



Office de la Propriété
Intellectuelle
du Canada

Un organisme
d'Industrie Canada

Canadian
Intellectual Property
Office

An agency of
Industry Canada

CA 2544841 C 2014/03/25

(11)(21) **2 544 841**

(12) **BREVET CANADIEN
CANADIAN PATENT**

(13) **C**

(86) Date de dépôt PCT/PCT Filing Date: 2004/11/03

(87) Date publication PCT/PCT Publication Date: 2005/05/19

(45) Date de délivrance/Issue Date: 2014/03/25

(85) Entrée phase nationale/National Entry: 2006/05/02

(86) N° demande PCT/PCT Application No.: GB 2004/004643

(87) N° publication PCT/PCT Publication No.: 2005/045015

(30) Priorités/Priorities: 2003/11/03 (GB0325650.0);
2004/05/14 (GB0410871.8)

(51) Cl.Int./Int.Cl. *C12N 15/10* (2006.01),
C12N 15/54 (2006.01), *C12N 9/12* (2006.01),
C12Q 1/68 (2006.01)

(72) Inventeurs/Inventors:
HOLLIGER, PHILIPP, GB;
GHADESSY, FARID, SG;
D'ABBADIE, MARC, GB

(73) Propriétaire/Owner:
MEDICAL RESEARCH COUNCIL, GB

(74) Agent: MBM INTELLECTUAL PROPERTY LAW LLP

(54) Titre : ADN POLYMERASES POLA CAPABLE DE CONTOURNER UN SITE ABASIQUE

(54) Title: POLA DNA POLYMERASES CAPABLE OF ABASIC SITE BYPASS

(57) **Abrégé/Abstract:**

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



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
19 May 2005 (19.05.2005)

PCT

(10) International Publication Number
WO 2005/045015 A3

(51) International Patent Classification⁷: C12N 15/10,
9/12, 15/54, C12Q 1/68

(21) International Application Number:
PCT/GB2004/004643

(22) International Filing Date:
3 November 2004 (03.11.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0325650.0 3 November 2003 (03.11.2003) GB
0410871.8 14 May 2004 (14.05.2004) GB

(71) Applicant (for all designated States except US): **MEDICAL RESEARCH COUNCIL** [GB/GB]; 20 Park Crescent, London WIB 1AL (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HOLLIGER, Philipp** [CH/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB). **GHADESSY, Farid** [GB/SG]; 10B Keng Chin Road, Villa Margaux, Singapore 258712 (SG). **D'ABBADIE, Marc** [FR/GB]; MRC laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB).

(74) Agents: **SOAMES, Candida, Jane** et al.; D. Young & Co., 120 Holborn, London EC1N 2DY (GB).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv)) for US only

Published:

— with international search report

(88) Date of publication of the international search report:
22 December 2005

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

(54) Title: POLYMERASE

(57) Abstract: The present invention relates to DNA polymerases. In particular the invention relates to a method for the generation of DNA polymerases exhibiting a relaxed substrate specificity. Uses of mutant polymerases produced using the methods of the invention are also described.



WO 2005/045015 A3

POLA DNA POLYMERASES CAPABLE OF ABASIC SITE BYPASS

5 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.

10 **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
15 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 $\sim 10^5$ base pairs (Schaaper JBC 1993).

20 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
25 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
30 modified or damaged DNA templates, the incorporation of modified or unnatural

deoxynucleotide 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 attachment 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
5 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
10 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
15 replication or other methods.

Summary of the invention

The present inventors modified the principles of directed evolution, (in particular
20 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.

25 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
30 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.

- 5 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.

10

- 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.

- 20 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 specifically the term 'distorting 3' termini/ends' includes within its scope the phenomenon 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

30

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 unnatural 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.

15

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.

25

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:

30

- (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.
- 5 (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
- 10 (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.

15

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

20

25

30

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

5

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). Preferably, 'expanded substrate range' (of an engineered DNA polymerase) includes a wider spectrum of unnatural nucleotide substrates including α S 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 GCC-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,

30

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.

5 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.

10

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
15 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.

20

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
25 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
30 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.

- 5 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-
 10 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,
 15 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
 20 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.

25

- 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
 30 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.

10 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 engineered polymerases. Preferably the term 'blends' refers to a mixture of 6, 7, 8, 9 or 10 or more 'engineered
15 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 pol ζ (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
20 the polA-family polymerases that are proficient in mismatch extension (ME) and translesion synthesis (TLS).
25

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.
30

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.

5

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

10 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.

15 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.

20 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 unnatural 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.

25

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
30 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.

5 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.

10 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.

15

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
20 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
25 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.

30

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 the figures

15

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)

20 **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.

25 **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) *wtTaq*, (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 α S dNTPs. α S DNA amplification products of 0.4kb, 0.8kb and 2kb, are obtained with M1 but not with *wtTaq* (wt). ϕ X, HaeIII-digested phage ϕ X174 DNA marker. λ H, 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. *WtTaq* (wt) fails to generate amplification products larger than 8.8kb while M1 is able to amplify fragments of > 25kb. λ H, 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.

5 **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 (*Crocota spelaea*) were extracted and analysed.

10

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

15 **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
20 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.

25

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.

5 **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,
10 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
15 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
20 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
25 co-hybridisation experiment.

Figure 22. Fidelity. (A) MutS ELISA. Relative replication fidelity of *wtTaq*, *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 *wtTaq* or *M1*. Types of substitutions are given as % of total substitutions (*wtTaq*: 48, *M1*: 74). Equivalent substitutions on either strand (e.g. G->A, C->T) were added together (GC->AT). Observed -1 deletions (*wtTaq*: 3, *M1*: 1) are not shown.

10

Figure 23. Processivity of *wtTaq*, *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 *wtTaq* and *M1* and from 13-35% for *M4*, while at positions 6 and 7 the termination probability was 5% for *wtTaq*, 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 *wtTaq*,

20

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

Detailed description of the invention

5

(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

10 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;

15 (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.

20

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.

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

30 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., Gorochoy, G., Jones, P. T. & Winter, G. (1992) *Nucleic Acids*) but with low efficiencies.

5

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
10 details of the method of compartmentalised self replication are given below.

Microcapsules

The microcapsules used according to the method of the invention require appropriate
15 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
20 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
25 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
30 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
5 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
10 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
15 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
20 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
25 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
30 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.*, 5 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 10 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 15 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 20 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; 25 Lissant, 1974; Lissant, 1984).

Emulsions

Emulsions may be produced from any suitable combination of immiscible liquids. 30 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 discrete droplets (the internal phase). The external phase, being a hydrophobic oil,
5 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
10 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 (SpanTM80; ICI) and polyoxyethylenesorbitan monooleate (TweenTM 80; ICI) and Triton-X-100..

15 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
20 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.

25 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,
30 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
5 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
10 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.

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

Expression within microcapsules

20 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
25 microcapsule size. Preferably, the mean volume of the microcapsules is less than $5.2 \times 10^{-16} \text{ m}^3$, (corresponding to a spherical microcapsule of diameter less than $10 \mu\text{m}$, more preferably less than $6.5 \times 10^{-17} \text{ m}^3$ ($5 \mu\text{m}$), more preferably about $4.2 \times 10^{-18} \text{ m}^3$ ($2 \mu\text{m}$) and ideally about $9 \times 10^{-18} \text{ m}^3$ ($2.6 \mu\text{m}$).

30 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); Q β 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 $5.2 \times 10^{-16} \text{ m}^3$ (corresponding to a sphere of diameter 10 μm).

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 ($4.17 \times 10^{-22} \text{ m}^3$ which if spherical would have a diameter of 93nm).

30

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 $5.2 \times 10^{-22} \text{ m}^3$ and $5.2 \times 10^{-16} \text{ m}^3$ corresponding to a sphere of diameter between 0.1um and 10um, more preferably of between about $5.2 \times 10^{-19} \text{ m}^3$ and $6.5 \times 10^{-17} \text{ m}^3$ (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.

25

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.

30

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 DynabeadsTM), 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.

15

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

25
30

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
5 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
10 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
15 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
20 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
25 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
30 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
5 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.*,
10 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'
15 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)).
20

Microcapsules/sorting

In addition to the nucleic acids described above, the microcapsules according to the
25 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
30 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.

5 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.

10 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 TNTTM extract systems from Promega). The mixture of
15 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).

20

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
25 from the individual clones for further purification and assay.

Microcapsule identification

30 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
5 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
10 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
15 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
20 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
25 (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
30 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.

10 *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; Q β 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).

20

(B) DNA polymerases according to the invention.

(i) General

25 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 /
30 pyrimidine·pyrimidine) mismatches are extended up to 10⁶-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)
5 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.

10

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
15 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'
20 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.

25

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
30 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).

5

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.

10 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).

15

(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
20 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
25 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).

30

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%
5 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
10 (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
15 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
20 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 pol ζ (Johnson Nature), all extend 3' mismatches with elevated efficiency compared to high-fidelity polymerases. Thus, the selected polymerases share significant
25 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
30 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
5 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
10 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
15 varying degrees from the canonical nucleobase structure. Of these the α S 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 α S substitution. Fluorescently-labelled nucleotides like aS nucleotides retain base-pairing potential but include a bulky and hydrophobic substituent
20 that must be accommodated by the polymerase active site. Steric clashes in the active site are alleviated by the presence of a long, flexible linker. Indeed, we find biotin-16-dUTP a much 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
25 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
30 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
5 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
10 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
15 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
20 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
25 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
30 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.

- 5 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.
 - 10 (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
- 15 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
- 20 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

25 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
 5 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.

10 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
 15 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.

20

Example 1

General Methods

25 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
- 30 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 GG 3'; G•G mismatch
 7: 5'-GAA CTG CGG GTG ACG CCA AGC GCG-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'
 5 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'
 10 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'
 15 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 XCG GCA TCC GTC GCG ACC
 20 ACG GTC GCA GCG-3'; X= abasic site
 24: 5'-AGC TAC CAT GCC TGC ACG AAX XCG 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'
 25 27: 5'-ACT GXT 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
 30 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 MgCl₂), 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.

5

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 ³²P-labeled and annealed to one of template strands 18, 19, 20 (template base = C, G, A respectively) or 21 (template base C

30

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.

5

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 ³²P-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 / *Sa*II-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 ³²P-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 MgCl₂, 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 ³²P-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.

30

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 V_{max} values for the mutant polymerases, while K_m for nucleotide substrates remains largely unchanged. For most DNA polymerases the relative efficiency of extending a given mispair (f_{ext}) is similar to the relative efficiency of forming it (f_{inc}) (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)	$f^{\dagger}$	$f_{\text{ext}}^{\ddagger}$	Ratio of $f_{\text{ext}}^{\S}$
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 = $V_{\max} / K_m$

[‡] $f_{\text{ext}}, f(\text{mismatched 3' terminus}) / f(\text{matched terminus})$

5 [§] $f_{\text{ext}}(\text{mutant polymerase}) / f_{\text{ext}}(\text{wtTaq})$

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

Base pair *	Polymerase	$V_{\max}$ (%Min ⁻¹)	K_m (μ M)	$f^{\dagger}$	$f_{\text{inc}}^{\ddagger}$	Ratio of $f_{\text{inc}}^{\S}$
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 = $V_{\max} / K_m$

[‡] $f_{\text{inc}}, f(\text{incorrect dNTP}) / f(\text{correct dNTP})$

5 [§] $f_{\text{inc}}(\text{mutant polymerase}) / f_{\text{inc}}(\text{wtTaq})$

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 pyrimidine 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.

5

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 (α S dNTPs), in which one of the oxygen atoms in the α phosphate group is replaced by sulfur. As part of a dNTP mixture, α S dNTPs are generally well accepted as substrates by DNA polymerases but when we replaced all four dNTPs with their α S 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 α S dNTPs with much increased efficiency compared to the wt enzyme (Fig. 6). As expected, the resulting α S 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 premature 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 candidate "universal base" 5-nitroindole (5NI) (Loakes & Brown 96), which, like other hydrophobic, strongly stacking base analogues, is readily accepted as a substrate, but

30

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).

5 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
10 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
15 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

20 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. .

25

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
30 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
 5 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.

10 **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
 15 CCG AAC TGC-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 *DpnI* to remove methylated plasmid DNA,
 treated with ExoSAP (USB) to remove residual primers, reamplified with outnested
 20 primers and recloned and transformed into *E. coli* as above.

The resultant clones were screened and ranked by PCR assay. Briefly, 2 μ l of
 induced cells were added to 20 μ l 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.

25

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,
 30 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
5 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% GmHCl. 7 µl of purified replication products (from 48) were digested with 1µl *DpnI* (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
10 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 Sall, recloned into pASK75 and transformed into *E. coli* as above.

15

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
20 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
25 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
30 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.

20

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.

30

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.

5

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

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
 5 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.

10

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).

15

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

20

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) .

25

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:

- 5 Extension reactions are carried out in 1x Taq buffer including 1-100nM of hairpin primer and 100 μ M 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. (1-10 μ l) Reaction products are added to Streptavidin coated ELISA plates (Streptawell, Roche) in 200 μ l PBS, 0.2% Tween20 (PBST) and
- 10 incubated at room temperature for 10min to 1h. ELISA plates are washed 3x in PBST and 200 μ l 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.

15 **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.

20

Clones were grown at 30°C overnight in 200 μ l 2XTY + ampicillin (100 μ g/ml).

A 150 μ l (2xTY + ampicillin 100 μ g/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 50 μ l of 2XTY + anhydrous tetracycline (8ng/ml) to the culture which was

25 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 100 μ l 1xTaq buffer (10mM Tris-HCl, pH 9.0, 1.5mM MgCl₂, 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

30 was pelleted at 2254xg for 5 minutes.

ELISA protocol:**Extension reaction.**

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

- 5 1x Taq buffer (10mM Tris-HCl, pH 9.0, 1.5mM MgCl₂, 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.
10 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.

15

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
20 200µl PBS-Tween + anti-digoxigenin-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 H₂O₂, the reaction was allowed to
25 develop after which it was stopped by adding 100µl of 1M H₂SO₄.

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
30 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.

5

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.

10

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).

15

Experiment V. Abasic site ELISA ...

20 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.

25

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

A1= *Tth* Wild-type

30 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

5

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

Experiment V. Abasic site and 5-hydroxyhydantoin bypass

10 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.

15 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
20 nucleoside which was dimethoxytritylated and phosphitylated 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.

25 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.

30 2F3:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
35 CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCGGCTACGAGGCGGACGACGTTCTCGCCACCGTGGCCAAG

AAGGCGGAAAAGGAGGGGTACGAGGTGGGCATCCTACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGC
GTCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGC
AGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGTCAAGGGCATCGGGGAGAAGA
CCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAA
5 ACGTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCACACCGACCT
CCCCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGA
GTTTGGCAGCCTCCTCCACGAGTTCCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCG
GAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAG
GGGGGGCCGGGTCCACCGGGCCCCGAGCCTTATAAAGCCCTCAGAGACCTGAAGGAGGCGCGGGGGCTTCTCGC
10 CAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTAC
CTCCTGGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGG
GAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGC
TTTACCGGGAGGTGGAGAGGCCCCCTTTCCGTTGTCTGGCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGC
CTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCAGAGCCGAGGTCTTCCGCCTGGCCGGC
15 CACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCA
AGACGGAGAAGACCGGCAAGCGCTCCACCGGCGCCGCCGTCTGGAGGCCCTCCACGAGGCCACCCCATCGTGG
AGAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCC
CAGGACGGGCCGCCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAAC
CTCCAGAACATCCCCGTCCGCACCCAGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTAT
20 TGGTGGTCTCTGACTATAGCCAGATAGAGCTCAGGGTGTCTGGCCACCTCTCCGGCGACGAGAACCTGATCCGGGT
CTTCCAGGAGGGGCGGGACATCCACACGGAACCGCCAGCTGGATGTTCCGGCGTCCCCCAGGAGGCCGTGGACCC
CCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCCGGCCTACCGCCTCTCCAGGAG
CTAGCCATCCCTTACGAGGAGGCCCAGGCCTTCAATTGAGCGCTACTTTCAGAGCTTCCCAAGGTGCGGGCCTGGAT
TGGGAAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCCGGCCGCCCGCTACGTGCCAGA
25 CCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACACGCCCGTCCAGGGCACCGC
CGCCGACCTCATGAAGCTAGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTACG
GTCCACGACGAGCTGGTCTCGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATG
GAGGGGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAG
TGA
30 MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVPQVVYGFASLLKALKEDGYKAVFVVFDAKAPSRHKAY
EAYRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATVAKKAEKEGYEVGILTADRGLYQLVSDRVAVLH
PEGLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGEKALKLLKEWGSLENLLKNLDRVKPENVREKIKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFFVGFVLSRKE
35 PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERALSERLFANLWGRLEGEERLLWLYREVERPLSVVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLFDLGLPAIGKTEKTGKRSTGAHVLEALHEAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTQLGQIRRAFIAEEGWLLVLDYSQIELRVLAHLSGDNELIRVFQE
GRDIHTETASWMFGVPQEAVDPLMRRAAKTINFGVLYGMSAYRLSQELAIPYEEAQAFIERFYQSFPKVRWIGKTLLE
40 GRRRGYVETLFGRRRYVPDLEARVKSREAAERMAFNTPVQGTAAADLMKLMVKLFPRLEEMGARMMLQVHDELVL
APKERAEAVARLAKEVMGCVYPLAVPLEVEVGIGEDWLSAKE*

45 1A10:
ATGCGTGGTATGCCTCCTCTTTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACACGAGCCGGGGCGAACCAGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCCTCCGCCACGAGG
50 CTTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCGAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCTGGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTACCGCCGACCGCGACCTCTACCAACTCGTCTCCGACCGC
GTCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGC
AGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGTCAAGGGCATCGGGGAGAGGA
CCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAA
55 ACGTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCACACCGACCT
CCCCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGA
GTTTGGCAGCCTCCTCCACGAGTTCCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCG
GAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCCAG
GGGTGGTCCGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCC
60 AAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACC
TCCTGGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGG
AGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAAGCTTGAGGGGGAGGAGAGGCTCCTTTGGCT
TTACCGGGAGGTGGATAGGCCCTTTCCGCTGTCTGGCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCC
TATCTCAGGGCCTCGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCC
65 ACCCCTTCAACCTCAACTCCCGGGACAGCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAA
GACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCCGCTCTGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGGA
GAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCC
AGGACGGGCCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACAGGCAGGCTAAGTAGCTCCGATCCCAACC
TCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATT
70 GGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGTGGCCACCTCTCCGGCGACGAGAACCTGATCCGGGTC
TTCCAGGAGGGGCGGGACATCCACACGAGACCGCCAGTTGGATGTTCCGGCGTCCCCCGGGAGGCGGTGGACCCCC

- 5 TGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCCTCTACGGCATGTGCGGCCCGCCGCCTCTCCCAGGAGCT
AGCCATCCCTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATT
GAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCGCCGCTACGTGCCAGAC
CTAGAGGCCCGGTGAAGAGCGTGCAGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCCGTCCAGGGCACCGCC
10 GCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGG
TCCACGACGAGCTGGTCCTCGAGGCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGG
AGGGGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGT
GA
- 15 MRGMPPLFEPKGRVLLVDGHLAYRTFFALKGPTTSRGEVPQAVYGFASLLKALKEDGYKAVFVVFDAKAPSLRHEAY
EAYKAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGERTALKLLKEWGSLENLLKNLDRVKPENVREKIK
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDRELRRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVAR
15 YGGEWTEEAGERAAALSERLFANLWGKLEGEERLLWLYREVDRLSAVLAHMEATGVRLDVAYLRASSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLDFELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQ
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSARRLSQELAIPEYEEAQAFIERYFQSPFKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSUREAAERMAFNMPVQGTADLMKLAMVKLFPRLLEEMGARMMLLQVHDELVL
20 APKERAEEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE*
- 1A9:
- 25 ATGCGTGGTATGCATCCTCTTTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCAC
CTTCCACGCCCTGAAGGGGCTCACCACCAGCCGGGGGAGCCGGTGCAGGCGGTCCACGGCTTCGCCAAGAGCCTC
CTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCTTCCGCCACG
AGGCCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGG
AGCTGGTGGACCTCCTGGGGTTTACCCGCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGC
30 CAAGAAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCAGCGACCGCGACCTCTACCAACTCGTCTCCGA
CCGCGTCGCCGTCTTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCG
GAGCAGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAG
AAGACCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGCTGAAGCCC
GCCATCCGGGAGAAGATCCTGGCCACATGGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGTGCGCACCGACC
35 TGCCCTAGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTG
AGCTTGCGACGCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGACCCTGGAGGAGGCCTCCTGGCCCCCGCC
GGAAGGGGCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCGGATCTTCTGGCCCTGGCCGCCGCCA
GGGGGGGCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGAGACCTGAAGGAGGCGCGGGGGCTTCTCG
CCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTA
40 CCTCCTGGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGG
GGAGCGGGCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGG
CTTTACCGGGAGGTGGAGAGGCCCCTTCCGTTGTCTGGCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGG
CCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGG
CCACCCCTTCAACCTCAACTCCCGGGACAGCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGC
45 AAGACGGAGAAGACCGGCAAGCGCTCCACCGGCGCCGCGCTCCTGGAGGCCCTCCGCGAGGCCACCCCATCGTG
GAGAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACC
CCAGGACGGGCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAA
CCTCCAGAACATCCCCGTCCGCACCCAGCTTGGGCAGAGGATCCGCGGGGCTTCATCGCCGAGGAGGGGTGGCTA
50 TTGGTGGTCTTGGACTATAGCCAGATAGAGCTCAGGGTGTGGCCACCTCTCCGGCGACGAGAACCTGATCCGGG
TCTTCCAGGAGGGGCGGGACATCCACACGAAACCGCCAGCTGGATGTTCCGGCGTCCCCCAGGAGGCCGTGGACCC
CCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTGCGCCTACCGCCTCTCCCAGGAG
CTAGCCATCCCTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGAT
TGGGAAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCGCCGCTACGTGCCAGA
CCTAGAGGCCCGGGTGAAGAGCGTGCAGGAGGCGGCCGAGCGCATGGCCTTCAACACGCCCCGTCCAGGGCACCGC
55 CGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAG
GTCCACGACGAGCTAGTCTCGAGGCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATG
GAGGGGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAG
TGA
- 60 MRGMHPLFEPKGRVLLVDGHLAYRTFHALKGLTTSRGEVPRAVHGFASLLKALKEDGYKAVFVVFDAKAPSRHE
AYEAYKAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAV
LHPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGIKTALKLLKEWGSLENLLKNLDRKPAIRKILA
HMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLELGSLLHEFGLLESPKTEEASWPPPEGAFVGFVLSRK
EPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVAR
65 RYGGEWTEEAGERAAALSERLFANLWGRLEGEERLLWLYREVERPLSVLAHMEATGVRLDVAYLRALSLEVAEEIARL
EAEVFRLAGHPFNLNSRDQLERVLDFELGLPAIGKTEKTGKRSTGAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLD
LIHPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTQLGQRIRRAFIAEEGWLLVLDYSQIELRVLAHLSGDENLIRVFQ
EGRDIHTETASWMFGVPQEAVDPLMRRAAKTINFGVLYGMSAYRLSQELAIPEYEEAQAFIERYFQSPFKVRAWIGKTL
EGRRGYVETLFGRRRYVPDLEARVKSUREAAERMAFNTPVQGTADLMKLAMVKLFPRLLEEMGARMMLLQVHDELVL
70 APKERAEEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE*

2F12:

5 ATGCGTGGTATGCTTCCTCTTTTGTAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCAC
CTTCTTCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCAGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTC
CTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCCTCCGCCACG
AGGCCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGG
10 AGCTGGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGC
CAAGAAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGACCTCTACCAACTCGTCTCCGA
CCGCGTCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCG
GAGCAGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAG
AAGACCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGCTGAAGCCC
GCCATCCGGGAGAAGATCCTGGCCACATGGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGTGCGCACCGACC
15 TGCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCCCGACGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTG
AGCTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCTCCTGGCCCCCGCC
GGAAGGGGCCCTTCGTGGGCTTTGTGCTTACCCGCAAGGAGCCCATGTGGGCGGATCTTCTGGCCCTGGCCGCCGCC
AGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTC
GCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCT
20 ACCTCCTGGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGG
GGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTG
GCTTTACCGGGAGGTGGAGAGACCCCTTTCCGCTGTCTGGCCACATGGAGGCCACGGGGGTGCGCCTGGACGTG
GCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCG
GCCACCCCTTCAACCTCAACTCCCGAGACCAGCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGG
25 CAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCGTCTGGAGGCCCTCCGCGAGGCCACCCCATCGT
GGAGAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCAC
CCCAGGACGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCA
ACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCTTTCATCGCCGAGGAGGGGTGGCT
ATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGACGAGAACCTGATCCGG
30 GTCTTCCAGGAGGGGCGGGACATCCACACGAGACCGCCAGCTGGATGTTTCGGCGTCCCCCGGGAGGCCGTGGAC
CCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTGCGCCACCGCCTCTCCCAGG
AGCTAGCCATCCCTTACGAGGAGGCCAGGCCTTATTGAGCGCTACTTTAGAGCTTCCCCAAGGTGCGGGCCTG
GATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCC
AGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCAC
35 CGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
AGGTCCACGACGAGCTCCTCCTGGAGGGCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCA
TGGAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGG
AGTGA

40 MRGMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGFAKSLKALKEDGYKAVFVVFDAKAPSLRHEA
YEAYKAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRLTADRDLVQLVSDRVAVL
HPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGIEKTALKLLKEWGSLENLLKNLDRKPAIRKILAH
MDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERLLEGLSLLHEFGLLESPKALEEASWPPPEGAFFVGFVLRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLPLPGDDPMLLAYLLDPSNTTPEGVARR
45 YGGEWTEEAGERALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLSRDQLERVLDFELGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPLDI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPIYEEAQAFIERYFQSPFKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSUREAERMAFNPVQGTAAADLMKMLAMVKLFPRLREMGARMMLLQVHDELLLE
50 APQARAEVAALAKEAMEKAYPLAVPLEVKVGIGEDWLSAKE*

1C2:

55 ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCAGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
60 CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCCTATACCAACTCGTCTATGACCGC
GTCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGC
AGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGA
CCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAA
65 ACGTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGGCGACCGACCT
CCCCCTGGAGGTGGACCTCGCCAGGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGA
GTTTGGCAGCCTCCTCCACGAGTTCCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCG
GAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCGGATCTTCTGGCCCTGGCCGCCGCCAG
GGGTGGTGCAGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCC

5 AAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACC
 TCCTGGACCCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGG
 AGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCT
 TTACCGGGAGGTGGAGAGGCCCCCTTTCCGCTGTCTTGGCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCC
 10 TATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCC
 ACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAATGGTGCTCTTTGACGAGCTTAGGCTTCCCGCCTTGGGGAAG
 ACGCAAAAGACGGGCAAGCGCTCCACCAGCGCCGCGTCTTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGAG
 AAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGTGCGACCTCATCCACCCA
 GGACGGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCT
 15 CCAGAACATCCCCGTCCGACCCCGCTTGGGCAGAGGATCCGCGGGGCTTCATCGCCGAGGAGGGGTGGCTACTG
 GTGGTCTTGGACTATAGCCAGATAGAGCTCAGGGTGTGGCCACCTCTCCGCGACGAAAACCTGATCAGGGTCT
 TCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTGCGCGTCCCCCGGAGGCCGTGGACCCCC
 TGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTGCGCCACCGCCTCTCCAGGAGCT
 AGCCATCCCTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCAAGGTGCGGGCCTGGATT
 20 GAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCGCCGCTACGTGCCAGAC
 CTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCC
 GCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGG
 TCCACGACGAGCTGGTCTCGAGGCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTGATGG
 AGGGGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGT
 GA

25 MAMLPFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVPQVVYGFASLLKALKEDGYKAVFVVFDAKAPSFRHKAY
 EAYRAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRLTADRGLYQLVYDRVAVLH
 PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGEKTALKLLKEWGSLENLLKNLDRVKNPVREKIK
 HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFFVGFVLSRKE
 PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
 YGGEWTEEAGERALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEBELARLE
 AEFRLAGHPFNLSRDQLEMVLFDELRLPALGKTQKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLSDL
 30 IHPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQIRRAFIAEEGWLLVVDYSQIELRVLAHLSGDENLIRVQE
 GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPIYEEAQAFIERFYQSFPKVRAWIEKTLEEG
 RRRGYVETLFGRRRYVPDLEARVKSREAAERMAFNPVQGTAAADLMKMLAMVKLFPRLLEMGARMMLQVHDELVLE
 APKERAEEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE*

35

2G6:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
 CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCAGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
 40 AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
 CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGACGCTCGCCCTCATCAAGGAGCT
 GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGGCGGACGACGTTCTCGCCACCTTCGCCAAG
 AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCG
 TCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
 GTGGGTGGACTTCCGCGCCCTCGTGGGGAACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
 45 GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
 GTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCC
 CCCTGGAGGTGGACCTCGCCAGGGGCGGGAGCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
 TGGCAGCCTCCTCCACGAGTTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAA
 50 GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGGCGATCTTCTGGCCCTGGCCGCCCGCCAGGGG
 TGGTTCGAGTCCACCGGGCCCCCGAGCCTTATAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
 GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
 GGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
 GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTAC
 55 CGGGAGGTGGAGAGGCCCCCTTTCCGCTGTCTTGGCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATC
 TCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCTGGCCGGCCACCC
 CTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
 GAGAAGACCGGCAAGCGCTCCACCAGCGCCGCGTCTTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGAGAAG
 ATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGGCCGACCTCATCCACCCAGGA
 CGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
 60 GAACATCCCCGTCCGCACCCGCTCGGGCAGAGGATCCGCCGGGCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
 GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGTGGCCACCTCTCCGCGACGAGAACCTGATCCGGGTCTTCC
 AGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTTCGGCGTCCCCCGGAGGCCGTGGACCCCTAA
 TGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTGCGCCCGCGCCTCTCCAGGAGCTAGC
 CATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCAAGGTGCGGGCCTGGATTGAG
 65 AAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCGCCGCTACGTGCCAGACCTA
 GAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
 GACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTCC
 ACGACGAGCTGGTCTCGAGGCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTGATGGAGG
 GGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTTTCCGCCAAGGGTTAG

70

Above: nucleic acid sequence of the clone

5 MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVQVVGFAKSLKALKEDGYKAVFVVFDAKAPSFRHKAY
 EAYRAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATFAKKAKEGYEVRLTADRGLYQLVSDRVAVLH
 PEGHLITPEWLWEKYGLRPEQWVDFRALVGNPSDNLPGVKIGIEKTALKLLKEWGSLENLLKNLDRVKPENVRKIKAKA
 HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPPKALEEAPWPPPEGAFVGFVLSRKE
 PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
 YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
 10 AEVFRLAGHPFNLNSRDQLERVLFDLGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
 HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRIRRAFIIEGWLLVVDYSQIELRVLAHLSGDNELIRVFQE
 GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSARLSQELAIPIYEEAQAFIERFYQSFPKVRWIEKTLEEG
 RRRGYVETLFGRRRYVPDLEARVKSVERAAERMAFNMPVQGTAAADLMKLAMVKLFPRLEEMGARMMLQVHDELVLE
 APKERAEAVARLAKEVMGVYPLAVPLEVEVGIGEDWLSAKG*

15 Above is the amino acid sequence of the clone

1A8:

20 ATGGTGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
 CTTCCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCAGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTC
 AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCTCCTCCTTCCGCCACGAGG
 CCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
 GGTGGACCTCCTGGGGTTTACCCGCTCGAGGTCCCCGGCTACGAGGTGGACGACGTCTGGCCAGCCTGGCCAAG
 AAGGTGGAAGAGGAGGGGTACGAGGTGCGCATCCTCACCAGCCGACCGCGACCTCTACCAACTCGTCTCCGACCGCG
 25 TCGCCGTCCTCCACCCCGAGGGCCACCTCATCACCAGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
 GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
 GCCCTCAAGCTCCTCAAGGAGTGGGGAGGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
 GTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCACACCGACCTCC
 CCCTGGAGGTGGACCTCGCCAGGGGCGGGAACCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
 30 TGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGAA
 GGGGCTTCTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCCGCCGAGGGG
 TGGTCGGGTCCACCGGACCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
 GACCTGAGCGTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
 GGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
 35 GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTAC
 CGGGAGGTGGATAGGCCCTTTCCGCTGTCTGGCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCCTACC
 TCAGGGCCTTGTCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCTGGCCGGCCACCC
 CTTCAACCTCAACTCCCGGGACAGCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
 GAGAAGACCGGCAAGCGCTCCACCAGCGCCCGCTCTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGAGAAG
 40 ATCCTGCAGTACCGGGAGCTCACCAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCAGGA
 CGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGAGGCTAAGTAGCTCCGATCCCAACCTCCA
 GAACATCCCCGTCCGACCCCGCTCGGGCAGAGGATCCGCCGGGCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
 GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCC
 AGGAGGGGCGGGACATCCACACGGAACCGCCAGCTGGATGTTTCGGCGTCCCCCGGGAGGCGGTGGACCCCTAA
 45 TGGCCCGGGCGGCAAGACCATCAACTTCGGGGTCTCTACGGCATGTTCGGCCACCGCCTCTCCAGGAGCTAGC
 CATCCCTTACGAGGAGGCCAGGCCTTATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAG
 AAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCGCTCGCTACGTGCCAGACCTA
 GAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCGCC
 GACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCAGGCTGGAAGAAACGGGGGCCAGGATGCTCCTTCAGGTCC
 50 ACGACGAGCTGGTCTCGAGGCCCAAAAGAGAGGGCGGAGGCGGTGGCCCGGCTGGCCAAGGAGGCCATGGAGG
 GGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

55 MVMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVQAVYGFAKSLKALKEDGYKAVFVVFDAKASSFRHEAY
 EAYKAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEVDDVLASLAKKVEKEGYEVRLTADRDLYQLVSDRVAVLH
 PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGIEKTALKLLKEWGGLLENLLKNLDRVKPENVRKIKAKA
 HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPPKALEEAPWPPPEGAFVGFVLSRKE
 PMWADLLALAAARGGRVHRTPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
 YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
 60 AEVFRLAGHPFNLNSRDQLERVLFDLGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
 HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRIRRAFIIEGWLLVVDYSQIELRVLAHLSGDNELIRVFQE
 GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPIYEEAQAFIERFYQSFPKVRWIEKTLEEG
 RRRGYVETLFGRRRYVPDLEARVKSVERAAERMAFNMPVQGTAAADLMKLAMVKLFPRLEETGARMMLQVHDELVLEA
 PKERAEAVARLAKEAMEGVYPLAVPLEVEVGIGEDWLSAKE*

65

2H1:

ATGGTGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
 CTTTCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTC
 AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCTCCTCCTTCCGCCACGAGG
 5 CCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
 GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGTGGACGACGTCTGGCCAGCCTGGCCAAG
 AAGGTGGAAGAGGAGGGGTACGAGGTGCGCATCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCG
 TCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCTCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
 GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
 10 GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
 GTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGTGCGCACCGACCTCC
 CCCTGGAGGTGGACCTCGCCAGGGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
 TGGCAGCCTCCTCCACGAGTTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGAA
 GGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCCGCAGGGG
 15 TGGTCGGGTCCACCGGGCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
 GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
 GGACCCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
 GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTAC
 CGGGAGGTGGATAGGCCCTTTCCGCTGTCTGGCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCCTATC
 TCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCTGGCCGCCACCC
 20 CTTCAACCTCAACTCCCGGGACCACTGGAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
 GAGAAGACCGGCAAGCGCTCCACCAAGCGCCGCCATCCTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGAGAAG
 ATCCTGCAGTACCGGGAGCTCACCAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCAGGA
 CGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
 GAACATCCCCGTCCGACCCCGCTCGGGCAGAGGATCCGCGGGGCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
 25 GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGTCTGGCCACCTCTCCGGCGACGAGAACCTGACCCGGGTCTTCC
 AGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTTCGGCGTCCCCCGGGAGGCCGTGGACCCCTGA
 TCGCCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCTGGCCACCGCCTCTCCAGGAGCTGGC
 CATCCCTTACGAGGAGGCCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGGATAGAA
 AAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAACCTCTTCGGAAGAAGGCGCTACGTGCCCCGACCTC
 30 AACGCCCGGTGAAGAGTGTACGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
 GACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCTCCGGGAGATGGGGGCCCGCATGCTCCTCCAGGTCC
 ACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATGGAGA
 AGGCCTATCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCCAAGGAGTGAG
 35 TCGACCTGCAGGCAGCGCTTGGCGTCACCCGCAGTTCGGTGGTTAA

MVMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGFASLLKALKEDGYKAVFVVFDAKASSFRHEAY
 EAYKAGRAPTEDFRQLALIKELVDLLGFTRLEVPGYEVDDVLASLAKKVEKEGYEVRLTADRGLYQLVSDRVAVLH
 PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPNDLPGVKIGEKTAALKLKEWGSLENLLKNLDRVKPENVREKIK
 40 HLEDLRLSLELSRVRTDLPLEVDLAQGREPDRELRRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFFVGFVLSRKE
 PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
 YGGEWTEEAGERAAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
 AEVFRLAGHPFNLNSRDQLERVLFDLGLPAIGKTEKTGKRSTSAAILREALREAHPIVEKILQYRELTKLKSTYIDPLDLI
 45 HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRIRRAFAIEEGWLLVLDYSQIELRVLAHLSGDENLTRVFQE
 GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPIYBEAQAFIERYFQSFPKVRAWIEKTLEEG
 RKRGYVETLFGRRRYVPDLNARVKSVERAAERMAFNMPVQGTADLMKLAMVKLFRLREMGARMMLLQVHDELLLE
 APQARAEVAALAKEAMEKAYPLAVPLEVKVGIGEDWLSAQGVSRPAGSAWRHPQFGG*

50

2F11:

