:

95°C 5 minutes; 50 °C 5 minutes; 72°C 5 minutes.

Detection reaction:

5 µl of the extension reaction was added to 200µl of PBS-Tween (1x PBS; 0.2% Tween 20) in StreptaWell high bind plates (Roche) and allowed to bind for 30 minutes at room temperature.The plate was washed 3X in PBS-Tween after which was added 200µl PBS-Tween + anti-digioxigenin-POD Fab fragments (antibody diluted 1/2000; Roche). The antibody was allowed to bind for 30 minutes at room temperature.

The plate was washed 3X in PBS-Tween and 200µl of the substrate added (per ml 100µl of 1M NaAc pH 6.0, 10µl of DAB, 1µl of H_2O_2, the reaction was allowed to develop after which it was stopped by adding 100µl of 1M H_2SO_4.

Experiment I. ELISA with Fluorescein 12-dATP:

The ability of clones selected for 4-mismatch extension to incorporate Fluorescein 12-dATP (Perkin Elmer) was assayed using the primer FITC4. The lysates used were concentrated 4-fold.

Experiment II. ELISA with Biotin 11-dATP:

The ability of clones selected for 4-mismatch extension to incorporate Biotin 11-dATP (Perkin Elmer) was assayed using the primer FITC10. The lysates used were concentrated 4-fold.

Experiment III. ELISA with CyDye 5-dCTP:

The ability of clones selected for 4-mismatch extension to incorporate Cy5-dCTP (Amersham Biosciences) was assayed using the primer ELISAC4P. The lysates used were concentrated 4-fold.

Experiment IV. ELISA with CyDye 3-dUTP:

The ability of clones selected for 4-mismatch extension to incorporate CyDye 3-dUTP (Amersham Biosciences) was assayed using the primer ELISAT3P. The lysates used were concentrated 4-fold. The DIG labelled dUTP in the extension reaction was replaced with Fluorescein 12-dATP and the incorporation of Fluorescein 12-dATP was detected by anti-Fluorescein-POD Fab fragments (Roche).

Experiment V. Abasic site ELISA ...

The ability of clones selected for 4-mismatch extension to bypass abasic sites was assayed using the primer Pscreen1Abas (AGC TAC CAT GCC TGC ACG CAG 1CG GCA TCC GTC GCG ACC ACG TT5 TTC GTG GTC GCG ACG GAT GCC G, 1= abasic site
5= dU biotin). The lysates used were concentrated 4-fold.

Clones selected for 4-mismatch extension were assayed for activity with different substrates using an ELISA assay.

A1= *Tth* Wild-type
A2= *Tfl* Wild-type
A3= *Taq* Wild-type
A4= *Taq* mutant M1

A5= *Taq* mutant M4

A6= *Taq* mutant H10

Rows A-D Clones isolated after 1 round of 4-mismatch selection

Rows E-H Clones isolated after 2 rounds of 4-mismatch selection

The results are shown in **Figure 8.**

Experiment V. Abasic site and 5-hydroxyhydantoin bypass

Polymerases 3A10 and 3D1 were investigated further for their ability to bypass abasic sites and 5-hydroxy hydantoins, which are both known to exist in damaged DNA such as found in ancient samples , using the ELISA based activity screen as described above. Both polymerases were more proficient at lesion bypass than wild type *Taq* by up to two orders of magnitude.

The hydantion phosphoramidite was synthesised by standard procedures starting from the hydantoin free base. Glycosylation of the silylated hydantoin base in the presence of tin(IV) chloride with the ditoluoyl (alpha) chlorosugar gave rise to two N-glycosylated products which were separated and characterised by 2D-NMR experiments. The tolyl groups were removed with ammonia to yield the free nucleoside which was dimethoxytritylated and phosphytylated in the usual manner. The hairpin primer to assay hydantoin bypass was: 5'-AGC TAC CAT GCC TGC ACG CAG XCG GCA TCC GTC GCG ACC ACG TTY TTC GTG GTC GCG ACG GAT GCC G-3', X=hydantoin, Y=Biotin-dU.

The sequences of the clones referred to in Examples are shown below: For the avoidance of any doubt, the first sequence provided in each section is the nucleic acid sequence. The second sequence provided is the corresponding amino acid sequence of the clone.

2F3:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCGTGGCCAAG

AAGGCGGAAAAGGAGGGGTACGAGGTGGGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGC
GTCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGC
AGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGA
CCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAA
ACGTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCT
CCCCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGA
GTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCG
GAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAG
GGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGAGACCTGAAGGAGGCGCGGGGGCTTCTCGC
CAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTAC
CTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGG
GAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGC
TTTACCGGGAGGTGGAGAGGCCCCTTTCCGTTGTCCTGGCCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGC
CTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGC
CACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCA
AGACGGAGAAGACCGGCAAGCGCTCCACCGGCGCCGCCGTCCTGGAGGCCCTCCACGAGGCCCACCCCATCGTGG
AGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCC
CAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAAC
CTCCAGAACATCCCCGTCCGCACCCAGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTAT
TGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGT
CTTCCAGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCAGGAGGCCGTGGACCC
CCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCTACCGCCTCTCCCAGGAG
CTAGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGAT
TGGGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGA
CCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACACGCCCGTCCAGGGCACCGC
CGCCGACCTCATGAAGCTAGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAG
GTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATG
GAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAG
TGA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHKAY
EAYRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATVAKKAEKEGYEVGILTADRGLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSVVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTGAAVLEALHEAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTQLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPQEAVDPLMRRAAKTINFGVLYGMSAYRLSQELAIPYEEAQAFIERYFQSFPKVRAWIGKTLEE
GRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNTPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLE
APKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE*

1A10:

ATGCGTGGTATGCCTCCTCTTTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCCTCCGCCACGAGG
CCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGACCTCTACCAACTCGTCTCCGACCGC
GTCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGC
AGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAGGA
CCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAA
ACGTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCT
CCCCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGA
GTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCG
GAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAG
GGGTGGTCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCC
AAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACC
TCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGG
AGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAAGCTTGAGGGGGAGGAGAGGCTCCTTTGGCT
TTACCGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCC
TATCTCAGGGCCTCGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCC
ACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAA
GACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGA
GAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCC
AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACAGGCAGGCTAAGTAGCTCCGATCCCAACC
TCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATT
GGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTC
TTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGTTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCC

TGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCGCCGCCTCTCCCAGGAGCT
AGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATT
GAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGAC
CTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCC
GCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGG
TCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGG
AGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGT
GA

MRGMPPLFEPKGRVLLVDGHLAYRTFFALKGPTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSLRHEAY
EAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGERTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGKLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRASSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSARRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLE
APKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE*

1A9:

ATGCGTGGTATGCATCCTCTTTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCAC
CTTCCACGCCCTGAAGGGGCTCACCACCAGCCGGGGGGAGCCGGTGCGGGCGGTCCACGGCTTCGCCAAGAGCCTC
CTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACG
AGGCCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGG
AGCTGGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGC
CAAGAAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGACCTCTACCAACTCGTCTCCGA
CCGCGTCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCG
GAGCAGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAG
AAGACCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGCTGAAGCCC
GCCATCCGGGAGAAGATCCTGGCCCACATGGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGTGCGCACCGACC
TGCCCCTAGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTG
AGCTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGACCCTGGAGGAGGCCTCCTGGCCCCCGCC
GGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCA
GGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGAGACCTGAAGGAGGCGCGGGGGCTTCTCG
CCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTA
CCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGG
GGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGG
CTTTACCGGGAGGTGGAGAGGCCCCTTTCCGTTGTCCTGGCCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGG
CCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGG
CCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGC
AAGACGGAGAAGACCGGCAAGCGCTCCACCGGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTG
GAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACC
CCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAA
CCTCCAGAACATCCCCGTCCGCACCCAGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTA
TTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGG
TCTTCCAGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCAGGAGGCCGTGGACCC
CCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCTACCGCCTCTCCCAGGAG
CTAGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGAT
TGGGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGA
CCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACACGCCCGTCCAGGGCACCGC
CGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAG
GTCCACGACGAGCTAGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATG
GAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAG
TGA

MRGMHPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVRAVHGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHE
AYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAV
LHPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRLKPAIREKILA
HMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLELGSLLHEFGLLESPKTLEEASWPPPEGAFVGFVLSRK
EPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVAR
RYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSVVLAHMEATGVRLDVAYLRALSLEVAEEIARL
EAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTGAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPD
LIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTQLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQ
EGRDIHTETASWMFGVPQEAVDPLMRRAAKTINFGVLYGMSAYRLSQELAIPYEEAQAFIERYFQSFPKVRAWIGKTLEE
GRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNTPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLE
APKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE*

2F12:

ATGCGTGGTATGCTTCCTCTTTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCAC
CTTCTTCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTC
CTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCCTCCGCCACG
AGGCCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGG
AGCTGGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGC
CAAGAAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGACCTCTACCAACTCGTCTCCGA
CCGCGTCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCG
GAGCAGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAG
AAGACCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGCTGAAGCCC
GCCATCCGGGAGAAGATCCTGGCCCACATGGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGTGCGCACCGACC
TGCCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTG
AGCTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCTCCTGGCCCCCGCC
GGAAGGGGCCTTCGTGGGCTTTGTGCTTACCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCC
AGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTC
GCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCT
ACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGG
GGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTG
GCTTTACCGGGAGGTGGAGAGACCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGCCTGGACGTG
GCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCG
GCCACCCCTTCAACCTCAACTCCCGAGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGG
CAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGT
GGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCAC
CCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCA
ACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCT
ATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGG
GTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGAC
CCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGG
AGCTAGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTG
GATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCC
AGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCAC
CGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCA
TGGAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGG
AGTGA

MRGMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSLRHEA
YEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVL
HPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRLKPAIREKILAH
MDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLELGSLLHEFGLLESPKALEEASWPPPEGAFVGFVLTRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELLLE
APQARAEEVAALAKEAMEKAYPLAVPLEVKVGIGEDWLSAKE*

1C2:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTATACCAACTCGTCTATGACCGC
GTCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGC
AGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGA
CCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAA
ACGTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCT
CCCCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGA
GTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCG
GAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAG
GGGTGGTCGAGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCC

AAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACC
TCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGG
AGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCT
TTACCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGGTGCGCCTGGACGTGGCC
TATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCC
ACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAATGGTGCTCTTTGACGAGCTTAGGCTTCCCGCCTTGGGGAAG
ACGCAAAAGACGGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAG
AAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGTCGGACCTCATCCACCCCA
GGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCT
CCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTACTG
GTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAAAACCTGATCAGGGTCT
TCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCC
TGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCT
AGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATT
GAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGAC
CTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCC
GCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGG
TCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGG
AGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGT
GA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHKAY
EAYRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVYDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLEMVLFDELRLPALGKTQKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLSDL
IHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLE
APKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE*

2G6:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCTTCGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCG
TCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
GTGGGTGGACTTCCGCGCCCTCGTGGGGAACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
GTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCC
CCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
TGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAA
GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGG
TGGTCGAGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
GGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTAC
CGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGGTGCGCCTGGACGTGGCCTATC
TCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCC
CTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
GAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAGAAG
ATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGA
CGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
GAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCC
AGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTAA
TGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCGCCGCCTCTCCCAGGAGCTAGC
CATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAG
AAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTA
GAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
GACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTCC
ACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGG
GGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTTTCCGCCAAGGGTTAG

Above: nucleic acid sequence of the clone

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHKAY
EAYRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATFAKKAEKEGYEVRILTADRGLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGNPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSARRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLE
APKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKG*

Above is the amino acid sequence of the clone

1A8:

ATGGTGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCTCCTCCTTCCGCCACGAGG
CCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGTGGACGACGTCCTGGCCAGCCTGGCCAAG
AAGGTGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGACCTCTACCAACTCGTCTCCGACCGCG
TCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
GCCCTCAAGCTCCTCAAGGAGTGGGGAGGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
GTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCC
CCCTGGAGGTGGACCTCGCCCAGGGGCGGGAACCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
TGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAA
GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGG
TGGTCGGGTCCACCGGACCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
GGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTAC
CGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCCTACC
TCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCC
CTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
GAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAGAAG
ATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGA
CGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
GAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCC
AGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTAA
TGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGC
CATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAG
AAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGTCGCTACGTGCCAGACCTA
GAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
GACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAAGAAACGGGGGCCAGGATGCTCCTTCAGGTCC
ACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGCCATGGAGG
GGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

MVMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKASSFRHEAY
EAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEVDDVLASLAKKVEKEGYEVRILTADRDLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGGLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRTPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEETGARMLLQVHDELVLEA
PKERAEAVARLAKEAMEGVYPLAVPLEVEVGIGEDWLSAKE*

2H1:

ATGGTGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCTCCTCCTTCCGCCACGAGG
CCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGTGGACGACGTCCTGGCCAGCCTGGCCAAG
AAGGTGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCG
TCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
GTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCC
CCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
TGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAA
GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGG
TGGTCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
GGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTAC
CGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCCTATC
TCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCC
CTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
GAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAGAAG
ATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGA
CGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
GAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGACCCGGGTCTTCC
AGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGA
TGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTGGC
CATCCCTTACGAGGAGGCCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGGATAGAA
AAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTACGTGCCCGACCTC
AACGCCCGGGTGAAGAGTGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
GACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCCAGGTCC
ACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATGGAGA
AGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCCAAGGAGTGAG
TCGACCTGCAGGCAGCGCTTGGCGTCACCCGCAGTTCGGTGGTTAA

MVMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKASSFRHEAY
EAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEVDDVLASLAKKVEKEGYEVRILTADRGLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAILEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLTRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEG
RKRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELLLE
APQARAEEVAALAKEAMEKAYPLAVPLEVKVGIGEDWLSAQGVSRPAGSAWRHPQFGG*

2F11:

ATGCGTGGTATGCTTCCTCTTTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCAC
CTTCTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTC
CTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCGCCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACG
AGGCCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGG
AGCTGGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCTGGCTACGAGGCGGACGACGTCCTCGCCACCCTGGC
CAAGAAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGACCTCTACCAACTCGTCTCCGA
CCGCGTCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCG
GAGCAGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAG
AAGACCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCA
GAAAACGTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCG
ACCTCCCCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCTCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCT
TGAGTTTGGCGGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCG
CCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGC
CAGGGGTGGTCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTC
GCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCT
ACCTCCTGGACCCTTCCAACACCGCCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGG
GGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTG
GCTTTACCGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACAGGGGTACGGCTGGACGTG
GCCTGCCTGCAGGCCCTTTCCCTGGAGCTTGCGGAGGAGATCCGCCGCCTCGAGGAGGAGGTCTTCCGCTTGGCGG
GCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGG

CAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGT
GGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCAC
CCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCA
ACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCGTCGCCGAGGAGGGGTGGCT
ATTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGACCCGG
GTCTTCCTGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACC
CCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGA
GCTGGCCATCCCTTACGAGGAGGCCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGG
ATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTACGTGCCC
GACCTCAACGCCCGGGTGAAGAGTGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACC
GCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCA
TGGAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGG
AGTGA

MRGMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQAVYGFAKSLLKALKEDGYKAAFVVFDAKAPSFRHEA
YEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVL
HPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIK
AHLEDLRLSLELSRVRTDLPLEVDLAQGRELDRERLRAFLERLEFGGLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRK
EPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTAPEGVA
RRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVACLQALSLELAEEIRR
LEEEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAILEALREAHPIVEKILQYRELTKLKSTYIDPLPD
LIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFVAEEGWLLVVLDYSQIELRVLAHLSGDENLTRVF
LEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLE
EGRKRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELL
LEAPQARAEEVAALAKEAMEKAYPLAVPLEVKVGIGEDWLSAKE*

2H4:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCG
TCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
GTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCC
CCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
TGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAA
GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGG
TGGTCGAGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
GGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTAC
CGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATC
TCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCC
CTTCAACCTCAACTCCCGGGACCAGCTGGAAATGGTGCTCTTTGACGAGCTTAGGCTTCCCGCCTTGGGGAAGACG
CAAAAGACGGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAGAAG
ATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGTCGGACCTCATCCACCCCAGGA
CGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
GAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTACTGGTG
GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAAAACCTGATCAGGGTCTTCC
AGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGA
TGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGC
CATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAG
AAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTA
GAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
GACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAACGGGGGCCAGGATGCTCCTTCAGGTCC
ACGACGAGCTGGTCCTTGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGG
GGGTGTATCCCCTGGCCGTGTCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHKAY
EAYRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE

PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLEMVLFDELRLPALGKTQKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLSDL
IHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEETGARMLLQVHDELVLEA
PKERAEAVARLAKEVMEGVYPLAVSLEVEVGIGEDWLSAKE*

2H9:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCGCGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCCCCCTGGCCAAG
AAGGCGGAAAAGGAGGGGTTCGAGGTGCGCATCCTCCCCGCCGTCCGCGGCCTCTGCCCTCTCGTCTCTGACCGCG
TCGCCGTCCTCCTCCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGAAGAAGACC
GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
GTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCC
CCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
TGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAA
GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGG
TGGTCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
GGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTGTGGCTTTAC
CGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACAGGGGTACGGCTGGACGTGGCCTGCC
TGCAGGCCCTTTCCCTGGAGCTTGCGGAGGAGATCCGCCGCCTCGAGGAGGAGGTCTTCCGCTTGGCGGGCCACCC
CTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
GAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAGAAG
ATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGA
CGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
GAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGACCCGGGTCTTCC
AGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGA
TGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTGGC
CATCCCTTACGAGGAGGCCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGGATAGAA
AAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTACGTGCCCGACCTC
AACGCCCGGGTGAAGAGTGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
GACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCCAGGTCC
ACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATGGAGA
AGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTASRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHKA
YEAYRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLAPLAKKAEKEGFEVRILPAVRGLCPLVSDRVAVLL
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGKKTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVACLQALSLELAEEIRRLE
EEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAILEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIH
PRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLTRVFQEG
RDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGR
KRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELLLEA
PQARAEEVAALAKEAMEKAYPLAVPLEVKVGIGEDWLSAKE*

1B12:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAAGGCCGGGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCTCATCACGAGCCGGGGCGAACCGGTGCAGGCGGTCTACGGTTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACGAGG
CCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCAAGGCTACGAGGCGGACGACGTCCTCGCCACCCTGGCCAAG
AAGGCGGAAAAAGAAGGGTACGAGGTGCGCATCCTCACCGCCGACCGGGACCTCTACCAGCTCGTCTCCGACCGC
GTCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGC
AGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGA
CCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAATCTCCTCAAGAACCTGGATCGGGTAAAGCCGGAAA

ACGTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGTACCGACCT
CCCCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAAGGGCTTAGGGCCTTCCTGGAGAGGCTGGA
GTTCGGCAGCCTCCTCCATGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCG
GAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAG
GGGTGGTCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCC
AAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACC
TCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGG
AGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCT
TTACCGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCC
TATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCC
ACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGTTAGGGCTTCCCGCCATCGGCAA
GACGGAGAGGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGA
GAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCC
AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACC
TCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATT
GGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTC
TTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGTGTCCCCCCGGAGGCCGTGGACCCCC
TGATGCGCCGGGCGGCCAAGACGGTGAACTTCGGCGTCCTCTACGGCATGTCCGCCCATAGGCTCTCCCAGGAGCT
TTCCATCCCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGGATA
GAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTACGTGCCCGAC
CTCAACGCCCGGGTGAAGAGCGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCC
GCCGACCTCATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCCAGG
TCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATGG
AGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGGAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGT
GA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLITSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAY
EAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVQGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTERTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPPEAVDPLMRRAAKTVNFGVLYGMSAHRLSQELSIPYEEAVAFIERYFQSFPKVRAWIEKTLEEG
RKRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELLLE
APQARAEEVAALAKEAMEKAYPLAVPLEVEVGIGEDWLSAKE*

2H2:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCG
TCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
GTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCC
CCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
TGGCGGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAA
GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGG
TGGTCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
GACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
GGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTAC
CGGGAGGTGGAGAGGCCCCTTTCCGTTGTCCTGGCCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCCTATC
TCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCC
CTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
GAGAAGACCGGCAAGCGCTCCACCGGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCACCGTGGAGAAG
ATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGA
CGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGACCCCAACCTCCA
GAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCC
AGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTAA
TGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGC

CATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACATTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAG
AAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGTCGCTACGTGCCAGACCTA
GAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
GACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAAGAAACGGGGGCCAGGATGCTCCTTCAGGTCC
ACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGCCATGGAGG
GGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHKAY
EAYRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDRERLRAFLERLEFGGLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSVVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTGAAVLEALREAHPTVEKILQYRELTKLKSTYIDPLPDL
IHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYIQSFPKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEETGARMLLQVHDELVLEA
PKERAEAVARLAKEAMEGVYPLAVPLEVEVGIGEDWLSAKE*

1C8:

ATGCGTGGTATGCTTCCTCTTTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCAC
CTTCTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTC
CTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCGCCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACG
AGGCCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGG
AGCTGGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCTGGCTACGAGGCGGACGACGTCCTCGCCACCCTGGC
CAAGAAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGACCTCTACCAACTCGTCTCCGA
CCGCGTCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCG
GAGCAGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAG
AAGACCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCA
GAAAACGTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCG
ACCTCCCCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCT
TGAGTTTGGCGGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCG
CCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGC
CAGGGGTGGTCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTC
GCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCT
ACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGG
GGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTG
GCTTTACCGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACAGGGGTACGGCTGGACGTG
GCCTGCCTGCAGGCCCTTTCCCTGGAGCTTGCGGAGGAGATCCGCCGCCTCGAGGAGGAGGTCTTCCGCTTGGCGG
GCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGG
CAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGT
GGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCAC
CCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCA
ACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCT
ATTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGACCCGG
GTCTTCCAGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGAC
CCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGG
AGCTGGCCATCCCTTACGAGGAGGCCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTG
GATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTACGTGCC
CGACCTCAACGCCCGGGTGAAGAGTGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCAC
CGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCA
TGGAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGG
AGTGA

MRGMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQAVYGFAKSLLKALKEDGYKAAFVVFDAKAPSFRHEA
YEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVL
HPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIK
AHLEDLRLSLELSRVRTDLPLEVDLAQGREPDRERLRAFLERLEFGGLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRK
EPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVAR
RYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVACLQALSLELAEEIRRL
EEEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAILEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLTRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEG
RKRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELLLE
APQARAEEVAALAKEAMEKAYPLAVPLEVKVGIGEDWLSAKE*

2H10X:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGTGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCCCATTCCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCG
TCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
GTCCGGGAGAAGATCAAGGCCCACCTGGAAGATCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCC
CCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
TGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAA
GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGG
TGGTCGAGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
GGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTAC
CGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATC
TCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCC
CTTCAACCTCAACTCCCGGGACCAGCTGGAAATGGTGCTCTTTGACGAGCTTAGGCTTCCCGCCTTGGGGAAGACG
CAAAAGACGGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAGAAG
ATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGTCGGACCTCATCCACCCCAGGA
CGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
GAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTACTGGTG
GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAAAACCTGATCAGGGTCTTCC
AGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGA
TGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGC
CATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAG
AAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTA
GAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
GACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTCC
ACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGG
GGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPPFRHKAY
EAYRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLEMVLFDELRLPALGKTQKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLSDL
IHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLE
APKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE*

3A10

ATGGCGATGCTTCCCCTCTTTGAGCCCAAAGGCCGGGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCCTCATTCCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCCCAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCCGACCGC
GTCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGC
AGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGA
CCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAA
ACGTCCGGGAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCT
CCCCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGA
GTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCG
GAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAG
GGGTGGTCGAGTCCACCAGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCC
AAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACC
TCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGG
AGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCT
TTACCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGACCACGGGGGTGCGCCTGGACGTGGCC
TATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCC
GCCCCTTCAACCTCAACTCCCGAGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAA
GACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGA

GAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCC
AGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACC
TCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATT
GGTGGTCCTGGACTATAGCCAGATGGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCAGGGTC
TTCCAGGAGGGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCGGTGTCCCCCCGGAGGCCGTGGACCCCC
TGATGCGCCCGGGCGGCCAAGACGGTGAACTTCGGCGTCCTCTACGGCATGTCCGCCCATAGGCTCTCCCAGGAGCT
TTCCATCCCCTACGAGGAGGCGGTGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGGATT
GAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCCGAC
CTCAACGCCCGGATGAAGAGCGTCAGGGGGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCC
GCCGACCTCATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCCAGG
TCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATGG
AGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGGAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGT
GA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHKAY
EAYRAGRAPTPQDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHQAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVAR
RYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMETTGVRLDVAYLRALSLEVAEEIARL
EAEVFRLAGRPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDL
IHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQMELRVLAHLSGDENLIRVFQ
EGKDIHTQTASWMFGVPPEAVDPLMRRAAKTVNFGVLYGMSAHRLSQELSIPYEEAVAFIERYFQSFPKVRAWIEKTLE
EGRRRGYVETLFGRRRYVPDLNARMKSVRGAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELL
LEAPQARAEEVAALAKEAMEKAYPLAVPLEVEVGIGEDWLSAKE*

3B5

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGTGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTCCT
TCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
CCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCCCATTCCGCCACAAGGCCTACG
AGGCCTACAGGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCGTCAAGGAGCTGGTGGACC
TCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAGAAGGCGGAAA
AGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCGTCGCCGTCCTCCA
CCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCCG
CGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCAA
GGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAGG
CCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCCCCCTGGAGGTGGACCTCGCCCA
GGGGCGGGAGCCCGACCGGGAAAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCATGAGTTCGG
CCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCC
CGCAAGGCGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGTGGTCGGGTCTACCGGGCCCCCGAGCCT
TATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGC
CTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGG
CCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGT
GGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGCCC
ACATGGAGGCCACAGGGGTACGGCTGGACGTGGCCTGCCTGCAGGCCCTTTCCCTGGAGCTTGCGGAGGAGATCCGCC
GCCTCGAGGAGGAGGTCTTCCGCTTGGCGGGCCACACCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTT
TGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGC
CCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGA
CCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
GCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATC
GCCGAGGAGGGGTGGCTACTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCTCACCTCTCCGGCGAC
GAAAACCTGATCAGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGG
GAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGC
CTCTCCCAGGAGCTAGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGC
GGGCCTGGATTGAGAAGGCCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGAAGAAGGCGCTAC
GTGCCCGACCTCAACGCCCGGGTGAAGAGTGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGC
ACCGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATG
GAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGTGA
MAMLPLFEPKGRVLLVDGHHLAYRTSFALKGPTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPPFRHKAYEA
YRAGRAPTPEDFPRQLALVKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVSDRVAVLHPEGH
LITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKAHLEDLRL
SLELSRVRTDLPLEVDLAQGREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKAPMWADLLAL
AAARGGRVYRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGE
RAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVACLQALSLELAEEIRRLEEEVFRLAGHTFNLN
SRDQLERVLFDELGLPAIGKTEKTGKRSTSAAILEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATA
TGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREA
VDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKALEEGRRRGYVETLFGRRRYVPDLNA
RVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELLLEAPQARAEEVAALAKEAMEKAYPL
AVPLEVKVGIGEDWLSAKE*

3B6

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCGCCTTCT
TCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
CCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACGAGGCCTACG
AGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACC
TCCTGGGGTTTACCCGCCTCGAGGTCCAAGGCTACGAGGCGGACGACGTCCTCGCCACCCTGGCCAAGAAGGCGGAAA
AAGAAGGGTACGAGGTGCGCATCCTCACCGCCGACCGGGACCTCTACCAGCTCGTCTCCGACCGCGTCGCCGTCCTCC
ACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCC
GCGCCCTCGTGGGGGACCCCTCCAACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCA
AGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAG
GCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCCCCCTGGAGGTGGACCTCGCCC
AGGGGCGGGAGCTCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCGGCCTCCTCCACGAGTTCG
GCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTC
CCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGTGGTCGGGTCCACCGGGCCCCCGAGCC
TTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGG
CCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCGCCCCCGAGGGGGTG
GCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCT
GTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGC
CCACATGGAGGCCACAGGGGTACGGCTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGC
GCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGAGACCAGCTGGAAAGGGTCCTC
TTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAG
GCCCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATT
GACCCCTTGCCGAACCTCATCCATCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGC
AGGCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTC
ATCGCCGAGGAGGGGTGGCTATTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGC
GACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCC
CGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCCAC
CGCCTCTCCCAGGAGCTAGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGG
TGCGGGCCTGGATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGC
TACGTGCCCGACCTCAACGCCCGGGTGAAGGGCGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAG
GGCACCGCCGCCGACCTCATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTC
CTCCAGGTCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGGGGAGGTGGCGGCTTTGGCCAAGGAGGCC
ATGGAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAG
TGA

MAMLPLFEPKGRVLLVDGHHLAYRAFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHEAYEA
YKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVQGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVLHPEGH
LITPEWLWEKYGLRPEQWVDFRALVGDPSNNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKAHLEDLRL
SLELSRVRTDLPLEVDLAQGRELDRERLRAFLERLEFGGLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLA
LAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTAPEGVARRYGGEWTEEAG
ERAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNL
NSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPNLIHPRTGRLHTRFNQTA
TATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPRE
AVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLN
ARVKGVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELLLEAPQARAGEVAALAKEAMEKAY
PLAVPLEVKVGIGEDWLSAKE*

3B8

ATGGCGATGCTTCCCCTCTTTGAGCCCAAAGGCCGGGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCT
TCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
CCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCCTCCGCCACGAGGCCTACG
AGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCTCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACC
TCCTGGGGTTTACCCGCCTCGAGGTCCAAGGCTACGAGGCGGACGACGTCCTCGCCACCCTGGCCAAGAAGGCGGAAA
AAGAAGGGTACGAGGTGCGCATCCTCACCGCCGACCGGGACCTCTACCAGCTCGTCTCCGACCGCGTCGCCGTCCTCC
ACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCC
GCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCA
AGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGCTGAAGCCCGCCATCCGGGAGAAGATCCTGGCC
CACATGGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGTGCGCACCGACCTGCCCCTAGAGGTGGACTTCGCCAAA
AGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGCTTGGCAGCCTCCTCCACGAGTTCGGC
CTTCTGGAAAGCCCCAAGACCCTGGAGGAGGCCTCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCC
GCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTT
ATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCC
TTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGC
CCGGCGCTACGGCGGGGAGTGGACGAAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGT
GGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGCCC
ACATGGAGGCCACAGGGGTGCGCTTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCC
GCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCATCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTT
TGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGC
CCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGA
CCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
GCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCGTC

GCCGAGGAGGGGTGGCTATTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGAC
GAGAACCTGACCCGGGTCTTCCTGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCGG
GAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCCACCGC
CTCTCCCAGGAGCTGGCCATCCCTTACGAGGAGGCCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGC
GGGCCTGGATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTAC
GTGCCCGACCTCAACGCCCGGGTGAAGAGTGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGC
ACCGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATG
GAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGAGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGTG
A

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSLRHEAYEA
YKAGRAPTPEDFLRQLALIKELVDLLGFTRLEVQGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVLHPEGH
LITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRLKPAIREKILAHMDDLKLS
WDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLELGSLLHEFGLLESPKTLEEASWPPPEGAFVGFVLSRKEPMWADLLAL
AAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTKEAGE
RAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLN
SRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTAT
ATGRLSSSDPNLQNIPVRTPLGQRIRRAFVAEEGWLLVVLDYSQIELRVLAHLSGDENLTRVFLEGRDIHTETASWMFGVPRE
AVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLN
ARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELLLEAPQARAEEVAALAKEAMEKAYP
LAVPLEVKEGIGEDWLSAKE*

3B10
ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCT
TCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
CCCTGAAAGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGGCCTACG
AGGCCTACAGGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACC
TCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAGAAGGCGGAAA
AGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCGTCGCCGTCCTCCA
CCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCCG
CGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCAA
GGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAGG
CCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCCCCCTGGAGGTGGACCTCGCCCA
GGGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCGGCCTCCTCCACGAGTTCGG
CCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCC
CGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGTGGTCGGGTCCACCGGGCCCCTGAGCCT
TATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGC
CTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGG
CCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGT
GGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGACCCCTTTCCGCTGTCCTGGCCC
ACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCC
GCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGAGACCAGCTGGAAAGGGTCCTCTT
TGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGC
CCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGA
CCCCTTGCCGGACCACATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
GCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCATC
GCCGAGGAGGGGTGGCTATTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGAC
GAGAACCTGACCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCGG
GAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCCACCGC
CTCTCCCAGGAGCTGGCCATCCCTTACGAGGAGGCCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGC
GGGCCTGGATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTAC
GTGCCCGACCTCAACGCCCGGGTGAAGAGTGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGC
ACCGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTACCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATG
GAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCCAAGGAGTG
AGTCGACCTGCAGGCAGCGCTTGGCGTCACCCGCAGTTCGGTGGTTAA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHKAYEA
YRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVSDRVAVLHPEGHL
ITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKAHLEDLRLS
LELSRVRTDLPLEVDLAQGREPDRERLRAFLERLEFGGLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLAL
AAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGE
RAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLN
SRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDHIHPRTGRLHTRFNQTAT
ATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLTRVFQEGRDIHTETASWMFGVPRE
AVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLN
ARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLYPRLREMGARMLLQVHDELLLEAPQARAEEVAALAKEAMEKAYP
LAVPLEVKVGIGEDWLSAQGVSRPAGSAWRHPQFGG*

3C12

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCT
TCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
CCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGGCCTACG
AGGCCTACAGGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACC
TCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAGAAGGCGGAAA
AGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCGTCGCCGTCCTCCA
CCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTAGACTTCCG
CGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCAA
GGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAGG
CCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCCCCCTGGAGGTGGACCTCGCCCA
GGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGTTCGG
CCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCA
CGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGTGGTCGGGTCCACCGGGCCCCCGAGCCT
TATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGC
CTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCGCCCCCGAGGGGGTGG
CCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGT
GGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGATAGGCCCCTTTCCGCTGTCCTGGCCC
ACATGGAGGCCACAGGGGTACGGCTGGACGTGGCCTGCCTGCAGGCCCTTTCCCTGGAGCTTGCGGAGGAGATCCGCC
GCCTCGAGGAGGAGGTCTTCCGCTTGGCGGGCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTT
TGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGC
CCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGA
CCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
GCTAAGTAGCTCCGGTCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCGTC
GCCGAGGAGGGGTGGCTATTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGAC
GAGAACCTGACCCGGGTCTTCCTGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCGGCGTCCCCCGG
GAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCCACCGC
CTCTCCCAGGAGCTGGCCATCCCTTACGAGGAGGCCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGC
GGGCCTGGATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTAC
GTGCCCGACCTCAACGCCCGGGTGAAGAGTGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGC
ACCGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCCAAGCGCGGGCCGAGGAAGTGGCGGCTTTGGCCAAGGAGGCCATG
GAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHKAYEA
YRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVSDRVAVLHPEGHL
ITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKAHLEDLRLS
LELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLAL
AAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTAPEGVARRYGGEWTEEAGE
RAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVACLQALSLELAEEIRRLEEEVFRLAGHPFNLN
SRDQLERVLFDELGLPAIGKTEKTGKRSTSAAILEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATA
TGRLSSSGPNLQNIPVRTPLGQRIRRAFVAEEGWLLVVLDYSQIELRVLAHLSGDENLTRVFLEGRDIHTETASWMFGVPREA
VDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLNA
RVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLREMGARMLLQVHDELLLEAPQARAEEVAALAKEAMEKAYPL
AVPLEVKVGIGEDWLSAKE*

3D1
ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCT
TCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
CCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGGCCTACG
AGGCCTACAGGGCGGGGAGGGCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACC
TCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAGAAGGCGGAAA
AGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCGTCGCCGTCCTCCA
CCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCCG
CGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCAA
GGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAGG
CCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCCCCCTGGAGGTGGACCTCGCCCA
GAGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCTTCCACGAGTTCGG
CCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCC
CGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAGGGGTGGTCGAGTCCACCGGGCCCCCGAGCCT
TATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGC
CTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGG
CCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGT
GGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCC
ACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCC
GCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAATGGTGCTCTT
TGACGAGCTTAGGCTTCCCGCCTTGGGGAAGACGCAAAAGACGGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGC
CCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGA
CCCCTTGTCGGACCTCATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
GCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATC
GCCGAGGAGGGGTGGCTACTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGAC
GAAAACCTGATCAGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCCCGG

GAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGC
CTCTCCCAGGAGCTAGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGC
GGGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTAC
GTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGG
CACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGGAGAAACGGGGGCCAGGATGCTCCT
TCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGCCA
TGGAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGGTT
AG

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEPVQVVYGFAKSLLKALKEDGYKAVFVVFDAKAPSFRHKAYEA
YRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVSDRVAVLHPEGHL
ITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKAHLEDLRLS
LELSRVRTDLPLEVDLAQRREPDREGLRAFLERLEFGSLFHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALA
AARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGER
AALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNS
RDQLEMVLFDELRLPALGKTQKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLSDLIHPRTGRLHTRFNQTAT
ATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPRE
AVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLE
ARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLGETGARMLLQVHDELVLEAPKERAEAVARLAKEAMEGVYP
LAVPLEVEVGIGEDWLSAKG*

Example 11: Abasic site bypass by mismatch extension clone in PCR

A list of polymerases selected to extend four mismatches were assayed for their ability to extend abasic sites in PCR (Figure 10). C12 and D1, which can also extend four mismatched primers in PCR, as well as A10, B6 and B8, which cannot, all produced an amplification product.

Example 12: Abasic site bypass by mismatch extension clone in PCR

A list of polymerases selected to extend four mismatches were assayed for their ability to extend abasic sites in PCR (Figure 10). C12 and D1, which can also extend four mismatched primers in PCR, as well as A10, B6 and B8, which cannot, all produced an amplification product.

Example 13: Translesion synthesis activity by mismatch extension clone as determined by primer extension assays.

Seven polymerases were assayed for their ability to bypass abasic sites in a primer extension assay (Figure 11).

Primer extension assays were essentially as described in (Ghadessy et al., 2004). Briefly, undamaged oligonucleotides and a 51mer containing a synthetic abasic site were synthesized by Lofstrand Laboratories (Gaithersburg, MD) using standard techniques and were gel purified prior to use. A 20mer primer (LES_20P) with the sequence 5'-CGTGGTCGCGACGGATGCCG-3' was 5'-labeled with [^{32}P]ATP (5000 Ci/mmole; 1 Ci=37 GBq) (Pharmacia) using T4 polynucleotide kinase (Invitrogen, Carlsbad CA). Radiolabeled primer-template DNAs were prepared by annealing the 5'[^{32}P] labeled 20mer primer to one of the two following 51mer templates (at a primer template ratio of molar 1:1.5). 1) undamaged DNA (UNDT51T); 5'-AGC TAC CAT GCC TGC ACG AAT TCG GCA TCC GTC GCG ACC ACG GTC GCA GCG-3'; 2) an oligo (LABA51T) containing a synthetic abasic site (indicated as an X in bold font); 5'-AGC TAC CAT GCC TGC ACG ACA **X**CG GCA TCC GTC GCG ACC ACG GTC GCA GCG-3'. Standard replication reactions of 10 µl contained 40mM Tris•HCl at pH 8.0, 5mM $MgCl_2$, 100µM of each ultrapure dNTP (Amersham Pharmacia Biotech, NJ), 10mM DTT, 250 µg/ml BSA, 2.5% glycerol, 10 nM 5'[32P] primer-template DNA and 0.1 Unit of polymerase. After incubation at 60°C for various times reactions were terminated by the addition of 10 µl of 95% formamide/ 10 mM EDTA and the samples heated to 100°C for 5 min. Reaction mixtures (5 µl) were subjected to 20 % polyacrylamide/7 M Urea gel electrophoresis and replication products visualized by PhosphorImager analysis.

Polymerases A10 was the most active and was chosen for further analysis (Figure 26JRF nomenclature) on abasic sites and cyclobutane thymine-thymine dimers (CPD). A10 was clearly better at both abasic site and CPD extension and bypass than both wild type and M1.

Example 14: Error rate investigation of mismatch extension clones as determined by MutS ELISA.

Relaxed specificity might be expected to be achieved at the cost of lower fidelity. We used a MutS ELISa to investigate this possibility.

MutS is an *E. coli* derived mismatch binding protein that binds single base pair mismatches or small (1-4 base) additions or deletions. It can be used to monitor PCR fidelity in an ELISA based assay (Debbie et al., 1997).

Immobilised Mismatch Binding protein plates (Genecheck, Ft Collins, USA) were used for fidelity measurements as per manufacturer's instructions, essentially as described in (Debbie et al., 1997).