ATCGTGATGCTTCCCTCTTTTGGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCAC
 CTTCTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTC
 55 CTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCGCTTCGTGGTCTTTGACGCCAAGGCCCTCCTTCCGCCACG
 AGGCCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGG
 AGCTGGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCTGGCTACGAGGCGGACGACGTCTCGCCACCCTGGC
 CAAGAAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCTCACCGCCGACCGCGACCTCTACCAACTCGTCTCCGA
 CCGCGTCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCTCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCG
 60 GAGCAGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAG
 AAGACCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCA
 GAAAACGTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGGCACC
 ACCTCCCCCTGGAGGTGGACCTCGCCAGGGGCGGGAGCTCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCT
 TGAGTTTGGCGGCCTCCTCCACGAGTTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCG
 65 CCGGAAGGGGCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCG
 CAGGGGTGGTGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTC
 GCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCT
 ACCTCCTGGACCTTCCAACACCGCCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGG
 GGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTG
 70 GCTTTACCGGGAGGTGGATAGGCCCTTTCCGCTGTCTGGCCACATGGAGGCCACAGGGGTACGGCTGGACGTG
 GCCTGCCTGCAGGCCCTTTCCCTGGAGCTTGGCGAGGAGATCCGCCCGCTCGAGGAGGAGGTCTTCCGCTTGGCGG
 GCCACCCCTTCAACCTCAACTCCCGGGACCACTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGG

CAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGCCCTCCGCGAGGCCACCCCATCGT
 GGAGAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCAC
 CCCAGGACGGGCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCA
 ACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCTTCGTCGCCGAGGAGGGGTGGCT
 5 ATTGGTGGTCTGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGACGAGAACCTGACCCGG
 GTCTTCTGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCCGGCGTCCCCCGGAGGCCGTGGACC
 CCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCCTACGGCATGTCGGCCACCGCCTCTCCCAGGA
 GCTGGCCATCCCTTACGAGGAGGCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGG
 ATAGAAAAGACCCTGGAGGAGGGAGGAAGCGGGGCTACGTGGAAACCTCTTCGGAAGAAGGCGCTACGTGCC
 10 GACCTCAACGCCCCGGTGAAGAGTGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCCGTCAGGGCACC
 GCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
 AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCA
 TGGAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGG
 AGTGA
 15 MRGMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVPQAVYGFASLLKALKEDGYKAAFVVFDAKAPSRHEA
 YEAYKAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRDLYQLVSDRVAVL
 HPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGIEKTALKLLKEWGSLENLLKNLDRVKPNVREKIK
 AHLEDLRLSLELSRVRTDLPLEVDLAQGRELDRELRALFLERLEFGGLLHEFGLLLESPKALEEAPWPPPEGAFVGFVLSRK
 20 EPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTAPEGVA
 RRYGGEWTEEAGERAAALSERLFANLWGRLEGEERLLWLYREVDRLSAVLAHMEATGVRLDVACLQALSLELAEEIRR
 LEEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAILREALREAHPIVEKILQYRELTKLKSTYIDPLPD
 LIHPRTGRLHTRFNQTATATGRLSSDPNLQNIPTPLGQRIRRAFVAEEGWLLVLDYSQIELRVLAHLSGDENLTRVF
 25 LEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHLRSQELAIPIYEEAQAFIERYFQSFPKVRAWIEKTLE
 EGRKRGYVETLFGRRRYVPDLNARVKSREAAERMAFNPVQGTAAADLMKMLAMVKLFPRLREMGARMMLLQVHDELL
 LEAPQARABEVAALAKEAMEKAYPLAVPLEVKVGIGEDWLSAKE*

30

2H4:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
 35 CTTCCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
 AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
 CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
 GGTGGACCTCCTGGGGTTTACCCGCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
 AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCG
 40 TCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
 GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
 GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
 GTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGTGCGCACCGACCTCC
 CCCTGGAGGTGGACCTCGCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
 45 TGGCAGCCTCCTCCACGAGTTCCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAA
 GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCGCCAGGGG
 TGGTCGAGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
 GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
 50 GGACCCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
 GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTAC
 CGGGAGGTGGAGAGGCCCCCTTCCGCTGTCTGGCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATC
 TCAGGGCCTTGTCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCTGGCCGGCCACCC
 CTTCAACCTCAACTCCCGGGACCACTGGAAATGGTGCTCTTTGACGAGCTTAGGCTTCCCGCCTTGGGGAAGACG
 CAAAAGACGGGCAAGCGCTCCACCAGCGCCGCGTCTTGAGAGGCCCTCCGCGAGGCCACCCCATCGTGGAGAAG
 55 ATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGTTCGGACCTCATCACCCAGGA
 CGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
 GAACATCCCCGTCCGCACCCCGCTTGGGCAAGGATCCGCCGGGCTTCATCGCCGAGGAGGGGTGGCTACTGGTG
 GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGACGAAAACCTGATCAGGGTCTTCC
 AGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCCGGCGTCCCCCGGAGGCCGTGGACCCCTGA
 60 TGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTCCGGCCACCGCCTCTCCAGGAGCTAGC
 CATCCCTTACGAGGAGGCCACGGCCTTCAATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAG
 AAGACCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCTCTTCGGCCGCGCCGCTACGTGCCAGACCTA
 GAGGCCCGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
 GACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAACGGGGGCCAGGATGCTCCTTCAGGTCC
 65 ACGACGAGCTGGTCTTGAGGCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGTGGCCAAGGAGGTGATGGAGG
 GGGTGTATCCCCTGGCCGTGTCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

70

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVPQVVYGFASLLKALKEDGYKAVFVVFDAKAPSRHKAY
 EAYRAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGYQLVSDRVAVLH
 PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGIEKTALKLLKEWGSLENLLKNLDRVKPNVREKIK
 HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLLESPKALEEAPWPPPEGAFVGFVLSRKE

5 PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
 YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
 AEFVRLAGHPFNLNSRDQLEMVLFDELRLPALGKTQKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLSDL
 IHPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQIRRAFIAEEGWLLVVDYSQIELRVLAHLSGDENLIRVFQE
 GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPEYEEAQAFIERYFQSFPKVRWIEKTLEEG
 RRRGYVETLFGRRRYVPDLNARVKSUREAAERMAFNMPVQGTAAADLMKLAMVKLFPRLEETGARMMLQVHDELVLEA
 PKERAEAVARLAKEVMIEGVYPLAVSLEVEVGIGEDWLSAKE*

10

2H9:

ATGGCGATGCTTCCCCCTCTTTGAGCCCAAGGGCCGCGTCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
 CTTCGCCCTGAAGGGCCCCACCGCGAGCCGGGGCGAACCAGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCTC
 AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGG
 15 CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
 GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCCCCTGGCCAAG
 AAGGCGGAAAAGGAGGGGTTTCGAGGTGCGCATCTCCCCCGCGTCCGCGGCCTCTGCCCTCTCGTCTCTGACCGCG
 TCGCCGTCCTCCTCCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
 GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGAAGAAGACC
 20 GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
 GTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGGCACCAGCCTCC
 CCCTGGAGGTGGACCTCGCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
 TGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGAA
 GGGGCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCGCCAGGGG
 25 TGGTCCGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
 GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
 GGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
 GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTGTGGCTTTAC
 CGGGAGGTGGATAGGCCCTTTCCGCTGTCTGGCCACATGGAGGCCACAGGGGTACGGCTGGACGTGGCCTGCC
 30 TGCAGGCCCTTTCCCTGGAGCTTGGCGAGGAGATCCGCCGCTCGAGGAGGAGGTCTTCCGCTTGGCGGGCCACCC
 CTCAACCTCAACTCCCGGGACAGCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
 GAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGAAGA
 ATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCAGGA
 CGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
 35 GAACATCCCCGTCCGCACCCGCTCGGGCAGAGGATCCGCCGGGCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
 GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGACGAGAACCTGACCCGGGTCTTCC
 AGGAGGGGCGGGACATCCACACGGAACCGCCAGCTGGATGTTTCGGCGTCCCCCGGGAGGCCGTGACCCCTGA
 TCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTGCGGCCACCGCCTCTCCAGGAGCTGGC
 CATCCCTTACGAGGAGGCCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGGATAGAA
 40 AAGACCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCTCTTCGGAAGAAGGCGCTACGTGCCCCGACCTC
 AACGCCCGGTGAAGAGTGTGAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCCGTCCAGGGCACCGCCGCC
 GACCTTATGAAGCTCGCCATGGTGAAGCTTCCCCCGCCTCCGGGAGATGGGGGCCCCGCATGCTCCTCCAGGTCC
 ACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATGGAGA
 AGGCCTATCCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

45

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTASRGEPVQVVYGFAKSLLKALKEDGYKAVFVFDKAPSFRHKA
 YEAYRAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLAPLAKKAEKEGFEVRILPAVRGLCPLVSDRVAVLL
 PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPNDLPGVKIGKKTALKLLKEWGSLENLLKNLDRVKPENVREKIK
 50 HLEDLRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLESPLKALEEAPWPPPEGAFFVGFVLSRKE
 PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
 YGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVDRLSAVLAHMEATGVRLDVACLQALSLELAEEIRRL
 EEFVRLAGHPFNLNSRDQLERVLFDLGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLDLI
 PRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQIRRAFIAEEGWLLVVDYSQIELRVLAHLSGDENLTRVFQEG
 55 RDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPEYEEAQAFIERYFQSFPKVRWIEKTLEEG
 KRGYVETLFGRRRYVPDLNARVKSUREAAERMAFNMPVQGTAAADLMKLAMVKLFPRLEEMGARMMLQVHDELLLEA
 PQARAEVAALAKEAMEKAYPLAVPLEVKVGIGEDWLSAKE*

60

1B12:

ATGGCGATGCTTCCCCCTCTTTGAGCCCAAGGGCCGGGTCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
 CTTCGCCCTGAAGGGCCTCATCACGAGCCGGGGCGAACCAGGTGCAGGCGGTCTACGGTTTCGCCAAGAGCCTCTC
 AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACGAGG
 65 CCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
 GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCAAGGCTACGAGGCGGACGACGTCCTCGCCACCCTGGCCAAG
 AAGGCGGAAAAGAAGGGTACGAGGTGCGCATCTCACCGCCGACCGGGACCTCTACCAGCTCGTCTCCGACCGC
 GTCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGC
 AGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGA
 70 CCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAATCTCCTCAAGAACCTGGATCGGGTAAAGCCGGAAA

ACGTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGTACCGACCT
CCCCCTGGAGGTGGACCTCGCCAGGGGCGGGAGCCCGACCGGGAAGGGCTTAGGGCCTTCCTGGAGAGGCTGGA
GTTTCGGCAGCCTCCTCCATGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCG
5 GAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCGGCCAG
GGGTGGTCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCC
AAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACC
TCCTGGACCCCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGG
AGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCT
10 TTACCGGGAGGTGGATAGGCCCTTTCCGCTGTCCTGGCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCC
TATCTCAGGGCCTTGTCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCC
ACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCTCTTTGACGAGTTAGGGCTTCCCGCCATCGGCAA
GACGGAGAGGACCGGCAAGCGCTCCACCAGCGCCGCGCTCTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGA
GAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCC
15 AGGACGGGCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACC
TCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCTTCATCGCCGAGGAGGGGTGGCTATT
GGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGACGAGAACCTGATCCGGGTC
TTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTTCGGTGTCCCCCGGAGGCCGTGGACCC
TGATGCGCCGGGCGGCCAAGACGGTGAACCTTCGGCGTCTCTACGGCATGTCCGCCCATAGGCTCTCCAGGAGCT
20 TTCCATCCCTACGAGGAGGCGGTGGCCTTTATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGGATA
GAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAACCCCTCTTCGGAAGAAGGCGCTACGTGCCCGAC
CTCAACGCCCGGGTGAAGAGCGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCC
GCCGACCTCATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCCAGG
TCCACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATGG
25 AGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGGAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGT
GA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLITSRGEFVQAVYGFASLLKALKEDGYKAVFVFDKAPSFRHEAY
EAYKAGRPTPEDFPRQLALIKELVDLLGFTRLEVQGYEADDVLATLAKKAEKEGYEVRLTADRDLYQLVSDRVAVLH
30 PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGIEKTALKLLKEWGSLENLLKNLDRVKPNVREKIK
HLEDRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRALSLEVAEBIARLE
AEVFRLAGHPFNLNSRDQLERVLDELGLPAIGKTERTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLD
35 HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVTRPLGQRIRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQ
GRDIHTETASWMFGVPPEAVDPLMRRAAKTVNFGVLYGMSAHRLSQELSIPIYEEAVAFIERYFQSPKPVRAWIEKTLEEG
RKRGYVETLFGRRRYVPDLNARVKSUREAAERMAFNPVQGTAAADLMKLAMVKLFPRLREMGARMLLQVHDELLLE
APQARABEEVAALAKEAMEKAYPLAVPLEVEVGIGEDWLSAKE*

40

2H2:
45 ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCAGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGGCCCCCTCATTCCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
50 GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTACCGCCGACCGCGGCTCTACCAACTCGTCTCTGACCGCG
TCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
GTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGGCGACCGACCTCC
55 CCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
TGGCGGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAA
GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCGGCCAGGGG
TGGTCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
60 GACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCT
GGACCCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTAC
CGGGAGGTGGAGAGGCCCTTTCCGTTGTCCTGGCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCCTATC
TCAGGGCCTTGTCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCC
65 CTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
GAGAAGACCGGCAAGCGCTCCACCGGCGCCGCGCTCTGGAGGCCCTCCGCGAGGCCACCCACCGTGGAGAAG
ATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCAGGA
CGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGACCCCAACCTCCA
GAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCTTCATCGCCGAGGAGGGGTGGCTATTGGTG
70 GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTCC
AGGAGGGGCGGGACATCCACACGGAACCGCCAGCTGGATGTTTCGGCGTCCCCCGGGAGGCCGTGGACCCCTAA
TGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTCCGCCACCGCCTCTCCAGGAGCTAGC

5 CATCCCTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACATTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAG
AAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCGCTCGCTACGTGCCAGACCTA
GAGGCCCGGGTGAAGAGCGTGCAGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCCGTCCAGGGCACCAGCCGCC
GACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAAGAAACGGGGGCCAGGATGCTCCTTCAGGTCC
ACGACGAGCTGGTCCTCGAGGCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGCCATGGAGG
GGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

10 MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVPVQVYGFASLLKALKEDGYKAVFVFDKAPSFRHKAY
EAYRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRLTADRGLYQLVSDRVAVLH
PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGEKTAALKLLKEWGSLENLLKNLDRVKPENVREKJKA
HLEDLRLSLELSRVRTDLPLEVDLAQGREPDRLRAFLERLEFGGLLHEFGLLESPKALEEAPWPPPEGAFFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERAAALSERLFANLWGRLEGEERLLWLYREVERPLSVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
15 AEVFRLAGHPFNLNSRDQLERVLFDLGLPAIGKTEKTGKRSTGAAVLEALREAHPTVEKILQYRELTKLKSTYIDPLPDL
IHPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRIRRAFIAEEGWLLVLDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHLRSQELAIPYEEAQAFIERIYQSFQKVRWIEKTLEEG
RRRGYVETLFGRRRYVPDLNARVKSAREMAFNPVQGTAAADLMKMLAMVKLFPRLREMGARMLLQVHDELVEA
PKERAEEVARLAKEAMEGVYPLAVPLEVEVGIGEDWLSAKE*

20

1C8:

25 ATGCGTGGTATGCTTCCTCTTTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCAC
CTTCTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCAGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTC
CTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCGCTTCGTGGTCTTTGACGCCAAGGGCCCCCTCCTCCGCCACG
AGGCCTACGAGGCCTACAAGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGG
AGCTGGTGGACCTCCTGGGGTTTACCCGCTCGAGGTCCCTGGCTACGAGGCGGACGACGTCTCGCCACCCTGGC
CAAGAAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTACCGCCGACCGCGACCTCTACCAACTCGTCTCCGA
30 CCGCGTCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCG
GAGCAGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGTCAAGGGCATCGGGGAG
AAGACCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCA
GAAAACGTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCACCG
ACCTCCCCCTGGAGGTGGACCTCGCCAGGGGCGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCT
TGAGTTTGGCGGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGCCCCCTGGCCCCCG
35 CCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGGCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCG
CAGGGGTGGTCCGGGTCCACCGGGCCCCGAGCCTTATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTC
GCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCGCGACGACCCCATGCTCCTCGCCT
ACCTCCTGGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGAGGAGGCGG
GGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTG
40 GCTTTACCGGGAGGTGGATAGGCCCTTTCCGCTGTCTGGCCACATGGAGGCCACAGGGGTACGGCTGGACGTG
GCCTGCCTGCAGGCCCTTTCCCTGGAGCTTGCGGAGGAGATCCGCCGCTCGAGGAGGAGGTCTTCCGCTTGGCGG
GCCACCCCTTCAACCTCAACTCCCGGGACAGCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGG
CAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGCCCTCCGCGAGGCCACCCCATCGT
GGAGAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCAC
45 CCCAGGACGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCA
ACCTCCAGAACATCCCCGTCCGCACCCGCTCGGGCAGAGGATCCGCCGGGCTTCATCGCCGAGGAGGGGTGGCT
ATTGGTGGTCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGACGAGAACCTGACCCGG
GTCTTCCAGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCCGGCGTCCCCCGGGAGGCCGTGGAC
CCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTGCGCCACCGCCTCTCCCAGG
50 AGCTGGCCATCCCTTACGAGGAGGCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTG
GATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGTACGTGGAAACCTCTTCGGAAGAAGGCGCTACGTGCC
CGACCTCAACGCCCGGGTGAAGAGTGTGAGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCCGTCCAGGGCAC
CGCCGCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
AGGTCCACGACGAGCTCCTCCTGGAGGCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCA
55 TGGAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGG
AGTGA

60 MRGMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVPVQAVYGFASLLKALKEDGYKAAFVFDKAPSFRHEA
YEAYKAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRLTADRDLYQLVSDRVAVL
HPEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGEKTAALKLLKEWGSLENLLKNLDRVKPENVREKJKA
AHLEDLRLSLELSRVRTDLPLEVDLAQGREPDRLRAFLERLEFGGLLHEFGLLESPKALEEAPWPPPEGAFFVGFVLSRKE
EPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVAR
RYGGEWTEEAGERAAALSERLFANLWGRLEGEERLLWLYREVDRLSAVLAHMEATGVRLDVACLQALSLELAEEIRRL
65 EEEVFRLAGHPFNLNSRDQLERVLFDLGLPAIGKTEKTGKRSTSAILEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRIRRAFIAEEGWLLVLDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHLRSQELAIPYEEAQAFIERIYQSFQKVRWIEKTLEEG
RKRGYVETLFGRRRYVPDLNARVKSAREMAFNPVQGTAAADLMKMLAMVKLFPRLREMGARMLLQVHDELLLE
APQARAEEVAALAKEAMEKAYPLAVPLEVKVGIGEDWLSAKE*

70

2H10X:

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGTGTCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCAGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
5 AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGGCCCCCATTCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCG
10 TCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCA
GTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACC
GCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAAC
GTCCGGGAGAAGATCAAGGCCACCTGGAAGATCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCACACCGACCTCC
CCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTT
15 TGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGGCCCCCTGGCCCCCGCCGAA
GGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGAGGGG
TGGTCGAGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAA
GACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCTACCTCCT
GGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCG
20 GGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTAC
CGGGAGGTGGAGAGGCCCTTTCCGCTGTCTGGCCACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATC
TCAGGGCCTTGTCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGCCACCC
CTTCAACCTCAACTCCCGGGACAGCTGGAAATGGTGTCTTTGACGAGCTTAGGCTTCCCGCCTTGGGGAAGACG
25 CAAAAGACGGGCAAGCGCTCCACCAGCGCCCGCTCCTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGAGAAG
ATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGACCCCTTGTTCGGACCTCATCCACCCAGGA
CGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCA
GAACATCCCCGTCCGACCCCGCTTGGGCAGAGGATCCGCCGGGCTTCATCGCCGAGGAGGGGTGGCTACTGGTG
GTCCTGGACTATAGCCAGATAGAGCTCAGGGTGTGGCCACCTCTCCGGCGACGAAAACCTGATCAGGGTCTTCC
AGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTTCGGCGTCCCCCGGGAGGCCGTGGACCCCTGA
30 TCGCCCGGGCGGCAAGACCATCAACTTCGGGGTCTCTACGGCATGTTCGGCCACCGCCTCTCCAGGAGCTAGC
CATCCCTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACTTTTCAGAGCTTCCCCAAGGTGCGGGCCTGGATTGAG
AAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCTCTTCGGCCGCCGCGCTACGTGCCAGACCTA
GAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCC
GACCTCATGAAGCTGGCTATGGTGAAGCTTTCGCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTCC
35 ACGACGAGCTGGTCTCGAGGCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGG
GGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVQVVYGFASLLKALKEDGYKAVFVFDKAPFRHKAY
EAYRAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRLTADRGLYQLVSDRVAVLH
40 PEGHLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGEKALKLLKEWGSLENLLKNLDRVKPENVRKIK
HLEDRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARR
YGGEWTEEAGERALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLSRDQLEMVLFDELRLPALGKTQKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLSDL
45 IHPRTGRLHTRFNQTATATGRLSSSDPNLQNPVRTPLGQRIRRAFIAEEGWLLVLDYSQIELRVLAHLSGDENLIRVFQE
GRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHLRSQELAIPYEEAQAFIERYFQSPFKVRAWIEKTLEEG
RRRGYVETLFGRRRYVPDLEARVKSVRERMAFNMVQGTAAADLMKLAAMVKLFPRLEEMGARMMLQVHDELVL
APKERAEAVARLAKEVMGCVYPLAVPLEVEVGIGEDWLSAKE*

3A10
ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGCCGGGTCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTT
CTTCGCCCTGAAGGGCCTACCACGAGCCGGGGCGAACCAGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTC
5 AAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGGCCCCCATTCGCCACAAGG
CCTACGAGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCT
GGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAG
55 AAGGCGGAAAAGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCCGACCGC
GTCGCCGTCTCCACCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGC
AGTGGGTGGACTTCCGCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGA
CCGCCCTCAAGCTCCTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAA
ACGTCCGGGAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCACACCGACCT
60 CCCCCTGGAGGTGGACCTCGCCCAGGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGA
GTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGGCCCCCTGGCCCCCGCG
GAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCGCGCAG
GGGTGGTCGAGTCCACAGGCCCCCGAGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCC
AAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACC
65 TCCTGGACCTTCCAACACCACCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGG
AGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCT
TTACCGGGAGGTGGAGAGGCCCTTTCCGCTGTCTGGCCACATGGAGACCACGGGGGTGCGCCTGGACGTGGCC
TATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGCC
70 GCCCTTCAACCTCAACTCCCGAGACAGCTGGAAAGGGTCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAA
GACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCCGCTCCTGGAGGCCCTCCGCGAGGCCACCCCATCGTGGA

GAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCC
AGGACGGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACC
TCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGGCTTCATCGCCGAGGAGGGGTGGCTATT
GGTGGTCTGGACTATAGCCAGATGGAGCTCAGGGTGCTGGCCACCTCTCCGGCGACGAGAACCTGATCAGGGTC
5 TTCCAGGAGGGGAAGGACATCCACACCCAGACCGCAAGCTGGATGTTCCGGTGTCCCCCGGAGGCCGTGGACCCC
TGATGCGCCGGGCGGCCAAGACGGTGAACCTCGGCGTCTCTACGGCATGTCCGCCCATAGGCTCTCCAGGAGCT
TTCCATCCCCTACGAGGAGGCGGTGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGGATT
GAGAAGACCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCTTCGGCCGCGCCGCTACGTGCCCGAC
CTCAACGCCCGGATGAAGAGCGTCAGGGGGGCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCC
10 GCCGACCTCATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCTCCGGGAGATGGGGGCGCCGATGCTCCTCCAGG
TCCACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATGG
AGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGGAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGT
GA

15 MAMLPLEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQVVYGFASLLKALKEDGYKAVFVVFDAKAPSFRHKAY
EAYRAGRPTPQDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVSDRVAVLH
PEGLITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGIEKTALKLLKEWGSLENLLKNLDRVKPENVREKIK
HLEDRLSLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKE
PMWADLLALAAARGGRVHQAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVAR
20 RYGGEWTEEAGERALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMETTGVRLDVAYLRALSLEVAEEIARL
EAEVFRLAGRPFNLSRDQLERVLFDLGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDL
IHPRTGRLHTRFNQTATATGRLSSSDPNLQNPVTRPLGQRIIRAFIAEEGWLLVLDYSQMELRVLAHLSGDNELIRVFQ
EGKDIHTQTASWMFGVPPEAVDPLMRRAAKTVNFGVLYGMSAHLRSQELSIPIYEEAVAFIERYFQSFVKVRAWIEKTL
EGRRRGYVETLFGRRRYVPDLNARMKSVRGAAERMAFNMPVQGTAAADLMKLAMVKLFPRLREMGARMMLQVHDELL
25 LEAPQARAEVAALAKEAMEKAYPLAVPLEVEVGIGEDWLSAKE*

3B5
ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGTGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTCCT
TCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
30 CCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCCCATTCGCCACAAGGCCTACG
AGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCGTCAAGGAGCTGGTGGACC
TCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAGAAGGCGGAAA
AGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCGTCGCCGTCTCCA
CCCCGAGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCCG
35 CGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCAA
GGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAGG
CCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCCCCCTGGAGGTGGACCTCGCCCA
GGGGCGGGAGCCCGACCGGAAAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCATGAGTTTCGG
CCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGGCCCCCTGGCCCCCGCCGGAAGGGGGCTTTCGTGGGCTTTGTGCTTTCC
40 CGCAAGGCGCCCATGTGGGGCCGATCTTCTGGCCCTGGCCGCCCGCCAGGGGTGGTGGGTCTACCGGGCCCCCGAGCCT
TATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGC
CTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCTTCCAACACCACCCCGAGGGGGTGG
CCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGT
GGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGATAGGCCCTTTCCGCTGTCTGGCCCC
45 ACATGGAGGCCACAGGGGTACGGCTGGACGTGGCCTGCCTGCAGGCCCTTTCCCTGGAGCTTGCGGAGGAGATCCGCC
GCCTCGAGGAGGAGGTCTTCCGCTTGGCGGGCCACACCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCTCTT
TGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGC
CCTCCGCGAGGCCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGA
CCCCCTTGCCGGACCTCATCCACCCAGGACGGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
50 GCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCCTTCATC
GCCGAGGAGGGGTGGCTACTGGTGGTCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCTCACCTCTCCGGCGAC
GAAAACCTGATCAGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCCGGCGTCCCCCGG
GAGGCCGTGGACCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTGCGGCCACCGC
CTCTCCCAGGAGCTAGCCATCCCTTACGAGGAGGGCCAGGCCTTCAATTGAGCGCTACTTTAGAGCTTCCCCAAGGTGC
55 GGGCCTGGATTGAGAAGGCCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCTCTTCGGAAGAAGGCGCTAC
GTGCCCGACCTCAACGCCCGGGTGAAGAGTGTACGGAGGCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGC
ACCGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCTCCGGGAGATGGGGGCGCGCATGCTCCTCC
AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATG
GAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGTGA
60 MAMLPLEPKGRVLLVDGHHLAYRTSFALKGPTTSRGEVPQVVYGFASLLKALKEDGYKAVFVVFDAKAPPFRHKAYEA
YRAGRPTPEDFPRQLALVKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRILTADRGLYQLVSDRVAVLHPEGH
LITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGIEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKAHLEDRL
SLELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKAPMWADLLAL
AAARGGRVYRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGE
65 RAALSERLFANLWGRLEGEERLLWLYREVDRLSAVLAHMEATGVRLDVACLQALSLELABEIRLEEEVFRLAGHTFNLN
SRDQLERVLFDLGLPAIGKTEKTGKRSTSAILEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATA
TGRLLSSSDPNLQNPVTRPLGQRIIRAFIAEEGWLLVLDYSQIELRVLAHLSGDNELIRVFQEGRDIHTETASWMFGVPREA
VDPLMRRAAKTINFGVLYGMSAHLRSQELAIPIYEEAQAFIERYFQSFVKVRAWIEKALEEGRRRGYVETLFGRRRYVPDLNA
RVKSVREAAERMAFNMPVQGTAAADLMKLAMVKLFPRLREMGARMMLQVHDELLLEAPQARAEVAALAKEAMEKAYPL
70 AVPLEVKVGIGEDWLSAKE*

3B6

5 ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCTCCTGGTGGACGGCCACCACCTGGCCTACCGCGCCTTCT
TCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCAGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
CCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACGAGGCCTACG
10 AGGCCTACAAGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACC
TCCTGGGGTTTACCCGCCTCGAGGTCCAAGGCTACGAGGCGGACGACGTCTCGCCACCCTGGCCAAGAAGGCGGAAA
AAGAAGGGTACGAGGTGCGCATCCTCACCGCCGACCGGGACCTCTACCAGCTCGTCTCCGACCGCGTCCGCGTCTCTCC
ACCCCGAGGGCCACCTCATCACCCCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCC
15 GCGCCCTCGTGGGGGACCCCTCCAACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCA
AGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAG
GCCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCACACCGACCTCCCCCTGGAGGTGGACCTCGCCC
AGGGGCGGGAGCTCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCGGCCTCCTCCACGAGTTTCG
GCCTTCTGGAAGGCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTTC
20 CCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCGGCCAGGGGTGGTGGGTCCACCGGGCCCCCGAGCC
TTATAAAGCCCTCAGGGAAGTGAAGGAGGCGCGGGGGCTTCTCGCCAAGACCTGAGCGTTCTGGCCCTAAGGGAAGG
CCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCGCCCCCGAGGGGGTG
GCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCT
GTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGATAGGCCCTTTCCGCTGTCTGGC
25 CCACATGGAGGCCACAGGGGTACGGCTGGACGTGGCCTATCTCAGGGCCTTGTCCTGGAGGTGGCCGAGGAGATCGC
GCGCCTCGAGGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGAGACCAGCTGGAAAGGGTCCTC
TTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCCGCGTCTGGAG
GCCCTCCGCGAGGCCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATT
GACCCCTTGCCGAACCTCATCCATCCAGGACGGGCCGCCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGC
30 AGGCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTC
ATCGCCGAGGAGGGGTGGCTATTGGTGGTCTTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGC
GACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCCGGCTCCCC
CGGGAGGCCGTGGACCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCCGGCCAC
CGCCTCTCCAGGAGCTAGCCATCCCTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGG
35 TGCGGGCCTGGATAGAAAAGACCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCTCTTCGGAAGAAGGCGC
TACGTGCCCGACCTCAACGCCCGGGTGAAGGGCGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAG
GGCACCGCCCGGACCTCATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGGCCCGCATGCTC
CTCCAGGTCCACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGGGAGGTGGCGGCTTTGGCCAAGGAGGCC
ATGGAGAAGGCCTATCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAG
40 TGA
MAMLPLFEPKGRVLLVDGHHLAYRAFFALKGLTTSRGEVPQAVYGFASLLKALKEDGYKAVFVFDKAPSFRHEAYEA
YKAGRPTPEDFPRQLALIKELVDLLGFTRLEVQGYEADDVLATLAKKAEKEGYEVRLTADRDLVQLVSDRVAVLHPEGH
LITPEWLWEKYGLRPEQWVDFRALVGDPSNNLPVKGIGEKALKLLKEWGSLENLLKNLDRVKPENREKIKAHLEDLRL
45 SLELSRVRTDLPLEVDLAQGRELDRELRALFLERLEFGGLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLA
LAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLPLPGDDPMLLAYLLDPSNTAPEGVARRYGGEWTEEAG
ERAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVAYLRLSLEVAEEIARLEAEVFRLAGHPFNL
NSRDQLERVLFDLGLPAIGKTEKTGKRSTSAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPNLIHPRTGRLHTRFNQTA
TATGRLSSSDPNLQNPVTRPLGQRIRRAFIIEGWLLVVDYSQIELRVLAHLSGDENLIRVFEGRDIHTETASWMFGVPRE
45 AVDPLMRRAAKTINFGVLYGMSAHLRSQELAIPIYEEAQAFIERFYQSFQKPVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLN
ARVKGVREAAERMAFNMPVQGTAAADLMKLAMVKLFPRLREMGARMMLLQVHDELLLEAPQARAGEVAALAKEAMEKAY
PLAVPLEVKVGIGEDWLSAKE*

3B8

50 ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGGGTCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCT
TCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCAGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
CCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACGAGGCCTACG
AGGCCTACAAGGCGGGGAGGGCCCCGACCCCGAGGACTTCTTCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACC
55 TCCTGGGGTTTACCCGCCTCGAGGTCCAAGGCTACGAGGCGGACGACGTCTCGCCACCCTGGCCAAGAAGGCGGAAA
AAGAAGGGTACGAGGTGCGCATCCTCACCGCCGACCGGGACCTCTACCAGCTCGTCTCCGACCGCGTCCGCGTCTCTCC
ACCCCGAGGGCCACCTCATCACCCCGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCC
GCGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCA
AGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGTGAAGCCCGCCATCCGGGAGAAGATCCTGGCC
CACATGGACGATCTGAAGCTCTCTGGGACCTGGCCAAGGTGCGCACCGACCTGCCCCTAGAGGTGGACTTCGCCAAA
60 AGGCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGCTTGGCAGCCTCCTCCACGAGTTTCGGC
CTTCTGGAAGGCCCAAGACCCTGGAGGAGGCCTCCTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCC
GCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCGGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTT
ATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCC
TTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCGAGGGGGTGGC
65 CCGGCGCTACGGCGGGGAGTGGACGAAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGT
GGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGATAGGCCCTTTCCGCTGTCTGGCCC
ACATGGAGGCCACAGGGGTGCGCTTGGACGTGGCCTATCTCAGGGCCTTGTCCTGGAGGTGGCCGAGGAGATCGCCC
GCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCATCCCTTCAACCTCAACTCCCGGGACCAAGCTGGAAAGGGTCTCTT
TGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCGCTCTGGAGGC
CCTCCGCGAGGCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGA
70 CCCCTTGCCGGACCTCATCACCCAGGACGGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
GCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCGTC

5 GCGGAGGAGGGGTGGCTATTGGTGGTCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGAC
 GAGAACCTGACCCGGGTCTTCTGGAGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCCGGCGTCCCCCGG
 GAGGCCGTGGACCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCACCGC
 CTCTCCCAGGAGCTGGCCATCCCTTACGAGGAGGGCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGC
 10 GGGCCTGGATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTAC
 GTGCCCCGACCTCAACGCCCCGGGTGAAGAGTGTGAGGGAGGCGCGGAGCGCATGGCCTTCAACATGCCCCGTCCAGGGC
 ACCGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
 AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATG
 GAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGAGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGTG
 15 A
 MAMLPFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQVVYGFASLLKALKEDGYKAVFVFDKAPSLRHEAYEA
 YKAGRAPTPEDFLRQLALIKELVDLLGFTRLEVQGYEADDVLATLAKKAEKEGYEVRLTADRDLYQLVSDRVAVLHPEGH
 LITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGIEKTALKLLKEWGSLENLLKNLDRKPAIREKILAHMDDLKLS
 20 WDLAKVRTDLPLEVDFAKRREPDRERLRAFLELRLGSLLEHFGLLLESPKLEEASWPPPEGAFVGFVLSRKEPMWADLLAL
 AAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTKAGE
 RAALSERLFANLWGRLEGEERLLWLYREVDRLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLN
 SRDQLERVLFDLGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTAT
 ATGRLSSSDPNLQNIPTPLGQRIRRAFVAEEGWLLVLDYSQIELRVLAHLSGDENLTVFLEGRDIHTETASWMFGVPRE
 25 AVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPEEAQAFIERFYQSFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLN
 ARVKSUREAAERMAFNMPVQGTAAADLMKLAMVKLFPRLREMGARMLLQVHDELLLEAPQARAEVAALAKEAMEKAYP
 LAVPLEVKEGIGEDWLSAKE*
 3B10
 30 ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCT
 TCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
 CCCTGAAAGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCTCATTCCGCCACAAGGCCTACG
 AGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACC
 TCCTGGGGTTTACCCGCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAGAAGGCGGAAA
 35 AGGAGGGGTACGAGGTGCGCATCCTACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCGTCGCCGTCTCCA
 CCCCAGGGGCCACCTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGAGCAGTGGGTGGACTTCCG
 CGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCAA
 GGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAGG
 CCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGGCACCAGCTCCCCCTGGAGGTGGACCTCGCCCA
 40 GGGGCGGGAGCCCCAGCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCGGCCTCCTCCACGAGTTCGG
 CCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCTTGGCCCCCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCC
 CGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCGCCAGGGGTGGTCCGGGTCCACCGGGGCCCTGAGCCT
 TATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAAGGC
 CTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCTTCCAACACCACCCCGAGGGGGTGG
 45 CCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGT
 GGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGACCCCTTTCCGCTGTCCTGGCCC
 ACATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCTTGGAGGTGGCCGAGGAGATCGCCC
 GCCTCGAGGCCGAGGTCTTCCGCTGGCCGGCCACCCCTTCAACCTCAACTCCCGAGACCAGCTGGAAAGGGTCTCTT
 TGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCGCTCCTGGAGGC
 50 CCTCCGCGAGGCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTACCAAGCTGAAGAGCACCTACATTGA
 CCCCTTGCCGGACCACATCCACCCAGGACGGGGCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
 GCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTCATC
 GCCGAGGAGGGGTGGCTATTGGTGGTCTTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGAC
 GAGAACCTGACCCGGGTCTTCCAGGAGGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCCGGCGTCCCCCGG
 55 GAGGCCGTGGACCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTTCTCTACGGCATGTCGGCCACCGC
 CTCTCCCAGGAGCTGGCCATCCCTTACGAGGAGGGCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGC
 GGGCCTGGATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCCTCTTCGGAAGAAGGCGCTAC
 GTGCCCCGACCTCAACGCCCCGGGTGAAGAGTGTGAGGGAGGCGCGGAGCGCATGGCCTTCAACATGCCCCGTCCAGGGC
 ACCGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTACCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
 60 AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGAGGAGGTGGCGGCTTTGGCCAAGGAGGCCATG
 GAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGTG
 AGTCGACCTGCAGGCAGCGCTTGGCGTCACCCCGCAGTTCGGTGGTTAA
 MAMLPFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVPQVVYGFASLLKALKEDGYKAVFVFDKAPSFHRHKAYEA
 65 YRAGRAPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRLTADRGYQLVSDRVAVLHPEGHL
 ITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKIGIEKTALKLLKEWGSLENLLKNLDRVKPENVREKIKAHLEDLRLS
 LELSRVRTDLPLEVDLAQGREPDRELRRAFLELRLFGGLLEHFGLLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLAL
 AAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGE
 RAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLN
 70 SRDQLERVLFDLGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDHIHPRTGRLHTRFNQTAT
 ATGRLSSSDPNLQNIPTPLGQRIRRAFIAEEGWLLVLDYSQIELRVLAHLSGDENLTVFQEGRDIHTETASWMFGVPRE
 AVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPEEAQAFIERFYQSFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLN
 ARVKSUREAAERMAFNMPVQGTAAADLMKLAMVKLYPRLREMGARMLLQVHDELLLEAPQARAEVAALAKEAMEKAYP
 LAVPLEVKVGIGEDWLSAQGVSRPAGSAWRHPQFGG*
 3C12

ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCT
 TCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCAGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
 CCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGGCCTACG
 AGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACC
 5 TCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAGAAGGCGGAAA
 AGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCGTCGCCGTCTCCA
 CCCCAGGGGCCACCTCATCACC CGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTAGACTTCCG
 CGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCAA
 GGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAGG
 10 CCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCCCCCTGGAGGTGGACCTCGCCCA
 GGGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGTTTCGG
 CCTTCTGGAAAAGCCCCAAGGCCCTGGAGGAGGGCCCCCTGGCCCCCGCCGGAAGGGGCTTCGTGGGCTTTGTGCTTTCA
 CGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCCGCAGGGGTGGTCGGGTCCACCGGGCCCCCGAGCCT
 15 TATAAAGCCCTCAGGGACTTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGC
 CTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCTACCTCCTGGACCTTCCAACACCGCCCCCGAGGGGGTGG
 CCGGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGT
 GGGGGAGGCTTGAGGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGATAGGCCCCCTTTCCGCTGTCTGGCCC
 ACATGGAGGCCACAGGGGTACGGCTGGACGTGGCCTGCCTGCAGGCCCTTTCCCTGGAGCTTGCGGAGGAGATCCGCC
 GCCTCGAGGAGGAGGTCTTCCGCTTGCGGGGCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAAGGGTCTCTT
 20 TGACGAGCTAGGGCTTCCCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCATCCTGGAGGC
 CCTCCGCGAGGCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGA
 CCCCTTGCCGACCTCATCCACCCAGGACGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
 GCTAAGTAGCTCCGGTCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCCTTCGT
 25 GCCGAGGAGGGGTGGCTATTGGTGGTCTTGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGAC
 GAGAACCTGACCCGGGTCTTCTTGAGGGGGCGGGACATCCACACGGAAACCGCCAGCTGGATGTTCCGGCGTCCCCCGG
 GAGGCCGTGGACCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTCCGGCCACCGC
 CTCTCCCAGGAGCTGGCCATCCCTTACGAGGAGGCCAGGCCTTCATAGAGCGCTACTTCCAAAGCTTCCCCAAGGTGC
 GGGCCTGGATAGAAAAGACCCTGGAGGAGGGGAGGAAGCGGGGCTACGTGGAAACCTCTTCGGAAGAAGGCGCTAC
 GTGCCCGACCTCAACGCCCCGGGTGAAGAGTGTGAGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGC
 30 ACCGCCGCCGACCTTATGAAGCTCGCCATGGTGAAGCTCTTCCCCCGCCTCCGGGAGATGGGGGCCCGCATGCTCCTCC
 AGGTCCACGACGAGCTCCTCCTGGAGGCCCCCAAGCGCGGGCCGAGGAAGTGGCGGCTTTGGCCAAGGAGGCCATG
 GAGAAGGCCTATCCCCTCGCCGTACCCCTGGAGGTGAAGGTGGGGATCGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

 MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVPVQVYGFAKSLLKALKEDGYKAVFVVFDAKAPSRHKAYEA
 35 YRAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRLTADRGLYQLVSDRVAVLHPEGHL
 ITPEWLWEKYGLRPEQWVDFRALVGDPSNLPVKGIGEKALKLLKEWGSLENLLKNLDRVKPENREKIKAHLEDLRLS
 LELSRVRTDLPLEVDLAQGREPDREGLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLAL
 AAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTAPEGVARRYGGEWTEEAGE
 RAALSERLFANLWGRLEGEERLLWLYREVDRPLSAVLAHMEATGVRLDVACLQALSLELAEEIRLEEEVFRLAGHPFNLN
 40 SRDQLERVLDFDELGLPAIGKTEKTKRSTSAAILALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATA
 TGRLLSSSGPNLQNPVRTPLGQRIRRAFAEEGWLLVLDYSQIELRVLALHLSGDNELTRVFLEGRDIHTETASWMFGVPREA
 VDPLMRRAAKTINFGVLYGMSAHRLSQELAIPIYEEAQAIFIERYFSFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDLNA
 RVKSVREAAERMAFNMPVQGTADLMKLAMVKLFPRLREMGARMLLQVHDELLLEAPQARAEEVAALAKEAMEKAYPL
 AVPLEVKVGIGEDWLSAKE*
 45 3D1
 ATGGCGATGCTTCCCCTCTTTGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTTCT
 TCGCCCTGAAGGGCCCCACCACGAGCCGGGGCGAACCAGGTGCAGGTGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
 CCCTGAAGGAGGACGGGTACAAGGCCGTCTTCGTGGTCTTTGACGCCAAGGCCCCCTCATTCCGCCACAAGGCCTACG
 50 AGGCCTACAGGGCGGGGAGGGCCCCGACCCCGAGGACTTCCCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACC
 TCCTGGGGTTTACCCGCCTCGAGGTCCCCGGCTACGAGGCGGACGACGTTCTCGCCACCCTGGCCAAGAAGGCGGAAA
 AGGAGGGGTACGAGGTGCGCATCCTCACCGCCGACCGCGGCCTCTACCAACTCGTCTCTGACCGCGTCGCCGTCTCCA
 CCCCAGGGGCCACCTCATCACC CGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCCG
 CGCCCTCGTGGGGGACCCCTCCGACAACCTCCCCGGGGTCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTCCTCAA
 55 GGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAACGTCCGGGAGAAGATCAAGG
 CCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTCCCCCTGGAGGTGGACCTCGCCCA
 GAGGCGGGAGCCCGACCGGGAGGGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCTTCCACGAGTTTCGG
 CCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGGCCCCCTGGCCCCCGCCGGAAGGGGCTTCGTGGGCTTTGTGCTTTCC
 CGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGGCCGCCCGCAGGGGTGGTCGAGTCCACCGGGCCCCCGAGCCT
 60 TATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGC
 CTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCTACCTCCTGGACCTTCCAACACACCCCGAGGGGGTGG
 CCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGT
 GGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGCCCTTTCCGCTGTCTGGCCC
 ACATGGAGGCCACGGGGGTGCGCTGGACGTGGCCTATCTCAGGGCCTTGTCCTGGAGGTGGCCGAGGAGATCGCCC
 65 GCCTCGAGGCCGAGGTCTTCCGCTTGCCCGGCCACCCCTTCAACCTCAACTCCCGGGACCAGCTGGAAATGGTGTCTT
 TGACGAGCTTAGGCTTCCCGCCTTGGGGAAGACGCAAAAGACGGGCAAGCGCTCCACCAGCGCCGCCGTCTGGAGGC
 CCTCCGCGAGGCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACCTACATTGA
 CCCCTTGTCGGACCTCATCCACCCAGGACGGGCCGCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
 GCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCTTCATC
 70 GCCGAGGAGGGGTGGCTACTGGTGGTCTTGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCGGCGAC
 GAAAACCTGATCAGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCCGGCGTCCCCCGG

5 GAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGGGTCTCTACGGCATGTGCGGCCACCGC
 CTCTCCAGGAGCTAGCCATCCCTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGC
 GGGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGTACGTGGAGACCCTCTTCGGCCGCCGCGCTAC
 GTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCCGTCCAGGG
 CACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGGAGAAACGGGGGCCAGGATGCTCCT
 TCAGGTCCACGACGAGCTGGTCCTCGAGGCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGCCA
 TGGAGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCAAGGGTT
 AG
 10 MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGPTTSRGEVQVVYGFAKSLKALKEDGYKAVFVFDKAPSFRHKAYEA
 YRAGRPTPEDFPRQLALIKELVDLLGFTRLEVPGYEADDVLATLAKKAEKEGYEVRLTADRGLYQLVSDRVAVLHPEGHL
 ITPEWLWEKYGLRPEQWVDFRALVGDPSDNLPGVKGIGECTALKLLKEWGSLENLLKNLDRVKPENVREKIKAHLEDLRLS
 LELSRVRTDLPLEVDLAQRREPDREGLRAFLEFGLFHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALA
 AARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGER
 15 AALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLS
 RDQLEMVLFDELRLPALGKTQKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLSDLIHPRTGRLHTRFNQTAT
 ATGRLSSSDPNLQNIPVRTPLGQRIRRAFAEEGWLLVVDYSQIELRVLAHLSGDNELIRVFQEGRDIHTETASWMFGVPRE
 AVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFQKVRWIEKTLEEGRRRGYVETLFGRRRYVPDLE
 ARVKSREAAERMAFNMPVQGTAAADLMKLMVKLFPRLGETGARMMLQVHDELVEAPKERAEEAVARLAKEAMEGVYP
 20 LAVPLEVEVGIGEDWLSAKG*

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

25

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.

30

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

35 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.

40 **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).

- 5 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 [³²P]ATP (5000
- 10 Ci/mmol; 1 Ci=37 GBq) (Pharmacia) using T4 polynucleotide kinase (Invitrogen, Carlsbad CA). Radiolabeled primer-template DNAs were prepared by annealing the 5' [³²P] 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)
- 15 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 MgCl₂, 100µM of each ultrapure dNTP (Amersham Pharmacia Biotech, NJ), 10mM DTT, 250 µg/ml BSA, 2.5% glycerol, 10 nM 5' [32P]
- 20 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
5 both wild type and M1.

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

10 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
15 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).

20

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
25 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.

5 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.

10 **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
15 product.

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

20 M1 has a slightly reduced k_{cat}/K_m , 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,
5 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
10 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
15 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.

20

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
5 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.

10 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
15 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
20 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 μ l of ancient sample were added to a 20 μ l PCR in SuperTaq buffer (HT
5 Biotech) with 1 μ M of the appropriate primers (see Figure 13), 2 μ M of each deoxyribonucleoside triphosphate (dNTP) as well as 0.1 μ l 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
10 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
15 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
20 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
5 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.

10 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.

15 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
20 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
5 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
10 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
15 (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
20 5D4:

4D11: 5' -
ATGGCGATGCTTCCCCTCTTTGAGCCCAAAGGCCGGGTCTCCTGGTGGACGGCCACCACCTGGCCTAC
CGCACCTTCTTCGCCCTGAAGGGCCTCACCACGAGCCGGGGCGAACCAGGTGCAGGCGGTTTACGGCTTC
25 GCCAAGAGCCTCCTCAAGGCCCTGAAGGAGGACGGGTACAAGGCCGTCTTTCGTGGTCTTTGACGCCAAG
GCCCCCTCCTTCCGCCACGAGGCCTACGAGGCCTACAAGGCCGGGGAGGGCCCCGACCCCCGAGGACTTC
CCCCGGCAGCTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGTTTACCCGCCTCGAGGTCCAAGGC
TACGAGGCGGACGACGTCCTCGCCACCCTGGCCAAGAAGGCCGGAAGAGGGTACGAGGTGCGCATC
CTCACCGCCGACCGGGACCTCTACCAGCTCGTCTCCGACCGCGTCGCCGTCTCCACCCCGAGGGCCAC
30 CTCATCACCCCGGAGTGGCTTTGGGAGAAGTACGGCCTCAGGCCGGAGCAGTGGGTGGACTTCCGCGCC
CTCGTGGGGGACCCCTCCGACAACCTCCCCGGGATCAAGGGCATCGGGGAGAAGACCGCCCTCAAGCTC
CTCAAGGAGTGGGGAAGCCTGGAAAACCTCCTCAAGAACCTGGACCGGGTAAAGCCAGAAAATGTCCGG
GAGAAGATCAAGGCCACCTGGAAGACCTCAGGCTCTCCTTGGAGCTCTCCCGGGTGCGCACCGACCTC
CCCCTGGAGGTGGACTTCGCCAAAAGGCCGGGAGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGG
35 CTTGAGTTTGGCAGCCTCCTCCACGAGTTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCC
TGGCCCCCGCCGGAAGGGGCTTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTT
CTGGCCCTGGCCGCCGCCAAGGGTGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGGGAC
TTGAAGGAGGCGCGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTAAGGGAAGGCCTTGGCCTC
CCGCCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCGAGGGGGTG

G C C C G G C G C T A C G G C G G G A G T G G A C G G A G G A G G C G G G G A G C G G G C C G C C T T T C C G A G A G G C T C T T C
 G C C A A C C T G T G G G G A G G C T T G A G G G G A G G A G A G G C T C C T T T G G C T T T A C C G G G A G G T G G A G A G G C C C
 C T T T C C G C T G T C C T G G C C C A C A T G G A G G C C A C G G G G T G C G C T G G A C G T G G C C T A T C T C A G G G C C T T G
 T C C C T G G A G G T G G C C G A G G A G A T C G C C C G C C T C G A G G C C G A G G T C T T C C G C C T G G C C G G C C A C C C C T T C
 5 A A C C T C A A C T C C C G A G A C C A G C T G G A A A G G G T C C T C T T T G A C G A G C T A G G G C T T C C C G C C A T C G G C A A G
 A C G G A G A A G A C C G G C A A G C G C T C C A C C A G C G C C G C C G T C C T G G A G G C C C T C C G C G A G G C C C A C C C C A T C
 G T G G A G A A G A T C C T G C A G T A C C G G G A G C T C A C C A A G C T G A A G A G C A C C T A C A T T G A C C C C T T G C C G G A C
 C T C A T C C A C C C C A G G A C G G G C C G C C T C C A C A C C C G C T T C A A C C A G A C G G C C A C G G C C A C G G G C A G G C T A
 A G T A G C T C C G A T C C C A A C C T C C A G A A C A T C C C C G T C C G C A C C C C G C T C G G G C A G A G G A T C C G C C G G G C C
 10 T T C A T C G C C G A G G G G G G T G G C T A T T G G T G G T C C T G G A C T A T A G C C A G A T G G A G C T C A G G G T G C T G G C C
 C A C C T C T C C G G C G A C G A G A A C C T G A T C C G G G T C T T C C A G G A G G G C G G G A C A T C C A C A C G G A A A C C G C C
 A G C T G G A T G T T C G G C G T C C C C C G G A G G C C G T G G A C C C C T G A T G C G C C G G G C G G C C A A G A C C A T C A A C
 T T C G G G G T T C T C T A C G G C A T G T C G G C C C A C C G C C T C T C C C A G G A G C T A G C C A T C C C T T A C G A G G A G G C C
 C A G G C C T T C A T T G A G C G C T A C T T T C A G A G C T T C C C C A A G G T G C G G G C C T G G A T T G A G A A G A C C C T G G A G
 15 G A G G G C A G A G G C G G G G T A C G T G G A G A C C C T C T T C G G C C G C C G C C G C T A C G T G C C A G A C C T A G A G G C C
 C G G G T G A A G A G C G T G C G G G A G G C G G C C G A G C G C A T G G C C T T C A A C A T G C C C G T C C A G G G C A C C G C C G C C
 G A C C T C A T G A A G C T G G C T A T G G T G A A G C T C T T C C C C A G G C T G G A G G A A A C G G G G G C C A G G A T G C T C C T T
 C A G G T C C A C G A C G A G C T G G T C C T C G A G G C C C C A A A G A G A G G G C G G A G G C C G T G G C C C G G C T G G C C A A G
 G A G G T C A T G G A G G G G T G T A T C C C C T G G C C G T G C C C C T G G A G G T G G A G G T G G G G A T A G G G G A G G A C T G G
 20 C T C T C C G C C A A G G A G T G A - 3 '

4D11 amino acid sequence:

M A M L P L F E P K G R V L L V D G H H L A Y R T F F A L K G L T T S R G E P V Q A V Y G F A K S L L K A L K E D G Y K A V F V V F D A K
 25 A P S F R H E A Y E A Y K A G R A P T P E D F P R Q L A L I K E L V D L L G F T R L E V Q G Y E A D D V L A T L A K K A E K E G Y E V R I
 L T A D R D L Y Q L V S D R V A V L H P E G H L I T P E W L W E K Y G L R P E Q W V D F R A L V G D P S D N L P G I K G I G E K T A L K L
 L K E W G S L E N L L K N L D R V K P E N V R E K I K A H L E D L R L S L E L S R V R T D L P L E V D F A K R R E P D R E R L R A F L E R
 L E F G S L L H E F G L L E S P K A L E E A P W P P P E G A F V G F V L S R K E P M W A D L L A L A A A K G G R V H R A P E P Y K A L R D
 L K E A R G L L A K D L S V L A L R E G L G L P P G D D P M L L A Y L L D P S N T T P E G V A R R Y G G E W T E E A G E R A A L S E R L F
 30 A N L W G R L E G E E R L L W L Y R E V E R P L S A V L A H M E A T G V R L D V A Y L R A L S L E V A E E I A R L E A E V F R L A G H P F
 N L N S R D Q L E R V L F D E L G L P A I G K T E K T G K R S T S A A V L E A L R E A H P I V E K I L Q Y R E L T K L K S T Y I D P L P D
 L I H P R T G R L H T R F N Q T A T A T G R L S S D P N L Q N I P V R T P L G Q R I R R A F I A E G G W L L V L D Y S Q M E L R V L A
 H L S G D E N L I R V F Q E G R D I H T E T A S W M F G V P R E A V D P L M R R A A K T I N F G V L Y G M S A H R L S Q E L A I P Y E E A
 Q A F I E R Y F Q S F P K V R A W I E K T L E E G R R R G Y V E T L F G R R R Y V P D L E A R V K S V R E A A E R M A F N M P V Q G T A A
 35 D L M K L A M V K L F P R L E E T G A R M L L Q V H D E L V L E A P K E R A E A V A R L A K E V M E G V Y P L A V P L E V E V G I G E D W
 L S A K E *

40 5D4 : 5' -
 A T G G C G A T G C T T C C C C T C T T T G A G C C C A A A G G C C G G G T C C T C C T G G T G G A C G G C C A C C A C C T G G C C T A C
 C G C A C C T T C T T C G C C C T G A A G G G C C T C A C C A C G A G T C G G G G C G A A C C G G T G C A G G C G G T C T A C G G C T T C
 G C C A A G A G C C T C C T C A A G G C C C T G A A G G A G A C G G G T A C A A G G C C A T C T T C G T G G T C T T T G A C G C C A A G
 G C C C C C T C C T T C C G C C A C G A G G C C C A C G A G G C C T A C A A G G C G G G G A G G G C C C G A G C C C C G A G G A C T T C
 45 C C C C G G C A G C T C G C C C T C A T C A A G G A G C T G G T G G A C C T C C T G G G G T T T A C C C G C C T C G A G G T C C A A G G C
 T A C G A G G C G G A C G A C G T C C T C G C C A C C C T G G C C A A G A A G G C G G A A A A G A A G G G T A C G A G G T G C G C A T C
 C T C A C C G C C G A C C G G G A C C T C T A C C A G C T C G T C T C C G A C C G C G T C G C C G T C C T C C A C C C C G A G G G C C A C
 C T C A T C A C C C C G G A G T G G C T T T G G G A G A A G T A C G G C C T C A G G C C G G A G C A G T G G G T G G A C T T C C G C G C C
 C T C G T G G G G G A C C C C T C C G A C A A C C T C C C C G G G G T C A A G G G C A T C G G G G A G A A G A C C G C C C T C A A G C T C
 50 C T C A A G G A G T G G G G A A G C C T G G A A A A C C T C C T C A A G A A C C T G G A C C G G C T G A A G C C C G C C A T C C G G G A G
 A A G A T C C T G G C C C A C A T G G A C G A T C T G A A G C T C T C C T G G G A C C T G G C C A A G G T G C G C A C C G A C C T G C C C
 C T G G A G G T G G A C T T C G C C A A A A G G C G G G A G T C C G A T C G G G A G A G G C T T A G G G C C T T T C T G G A G A G G C T T
 G A G T T T G G C A G C C T C C T C C A C G A G T T C G G C C T T C T G G A A A G C C C C A A G G C C C T G G A G G A G G C C C C C T G G
 C C C C C G C C G G T A G G G G C C T T C G T G G G C T T T G T G C T T T C C C G C A A G G A G C C C A T G T G G G C C G A T C T T C T G
 55 G C C C T G G C C G C C G C C A G G G G T G G T C G G G T C C A C C G G G C C C C G A G C C T T A T A A A G C C C T C A G A G A C C T G
 A A G G A G G C G C G G G G G C T T C T C G C C A A A G A C C T G A G C G T T C T G G C C C T G A G G G A A G G C C T T G G C C T C C C G
 C C C G G C G A C G A C C C A T G C T C C T C G C C T A C C T C C T G G A C C C T T C C A A C A C C A C C C C G A G G T G G T G G C C

CGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCC
 AACCTGTGGGGGAGGCTTGAGGGGGAGGGGAGGCTCCTTTGGCTTTACCGGGGGGTGGAGAGGCCCTT
 TCCGCTGTCTTGGCCACATGGAGGCCACAGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTC
 CTGGAGGTGGCCGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAAC
 5 CTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAGACG
 GAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCTTGAGAGGCCCTCCGCGAGGCCACCCCATCGTG
 GAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGCACTTACATTGACCCCTTGCCGGACCTC
 ATCCACCCAGGACGGGCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGT
 AGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTCGGGCAGAGGATCCGCCGGGCCTTC
 10 ATCGCCGAGGGGGGTGGCTATTGGTGGTCCTGGACTATAGCCAGATGGAGCTCAGGGTGCTGGCCAC
 CTCTCCGGCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAAACCGCCAGC
 TGGATGTTTCGGCGTCCCCCGGGAGGCCGTGGACCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTC
 GGGGTTCTCTACGGCATGTTCGGCCACCGCCTCTCCAGGAGCTAGCCATCCCTTACGAGGAGGCCAG
 GCCTTCATTGAGCGCTACTTCCAAAGCTTCCCCAAGGTGCGGGCCTGGATAGAAAAGACCTTGAGGAG
 15 GGGAGGAAGCGGGGCTACGTGGAAACCTCTTCGGAAGAAGGCGCTACGTGCCCGACCTCAACGCCCGG
 GTGAAGAGCGTCAGGGAGGCCGCGGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCGAC
 CTCACGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAACGGGGGCCAGGATGCTCCTTCAG
 GTCCACGACGAGCTGGTCCTCGAGGCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAG
 GTCATGGAGGGGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTT
 20 TCCGCCAAGGGTTAG-3'

5D4 amino acid sequence:

MAMLPLFEPKGRVLLVDGHHLAYRTFFALKGLTTSRGEVPQAVYGFASLLKALKEDG
 25 YKAI FVVFDAKAPSFRHEAHEAYKAGRAPSPEDFPRQLALIKELVDLLGFTRLEVQGY
 EADDVLATLAKKAEKEGYEVRI LTADRDLYQLVSDRVAVLHPEGHLITPEWLWEKYGL
 RPEQWVDFRALVGDPSDNLPGVKIGIKTALKLLKEWGSLENLLKNLDRLKPAIREKI
 LAHMDDLKLSWDLAKVRTDLPLEVDFAKRRESDRERLRAFLEFGLSLLHEFGLLES
 PKALEEAPWPPPVGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEA
 30 RGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEVVARRYGGEWTEEAGERA
 ALSERLFANLWGRLEGEGRLLWLYRGVERPLSAVLAHMEATGVRLDVAYLRALSLEVA
 EEIARLEAEVFRLAGHPFNLSRDQLERVLFDLGLPAIGKTEKTGKRSTSAVLEAL
 REAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPN
 LQNI PVRTPLGQRI RRAFI AEGGWLLVVDYSQMELRVLAHLSGDENLIRVFQEGRDI
 35 HTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERY
 FQSFPKVRWIEKTLEEGRKRGYVETLFGRRRYVPDLNARVKS VREA AERMAFNMPVQ
 GTAADLTKLAMVKLFPRLEETGARMLLQVHDEL VLEAPKERA EAVARLAKEVM EGVYP
 LAVPLEVEVGIGEDWLSAKG*

40

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
 45 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' ,
 5 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 as 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
 10 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
 15 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
 25 CGA CGG AAG CCG -3' , (X=dU-biotin (Glen research)). The lysates used were concentrated 4-fold.

4. Basic site bypass ELISA

30 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.

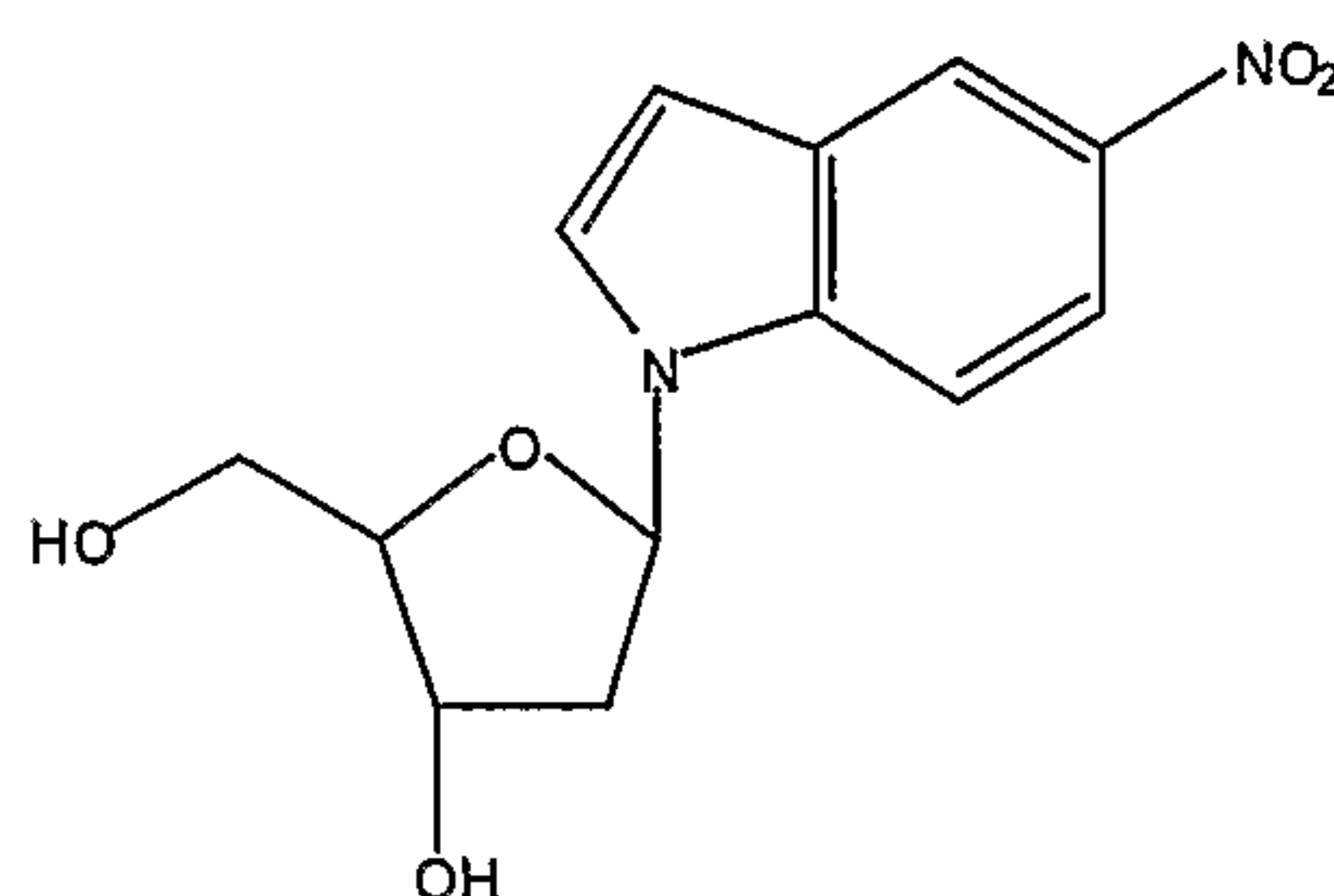
5

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

1: Extension opposite 5-nitroindole.

10 Primer 5'- TAATACGACTCACTATAGGGAGA
 Template 3'- ATTATGCTGAGTGATATCCCTCT5ATCGAT

5 = 5-Nitroindole



15

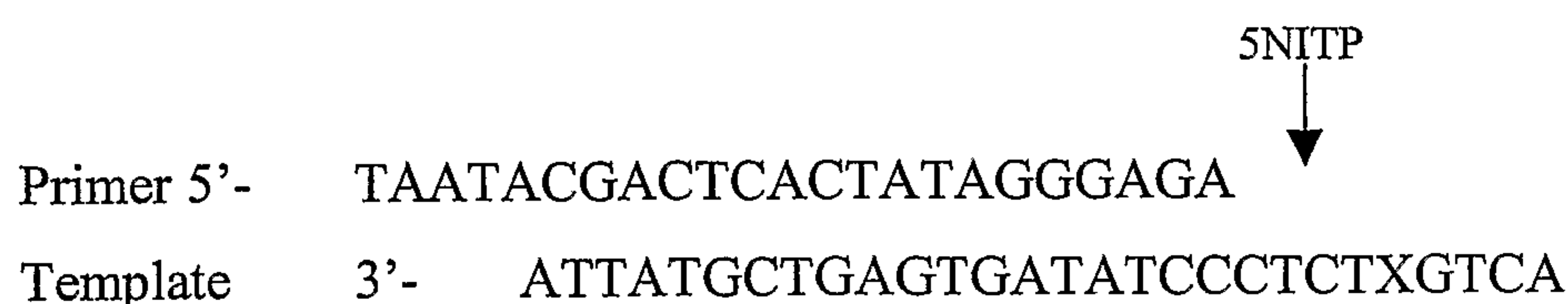
Primer extension reactions were carried out as follows:

50pmol of ³²P-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
 20 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
 25 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).

5



10

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 44 μl were
 15 annealed in 1X Taq buffer. 4D11 or 5D4 polymerase as cell lysate (6 μl) was added
 and reactions were incubated at 50°C for 15 minutes followed by addition of d5NITP
 (1 μl in total volume of 50 μl , final dNTP concentration 40 μM). 8 μl samples were taken
 at various time points and added to 8 μl stop solution (7M urea, 100mM EDTA
 containing xylene cyanol F). At the end of the time course the 4 native dNTPs were
 20 added (final concentration each dNTP 40 μM) 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) :

25

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).

30

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

5 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

10 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,

15 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

20 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
5 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
10 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
15 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
20 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 assay²⁶ (Genecheck, Ft. Collins, CO) according to manufacturers instructions. Alternatively, amplification products derived from 2x50 cycles of PCR of 2 targets with different GC content
5 (HIV *pol* (38% GC), *Taq* (68% GC)) were cloned, 40 clones (800bp each) were sequenced and mutations (*wtTaq* (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
10 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 *wtTaq*, with a clearly increased propensity for transversions, in particular G/C->C/G transversions (Fig 22B).

15

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*
20 and M4 closely matched those of *wtTaq* (Fig.23), indicating that both *M1* and *M4* exhibit processivity equal (or higher) than *wtTaq*. 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.

- 5 Oligonucleotide primer 32 (5'-GCG GTG TAG AGA CGA GTG CGG AG-3') was ³²P-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
- 10 DNA substrates (10nM) in 10mM Tris-HCl at pH 9.0, 5mM MgCl₂, 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
- 15 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: $I_n \times 100\% / (I_1 + I_2 + \dots + I_n)$, where I_n is the intensity of the band at position 22 or 23; $I_1, I_2, \dots$ is the intensity of the band at position 1, 2...
- 20 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.

5

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
10 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.

15 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.
- 20 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.
- 25 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.
- 5 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.
- 10 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.
- 15 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.
- 20 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.
- 25 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.
- 30 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*
5 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*
10 *Biotechnol.* **16**, 258-61.
34. Murata, T., Iwai, S. & Ohtsuka, E. (1990) *Nucleic Acids Res* **18**, 7279-7286.

SEQUENCE TABLE

<110> Medical Research Council

<120> Polymerase

<130> 757-143

<140> 2,544,841

<141> 2004-11-03

<150> GB0325650.0

<151> 2003-11-03

<150> GB0410871.8

<151> 2004-05-14

<160> 121

<170> PatentIn version 3.0

<210> 1

<211> 832

<212> PRT

<213> Artificial Sequence

<220>

<223> Engineered DNA polymerase

<400> 1

Met	Arg	Gly	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu
1				5					10					15	

Val	Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	His	Ala	Leu	Lys	Gly
			20					25					30		

Leu	Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Ala	Val	Tyr	Gly	Phe	Ala
			35				40					45			

Lys	Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Asp	Ala	Val	Ile	Val
	50					55					60				

Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Glu	Ala	Tyr	Gly	Gly
65					70					75				80	

Tyr	Lys	Ala	Ala	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu
				85					90					95	

Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Leu	Ala	Arg	Leu	Glu
			100					105					110		

Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Ser	Leu	Ala	Lys	Lys
		115					120					125			

Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Lys	Gly	130	135	140	
Leu	Tyr	Gln	Leu	Leu	Ser	Asp	Arg	Ile	His	Val	Leu	His	Pro	Glu	Gly	145	150	155	160
Tyr	Leu	Ile	Thr	Pro	Ala	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	165	170	175	
Asp	Gln	Trp	Ala	Asp	Tyr	Arg	Ala	Leu	Thr	Gly	Asp	Glu	Ser	Asp	Asn	180	185	190	
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Arg	Lys	Leu	Leu	195	200	205	
Glu	Glu	Trp	Gly	Ser	Leu	Glu	Ala	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Leu	210	215	220	
Lys	Pro	Ala	Ile	Arg	Glu	Lys	Ile	Leu	Ala	His	Met	Asp	Asp	Leu	Lys	225	230	235	240
Leu	Ser	Trp	Asp	Leu	Ala	Lys	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu	Val	245	250	255	
Asp	Phe	Ala	Lys	Arg	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala	Phe	260	265	270	
Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	Leu	275	280	285	
Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	Gly	290	295	300	
Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Arg	Glu	Pro	Met	Trp	Ala	Asp	305	310	315	320
Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	Pro	325	330	335	
Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	Leu	340	345	350	
Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	Pro	355	360	365	
Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	Asn	370	375	380	
Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	Glu	385	390	395	400
Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	Leu	405	410	415	
Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	Glu	420	425	430	

Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr	Gly
		435					440					445			
Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val	Ala
	450					455					460				
Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly	His
465					470					475					480
Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe	Asp
				485					490					495	
Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys	Arg
			500					505					510		
Ser	Thr	Ser	Ala	Ala	Val	Leu	Gly	Ala	Leu	Arg	Glu	Ala	His	Pro	Ile
		515					520					525			
Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	Thr
	530					535					540				
Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg	Leu
545					550					555					560
His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	Ser
				565					570					575	
Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly	Gln
			580					585					590		
Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val	Val
		595					600					605			
Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser	Gly
	610					615					620				
Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His	Thr
625					630					635					640
Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp	Pro
				645					650					655	
Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr	Gly
			660					665					670		
Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu	Glu
		675					680					685			
Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val	Arg
		690				695					700				
Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Arg	Arg	Gly	Tyr	Val
705					710					715					720
Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Glu	Ala	Arg
				725					730					735	

Val Lys Ser Val Arg Gly Ala Ala Glu Arg Met Ala Phe Asn Met Pro
740 745 750

Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys Leu
755 760 765

Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His
770 775 780

Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala
785 790 795 800

Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro
805 810 815

Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu
820 825 830

<210> 2

<211> 832

<212> PRT

<213> Artificial Sequence

<220>

<223> Artificial DNA polymerase

<400> 2

Met Arg Gly Met Leu Pro Leu Tyr Glu Pro Lys Gly Arg Val Leu Leu
1 5 10 15

Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys Gly
20 25 30

Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala
35 40 45

Lys Ser Leu Leu Lys Ala Leu Lys Glu Gly Gly Asp Ala Val Ile Val
50 55 60

Val Phe Asp Ala Lys Ala Pro Ser Phe Pro His Glu Ala Tyr Gly Gly
65 70 75 80

Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
85 90 95

Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Thr Arg Leu Glu
100 105 110

Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys Lys
115 120 125

Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys Asp
130 135 140

Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu Gly
145 150 155 160

Tyr	Leu	Ile	Thr	Pro	Ala	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro			
				165					170					175				
Asp	Gln	Trp	Ala	Asp	Tyr	Arg	Ala	Leu	Thr	Gly	Asp	Glu	Ser	Asp	Asn			
			180					185					190					
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Arg	Lys	Leu	Leu			
		195					200					205						
Glu	Glu	Trp	Gly	Ser	Leu	Glu	Ala	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Leu			
	210					215					220							
Lys	Pro	Ala	Ile	Arg	Glu	Lys	Ile	Leu	Ala	His	Met	Asp	Asp	Leu	Lys			
225					230					235					240			
Leu	Ser	Trp	Asp	Arg	Ala	Lys	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu	Val			
				245					250					255				
Asp	Phe	Ala	Lys	Arg	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala	Phe			
			260					265					270					
Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	Leu			
		275					280					285						
Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	Gly			
	290					295				300								
Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	Asp			
305					310					315					320			
Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	Pro			
				325					330					335				
Glu	Pro	Tyr	Lys	Ala	Leu	Gly	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	Leu			
			340					345					350					
Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	Pro			
		355					360					365						
Pro	Asp	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	Asn			
	370					375					380							
Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	Glu			
385					390					395					400			
Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	Leu			
				405					410					415				
Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	Glu			
			420					425					430					
Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr	Gly			
		435					440					445						
Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val	Ala			
	450					455					460							

Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly	His	465	470	475	480
Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe	Asp	485	490	495	
Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys	Arg	500	505	510	
Ser	Thr	Ser	Ala	Ala	Val	Leu	Gly	Ala	Leu	Arg	Glu	Ala	His	Pro	Ile	515	520	525	
Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	Thr	530	535	540	
Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg	Leu	545	550	555	560
His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	Ser	565	570	575	
Ser	Asp	Pro	Asn	Leu	Gln	Ser	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly	Gln	580	585	590	
Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val	Ala	595	600	605	
Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser	Gly	610	615	620	
Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His	Thr	625	630	635	640
Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp	Pro	645	650	655	
Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr	Gly	660	665	670	
Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu	Glu	675	680	685	
Ala	Gln	Ala	Phe	Ile	Lys	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val	Arg	690	695	700	
Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Arg	Arg	Gly	Tyr	Val	705	710	715	720
Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Glu	Ala	Arg	725	730	735	
Val	Lys	Ser	Val	Arg	Glu	Pro	Ala	Glu	Arg	Met	Ala	Phe	Asn	Met	Pro	740	745	750	
Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val	Lys	Leu	755	760	765	

Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val His
 770 775 780

Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val Ala
 785 790 795 800

Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val Pro
 805 810 815

Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys Glu
 820 825 830

<210> 3

<211> 2490

<212> DNA

<213> Artificial Sequence

<220>

<223> Engineered polymerase

<400> 3

```

atgctccctc tttttgagcc caaaggccgc gtcctcctgg tggacggcca ccacctggcc      60
taccgcacct tccacgccct gaagggcctc accaccagcc ggggggagcc ggtgcaggcg      120
gtctacggct tcgccaagag cctcctcaag gccctcaagg aggacgggga cgcggtgatc      180
gtgggtctttg acgccaaggc cccctccttc cgccacgagg cctacggggg gtacaaggcg      240
gcccggggccc ccacgccgga ggactttccc cggcaactcg cctcatcaa ggagctggtg      300
gatctcctgg ggctggcgcg cctcgaggtc ccgggctacg aggcggacga cgtcctggcc      360
agcctggcca agaaggcgga aaaggagggc tacgagggtc gcatectcac cgccgacaaa      420
ggcctttacc agctcctttc cgaccgcata cacgtcctcc accccgaggg gtacctcatc      480
accccggcct ggctttggga aaagtacggc ctgaggcccc accagtgggc cgactaccgg      540
gccctgaccg gggacgagtc cgacaacctt cccgggggtca agggcatcgg ggagaagacg      600
gcgaggaagc ttctggagga gtggggggagc ctggaagccc tctcaagaa cctggaccgg      660
ctgaagcccc ccatccggga gaagatcctg gccacatgg acgatctgaa gctctcctgg      720
gatctggcca aggtgcgcac cgacctgccc ctggagggtg acttcgccaa aaggcgggag      780
cccgaccggg agaggcttag ggcctttctg gagaggcttg agtttggcag cctcctccac      840
gagttcggcc ttctggaaag cccaaggcc ctggaggagg cccctggcc cccgccggaa      900
ggggccttcg tgggctttgt cctttcccgc agggagccca tgtgggccga tcttctggcc      960
ctggccgccc ccaggggggg ccgggtccac cgggcccccg agccttataa agcctcagg      1020
gacctgaagg aggcgcgggg gcttctcgcc aaagacctga gcgttctggc cctgagggaa      1080

```

```

ggccttggcc tcccgcccg cgacgacccc atgctcctcg cctacctcct ggacccttcc 1140
aacaccaccc ccgaggggggt ggcccggcgc tacggcgggg agtggacgga ggaggcgggg 1200
gagcggggccg ccttttccga gaggtctctc gccaacctgt gggggaggct tgagggggag 1260
gagaggctcc tttggcttta ccgggaggtg gagaggcccc ttcccgctgt cctggcccac 1320
atggaggcca cgggggtgcg cctggacgtg gcctatctca gggccttgtc cctggaggtg 1380
gccgaggaga tcgcccgcct cgaggccgag gtcttccgcc tggccggcca ccccttcaac 1440
ctcaactccc gggaccagct ggaaagggtc ctctttgacg agctagggtt tcccgccatc 1500
ggcaagacgg agaagaccgg caagcgctcc accagcgccg ccgtcctggg ggccctccgc 1560
gaggcccacc ccacgtgga gaagatcctg cagtaccggg agctcaccaa gctgaagagc 1620
acctacattg accccttacc ggacctcacc caccacagga cgggccgcct ccacacccgc 1680
ttcaaccaga cggccacggc cacgggcagg ctaagtagct ccgatcccaa cctccagaac 1740
atccccgtcc gcaccccgct tgggcagagg atccgcccgg ccttcacgc cgaggagggg 1800
tggctattgg tggctctgga ctatagccag atagagctca gggtgctggc ccacctctcc 1860
ggcgacgaga acctgatccg ggtcttccag gaggggcggg acatccacac ggagaccgcc 1920
agctggatgt tcggcgctcc ccgggaggcc gtggaccccc tgatgcgccg ggcggccaag 1980
accatcaact tcggggctct ctacggcatg tcggcccacc gcctctccca ggagctagcc 2040
atcccttacg aggaggccca ggccttcatt gagcgctact ttcagagctt cccaagggtg 2100
cgggcctgga ttgagaagac cctggaggag ggcaggaggc gggggtacgt ggagaccctc 2160
ttcggccgcc gcgctacgt gccagaccta gagggccggg tgaagagcgt gcggggggcg 2220
gccgagcgca tggccttcaa catgcccgtc cagggcaccc ccgccgacct catgaagctg 2280
gctatgggtga agctcttccc caggctggag gaaatggggg ccaggatgct ccttcaggtc 2340
cacgacgagc tggctctcga ggcccacaaa gagagggcgg aggcctggc ccggctggcc 2400
aaggaggtca tggaggggggt gtatccctg gccgtgcccc tggaggtgga ggtggggata 2460
ggggaggact ggctctccgc caaggagtga 2490

```

```

<210> 4
<211> 2490
<212> DNA
<213> Artificial Sequence

<220>
<223> Engineered polymerase

```

<400> 4
atgctccctc tttatgagcc caagggccgc gtctctctgg tggacggcca ccacctggcc 60
taccgcacct tccacgccct gaagggcctc accaccagcc ggggggagcc ggtgcaggcg 120
gtctacggct tcgccaagag cctcctcaag gccctcaagg agggcgggga cgcggtgatc 180
gtggtctttg acgccaaggc cccctccttc ccccatgagg cctacggggg gtacaaggcg 240
ggccggggccc ccacgccgga ggactttccc cgacaactcg cctcatcaa ggagctggtg 300
gacctcctgg ggctgacgcg cctcgaggtc ccgggctacg aggcggacga cgtcctggcc 360
agcctggcca agaaggcgga aaaggagggc tacgagggtc gcatectcac cgccgacaaa 420
gacctttacc agctcctttc cgaccgcata cacgtcctcc accccgaggg gtacctcatc 480
accccggcct ggctttggga aaagtacggc ctgaggcccg accagtgggc cgactaccgg 540
gcctgaccg gggacgagtc cgacaacctt cccgggggtca agggcatcgg ggagaagacg 600
gcgaggaagc ttctggagga gtgggggagc ctggaagccc tctcaagaa cctggaccgg 660
ctgaagcccg ccatccggga gaagatcctg gccacatgg acgatctgaa gctctcctgg 720
gaccgggcca aggtgcgcac cgacctgcc ctggagggtg acttcgccaa aaggcgggag 780
cccgaccggg agaggcttag ggcctttctg gagaggcttg agtttggcag cctcctccac 840
gagttcggcc ttctggaaag cccaaggcc ctggaggagg cccctggcc cccgccggaa 900
ggggccttcg tgggctttgt gctttccgc aaggagccca tgtgggccga tcttctagcc 960
ctggccgccc ccaggggggg cgggtccac cgggcccccg agccttataa agcctcggg 1020
gacctgaagg aggcgcgggg gcttctcgcc aaagacctga gcgttctggc cctgagggaa 1080
ggccttggcc tcccgcccga cgacgacccc atgctcctcg cctacctct ggaccttcc 1140
aacaccacc ccgaggggggt ggcccggcgc tacggcgggg agtggacgga ggaggcaggg 1200
gagcggggccg ccctttccga gaggtcttc gccaacctgt gggggaggct tgagggggag 1260
gaaaggctcc tttggcttta ccgggagggt gagaggcccc ttccgctgt cctggcccac 1320
atggaggcca cgggggtgcg cctggacgtg gcctatctca gggccttgtc cctggagggtg 1380
gccgaggaga tcgcccgcct cgaggccgag gtcttccgcc tggccggcca cccctcaac 1440
ctcaactccc gggaccagct ggaaagggtc ctctttgacg agctagggt tcccgccatc 1500
ggcaagacgg agaagaccgg caagcgtcc accagcgcg ccgtcctggg ggccctccgc 1560
gaggcccacc ccatcgtgga gaagatcctg cagtaccggg agctaccaa gctgaagagc 1620
acctacattg accccttgcc ggacctcatc cccccagga cgggcgcct ccacaccgc 1680

```

ttcaaccaga cggccacggc cacgggcagg ctaagtagct ccgatcccaa cctccagagc 1740
atccccgtcc gcaccccgtc tgggcagagg atccgccggg ccttcacgc cgaggagggg 1800
tggctattgg tggccctgga ctatagccag atagagctca gggtgctggc ccacctctcc 1860
ggcgacgaga acctgatccg ggtcttccag gagggggcggg acatccacac ggagaccgcc 1920
agctggatgt tcggcgtccc ccgggaggcc gtggaccccc tgatgcgccg ggcgggccaag 1980
accatcaact tcggggtcct ctacggcatg tcggcccacc gcctctccca ggagctagcc 2040
atcccttacg aggaggccca ggccttcatt aagcgctact ttcagagctt cccaaggtg 2100
cgggcctgga ttgagaagac cctggaggag ggcaggaggc ggggggtacgt ggagaccctc 2160
ttcggccgcc gccgctacgt gccagaccta gagggccggg tgaagagcgt gcgggagccg 2220
gccgagcgca tggccttcaa catgcccgtc caggggtaccg ccgccgacct catgaagctg 2280
gctatgggtga agctcttccc caggctggag gaaatggggg ccaggatgct ccttcaggtc 2340
cacgacgagc tggctctcga ggccccaaaa gagagggcgg aggccgtggc ccggctggcc 2400
aaggaggtca tggagggggg gtatcccctg gccgtgcccc tggaggtgga ggtggggata 2460
ggggaggact ggctctccgc caaggagtga 2490

```

```

<210> 5
<211> 38
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> primer

```

```

<400> 5
caggaaacag ctatgacaaa aatctagata acgaggga 38

```

```

<210> 6
<211> 45
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> primer

```

```

<400> 6
gtaaaacgac ggccagtacc accgaactgc gggtagcgcc aagcc 45

```

```

<210> 7
<211> 23
<212> DNA

```

<213> Artificial Sequence

<220>

<223> primer

<400> 7

aaaaatctag ataacgaggg caa

23

<210> 8

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 8

accaccgaac tgcgggtgac gccaaagcg

28

<210> 9

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 9

gaactgcggg tgacgccaaag cgca

24

<210> 10

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 10

ccgaactgcg ggtgacgcca agcgg

25

<210> 11

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 11

gaactgcggg tgacgccaaag cgcg

24

<210> 12
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 12
aaaaatctag ataacgaggg caa 23

<210> 13
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 13
ccgactggcc aagattagag agtatgg 27

<210> 14
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 14
gatttccacg gataagactc cgcatcc 27

<210> 15
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 15
ggcagacgat gatgcagata accagag 27

<210> 16
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 16

gccgatagat agccacggac ttcgtag 27

<210> 17
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 17
ggagtagatg cttgcttttc tgagcc 26

<210> 18
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 18
gagttcgtgc ttaccgcaga atgcag 26

<210> 19
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 19
accgaactgc gggtgacgcc aagcg 25

<210> 20
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 20
accgaactgc gggtgacgcc aagcc 25

<210> 21
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 21
accgaactgc gggtgacgcc aagca 25

<210> 22
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 22
aaacagcgct tggcgtcacc cgcagttcgg t 31

<210> 23
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 23
aaacagggct tggcgtcacc cgcagttcgg t 31

<210> 24
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 24
aaacagagct tggcgtcacc cgcagttcgg t 31

<210> 25
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 25
aaacaccgct tggcgtcacc cgcagttcgg t 31

<210> 26
<211> 51
<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 26

agctaccatg cctgcacgaa ttcggcatcc gtcgcgacca cggtcgcagc g

51

<210> 27

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> misc_feature

<222> (22)..(22)

<223> n is an abasic site

<400> 27

agctaccatg cctgcacgac ancgcatcc gtcgcgacca cggtcgcagc g

51

<210> 28

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<220>

<221> misc_feature

<222> (21)..(22)

<223> n is a CPD dimer

<400> 28

agctaccatg cctgcacgaa nncgcatcc gtcgcgacca cggtcgcagc g

51

<210> 29

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 29

cgtggtcgcg acggatgccg

20

<210> 30

<211> 23

<212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer

<400> 30
 taatacgact cactataggg aga

23

<210> 31
 <211> 28
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer

<220>
 <221> misc_feature
 <222> (5)..(5)
 <223> n is 5NI

<400> 31
 actgntctcc ctatagtgag tcgtatta

28

<210> 32
 <211> 40
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 32
 caggaaacag ctatgacaaa aatctagata acgagggcaa

40

<210> 33
 <211> 45
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Primer

<400> 33
 gtaaaacgac ggccagtacc accgaactgc gggtagcgcc aagcg

45

<210> 34
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Primer

<400> 34

gtaaaacgac ggccagttta ttaaccaccg aactgc

36

<210> 35

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 35

caggaaacag ctatgactcg acaaaaatct agataacgac c

41

<210> 36

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 36

gtaaaacgac ggccagt

17

<210> 37

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 37

caggaaacag ctatgac

17

<210> 38

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 38

caggaaacag ctatgacaaa agtgaaatga atagttagac tttt

44

<210> 39

112

<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 39
gtaaaacgac ggccagtctt cacaggtcaa gcttattaag gtg 43

<210> 40
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 40
caggaaacag ctatgaccat tgatagagtt attttaccac aggg 44

<210> 41
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 41
gtaaaacgac ggccagtctt cacaggtcaa gcttattaag gtg 43

<210> 42
<211> 38
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 42
caggaaacag ctatgacaaa aatctagata acgagggg 38

<210> 43
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 43
gtaaaacgac ggccagtacc accgaactgc gggtgacgcc aagcc 45

113

<210> 44
 <211> 41
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 44
 caggaaacag ctatgactcg acaaaaatct agataacgac c 41

<210> 45
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 45
 gtaaaacgac ggccagttta ttaaccaccg aactgc 36

<210> 46
 <211> 67
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<220>
 <221> misc_feature
 <222> (45)..(45)
 <223> n is dU-biotin

<400> 46
 agctaccatg cctgcacgca gtcggcatcc gtcgcgacca cgttnttctg ggtcgcgacg 60
 gatgccg 67

<210> 47
 <211> 67
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<220>
 <221> misc_feature
 <222> (22)..(22)

<223> n is an abasic site

<220>

<221> misc_feature

<222> (45)..(45)

<223> n is dU- biotin

<400> 47

agctaccatg cctgcacgca gncggcatcc gtcgcgacca cgttnttcgt ggtcgcgacg 60

gatgccg 67

<210> 48

<211> 67

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<220>

<221> misc_feature

<222> (22)..(22)

<223> n is hydantoin

<220>

<221> misc_feature

<222> (45)..(45)

<223> n is biotin-dU

<400> 48

agctaccatg cctgcacgca gncggcatcc gtcgcgacca cgttnttcgt ggtcgcgacg 60

gatgccg 67

<210> 49

<211> 2502

<212> DNA

<213> Artificial Sequence

<220>

<223> Cloned sequence

<400> 49

atggcgatgc ttccctcttt tgagcccaag ggccgcgtcc tcctggtgga cggccaccac 60

ctggcctacc gcaccttctt cgcctgaag ggcccacca cgagccgggg cgaaccggtg 120

caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180

gccgtcttcg tggcttttga cgccaaggcc cctcattcc gccacaaggc ctacgaggcc 240

tacagggcgg ggagggcccc gacccccgag gacttcccc ggcagctcgc cctcatcaag 300

gagctggtgg acctcctggg gtttaccgc ctcgaggtcc ccggctacga ggcggaacgac	360
gttctcgcca ccgtggccaa gaaggcggaa aaggaggggt acgaggtggg catcctcacc	420
gccgaccgcg gcctctacca actcgtctct gaccgcgtcg ccgtcctcca ccccgagggc	480
cacctcatca ccccgagtg gctttgggag aagtacggcc tcaggccgga gcagtgggtg	540
gacttcgcg ccctcgtggg ggacccctcc gacaacctcc ccgggggtcaa gggcatcggg	600
gagaagaccg ccctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac	660
ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagacctc	720
aggctctcct tggagctctc ccgggtgcgc accgacctcc ccctggaggt ggacctcgcc	780
caggggcggg agcccgaccg ggaggggctt agggcctttc tggagaggct tgagtttggc	840
agcctcctcc acgagttcgg ccttctggaa agccccaagg ccctggagga ggccccctgg	900
ccccgcgcg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc	960
gatcttctgg ccctggccgc cgccaggggg ggccgggtcc accgggcccc cgagccttat	1020
aaagccctca gagacctgaa ggaggcgcgg gggcttctcg ccaaagacct gagcgttctg	1080
gccctgaggg aaggccttgg cctcccgccc ggcgacgacc ccattgctct cgcctacctc	1140
ctggaccctt ccaacaccac ccccgagggg gtggcccggc gctacggcgg ggagtggacg	1200
gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg	1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggagaggcc cctttccggt	1320
gtcctggccc acatggaggc cacaggggtg cgctggacg tggcctatct cagggccttg	1380
tccttgagg tggccgagga gatcgccgc ctcgaggccg aggtcttccg cctggccggc	1440
cacccttca acctcaactc ccgggaccag ctggaaaggg tcctctttga cgagctaggg	1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgt ccaccgggc cgccgtcctg	1560
gaggccctcc acgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc	1620
aagctgaaga gcacctacat tgaccccttg ccggacctca tccaccccag gacgggccgc	1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc	1740
aacctccaga acatccccgt ccgcacccag cttgggcaga ggatecgccg ggccttcate	1800
gccgaggagg ggtggtatt ggtggtcctg gactatagcc agatagagct cagggtgctg	1860
gccacctct ccggcgacga gaacctgac cggtcttcc aggaggggcg ggacatccac	1920
acggaaaccg ccagctggat gtcggcgtc cccagaggag ccgtggaccc cctgatgcgc	1980
cgggcggcc aaccatcaa ctccggggtt ctctacggca tgcggccta ccgcctctcc	2040

```

caggagctag ccatccctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttccccaagg tgcgggcctg gattgggaag accctggagg agggcaggag gcgggggtac 2160
gtggagaccc tcttcggccg ccgccgctac gtgccagacc tagaggcccg ggtgaagagc 2220
gtgcgggagg cggccgagcg catggccttc aacacgcccg tccagggcac cgccgccgac 2280
ctcatgaagc tagctatggt gaagctcttc cccaggctgg aggaaatggg ggccaggatg 2340
ctccttcagg tccacgacga gctggtcctc gaggcccca aagagagggc ggaggccgtg 2400
gcccggtg ccaaggaggt catggagggg gtgtatcccc tggccgtgcc cctggaggtg 2460
gaggtgggga taggggagga ctggctctcc gcccaaggagt ga 2502

```

```

<210> 50
<211> 833
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 50

```

```

Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1          5          10          15

Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Pro
          20          25          30

Thr Thr Ser Arg Gly Glu Pro Val Gln Val Val Tyr Gly Phe Ala Lys
          35          40          45

Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe Val
          50          55          60

Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Lys Ala Tyr Glu Ala
65          70          75          80

Tyr Arg Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
          85          90          95

Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu Glu
          100          105          110

Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Val Ala Lys Lys
          115          120          125

Ala Glu Lys Glu Gly Tyr Glu Val Gly Ile Leu Thr Ala Asp Arg Gly
          130          135          140

Leu Tyr Gln Leu Val Ser Asp Arg Val Ala Val Leu His Pro Glu Gly
145          150          155          160

```

His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	
				165					170					175		
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn	
			180					185					190			
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu	
		195					200					205				
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val	
	210					215					220					
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu	
225					230					235					240	
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu	
				245					250					255		
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	Ala	
			260					265					270			
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	
		275					280					285				
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	
	290					295					300					
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	
305					310					315					320	
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	
				325					330					335		
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	
			340					345					350			
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	
		355					360					365				
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	
	370					375					380					
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	
385					390					395					400	
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	
				405					410					415		
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	
			420					425					430			
Glu	Val	Glu	Arg	Pro	Leu	Ser	Val	Val	Leu	Ala	His	Met	Glu	Ala	Thr	
		435					440					445				
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val	
	450					455					460					

Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly	465	470	475	480
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe	485	490	495	
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys	500	505	510	
Arg	Ser	Thr	Gly	Ala	Ala	Val	Leu	Glu	Ala	Leu	His	Glu	Ala	His	Pro	515	520	525	
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	530	535	540	
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg	545	550	555	560
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	565	570	575	
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Gln	Leu	Gly	580	585	590	
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val	595	600	605	
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser	610	615	620	
Gly	Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His	625	630	635	640
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Gln	Glu	Ala	Val	Asp	645	650	655	
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr	660	665	670	
Gly	Met	Ser	Ala	Tyr	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu	675	680	685	
Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val	690	695	700	
Arg	Ala	Trp	Ile	Gly	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Arg	Arg	Gly	Tyr	705	710	715	720
Val	Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Glu	Ala	725	730	735	
Arg	Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn	Thr	740	745	750	
Pro	Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val	Lys	755	760	765	

Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
 770 775 780

His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
 785 790 795 800

Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
 805 810 815

Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Glu

<210> 51
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 51
 atgcgtggta tgcctcctct ttttgagccc aagggccgcg tcctcctggt ggacggccac 60
 ctggcctacc gcaccttctt cgccttgaag ggccccacca cgagccgggg cgaaccggtg 120
 caggcgggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc cctccctcc gccacgaggc ctacgaggcc 240
 tacaaggcgg ggaggggccc gacccccgag gacttcccc ggcagctcgc cctcatcaag 300
 gagctggtgg acctcctggg gtttaccgcg ctcgaggtcc ccggctacga ggcggaacgac 360
 gttctcgcca ccctggccaa gaaggcggaa aaggaggggt acgaggtgcg catcctcacc 420
 gccgaccgcg acctctacca actcgtctcc gaccgcgtcg ccgtcctcca ccccgagggc 480
 cacctcatca ccccgagagt gctttgggag aagtacggcc tcaggccgga gcagtgggtg 540
 gacttccgcg ccctcgtggg ggacccctcc gacaacctcc ccgggggtcaa gggcatcggg 600
 gagaggaccg ccctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac 660
 ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagacctc 720
 aggctctcct tggagctctc ccgggtgcgc accgacctcc ccctggaggt ggacctcgcc 780
 caggggcggg agcccgaccg ggagaggctt agggccttcc tggagaggct tgagtttggc 840
 agcctcctcc acgagttcgg ccttctggaa agccccaagg ccctggagga ggccccctgg 900
 ccccgccggg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc 960
 gatcttctgg ccctggccgc cgccaggggt ggtcgggtcc accgggcccc cgagccttat 1020

```

aaagccctca gggacttgaa ggaggcgcg gggcttctcg ccaaagacct gagcgttctg 1080
gccctaaggg aaggccttgg cctcccgccc ggcgacgacc ccatgctcct cgcctacctc 1140
ctggaccctt ccaacaccac ccccgagggg gtggcccggc gctacggcgg ggagtggacg 1200
gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggaag 1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggataggcc cctttccgct 1320
gtcctggccc acatggaggc cacaggggtg cgcttgacg tggcctatct cagggcctcg 1380
tccttgaggg tggccgagga gatcgccgc ctcgaggccg aggtcttccg cctggccggc 1440
cacccttca acctcaactc ccgggaccag ctggaaaggg tcctctttga cgagctaggg 1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgct ccaccagcgc cgccgtcctg 1560
gagggcctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgacccttg ccggacctca tccaccccag gacgggcccgc 1680
ctccacaccc gcttcaacca gacggccacg gccacaggca ggctaagtag ctccgatecc 1740
aacctccaga acatccccgt ccgcaccccg cttgggcaga ggatccgccg ggccttcate 1800
gccgaggagg ggtggctatt ggtggccctg gactatagcc agatagagct cagggtgctg 1860
gcccacctct ccggcgacga gaacctgate cgggtcttcc aggaggggcg ggacatccac 1920
acggagaccg ccagttggat gttcggcgtc ccccgaggagg ccgtggacce cctgatgcgc 1980
cgggcggcca agaccatcaa cttcgggggtc ctctacggca tgcgggcccg ccgcctctcc 2040
caggagctag ccatacctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttccccaagg tgcgggcctg gattgagaag acctggagg agggcaggag gcgggggtac 2160
gtggagacce tcttcggccg ccgccgtac gtgccagacc tagaggcccg ggtgaagagc 2220
gtgcgggagg cggccgagcg catggccttc aacatgcccg tccagggcac cgccgccgac 2280
ctcatgaagc tggctatggt gaagctcttc ccaggctgg aggaaatggg ggccaggatg 2340
ctccttcagg tccacgacga gctggtcctc gaggcccaa aagagagggc ggaggccgtg 2400
gcccggctgg ccaaggaggt catggagggg gtgtatcccc tggccgtgcc cctggaggtg 2460
gaggtgggga taggggagga ctggctctcc gccaaaggagt ga 2502

```

```

<210> 52
<211> 833
<212> PRT
<213> Artificial Sequence

```

<220>

<223> cloned sequence

<400> 52

Met	Arg	Gly	Met	Pro	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu
1				5					10					15	
Val	Asp	Gly	His	Leu	Ala	Tyr	Arg	Thr	Phe	Phe	Ala	Leu	Lys	Gly	Pro
			20					25					30		
Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Ala	Val	Tyr	Gly	Phe	Ala	Lys
		35					40					45			
Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	Val
	50					55					60				
Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Leu	Arg	His	Glu	Ala	Tyr	Glu	Ala
65					70					75					80
Tyr	Lys	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu
				85					90					95	
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu
			100					105					110		
Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys
		115					120					125			
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Asp
	130					135					140				
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly
145					150					155					160
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro
				165					170					175	
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn
			180					185					190		
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Arg	Thr	Ala	Leu	Lys	Leu	Leu
		195					200					205			
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val
	210					215					220				
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu
225					230					235					240
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu
				245					250					255	
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala
			260					265					270		
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu
		275					280					285			

Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	
	290					295					300					
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	
305					310					315					320	
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	
				325					330					335		
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	
			340					345					350			
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	
		355					360					365				
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	
	370					375					380					
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	
385					390					395					400	
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	
				405					410					415		
Leu	Trp	Gly	Lys	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	
			420					425					430			
Glu	Val	Asp	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr	
		435					440					445				
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Ser	Ser	Leu	Glu	Val	
	450					455					460					
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly	
465					470					475					480	
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe	
			485						490					495		
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys	
			500					505						510		
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro	
		515					520					525				
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	
	530					535					540					
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg	
545					550					555					560	
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	
				565					570					575		
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly	
			580					585					590			

Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
 595 600 605
 Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
 610 615 620
 Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
 625 630 635 640
 Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
 645 650 655
 Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
 660 665 670
 Gly Met Ser Ala Arg Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
 675 680 685
 Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
 690 695 700
 Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
 705 710 715 720
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
 725 730 735
 Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
 740 745 750
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765
 Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
 770 775 780
 His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
 785 790 795 800
 Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
 805 810 815
 Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Glu

<210> 53
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 53
atgcgtggta tgcatacctct ttttgagccc aagggcgcgc tcctcctggt ggacggccac 60
cacctggcct accgcacctt ccacgccctg aaggggctca ccaccagccg gggggagccg 120
gtgcgggcgg tccacggctt cgccaagagc ctctcaagg ccctgaagga ggacgggtac 180
aaggccgtct tcgtggtctt tgacgccaaag gccccctcct tccgccacga ggcctacgag 240
gcctacaagg cggggagggc cccgaccccc gaggacttcc cccggcagct cgcctcctc 300
aaggagctgg tggacctcct ggggtttacc cgcctcgagg tccccggcta cgaggcggac 360
gacgttctcg ccacctggc caagaaggcg gaaaaggagg ggtacgaggt gcgcatactc 420
accgccgacc gcgacctcta ccaactcgtc tccgaccgcg tcgccgtcct ccaccccgag 480
ggccacctca tcaccccgga gtggctttgg gagaagtacg gcctcaggcc ggagcagtgg 540
gtggacttcc gcgcctcgt gggggacccc tccgacaacc tccccggggt caagggcatc 600
ggggagaaga ccgccctcaa gctcctcaag gagtggggaa gcctggaaaa cctcctcaag 660
aacctggacc ggctgaagcc cgcctccgg gagaagatcc tggcccacat ggacgatctg 720
aagctctcct gggacctggc caaggtgcgc accgacctgc ccctagaggt ggacttcgcc 780
aaaaggcggg agcccgaccg ggagaggctt agggcctttc tggagaggct tgagcttggc 840
agcctcctcc acgagttcgg ccttctggaa agccccaaga ccctggagga ggctcctgg 900
ccccgcgcg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc 960
gatcttctgg ccctggccgc cgccaggggg ggccgggtcc accgggcccc cgagccttat 1020
aaagccctca gagacctgaa ggaggcgcgg gggcttctcg ccaaagacct gagcgttctg 1080
gcctgaggg aaggccttgg cctcccgcgc ggcgacgacc ccatgctcct cgcctacctc 1140
ctggaccctt ccaacaccac ccccgagggg gtggcccggc gctacggcgg ggagtggacg 1200
gaggaggcgg gggagcgggc cgccttttcc gagaggctct tcgccaacct gtgggggagg 1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggagaggcc cctttccggt 1320
gtcctggccc acatggaggc cacaggggtg cgcctggacg tggcctatct cagggccttg 1380
tcctggagg tggccgagga gatcgcccgc ctcgaggccg aggtcttccg cctggccggc 1440
cacccttca acctcaactc ccgggaccag ctggaaaggg tcctctttga cgagctaggg 1500
cttcccgcga tcggcaagac ggagaagacc ggcaagcgt ccaccggcgc cgcgtcctg 1560
gaggccctcc gcgaggccca cccatcgctg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgacccttg ccggacctca tccaccccag gacgggccgc 1680

```

ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
aacctccaga acatccccgt ccgcacccag cttgggcaga ggatccgccg ggccttcac 1800
gccgaggagg ggtggctatt ggtggtcctg gactatagcc agatagagct cagggtgctg 1860
gcccacctct ccggcgacga gaacctgatc cgggtcttcc aggaggggcg ggacatccac 1920
acggaaaccg ccagctggat gttcggcgtc cccagggagg ccgtggaccc cctgatgcgc 1980
cgggcggcca agaccatcaa cttcgggggt ctctacggca tgtcggccta ccgcctctcc 2040
caggagctag ccatccctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttccccaagg tgcgggcctg gattgggaag accctggagg agggcaggag gcgggggtac 2160
gtggagaccc tcttcggccg ccgccgctac gtgccagacc tagaggcccg ggtgaagagc 2220
gtgcgggagg cggccgagcg catggccttc aacacgcccg tccagggcac cgccgccgac 2280
ctcatgaagc tggctatggt gaagctcttc cccaggctgg aggaaatggg ggccaggatg 2340
ctccttcagg tccacgacga gctagtcctc gaggccccaa aagagagggc ggaggccgtg 2400
gcccggctgg ccaaggaggt catggagggg gtgtatcccc tggccgtgcc cctggaggtg 2460
gaggtgggga taggggagga ctggctctcc gccaaggagt ga 2502

```

```

<210> 54
<211> 833
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 54
Met Arg Gly Met His Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu
1          5          10          15
Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys Gly
          20          25          30
Leu Thr Thr Ser Arg Gly Glu Pro Val Arg Ala Val His Gly Phe Ala
          35          40          45
Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe
          50          55          60
Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu
65          70          75          80
Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
          85          90          95
Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu

```

126

100						105						110					
Glu	Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys		
		115					120					125					
Lys	Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg		
	130					135					140						
Asp	Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu		
145					150					155					160		
Gly	His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg		
				165					170					175			
Pro	Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp		
			180					185				190					
Asn	Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu		
		195					200					205					
Leu	Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg		
	210					215					220						
Leu	Lys	Pro	Ala	Ile	Arg	Glu	Lys	Ile	Leu	Ala	His	Met	Asp	Asp	Leu		
225					230					235					240		
Lys	Leu	Ser	Trp	Asp	Leu	Ala	Lys	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu		
				245					250					255			
Val	Asp	Phe	Ala	Lys	Arg	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala		
			260					265					270				
Phe	Leu	Glu	Arg	Leu	Glu	Leu	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu		
		275					280					285					
Leu	Glu	Ser	Pro	Lys	Thr	Leu	Glu	Glu	Ala	Ser	Trp	Pro	Pro	Pro	Glu		
	290					295					300						
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala		
305					310					315					320		
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala		
				325					330					335			
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu		
			340					345					350				
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu		
		355					360					365					
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser		
	370					375					380						
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr		
385					390					395					400		
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn		

405					410					415					
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg
			420					425					430		
Glu	Val	Glu	Arg	Pro	Leu	Ser	Val	Val	Leu	Ala	His	Met	Glu	Ala	Thr
		435					440					445			
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val
	450					455					460				
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly
465						470					475				480
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe
			485					490					495		
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys
			500					505					510		
Arg	Ser	Thr	Gly	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro
		515					520					525			
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser
	530					535					540				
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg
545						550					555				560
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser
				565				570						575	
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Gln	Leu	Gly
			580					585					590		
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val
		595					600					605			
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser
	610					615					620				
Gly	Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His
625						630					635				640
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Gln	Glu	Ala	Val	Asp
				645					650					655	
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr
			660					665					670		
Gly	Met	Ser	Ala	Tyr	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu
		675					680					685			
Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val
	690					695					700				
Arg	Ala	Trp	Ile	Gly	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Arg	Arg	Gly	Tyr

128

705		710		715		720									
Val	Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Glu	Ala
				725					730					735	
Arg	Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn	Thr
			740					745					750		
Pro	Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val	Lys
		755					760					765			
Leu	Phe	Pro	Arg	Leu	Glu	Glu	Met	Gly	Ala	Arg	Met	Leu	Leu	Gln	Val
	770					775					780				
His	Asp	Glu	Leu	Val	Leu	Glu	Ala	Pro	Lys	Glu	Arg	Ala	Glu	Ala	Val
785					790					795					800
Ala	Arg	Leu	Ala	Lys	Glu	Val	Met	Glu	Gly	Val	Tyr	Pro	Leu	Ala	Val
			805						810					815	
Pro	Leu	Glu	Val	Glu	Val	Gly	Ile	Gly	Glu	Asp	Trp	Leu	Ser	Ala	Lys
			820					825					830		

Glu

<210> 55
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 55
 atgcgtggta tgcttctct ttttgagccc aagggccgcg tcctctggt ggacggccac 60
 cacctggcct accgcacctt ctctgccctg aagggcctca ccacgagccg gggcgaaccg 120
 gtgcaggcgg tctacggctt cgccaagagc ctctcaagg cctgaagga ggacgggtac 180
 aagggcgtct tcgtggtctt tgacgccaaag gccccctccc tccgccacga ggcttacgag 240
 gcctacaagg cggggagggc cccgaccccc gaggacttcc cccggcagct cgccctcatc 300
 aaggagctgg tggacctcct ggggtttacc cgctcgagg tccccggcta cgaggcggac 360
 gacgttctcg ccacctggc caagaaggcg gaaaaggagg ggtacgaggt gcgcctcctc 420
 accgccgacc gcgacctcta ccaactcgtc tccgaccgcg tcgccgtcct ccaccccgag 480
 ggccacctca tcaccccgga gtggctttgg gagaagtacg gcctcaggcc ggagcagtgg 540
 gtggacttcc gcgcctcgt gggggacccc tccgacaacc tccccggggt caagggcatc 600
 ggggagaaga ccgcctcaa gctcctcaag gagtggggaa gcctggaaaa cctcctcaag 660

aacctggacc ggctgaagcc cgccatccgg gagaagatcc tggcccacat ggacgatctg	720
aagctctcct gggacctggc caaggtgcgc accgacctgc ccctggaggt ggacttcgcc	780
aaaaggcggg agcccgaccg ggagaggctt agggcctttc tggagaggct tgagcttggc	840
agcctcctcc acgagttcgg ccttctggaa agccccaagg ccctggagga ggcctcctgg	900
ccccgcggg aaggggcctt cgtgggcttt gtgcttacc gcaaggagcc catgtgggccc	960
gatcttctgg ccctggccgc cgccaggggg ggccgggtcc accgggcccc cgagccttat	1020
aaagccctca gggacctgaa ggaggcgcgg gggcttctcg ccaaagacct gagcgttctg	1080
gccctgaggg aaggccttgg cctcccgccc ggcgacgacc ccatgctcct cgcctacctc	1140
ctggaccctt ccaacaccac ccccgagggg gtggcccggc gctacggcgg ggagtggacg	1200
gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg	1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggagagacc cctttccgct	1320
gtcctggccc acatggaggc cacgggggtg cgcctggacg tggcctatct cagggccttg	1380
tccttgagg tggccgagga gatcgccgc ctcgaggccg aggtcttccg cctggccggc	1440
cacccttca acctcaactc ccgagaccag ctggaaaggg tcctctttga cgagctaggg	1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgct ccaccagcgc cgccgtcctg	1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc	1620
aagctgaaga gcacctacat tgaccttctg ccggacctca tccaccccag gacgggcccgc	1680
ctccacacc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc	1740
aacctccaga acatccccgt ccgcaccccg cttgggcaga ggatccgccc ggccttcate	1800
gccgaggagg ggtggctatt ggtggccctg gactatagcc agatagagct cagggtgctg	1860
gcccacctct ccggcgacga gaacctgatc cgggtcttcc aggaggggcg ggacatccac	1920
acggagaccg ccagctggat gttcggcgtc ccccgggagg ccgtggaccc cctgatgcgc	1980
cgggcggcca agaccatcaa ctteggggtc ctctacggca tgtcggccca ccgcctctcc	2040
caggagctag ccatccctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc	2100
ttccccaagg tgcgggcctg gattgagaag acctggagg agggcaggag gcgggggtac	2160
gtggagaccc tcttcggccg ccgcgcctac gtgccagacc tagaggcccc ggtgaagagc	2220
gtgcgggagg cggccgagcg catggccttc aacatgcccg tccagggcac cgccgccgac	2280
cttatgaagc tcgccatggt gaagctcttc cccgcctcc gggagatggg ggcccgcctg	2340
ctcctccagg tccacgacga gctcctcctg gaggcccccc aagcgcgggc cgaggagggtg	2400

gcggcctttgg ccaaggaggc catggagaag gcctatcccc tcgccgtacc cctggaggtg 2460
 aaggtgggga tcggggagga ctggctctcc gcccaaggagt ga 2502

<210> 56
 <211> 833
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 56

Met	Arg	Gly	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu	1	5	10	15
Val	Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	Phe	Ala	Leu	Lys	Gly	20	25	30	
Leu	Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Ala	Val	Tyr	Gly	Phe	Ala	35	40	45	
Lys	Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	50	55	60	
Val	Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Leu	Arg	His	Glu	Ala	Tyr	Glu	65	70	75	80
Ala	Tyr	Lys	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	85	90	95	
Leu	Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	100	105	110	
Glu	Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	115	120	125	
Lys	Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	130	135	140	
Asp	Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	145	150	155	160
Gly	His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	165	170	175	
Pro	Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	180	185	190	
Asn	Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	195	200	205	
Leu	Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	210	215	220	