The mutation rate of D1 was compared that of wt*Taq* and M1 M1 was already known to have a modestly increased mutation rate (approximately 2 fold) (Ghadessy et al., 2004). The data presented here suggests that D1 has a 2 fold increased error rate compared to M1 and a four fold increased error rate compared to wt*Taq*. This corresponds approximately to a 1 in 2500 error ratio and is sufficiently low to not be problematic for many applications.

Example 15: Investigation of mismatch extension clones for the amplification of damaged DNA such as is found in ancient samples.

DNA recovered from ancient samples is invariably damaged, limiting the information it can yield. Polymerases that can bypass damage (such as abasic site or hydantoins) might therefore be useful in increasing the information that can be recovered from ancient samples of DNA.

Experiment 1: A mismatch extending polymerase can amplify previously un-amplifiable cave hyena DNA

Several samples of cave hyena (*Crocuta spelaea)* were extracted and analysed. Of those, seven samples (see Figure 12 for the list) failed to ever produce an amplification product.

These samples were chosen to test the efficacy of the expanded substrate spectrum polymerases.

M1 has a slightly reduced kcat/Km, 14% of *Taq* wild type, and is hence slightly less efficient in PCR. Therefore, M1 was blended with a commercial preparation of *Taq* (SuperTaq (HT biotechnology Ltd)) in a ratio of 1 unit to 10 and compared to Taq in the absence of M1. It was hoped that if M1 could bypass the blocking lesions, then the wild type *Taq* would amplify the resulting translesion synthesis product. On two

separate occasions, the M1/SuperTaq mix was able to produce an amplification product whereas SuperTaq alone did not (see figure 12 for one example)

The DNA was cloned and sequence and found to differ in two positions (A71→G, 77A →G) from the expected sequence. This could either be a miscoding lesion resulting from a deamination of C or a population variant sequence not seen previously in aDNA. Indeed, both mutations exist in modern spotted hyena (*Crocuta crocuta*), arguing for the second interpretation. Of the 10 sequences obtained from the same successful PCR, two each had a further unique single mutation, an A to G in different places. These are most likely errors incurred during amplification. Such errors are frequently seen in aDNA PCR and are one reason why multiple sequences need to be obtained from the same PCR product.

Contamination problems prevented an exhaustive analysis of the benefits of M1 polymerase. However, this result strongly suggested that a suitable altered polymerase could be usefully applied to aDNA.

Experiment 2: A blend of mismatch extending polymerase needs less ancient DNA for a successful PCR.

Polymerases that displayed interesting properties: B5, B8, C12 and D1, which can extend mismatches as well as A10, B6 and B10 which are proficient at abasic site bypass were purified. In order to keep the number of experiments manageable, they were blended in equal volumes with M1, SuperTaq and heparin purified wild-type

Taq. This mix of polymerases was used in almost all subsequent experiments and is referred to as the blend.

To ensure that no polymerase would negatively affect the PCR through its mutant activity, each one was individually blended with SuperTaq and used to perform an aDNA PCR with an ancient sample known to contain amplifiable DNA. All PCRs were successful (data not shown), indicating that it was unlikely that any of the mutant enzymes would be a liability in the blend.

The activity of the blend was checked against the activity of SuperTaq by a PCR activity dilution series. By this measure, the blend was less active than SuperTaq, by a factor of two.

The conditions that are usually used in aDNA PCR did not transfer readily to the blend or to SuperTaq as they had been optimised for AmpliTaqGold (Applied Biosystems), a chemically modified version of Taq that allows a hot start and slow enzyme release through heat activation. Manual hot starts are not advisable in aDNA analysis because opening the PCR tube outside the clean room prior to thermocycling carries a high risk of contamination. Furthermore, alternative hot start techniques could not be utilised either: antibodies used to inactivate wt*Taq* at low temperatures might not bind to the chimerical proteins selected from the Molecular Breeding library and hot start buffers proved ineffective (data not shown). A new two step nested PCR strategy was used. In the first step, the aDNA is amplified over 28 cycles with either SuperTaq or the blend. In the second step, the first PCR is diluted 20 fold in a secondary clean room and

amplified with SuperTaq using in-nested primers. This is the approach subsequently used to compare SuperTaq and the blend

Briefly, 2 µl of ancient sample were added to a 20 µl PCR in SuperTaq buffer (HT Biotech) with 1 µM of the appropriate primers (see Figure 13), 2µM of each deoxyribonucleoside triphosphate (dNTP) as well as 0.1 µl of SuperTaq or an equal volume of mutant polymerases and amplified for 28 cycles. This PCR was set up in a clean room following precautions appropriate for aDNA.. The first step PCR was then diluted 1 in 20 in a secondary clean room and thermocycled for a further 32 cycles with the same buffer and dNTPs conditions, using in-nested primers and SuperTaq. No template controls were used to test for contamination.

A two fold dilution series of aDNA with equal volumes of SuperTaq and the blend (and therefore approximately equal activities, with the blend slightly less active) was performed and repeated this four times

This experiment showed that the blend was more likely to produce a band at a lower concentration of aDNA than SuperTaq. This therefore represented the second experiment that indicated that the mismatch extension polymerases were more proficient at amplifying aDNA than wild-type Taq.

Experiment 3: The mismatch extension polymerases perform consistently better in ancient DNA PCR.

Sample heterogeneity and the inherent stochasticity of aDNA analysis make the interpretation of a single positive or negative PCR problematic. To address this, multiple PCRs of a same sample and count the number of successful PCR amplifications at a limiting sample dilution were performed. Comparison of SuperTaq with the blend would allowed a statistical analysis. As the amount of aDNA required for this type of approach is large, samples previously shown to be of high quality were chosen and tested at limiting dilutions to increase the amount of material available for analysis. A short target sequence was chosen to allow maximal dilutions.

This has the additional advantage that at a sufficiently high dilution, the undamaged DNA will have been diluted out, leaving only damaged template. In such conditions, the difference between a polymerase that can bypass blocking lesions and one that cannot should become clearly apparent.

A total of nine experiments at limiting amounts of aDNA, where the PCR would only be stochastically successful (Figures 14 and 15) were performed. In eight out of nine experiments, the blend resulted in more successful PCRs than SuperTaq. The probability of this occurring by chance is 1.76%, as determined by binomial distribution analysis. It is commonly accepted that chance can be dismissed as an explanation when an event is expected to occur at 5% probability or less.

We can therefore state that this effect is not due to chance and that the blend is repeatedly performing better than SuperTaq in the conditions of the experiment. This

proves beyond reasonable doubt that the mismatch extension polymerases are a more sensitive tool for the recovery of ancient DNA sequences.

Example 16: Selection of a polymerases capable of replicating the unnatural base analogue 5-nitroindol (5NI)

We selected for extension and bypass of 5NI directly from the polymerase chimera library described in example 8 using an analogous strategy to the mismatch selection using flanking primers (5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GCA 5NI-3', 5'-GTA AAA CGA CGG CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GC5NI-3') comprising 5NI (or a derivative) at their 3' ends. After round 3 , we used flanking primers (5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA 5NICG AGG GCA 5NI-3', 5'-GTA AAA CGA CGG CCA GTA CCA C5NIG AAC TGC GGG TGA CGC CAA GC5NI-3') comprising internal 5NI (or a derivative) as well as 3' terminal 5NI (or a derivative) to increase selection pressure for 5NI replication.

Five rounds of selection yielded a number of clones with greatly increased ability to replicate 5NI. Among the best clones were round 4 clone 4D11 and round 5 clone 5D4:

4D11: 5'-
ATGGCGATGCTTCCCCTCTTTGAGCCCAAAGGCCGGGTCCTCCTGGTGGACGGCCACCACCTGGCCTAC
CGCACCTTCTTCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCGGTGCAGGCGGTTTACGGCTTC
GCCAAGAGCCTCCTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAG
GCCCCCTCCTTCCGCCACGAGGCCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTC
CCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCAAGGC
TACGAGGCGGACGACGTCCTCGCCACCCTGGCCAAGAAGGCGGAAAAAGAAGGGTACGAGGTGCGCATC
CTCACCGCCGACCGGGACCTCTACCAGCTCGTCTCCGACCGCGTCGCCGTCCTCCACCCCGAGGGCCAC
CTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCCGCGCC
CTCGTGGGGGACCCCTCCGACAACCTCCCCGGGATCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTC
CTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAATGTCCGG
GAGAAGATCAAGGCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTC
CCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGG
CTTGAGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCC
TGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTT
CTGGCCCTGGCCGCCGCCAAGGGTGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGAC
TTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTC
CCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTG

GCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTC
GCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCC
CTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTG
TCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTC
AACCTCAACTCCCGAGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAG
ACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATC
GTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGAC
CTCATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTA
AGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCC
TTCATCGCCGAGGGGGGGTGGCTATTGGTGGTCCTGGACTATAGCCAGATGGAGCTCAGGGTGCTGGCC
CACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAAACCGCC
AGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAAC
TTCGGGGTTCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTACGAGGAGGCC
CAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAGAAGACCCTGGAG
GAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCC
CGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
GACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAACGGGGGCCAGGATGCTCCTT
CAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAG
GAGGTCATGGAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGG
CTCTCCGCCAAGGAGTGA-3'

4D11 amino acid sequence:

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDGYKAVFVVFDAK
APSFRHEAYEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVQGYEADDVLATLAKKAEKEGYEVRI
LTADRDLYQLVSDRVAVLHPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGIKGIGEKTALKL
LKEWGSLENLLKNLDRVKPENVREKIKAHLEDLRLSLELSRVRTDLPLEVDFAKRREPDRERLRAFLER
LEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAAKGGRVHRAPEPYKALRD
LKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLF
ANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPF
NLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPD
LIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEGGWLLVVLDYSQMELRVLA
HLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEA
QAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAA
DLMKLAMVKLFPRLEETGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDW
LSAKE*

5D4: 5'-
ATGGCGATGCTTCCCCTCTTTGAGCCCAAAGGCCGGGTCCTCCTGGTGGACGGCCACCACCTGGCCTAC
CGCACCTTCTTCGCCCTGAAGGGCCTCACCACGAGTCGGGGCGAACCGGTGCAGGCGGTCTACGGCTTC
GCCAAGAGCCTCCTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCATCTTCGTGGTCTTTGACGCCAAG
GCCCCCTCCTTCCGCCACGAGGCCCACGAGGCCTACAAGGCGGGGAGGGCCCCGAGCCCCGAGGACTTC
CCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCAAGGC
TACGAGGCGGACGACGTCCTCGCCACCCTGGCCAAGAAGGCGGAAAAAGAAGGGTACGAGGTGCGCATC
CTCACCGCCGACCGGGACCTCTACCAGCTCGTCTCCGACCGCGTCGCCGTCCTCCACCCCGAGGGCCAC
CTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCCGCGCC
CTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTC
CTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGCTGAAGCCCGCCATCCGGGAG
AAGATCCTGGCCCACATGGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGTGCGCACCGACCTGCCC
CTGGAGGTGGACTTCGCCAAAAGGCGGGAGTCCGATCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTT
GAGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGG
CCCCCGCCGGTAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTG
GCCCTGGCCGCCGCCAGGGGTGGTCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGAGACCTG
AAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCG
CCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGTGGTGGCC

CGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCC
AACCTGTGGGGGAGGCTTGAGGGGGAGGGGAGGCTCCTTTGGCTTTACCGGGGGGTGGAGAGGCCCCTT
TCCGCTGTCCTGGCCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCC
CTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAAC
CTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
GAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTG
GAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACTTACATTGACCCCTTGCCGGACCTC
ATCCACCCCAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGT
AGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTC
ATCGCCGAGGGGGGGTGGCTATTGGTGGTCCTGGACTATAGCCAGATGGAGCTCAGGGTGCTGGCCCAC
CTCTCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAAACCGCCAGC
TGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTC
GGGGTTCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTACGAGGAGGCCCAG
GCCTTCATTGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGGATAGAAAAGACCCTGGAGGAG
GGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTACGTGCCCGACCTCAACGCCCGG
GTGAAGAGCGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCGAC
CTCACGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAACGGGGGCCAGGATGCTCCTTCAG
GTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAG
GTCATGGAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTT
TCCGCCAAGGGTTAG-3'

5D4 amino acid sequence:

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEPVQAVYGFAKSLLKALKEDG
YKAIFVVFDAKAPSFRHEAHEAYKAGRAPSPEDFPRQLALIKELVDLLGFTRLEVQGY
EADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVLHPEGHLITPEWLWEKYGL
RPEQWVDFRALVGDPSDNLPGVKGIGEKTALKLLKEWGSLENLLKNLDRLKPAIREKI
LAHMDDLKLSWDLAKVRTDLPLEVDFAKRRESDRERLRAFLERLEFGSLLHEFGLLES
PKALEEAPWPPPVGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEA
RGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEVVARRYGGEWTEEAGERA
ALSERLFANLWGRLEGEGRLLWLYRGVERPLSAVLAHMEATGVRLDVAYLRALSLEVA
EEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEAL
REAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPN
LQNIPVRTPLGQRIRRAFIAEGGWLLVVLDYSQMELRVLAHLSGDENLIRVFQEGRDI
HTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERY
FQSFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLNARVKSVREAAERMAFNMPVQ
GTAADLTKLAMVKLFPRLEETGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYP
LAVPLEVEVGIGEDWLSAKG*

Example 17: Expanded spectrum of polymerases selected for replication of 5NI

Round 5 polymerases selected for replication of 5NI were tested for activity with a range of substrates using the hairpin ELISA assay described in example 8. tUTP and ceATP were kind gifts from the laboratory of P. Herdewijin, Rega Institute, Katholieke Universiteit Leuven, Belgium. Results are shown in Figure 14

1. ELISA with tUTP:

The ability of round 5 clones selected for 5NI replication extension to sequentially incorporate 2 or 3 of the TNA UTP derivative (3', 2')-beta-L-threonyl–UTP was assayed using the hairpin primers (ELISAT2p: 5'-TAG CTC GGT AA CGC CGG CTT CCG TCG CGA CCA CGT TX TTC GTG GTC GCG ACG GAA GCC G-3' , ELISAT3p: 5'-TAG CTC GGT AAA CGC CGG CTT CCG TCG CGA CCA CGT TX TTC GTG GTC GCG ACG GAA GCC G-3' (X=dU-biotin (Glen research)). The lysates used were concentrated 4-fold. ELISA protocol was a described except that The DIG labelled dUTP in the extension reaction was replaced with Fluorescein 12-dATP (Perkin-Elmer) (at 3% of dATP) and the incorporation of Fluorescein 12-dATP was detected by anti-Fluorescein-POD Fab fragments (Roche).

2. ELISA with ceATP:

The ability of round 5 clones selected for 5NI replication extension to sequentially incorporate the cyclohexenyl ATP derivative ceATP was assayed using the hairpin primers (ELISA2p: 5'-TAG CTC GGA TTTT CGC CGG CTT CCG TCG CGA CCA CGT TX TTC GTG GTC GCG ACG GAA GCC G-3' , (X=dU-biotin (Glen research)). The lysates used were concentrated 4-fold.

3. ELISA with CyDye 5-dCTP and CyDye 3-dCTP:

The ability of round 5 clones selected for 5NI replication extension to sequentially incorporate the fluorescent dye-labelled nucleotides Cy5-dCTP and Cy3-dCTP (Amersham Biosciences) was assayed using the hairpin primers (ELISA2p: 5'- TAG CTA CCA GGG CTC CGG CTT CCG TCG CGA CCA CGT TXT TCG TGG TCG CGA CGG AAG CCG -3' , (X=dU-biotin (Glen research)). The lysates used were concentrated 4-fold.

4. Basic site bypass ELISA

The ability of round 5 clones selected for 5NI replication extension to bypass an abasic site was assayed using the hairpin primer (PScreen1abas: 5'-AGC TAC CAT GCC TGC ACG CAG YCG GCA TCC GTC GCG ACC ACG TTX TTC GTG GTC GCG

ACG GAT GCC G -3' , (X=dU-biotin, Y=abasic site (Glen research)). The lysates used were concentrated 4-fold.

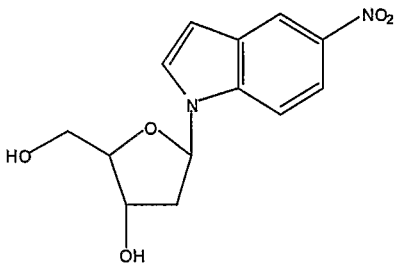
Example 18: Primer extension reaction with polymerases 4D11 and 5D4

1: Extension opposite 5-nitroindole.

Primer 5'- TAATACGACTCACTATAGGGAGA

Template 3'- ATTATGCTGAGTGATATCCCTCT5ATCGAT

5 = 5-Nitroindole



Primer extension reactions were carried out as follows:

50pmol of ^{32}P-labelled primer and 100pmol of template in a volume of 44µl were annealed in 1X Taq buffer. 4D11 or 5D4 polymerase as cell lysate (6µl) was added and reactions were incubated at 50°C for 15 minutes followed by addition of one dNTP (1µl in total volume of 50µl, final dNTP concentration 40µM). 8µl samples were taken at various time points and added to 8µl stop solution (7M urea, 100mM EDTA containing xylene cyanol F). At the end of the time course the remaining 3 dNTPs were added (final concentration each dNTP 40µM) and reactions incubated at 50°C for a further 30 minutes. Reaction samples were electrophoretically separated using 20% polyacrylamide gels at 25W for 4 hours. The resultant gels were dried and scanned using a phosphorimager (Molecular Dynamics). Data was processed using the program ImageQuant (Molecular Dynamics). Results are shown in Figures 35, 36:

Similar reactions using Taq, Tth and Tfl wild-type polymerases under identical conditions leads to almost undetectable extension reactions (data not shown).

2. Incorporation and extension of 5-nitroindole-5'-triphosphate (5NITP).

5NITP ↓

Primer 5'- TAATACGACTCACTATAGGGAGA

Template 3'- ATTATGCTGAGTGATATCCCTCTXGTCA

X = A, T, C, G

Primer extension reactions were carried out as follows:

50pmol of ^{32}P-labelled primer and 100pmol of template in a volume of 44µl were annealed in 1X Taq buffer. 4D11 or 5D4 polymerase as cell lysate (6µl) was added and reactions were incubated at 50°C for 15 minutes followed by addition of d5NITP (1µl in total volume of 50µl, final dNTP concentration 40µM). 8µl samples were taken at various time points and added to 8µl stop solution (7M urea, 100mM EDTA containing xylene cyanol F). At the end of the time course the 4 native dNTPs were added (final concentration each dNTP 40µM) and reactions incubated at 50°C for a further 30 minutes. Reaction samples were electrophoretically separated using 20% polyacrylamide gels at 25W for 4 hours. The resultant gels were dried and scanned using a phosphorimager (Molecular Dynamics). Data was processed using the program ImageQuant (Molecular Dynamics). Results are shown in Figures 17, 18) :

The NI-NI self-pair is also formed exceptionally well, though further extension is reduced (data not shown). Similar reactions using Taq, Tth and Tfl wild-type polymerases under identical conditions leads to almost undetectable extension reactions (data not shown).

Example 19: Array manufacture and hybridization using M1.

Targets were prepared by PCR amplification of 2.5kb *Taq* gene using primers 29, 28 or 2kb of the HIV *pol* gene using primers 30, 31. Salmon sperm DNA (Invitrogen) was prepared at 100ng/ul in 50% DMSO. FITC and Cy5 probes were prepared by PCR amplification of 0.4kb fragment of Taq using primers 8, 28 with either 100% ($FITC100_{M1}$) or 10% of dATP ($FITC10_{M1}$, $FITC10_{Taq}$) replaced by FITC-12-dATP or 10% of dCTP replaced by Cy5-dCTP ($Cy5_{Taq}$). Cy5 and Cy3 random 20mers (MWG) were used at 250nM. Targets were purified using PCR purification kit (Qiagen) and prepared in 50% DMSO and spotted onto GAPSII aminosilane-coated glass slides (Corning) using a MicroGrid (BioRobotics). Array hybridizations were performed according to standard protocols :

Printed slides were baked for 2hr at 80°C, incubated with agitation for 30min at 42°C in 5x SSC/0.1% BSA Fraction V (Roche)/0.1% SDS, boiled for 2min in ultrapure water, washed 20x in ultrapure water at room temperature (RT), rinsed in propan-2-ol and dried in a clean airstream. 50ng of FITC- and Cy5-labelled probes were prepared in 20µl of hybridization buffer (1mM Tris-HCl pH7.4, 50mM tetrasodium pyrophosphate, 1x Denhardts solution, 40% deionised formamide, 0.1% SDS, 100µg/ml sheared salmon sperm DNA). Each sample was heated to 95°C for 5min, centrifuged for 2min, applied to the surface of an array and covered with a 22x22 mm HybriSlip (Sigma). Hybridizations were performed at 48°C for 16hr in a hybridization chamber (Corning). Arrays were washed once with 2xSSC/0.1% SDS at 65°C for 5min once with 0.2xSSC at RT for 5min and twice with 0.05xSSC at RT for 5min. Slides were dried in a clean airstream, scanned

with an ArrayWoRx autoloader (Applied Precision Instruments) and the array images analysed using SoftWoRx tracker (Molecularware).