Leu	Lys	Pro	Ala	Ile	Arg	Glu	Lys	Ile	Leu	Ala	His	Met	Asp	Asp	Leu
225					230					235					240
Lys	Leu	Ser	Trp	Asp	Leu	Ala	Lys	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu
				245					250					255	
Val	Asp	Phe	Ala	Lys	Arg	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala
			260					265					270		
Phe	Leu	Glu	Arg	Leu	Glu	Leu	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu
		275					280					285			
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Ser	Trp	Pro	Pro	Pro	Glu
	290					295					300				
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Thr	Arg	Lys	Glu	Pro	Met	Trp	Ala
305					310					315					320
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala
				325					330					335	
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu
			340					345					350		
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu
		355					360					365			
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser
	370					375					380				
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr
385					390					395					400
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn
				405					410					415	
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg
			420					425					430		
Glu	Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr
		435					440					445			
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val
	450					455					460				
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly
465					470					475					480
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe
			485						490					495	
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys
			500					505					510		
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro
		515					520					525			

Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser
530						535					540				
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg
545					550					555					560
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser
				565					570					575	
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly
			580					585					590		
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val
		595					600					605			
Ala	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser
	610					615					620				
Gly	Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His
625					630					635					640
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp
				645					650					655	
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr
			660					665					670		
Gly	Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu
		675					680					685			
Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val
	690					695					700				
Arg	Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Arg	Arg	Gly	Tyr
705					710					715					720
Val	Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Glu	Ala
				725					730					735	
Arg	Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn	Met
			740					745					750		
Pro	Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val	Lys
		755					760					765			
Leu	Phe	Pro	Arg	Leu	Arg	Glu	Met	Gly	Ala	Arg	Met	Leu	Leu	Gln	Val
	770					775					780				
His	Asp	Glu	Leu	Leu	Leu	Glu	Ala	Pro	Gln	Ala	Arg	Ala	Glu	Glu	Val
785					790					795					800
Ala	Ala	Leu	Ala	Lys	Glu	Ala	Met	Glu	Lys	Ala	Tyr	Pro	Leu	Ala	Val
				805					810					815	
Pro	Leu	Glu	Val	Lys	Val	Gly	Ile	Gly	Glu	Asp	Trp	Leu	Ser	Ala	Lys
			820					825					830		

Glu

<210> 57
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 57
 atggcgatgc ttccctcttt tgagcccaag ggccgcgtcc tcctggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgccctgaag ggccccacca cgagccgggg cgaaccggtg 120
 caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc cctcatctcc gccacaaggc ctacgaggcc 240
 tacagggcgg ggagggcccc gacccccgag gacttcccc ggagctcgc cctcatcaag 300
 gagctggtgg acctcctggg gtttaccgc ctcgaggtec ccggtacga ggcgacgac 360
 gttctcgcca cctgggcaa gaaggcgga aaggaggggt acgaggtgcg catcctcacc 420
 gccgaccgcg gcctatacca actcgtctat gaccgcgtcg ccgtcctcca ccccgagggc 480
 cacctcatca ccccgagtg gctttgggag aagtaaggcc tcaggccgga gcagtgggtg 540
 gacttcgcgc cctcgtggg ggacccctcc gacaacctcc ccggggtcaa gggcatcggg 600
 gagaagaccg cctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac 660
 ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagacctc 720
 aggtctcctc tggagctctc ccgggtgcgc accgacctcc cctggagggt ggacctcgcc 780
 caggggcggg agcccgaccg ggaggggctt agggcctttc tggagaggct tgagtttggc 840
 agcctcctcc acgagttcgg cttcttgaa agccccaagg cctggagga ggccccctgg 900
 ccccgccgg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc 960
 gatcttctgg cctggccgc cgccaggggt ggtcgagtec accgggcccc cgagccttat 1020
 aaagccctca gggacctgaa ggaggcgcgg gggcttctcg ccaaagacct gagcgttctg 1080
 gccctaaggg aaggccttgg cctcccgccc ggcgacgacc ccatgctcct cgcctacctc 1140
 ctggaccctt ccaacaccac ccccgagggg gtggccccgc gctacggcgg ggagtggacg 1200
 gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg 1260
 cttgaggggg aggagaggct cttttggctt taccgggagg tggagaggcc cttttccgct 1320
 gtcttgcccc acatggaggc cacgggggtg cgctggacg tggcctatct cagggccttg 1380

```

tccctggagg tggccgagga gatcgcccg ctcgaggccg aggtcttccg cctggccggc 1440
cacccttca acctcaactc ccgggaccag ctggaaatgg tgctctttga cgagcttagg 1500
cttcccgcct tggggaagac gcaaaagacg ggcaagcgt ccaccagcgc cgccgtcctg 1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgacctcttg tcggacctca tccaccccag gacgggcccgc 1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
aacctccaga acatccccgt ccgcaccccgc cttgggcaga ggatccgccg ggccttcac 1800
gccgaggagg ggtggctact ggtggtcctg gactatagcc agatagagct cagggtgctg 1860
gcccacctct ccggcgacga aaacctgatc agggctcttc aggagggggcg ggacatccac 1920
acggagaccg ccagctggat gttcggcgtc ccccgaggagg ccgtggaccc cctgatgcgc 1980
cgggcggcca agaccatcaa cttcgggggtc ctctacggca tgctggccca ccgcctctcc 2040
caggagctag ccatccctta ccaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttcccgaagg tgcgggcctg gattgagaag acctggagg agggcaggag gcgggggtac 2160
gtggagaccc tcttcggccg ccgccgtac gtgccagacc tagaggcccg ggtgaagagc 2220
gtgcgggagg cggccgagcg catggccttc aacatgcccg tccagggcac cgccgccgac 2280
ctcatgaagc tggctatggt gaagctcttc ccaggctgg aggaaatggg ggccaggatg 2340
ctccttcagg tccacgacga gctggtcctc gaggcccca aagagagggc ggaggccgtg 2400
gcccggtgg ccaaggaggt catggagggg gtgtatcccc tggccgtgcc cctggaggtg 2460
gaggtgggga taggggagga ctggctctcc gccaaaggag ga 2502

```

```

<210> 58
<211> 833
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 58
Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1          5          10          15

Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Pro
          20          25          30

Thr Thr Ser Arg Gly Glu Pro Val Gln Val Val Tyr Gly Phe Ala Lys
35          40          45

```

135

Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	Val
50						55					60				
Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Lys	Ala	Tyr	Glu	Ala
65					70					75					80
Tyr	Arg	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu
				85					90					95	
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu
			100					105					110		
Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys
		115					120					125			
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Gly
	130					135					140				
Leu	Tyr	Gln	Leu	Val	Tyr	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly
145					150					155					160
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro
				165					170					175	
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn
			180					185					190		
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu
		195					200					205			
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val
	210					215					220				
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu
225					230					235					240
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu
				245					250					255	
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	Ala
			260					265					270		
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu
		275					280					285			
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu
	290					295					300				
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala
305					310					315					320
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala
				325					330					335	
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu
			340					345					350		

Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	
		355					360					365				
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	
	370					375					380					
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	
385					390					395					400	
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	
				405					410					415		
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	
			420					425					430			
Glu	Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr	
		435					440					445				
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val	
	450					455					460					
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly	
465					470					475					480	
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Met	Val	Leu	Phe	
			485						490					495		
Asp	Glu	Leu	Arg	Leu	Pro	Ala	Leu	Gly	Lys	Thr	Gln	Lys	Thr	Gly	Lys	
			500					505					510			
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro	
		515					520					525				
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	
	530					535					540					
Thr	Tyr	Ile	Asp	Pro	Leu	Ser	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg	
545					550					555					560	
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	
				565					570					575		
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly	
			580					585					590			
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val	
		595					600					605				
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser	
	610					615					620					
Gly	Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His	
625					630					635					640	
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp	
				645					650					655		

Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
 660 665 670
 Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
 675 680 685
 Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
 690 695 700
 Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
 705 710 715 720
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
 725 730 735
 Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
 740 745 750
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765
 Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
 770 775 780
 His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
 785 790 795 800
 Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
 805 810 815
 Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Glu

<210> 59
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 59
 atggcgatgc ttccctctt tgagcccaag ggccgcgtcc tctggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgccctgaag ggccccacca cgagccgggg cgaaccggtg 120
 caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc cctcattcc gccacaaggc ctacgaggcc 240
 tacagggcgg ggagggcccc gacccccgag gacttcccc ggcagctcgc cctcatcaag 300
 gagctggtgg acctcctggg gtttaccgc ctcgaggtcc ccggctacga ggcggacgac 360

gttctcgcca ccttcgcca gaaggcgga aaggaggggt acgaggtgcg catcctcacc	420
gccgaccgcg gcctctacca actcgtctct gaccgcgtcg ccgtcctcca ccccgagggc	480
cacctcatca ccccgagagt gctttgggag aagtaaggcc tcaggccgga gcagtgggtg	540
gacttccgcg ccctcgtggg gaacccctcc gacaacctcc ccgggggtcaa gggcatcggg	600
gagaagaccg ccctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac	660
ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagacctc	720
aggctctcct tggagctctc ccgggtgcgc accgacctcc ccctggaggt ggacctcgcc	780
caggggcggg agcccgaccg ggaggggctt agggcctttc tggagaggct tgagtttggc	840
agcctcctcc acgagttcgg ccttctggaa agccccaagg ccctggagga ggccccctgg	900
ccccgcggg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc	960
gatcttcttg ccctggccgc cgcaggggtt ggtcgagtc accgggcccc cgagccttat	1020
aaagccctca gggacctgaa ggaggcgcg gggttctctg ccaaagacct gagcgttctg	1080
gccctaaggg aaggccttg cctcccgccc ggcgacgacc ccatgctcct cgcctacctc	1140
ctggaccctt ccaacaccac ccccgagggg gtggccccgc gctacggcgg ggagtggacg	1200
gaggaggcgg gggagcgggc cgcctttcc gagaggctct tcgccaacct gtgggggagg	1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggagaggcc cctttccgct	1320
gtcctggccc acatggaggc cacgggggtg cgcctggacg tggcctatct cagggccttg	1380
tccctggagg tggccgagga gatcgccgc ctcgaggccg aggtcttccg cctggccggc	1440
cacccttca acctcaactc ccgggaccag ctggaaaggg tcctctttga cgagctaggg	1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgt ccaccagcgc cgcgctcctg	1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc	1620
aagctgaaga gcacctacat tgacccttg ccggacctca tccaccccag gacgggccgc	1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc	1740
aacctccaga acatccccgt ccgcaccccg ctcgggcaga ggatccgccg ggccttcac	1800
gccgaggagg ggtggctatt ggtggctcctg gactatagcc agatagagct cagggtgctg	1860
gccacctct ccggcgacga gaacctgatc cgggtcttcc aggaggggcg ggacatccac	1920
acggaaaccg ccagctggat gttcggcgtc ccccgggagg ccgtggaccc cctaatacgc	1980
cgggcggcca agaccatcaa cttcggggtc ctctacggca tgcgggccg ccgcctctcc	2040

caggagctag ccattccctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
 ttccccaagg tgcgggcctg gattgagaag accctggagg agggcaggag gcgggggtac 2160
 gtggagaccc tcttcggccg ccgccgctac gtgccagacc tagaggcccg ggtgaagagc 2220
 gtgcgggagg cggccgagcg catggccttc aacatgcccg tccagggcac cgccgccgac 2280
 ctcatgaagc tggctatggt gaagctcttc cccaggctgg aggaaatggg ggccaggatg 2340
 ctcttcagg tccacgacga gctggctctc gagggcccaa aagagagggc ggaggccgtg 2400
 gcccggtgg ccaaggaggt catggagggg gtgtatcccc tggccgtgcc cctggaggtg 2460
 gaggtgggga taggggagga ctggctttcc gcccaagggtt ag 2502

<210> 60
 <211> 833
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 60
 Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
 1 5 10 15
 Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Pro
 20 25 30
 Thr Thr Ser Arg Gly Glu Pro Val Gln Val Val Tyr Gly Phe Ala Lys
 35 40 45
 Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe Val
 50 55 60
 Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Lys Ala Tyr Glu Ala
 65 70 75 80
 Tyr Arg Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
 85 90 95
 Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu Glu
 100 105 110
 Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Phe Ala Lys Lys
 115 120 125
 Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg Gly
 130 135 140
 Leu Tyr Gln Leu Val Ser Asp Arg Val Ala Val Leu His Pro Glu Gly
 145 150 155 160
 His Leu Ile Thr Pro Glu Trp Leu Trp Glu Lys Tyr Gly Leu Arg Pro

140

165							170					175			
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asn	Pro	Ser	Asp	Asn
			180					185					190		
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu
		195					200					205			
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val
	210					215					220				
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu
225					230					235					240
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu
				245					250					255	
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	Ala
			260					265					270		
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu
		275					280					285			
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu
	290					295					300				
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala
305					310					315					320
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala
				325					330					335	
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu
			340					345					350		
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu
		355					360					365			
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser
	370					375					380				
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr
385					390					395					400
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn
				405					410					415	
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg
			420					425					430		
Glu	Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr
			435				440					445			
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val
	450					455					460				
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly

465					470						475				480
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe
				485					490					495	
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys
			500					505					510		
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro
		515					520					525			
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser
	530					535					540				
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg
545					550					555					560
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser
				565					570					575	
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly
			580					585					590		
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val
		595					600					605			
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser
	610					615					620				
Gly	Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His
625				630						635					640
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp
				645					650					655	
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr
			660					665					670		
Gly	Met	Ser	Ala	Arg	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu
		675					680					685			
Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val
	690					695					700				
Arg	Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Arg	Arg	Gly	Tyr
705					710					715					720
Val	Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Glu	Ala
				725					730					735	
Arg	Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn	Met
			740					745					750		
Pro	Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val	Lys
		755					760					765			
Leu	Phe	Pro	Arg	Leu	Glu	Glu	Met	Gly	Ala	Arg	Met	Leu	Leu	Gln	Val

770		775		780
His Asp Glu Leu Val	Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val			
785		790		795
				800
Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val				
		805		810
				815
Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys				
		820		825
				830

Gly

<210> 61
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 61
 atggtgatgc ttccctcttt tgagcccaag ggccgcgtcc tcttgggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgcctgaag ggcctcacca cgagccgggg cgaaccgggtg 120
 caggcgggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc tcttccttcc gccacgaggc ctacgaggcc 240
 tacaaggcgg ggagggcccc gacccccgag gacttcccc ggagctcgc cctcatcaag 300
 gagctgggtg acctcctggg gtttaccgc ctcgagggtc ccggctacga ggtggacgac 360
 gtcctggcca gcctggccaa gaaggtggaa aaggaggggt acgaggtgcg catcctcacc 420
 gccgaccgcg acctctacca actcgtctcc gaccgcgtcg ccgtcctcca ccccgagggc 480
 cacctcatca ccccgagtg gctttgggag aagtacggcc tcaggccgga gcagtgggtg 540
 gacttccgcg ccctcgtggg ggacccctcc gacaacctcc ccgggggtcaa gggcatcggg 600
 gagaagaccg ccctcaagct cctcaaggag tggggaggcc tggaaaacct cctcaagaac 660
 ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagacctc 720
 aggctctcct tggagctctc ccgggtgctc accgacctcc ccctggaggt ggacctcgcc 780
 caggggcggg aaccgaccg ggagaggctt agggcctttc tggagaggct tgagtttggc 840
 agcctcctcc acgagttcgg ccttctggaa agccccaagg ccctggagga ggccccctgg 900
 ccccgccgg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc 960
 gatcttctgg ccctggccgc cgccaggggt ggtcgggtcc accggacccc cgagccttat 1020

```

aaagccctca gggacttgaa ggaggcgcgg gggcttctcg ccaaagacct gagcggtctg 1080
gccctaaggg aaggccttgg cctcccgccc ggcgacgacc ccattgctcct cgcctacctc 1140
ctggaccctt ccaacaccac ccccgagggg gtggcccggc gctacggcgg ggagtggacg 1200
gaggaggcgg gggagcgggc cgccttttcc gagaggctct tcgccaacct gtgggggagg 1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggataggcc cctttccgct 1320
gtcctggccc acatggaggc cacaggggtg cgcctggacg tggcctacct cagggccttg 1380
tccctggagg tggccgagga gatcgcccg ctcgaggccg aggtcttccg cctggccggc 1440
cacccttca acctcaactc ccgggaccag ctggaaaggg tcctctttga cgagctaggg 1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgct ccaccagcgc cgcgctcctg 1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgacctcttg ccggacctca tccaccccag gacgggcccgc 1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatecc 1740
aacctccaga acatccccgt ccgcaccccg ctcgggcaga ggatccgccc ggccttcac 1800
gccgaggagg ggtggctatt ggtggctctg gactatagcc agatagagct cagggtgctg 1860
gccacctct ccggcgacga gaacctgatc cgggtcttcc aggagggggc ggacatccac 1920
acggaaaccg ccagctggat gttcggcgtc ccccgggagg ccgtggacct cctaattgcgc 1980
cgggcggcca agaccatcaa cttcgggggtt ctctacggca tgcgggcca ccgcctctcc 2040
caggagctag ccattcctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttccccaagg tgcgggcctg gattgagaag accctggagg agggcaggag gcgggggtac 2160
gtggagacct tcttcggccg ccgtcgctac gtgccagacc tagaggcccg ggtgaagagc 2220
gtgcgggagg cggccgagcg catggccttc aacatgcccg tccagggcac cgcgcccgc 2280
ctcatgaagc tggctatggt gaagctcttc ccaggctgg aagaaacggg ggccaggatg 2340
ctccttcagg tccacgacga gctggctctc gaggcccaaa aagagagggc ggaggccgtg 2400
gcccggtgg ccaaggaggc catggagggg gtgtatcccc tggccgtgcc cctggagggtg 2460
gaggtgggga taggggagga ctggctctcc gccaggagt ga 2502

```

```

<210> 62
<211> 833
<212> PRT
<213> Artificial Sequence

<220>

```

<223> cloned sequence

<400> 62

Met	Val	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu	Val
1				5					10					15	
Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	Phe	Ala	Leu	Lys	Gly	Leu
			20					25					30		
Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Ala	Val	Tyr	Gly	Phe	Ala	Lys
		35					40					45			
Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	Val
	50					55					60				
Val	Phe	Asp	Ala	Lys	Ala	Ser	Ser	Phe	Arg	His	Glu	Ala	Tyr	Glu	Ala
65					70					75					80
Tyr	Lys	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu
				85					90					95	
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu
			100					105					110		
Val	Pro	Gly	Tyr	Glu	Val	Asp	Asp	Val	Leu	Ala	Ser	Leu	Ala	Lys	Lys
		115					120					125			
Val	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Asp
	130					135					140				
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly
145					150					155					160
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro
				165					170					175	
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn
			180					185					190		
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu
		195					200					205			
Lys	Glu	Trp	Gly	Gly	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val
	210					215					220				
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu
225					230					235					240
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu
				245					250					255	
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala
			260					265					270		
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu
		275					280					285			

Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu
290						295					300				
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala
305					310					315					320
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Thr
				325					330					335	
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu
			340					345					350		
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu
		355					360					365			
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser
	370					375					380				
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr
385					390					395					400
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn
				405					410					415	
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg
			420					425					430		
Glu	Val	Asp	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr
		435					440					445			
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val
	450					455					460				
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly
465					470					475					480
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe
			485						490					495	
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys
			500					505					510		
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro
		515					520					525			
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser
	530					535					540				
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg
545					550					555					560
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser
				565					570					575	
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly
			580					585					590		

Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
 595 600 605
 Val Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
 610 615 620
 Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
 625 630 635 640
 Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
 645 650 655
 Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
 660 665 670
 Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
 675 680 685
 Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
 690 695 700
 Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
 705 710 715 720
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
 725 730 735
 Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
 740 745 750
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765
 Leu Phe Pro Arg Leu Glu Glu Thr Gly Ala Arg Met Leu Leu Gln Val
 770 775 780
 His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
 785 790 795 800
 Ala Arg Leu Ala Lys Glu Ala Met Glu Gly Val Tyr Pro Leu Ala Val
 805 810 815
 Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Glu

<210> 63
 <211> 2550
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 63

atggtgatgc ttccctctt tgagcccaag ggccgcgtcc tcctggtgga cggccaccac	60
ctggcctacc gcaccttctt cgcctgaag ggcctcacca cgagccgggg cgaaccggtg	120
caggcggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag	180
gccgtcttcg tggctcttga cgccaaggcc tcctccttcc gccacgaggc ctacgaggcc	240
tacaaggcgg ggagggcccc gacccccgag gacttcccc gccagctcgc cctcatcaag	300
gagctggtgg acctcctggg gtttaccgc ctcgaggtcc ccggctacga ggtggacgac	360
gtcctggcca gcctggccaa gaaggtggaa aaggaggggt acgaggtgcg cctcctcacc	420
gccgaccgcg gcctctacca actcgtctct gaccgcgtcg ccgtcctcca ccccgagggc	480
cacctcatca ccccgagtg gctttgggag aagtacggcc tcaggccgga gcagtgggtg	540
gacttcgcg cctcgtggg ggacccctcc gacaacctcc ccgggggtcaa gggcatcggg	600
gagaagaccg cctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac	660
ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagacctc	720
aggctctcct tggagctctc ccgggtgcgc accgacctcc cctggaggt ggacctgcc	780
caggggcggg agcccgaccg ggagaggctt agggccttcc tggagaggct tgagtttggc	840
agcctcctcc acgagttcgg ccttctggaa agccccaagg cctggagga ggccccctgg	900
ccccgcggg aaggggcctt cgtgggcttt gtgcttccc gcaaggagcc catgtgggcc	960
gatcttctgg cctggccgc cgccaggggt ggtcgggtcc accgggcccc cgagccttat	1020
aaagccctca gggacttgaa ggaggcgcgg gggcttctcg ccaaagacct gacggtctg	1080
gccctaaggg aaggccttgg cctcccgccc ggcgacgacc ccatgctcct cgcctacctc	1140
ctggaccctt ccaacaccac ccccgagggg gtggccccgc gctacggcgg ggagtggacg	1200
gaggaggcgg gggagcgggc cgcccttccc gagaggctct tcgccaacct gtgggggagg	1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggataggcc cctttccgct	1320
gtcctggccc acatggaggc cacaggggtg cgctggacg tggcctatct cagggccttg	1380
tcctggagg tggccgagga gatcgcgcgc ctcgaggccg aggtcttccg cctggccggc	1440
caccttca acctcaactc ccgggaccag ctggaaaggg tcctctttga cgagctaggg	1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgt ccaccagcgc cgccatcctg	1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc	1620
aagctgaaga gcacctacat tgaccttgg ccggacctca tccaccccag gacgggcgc	1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc	1740

```

aacctccaga acatccccgt ccgcacccccg ctcgggcaga ggatccgccg ggccttcac 1800
gccgaggagg ggtggctatt ggtggctctg gactatagcc agatagagct cagggtgctg 1860
gccacacctt ccggcgacga gaacctgacc cgggtcttcc aggagggggc ggacatccac 1920
acggaaaccg ccagctggat gttcggcgtc ccccgggagg ccgtggaccc cctgatgcgc 1980
cgggcggcca agaccatcaa cttcgggggt ctctacggca tgcggccca ccgcctctcc 2040
caggagctgg ccatccctta cgaggaggcc caggccttca tagagcgcta cttccaaagc 2100
ttccccaagg tgcgggcctg gatagaaaag accctggagg aggggaggaa gcggggctac 2160
gtggaaaccc tcttcggaag aaggcgctac gtgcccgacc tcaacgcccg ggtgaagagt 2220
gtcagggagg ccgcgagcgc catggccttc aacatgcccg tccagggcac cgccgccgac 2280
cttatgaagc tcgccatggt gaagctcttc cccgcctcc gggagatggg ggcccgcatg 2340
ctcctccagg tccacgacga gctcctctg gagggccccc aagcgcgggc cgaggaggtg 2400
gcggctttgg ccaaggaggc catggagaag gcctatcccc tcgccgtacc cctggaggtg 2460
aagggtggga tcggggagga ctggctctcc gcccaaggag tgagtcgacc tgcaggcagc 2520
gcttggcgtc acccgagtt cgggtggttaa 2550

```

```

<210> 64
<211> 849
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 64
Met Val Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1 5 10 15
Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Leu
20 25 30
Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala Lys
35 40 45
Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe Val
50 55 60
Val Phe Asp Ala Lys Ala Ser Ser Phe Arg His Glu Ala Tyr Glu Ala
65 70 75 80
Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
85 90 95

```

Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu	
			100					105					110			
Val	Pro	Gly	Tyr	Glu	Val	Asp	Asp	Val	Leu	Ala	Ser	Leu	Ala	Lys	Lys	
		115					120					125				
Val	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Gly	
	130					135					140					
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly	
145					150					155					160	
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	
				165					170					175		
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn	
			180					185					190			
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu	
		195					200					205				
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val	
	210					215					220					
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu	
225					230					235					240	
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu	
				245					250					255		
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala	
			260					265					270			
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	
		275					280					285				
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	
	290					295					300					
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	
305					310					315					320	
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	
				325					330					335		
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	
			340					345					350			
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	
		355					360					365				
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	
	370					375					380					
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	
385					390					395					400	

Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	
			405						410					415		
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	
			420					425					430			
Glu	Val	Asp	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr	
		435					440					445				
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val	
	450					455					460					
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly	
465					470					475					480	
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe	
			485					490						495		
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys	
			500					505					510			
Arg	Ser	Thr	Ser	Ala	Ala	Ile	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro	
		515					520					525				
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	
	530					535					540					
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg	
545					550					555					560	
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	
				565				570						575		
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly	
			580					585					590			
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val	
		595					600					605				
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser	
	610					615					620					
Gly	Asp	Glu	Asn	Leu	Thr	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His	
625					630					635				640		
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp	
				645				650						655		
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr	
			660					665					670			
Gly	Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu	
		675					680					685				
Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val	
	690					695					700					

Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly Tyr
 705 710 715 720
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala
 725 730 735
 Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
 740 745 750
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765
 Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln Val
 770 775 780
 His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu Val
 785 790 795 800
 Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala Val
 805 810 815
 Pro Leu Glu Val Lys Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Gln
 820 825 830
 Gly Val Ser Arg Pro Ala Gly Ser Ala Trp Arg His Pro Gln Phe Gly
 835 840 845

Gly

<210> 65
 <211> 2505
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 65
 atgcgtggta tgcttcctct ttttgagccc aagggccgcg tcctcctggt ggacggccac 60
 cacctggcct accgcacctt ctctgccttg aagggcccca ccacgagccg gggcgaaccg 120
 gtgcaggcgg tctacggctt cgccaagagc ctctcaagg ccctgaagga ggacgggtac 180
 aagggcgctt tcgtggtctt tgacgccaag gccccctcct tccgccacga ggcctacgag 240
 gcctacaagg cggggagggc cccgaccccc gaggacttcc cccggcagct cgccctcatc 300
 aaggagctgg tggacctcct ggggtttacc cgctcgagg tccttggtta cgaggcggac 360
 gacgtcctcg ccaccctggc caagaaggcg gaaaaggagg ggtacgaggt gcgcatactc 420
 accgccgacc gcgacctcta ccaactcgtc tccgaccgcg tcgccgtcct ccaccccgag 480
 ggccacctca tcaccccgga gtggctttgg gagaagtacg gcctcaggcc ggagcagtgg 540

gtggacttcc ggccectcgt gggggacccc tccgacaacc tccccggggt caagggcatc	600
ggggagaaga ccgcectcaa gctcctcaag gagtggggaa gcctggaaaa cctcctcaag	660
aacctggacc gggtaaagcc agaaaacgtc cgggagaaga tcaaggccca cctggaagac	720
ctcaggctct ccttggagct ctcccgggtg cgcaccgacc tccccctgga ggtggacctc	780
gcccaggggc gggagctcga ccgggagagg cttagggcct ttctggagag gcttgagttt	840
ggcggcctcc tccacgagtt cggccttctg gaaagcccca aggccctgga ggaggccccc	900
tggccccgc cggaaggggc cttcgtgggc tttgtgcttt cccgcaagga gcccatgtgg	960
gccgatcttc tggccctggc cgcgcgcagg ggtggtcggg tccaccgggc ccccgagcct	1020
tataaagccc tcagggactt gaaggaggcg cgggggcttc tcgccaaaga cctgagcgtt	1080
ctggccctaa gggaaggcct tggcctcccg cccggcgacg accccatgct cctcgcctac	1140
ctcctggacc cttccaacac cgcccccgag ggggtggccc ggcgctacgg cggggagtg	1200
acggaggagg cgggggagcg ggccgcctt tccgagagge tcttcgcca cctgtggggg	1260
aggcttgagg gggaggagag gctcctttgg ctttaccggg aggtggatag gccctttcc	1320
gctgtcctgg ccacatgga ggccacaggg gtacggctgg acgtggcctg cctgcaggcc	1380
ctttccctgg agcttgcgga ggagatccgc cgcctcgagg aggaggtctt ccgcttggcg	1440
ggccaccct tcaacctcaa ctcccgggac cagctggaaa gggtcctctt tgacgagcta	1500
gggcttcccg ccatcggcaa gacggagaag accggcaagc gctccaccag cgccgccatc	1560
ctggaggccc tccgcgaggc ccaccccatc gtggagaaga tcctgcagta ccgggagctc	1620
accaagctga agagcaccta cattgacccc ttgcgggacc tcatccaccc caggacgggc	1680
cgcctccaca cccgcttcaa ccagacggcc acggccacgg gcaggctaag tagctccgat	1740
cccaacctcc agaacatccc cgtccgcacc ccgctcgggc agaggatccg ccgggccttc	1800
gtcgccgagg aggggtggct attggtggtc ctggactata gccagataga gctcagggtg	1860
ctggcccacc tctccggcga cgagaacctg acccgggtct tcctggaggg gcgggacatc	1920
cacacggaaa ccgccagctg gatgttcggc gtcccccggg aggccgtgga cccctgatg	1980
cgccgggchg ccaagaccat caacttcggg gttctctacg gcatgtcggc ccaccgcctc	2040
tcccaggagc tggccatccc ttacgaggag gcccaggcct tcatagagcg ctacttccaa	2100
agcttcccca aggtgcgggc ctggatagaa aagaccctgg aggaggggag gaagcggggc	2160
tacgtggaaa ccctcttcgg aagaaggcg tacgtgcccg acctcaacgc ccgggtgaag	2220
agtgtcaggg aggccgcgga gcgcatggcc ttcaacatgc ccgtccaggg caccgccgcc	2280

gaccttatga agctcgccat ggtgaagctc ttcccccgcc tccgggagat gggggcccgc 2340
 atgctcctcc aggtccacga cgagctcctc ctggaggccc cccaagcgcg ggccgaggag 2400
 gtggcgggctt tggccaagga ggccatggag aaggcctatc ccctcgccgt acccctggag 2460
 gtgaagggtgg ggatcgggga ggactggctc tccgccaagg agtga 2505

<210> 66
 <211> 834
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 66
 Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu
 1 5 10 15
 Val Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly
 20 25 30
 Pro Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala
 35 40 45
 Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Ala Phe
 50 55 60
 Val Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu
 65 70 75 80
 Ala Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln
 85 90 95
 Leu Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu
 100 105 110
 Glu Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys
 115 120 125
 Lys Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg
 130 135 140
 Asp Leu Tyr Gln Leu Val Ser Asp Arg Val Ala Val Leu His Pro Glu
 145 150 155 160
 Gly His Leu Ile Thr Pro Glu Trp Leu Trp Glu Lys Tyr Gly Leu Arg
 165 170 175
 Pro Glu Gln Trp Val Asp Phe Arg Ala Leu Val Gly Asp Pro Ser Asp
 180 185 190
 Asn Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Leu Lys Leu
 195 200 205

Leu	Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	210	215	220	
Val	Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	225	230	235	240
Leu	Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	245	250	255	
Glu	Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Leu	Asp	Arg	Glu	Arg	Leu	Arg	260	265	270	
Ala	Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Gly	Leu	Leu	His	Glu	Phe	Gly	275	280	285	
Leu	Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	290	295	300	
Glu	Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	305	310	315	320
Ala	Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	325	330	335	
Ala	Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	340	345	350	
Leu	Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	355	360	365	
Leu	Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	370	375	380	
Ser	Asn	Thr	Ala	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	385	390	395	400
Thr	Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	405	410	415	
Asn	Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	420	425	430	
Arg	Glu	Val	Asp	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	435	440	445	
Thr	Gly	Val	Arg	Leu	Asp	Val	Ala	Cys	Leu	Gln	Ala	Leu	Ser	Leu	Glu	450	455	460	
Leu	Ala	Glu	Glu	Ile	Arg	Arg	Leu	Glu	Glu	Glu	Val	Phe	Arg	Leu	Ala	465	470	475	480
Gly	His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	485	490	495	
Phe	Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	500	505	510	

Lys	Arg	Ser	Thr	Ser	Ala	Ala	Ile	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His
		515					520					525			
Pro	Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys
	530					535					540				
Ser	Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly
545					550					555					560
Arg	Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu
				565					570					575	
Ser	Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu
			580					585					590		
Gly	Gln	Arg	Ile	Arg	Arg	Ala	Phe	Val	Ala	Glu	Glu	Gly	Trp	Leu	Leu
		595					600					605			
Val	Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu
	610					615					620				
Ser	Gly	Asp	Glu	Asn	Leu	Thr	Arg	Val	Phe	Leu	Glu	Gly	Arg	Asp	Ile
625					630					635					640
His	Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val
				645					650					655	
Asp	Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu
			660					665					670		
Tyr	Gly	Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr
		675					680					685			
Glu	Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys
	690					695					700				
Val	Arg	Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Lys	Arg	Gly
705					710					715					720
Tyr	Val	Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Asn
				725					730					735	
Ala	Arg	Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn
			740					745					750		
Met	Pro	Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val
		755					760					765			
Lys	Leu	Phe	Pro	Arg	Leu	Arg	Glu	Met	Gly	Ala	Arg	Met	Leu	Leu	Gln
	770					775					780				
Val	His	Asp	Glu	Leu	Leu	Leu	Glu	Ala	Pro	Gln	Ala	Arg	Ala	Glu	Glu
785					790					795					800
Val	Ala	Ala	Leu	Ala	Lys	Glu	Ala	Met	Glu	Lys	Ala	Tyr	Pro	Leu	Ala
			805						810					815	

Val Pro Leu Glu Val Lys Val Gly Ile Gly Glu Asp Trp Leu Ser Ala
 820 825 830

Lys Glu

<210> 67
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 67
 atggcgatgc ttccctcttt tgagcccaag ggccgcgtcc tcctggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgccctgaag ggccccacca cgagccgggg cgaaccggtg 120
 caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc ccctcattec gccacaaggc ctacgaggcc 240
 tacagggcgg ggagggcccc gacccccgag gacttcccc ggcagctcgc cctcatcaag 300
 gagctggtgg acctcctggg gtttaccgcg ctcgaggtec ccggctacga ggcggaacgac 360
 gttctcgcca ccctggccaa gaaggcggaa aaggaggggt acgaggtgcg catcctcacc 420
 gccgaccgcg gcctctacca actcgtctct gaccgcgtcg ccgtcctcca ccccgagggc 480
 cacctcatca ccccgaggtg gctttgggag aagtacggcc tcaggccgga gcagtgggtg 540
 gacttccgcg ccctcgtggg ggacccctcc gacaacctcc ccgggggtcaa gggcatcggg 600
 gagaagaccg ccctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac 660
 ctggaccggg taaagccaga aaacgtccgg gagaagatca aggcccacct ggaagacctc 720
 aggctctcct tggagctctc ccgggtgcgc accgacctcc ccctggaggt ggacctcgcc 780
 cagggggcgg agcccgaccg ggaggggctt agggcctttc tggagaggct tgagtttggc 840
 agcctcctcc acgagttcgg ctttctggaa agccccaagg ccctggagga ggccccctgg 900
 ccccgcccgg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc 960
 gatcttctgg ccctggcgcg cgccaggggt ggtcgagtcc accgggcccc cgagccttat 1020
 aaagccctca gggacctgaa ggaggcgcgg gggcttctcg ccaaagacct gagcgttctg 1080
 gccctaaggg aaggccttgg cctcccggcc ggcgacgacc ccatgctcct cgcctacctc 1140
 ctggaccctt ccaacaccac ccccgagggg gtggccccgc gctacggcgg ggagtggacg 1200
 gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg 1260

```

cttgagggggg aggagagggt cctttgggtt taccgggagg tggagaggcc cctttccgct 1320
gtcctggccc acatggaggc cacgggggtg cgcctggacg tggcctatct cagggccttg 1380
tccctggagg tggccgagga gatcgcccg ctcgaggccg aggtcttccg cctggccggc 1440
cacccttca acctcaactc ccgggaccag ctggaaatgg tgctctttga cgagcttagg 1500
cttcccgctt tggggaagac gcaaaagacg ggcaagcgct ccaccagcgc cgccgtcctg 1560
gaggccctcc gcgaggccca ccccatcggtg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgacccttg tcggacctca tccaccccag gacgggccgc 1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
aacctccaga acatccccgt ccgcaccccg cttgggcaga ggatccgccc ggccttcac 1800
gccgaggagg ggtggctact ggtggctcctg gactatagcc agatagagct cagggtgctg 1860
gccacctct ccggcgacga aaacctgatc agggctcttc aggaggggcg ggacatccac 1920
acggagaccg ccagctggat gttcggcgtc ccccgaggagg ccgtggacce cctgatgcgc 1980
cgggcggcca agaccatcaa cttcggggtc ctctacggca tgtcggccca ccgcctctcc 2040
caggagctag ccacccctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttccccaagg tgccggcctg gattgagaag accctggagg agggcaggag gcgggggtac 2160
gtggagacce tcttcggccg ccgcgcctac gtgccagacc tagaggcccg ggtgaagagc 2220
gtgcgggagg cggccgagcg catggccttc aacatgcccg tccagggcac cgccgccgac 2280
ctcatgaagc tggctatggt gaagctcttc ccaggctgg aggaaacggg ggccaggatg 2340
ctccttcagg tccacgacga gctggctcctt gaggccccaa aagagagggc ggaggccgtg 2400
gcccggtgg ccaaggaggt catggagggg gtgtatcccc tggccgtgtc cctggagggtg 2460
gaggtgggga taggggagga ctggctctcc gccaaaggagt ga 2502