Complete substitution of natural nucleotides with their unnatural counterparts altered the properties of the resulting amplification products. For example, fully alphaS substituted DNA was completely resistant to nuclease digestion (not shown).

The 0.4kb fragment, in which all adenines (dA) on both strands had been replaced with FITC-12-dAMP ($FITC100_{M1}$), displayed extremely bright fluorescence. The frequency of fluorophore incorporation per 1000 nucleotides (FOI) is commonly used to specify the fluorescence intensity of a probe. FOIs of microarray probes commonly range from 10-50, while $FITC100_{M1}$ has an FOI of 295. To investigate if such a high level of fluorophore substitution would affect hybridisation characteristics we performed a series of microarray experiments. We compared the fluorescent signal generated by $FITC100_{M1}$ with equivalent probes generated using either wt*Taq* or *M1* and replacing only 10% of dAMP with FITC-12-dAMP ($FITC10_{Taq}$, $FITC10_{M1}$ (FOI=30)). In competitive co-hybridisation with a standard Cy5-labelled probe ($Cy5_{Taq}$), $FITC100_{M1}$ hybridised specifically only with its cognate *Taq* polymerase target sequence and not with any non-cognate control DNA. Hybridisation of $FITC100_{M1}$ generated an up to 20-fold higher specific signal than equimolar amounts of the FITC10 probes (**Fig. 20**) without showing increased background binding (**Figs. 19, 21**).

Example 20 : Mutation rates & spectra of selected polymerases M1 and M4. Mutation rates were determined using the mutS ELISA assay[26] (Genecheck, Ft. Collins, CO) according to manufacturers instructions. Alternatively, amplification products derived from 2x50 cycles of PCR of 2 targets with different GC content (HIV *pol* (38% GC), *Taq* (68% GC)) were cloned, 40 clones (800bp each) were sequenced and mutations (wt*Taq* (51), *M1* (75)) analyzed.

Promiscuous mismatch extension might be expected to come at the price of reduced fidelity, as misincorporation no longer leads to termination. Measurement of the overall mutation rate using both the MutS assay (Fig 22A) and direct sequencing of amplification products, however, indicated an only modestly (1.6 fold) increased mutation rate in *M1* (or *M4*). However, *M1* displays a significantly altered mutation spectrum compared to wt*Taq*, with a clearly increased propensity for transversions, in particular G/C->C/G transversions (Fig 22B).

Example 21: Processivity

Naturally occurring translesion polymerases are mostly poorly processive. We therefore investigated, if processivity of M1 and M4 was similarly reduced but found that, even at the lowest enzyme concentrations, primer extension and termination probabilities by *M1* and M4 closely matched those of wtTaq (Fig.23), indicating that both *M1* and *M4* exhibit processivity equal (or higher) than wt*Taq*. This is also reflected in the striking proficiency of M1 in long-range PCR (see example 6).

Processivity was measured using a primer extension assay the presence and absence of trap DNA. Termination probabilities were calculated according to the method of Kokoska et al.

Oligonucleotide primer 32 (5'-GCG GTG TAG AGA CGA GTG CGG AG-3') was ^{32}P-labelled and annealed to the template 33 (5'-CTC TCA CAA GCA GCC AGG CAA GCT CCG CAC TCG TCT CTA CAC CGC TCC GC-3') (at a primer/template ratio of molar 1/1.5). wt*Taq* (0.0025nM; 0.025nM; 0.25nM), *M1*(0.05nM; 0.5nM; 5nM), and *M4* (0.05nM; 0.5nM; 5nM) were preincubated with the primer-template DNA substrates (10nM) in 10mM Tris-HCl at pH 9.0, 5mM $MgCl_2$, 50mM KCl, 0.1% Triton X 100 at 25°C for 15min. Reactions were initiated by addition of 100μM dNTPs with or without trap DNA (1000-fold excess of unlabeled primer-templates). Reactions were performed at 60°C for 2min. Preincubation of polymerases with the trap DNA substrate and labelled primer-template before the addition of dNTPs completely abolished primer extension (not shown) demonstrating trap effectiveness. Thus, in the presence of trap DNA, all DNA synthesis resulted from a single DNA binding event. Gel band intensities were calculated using a Phosphoimager and ImageQuant (both Molecular Dynamics) software. Percentage of polymerase molecules, which extended primers to the end of the template was calculated using the formula: In x 100% / (I1+I2+...+In), where In is the intensity of the band at position 22 or 23; I1, I2... is the intensity of the band at position 1, 2... Termination probabilities (τ) were calculated according to the method of Kokoska *et al*[1], whereby τ at a particular template position was calculated as the intensity of the band at this position divided by the sum of the intensity of this band and the band intensities of all longer products.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the present invention will be apparent to those skilled in the art without departing from the scope and spirit of the present invention. Although the present invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in biochemistry, molecular biology and biotechnology or related fields are intended to be within the scope of the following claims.

References

1. Schaaper, R. M. (1993) *J. Biol. Chem.* **268,** 23762-23765.
2. Li, Y., Korolev, S. & Waksman, G. (1998) *Embo J.* **17,** 7514-7525.
3. Doublié, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998) *Nature* **391,** 251-258.
4. Johnson, S. J., Taylor, J. S. & Beese, L. S. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 38895-38900.
5. Li, Y., Mitaxov, V. & Waksman, G. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9491-9496.
6. Astatke, M., Ng, K., Grindley, N. D. & Joyce, C. M. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 3402-3407.
7. Patel, P. H. & Loeb, L. A. (2000) *J. Biol. Chem.* **275,** 40266-40272.

8. Jestin, J. L., Kristensen, P. & Winter, G. (1999) *Angew. Chem. Int. Ed.* **38,** 1124-1127.

9. Xia, G., Chen, L., Sera, T., Fa, M., Schultz, P. G. & Romesberg, F. E. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 6597-6602.

10. Ghadessy, F. J., Ong, J. L. & Holliger, P. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4552-4557.

11. Tawfik, D. S. & Griffiths, A. D. (1998) *Nature Biotechnol* **16,** 652-656.

12. Huang, M.-M., Arnheim, N. & Goodman, M. F. (1992) *Nucleic Acids Res.* **20,** 4567-4573.

13. Kool, E. T. (2000) *Curr. Op. Chem. Biol.* **4,** 602-608.

14. Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. & Sninsky, J. J. (1990) *Nucleic Acids Res* **18,** 999-1005.

15. Eom, S. H., Wang, J. & Steitz, T. A. (1996) *Nature* **382,** 278-281.

16. Creighton, S., Bloom, L. B. & Goodman, M. F. (1995) *Meth. Enzymol.* **262,** 232-56.

17. Mendelman, L. V., Petruska, J. & Goodman, M. F. (1990) *J. Biol. Chem.* **265,** 2338-2346.

18. Boudsocq, F., Iwai, S., Hanaoka, F. & Woodgate, R. (2001) *Nucleic Acids Res* **29,** 4607-4616.

19. Verma, S. & Eckstein, F. (1998) *Annu. Rev. Biochem.* **67,** 99-134.

20. Loakes, D. (2001) *Nucleic Acids Res* **29,** 2437-2447.

21. Berger, M., Wu, Y., Ogawa, A. K., McMinn, D. L., Schultz, P. G. & Romesberg, F. E. (2000) *Nucleic Acids Res,* 2911-2914.

22. Barnes, W. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 2216-2220.

23. Goodman, M. F. (2002) *Annu. Rev. Biochem.* **71,** 17-50.

24. Kunkel, T. A. & Bebenek, K. (2000) *Annu. Rev. Biochem.* **69,** 497-529.

25. Patel, P. H., Suzuki, M., Adman, E., Shinkai, A. & Loeb, L. A. (2001) *J. Mol. Biol.* **18,** 823-837.

26. Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S. Y., Landre, P. A., Abramson, R. D. & Gelfand, D. H. (1993) *PCR Meth. Appl.* **2,** 275-87.

27. Tada, M., Omata, M., Kawai, S., Saisho, H., Ohto, M., Saiki, R. K. & Sninsky, J. J. (1993) *Cancer Res* **53,** 2472-2474.

28. Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. (2001) *Cell* **107,** 91-102.

29. Trincao, J., Johnson, R. E., Escalante, C. R., Prakash, S., Prakash, L. & Aggarwal, A. K. (2001) *Mol. Cell* **8,** 417-426.

30. Cho, Y. S., Zhu, F. C., Luxon, B. A. & Gorenstein, D. G. *J Biomol Struct Dyn* **11,** 685-702.

31. Eigen, M. (1971) *Naturwissenschaften* **58,** 465-523.

32. Engelke, D. R., Krikos, A., Bruck, M. E. & Ginsburg, D. (1990) *Anal. Biochem.* **191,** 396-400.

33. Zhao, H., Giver, L., Shao, Z., Affholter, J. A. & Arnold, F. H. (1998) *Nature Biotechnol.* **16,** 258-61.

34. Murata, T., Iwai, S. & Ohtsuka, E. (1990) *Nucleic Acids Res* **18,** 7279-7286.

CLAIMS

1. A method for the generation of an engineered DNA polymerase with an expanded substrate range which comprises the step of preparing and expressing nucleic acid encoding an engineered DNA polymerase wherein the method comprises the use of template nucleic acid and primers which bear one or more distorting 3' termini.

2. A method for the generation of a engineered DNA polymerase with an expanded substrate range which comprises the steps of:

(a) preparing nucleic acid encoding a engineered DNA polymerase, wherein the polymerase is generated using a repertoire of nucleic acid molecules encoding one or more DNA polymerases and primers which bear distorting 3 termini.

(b) compartmentalising the nucleic acid of step (a) into microcapsules;

(c) expressing the nucleic acid to produce their respective DNA polymerase within the microcapsules;

(d) sorting the nucleic acid encoding the engineered DNA polymerase which exhibits an expanded substrate range; and

(e) expressing the engineered DNA polymerase which exhibits an expanded substrate range.

3. A method according to claim 1 or claim 2 wherein the distorting 3' terminus of the primers is effected by any one or more of the following techniques: the presence of one nucleotide mismatch bases at the 3'end of the one or more primers (primer mismatch disortion); the presence of two nucleotide mismatch bases at the 3'end of the one or more flanking primers (primer mismatch disortion); the presence of three nucleotide mismatch bases at the 3'end of the one or more flanking primers (primer mismatch disortion); the presence of four nucleotide mismatch bases at the 3'end of the one or more flanking primers (primer mismatch disortion) and the presence of one or more unnatural bases at the 3' end of one or more flanking primers or combinations thereof.

4. A method according to claim 3 wherein the distorting 3' terminus of the primers is effected by the presence of one or more unnatural bases at the 3' end of one or more flanking primers; wherein the at least one unnatural base is 5nitroindole triphosphate (5NITP).

5. A method according to claim 4 wherein the flanking primers bear one or more nucleotide mismatches at their 3' end.

6 A method according to any preceding claim wherein the engineered DNA polymerase is a pol A DNA polymerase.

7. A method according to claim 6 wherein the pol A polymerase is generated from one or more repertoires of randomly mutated Taq genes.

8. A method according to claim 6 wherein the pol A polymerase is generated from repertoires generated by recombining related pol A genes.

9 A method according to claim 7 wherein the pol A polymerase is generated from repertoires generated by recombining one or more polA genes selected from the group consisting of the following: Thermus aquaticus (Taq), Thermus thermophilus (Tth) and Thermus flavus (Tfl).

10 A method according to any of claims 3 to 9 wherein the pol A polymerase is generated from repertoires of Taq genes and wherein the pol A DNA polymerase is generated using 3' mismatch flanking primers selected from the group consisting of the following: 5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GA-3'; **A•G** mismatch);5'GTA AAA CGA CGG CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GCC-3' C*C mismatch.

11. A method according to any of claims 3 to 9 wherein the pol A polymerase is generated from repertoires of Taq genes or from repertoires of chimeras of Taq , Tth and Tfl genes wherein the pol A DNA polymerase is generated using flanking primers with four mismatches at their 3' end (underlined) selected from the group consisting of the following: 5'-CAG GAA ACA GCT ATG ACA AAA GTG AAA TGA ATA GTT CGA CTTTT-3'; 5'-GTA AAA CGA CGG CCA GTC TTC ACA GGT CAA GCT TAT TAA GGTG-3' ; 5'-CAG GAA ACA GCT ATG ACC ATT GAT AGA GTT ATT TTA CCA CAGGG-3'; 5'-GTA AAA CGA CGG CCA GTC TTC ACA GGT CAA GCT TAT TAA GGTG-3'

12. A method according to any of claims 3 to 9 wherein the pol A polymerase is generated from repertoires of Taq genes or from repertoires of chimeras of Taq , Tth and Tfl genes wherein the pol A DNA polymerase is generated using flanking primers with containing unnatural base analogues at their 3' end (X), with X for example being 5-nitroindole or derivatives selected from the group consisting of the following: 5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA ACG AGG GCA X-3', 5'-GTA AAA CGA CGG CCA GTA CCA CCG AAC TGC GGG TGA CGC CAA GCX-3') ot primers with containing unnatural base analogues at their 3' ends and internally (X), with X for example being 5-nitroindole or derivatives, selected from the group consisting of the following: (5'-CAG GAA ACA GCT ATG ACA AAA ATC TAG ATA XCG AGG GCA X-3', 5'-GTA AAA CGA CGG CCA GTA CCA CXG AAC TGC GGG TGA CGC CAA GCX-3') .

13. A method for the generation of an engineered DNA polymerase with an expanded substrate range according to any of claims 1 to 12 which comprises the step of preparing and expressing nucleic acid encoding a blend of engineered polymerases.

14. An engineered DNA polymerase which exhibits an expanded substrate range.

15 An engineered DNA polymerase which exhibits an expanded substrate range obtainable using the method of any preceding claim.

16. A engineered DNA polymerase which exhibits an expanded substrate range which is obtained using the method of claim 12.

17. An engineered DNA polymerase according to claim 15 or claim 16 which is an engineered pol A polymerase.

18. A pol A DNA polymerase with an expanded substrate range, or the nucleic acid encoding it, wherein the DNA polymerase is designated M1 or M4 as shown in fig 1 and 2 respectively and depicted as SEQ No 1 and SEQ No 2 respectively.

19. A pol A DNA polymerase with an expanded substrate range, wherein the polymerase exhibits at least **95%** identity to one or more of the amino acid sequences designated M1 and M4 as shown in fig 1 and fig 2 respectively and depicted SEQ No 1 and SEQ No 2 respectively and which comprises any one or more of the following mutations: E520G, D144G, L254P, E520G, E524G, N583S, 1.1- D144G, L254P, E520G, E524G, N583S, V113I, A129V, L245R, E315K, G364D, G403R, E432D, P481A, I614M, R704W, D144G, G370D, E742G, K56E, I63T, K127R, M317I, Q680R, R343G, G370D, E520G, G12A, A109T, D251E, P387L, A608V, R617K, D655E, T710N, E742G, A109T, D144G, V155I, P298L, G370D, I614M, E694K, R795G, E39K, R343G, G370D, E520G, T539A, M747V, K767R, G84A, D144G, K314R, E520G, F598L, A608V, E742G, D58G, R74P, A109T, L245R, R343G, G370D, E520G, N583S, E694K, A743P.

20. A pol A DNA polymerase with an expanded substrate range, in particular which is capable of mismatch extension, wherein the DNA polymerase comprises, preferably consists of the amino acid sequence of any one or more of the clones designated herein as 3B5, 3B8, 3C12 and 3D1.

21. A pol A DNA polymerase with an expanded substrate range, in particular which is capable of abasic site bypass, wherein the DNA polymerase comprises, preferably consists of the amino acid sequence of any one or more of the clones designated herein as 3A10, 3B6 and 3B11.

22. A pol A DNA polymerase with an expanded substrate range, in particular which is capable of DNA replication involving the incorporation of unatural base analogues into the newly replicated DNA, wherein the pol A DNA polymerase comprises, preferably consists of the amino acid sequence of any one or more of the clones designated herein as 4D11 and 5D4.

23. A pol A DNA polymerase with an expanded substrate range, wherein the polymerase exhibits at least **95%** identity to one or more of the amino acid sequences designated 3B5, 3B8, 3C12, 3D1, 3A10, 3B6, 3B11, 4D11 and 5D4. which comprises any one or more of the mutations (with respect to either of the three parent genes Taq, Tth, Tfl) or gene segments found in clones 3B5, 3B8, 3C12, 3D1, 3A10, 3B6, 3B11, 4D11 and 5D4.

24. A nucleic acid construct encoding an engineered polymerase according to any of claims 15 to 23.

25. A nucleic acid construct encoding an engineered pol A DNA polymerase which exhibits an expanded substrate range, wherein said pol A DNA polymerase is depicted in fig 1 and fig 2 respectively as SEQ No 1 or SEQ No 2 and is designated M1 and M4 respectively.

26. A vector comprising a nucleic acid construct according to claim 24 or claim 25..

27. The use of an engineered DNA polymerase according to any of claims 15 to 23 in any one or more of the following applications selected from the group consisting of the following: PCR amplification, sequencing of damaged DNA templates, the incorporation of unatural base analogues into DNA and the creation of novel polymerase activities.

28. The use of a blend of engineered polymerases according to claim 27.

29. The use according to claim 27 or claim 28 wherein the engineered polymerase is derived from at least Taq polymerase.

30. The use according to claim 27 or claim 28 wherein the engineered pol A polymerase is M1 or M4 as depicted in fig 1 and fig 2 and designated SEQ No 1 and SEQ No 2 respectively.

31. The use according to claim 27 or claim 28of M1 DNA polymerase as depicted in fig 1 and designated SEQ No 1.

32. The use according to claim 27 or claim 28 of any one or more pol A polymerases in the list consisting of the following: 3B5, 3B8, 3C12, 3D1, 3A10, 3B6, 3B11, 4D11 and 5D4.

Figure 1a shows M1 nucleic acid sequence

ATGCTCCCTCTTTTTGAGCCCAAAGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCT
ACCGCACCTTCCACGCCCTGAAGGGCCTCACCACCAGCCGGGGGGAGCCGGTGCAGGCGG
TCTACGGCTTCGCCAAGAGCCTCCTCAAGGCCCTCAAGGAGGACGGGGACGCGGTGATCG
TGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACGAGGCCTACGGGGGGTACAAGGCGG
CCCGGGCCCCCCACGCCGGAGGACTTTCCCCGGCAACTCGCCCTCATCAAGGAGCTGGTGG
ATCTCCTGGGGCTGGCGCGCCTCGAGGTCCCGGGCTACGAGGCGGACGACGTCCTGGCCA
GCCTGGCCAAGAAGGCGGAAAAGGAGGGCTACGAGGTCCGCATCCTCACCGCCGACAAA
GGCCTTTACCAGCTCCTTTCCGACCGCATCCACGTCCTCCACCCCGAGGGGTACCTCATCA
CCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCAGTGGGCCGACTACCGGG
CCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGGCATCGGGGAGAAGACG
GCGAGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCAAGAACCTGGACCG
GCTGAAGCCCGCCATCCGGGAGAAGATCCTGGCCCACATGGACGATCTGAAGCTCTCCTG
GGATCTGGCCAAGGTGCGCACCGACCTGCCCCTGGAGGTGGACTTCGCCAAAAGGCGGGA
GCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCA
CGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGA
AGGGGCCTTCGTGGGCTTTGTCCTTTCCCGCAGGGAGCCCATGTGGGCCGATCTTCTGGCC
CTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGG
GACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAA
GGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCA
ACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGG
GAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAG
GAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACA
TGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGG
CCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACC
TCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCG
GCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGGGGCCCTCCGC
GAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGC
ACCTACATTGACCCCTTACCGGACCTCATCCACCCCAGGACGGGCCGCCTCCACACCCGCT
TCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCAGAACA
TCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGT
GGCTATTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCG
GCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCC
AGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAG
ACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCA
TCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCG
GGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCT
TCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGGGGCG
GCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCGACCTCATGAAGCTG
GCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTC
CACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGC
CAAGGAGGTCATGGAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGAT
AGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

FIG. 1a

SEQ 1

M1 amino acid sequence:

MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRH
EAYGGYKAARAPTPEDFPRQLALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKGLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRREPMWADLLALALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLGALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIA
EEGWLLVVLDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVRGAAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE

FIG. 1b

Figure 2a shows M4 nucleic acid sequence

ATGCTCCCTCTTTATGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCT
ACCGCACCTTCCACGCCCTGAAGGGCCTCACCACCAGCCGGGGGGAGCCGGTGCAGGCGG
TCTACGGCTTCGCCAAGAGCCTCCTCAAGGCCCTCAAGGAGGGCGGGGACGCGGTGATCG
TGGTCTTTGACGCCAAGGCCCCCTCCTTCCCCCATGAGGCCTACGGGGGGTACAAGGCGG
GCCGGGCCCCCACGCCGGAGGACTTTCCCCGACAACTCGCCCTCATCAAGGAGCTGGTGG
ACCTCCTGGGGCTGACGCGCCTCGAGGTCCCGGGCTACGAGGCGGACGACGTCCTGGCCA
GCCTGGCCAAGAAGGCGGAAAAGGAGGGCTACGAGGTCCGCATCCTCACCGCCGACAAA
GACCTTTACCAGCTCCTTTCCGACCGCATCCACGTCCTCCACCCCGAGGGGTACCTCATCA
CCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCAGTGGGCCGACTACCGGG
CCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGGCATCGGGGAGAAGACG
GCGAGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCAAGAACCTGGACCG
GCTGAAGCCCGCCATCCGGGAGAAGATCCTGGCCCACATGGACGATCTGAAGCTCTCCTG
GGACCGGGCCAAGGTGCGCACCGACCTGCCCCTGGAGGTGGACTTCGCCAAAAGGCGGG
AGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCC
ACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGG
AAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTAGC
CCTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCGG
GGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGA
AGGCCTTGGCCTCCCGCCCGACGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCC
AACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCAGG
GGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGA
GGAAAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCAC
ATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTG
GCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAAC
CTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATC
GGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGGGGCCCTCCG
CGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAG
CACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCCTCCACACCCGC
TTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCAGAGC
ATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGG
TGGCTATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCG
GCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCC
AGCTGGATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAG
ACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCA
TCCCTTACGAGGAGGCCCAGGCCTTCATTAAGCGCTACTTTCAGAGCTTCCCCAAGGTGCG
GGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCT
TCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGCCG
GCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGTACCGCCGCCGACCTCATGAAGCTG
GCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTC
CACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGC
CAAGGAGGTCATGGAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGAT
AGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

FIG. 2a

SEQ 2

M4 amino acid sequence:

MRGMLPLYEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEGGDAVIVVFDAKAPSFPH
EAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLTRLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLS
DRIHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLLEEWGSLEALLKNLDRLK
PAIREKILAHMDDLKLSWDRAKVRTDLPLEVDFAKRREPDRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALGDLKEARGLLAKDLSVLALREGLGLPPDDDPML
LAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVR
LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLGALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQSIPVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA
HRLSQELAIPYEEAQAFIKRYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREPAERMAFN
MPVQGTAADLMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGE
DWLSAKE

FIG. 2b

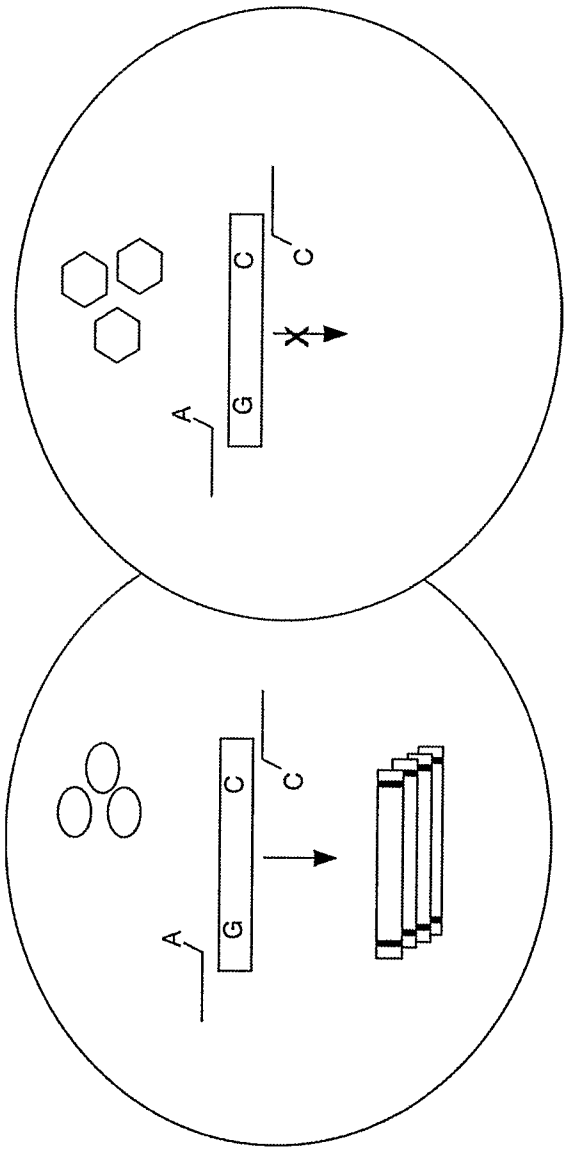
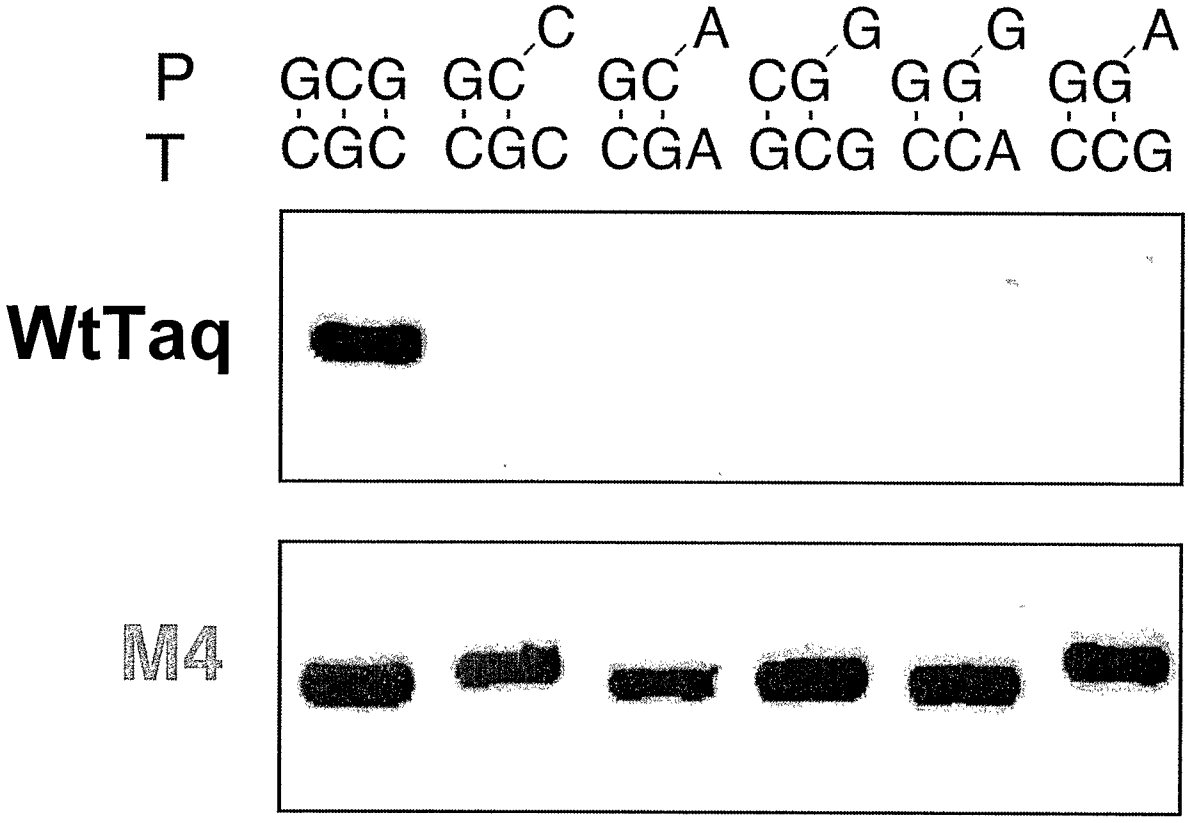
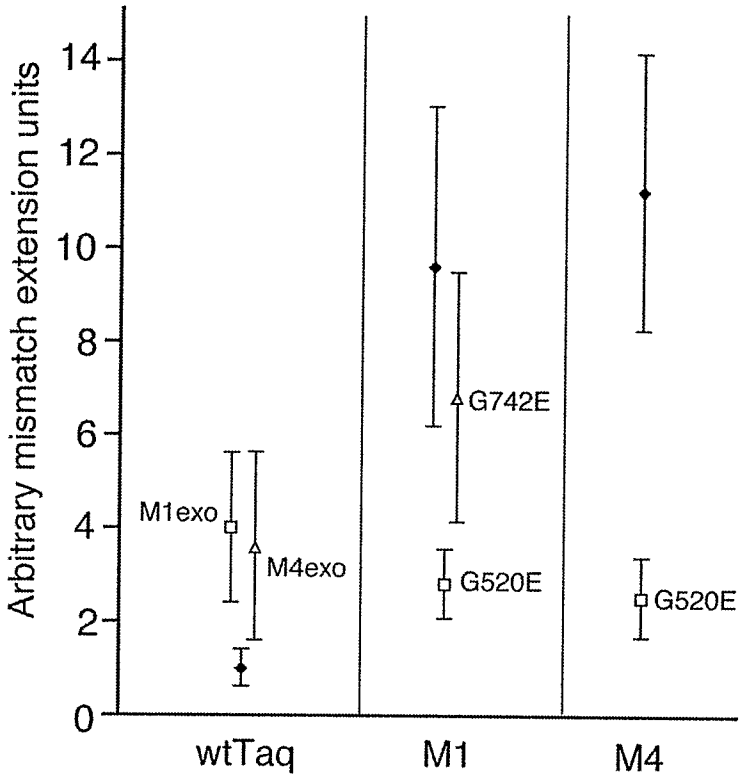


FIG. 3



Mismatch extension properties of selected polymerases.

FIG. 4a



Mismatch extension properties of selected polymerases.

FIG. 4b

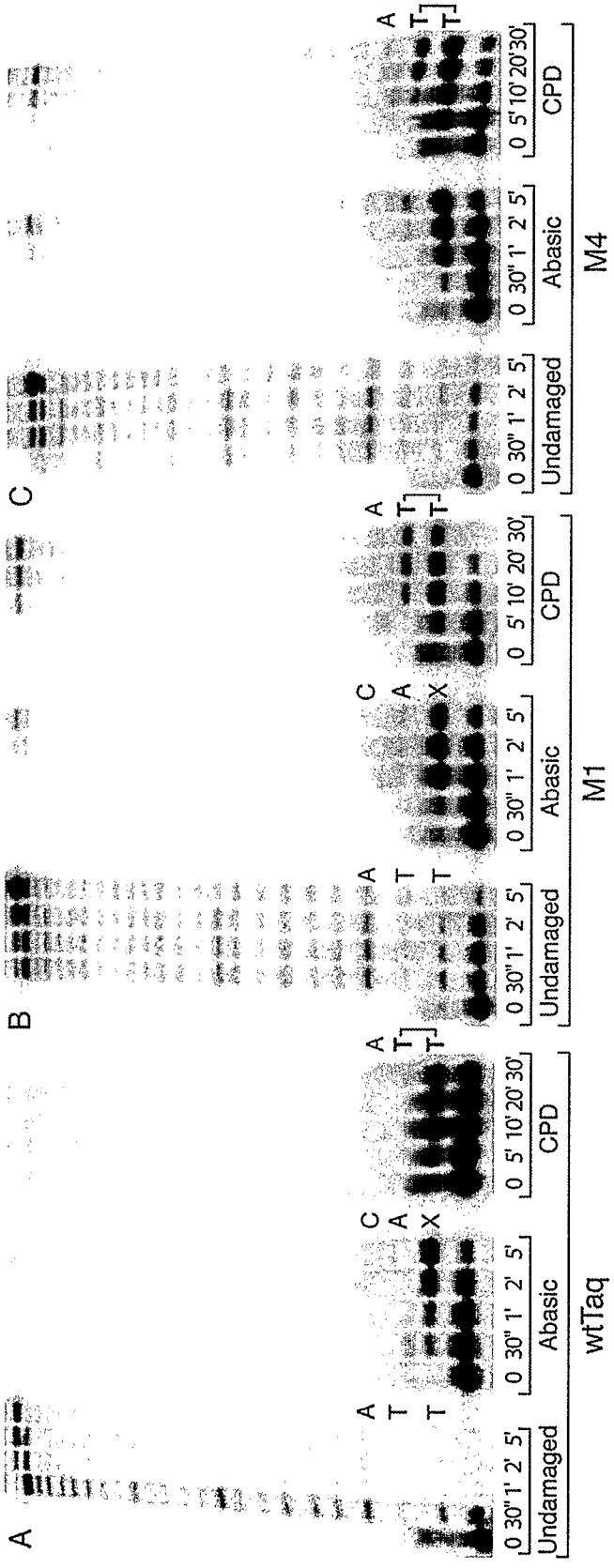
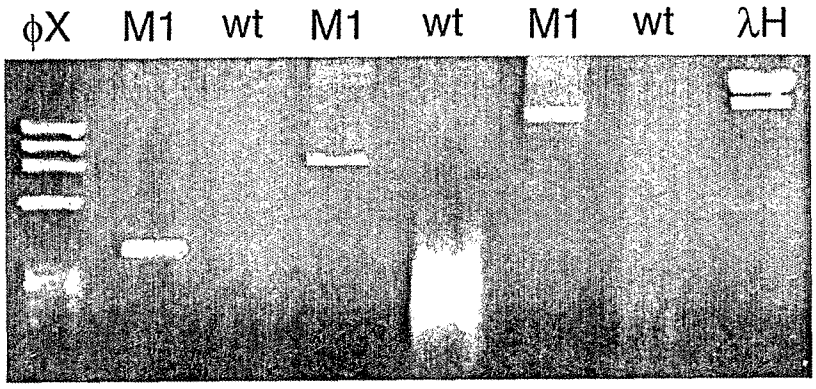
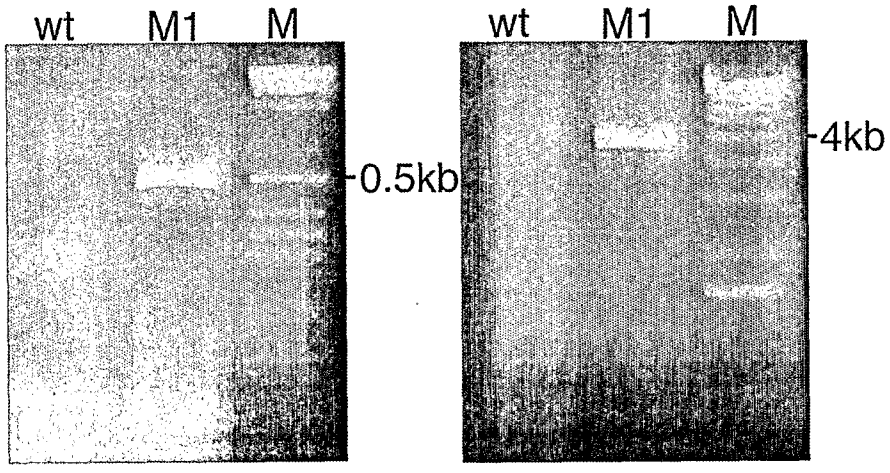

A
B
C
Undamaged
Abasic
CPD
0 30" 1' 2' 5'
0 5' 10' 20' 30'
wtTaq
M1
M4
Lesion bypass activity

FIG. 5



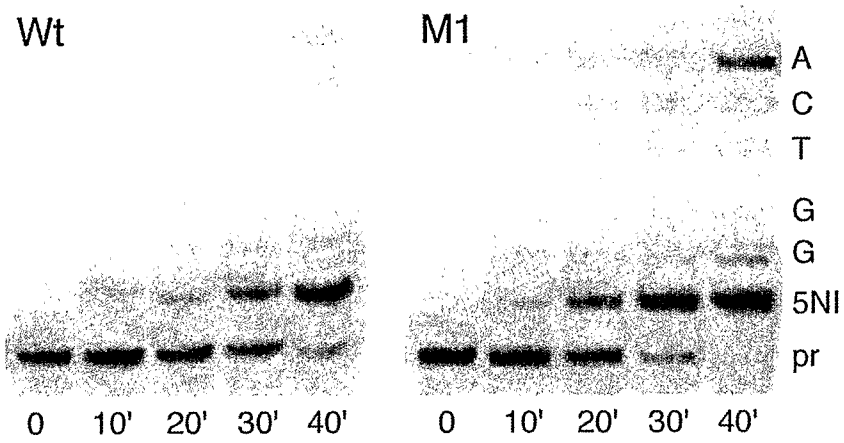
Polymerase activity on unnatural substrates (αS dNTP).

FIG. 6a



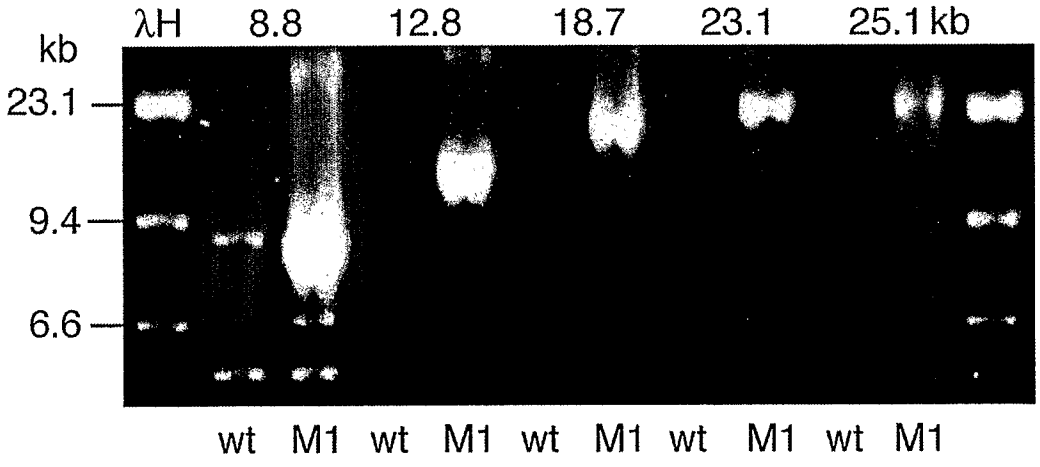
Polymerase activity on unnatural substrates.

FIG. 6b



Polymerase activity on unnatural substrates (5NI)

FIG. 6c



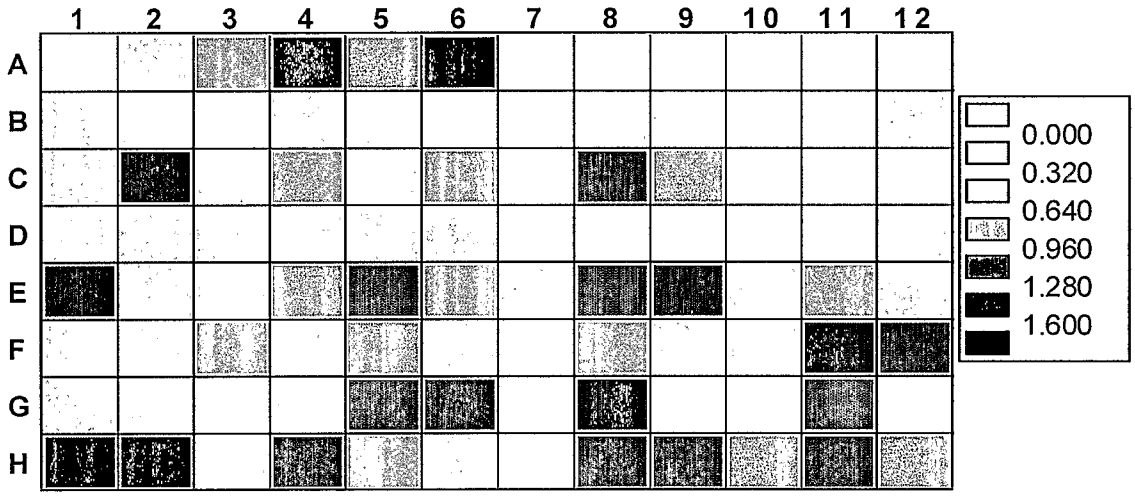
Long range PCR.