```

```

<210> 68
<211> 833
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 68
Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1           5           10           15

```

```

Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Pro

```

20					25					30					
Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Val	Val	Tyr	Gly	Phe	Ala	Lys
		35					40					45			
Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	Val
	50					55					60				
Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Lys	Ala	Tyr	Glu	Ala
65					70					75					80
Tyr	Arg	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu
				85					90					95	
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu
			100					105					110		
Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys
		115					120					125			
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Gly
	130					135					140				
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly
145					150					155					160
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro
				165					170					175	
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn
			180					185					190		
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu
		195					200					205			
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val
	210					215					220				
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu
225					230					235					240
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu
				245					250					255	
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	Ala
			260					265					270		
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu
		275					280					285			
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu
	290					295					300				
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala
305					310					315					320
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala

				325					330					335			
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu		
			340					345					350				
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu		
		355					360					365					
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser		
	370					375					380						
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr		
385					390					395					400		
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn		
				405					410					415			
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg		
			420					425					430				
Glu	Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr		
		435					440					445					
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val		
	450					455					460						
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly		
465					470					475					480		
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Met	Val	Leu	Phe		
			485					490					495				
Asp	Glu	Leu	Arg	Leu	Pro	Ala	Leu	Gly	Lys	Thr	Gln	Lys	Thr	Gly	Lys		
			500					505					510				
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro		
		515					520					525					
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser		
	530					535					540						
Thr	Tyr	Ile	Asp	Pro	Leu	Ser	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg		
545					550				555						560		
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser		
				565					570					575			
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly		
			580					585					590				
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val		
		595					600					605					
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser		
	610					615					620						
Gly	Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His		

160

625		630		635		640									
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp
			645						650					655	
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr
		660						665					670		
Gly	Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu
	675						680					685			
Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val
	690					695					700				
Arg	Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Arg	Arg	Gly	Tyr
705					710					715					720
Val	Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Glu	Ala
			725						730					735	
Arg	Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn	Met
		740						745					750		
Pro	Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val	Lys
	755						760					765			
Leu	Phe	Pro	Arg	Leu	Glu	Glu	Thr	Gly	Ala	Arg	Met	Leu	Leu	Gln	Val
	770					775				780					
His	Asp	Glu	Leu	Val	Leu	Glu	Ala	Pro	Lys	Glu	Arg	Ala	Glu	Ala	Val
785					790					795					800
Ala	Arg	Leu	Ala	Lys	Glu	Val	Met	Glu	Gly	Val	Tyr	Pro	Leu	Ala	Val
			805						810					815	
Ser	Leu	Glu	Val	Glu	Val	Gly	Ile	Gly	Glu	Asp	Trp	Leu	Ser	Ala	Lys
		820						825					830		

Glu

<210> 69
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 69
 atggc gatgc ttccctctt tgagcccaag ggccgcgtcc tcttggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgccctgaag ggccccaccg cgagccgggg cgaaccggtg 120
 caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tgggtcttga cgccaaggcc ccttcattcc gccacaaggc ctacgaggcc 240

tacagggcgg ggagggcccc gacccccgag gacttcccc	ggcagctcgc cctcatcaag	300
gagctggtgg acctcctggg gtttaccgc ctcgaggtcc	ccggctacga ggcggacgac	360
gttctcgccc cctggccaa gaaggcggaa aaggaggggt	tcgaggtgcg catcctcccc	420
gccgtccgcg gcctctgccc tctcgtctct gaccgcgtcg	ccgtcctcct ccccgagggc	480
cacctcatca ccccgagtg gctttgggag aagtacggcc	tcaggccgga gcagtgggtg	540
gacttccgcg cctcgtggg ggacccctcc gacaacctcc	ccgggggtcaa gggcatcggg	600
aagaagaccg cctcaagct cctcaaggag tggggaagcc	tggaaaacct cctcaagaac	660
ctggaccggg taaagccaga aaacgtccgg gagaagatca	aggcccacct ggaagacctc	720
aggctctcct tggagctctc ccgggtgcgc accgacctcc	cctggaggt ggacctcgcc	780
caggggcggg agcccgaccg ggaggggctt agggccttcc	tggagaggct tgagtttggc	840
agcctcctcc acgagttcgg ccttctggaa agccccaagg	cctggagga ggccccctgg	900
ccccgcggg aaggggcctt cgtgggcttt gtgctttccc	gcaaggagcc catgtgggcc	960
gatcttctgg cctggccgc cgcaggggt ggtcgggtcc	accgggcccc cgagccttat	1020
aaagccctca gggacttgaa ggaggcgcgg gggtctctcg	ccaaagacct gagcgttctg	1080
gccetaaggg aaggccttgg cctcccgccc ggcgacgacc	ccatgctcct cgcctacctc	1140
ctggaccctt ccaacaccac ccccgagggg gtggcccggc	gctacggcgg ggagtggacg	1200
gaggaggcgg gggagcgggc cgcctttcc gagaggctct	tcgccaacct gtgggggagg	1260
cttgaggggg aggagaggct cctgtggctt taccgggagg	tggataggcc cctttccgct	1320
gtcctggccc acatggaggc cacaggggta cggctggacg	tggcctgcct gcaggccctt	1380
tccctggagc ttgcggagga gatccgcgc ctcgaggagg	aggtcttccg cttggcgggc	1440
cacccttca acctcaactc ccgggaccag ctggaaaggg	tcctctttga cgagctaggg	1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgct	ccaccagcgc cgccatcctg	1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc	tgcagtaccg ggagctcacc	1620
aagctgaaga gcacctacat tgaccccttg ccggacctca	tccaccccag gacgggcccgc	1680
ctccacaccc gcttcaacca gacggccacg gccacgggca	ggctaagtag ctccgatecc	1740
aacctccaga acatccccgt ccgcaccccg ctcgggcaga	ggatccgcgg ggccttcac	1800
gccgaggagg ggtggctatt ggtggtcctg gactatagcc	agatagagct cagggtgctg	1860
gccacctct ccggcgacga gaacctgacc cgggtcttcc	aggaggggcg ggacatccac	1920

acggaaaccg ccagctggat gttcggcgtc ccccgggagg ccgtggaccc cctgatgcgc 1980
 cgggcggcca agaccatcaa cttcgggggtt ctctacggca tgcgggcca ccgcctctcc 2040
 caggagctgg ccatacctta cgaggaggcc caggccttca tagagcgcta cttccaaagc 2100
 ttccccaagg tgcgggcctg gatagaaaag accctggagg aggggaggaa gcggggctac 2160
 gtggaaaccc tcttcggaag aaggcgctac gtgcccgacc tcaacgcccg ggtgaagagt 2220
 gtcaggagg cgcgggagcg catggccttc aacatgcccg tccagggcac cgccgccgac 2280
 cttatgaagc tcgcatggt gaagctcttc cccgcctcc gggagatggg ggcccgcacg 2340
 ctctccagg tccacgacga gctcctcctg gagggccccc aagcgcgggc cgaggaggtg 2400
 gcggctttgg ccaaggaggc catggagaag gcctatcccc tcgccgtacc cctggaggtg 2460
 aaggtgggga tcggggagga ctggctctcc gccaaagagt ga 2502

<210> 70
 <211> 833
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 70
 Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
 1 5 10 15
 Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Pro
 20 25 30
 Thr Ala Ser Arg Gly Glu Pro Val Gln Val Val Tyr Gly Phe Ala Lys
 35 40 45
 Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe Val
 50 55 60
 Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Lys Ala Tyr Glu Ala
 65 70 75 80
 Tyr Arg Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
 85 90 95
 Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu Glu
 100 105 110
 Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Pro Leu Ala Lys Lys
 115 120 125
 Ala Glu Lys Glu Gly Phe Glu Val Arg Ile Leu Pro Ala Val Arg Gly
 130 135 140

Leu	Cys	Pro	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	Leu	Pro	Glu	Gly	
145					150					155					160	
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	
				165					170					175		
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn	
			180					185					190			
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Lys	Lys	Thr	Ala	Leu	Lys	Leu	Leu	
		195					200					205				
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val	
	210					215					220					
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu	
225					230					235					240	
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu	
				245					250					255		
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	Ala	
			260					265					270			
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	
		275					280					285				
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	
	290					295					300					
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	
305					310					315					320	
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	
				325					330					335		
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	
			340					345					350			
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	
		355					360					365				
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	
	370					375					380					
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	
385					390					395					400	
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	
			405						410					415		
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	
			420					425					430			
Glu	Val	Asp	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr	
		435					440					445				

Gly	Val	Arg	Leu	Asp	Val	Ala	Cys	Leu	Gln	Ala	Leu	Ser	Leu	Glu	Leu			
450						455					460							
Ala	Glu	Glu	Ile	Arg	Arg	Leu	Glu	Glu	Glu	Val	Phe	Arg	Leu	Ala	Gly			
465					470					475					480			
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe			
			485						490					495				
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys			
			500					505					510					
Arg	Ser	Thr	Ser	Ala	Ala	Ile	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro			
		515					520					525						
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser			
	530					535					540							
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg			
545					550					555					560			
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser			
				565					570					575				
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly			
			580					585					590					
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val			
		595					600					605						
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser			
	610					615					620							
Gly	Asp	Glu	Asn	Leu	Thr	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His			
625					630					635					640			
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp			
				645					650					655				
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr			
			660					665					670					
Gly	Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu			
		675					680					685						
Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val			
	690					695					700							
Arg	Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Lys	Arg	Gly	Tyr			
705					710					715					720			
Val	Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Asn	Ala			
				725					730					735				
Arg	Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn	Met			
			740					745					750					

165

Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765

Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln Val
 770 775 780

His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu Val
 785 790 795 800

Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala Val
 805 810 815

Pro Leu Glu Val Lys Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Glu

<210> 71
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 71
 atggcgatgc ttccctcttt tgagcccaaa ggccgggtcc tcttgggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgcctgaag ggcctcatca cgagccgggg cgaaccggtg 120
 caggcggtct acggtttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc cctccttcc gccacgaggc ctacgaggcc 240
 tacaaggcgg ggaggggccc gacccccgag gacttcccc gccagctcgc cctcatcaag 300
 gagctgggtg acctcctggg gtttaccgc ctcgaggtcc aaggctacga ggcggacgac 360
 gtctctgcca ccctggccaa gaaggcggaa aaagaagggt acgaggtgcg catcctcacc 420
 gccgaccggg acctctacca gctcgtctcc gaccgcgtcg ccgtcctcca ccccgagggc 480
 cacctcatca ccccgagtg gctttgggag aagtacggcc tcaggccgga gcagtgggtg 540
 gacttccgcg ccctcgtggg ggacccctcc gacaacctcc ccgggggtcaa gggcatcggg 600
 gagaagaccg ccctcaagct cctcaaggag tggggaagcc tggaaaatct cctcaagaac 660
 ctggatcggg taaagccgga aaacgtccgg gagaagatca aggccacct ggaagacctc 720
 aggctctcct tggagctctc ccgggtgcgt accgacctcc ccctggaggt ggacctcgcc 780
 caggggctgg agcccgaccg ggaagggtt agggccttcc tggagaggct ggagttcggc 840
 agcctcctcc atgagttcgg ccttctggaa agccccaagg ccctggagga ggccccctgg 900

ccccgccgg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc 960
 gatcttctgg ccctggccgc cgccaggggt ggtcgggtcc accgggcccc cgagccttat 1020
 aaagccctca gggacttgaa ggaggcgcg gggcttctcg ccaaagacct gagcgttctg 1080
 gccctaaggg aaggccttgg cctcccgccc ggcgacgacc ccatgctcct cgcctacctc 1140
 ctggaccctt ccaacaccac ccccgagggg gtggcccggc gctacggcgg ggagtggacg 1200
 gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg 1260
 cttgaggggg aggagaggct cctttggctt taccgggagg tggataggcc cctttccgct 1320
 gtcttgccc acatggaggc cacaggggtg cgcttgacg tggcctatct cagggccttg 1380
 tccttgagg tggccgagga gatcgccgc ctcgaggccg aggtcttccg cctggccggc 1440
 cacccttca acctcaactc ccgggaccag ctggaaaggg tcctctttga cgagttaggg 1500
 cttcccgcca tcggcaagac ggagaggacc ggcaagcgt ccaccagcg cgccgtcctg 1560
 gaggccctcc gcgaggecca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc 1620
 aagctgaaga gcacctacat tgacccttg ccggacctca tccaccccag gacgggcccgc 1680
 ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
 aacctccaga acatccccgt ccgcaccccg cttgggcaga ggatccgccg ggccttcac 1800
 gccgaggagg ggtggctatt ggtggccctg gactatagcc agatagagct cagggtgctg 1860
 gccacctct ccggcgacga gaacctgatc cgggtcttcc aggaggggcg ggacatccac 1920
 acggagaccg ccagctggat gttcgggtgc ccccgaggagg ccgtggaccc cctgatgcgc 1980
 cgggcggcca agacggtgaa cttcggcgtc ctctacggca tgtecgccca taggctctcc 2040
 caggagcttt ccatecccta cgaggaggcg gtggccttta tagagcgcta cttccaaagc 2100
 ttccccaagg tcggggcctg gatagaaaag accctggagg aggggaggaa gcggggctac 2160
 gtggaaaccc tcttcggaag aaggcgctac gtgcccgacc tcaacgcccg ggtgaagagc 2220
 gtcagggagg ccgcggagcg catggccttc aacatgcccg tccagggcac cgccgccgac 2280
 ctcatgaagc tcgccatggt gaagctcttc cccgcctcc gggagatggg ggcccgcacg 2340
 ctctccagg tccacgacga gctcctcctg gagggccccc aagcgcgggc cgaggagggtg 2400
 gcggctttgg ccaaggaggc catggagaag gcctatcccc tcgccgtacc cctggagggtg 2460
 gaggtgggga tcggggagga ctggctctcc gccaaaggagt ga 2502

<210> 72

<211> 833

<400>	72														
Met	Ala	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu	Val
1				5					10					15	
Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	Phe	Ala	Leu	Lys	Gly	Leu
			20					25					30		
Ile	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Ala	Val	Tyr	Gly	Phe	Ala	Lys
		35					40					45			
Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	Val
	50					55					60				
Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Glu	Ala	Tyr	Glu	Ala
65					70					75					80
Tyr	Lys	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu
				85					90					95	
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu
			100					105					110		
Val	Gln	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys
		115					120					125			
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Asp
	130					135					140				
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly
145					150					155					160
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro
				165					170					175	
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn
			180					185					190		
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu
		195					200					205			
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val
	210					215					220				
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu
225					230					235					240
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu
				245					250					255	
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	Ala
			260					265					270		

Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	275	280	285
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	290	295	300
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	305	310	315
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	325	330	335
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	340	345	350
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	355	360	365
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	370	375	380
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	385	390	395
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	405	410	415
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	420	425	430
Glu	Val	Asp	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr	435	440	445
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val	450	455	460
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly	465	470	475
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe	485	490	495
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Arg	Thr	Gly	Lys	500	505	510
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro	515	520	525
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	530	535	540
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg	545	550	555
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	565	570	575

Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
 580 585 590
 Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
 595 600 605
 Ala Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
 610 615 620
 Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
 625 630 635 640
 Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val Asp
 645 650 655
 Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Phe Gly Val Leu Tyr
 660 665 670
 Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ser Ile Pro Tyr Glu
 675 680 685
 Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
 690 695 700
 Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly Tyr
 705 710 715 720
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala
 725 730 735
 Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
 740 745 750
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765
 Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln Val
 770 775 780
 His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu Val
 785 790 795 800
 Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala Val
 805 810 815
 Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Glu

<210> 73
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>

<223> cloned sequence

<400> 73

atggcgatgc ttccctctt tgagcccaag ggccgcgtcc tcctggtgga cggccaccac	60
ctggcctacc gcaccttctt cgccctgaag ggccccacca cgagccgggg cgaaccggtg	120
caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag	180
gccgtcttcg tggcttttga cgccaaggcc ccctcattcc gccacaaggc ctacgaggcc	240
tacagggcgg ggagggcccc gacccccgag gacttcccc ggagctcgc cctcatcaag	300
gagctggtgg acctcctggg gtttaccgc ctcgaggtcc ccggctacga ggccgacgac	360
gttctcgcca ccctggccaa gaaggcggaa aaggaggggt acgaggtgcg catcctcacc	420
gccgaccgcg gcctctacca actcgtctct gaccgcgtcg ccgtcctcca ccccgagggc	480
cacctcatca ccccgagtg gctttgggag aagtacggcc tcaggccgga gcagtgggtg	540
gacttcgcg ccctcgtggg ggacccctcc gacaacctcc ccggggtcaa gggcatcggg	600
gagaagaccg ccctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac	660
ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagacctc	720
aggctctcct tggagctctc ccgggtgcgc accgacctcc ccctggaggt ggacctcgcc	780
caggggcggg agcccgaccg ggagaggctt agggccttcc tggagaggct tgagtttggc	840
ggcctcctcc acgagttcgg ccttctggaa agccccaagg ccctggagga ggccccctgg	900
ccccgcgcg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc	960
gatcttctgg ccctggccgc cgccaggggt ggtcgggtcc accgggcccc cgagccttat	1020
aaagccctca gggacttgaa ggaggcgcgg gggcttctcg ccaaagacct gagcgttctg	1080
gcctgaggg aaggccttgg cctcccgccc ggcgacgacc ccatgctcct cgctacctc	1140
ctggaccctt ccaacaccac ccccgagggg gtggccccggc gctacggcgg ggagtggacg	1200
gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg	1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggagaggcc cctttccggt	1320
gtcctggccc acatggaggg cacaggggtg cgcttgagc tggcctatct cagggccttg	1380
tcctggagg tggccgagga gatcgccgc ctcgaggccg aggtcttccg cctggccggc	1440
cacccttca acctcaactc ccgggaccag ctggaaaggg tcctctttga cgagctaggg	1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgct ccaccggcgc cgccgtcctg	1560
gaggccctcc gcgaggccca cccaccgtg gagaagatcc tgcagtaccg ggagctcacc	1620

```

aagctgaaga gcacctacat tgaccccttg ccggacctca tccaccccag gacgggcccgc 1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgacccc 1740
aacctccaga acatccccgt ccgcacccccg ctcgggcaga ggatccgccg ggccttcac 1800
gccgaggagg ggtggctatt ggtggtcctg gactatagcc agatagagct cagggtgctg 1860
gcccacctct ccggcgacga gaacctgac cgggtcttcc aggagggggcg ggacatccac 1920
acggaaaccg ccagctggat gttcggcgtc ccccgaggagg ccgtggaccc cctaatacgc 1980
cgggcggcca agaccatcaa cttcgggggtt ctctacggca tgcgggcca ccgcctctcc 2040
caggagctag ccatccctta cgaggaggcc caggccttca ttgagcgcta cattcagagc 2100
ttccccaagg tgcgggcctg gattgagaag accctggagg agggcaggag gcgggggtac 2160
gtggagaccc tcttcggccg ccgtcgctac gtgccagacc tagaggcccc ggtgaagagc 2220
gtgcgggagg cggccgagcg catggccttc aacatgcccg tccagggcac cgcgcgcgac 2280
ctcatgaagc tggctatggt gaagctcttc ccaggctgg aagaaacggg ggccaggatg 2340
ctccttcagg tccacgacga gctggtcctc gaggcccaa aagagagggc ggaggccgtg 2400
gcccggctgg ccaaggaggc catggagggg gtgtatcccc tggccgtgcc cctggaggtg 2460
gaggtgggga taggggagga ctggctctcc gccaaaggag ga 2502

```

```

<210> 74
<211> 833
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 74
Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1          5          10         15

Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Pro
          20          25         30

Thr Thr Ser Arg Gly Glu Pro Val Gln Val Val Tyr Gly Phe Ala Lys
          35          40         45

Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe Val
          50          55         60

Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Lys Ala Tyr Glu Ala
65          70          75         80

Tyr Arg Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu

```

172

85								90				95			
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu
			100					105					110		
Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys
		115					120					125			
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Gly
	130					135					140				
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly
145					150					155					160
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro
				165					170					175	
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn
			180					185					190		
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu
		195					200					205			
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val
	210					215					220				
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu
225					230					235					240
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu
				245					250					255	
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala
			260					265					270		
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Gly	Leu	Leu	His	Glu	Phe	Gly	Leu
		275					280					285			
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu
	290					295					300				
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala
305					310					315					320
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala
				325					330					335	
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu
			340					345					350		
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu
		355					360					365			
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser
	370					375					380				
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr

385		390		395		400
Glu Glu Ala Gly	Glu Arg Ala Ala Leu Ser	Glu Arg Leu Phe Ala Asn				
	405	410		415		
Leu Trp Gly Arg	Leu Glu Gly Glu Glu Arg Leu Leu Trp	Leu Tyr Arg				
	420	425		430		
Glu Val Glu Arg	Pro Leu Ser Val Val Leu Ala His Met	Glu Ala Thr				
	435	440		445		
Gly Val Arg Leu Asp	Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val					
	450	455		460		
Ala Glu Glu Ile Ala Arg	Leu Glu Ala Glu Val Phe Arg Leu Ala Gly					
	465	470		475		480
His Pro Phe Asn	Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe					
	485	490		495		
Asp Glu Leu Gly	Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys					
	500	505		510		
Arg Ser Thr Gly	Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro					
	515	520		525		
Thr Val Glu Lys	Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser					
	530	535		540		
Thr Tyr Ile Asp	Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg					
	545	550		555		560
Leu His Thr Arg	Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser					
	565	570		575		
Ser Ser Asp Pro	Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly					
	580	585		590		
Gln Arg Ile Arg	Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val					
	595	600		605		
Val Leu Asp Tyr	Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser					
	610	615		620		
Gly Asp Glu Asn	Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His					
	625	630		635		640
Thr Glu Thr Ala	Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp					
	645	650		655		
Pro Leu Met Arg	Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr					
	660	665		670		
Gly Met Ser Ala	His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu					
	675	680		685		
Glu Ala Gln Ala	Phe Ile Glu Arg Tyr Ile Gln Ser Phe Pro Lys Val					

174

690	695	700
Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr		
705	710	715 720
Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala		
	725	730 735
Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met		
	740	745 750
Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys		
	755	760 765
Leu Phe Pro Arg Leu Glu Glu Thr Gly Ala Arg Met Leu Leu Gln Val		
	770	775 780
His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val		
	785	790 795 800
Ala Arg Leu Ala Lys Glu Ala Met Glu Gly Val Tyr Pro Leu Ala Val		
	805	810 815
Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys		
	820	825 830

Glu

<210> 75
 <211> 2505
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 75
 atgcgtggta tgcttcctct ttttgagccc aagggccgcg tctcctgggt ggacggccac 60
 cacctggcct accgcacctt ctctgccttg aagggcccca ccacgagccg gggcgaaccg 120
 gtgcaggcgg tctacggctt cgccaagagc ctctcaagg ccctgaagga ggacgggtac 180
 aagggcgcct tcgtggtctt tgacgccaaag gccccctcct tccgccacga ggcctacgag 240
 gcctacaagg cggggagggc cccgaccccc gaggacttcc cccggcagct cgccctcatc 300
 aaggagctgg tggacctcct ggggtttacc cgctcgagg tccttggtta cgaggcggac 360
 gacgtcctcg ccacctggc caagaaggcg gaaaaggagg ggtacgaggt gcgcatactc 420
 accgccgacc gcgacctcta ccaactcgtc tccgaccgcg tcgccgtcct ccaccccgag 480
 ggccacctca tcaccccgga gtggcttttg gagaagtacg gcctcaggcc ggagcagtgg 540
 gtggacttcc gcgccctcgt gggggacccc tccgacaacc tccccggggt caaggcatc 600

ggggagaaga ccgcctcaa gctcctcaag gagtggggaa gcctggaaaa cctcctcaag	660
aacctggacc gggtaaagcc agaaaacgtc cgggagaaga tcaaggccca cctggaagac	720
ctcaggctct ccttggagct ctcccgggtg cgcaccgacc tccccctgga ggtggacctc	780
gcccaggggc gggagcccga ccgggagagg cttagggcct ttctggagag gcttgagttt	840
ggcggcctcc tccacgagtt cggccttctg gaaagcccca aggccttga ggaggcccc	900
tggccccgc cggaaggggc cttcgtgggc tttgtgcttt cccgcaagga gcccatgtgg	960
gccgatcttc tggccctggc cgcgcacagg ggtggtcggg tccaccgggc ccccgagcct	1020
tataaagccc tcagggactt gaaggaggcg cgggggcttc tcgccaaaga cctgagcggt	1080
ctggccctaa gggaaggcct tggcctcccg cccggcgacg accccatgct cctcgcctac	1140
ctcctggacc ctccaacac ccccccgag ggggtggccc ggcgctacgg cggggagtg	1200
acggaggagg cgggggagcg ggccgcctt tccgagaggc tcttcgcca cctgtggggg	1260
aggcttgagg gggaggagag gctcctttgg ctttaccggg aggtggatag gccctttcc	1320
gctgtcctgg cccacatgga ggccacaggg gtacggctgg acgtggcctg cctgcaggcc	1380
ctttccctgg agcttgcgga ggagatccgc cgcctcgagg aggaggtctt ccgcttggcg	1440
ggccaccct tcaacctcaa ctcccgggac cagctggaaa gggtcctctt tgacgagcta	1500
gggcttcccg ccatcggcaa gacggagaag accggcaagc gctccaccag cgcgcctc	1560
ctggaggccc tccgcgaggc ccaccccatc gtggagaaga tcctgcagta ccgggagctc	1620
accaagctga agagcaccta cattgacccc ttgccggacc tcatccaccc caggacgggc	1680
cgcctccaca cccgcttcaa ccagacggcc acggccacgg gcaggctaag tagctccgat	1740
cccaacctcc agaacatccc cgtccgcacc ccgctcgggc agaggatccg ccgggccttc	1800
atcgccgagg aggggtggct attggtggtc ctggactata gccagataga gctcagggtg	1860
ctggcccacc tctccggcga cgagaacctg acccgggtct tccaggaggg gcgggacatc	1920
cacacggaaa ccgccagctg gatgttcggc gtcccccggg aggccttga cccctgatg	1980
cgcggggcgg ccaagaccat caacttcggg gttctctacg gcatgtcggc ccaccgcctc	2040
tcccaggagc tggccatccc ttacgaggag gcccaggcct tcatagagcg ctacttccaa	2100
agcttcccca aggtgcgggc ctggatagaa aagaccctgg aggaggggag gaagcggggc	2160
tacgtggaaa ccctcttcgg aagaaggcgc tacgtgcccc acctcaacgc ccgggtgaag	2220
agtgtcaggg aggcgcgga gcgcatggcc ttcaacatgc ccgtccaggg caccgccgc	2280

gaccttatga agctcgccat ggtgaagctc ttcccccgcc tccgggagat gggggcccgc 2340
 atgctcctcc aggtccacga cgagctcctc ctggaggccc cccaagcgcg ggccgaggag 2400
 gtggcggtt tggccaagga ggccatggag aaggcctatc ccctcgccgt acccctggag 2460
 gtgaagggtgg ggatcgggga ggactggctc tccgccaagg agtga 2505

<210> 76
 <211> 834
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 76

Met	Arg	Gly	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu	1	5	10	15
Val	Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	Phe	Ala	Leu	Lys	Gly	20	25	30	
Pro	Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Ala	Val	Tyr	Gly	Phe	Ala	35	40	45	
Lys	Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Ala	Phe	50	55	60	
Val	Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Glu	Ala	Tyr	Glu	65	70	75	80
Ala	Tyr	Lys	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	85	90	95	
Leu	Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	100	105	110	
Glu	Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	115	120	125	
Lys	Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	130	135	140	
Asp	Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	145	150	155	160
Gly	His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	165	170	175	
Pro	Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	180	185	190	
Asn	Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	195	200	205	

Leu	Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	210	215	220	
Val	Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	225	230	235	240
Leu	Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	245	250	255	
Glu	Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	260	265	270	
Ala	Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Gly	Leu	Leu	His	Glu	Phe	Gly	275	280	285	
Leu	Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	290	295	300	
Glu	Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	305	310	315	320
Ala	Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	325	330	335	
Ala	Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	340	345	350	
Leu	Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	355	360	365	
Leu	Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	370	375	380	
Ser	Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	385	390	395	400
Thr	Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	405	410	415	
Asn	Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	420	425	430	
Arg	Glu	Val	Asp	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	435	440	445	
Thr	Gly	Val	Arg	Leu	Asp	Val	Ala	Cys	Leu	Gln	Ala	Leu	Ser	Leu	Glu	450	455	460	
Leu	Ala	Glu	Glu	Ile	Arg	Arg	Leu	Glu	Glu	Glu	Val	Phe	Arg	Leu	Ala	465	470	475	480
Gly	His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	485	490	495	
Phe	Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	500	505	510	

Lys	Arg	Ser	Thr	Ser	Ala	Ala	Ile	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His
		515					520					525			
Pro	Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys
	530					535					540				
Ser	Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly
545					550					555					560
Arg	Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu
				565					570					575	
Ser	Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu
			580					585					590		
Gly	Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu
		595					600					605			
Val	Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu
	610					615					620				
Ser	Gly	Asp	Glu	Asn	Leu	Thr	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile
625					630					635					640
His	Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val
				645					650					655	
Asp	Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu
			660					665					670		
Tyr	Gly	Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr
		675					680					685			
Glu	Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys
	690					695					700				
Val	Arg	Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Lys	Arg	Gly
705					710					715					720
Tyr	Val	Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Asn
				725					730					735	
Ala	Arg	Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn
			740					745					750		
Met	Pro	Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val
		755					760					765			
Lys	Leu	Phe	Pro	Arg	Leu	Arg	Glu	Met	Gly	Ala	Arg	Met	Leu	Leu	Gln
	770					775					780				
Val	His	Asp	Glu	Leu	Leu	Leu	Glu	Ala	Pro	Gln	Ala	Arg	Ala	Glu	Glu
785					790					795					800
Val	Ala	Ala	Leu	Ala	Lys	Glu	Ala	Met	Glu	Lys	Ala	Tyr	Pro	Leu	Ala
			805						810					815	

Val Pro Leu Glu Val Lys Val Gly Ile Gly Glu Asp Trp Leu Ser Ala
 820 825 830

Lys Glu

<210> 77
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 77
 atggcgatgc ttcccctctt tgagcccaag ggccgtgtcc tcctggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgccctgaag ggccccacca cgagccgggg cgaaccggtg 120
 caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc ccccatctcc gccacaaggc ctacgaggcc 240
 tacagggcgg ggagggcccc gacccccgag gacttcccc ggagctcgc cctcatcaag 300
 gagctggtgg acctcctggg gtttaccgc ctcgaggtcc ccggctacga ggcggacgac 360
 gttctcgcca ccctggccaa gaaggcgga aaggaggggt acgaggtgcg catcctcacc 420
 gccgaccgcg gcctctacca actcgtctct gaccgcgtcg ccgtcctcca ccccgagggc 480
 cacctcatca ccccgagtg gctttgggag aagtacggcc tcaggccgga gcagtgggtg 540
 gacttccgcg ccctcgtggg ggacccctcc gacaacctcc ccgggggtcaa gggcatcggg 600
 gagaagaccg ccctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac 660
 ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagatctc 720
 aggctctcct tggagctctc ccgggtgcgc accgacctcc ccctggaggt ggacctcgcc 780
 caggggcggg agcccgaccg ggaggggctt agggcctttc tggagaggct tgagtctggc 840
 agcctcctcc acgagttcgg ccttctggaa agccccaagg ccctggagga ggccccctgg 900
 ccccgccgg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc 960
 gatcttctgg ccctggccgc cgccaggggt ggtcgagtcc accgggcccc cgagccttat 1020
 aaagccctca gggacctgaa ggaggcgcgg gggcttctcg ccaaagacct gagcgttctg 1080
 gccctaaggg aaggccttgg cctcccgccc ggcgacgacc ccatgctcct cgcctacctc 1140
 ctggaccctt ccaacaccac ccccgagggg gtggcccggc gctacggcgg ggagtggacg 1200
 gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg 1260

```

cttgagggggg aggagaggct cctttggctt taccgggagg tggagaggcc cctttccgct 1320
gtcctggccc acatggaggc cacgggggtg cgcttgacg tggcctatct cagggccttg 1380
tccctggagg tggccgagga gatcgcccg ctcgaggccg aggtcttccg cctggccggc 1440
cacccttca acctcaactc ccgggaccag ctggaaatgg tgctctttga cgagcttagg 1500
cttcccgcct tggggaagac gcaaaagacg ggcaagcgct ccaccagcgc cgccgtcctg 1560
gaggccctcc gcgaggcca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgacccttg tcggacctca tccaccccag gacgggcccgc 1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
aacctccaga acatccccgt ccgcaccccg cttgggcaga ggatccgccg ggccttcac 1800
gccgaggagg ggtggctact ggtggctcctg gactatagcc agatagagct cagggtgctg 1860
gccacctct cggcgacga aaacctgatc aggtcttcc aggaggggcg ggacatccac 1920
acggagaccg ccagctggat gttcggcgtc ccccgaggagg ccgtggaccc cctgatgcgc 1980
cgggcggcca agaccatcaa cttcgggggtc ctctacggca tgtcggccca ccgcctctcc 2040
caggagctag ccatccctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttccccaagg tgcgggcctg gattgagaag accctggagg agggcaggag gcgggggtac 2160
gtggagaccc tcttcggccg ccgcccgtac gtgccagacc tagaggcccg ggtgaagagc 2220
gtgcgggagg cggccgagcg catggccttc aacatgcccg tccagggcac cgccgccgac 2280
ctcatgaagc tggctatggt gaagctcttc ccaggctgg aggaaatggg ggccaggatg 2340
ctccttcagg tccacgacga gctggctctc gaggccccaa aagagagggc ggaggccgtg 2400
gcccggtgg ccaaggaggt catggagggg gtgtatcccc tggccgtgcc cctggagggtg 2460
gaggtgggga taggggagga ctggctctcc gccaaaggagt ga 2502

```

```

<210> 78
<211> 833
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 78
Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1           5           10          15
Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Pro
          20          25          30

```

Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Val	Val	Tyr	Gly	Phe	Ala	Lys	
		35					40					45				
Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	Val	
	50					55					60					
Val	Phe	Asp	Ala	Lys	Ala	Pro	Pro	Phe	Arg	His	Lys	Ala	Tyr	Glu	Ala	
65					70					75					80	
Tyr	Arg	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu	
				85					90					95		
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu	
			100					105					110			
Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys	
		115					120					125				
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Gly	
	130					135					140					
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly	
145					150					155					160	
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	
				165					170					175		
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn	
			180					185					190			
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu	
		195					200					205				
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val	
	210					215					220					
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu	
225					230					235					240	
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu	
				245					250					255		
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	Ala	
			260					265					270			
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	
		275					280					285				
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	
	290					295					300					
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	
305					310					315					320	
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	
				325					330					335		

Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu			
			340					345					350					
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu			
		355					360					365						
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser			
	370					375					380							
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr			
385					390					395					400			
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn			
				405					410					415				
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg			
			420					425					430					
Glu	Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr			
		435					440					445						
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val			
	450					455					460							
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly			
465					470					475					480			
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Met	Val	Leu	Phe			
			485					490						495				
Asp	Glu	Leu	Arg	Leu	Pro	Ala	Leu	Gly	Lys	Thr	Gln	Lys	Thr	Gly	Lys			
			500					505					510					
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro			
		515					520					525						
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser			
	530					535					540							
Thr	Tyr	Ile	Asp	Pro	Leu	Ser	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg			
545					550					555					560			
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser			
				565					570					575				
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly			
			580					585					590					
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val			
		595					600					605						
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser			
		610				615					620							
Gly	Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His			
625					630					635					640			

Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
 645 650 655
 Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
 660 665 670
 Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
 675 680 685
 Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
 690 695 700
 Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
 705 710 715 720
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
 725 730 735
 Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
 740 745 750
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765
 Leu Phe Pro Arg Leu Glu Glu Met Gly Ala Arg Met Leu Leu Gln Val
 770 775 780
 His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
 785 790 795 800
 Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
 805 810 815
 Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Glu

<210> 79
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 79
 atggcgatgc ttccctctt tgagcccaaa ggccgggtcc tcttggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgccctgaag ggcctcacca cgagccgggg cgaaccggtg 120
 caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tgggtcttga cgccaaggcc ccttcattcc gccacaaggc ctacgaggcc 240

tacagggcgg	ggagggcccc	gacccccag	gacttcccc	ggcagctcgc	cctcatcaag	300
gagctggtgg	acctcctggg	gtttaccgc	ctcgaggtcc	ccggctacga	ggcggacgac	360
gttctcgcca	ccctggccaa	gaaggcggaa	aaggaggggt	acgaggtgcg	catcctcacc	420
gccgaccgcg	gcctctacca	actcgtctcc	gaccgcgtcg	ccgtcctcca	ccccgagggc	480
cacctcatca	ccccggagtg	gctttgggag	aagtacggcc	tcaggccgga	gcagtgggtg	540
gacttccgcg	ccctcgtggg	ggacccctcc	gacaacctcc	ccgggggtcaa	gggcatcggg	600
gagaagaccg	ccctcaagct	cctcaaggag	tggggaagcc	tggaaaacct	cctcaagaac	660
ctggaccggg	taaagccaga	aaacgtccgg	gagaagatca	aggcccacct	ggaagacctc	720
aggctctcct	tggagctctc	ccgggtgcgc	accgacctcc	ccctggaggt	ggacctcgcc	780
caggggcggg	agcccgaccg	ggaggggctt	agggcctttc	tggagaggct	tgagtttggc	840
agcctcctcc	acgagttcgg	ccttctggaa	agccccaagg	ccctggagga	ggccccctgg	900
ccccgcggg	aaggggcctt	cgtgggcttt	gtgctttccc	gcaaggagcc	catgtgggcc	960
gatcttctgg	ccctggccgc	cgccaggggt	ggtcgagtcc	accaggcccc	cgagccttat	1020
aaagccctca	gggacctgaa	ggaggcgcgg	gggcttctcg	ccaaagacct	gagcgttctg	1080
gccctaaggg	aaggccttgg	cctcccgcgc	ggcgacgacc	ccatgctcct	cgcctacctc	1140
ctggaccctt	ccaacaccac	ccccgagggg	gtggcccggc	gctacggcgg	ggagtggacg	1200
gaggaggcgg	gggagcgggc	cgccctttcc	gagaggctct	tcgccaacct	gtgggggagg	1260
cttgaggggg	aggagaggct	cctttggctt	taccgggagg	tggagaggcc	cctttccgct	1320
gtcctggccc	acatggagac	cacgggggtg	cgccctggacg	tggcctatct	cagggccttg	1380
tccctggagg	tggccgagga	gatcgcccgc	ctcgaggccg	aggtcttccg	cctggccggc	1440
cgcccttca	acctcaactc	ccgagaccag	ctggaaaggg	tcctctttga	cgagctaggg	1500
cttcccgcga	tcggcaagac	ggagaagacc	ggcaagcgct	ccaccagcgc	cgccgtcctg	1560
gaggccctcc	gcgaggccca	ccccatcgtg	gagaagatcc	tgcagtaccg	ggagctcacc	1620
aagctgaaga	gcacctacat	tgacccttg	ccggacctca	tccaccccag	gacgggcccgc	1680
ctccacaccc	gcttcaacca	gacggccacg	gccacgggca	ggctaagtag	ctccgatccc	1740
aacctccaga	acatccccgt	ccgcaccccg	cttgggcaga	ggatccgccc	ggccttcac	1800
gccgaggagg	ggtggctatt	ggtggctcctg	gactatagcc	agatggagct	cagggtgctg	1860
gcccacctct	ccggcgacga	gaacctgatc	agggtcttcc	aggaggggaa	ggacatccac	1920
accagaccg	caagctggat	gttcggtgtc	ccccggagg	ccgtggaccc	cctgatgcgc	1980

```

cgggcggcca agacggtgaa ctctggcgctc ctctacggca tgtccgcca taggctctcc 2040
caggagcttt ccatccccta cgaggaggcg gtggccttca tagagcgcta cttccaaagc 2100
ttccccaagg tgcgggcctg gattgagaag accctggagg agggcaggag gcgggggtac 2160
gtggagaccc tcttcggccg ccgccgctac gtgcccgacc tcaacgcccg gatgaagagc 2220
gtcagggggg ccgaggagcg catggccttc aacatgcccg tccagggcac cgccgcccac 2280
ctcatgaagc tcgccatggt gaagctcttc ccccgctcc gggagatggg ggcccgcacg 2340
ctcctccagg tccacgacga gctcctctg gagggccccc aagcgcgggc cgaggagggtg 2400
gcggctttgg ccaaggaggc catggagaag gcctatcccc tcgccgtacc cctggagggtg 2460
gaggtgggga tcggggagga ctggctctcc gccaaaggagt ga 2502