FIG. 7

Experiment I.

ELISA with Fluorescein 12-dATP on clones selected for 4-mismatch extension using the primer FITC4

(5'-TAGCTACCATTTTCGCCGGCTTCCGTCGCGACCACGTTP1TTCGTGGTCGCGACGGAAGCCG-3', P1=Biotin)



High Limit = 1.600

Low Limit = 0.000

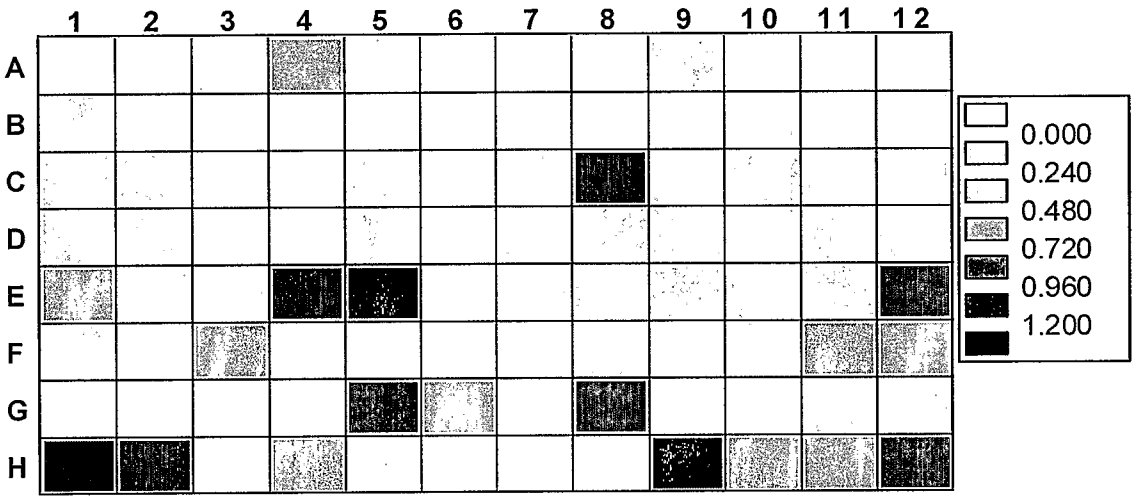
0.315	0.489	0.846	1.558	0.652	1.378	0.360	0.216	0.566	0.290	0.259	0.042
0.513	0.038	0.061	0.461	0.449	0.640	0.523	0.444	0.412	0.222	0.404	0.476
0.467	1.125	0.588	0.818	0.311	0.718	0.292	1.169	0.736	0.551	0.222	0.260
0.627	0.574	0.340	0.422	0.593	0.581	0.399	0.556	0.347	0.280	0.472	0.157
1.056	0.357	0.316	0.927	1.021	0.843	0.636	1.018	1.057	0.632	0.933	0.638
0.046	0.499	0.906	0.489	0.774	0.408	0.444	0.867	0.576	0.366	1.344	1.070
0.404	0.286	0.455	0.335	1.140	0.972	0.467	1.448	0.366	0.408	0.988	0.410
1.512	1.291	0.063	1.034	0.663	0.419	0.346	1.117	1.204	0.936	1.239	0.750

FIG. 8a

Experiment II.

ELISA with Bio-11-dATP on clones selected for 4-mismatch extension and displayed activity with Fluorescein 12-dATP in the ELISA assaying the primer FITC102

(5'-
TAGCTACCATTTTTTTTTTCGCCGGCTTCCGTCGCGACCACGTTP1TTCGTGGTCGCGACGGAA
GCCG--3', P1=Biotin)



High Limit = 1.200

Low Limit = 0.000

0.044	0.048	0.200	0.660	0.464	0.459	0.053	0.042	0.325	0.067	0.035	0.027
0.275	0.097	0.014	0.107	0.161	0.059	0.092	0.138	0.070	0.044	0.050	0.160
0.204	0.373	0.177	0.219	0.142	0.080	0.033	0.821	0.147	0.310	0.050	0.032
0.288	0.094	0.038	0.050	0.381	0.076	0.057	0.264	0.052	0.044	0.056	0.031
0.540	0.061	0.054	0.928	1.106	0.151	0.152	0.441	0.342	0.381	0.345	0.802
0.038	0.069	0.639	0.089	0.184	0.045	0.055	0.401	0.154	0.072	0.715	0.647
0.062	0.043	0.135	0.057	0.904	0.546	0.182	0.844	0.054	0.064	0.250	0.188
1.332	0.758	0.036	0.535	0.212	0.057	0.049	0.214	1.070	0.494	0.570	0.839

FIG. 8b

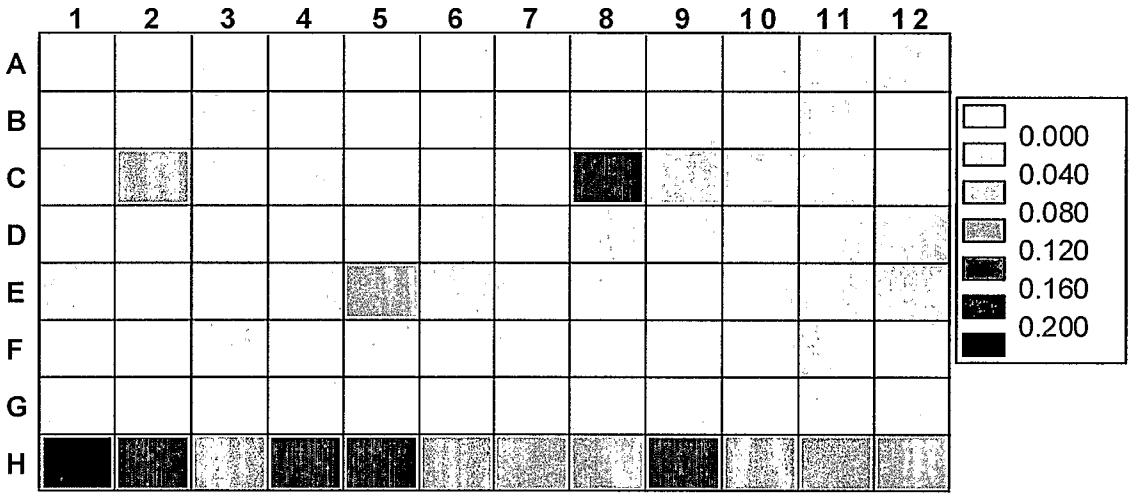
Experiment III.

ELISA with CyDye 5-dCTP on clones selected for 4-mismatch extension using the primer ELISAC4P

(5'-

TAGCTACCAGGGGCTCCGGCTTCCGTCGCGACCACGTTP1TTCGTGGTCGCGACGGAAGCC

G-3', P1=Biotin)



High Limit = 0.200

Low Limit = 0.000

0.053	0.062	0.054	0.058	0.048	0.083	0.063	0.043	0.081	0.066	0.065	0.058
0.079	0.057	0.055	0.084	0.058	0.078	0.060	0.061	0.061	0.058	0.054	0.075
0.113	0.136	0.134	0.129	0.120	0.122	0.111	0.176	0.109	0.119	0.112	0.115
0.081	0.075	0.068	0.068	0.083	0.076	0.065	0.072	0.060	0.061	0.065	0.065
0.119	0.079	0.074	0.092	0.142	0.106	0.091	0.103	0.084	0.092	0.077	0.101
0.086	0.079	0.112	0.092	0.113	0.079	0.075	0.089	0.081	0.086	0.133	0.131
0.076	0.085	0.077	0.084	0.105	0.115	0.088	0.126	0.068	0.074	0.090	0.101
0.422	0.209	0.143	0.175	0.197	0.142	0.144	0.155	0.182	0.168	0.156	0.156

FIG. 8c

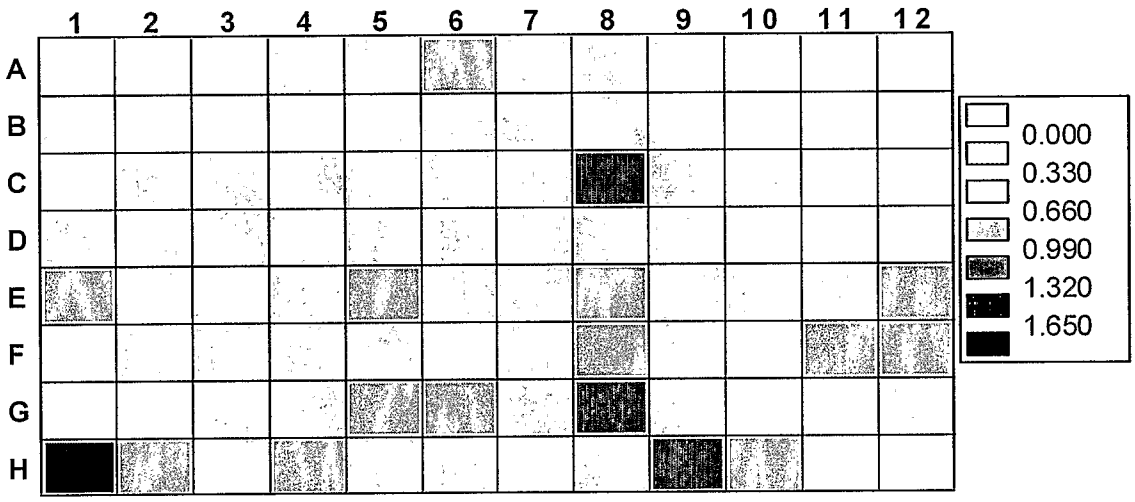
Experiment IV.

ELISA with CyDye 3-dUTP on clones selected for 4-mismatch extension using the ELISA primer ELISAT3P

(5'-

TAGCTCGGTAAACGCCGGCTTCCGTCGCGACCACGTTP5TTCGTGGTCGCGACGGAAGCCG

--3', P1=Biotin)



High Limit = 1.650

Low Limit = 0.000

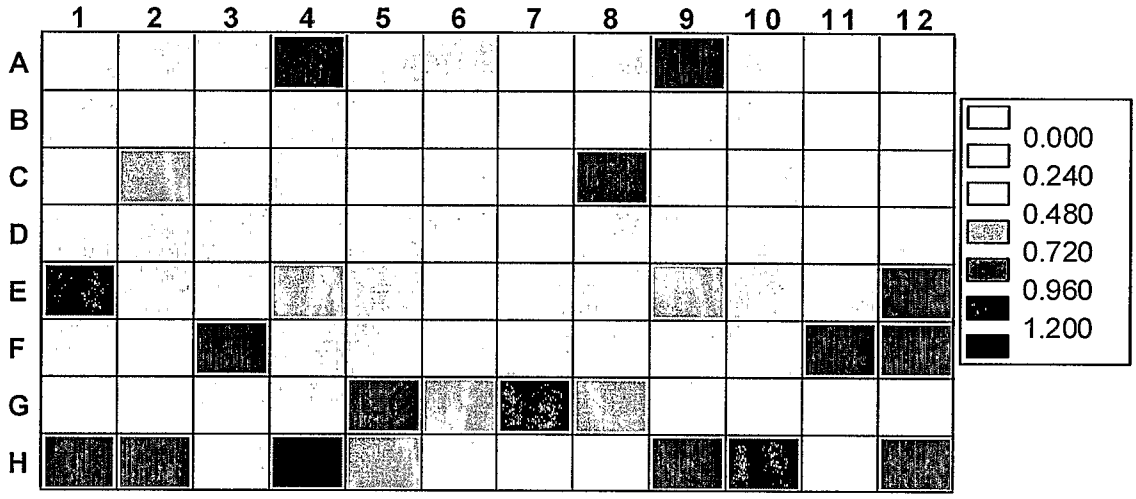
0.104	0.141	0.226	0.316	0.244	0.878	0.352	0.107	0.310	0.147	0.090	0.022
0.578	0.014	0.029	0.255	0.297	0.629	0.541	0.347	0.346	0.191	0.206	0.522
0.261	0.516	0.399	0.643	0.188	0.339	0.285	1.051	0.546	0.501	0.174	0.183
0.651	0.277	0.372	0.540	0.444	0.485	0.311	0.569	0.236	0.173	0.222	0.077
0.716	0.143	0.215	0.574	0.782	0.492	0.509	0.910	0.617	0.615	0.515	0.719
0.019	0.380	0.331	0.336	0.582	0.187	0.253	0.768	0.391	0.285	0.829	0.789
0.291	0.174	0.388	0.363	0.735	0.827	0.435	1.072	0.230	0.332	0.545	0.394
1.673	0.724	0.031	0.795	0.454	0.400	0.297	0.649	1.139	0.755	0.407	0.509

FIG. 8d

Experiment V.

ELISA on clones selected for 4-mismatch extension using a hairpin ELISA primer containing an abasic site

(Primer: Pscreen1Abas; 1= abasic site, 5= U biotin; sequence:
AGCTACCATGCCTGCACGCAG1CGGCATCCGTCGCGACCACGTT5TTCGTGGTCGCGAC
GGATGCCG)

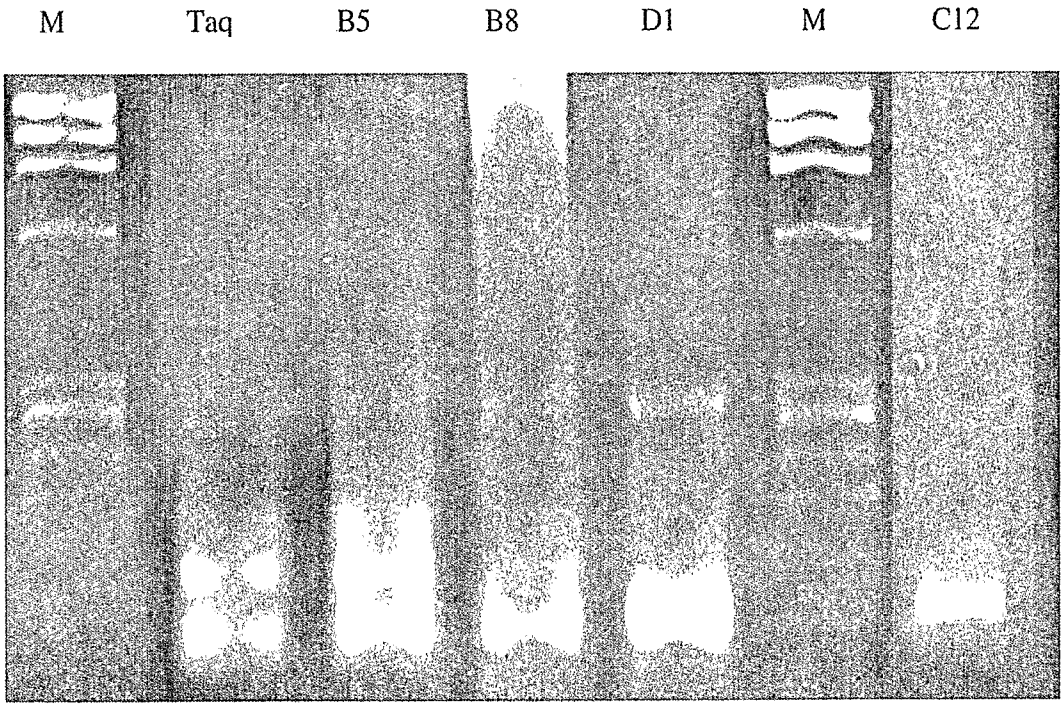


High Limit = 1.200

Low Limit = 0.000

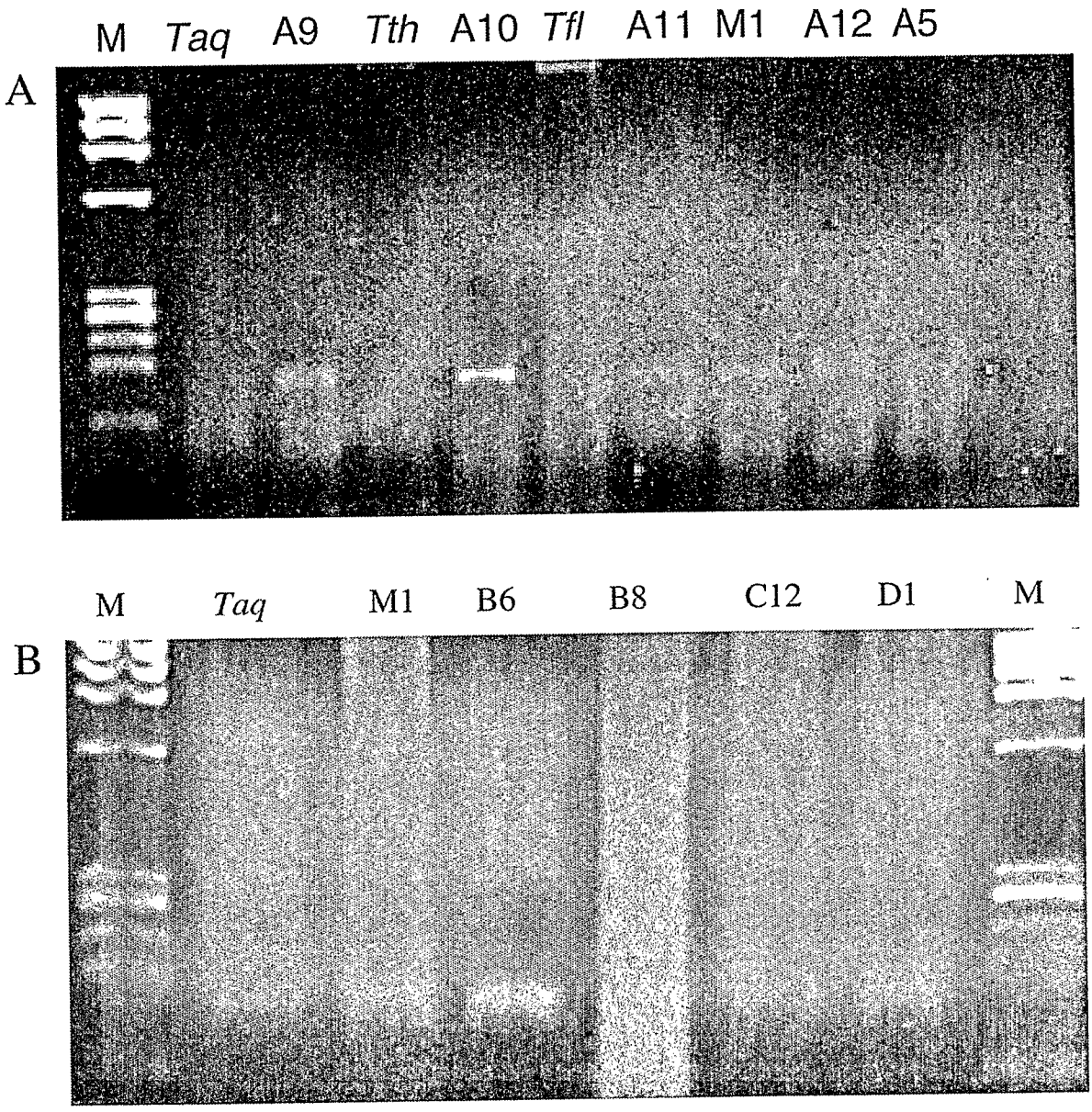
0.075	0.092	0.137	0.882	0.453	0.441	0.185	0.058	0.931	0.108	0.040	0.023
0.445	0.011	0.035	0.096	0.095	0.086	0.059	0.093	0.089	0.071	0.037	0.113
0.314	0.512	0.091	0.113	0.086	0.092	0.083	0.746	0.122	0.304	0.065	0.093
0.144	0.098	0.063	0.071	0.129	0.093	0.107	0.095	0.019	0.089	0.110	0.079
1.170	0.220	0.073	0.563	0.343	0.099	0.280	0.099	0.596	0.442	0.135	0.883
0.025	0.077	0.913	0.094	0.150	0.062	0.229	0.092	0.087	0.071	0.814	0.912
0.099	0.075	0.102	0.021	0.742	0.666	0.990	0.659	0.063	0.095	0.070	0.105
0.913	0.957	0.038	1.382	0.575	0.101	0.105	0.125	0.761	1.099	0.285	0.808

FIG. 8e



Four Mismatch Extension in PCR.

FIG. 9



Abasic site bypass in PCR.

FIG. 10

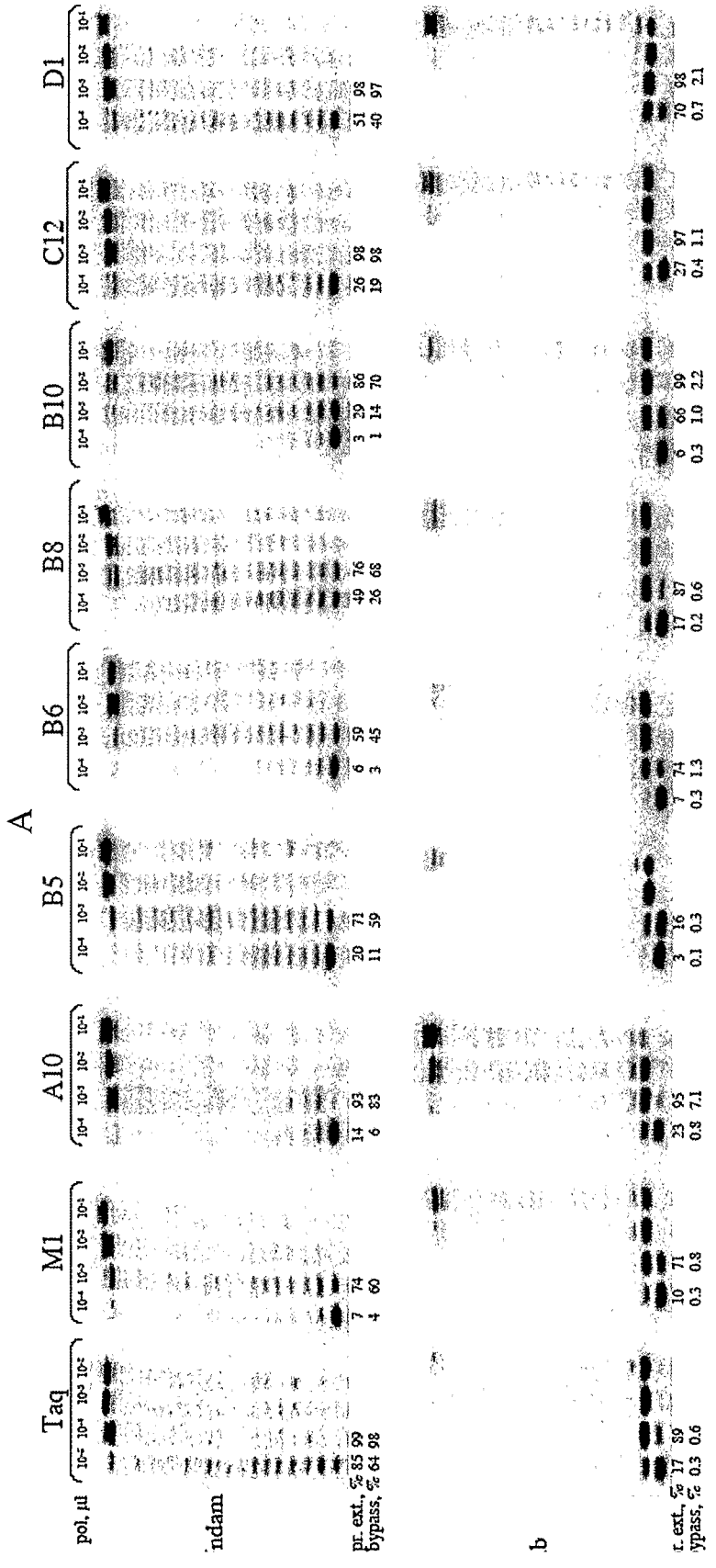


FIG. 11

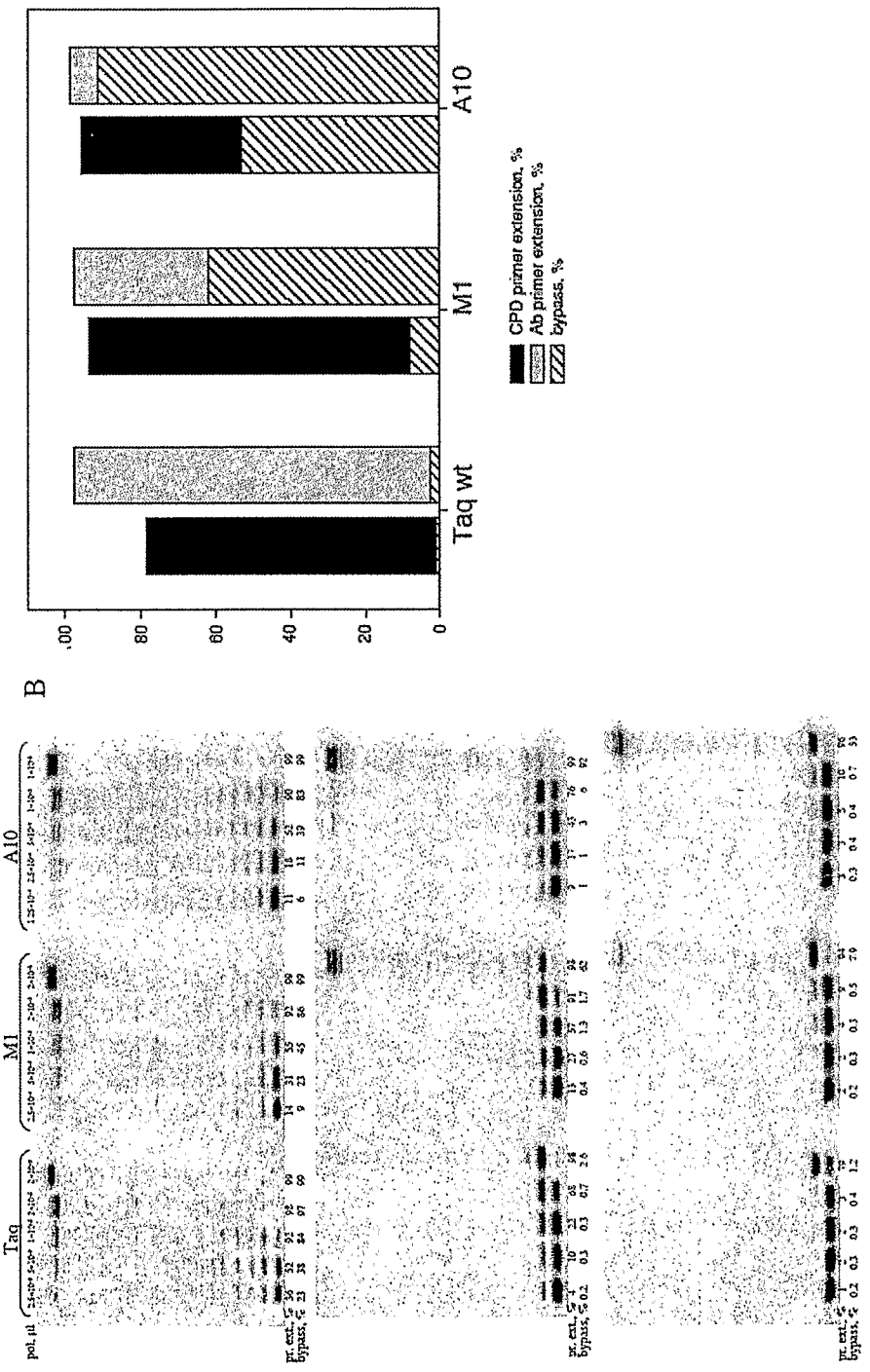
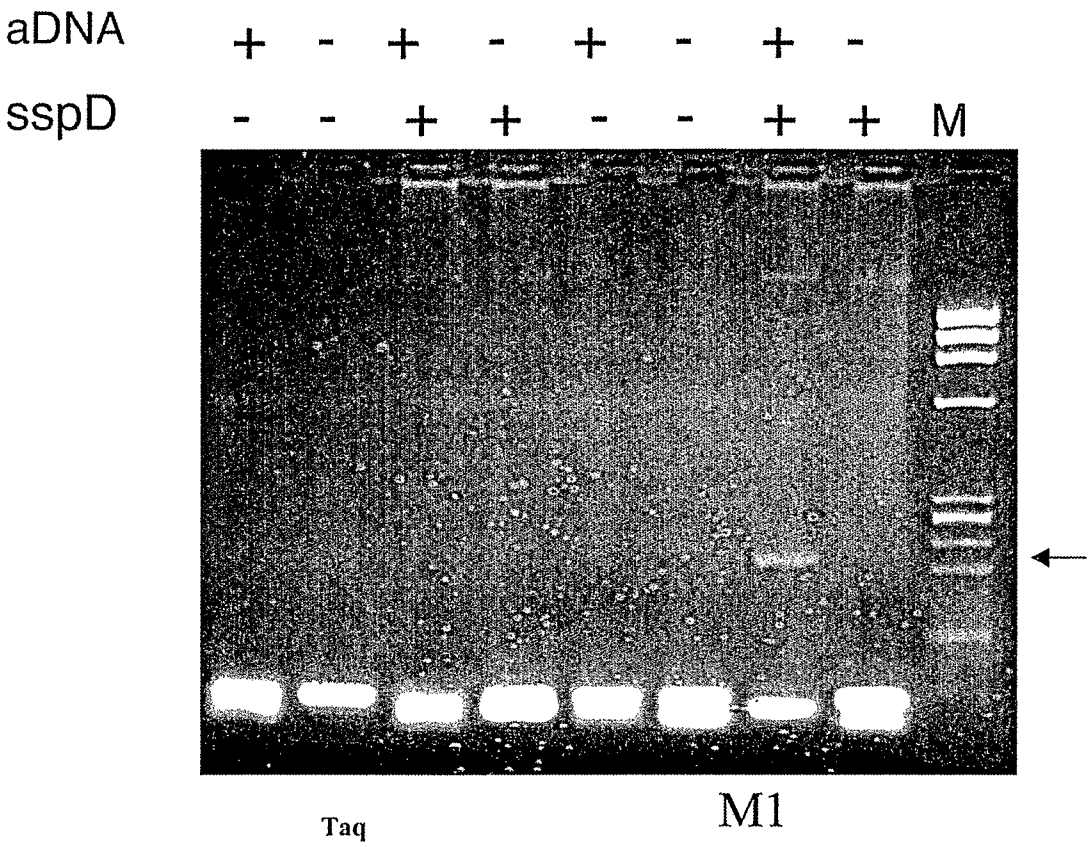
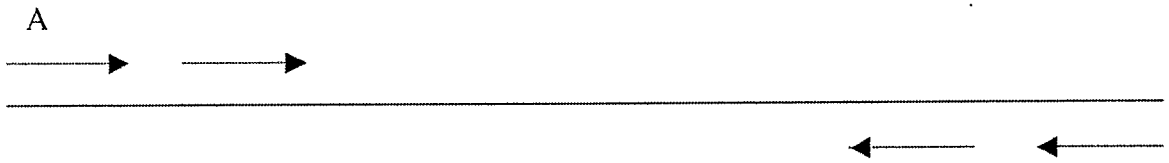

B
Taq wt
M1
A10
CPD primer extension, %
Ab primer extension, %
bypass, %
0
20
40
60
80
100
Taq
M1
A10
pol. µl
pr. ext.
bypass,

FIG. 11 CONT'D



PCR amplification of a mix of extracts of cave hyenas (*Crocuta spelea*) that had not previously produced an amplification product.

FIG. 12



B

ACCCCATGCATATAGGCATGTACATATTATGCTTGATCTTACATGAGGACTTACA
TCTCAAAAGTTTATTTCAAGTGTATAGTCTGTAAGCATGTATTTCACTTAGTCCA
GGAGCTTAATCACCAGGCCTCGAGAAACCAGCAACCCTTGCGAGT

FIG. 13

	Sample	Dilution	SuperTaq		Mutant blend		Improvement
Experiment			Positive PCR	Negative PCR	Positive PCR	Negative PCR	
1	GS3-7	500	24	12	28	8	16%
2	GS3-7	2000	2	22	5	19	150%
3	GS3-7	1000	21	27	24	24	14%
4	366	5	2	2	4	0	100%
5	366	10	12	12	16	8	33%
6	GS3-7	1000	12	36	14	34	16%
7	GS3-7	1000	10	14	7	17	-30%
8	GS3-7	1000	12	36	13	35	8%
9	GS3-7	1000	12	36	13	35	8%
Total			107	197	124	180	15%

PCRS in the presence of limiting amounts of cave bear aDNA

FIG. 14

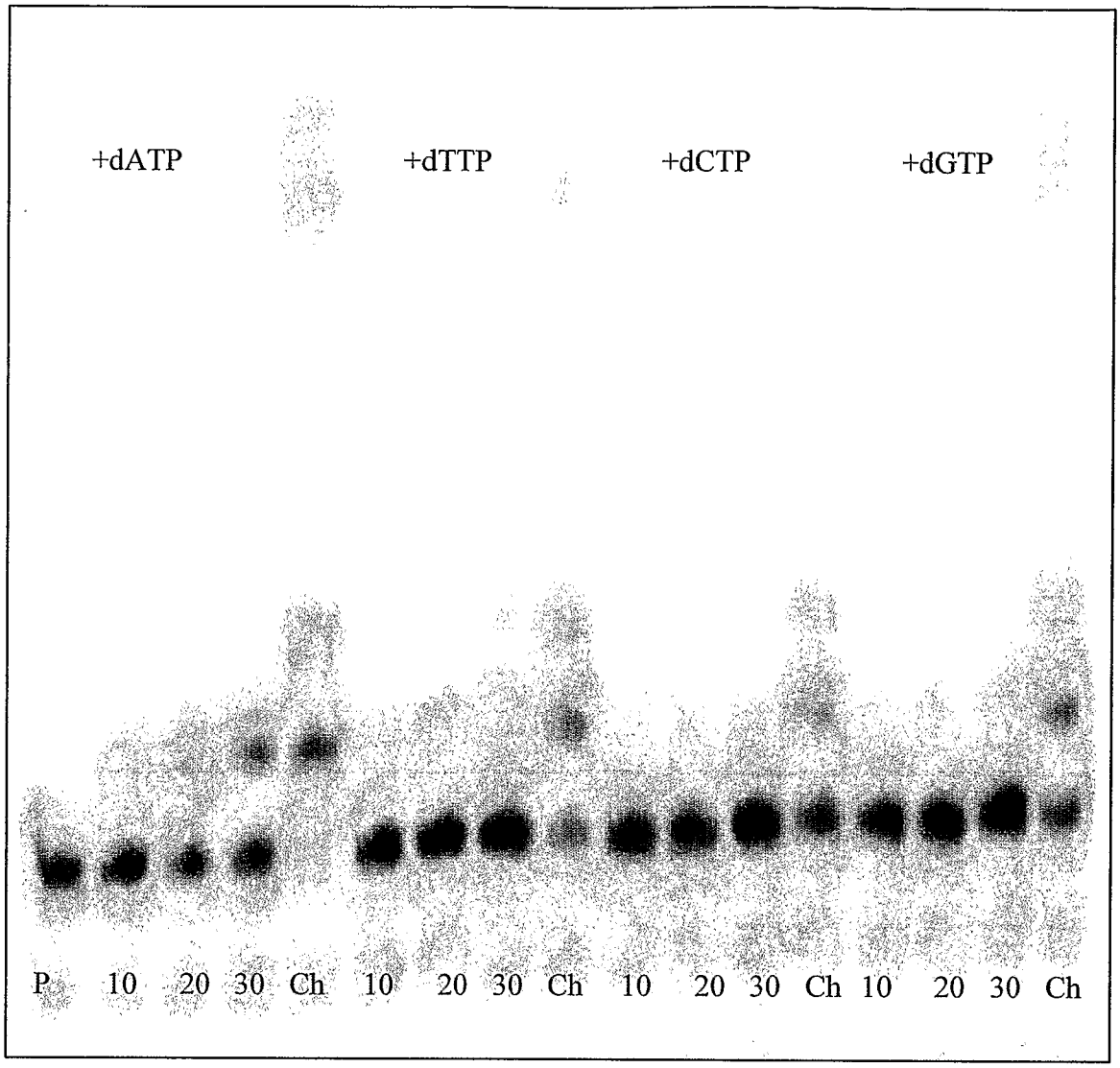
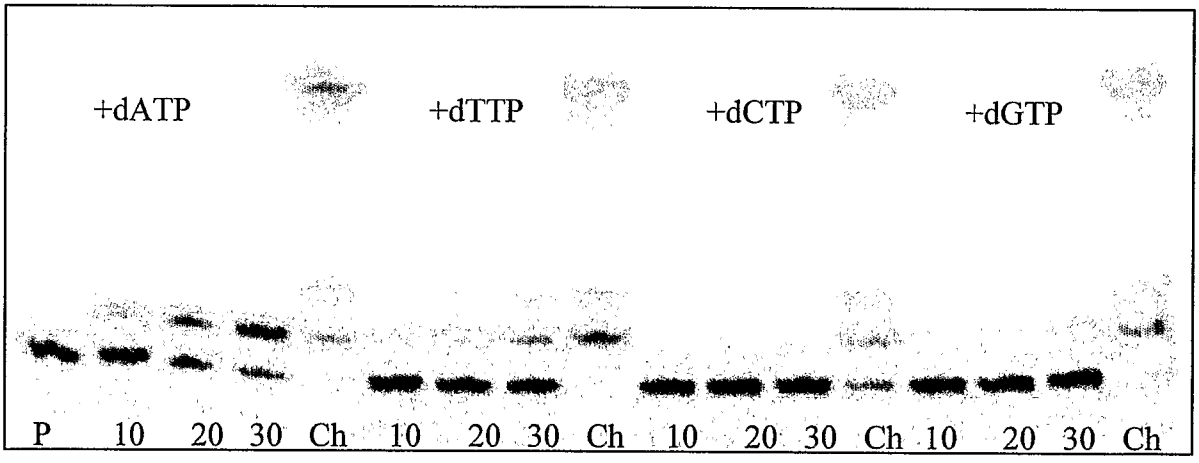

+dATP
+dTTP
+dCTP
+dGTP
P 10 20 30 Ch 10 20 30 Ch 10 20 30 Ch 10 20 30 Ch

FIG. 15


+dATP
+dTTP
+dCTP
+dGTP
P 10 20 30 Ch 10 20 30 Ch 10 20 30 Ch 10 20 30 Ch

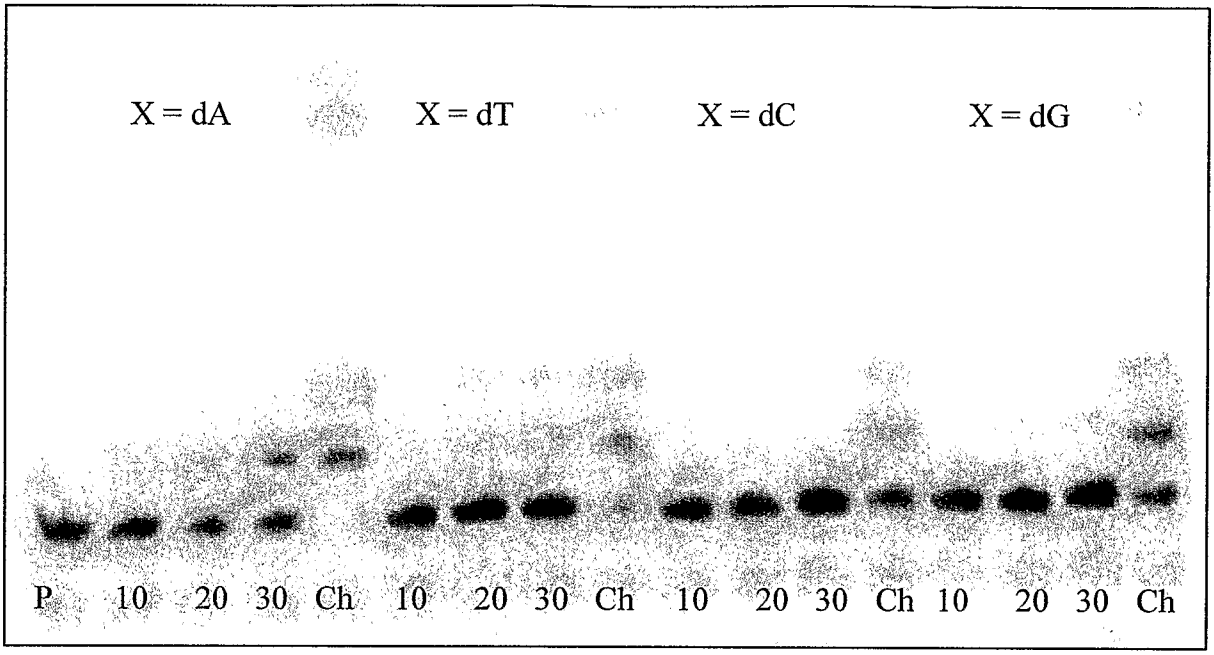

X = dA
X = dT
X = dC
X = dG
P 10 20 30 Ch 10 20 30 Ch 10 20 30 Ch 10 20 30 Ch

FIG. 17

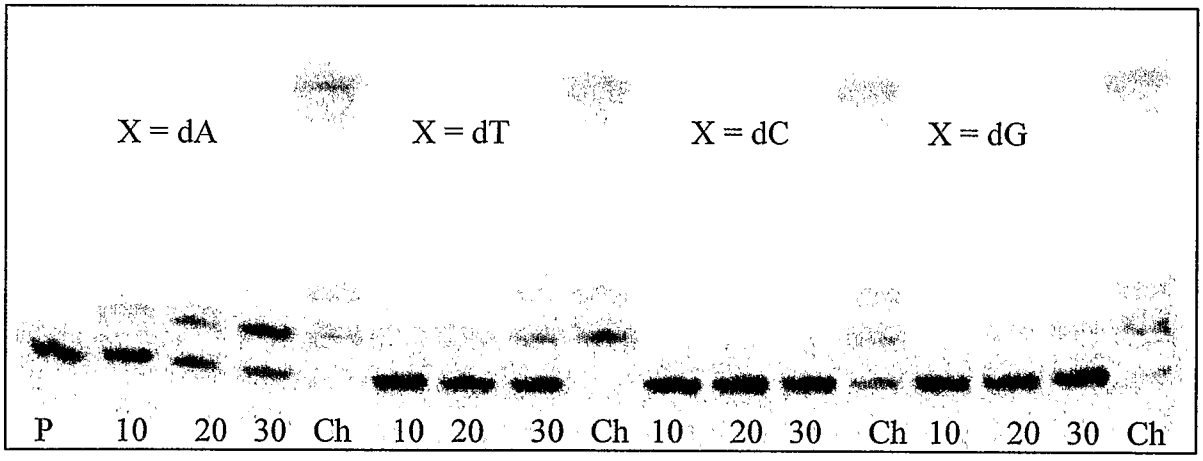

X = dA
X = dT
X = dC
X = dG
P 10 20 30 Ch 10 20 30 Ch 10 20 30 Ch 10 20 30 Ch

FIG. 18

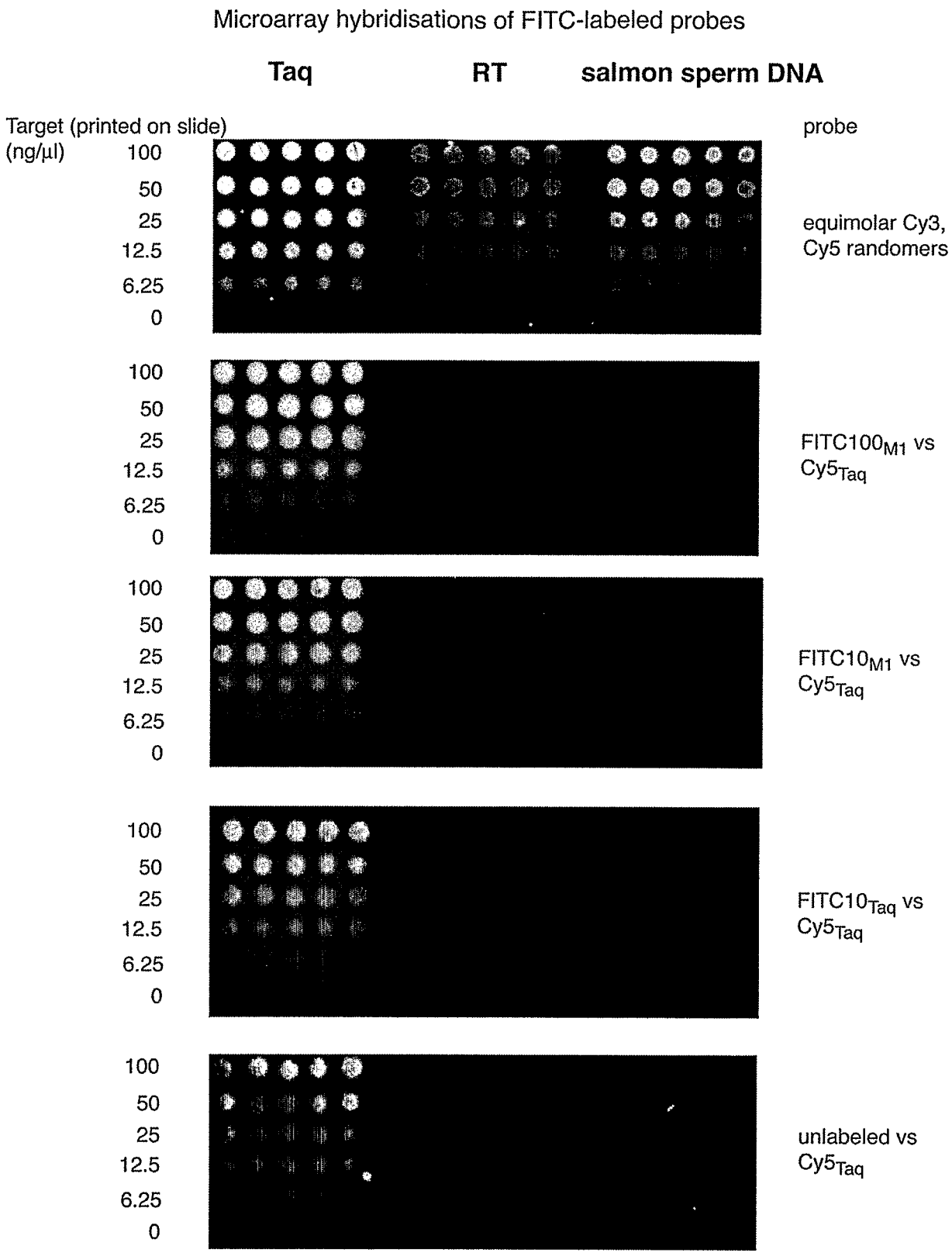
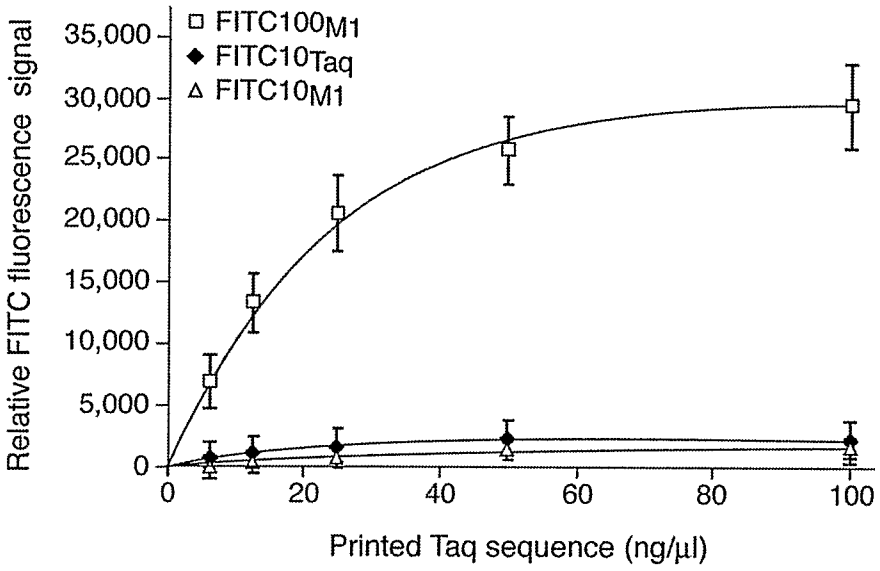

Microarray hybridisations of FITC-labeled probes
Taq
RT
salmon sperm DNA
Target (printed on slide)
(ng/µl)
probe
100
50
25
12.5
6.25
0
equimolar Cy3, Cy5 randomers
FITC100M1 vs Cy5Taq
FITC10M1 vs Cy5Taq
FITC10Taq vs Cy5Taq
unlabeled vs Cy5Taq

FIG. 19



Microarray analysis of FITC-labeled probes.

FIG. 20

Microarray signals from FITC-labelled probes.

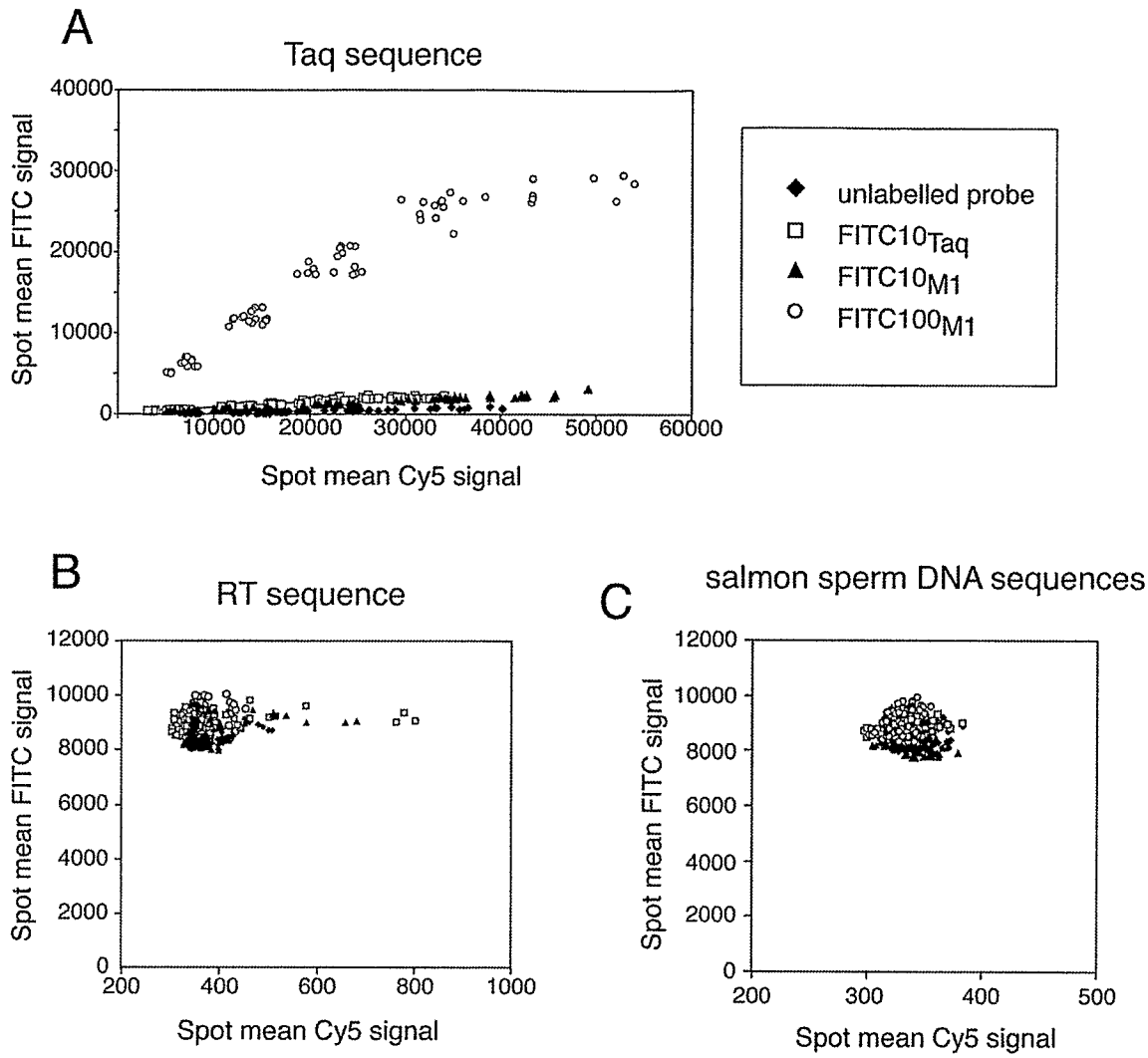


FIG. 21

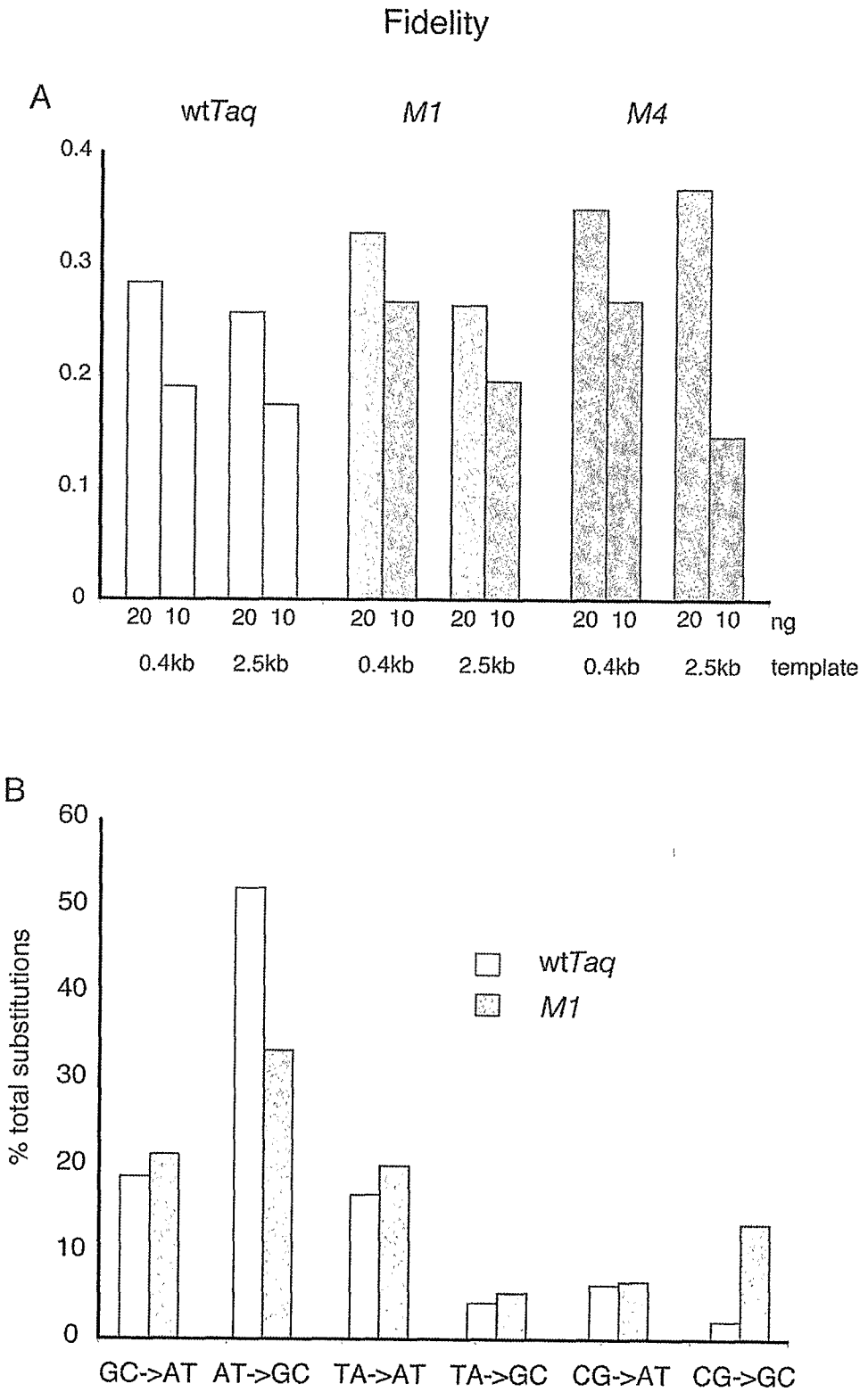
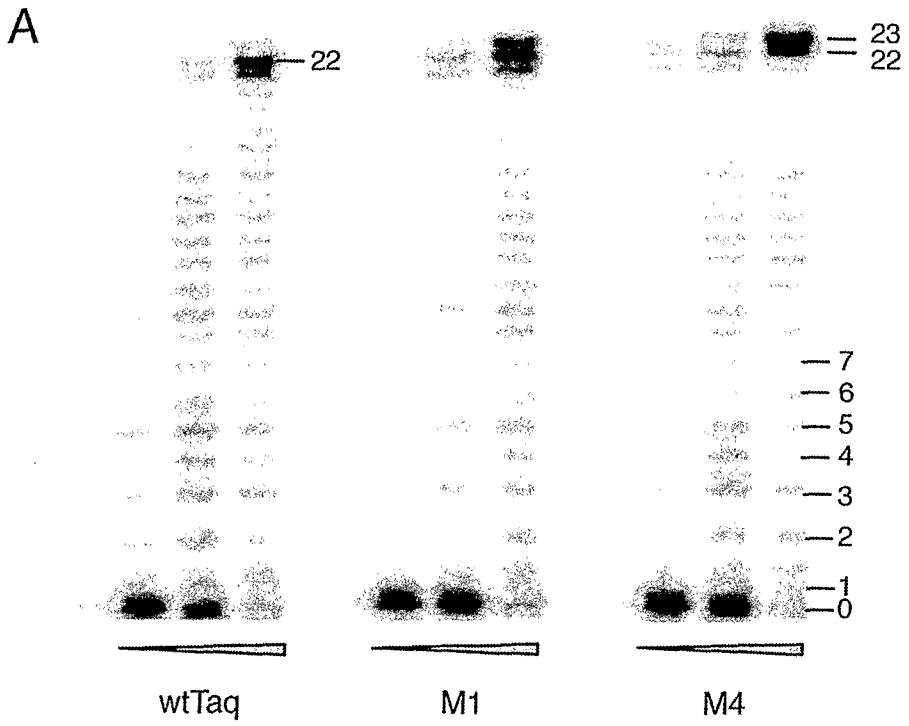

Fidelity
A
wtTaq
M1
M4
0.4
0.3
0.2
0.1
0
20 10 20 10 20 10 20 10 20 10 20 10 ng
0.4kb 2.5kb 0.4kb 2.5kb 0.4kb 2.5kb template
B
% total substitutions
60
50
40
30
20
10
0
wtTaq
M1
GC->AT AT->GC TA->AT TA->GC CG->AT CG->GC

FIG. 22

Processivity



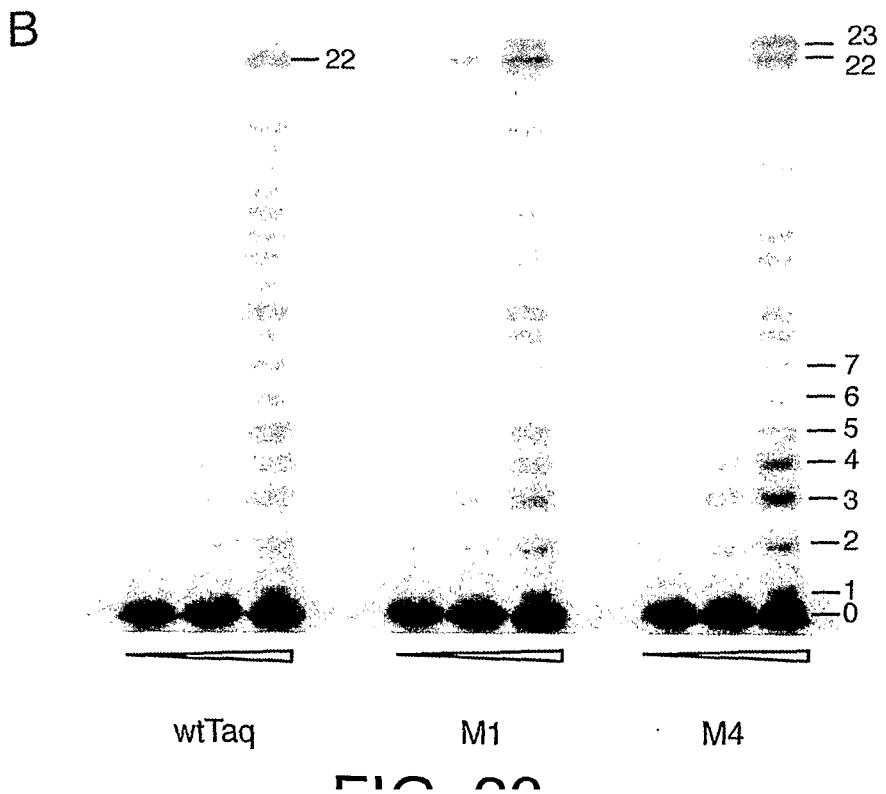


FIG. 2B