```

```

<210> 80
<211> 833
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 80
Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1          5          10         15

Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Leu
          20          25         30

Thr Thr Ser Arg Gly Glu Pro Val Gln Val Val Tyr Gly Phe Ala Lys
          35          40         45

Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe Val
          50          55         60

Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Lys Ala Tyr Glu Ala
65          70          75         80

Tyr Arg Ala Gly Arg Ala Pro Thr Pro Gln Asp Phe Pro Arg Gln Leu
          85          90         95

Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu Glu
          100         105        110

Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys Lys
          115        120        125

Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg Gly
          130        135        140

Leu Tyr Gln Leu Val Ser Asp Arg Val Ala Val Leu His Pro Glu Gly

```

145						150						155						160
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro			
				165					170					175				
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn			
				180					185					190				
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu			
				195					200					205				
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val			
				210					215					220				
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu			
				225					230					235				
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu			
				245					250					255				
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	Ala			
				260					265					270				
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu			
				275					280					285				
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu			
				290					295					300				
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala			
				305					310					315				
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Gln	Ala			
				325					330					335				
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu			
				340					345					350				
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu			
				355					360					365				
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser			
				370					375					380				
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr			
				385					390					395				
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn			
				405					410					415				
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg			
				420					425					430				
Glu	Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Thr	Thr			
				435					440					445				
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val			

450		455		460	
Ala Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly					
465		470		475	480
Arg Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe					
	485		490		495
Asp Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys					
	500		505		510
Arg Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro					
	515		520		525
Ile Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser					
	530		535		540
Thr Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg					
545		550		555	560
Leu His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser					
	565		570		575
Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly					
	580		585		590
Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val					
	595		600		605
Val Leu Asp Tyr Ser Gln Met Glu Leu Arg Val Leu Ala His Leu Ser					
	610		615		620
Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Lys Asp Ile His					
625		630		635	640
Thr Gln Thr Ala Ser Trp Met Phe Gly Val Pro Pro Glu Ala Val Asp					
	645		650		655
Pro Leu Met Arg Arg Ala Ala Lys Thr Val Asn Phe Gly Val Leu Tyr					
	660		665		670
Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ser Ile Pro Tyr Glu					
	675		680		685
Glu Ala Val Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val					
	690		695		700
Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr					
705		710		715	720
Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala					
	725		730		735
Arg Met Lys Ser Val Arg Gly Ala Ala Glu Arg Met Ala Phe Asn Met					
	740		745		750
Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys					

755		760		765
Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln Val				
770		775		780
His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu Val				
785		790		800
Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala Val				
	805		810	815
Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys				
	820		825	830

Glu

<210> 81
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 81
 atggcgatgc ttccctcttt tgagcccaag ggccgtgtcc tcctggtgga cggccaccac 60
 ctggcctacc gcacctcctt cgccctgaag ggccccacca cgagccgggg cgaaccggtg 120
 caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc ccccatctcc gccacaaggc ctacgaggcc 240
 tacagggcgg ggagggcccc gacccccgag gacttcccc ggagctcgc cctcgtcaag 300
 gagctggtgg acctcctggg gtttaccgc ctcgaggtcc ccggctacga ggcggaacgac 360
 gttctcgcca ccctggccaa gaaggcggaa aaggaggggt acgaggtgcg catcctcacc 420
 gccgaccgcg gcctctacca actcgtctct gaccgcgtcg ccgtcctcca ccccgagggc 480
 cacctcatca ccccgagagt gctttgggag aagtacggcc tcaggccgga gcagtgggtg 540
 gacttccgcg ccctcgtggg ggacccctcc gacaacctcc ccgggggtcaa gggcatcggg 600
 gagaagaccg ccctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac 660
 ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagacctc 720
 aggctctcct tggagctctc ccgggtgcgc accgacctcc ccctggaggt ggacctcgcc 780
 caggggcggg agcccgaccg ggaaaggctt agggcctttc tggagaggct tgagtttggc 840
 agcctcctcc atgagttcgg ccttctggaa agccccaagg ccctggagga ggccccctgg 900
 ccccgccgg aaggggcctt cgtgggcttt gtgctttccc gcaaggcgcc catgtgggccc 960

gatctttctgg ccttggccgc cgccaggggt ggtcgggtct accgggcccc cgagccttat 1020
aaagccctca gggacttgaa ggaggcgcg gggcttctcg ccaaagacct gagcgttctg 1080
gccctaaggg aaggccttgg cctcccgccc ggcgacgacc ccatgctcct cgcctacctc 1140
ctggaccctt ccaacaccac ccccgagggg gtggcccggc gctacggcgg ggagtggacg 1200
gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg 1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggataggcc cctttccgct 1320
gtcctggccc acatggaggc cacaggggta cggctggacg tggcctgcct gcaggccctt 1380
tccctggagc ttgcggagga gatccgccc ctcgaggagg aggtcttccg cttggcgggc 1440
cacaccttca acctcaactc ccgggaccag ctggaaaggg tcctctttga cgagctaggg 1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgct ccaccagcgc cgccatcctg 1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgacctcttg ccggacctca tccaccccag gacgggccgc 1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
aacctccaga acatccccgt ccgcaccccg cttgggcaga ggatccgcgc ggccttcac 1800
gccgaggagg ggtggctact ggtggctctg gactatagcc agatagagct cagggtgctg 1860
gctcacctct ccggcgacga aaacctgac agggctcttc aggaggggcg ggacatccac 1920
acggagaccg ccagctggat gttcggcgtc ccccgaggagg ccgtggaccc cctgatgcgc 1980
cgggcggcca agaccatcaa cttcgggggtc ctctacggca tgtcggccca ccgcctctcc 2040
caggagctag ccatccctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttccccaagg tgccggcctg gattgagaag gccctggagg agggcaggag gcgggggtac 2160
gtggagaccc tcttcggaag aaggcgctac gtgcccgcac tcaacgcccg ggtgaagagt 2220
gtcagggagg ccgcggagcg catggccttc aacatgcccg tccagggcac cgccgcccgc 2280
cttatgaagc tcgccatggt gaagctcttc cccgcctcc gggagatggg ggcccgcacg 2340
ctcctccagg tccacgacga gctcctcctg gaggcccccc aagcgcgggc cgaggaggtg 2400
gcggctttgg ccaaggaggc catggagaag gcctatcccc tcgccgtacc cctggaggtg 2460
aaggtgggga tcggggagga ctggctctcc gccaaaggagt ga 2502

<210> 82
<211> 833
<212> PRT

<213> Artificial Sequence

<220>

<223> cloned sequence

<400> 82

Met	Ala	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu	Val	1	5	10	15
Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Ser	Phe	Ala	Leu	Lys	Gly	Pro	20	25	30	
Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Val	Val	Tyr	Gly	Phe	Ala	Lys	35	40	45	
Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	Val	50	55	60	
Val	Phe	Asp	Ala	Lys	Ala	Pro	Pro	Phe	Arg	His	Lys	Ala	Tyr	Glu	Ala	65	70	75	80
Tyr	Arg	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu	85	90	95	
Ala	Leu	Val	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu	100	105	110	
Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys	115	120	125	
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Gly	130	135	140	
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly	145	150	155	160
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	165	170	175	
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn	180	185	190	
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu	195	200	205	
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val	210	215	220	
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu	225	230	235	240
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu	245	250	255	
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala	260	265	270	

Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu
		275					280					285			
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu
		290					295				300				
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Ala	Pro	Met	Trp	Ala
305					310					315					320
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	Tyr	Arg	Ala
				325					330					335	
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu
			340					345					350		
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu
		355					360					365			
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser
		370				375					380				
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr
385					390					395					400
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn
				405					410					415	
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg
			420					425					430		
Glu	Val	Asp	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr
		435					440					445			
Gly	Val	Arg	Leu	Asp	Val	Ala	Cys	Leu	Gln	Ala	Leu	Ser	Leu	Glu	Leu
		450				455					460				
Ala	Glu	Glu	Ile	Arg	Arg	Leu	Glu	Glu	Glu	Val	Phe	Arg	Leu	Ala	Gly
465					470					475					480
His	Thr	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe
			485						490					495	
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys
			500					505					510		
Arg	Ser	Thr	Ser	Ala	Ala	Ile	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro
		515					520					525			
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser
		530				535					540				
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg
545					550					555					560
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser
				565					570					575	

192

Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
 580 585 590
 Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
 595 600 605
 Val Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
 610 615 620
 Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
 625 630 635 640
 Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
 645 650 655
 Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
 660 665 670
 Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
 675 680 685
 Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
 690 695 700
 Arg Ala Trp Ile Glu Lys Ala Leu Glu Glu Gly Arg Arg Arg Gly Tyr
 705 710 715 720
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala
 725 730 735
 Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
 740 745 750
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765
 Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln Val
 770 775 780
 His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu Val
 785 790 795 800
 Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala Val
 805 810 815
 Pro Leu Glu Val Lys Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Glu

<210> 83
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>

<223> cloned sequence

<400> 83

atggcgatgc ttcccctctt tgagcccaag ggccgcgtcc tcctgggtgga cggccaccac	60
ctggcctacc gcgccttctt cgccctgaag ggcctcacca cgagccgggg cgaaccggtg	120
caggcggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag	180
gccgtcttcg tggctcttga cgccaaggcc cctccttcc gccacgaggc ctacgaggcc	240
tacaaggcgg ggagggcccc gacccccgag gacttcccc gccagctcgc cctcatcaag	300
gagctgggtg acctcctggg gtttaccgc ctcgaggctc aaggctacga ggcggaacgac	360
gtcctcgcca ccttggccaa gaaggcgga aaagaagggt acgagggtgc catcctcacc	420
gccgaccggg acctctacca gctcgtctcc gaccgcgtcg ccgtcctcca ccccgagggc	480
cacctcatca ccccgaggtg gctttgggag aagtacggcc tcaggccgga gcagtgggtg	540
gacttcgcgc cctcgtggg ggacccctcc aacaacctcc ccgggggtcaa gggcatcggg	600
gagaagaccg cctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac	660
ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagacctc	720
aggctctcct tggagctctc ccgggtgcgc accgacctcc cctggagggt ggacctcgcc	780
caggggcggg agctcgaccg ggagaggctt agggccttcc tggagaggct tgagtttggc	840
ggcctcctcc acgagttcgg ccttctggaa agccccaagg cctggagga ggccccctgg	900
ccccgcggg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc	960
gatcttcttg cctggccgc cgccaggggt ggtcgggtcc accgggcccc cgagccttat	1020
aaagccctca gggacttgaa ggaggcgcgg gggcttctcg ccaaagacct gagcgttctg	1080
gccctaaggg aaggccttgg cctcccgccc ggcgacgacc ccatgctcct cgcctacctc	1140
ctggaccctt ccaacaccgc ccccgagggg gtggcccggc gctacggcgg ggagtggacg	1200
gaggaggcgg gggagcgggc cgcccttcc gagaggctct tcgccaacct gtgggggagg	1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggataggcc cctttccgct	1320
gtcctggccc acatggaggc cacaggggta cggtggacg tggcctatct cagggccttg	1380
tccttgagg tggccgagga gatcgcgcgc ctcgaggccg aggtcttccg cctggccggc	1440
cacccttca acctcaactc ccgagaccag ctggaaaggg tcctctttga cgagctaggg	1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgct ccaccagcgc cgccgtcctg	1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc	1620

```

aagctgaaga gcacctacat tgaccccttg ccgaacctca tccatcccag gacgggccgc 1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
aacctccaga acatccccgt ccgcaccccg ctcgggcaga ggatccgccg ggccttcac 1800
gccgaggagg ggtggctatt ggtggctctg gactatagcc agatagagct cagggtgctg 1860
gccacctct cgggcgacga gaacctgac cgggtcttcc aggaggggag ggacatccac 1920
acggaaaccg ccagctggat gtccggcgtc ccccgggagg ccgtggaccc cctgatgcgc 1980
cgggcggcca agaccatcaa ctccgggggtt ctctacggca tgtcggccca ccgcctctcc 2040
caggagctag ccatccctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttccccaagg tgcgggcctg gatagaaaag accctggagg aggggaggaa gcggggctac 2160
gtggaaaccc tcttcggaag aaggcgctac gtgcccgaac tcaacgcccg ggtgaagggc 2220
gtcagggagg ccgcggagcg catggccttc aacatgcccg tccagggcac cgccgccgac 2280
ctcatgaagc tcgcatggg gaagctcttc cccgcctcc gggagatggg ggcccgcacg 2340
ctcctccagg tccacgacga gctcctcctg gagggccccc aagcgcgggc cggggagggtg 2400
gcggctttgg ccaaggaggc catggagaag gcctatcccc tcgccgtacc cctggagggtg 2460
aaggtgggga tcggggagga ctggctctcc gccaaaggagt ga 2502

```

```

<210> 84
<211> 833
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 84
Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1          5          10         15

Asp Gly His His Leu Ala Tyr Arg Ala Phe Phe Ala Leu Lys Gly Leu
          20          25         30

Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala Lys
          35          40         45

Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe Val
          50          55         60

Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Glu Ala
65          70          75         80

Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
          85          90         95

```

Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu	100	105	110
Val	Gln	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys	115	120	125
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Asp	130	135	140
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly	145	150	155
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	165	170	175
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asn	Asn	180	185	190
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu	195	200	205
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val	210	215	220
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu	225	230	235
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu	245	250	255
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Leu	Asp	Arg	Glu	Arg	Leu	Arg	Ala	260	265	270
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Gly	Leu	Leu	His	Glu	Phe	Gly	Leu	275	280	285
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	290	295	300
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	305	310	315
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	325	330	335
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	340	345	350
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	355	360	365
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	370	375	380
Asn	Thr	Ala	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	385	390	395
																		400

Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	
				405					410					415		
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	
			420					425					430			
Glu	Val	Asp	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr	
		435					440					445				
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val	
	450					455					460					
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly	
465					470					475					480	
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe	
			485					490						495		
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys	
			500					505					510			
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro	
		515					520					525				
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	
	530					535					540					
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asn	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg	
545					550					555					560	
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	
				565					570					575		
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly	
			580					585					590			
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val	
		595					600					605				
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser	
	610					615					620					
Gly	Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His	
625					630					635					640	
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp	
				645					650					655		
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr	
			660					665					670			
Gly	Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu	
		675					680					685				
Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val	
	690					695					700					

Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Lys Arg Gly Tyr
 705 710 715 720
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Asn Ala
 725 730 735
 Arg Val Lys Gly Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
 740 745 750
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765
 Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln Val
 770 775 780
 His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Gly Glu Val
 785 790 795 800
 Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala Val
 805 810 815
 Pro Leu Glu Val Lys Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Glu

<210> 85
 <211> 2499
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 85
 atggcgatgc ttccctcttt tgagcccaaa ggccgggtcc tcttggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgccctgaag ggcctcacca cgagccgggg cgaaccggtg 120
 cagggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tgggtctttga cgccaaggcc ccctccctcc gccacgaggc ctacgaggcc 240
 tacaaggcgg ggagggcccc gacccccgag gacttcctcc ggcagctcgc cctcatcaag 300
 gagctggtgg acctcctggg gtttaccgc ctcgaggtcc aaggctacga ggcggacgac 360
 gtctcgcca ccctggccaa gaaggcggaa aaagaagggt acgaggtgcg catcctcacc 420
 gccgaccggg acctctacca gctcgtctcc gaccgcgtcg ccgtcctcca ccccgagggc 480
 cacctcatca ccccgagtg gctttgggag aagtacggcc tcaggccgga gcagtgggtg 540
 gacttcgcgc ccctcgtggg ggacccctcc gacaacctcc ccgggggtcaa gggcatcggg 600

gagaagaccg cctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac	660
ctggaccggc tgaagcccgc catccgggag aagatcctgg ccacatgga cgatctgaag	720
ctctcctggg acctggccaa ggtgcgcacc gacctgcccc tagaggtgga cttcgccaaa	780
aggcgggagc ccgaccggga gaggcttagg gcctttcttg agaggcttga gcttggcagc	840
ctcctccacg agttcggcct tctggaaagc cccaagacct tggaggaggc ctctggccc	900
ccgccggaag gggccttcgt gggctttgtg ctttcccgca aggagcccat gtgggccgat	960
cttctggccc tggccgcccgc cagggggggc cgggtccacc gggccccga gccttataaa	1020
gccctcaggg acctgaagga ggcgcggggg cttctcgcca aagacctgag cgttctggcc	1080
ctaagggaag gccttggcct cccgcccggc gacgaccca tgctcctcgc ctacctctg	1140
gacccttcca acaccacccc cgaggggggtg gcccggcgt acggcgggga gtggacgaag	1200
gaggcggggg agcgggccgc cctttccgag aggtctctcg ccaacctgtg ggggaggctt	1260
gagggggagg agaggctcct ttggctttac cgggaggtgg ataggccct ttcgctgtc	1320
ctggcccaca tggaggccac aggggtgcgc ttggacgtgg cctatctcag ggccttgtec	1380
ctggaggtgg ccgaggagat cgcgcgctc gaggcgagg tcttcgcct ggccggccat	1440
cccttcaacc tcaactcccg ggaccagctg gaaagggtcc tctttgacga gctagggctt	1500
ccgcctatcg gcaagacgga gaagaccggc aagcgctcca ccagcgccgc cgtcctggag	1560
gcctccgcg agggccaccc catcgtggag aagatcctgc agtaccggga gctaccaag	1620
ctgaagagca cctacattga ccccttgccg gacctcatcc accccaggac gggccgcctc	1680
cacaccgct tcaaccagac ggccacggcc acgggcaggc taagtagctc cgatcccaac	1740
ctccagaaca tccccgtccg cccccgctc gggcagagga tccgccgggc cttcgtcgcc	1800
gaggaggggt ggctattggt ggtcctggac tatagccaga tagagctcag ggtgctggcc	1860
cacctctccg gcgacgagaa cctgaccggg gtcttcctgg aggggcggga catccacacg	1920
gaaaccgcca gctggatggt cggcgtcccc cgggaggccg tggacccct gatgcgccg	1980
gcggccaaga ccatcaactt cggggttctc tacggcatgt cggcccaccg cctctcccag	2040
gagctggcca tcccttacga ggaggcccag gccttcatag agcgctactt ccaaagcttc	2100
cccaagggtgc gggcctggat agaaaagacc ctggaggagg ggaggaagcg gggctacgtg	2160
gaaaccctct tcggaagaag gcgctacgtg cccgacctca acgcccgggt gaagagtgtc	2220
agggaggccg cggagcgcac ggccttcaac atgcccgtcc agggcaccgc cgcgacctt	2280
atgaagctcg ccatggtgaa gctcttcccc cgctccggg agatgggggc ccgcatgctc	2340

ctccaggtcc acgacgagct cctcctggag gccccccaag cgcgggccga ggaggtggcg 2400
 gctttggcca aggaggccat ggagaaggcc tatccctcgc ccgtaccctt ggaggtgaag 2460
 gaggggatcg gggaggactg gctctccgcc aaggagtga 2499

<210> 86
 <211> 832
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 86

Met	Ala	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu	Val	1	5	10	15
Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	Phe	Ala	Leu	Lys	Gly	Leu	20	25	30	
Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Val	Val	Tyr	Gly	Phe	Ala	Lys	35	40	45	
Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	Val	50	55	60	
Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Leu	Arg	His	Glu	Ala	Tyr	Glu	Ala	65	70	75	80
Tyr	Lys	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Leu	Arg	Gln	Leu	85	90	95	
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu	100	105	110	
Val	Gln	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys	115	120	125	
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Asp	130	135	140	
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly	145	150	155	160
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	165	170	175	
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn	180	185	190	
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu	195	200	205	
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Leu				

200

210	215	220
Lys Pro Ala Ile Arg Glu	Lys Ile Leu Ala His Met Asp Asp Leu Lys	
225	230	235 240
Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu Val		
	245	250 255
Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala Phe		
	260	265 270
Leu Glu Arg Leu Glu Leu Gly Ser Leu Leu His Glu Phe Gly Leu Leu		
	275	280 285
Glu Ser Pro Lys Thr Leu Glu Glu Ala Ser Trp Pro Pro Pro Glu Gly		
	290	295 300
Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala Asp		
305	310	315 320
Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala Pro		
	325	330 335
Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu Leu		
	340	345 350
Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu Pro		
	355	360 365
Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn		
	370	375 380
Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Lys		
385	390	395 400
Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn Leu		
	405	410 415
Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg Glu		
	420	425 430
Val Asp Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr Gly		
	435	440 445
Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala		
	450	455 460
Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His		
465	470	475 480
Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp		
	485	490 495
Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg		
	500	505 510
Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile		

515					520					525					
Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	Thr
530						535					540				
Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg	Leu
545					550					555					560
His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	Ser
				565					570					575	
Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly	Gln
			580					585					590		
Arg	Ile	Arg	Arg	Ala	Phe	Val	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val	Val
		595					600					605			
Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser	Gly
	610					615					620				
Asp	Glu	Asn	Leu	Thr	Arg	Val	Phe	Leu	Glu	Gly	Arg	Asp	Ile	His	Thr
625					630					635					640
Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp	Pro
				645					650					655	
Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr	Gly
			660					665					670		
Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu	Glu
		675					680					685			
Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val	Arg
	690					695					700				
Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Lys	Arg	Gly	Tyr	Val
705					710					715					720
Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Asn	Ala	Arg
				725					730					735	
Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn	Met	Pro
			740					745					750		
Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val	Lys	Leu
		755					760					765			
Phe	Pro	Arg	Leu	Arg	Glu	Met	Gly	Ala	Arg	Met	Leu	Leu	Gln	Val	His
	770					775					780				
Asp	Glu	Leu	Leu	Leu	Glu	Ala	Pro	Gln	Ala	Arg	Ala	Glu	Glu	Val	Ala
785					790					795					800
Ala	Leu	Ala	Lys	Glu	Ala	Met	Glu	Lys	Ala	Tyr	Pro	Leu	Ala	Val	Pro
				805					810					815	
Leu	Glu	Val	Lys	Glu	Gly	Ile	Gly	Glu	Asp	Trp	Leu	Ser	Ala	Lys	Glu

820	825	830	
<210>	87		
<211>	2550		
<212>	DNA		
<213>	Artificial Sequence		
<220>			
<223>	cloned sequence		
<400>	87		
atggcgatgc	ttcccctctt	tgagcccaag	ggccgcgtcc
tccctggtgga	cggccaccac		60
ctggcctacc	gcaccttctt	cgccctgaag	ggccccacca
cgagccgggg	cgaaccggtg		120
caggtggtct	acggcttcgc	caagagcctc	ctcaaggccc
tgaaagagga	cgggtacaag		180
gccgtcttcg	tggctcttga	cgccaaggcc	ccctcattcc
gccacaaggc	ctacgaggcc		240
tacagggcgg	ggagggcccc	gacccccgag	gacttcccc
ggcagctcgc	cctcatcaag		300
gagctggtgg	acctcctggg	gtttaccgcg	ctcgagggtc
ccggctacga	ggcggacgac		360
gttctcgcca	ccctggccaa	gaaggcggaa	aaggaggggt
acgaggtgcg	catcctcacc		420
gccgaccgcg	gcctctacca	actcgtctct	gaccgcgtcg
ccgtcctcca	ccccgagggc		480
cacctcatca	ccccggagtg	gctttgggag	aagtacggcc
tcaggccgga	gcagtgggtg		540
gacttccgcg	ccctcgtggg	ggacccctcc	gacaacctcc
ccgggggtcaa	gggcatcggg		600
gagaagaccg	ccctcaagct	cctcaaggag	tgggggaagcc
tggaaaacct	cctcaagaac		660
ctggaccggg	taaagccaga	aaacgtccgg	gagaagatca
aggccccacct	ggaagacctc		720
aggctctcct	tggagctctc	ccgggtgcgc	accgacctcc
ccctggaggt	ggacctcgcc		780
caggggcggg	agccccgaccg	ggagaggctt	agggcctttc
tggagaggct	tgagtttggc		840
ggcctcctcc	acgagttcgg	ccttctggaa	agcccccaagg
ccctggagga	ggccccctgg		900
cccccgccgg	aaggggcctt	cgtgggcttt	gtgctttccc
gcaaggagcc	catgtgggcc		960
gatcttctgg	ccctggccgc	cgccaggggt	ggtcgggtcc
accgggcccc	tgagccttat		1020
aaagccctca	gggacttgaa	ggaggcgcgg	gggcttctcg
ccaaagacct	gagcgttctg		1080
gccctgaggg	aaggccttgg	cctccccgcc	ggcgacgacc
ccatgctcct	cgcctacctc		1140
ctggaccctt	ccaacaccac	ccccgagggg	gtggccccggc
gctacggcgg	ggagtggacg		1200
gaggaggcgg	gggagcgggc	cgccctttcc	gagaggctct
tcgccaacct	gtgggggagg		1260
cttgaggggg	aggagaggct	cctttggctt	taccgggagg
tggagagacc	cctttccgct		1320
gtcctggccc	acatggaggc	cacgggggtg	cgcttgacg
tggcctatct	cagggccttg		1380

```

tccttgagg tggccgagga gatcgccgc ctcgaggccg aggtcttcg cctggccggc 1440
cacccttca acctcaactc ccgagaccag ctggaaaggg tcctctttga cgagctaggg 1500
cttcccgcga tcggcaagac ggagaagacc ggcaagcgct ccaccagcgc cgccgtcctg 1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgacctcttg ccggaccaca tccaccccag gacgggcccgc 1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
aacctccaga acatccccgt ccgcaccccg ctcgggcaga ggatccgccg ggccttcac 1800
gccgaggagg ggtggctatt ggtggctcctg gactatagcc agatagagct cagggtgctg 1860
gccacctct cgggcgacga gaacctgacc cgggtcttcc aggagggggc ggacatccac 1920
acggaaaccg ccagctggat gtccggcgtc ccccgaggag ccgtggacc cctgatgcgc 1980
cgggcggcca agaccatcaa ctccgggggt ctctacggca tgcggccca ccgcctctcc 2040
caggagctgg ccatccctta cgaggaggcc caggccttca tagagcgcta cttccaaagc 2100
ttcccgaagg tgcgggcctg gatagaaaag accctggagg aggggaggaa gcggggctac 2160
gtggaacccc tcttcggaag aaggcgctac gtgcccgcacc tcaacgcccg ggtgaagagt 2220
gtcagggagg ccgcggagcg catggccttc aacatgcccg tccagggcac cgccgccgac 2280
cttatgaagc tcgccatggg gaagctctac cccgcctcc gggagatggg ggcccgcctg 2340
ctcctccagg tccacgacga gctcctcctg gaggccccc aagcgcgggc cgaggaggtg 2400
gcggctttgg ccaaggaggc catggagaag gcctatcccc tcgccgtacc cctggaggtg 2460
aaggtgggga tcggggagga ctggctctcc gcccaaggag tgagtcgacc tgcaggcagc 2520
gcttggcgtc accgcagtt cggtgggttaa 2550

```

```

<210> 88
<211> 849
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 88
Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1          5          10          15

Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Pro
20          25          30

Thr Thr Ser Arg Gly Glu Pro Val Gln Val Val Tyr Gly Phe Ala Lys

```

		35				40						45					
Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	Val		
	50					55					60						
Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Lys	Ala	Tyr	Glu	Ala		
65					70					75					80		
Tyr	Arg	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu		
				85					90					95			
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu		
			100					105					110				
Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys		
		115					120					125					
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Gly		
	130					135					140						
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly		
145					150					155					160		
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro		
				165					170					175			
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn		
			180					185					190				
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu		
		195					200					205					
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val		
	210					215					220						
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu		
225					230					235					240		
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu		
				245					250					255			
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala		
			260					265					270				
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Gly	Leu	Leu	His	Glu	Phe	Gly	Leu		
		275					280					285					
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu		
	290					295					300						
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala		
305					310					315					320		
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala		
				325					330					335			
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu		

205

			340					345					350			
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	
		355					360					365				
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	
	370					375					380					
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	
385					390					395					400	
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	
				405					410					415		
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	
			420					425					430			
Glu	Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr	
		435					440					445				
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val	
	450					455					460					
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly	
465					470					475					480	
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe	
			485					490						495		
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys	
			500					505					510			
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro	
		515					520					525				
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	
	530					535					540					
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	His	Ile	His	Pro	Arg	Thr	Gly	Arg	
545					550					555					560	
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	
				565					570					575		
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly	
			580					585					590			
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val	
		595					600					605				
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser	
	610					615					620					
Gly	Asp	Glu	Asn	Leu	Thr	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His	
625					630					635				640		
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp	

206

645						650						655					
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr		
			660					665					670				
Gly	Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu		
		675					680							685			
Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val		
	690						695				700						
Arg	Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Lys	Arg	Gly	Tyr		
705					710					715					720		
Val	Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Asn	Ala		
				725					730					735			
Arg	Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn	Met		
			740					745					750				
Pro	Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Met	Lys	Leu	Ala	Met	Val	Lys		
		755					760						765				
Leu	Tyr	Pro	Arg	Leu	Arg	Glu	Met	Gly	Ala	Arg	Met	Leu	Leu	Gln	Val		
	770						775				780						
His	Asp	Glu	Leu	Leu	Leu	Glu	Ala	Pro	Gln	Ala	Arg	Ala	Glu	Glu	Val		
785					790					795					800		
Ala	Ala	Leu	Ala	Lys	Glu	Ala	Met	Glu	Lys	Ala	Tyr	Pro	Leu	Ala	Val		
				805					810					815			
Pro	Leu	Glu	Val	Lys	Val	Gly	Ile	Gly	Glu	Asp	Trp	Leu	Ser	Ala	Gln		
			820					825					830				
Gly	Val	Ser	Arg	Pro	Ala	Gly	Ser	Ala	Trp	Arg	His	Pro	Gln	Phe	Gly		
		835					840						845				

Gly

<210> 89
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 89
 atggcgatgc ttccctctt tgagcccaag ggccgcgtcc tcctggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgccctgaag ggccccacca cgagccgggg cgaaccggtg 120
 caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc ccctcattcc gccacaaggc ctacgaggcc 240

tacagggcgg ggagggcccc gacccccgag gacttcccc	ggcagctcgc cctcatcaag	300
gagctggtgg acctcctggg gtttaccgcg ctcgaggtec	ccggctacga ggcggacgac	360
gttctcgcca ccctggccaa gaaggcggaa aaggaggggt	acgaggcgcg catcctcacc	420
gccgaccgcg gcctctacca actcgtctct gaccgcgtcg	ccgtcctcca ccccgagggc	480
cacctcatca ccccgagagt gctttgggag aagtacggcc	tcaggccgga gcagtgggta	540
gacttccgcg ccctcgtggg ggacccctcc gacaacctcc	ccgggggtcaa gggcatcggg	600
gagaagaccg ccctcaagct cctcaaggag tggggaagcc	tggaaaacct cctcaagaac	660
ctggaccggg taaagccaga aaacgtccgg gagaagatca	aggcccacct ggaagacctc	720
aggctctcct tggagctctc ccgggtgcgc accgacctcc	ccctggaggt ggacctcgcc	780
caggggcggg agcccgaccg ggaggggctt agggcctttc	tggagaggct tgagtgttggc	840
agcctcctcc acgagttcgg ccttctggaa agccccaagg	ccctggagga ggccccctgg	900
ccccgcgg aaggggcctt cgtgggcttt gtgctttcac	gcaaggagcc catgtgggcc	960
gatcttctgg ccctggccgc cgccaggggt ggtcgggtcc	accgggcccc cgagccttat	1020
aaagccctca gggacttgaa ggaggcgcgg gggcttctcg	ccaaagacct gagcgttctg	1080
gccctaaggg aaggccttgg cctcccgccc ggcgacgacc	ccatgctcct cgcctacctc	1140
ctggaccctt ccaacaccgc ccccgagggg gtggcccggc	gctacggcgg ggagtggacg	1200
gaggaggcgg gggagcgggc cgccctttcc gagaggctct	tcgccaacct gtgggggagg	1260
cttgaggggg aggagaggct cctttggctt taccgggagg	tggataggcc cctttccgct	1320
gtcctggccc acatggaggc cacaggggta cggctggacg	tggcctgcct gcaggccctt	1380
tcctggagc ttgcggagga gatccgcgc ctcgaggagg	aggtcttccg cttggcgggc	1440
cacccttca acctcaactc ccgggaccag ctggaaaggg	tcctctttga cgagctaggg	1500
cttcccgcca tcggcaagac ggagaagacc ggcaagcgct	ccaccagcgc cgccatcctg	1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc	tgcagtaccg ggagctcacc	1620
aagctgaaga gcacctacat tgacctcttg ccggacctca	tcacccccag gacgggccgc	1680
ctccacaccc gcttcaacca gacggccacg gccacgggca	ggctaagtag ctccggtccc	1740
aacctccaga acatccccgt ccgcaccccg ctcgggcaga	ggatccgccc ggccttcgtc	1800
gccgaggagg ggtggctatt ggtggtcctg gactatagcc	agatagagct cagggtgctg	1860
gcccacctct ccggcgacga gaacctgacc cgggtcttcc	tggagggggc ggacatccac	1920

```

acggaaccg ccagctggat gttcggcgtc ccccgagg cegtggacc cctgatgcgc 1980
cgggcgcca agaccatcaa cttcggggtt ctctacggca tgcgggcca cgcctctcc 2040
caggagctgg ccattccctta cgaggaggcc caggccttca tagagcgcta cttccaaagc 2100
ttccccaagg tgcgggcctg gatagaaaag acctggagg aggggaggaa gcggggctac 2160
gtggaaaccc tcttcggaag aaggcgctac gtgcccgacc tcaacgccc ggtgaagagt 2220
gtcagggagg cgcgggagcg catggccttc aacatgcccg tccagggcac cgcgcgcgac 2280
cttatgaagc tcgccaatgg gaagctcttc ccccgctcc gggagatggg ggcccgcgac 2340
ctcctccagg tccacgacga gctcctcctg gagggcccc aagcgcgggc cgaggaagtg 2400
gcggctttgg ccaaggaggc catggagaag gcctatcccc tcgccgtacc cctggaggtg 2460
aaggtgggga tcggggagga ctggctctcc gccaaaggagt ga 2502

```

```

<210> 90
<211> 833
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 90
Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1          5          10          15
Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Pro
          20          25          30
Thr Thr Ser Arg Gly Glu Pro Val Gln Val Val Tyr Gly Phe Ala Lys
          35          40          45
Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe Val
          50          55          60
Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Lys Ala Tyr Glu Ala
65          70          75          80
Tyr Arg Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
          85          90          95
Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu Glu
          100          105          110
Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys Lys
          115          120          125
Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg Gly
          130          135          140

```

Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly
145					150					155					160
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro
				165					170					175	
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn
			180					185					190		
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu
		195					200					205			
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val
	210					215					220				
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu
225					230					235					240
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu
				245					250					255	
Val	Asp	Leu	Ala	Gln	Gly	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	Ala
			260					265					270		
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu
		275					280					285			
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu
	290					295					300				
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala
305					310					315					320
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala
				325					330					335	
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu
			340					345					350		
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu
		355					360					365			
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser
	370					375					380				
Asn	Thr	Ala	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr
385					390					395					400
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn
				405					410					415	
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg
			420					425					430		
Glu	Val	Asp	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr
		435					440					445			

Gly	Val	Arg	Leu	Asp	Val	Ala	Cys	Leu	Gln	Ala	Leu	Ser	Leu	Glu	Leu			
450						455					460							
Ala	Glu	Glu	Ile	Arg	Arg	Leu	Glu	Glu	Glu	Val	Phe	Arg	Leu	Ala	Gly			
465					470					475					480			
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe			
				485					490					495				
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys			
			500					505					510					
Arg	Ser	Thr	Ser	Ala	Ala	Ile	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro			
		515					520					525						
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser			
	530					535					540							
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg			
545					550					555					560			
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser			
				565				570						575				
Ser	Ser	Gly	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly			
			580					585					590					
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Val	Ala	Glu	Glu	Gly	Trp	Leu	Leu	Val			
		595					600					605						
Val	Leu	Asp	Tyr	Ser	Gln	Ile	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser			
	610					615					620							
Gly	Asp	Glu	Asn	Leu	Thr	Arg	Val	Phe	Leu	Glu	Gly	Arg	Asp	Ile	His			
625					630					635					640			
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp			
				645					650					655				
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr			
			660					665					670					
Gly	Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu			
		675					680					685						
Glu	Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val			
	690					695					700							
Arg	Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Lys	Arg	Gly	Tyr			
705					710					715					720			
Val	Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Asn	Ala			
				725					730					735				
Arg	Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn	Met			
			740					745					750					

Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765

Leu Phe Pro Arg Leu Arg Glu Met Gly Ala Arg Met Leu Leu Gln Val
 770 775 780

His Asp Glu Leu Leu Leu Glu Ala Pro Gln Ala Arg Ala Glu Glu Val
 785 790 795 800

Ala Ala Leu Ala Lys Glu Ala Met Glu Lys Ala Tyr Pro Leu Ala Val
 805 810 815

Pro Leu Glu Val Lys Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Glu

<210> 91
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 91
 atggcgatgc ttccctcttt tgagcccaag ggccgcgtcc tcttggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgccctgaag ggccccacca cgagccgggg cgaaccggtg 120
 caggtggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc ccctcattcc gccacaaggc ctacgaggcc 240
 tacagggcgg ggagggcccc gacccccgag gacttcccc ggagctcgc cctcatcaag 300
 gagctggtgg acctcctggg gtttaccgc ctcgaggtcc ccggctacga ggcggacgac 360
 gttctcgcca ccctggccaa gaaggcggaa aaggaggggt acgaggtgcg catcctcacc 420
 gccgaccgcg gcctctacca actcgtctct gaccgcgtcg ccgtcctcca ccccgagggc 480
 cacctcatca ccccgagtg gctttgggag aagtacggcc tcaggccgga gcagtgggtg 540
 gacttccgcg ccctcgtggg ggacccctcc gacaacctcc ccgggggtcaa gggcatcggg 600
 gagaagaccg ccctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac 660
 ctggaccggg taaagccaga aaacgtccgg gagaagatca aggccacct ggaagacctc 720
 aggctctct tggagctctc ccgggtgcgc accgacctcc ccctggaggt ggacctcgcc 780
 cagaggcggg agcccgaccg ggaggggctt agggcctttc tggagaggct tgagtttggc 840
 agcctcttcc acgagttcgg ccttctggaa agccccaagg ccctggagga ggccccctgg 900

ccccgcgcgg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggccc 960
gatcttcttg ccctggccgc cgccaggggt ggtcgagtec accgggcccc cgagccttat 1020
aaagccctca gggacctgaa ggaggcgcgg gggcttctcg ccaaagacct gagcgttctg 1080
gccctaaggg aaggccttgg cctcccgcgc ggcgacgacc ccatgctcct cgcctacctc 1140
ctggaccctt ccaacaccac ccccgagggg gtggcccggc gctacggcgg ggagtggacg 1200
gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg 1260
cttgaggggg aggagaggct cctttggctt taccgggagg tggagaggcc cctttccgct 1320
gtcctggccc acatggaggc cacgggggtg cgcttgacg tggcctatct cagggccttg 1380
tccttgaggg tggccgagga gatcgccgc ctcgaggccg aggtcttccg cctggccggc 1440
cacccttca acctcaactc ccgggaccag ctggaaatgg tgctctttga cgagcttagg 1500
cttcccgcct tggggaagac gcaaaagacg ggcaagcgt ccaccagcgc cgccgtcctg 1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgacccttg tcggacctca tccaccccag gacgggcccgc 1680
ctccacacce gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
aacctccaga acatcccgt ccgcaccccg cttgggcaga ggatccgccg ggccttcac 1800
gccgaggagg ggtggctact ggtggctcctg gactatagcc agatagagct cagggtgctg 1860
gcccacctct ccggcgacga aaacctgatc agggctcttc aggaggggcg ggacatccac 1920
acggagaccg ccagctggat gttcggcgtc cccggggagg ccgtggaccc cctgatgcgc 1980
cgggcggcca agaccatcaa cttcggggtc ctctacggca tgtcggccca ccgcctctcc 2040
caggagctag ccatecctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttccccaagg tgcgggcctg gattgagaag accctggagg agggcaggag gcgggggtac 2160
gtggagaccc tcttcggccg ccgccgtac gtgccagacc tagaggcccg ggtgaagagc 2220
gtgcgggagg cggccgagcg catggccttc aacatgcccg tccagggcac cgccgccgac 2280
ctcatgaagc tggctatggt gaagctcttc ccaggctgg gagaaacggg ggccaggatg 2340
ctccttcagg tccacgacga gctggctctc gaggcccaa aagagagggc ggaggccgtg 2400
gcccggttg ccaaggaggc catggagggg gtgtatcccc tggccgtgcc cctggaggtg 2460
gagggtggga taggggagga ctggctctcc gccaaagggt ag 2502

<210> 92

<211> 833

213

<212> PRT

<213> Artificial Sequence

<220>

<223> cloned sequence

<400> 92

Met	Ala	Met	Leu	Pro	Leu	Phe	Glu	Pro	Lys	Gly	Arg	Val	Leu	Leu	Val
1				5					10					15	

Asp	Gly	His	His	Leu	Ala	Tyr	Arg	Thr	Phe	Phe	Ala	Leu	Lys	Gly	Pro
			20					25					30		

Thr	Thr	Ser	Arg	Gly	Glu	Pro	Val	Gln	Val	Val	Tyr	Gly	Phe	Ala	Lys
		35					40					45			

Ser	Leu	Leu	Lys	Ala	Leu	Lys	Glu	Asp	Gly	Tyr	Lys	Ala	Val	Phe	Val
	50					55					60				

Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Lys	Ala	Tyr	Glu	Ala
65					70					75					80

Tyr	Arg	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu
				85					90					95	

Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu
			100					105					110		

Val	Pro	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys
		115					120					125			

Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Gly
	130					135					140				

Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly
145					150					155					160

His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro
				165					170					175	

Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn
			180					185					190		

Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu
		195					200					205			

Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val
	210					215					220				

Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu
225					230					235					240

Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu
				245					250					255	

Val	Asp	Leu	Ala	Gln	Arg	Arg	Glu	Pro	Asp	Arg	Glu	Gly	Leu	Arg	Ala
			260					265					270		

Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Phe	His	Glu	Phe	Gly	Leu
		275					280					285			
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu
		290					295				300				
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala
305					310					315					320
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala
				325					330					335	
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu
			340					345					350		
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu
		355					360					365			
Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser
		370				375					380				
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr
385					390					395					400
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn
			405						410					415	
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg
			420				425						430		
Glu	Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr
		435					440					445			
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val
		450				455					460				
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly
465					470					475					480
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Met	Val	Leu	Phe
			485						490					495	
Asp	Glu	Leu	Arg	Leu	Pro	Ala	Leu	Gly	Lys	Thr	Gln	Lys	Thr	Gly	Lys
			500					505					510		
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro
		515					520					525			
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser
		530				535					540				
Thr	Tyr	Ile	Asp	Pro	Leu	Ser	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg
545					550					555					560
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser
				565					570					575	

Ser Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly
 580 585 590
 Gln Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val
 595 600 605
 Val Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser
 610 615 620
 Gly Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His
 625 630 635 640
 Thr Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp
 645 650 655
 Pro Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr
 660 665 670
 Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
 675 680 685
 Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
 690 695 700
 Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
 705 710 715 720
 Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
 725 730 735
 Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
 740 745 750
 Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
 755 760 765
 Leu Phe Pro Arg Leu Gly Glu Thr Gly Ala Arg Met Leu Leu Gln Val
 770 775 780
 His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
 785 790 795 800
 Ala Arg Leu Ala Lys Glu Ala Met Glu Gly Val Tyr Pro Leu Ala Val
 805 810 815
 Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
 820 825 830

Gly

<210> 93
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>

<223> primer

<400> 93

cgtggtcgcg acggatgccg

20

<210> 94

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide

<400> 94

agctaccatg cctgcacgaa ttcggcatcc gtcgcgacca cggtcgcagc g

51

<210> 95

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> oligonucleotide

<220>

<221> misc_feature

<222> (22)..(22)

<223> n is an abasic site

<400> 95

agctaccatg cctgcacgac ancggcatcc gtcgcgacca cggtcgcagc g

51

<210> 96

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<220>

<221> misc_feature

<222> (40)..(40)

<223> n is 5nitroindol or a derivative

<400> 96

caggaaacag ctatgacaaa aatctagata acgagggcan

40

<210> 97

<211> 45

<212> DNA

<213> Artificial Sequence

<220>
<223> primer

<220>
<221> misc_feature
<222> (45)..(45)
<223> n is 5nitroindol or a derivative

<400> 97
gtaaaacgac ggccagtacc accgaactgc gggtgacgcc aagcn

45

<210> 98
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<220>
<221> misc_feature
<222> (31)..(31)
<223> n is 5nitroindol or a derivative

<220>
<221> misc_feature
<222> (40)..(40)
<223> n is 5nitroindol or a derivative

<400> 98
caggaaacag ctatgacaaa aatctagata ncgagggcan

40

<210> 99
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<220>
<221> misc_feature
<222> (23)..(23)
<223> n is 5nitroindol or a derivative

<220>
<221> misc_feature
<222> (45)..(45)
<223> n is 5nitroindol or a derivative

<400> 99
gtaaaacgac ggccagtacc acngaactgc gggtgacgcc aagcn

45

<210> 100
 <211> 2502
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> cloned sequence

<400> 100
 atggcgatgc ttcccctctt tgagcccaaa ggccgggtcc tcctggtgga cggccaccac 60
 ctggcctacc gcaccttctt cgccctgaag ggcctcacca cgagccgggg cgaaccggtg 120
 caggcggttt acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
 gccgtcttcg tggctcttga cgccaaggcc cctccttcc gccacgaggc ctacgaggcc 240
 tacaaggcgg ggagggcccc gacccccgag gacttcccc ggcagctcgc cctcatcaag 300
 gagctggtgg acctcctggg gtttaccgcg ctcgaggctc aaggctacga ggcggacgac 360
 gtctctgcca ccctggccaa gaaggcggaa aaagaagggt acgaggtgcg catcctcacc 420
 gccgaccggg acctctacca gctcgtctcc gaccgcgtcg ccgtcctcca ccccgagggc 480
 cacctcatca ccccgagagt gctttgggag aagtaaggcc tcaggccgga gcagtgggtg 540
 gacttcgcgc ccctcgtggg ggaccctcc gacaacctcc ccgggatcaa gggcatcggg 600
 gagaagaccg ccctcaagct cctcaaggag tggggaagcc tggaaaacct cctcaagaac 660
 ctggaccggg taaagccaga aaatgtccgg gagaagatca aggccacct ggaagacctc 720
 aggtctctct tggagctctc ccgggtgcgc accgacctcc ccctggaggt ggacttcgcc 780
 aaaaggcggg agcccgaccg ggagaggctt agggcctttc tggagaggct tgagtgtggc 840
 agcctcctcc acgagttcgg ccttctggaa agccccaagg ccctggagga ggccccctgg 900
 ccccgccggg aaggggcctt cgtgggcttt gtgctttccc gcaaggagcc catgtgggcc 960
 gatcttctgg ccctggcgcg cgccaagggt ggccgggtcc accgggcccc cgagccttat 1020
 aaagccctca gggacttgaa ggaggcgcgg gggcttctcg ccaaagacct gagcgttctg 1080
 gccctaaggg aaggccttgg cctcccgcgc ggcgacgacc ccatgctcct cgcctacctc 1140
 ctggaccctt ccaacaccac ccccgagggg gtggcccggc gctacggcgg ggagtggacg 1200
 gaggaggcgg gggagcgggc cgccctttcc gagaggctct tcgccaacct gtgggggagg 1260
 cttgaggggg aggagaggct cctttggctt taccgggagg tggagaggcc cctttccgct 1320
 gtcttgccc acatggaggc cacgggggtg cgctggacg tggcctatct cagggccttg 1380
 tccctggagg tggccgagga gatcgccgcg ctcgaggccg aggtcttccg cctggccggc 1440

219

```

cacccttca acctcaactc ccgagaccag ctggaaaggg tcctctttga cgagctaggg 1500
cttcccgcc tggcaagac ggagaagacc ggcaagcgct ccaccagcgc cgcgctcctg 1560
gaggccctcc gcgaggccca ccccatcgtg gagaagatcc tgcagtaccg ggagctcacc 1620
aagctgaaga gcacctacat tgaccccttg ccggacctca tccaccccag gacggggccgc 1680
ctccacaccc gcttcaacca gacggccacg gccacgggca ggctaagtag ctccgatccc 1740
aacctccaga acatccccgt ccgcaccccg ctcgggcaga ggateccgcg ggccttcac 1800
gccgaggggg ggtggctatt ggtggctcctg gactatagcc agatggagct cagggtgctg 1860
gcccacctct ccggcgacga gaacctgac cgggtcttcc aggagggggc ggacatccac 1920
acggaaaccg ccagctggat gttcggcgtc ccccgggagg ccgtggaccc cctgatgcgc 1980
cgggcggcca agaccatcaa cttcgggggtt ctctacggca tgcgggcca ccgcctctcc 2040
caggagctag ccatccctta cgaggaggcc caggccttca ttgagcgcta ctttcagagc 2100
ttccccaagg tgcgggcctg gattgagaag accctggagg agggcaggag gcgggggtac 2160
gtggagaccc tcttcggccg ccgcccgtac gtgccagacc tagaggcccg ggtgaagagc 2220
gtgcgggagg cggccgagcg catggccttc aacatgcccg tccagggcac cgcgcgcgac 2280
ctcatgaagc tggctatggt gaagctcttc cccaggctgg aggaaacggg ggccaggatg 2340
ctccttcagg tccacgacga gctggctctc gaggcccca aagagagggc ggaggccgtg 2400
gcccggtg ccaaggaggt catggagggg gtgtatcccc tggccgtgcc cctggaggtg 2460
gaggtgggga taggggagga ctggctctcc gccaaaggag ga 2502

```

```

<210> 101
<211> 833
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 101
Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1          5          10          15

Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Leu
20          25          30

Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala Lys
35          40          45

Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Val Phe Val
50          55          60

```

220

Val	Phe	Asp	Ala	Lys	Ala	Pro	Ser	Phe	Arg	His	Glu	Ala	Tyr	Glu	Ala	
65					70					75					80	
Tyr	Lys	Ala	Gly	Arg	Ala	Pro	Thr	Pro	Glu	Asp	Phe	Pro	Arg	Gln	Leu	
				85					90					95		
Ala	Leu	Ile	Lys	Glu	Leu	Val	Asp	Leu	Leu	Gly	Phe	Thr	Arg	Leu	Glu	
			100					105					110			
Val	Gln	Gly	Tyr	Glu	Ala	Asp	Asp	Val	Leu	Ala	Thr	Leu	Ala	Lys	Lys	
		115					120					125				
Ala	Glu	Lys	Glu	Gly	Tyr	Glu	Val	Arg	Ile	Leu	Thr	Ala	Asp	Arg	Asp	
	130					135					140					
Leu	Tyr	Gln	Leu	Val	Ser	Asp	Arg	Val	Ala	Val	Leu	His	Pro	Glu	Gly	
145					150					155					160	
His	Leu	Ile	Thr	Pro	Glu	Trp	Leu	Trp	Glu	Lys	Tyr	Gly	Leu	Arg	Pro	
				165					170					175		
Glu	Gln	Trp	Val	Asp	Phe	Arg	Ala	Leu	Val	Gly	Asp	Pro	Ser	Asp	Asn	
			180					185						190		
Leu	Pro	Gly	Ile	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu	
		195					200					205				
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Val	
	210					215					220					
Lys	Pro	Glu	Asn	Val	Arg	Glu	Lys	Ile	Lys	Ala	His	Leu	Glu	Asp	Leu	
225					230					235					240	
Arg	Leu	Ser	Leu	Glu	Leu	Ser	Arg	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu	
				245					250					255		
Val	Asp	Phe	Ala	Lys	Arg	Arg	Glu	Pro	Asp	Arg	Glu	Arg	Leu	Arg	Ala	
			260					265					270			
Phe	Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	
		275					280					285				
Leu	Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Glu	
	290					295					300					
Gly	Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	
305					310					315					320	
Asp	Leu	Leu	Ala	Leu	Ala	Ala	Ala	Lys	Gly	Gly	Arg	Val	His	Arg	Ala	
				325					330					335		
Pro	Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	
			340					345					350			
Leu	Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	
		355					360					365				

Pro	Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser
	370					375					380				
Asn	Thr	Thr	Pro	Glu	Gly	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr
385					390					395					400
Glu	Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn
				405					410					415	
Leu	Trp	Gly	Arg	Leu	Glu	Gly	Glu	Glu	Arg	Leu	Leu	Trp	Leu	Tyr	Arg
			420					425					430		
Glu	Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr
		435					440					445			
Gly	Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val
	450					455					460				
Ala	Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly
465					470					475					480
His	Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe
			485						490					495	
Asp	Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys
			500					505					510		
Arg	Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro
		515					520					525			
Ile	Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser
	530					535					540				
Thr	Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg
545					550					555					560
Leu	His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser
				565				570						575	
Ser	Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly
			580					585					590		
Gln	Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Gly	Gly	Trp	Leu	Leu	Val
		595					600					605			
Val	Leu	Asp	Tyr	Ser	Gln	Met	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser
		610				615					620				
Gly	Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His
625					630					635					640
Thr	Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp
				645					650					655	
Pro	Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr
			660					665					670		

Gly Met Ser Ala His Arg Leu Ser Gln Glu Leu Ala Ile Pro Tyr Glu
675 680 685

Glu Ala Gln Ala Phe Ile Glu Arg Tyr Phe Gln Ser Phe Pro Lys Val
690 695 700

Arg Ala Trp Ile Glu Lys Thr Leu Glu Glu Gly Arg Arg Arg Gly Tyr
705 710 715 720

Val Glu Thr Leu Phe Gly Arg Arg Arg Tyr Val Pro Asp Leu Glu Ala
725 730 735

Arg Val Lys Ser Val Arg Glu Ala Ala Glu Arg Met Ala Phe Asn Met
740 745 750

Pro Val Gln Gly Thr Ala Ala Asp Leu Met Lys Leu Ala Met Val Lys
755 760 765

Leu Phe Pro Arg Leu Glu Glu Thr Gly Ala Arg Met Leu Leu Gln Val
770 775 780

His Asp Glu Leu Val Leu Glu Ala Pro Lys Glu Arg Ala Glu Ala Val
785 790 795 800

Ala Arg Leu Ala Lys Glu Val Met Glu Gly Val Tyr Pro Leu Ala Val
805 810 815

Pro Leu Glu Val Glu Val Gly Ile Gly Glu Asp Trp Leu Ser Ala Lys
820 825 830

Glu

<210> 102
<211> 2499
<212> DNA
<213> Artificial Sequence

<220>
<223> cloned sequence

<400> 102
atggcgatgc ttccctcttt tgagcccaaa ggccgggtcc tcttggtgga cggccaccac 60
ctggcctacc gcaccttctt cgccctgaag ggcctcacca cgagtcgggg cgaaccggtg 120
caggcgggtct acggcttcgc caagagcctc ctcaaggccc tgaaggagga cgggtacaag 180
gccatcttcg tggctcttga cgccaaggcc ccctccttcc gccacgaggc ccacgaggcc 240
tacaaggcgg ggagggcccc gagccccgag gacttcccc ggcagctcgc cctcatcaag 300
gagctggtgg acctcctggg gtttaccgc ctcgaggtcc aaggctacga ggcggacgac 360
gtcctcgcca ccctggccaa gaaggcggaa aaagaagggt acgaggtgcg catcctcacc 420

gccgaccggg	acctctacca	gctcgtctcc	gaccgcgtcg	ccgtcctcca	ccccgagggc	480
cacctcatca	ccccggagtg	gctttgggag	aagtacggcc	tcaggccgga	gcagtgggtg	540
gacttccgcg	ccctcgtggg	ggaccctctc	gacaacctcc	ccgggggtcaa	gggcatcggg	600
gagaagaccg	ccctcaagct	cctcaaggag	tggggaagcc	tggaaaacct	cctcaagaac	660
ctggaccggc	tgaagcccg	catccgggag	aagatcctgg	cccacatgga	cgatctgaag	720
ctctcctggg	acctggccaa	ggtgcgccac	gacctgcccc	tggaggtgga	cttcgccaaa	780
aggcgggagt	ccgatcggga	gaggcttagg	gcctttcttg	agaggcttga	gtttggcagc	840
ctcctccacg	agttcggcct	tctggaaagc	cccaaggccc	tggaggaggc	cccctggccc	900
ccgccggtag	gggccttcgt	gggctttgtg	ctttcccgca	aggagcccat	gtgggccgat	960
cttctggccc	tggccgccgc	caggggtggg	cgggtccacc	gggccccga	gccttataaa	1020
gccctcagag	acctgaagga	ggcgcggggg	cttctcgcca	aagacctgag	cgttctggcc	1080
ctgaggggaag	gccttggcct	cccgcccggc	gacgacccca	tgtcctcgc	ctacctctg	1140
gacccttcca	acaccacccc	cgaggtgggt	gcccggcgct	acggcgggga	gtggacggag	1200
gaggcggggg	agcgggcccgc	cctttccgag	aggctcttcg	ccaacctgtg	ggggaggctt	1260
gagggggagg	ggaggtcct	ttggctttac	cgggggggtg	agaggccct	ttccgctgtc	1320
ctggcccaca	tggaggccac	aggggtgcgc	ctggacgtgg	cctatctcag	ggccttgtec	1380
ctggaggtgg	ccgaggagat	cgcccgcctc	gaggccgagg	tcttccgcct	ggccggccac	1440
cccttcaacc	tcaactcccg	ggaccagctg	gaaagggctc	tctttgacga	gctagggctt	1500
cccgccatcg	gcaagacgga	gaagaccggc	aagcgctcca	ccagcgccgc	cgtcctggag	1560
gccctccgcg	aggcccaccc	catcgtggag	aagatcctgc	agtaccggga	gctcaccaag	1620
ctgaagagca	cttacattga	ccccttgccg	gacctcatcc	accccaggac	gggccgcctc	1680
cacacccgct	tcaaccagac	ggccacggcc	acgggcaggc	taagtagctc	cgatcccaac	1740
ctccagaaca	tccccgtccg	caccccgcctc	gggcagagga	tccgccgggc	cttcacgcgc	1800
gagggggggg	ggctattggg	ggtcctggac	tatagccaga	tggagctcag	ggtgctggcc	1860
cacctctccg	gcgacgagaa	cctgatccgg	gtcttccagg	aggggcggga	catccacacg	1920
gaaaccgcca	gctggatggt	cggcgctccc	cgggaggccg	tggacccct	gatgcgccgg	1980
gcggccaaga	ccatcaactt	cgggggttctc	tacggcatgt	cggcccaccg	cctctcccag	2040
gagctagcca	tcccttacga	ggaggcccag	gccttcattg	agcgctactt	ccaaagcttc	2100
cccaaggtgc	gggcctggat	agaaaagacc	ctggaggagg	ggaggaagcg	gggctacgtg	2160

```

gaaaccctct tcggaagaag gcgctacgtg cccgacctca acgcccgggt gaagagcgtc 2220
agggaggccg cggagcgcat ggccttcaac atgcccgtcc agggcaccgc cgccgacctc 2280
acgaagctgg ctatggtgaa gctcttcccc aggctggagg aaacgggggc caggatgctc 2340
cttcagggtcc acgacgagct ggtcctcgag gccccaaaag agagggcgga ggccgtggcc 2400
cggctggcca aggaggtcat ggaggggggtg tatcccttgg ccgtgcccct ggaggtggag 2460
gtgggggatag gggaggactg gctttccgcc aagggttag 2499

```

```

<210> 103
<211> 832
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> cloned sequence

```

```

<400> 103
Met Ala Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu Val
1          5          10          15
Asp Gly His His Leu Ala Tyr Arg Thr Phe Phe Ala Leu Lys Gly Leu
20          25          30
Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala Lys
35          40          45
Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Tyr Lys Ala Ile Phe Val
50          55          60
Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala His Glu Ala
65          70          75          80
Tyr Lys Ala Gly Arg Ala Pro Ser Pro Glu Asp Phe Pro Arg Gln Leu
85          90          95
Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Phe Thr Arg Leu Glu
100         105         110
Val Gln Gly Tyr Glu Ala Asp Asp Val Leu Ala Thr Leu Ala Lys Lys
115         120         125
Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Arg Asp
130         135         140
Leu Tyr Gln Leu Val Ser Asp Arg Val Ala Val Leu His Pro Glu Gly
145         150         155         160
His Leu Ile Thr Pro Glu Trp Leu Trp Glu Lys Tyr Gly Leu Arg Pro
165         170         175
Glu Gln Trp Val Asp Phe Arg Ala Leu Val Gly Asp Pro Ser Asp Asn

```

225

180							185					190				
Leu	Pro	Gly	Val	Lys	Gly	Ile	Gly	Glu	Lys	Thr	Ala	Leu	Lys	Leu	Leu	
		195					200					205				
Lys	Glu	Trp	Gly	Ser	Leu	Glu	Asn	Leu	Leu	Lys	Asn	Leu	Asp	Arg	Leu	
	210					215					220					
Lys	Pro	Ala	Ile	Arg	Glu	Lys	Ile	Leu	Ala	His	Met	Asp	Asp	Leu	Lys	
225					230					235					240	
Leu	Ser	Trp	Asp	Leu	Ala	Lys	Val	Arg	Thr	Asp	Leu	Pro	Leu	Glu	Val	
				245					250					255		
Asp	Phe	Ala	Lys	Arg	Arg	Glu	Ser	Asp	Arg	Glu	Arg	Leu	Arg	Ala	Phe	
			260					265					270			
Leu	Glu	Arg	Leu	Glu	Phe	Gly	Ser	Leu	Leu	His	Glu	Phe	Gly	Leu	Leu	
		275					280					285				
Glu	Ser	Pro	Lys	Ala	Leu	Glu	Glu	Ala	Pro	Trp	Pro	Pro	Pro	Val	Gly	
	290					295					300					
Ala	Phe	Val	Gly	Phe	Val	Leu	Ser	Arg	Lys	Glu	Pro	Met	Trp	Ala	Asp	
305					310					315					320	
Leu	Leu	Ala	Leu	Ala	Ala	Ala	Arg	Gly	Gly	Arg	Val	His	Arg	Ala	Pro	
				325					330					335		
Glu	Pro	Tyr	Lys	Ala	Leu	Arg	Asp	Leu	Lys	Glu	Ala	Arg	Gly	Leu	Leu	
			340					345					350			
Ala	Lys	Asp	Leu	Ser	Val	Leu	Ala	Leu	Arg	Glu	Gly	Leu	Gly	Leu	Pro	
		355					360					365				
Pro	Gly	Asp	Asp	Pro	Met	Leu	Leu	Ala	Tyr	Leu	Leu	Asp	Pro	Ser	Asn	
	370					375					380					
Thr	Thr	Pro	Glu	Val	Val	Ala	Arg	Arg	Tyr	Gly	Gly	Glu	Trp	Thr	Glu	
385					390					395					400	
Glu	Ala	Gly	Glu	Arg	Ala	Ala	Leu	Ser	Glu	Arg	Leu	Phe	Ala	Asn	Leu	
				405					410					415		
Trp	Gly	Arg	Leu	Glu	Gly	Glu	Gly	Arg	Leu	Leu	Trp	Leu	Tyr	Arg	Gly	
			420					425					430			
Val	Glu	Arg	Pro	Leu	Ser	Ala	Val	Leu	Ala	His	Met	Glu	Ala	Thr	Gly	
		435					440					445				
Val	Arg	Leu	Asp	Val	Ala	Tyr	Leu	Arg	Ala	Leu	Ser	Leu	Glu	Val	Ala	
	450					455					460					
Glu	Glu	Ile	Ala	Arg	Leu	Glu	Ala	Glu	Val	Phe	Arg	Leu	Ala	Gly	His	
465					470					475					480	
Pro	Phe	Asn	Leu	Asn	Ser	Arg	Asp	Gln	Leu	Glu	Arg	Val	Leu	Phe	Asp	

485							490						495			
Glu	Leu	Gly	Leu	Pro	Ala	Ile	Gly	Lys	Thr	Glu	Lys	Thr	Gly	Lys	Arg	
			500				505						510			
Ser	Thr	Ser	Ala	Ala	Val	Leu	Glu	Ala	Leu	Arg	Glu	Ala	His	Pro	Ile	
		515				520						525				
Val	Glu	Lys	Ile	Leu	Gln	Tyr	Arg	Glu	Leu	Thr	Lys	Leu	Lys	Ser	Thr	
		530				535				540						
Tyr	Ile	Asp	Pro	Leu	Pro	Asp	Leu	Ile	His	Pro	Arg	Thr	Gly	Arg	Leu	
545					550				555			560				
His	Thr	Arg	Phe	Asn	Gln	Thr	Ala	Thr	Ala	Thr	Gly	Arg	Leu	Ser	Ser	
			565				570						575			
Ser	Asp	Pro	Asn	Leu	Gln	Asn	Ile	Pro	Val	Arg	Thr	Pro	Leu	Gly	Gln	
			580				585						590			
Arg	Ile	Arg	Arg	Ala	Phe	Ile	Ala	Glu	Gly	Gly	Trp	Leu	Leu	Val	Val	
		595				600						605				
Leu	Asp	Tyr	Ser	Gln	Met	Glu	Leu	Arg	Val	Leu	Ala	His	Leu	Ser	Gly	
610					615						620					
Asp	Glu	Asn	Leu	Ile	Arg	Val	Phe	Gln	Glu	Gly	Arg	Asp	Ile	His	Thr	
625					630				635			640				
Glu	Thr	Ala	Ser	Trp	Met	Phe	Gly	Val	Pro	Arg	Glu	Ala	Val	Asp	Pro	
			645				650						655			
Leu	Met	Arg	Arg	Ala	Ala	Lys	Thr	Ile	Asn	Phe	Gly	Val	Leu	Tyr	Gly	
			660				665						670			
Met	Ser	Ala	His	Arg	Leu	Ser	Gln	Glu	Leu	Ala	Ile	Pro	Tyr	Glu	Glu	
		675				680						685				
Ala	Gln	Ala	Phe	Ile	Glu	Arg	Tyr	Phe	Gln	Ser	Phe	Pro	Lys	Val	Arg	
690					695						700					
Ala	Trp	Ile	Glu	Lys	Thr	Leu	Glu	Glu	Gly	Arg	Lys	Arg	Gly	Tyr	Val	
705					710				715			720				
Glu	Thr	Leu	Phe	Gly	Arg	Arg	Arg	Tyr	Val	Pro	Asp	Leu	Asn	Ala	Arg	
			725				730						735			
Val	Lys	Ser	Val	Arg	Glu	Ala	Ala	Glu	Arg	Met	Ala	Phe	Asn	Met	Pro	
			740				745						750			
Val	Gln	Gly	Thr	Ala	Ala	Asp	Leu	Thr	Lys	Leu	Ala	Met	Val	Lys	Leu	
		755				760						765				
Phe	Pro	Arg	Leu	Glu	Glu	Thr	Gly	Ala	Arg	Met	Leu	Leu	Gln	Val	His	
770					775						780					
Asp	Glu	Leu	Val	Leu	Glu	Ala	Pro	Lys	Glu	Arg	Ala	Glu	Ala	Val	Ala	

227

785		790		795		800									
Arg	Leu	Ala	Lys	Glu	Val	Met	Glu	Gly	Val	Tyr	Pro	Leu	Ala	Val	Pro
				805					810					815	
Leu	Glu	Val	Glu	Val	Gly	Ile	Gly	Glu	Asp	Trp	Leu	Ser	Ala	Lys	Gly
			820					825					830		

<210> 104
 <211> 59
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<220>
 <221> misc_feature
 <222> (37)..(37)
 <223> n is dU-biotin

<400> 104
 tagctcggta acgccggctt ccgtcgcgac cacgttnttc gtggtcgcga cggaagccg 59

<210> 105
 <211> 60
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<220>
 <221> misc_feature
 <222> (38)..(38)
 <223> n is dU-biotin

<400> 105
 tagctcggta aacgccggct tccgtcgcga ccacgttntt cgtggtcgcg acggaagccg 60

<210> 106
 <211> 61
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<220>
 <221> misc_feature
 <222> (39)..(39)
 <223> n is dU-biotin

<400> 106

tagctcggat ttctgccggc ttccgtcgcg accacgttnt tcgtggtcgc gacggaagcc 60

g 61

<210> 107
 <211> 60
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<220>
 <221> misc_feature
 <222> (38)..(38)
 <223> n is dU-biotin

<400> 107
 tagctaccag ggctccggct tccgtcgcga ccacgttntt cgtggtcgcg acggaagccg 60

<210> 108
 <211> 67
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<220>
 <221> misc_feature
 <222> (22)..(22)
 <223> n is an abasic site

<220>
 <221> misc_feature
 <222> (45)..(45)
 <223> n is dU-biotin

<400> 108
 agctaccatg cctgcacgca gncggcatcc gtcgcgacca cgttnttcgt ggtcgcgacg 60

gatgccg 67

<210> 109
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<400> 109
 taatacgact cactataggg aga 23

<210> 110
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide template

<400> 110
attatgctga gtgatatccc tctatcgat

29

<210> 111
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 111
taatacgact cactataggg aga

23

<210> 112
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> oligonucleotide template

<220>
<221> misc_feature
<222> (24)..(24)
<223> n is a, t, c or g

<400> 112
attatgctga gtgatatccc tctngtca

28

<210> 113
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> primer

<400> 113
gcggtgtaga gacgagtgcg gag

23

<210> 114
<211> 50
<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 114

ctctcacaag cagccaggca agctccgcac tcgtctctac accgctccgc 50

<210> 115

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 115

ctctcacaag cagccaggca agctccgcac tcgtctctac accgctccgc 50

<210> 116

<211> 61

<212> DNA

<213> Artificial Sequence

<220>

<223> FITC primer

<220>

<221> misc_feature

<222> (39)..(39)

<223> n is dU-biotin

<400> 116

tagctaccat ttctgccggc ttccgtcgcg accacgttnt tcgtggtcgc gacggaagcc 60

g 61

<210> 117

<211> 67

<212> DNA

<213> Artificial Sequence

<220>

<223> FITC primer

<220>

<221> misc_feature

<222> (45)..(45)

<223> n is dU-biotin

<400> 117

tagctaccat ttttttttttc gccggcttcc gtcgcgacca cgttnttcgt ggtcgcgacg 60

gaagccg

67

<210> 118
 <211> 61
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> ELISA primer

<220>
 <221> misc_feature
 <222> (39)..(39)
 <223> n is dU-biotin

<400> 118
 tagctaccag gggctccggc ttccgtcgcg accacgttnt tcgtggtcgc gacggaagcc 60

g 61

<210> 119
 <211> 60
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> ELISA primer

<220>
 <221> misc_feature
 <222> (38)..(38)
 <223> n is dU-biotin

<400> 119
 tagctcggta aacgccggct tccgtcgcga ccacgttntt cgtggtcgcg acggaagccg 60

<210> 120
 <211> 67
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> primer

<220>
 <221> misc_feature
 <222> (22)..(22)
 <223> n is an abasic site

<220>
 <221> misc_feature
 <222> (45)..(45)
 <223> n is dU-biotin

<400> 120
agctaccatg cctgcacgca gncggcatcc gtcgcgacca cgttnttcgt ggtcgcgacg 60
gatgccg 67

<210> 121
<211> 155
<212> DNA
<213> Artificial Sequence

<220>
<223> Sequence amplified from ancient DNA

<400> 121
accccatgca tataggcatg tacatattat gcttgatctt acatgaggac ttacatctca 60
aaagtttatt tcaagtgtat agtctgtaag catgtatttc acttagtcca ggagcttaat 120
caccaggcct cgagaaacca gcaacccttg cgagt 155

**THE EMBODIMENTS FOR WHICH AN EXCLUSIVE PROPERTY AND
PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:**

1. A pol A DNA polymerase with an expanded substrate range, which is capable
5 of abasic site bypass, wherein the polymerase exhibits at least 95% identity to the
amino acid sequence set forth in SEQ ID NO: 80 and comprises one or more of the
following amino acids: V at position 42, K at position 76, R at position 82, Q at
position 90, G at position 144, Q at position 335, T at position 447, R at position 481,
V at position 609, M at position 615, S at position 684, M at position 738, G at
10 position 743 with reference to the amino acid positions of SEQ ID NO: 80.
2. A pol A DNA polymerase with an expanded substrate range, which is capable
of abasic site bypass, wherein the polymerase exhibits at least 95% identity to the
amino acid sequence set forth in SEQ ID NO: 84 and comprises one or more of the
15 following amino acids: A at position 25, Q at position 114, N at position 191, L at
position 265, A at position 387, D at position 435, N at position 552, V at position
609, G at position 798, K at position 821 with reference to the amino acid positions of
SEQ ID NO: 84.
- 20 3. A pol A DNA polymerase with an expanded substrate range, which is capable
of abasic site bypass, wherein the polymerase exhibits at least 95% identity to the
amino acid sequence set forth in SEQ ID NO: 80 or the amino acid sequence set for in
SEQ ID NO:84 and comprises the amino acid V at position 609 with reference to the
amino acid positions of SEQ ID NO: 80 or SEQ ID NO: 84.
- 25 4. A pol A DNA polymerase with an expanded substrate range, which is capable
of abasic site bypass, wherein the polymerase exhibits at least 98% identity to one or
more of the amino acid sequences set forth in SEQ ID NO: 80 and SEQ ID NO: 84.
- 30 5. A pol A DNA polymerase according to any one of claims 1 to 4, wherein the
DNA polymerase comprises the amino acid sequence as set forth in SEQ ID NO: 80
or SEQ ID NO: 84.

6. A pol A DNA polymerase according to claim 5, wherein the DNA polymerase which consists of the amino acid sequence as set forth in SEQ ID NO: 80 or SEQ ID NO: 84.
- 5 7. A nucleic acid construct encoding the pol A DNA polymerase according to any one of claims 1 to 6.
8. A vector comprising a nucleic acid construct according to claim 7.
- 10 9. Use of one or more of the pol A DNA polymerase according to any one of claims 1 to 6 in any one or more of the following applications selected from the group consisting of: PCR amplification, sequencing of damaged DNA templates, the incorporation of unnatural base analogues into DNA and the creation of novel polymerase activities.
- 15 10. The use according to claim 9 of a blend of the pol A DNA polymerases is used.

1 / 29

Figure 1a shows M1 nucleic acid sequence

ATGCTCCCTCTTTTTGAGCCCAAAGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCT
ACCGCACCTTCCACGCCCTGAAGGGCCTCACCACCAGCCGGGGGGAGCCGGTGCAGGCGG
TCTACGGCTTCGCCAAGAGCCTCCTCAAGGCCCTCAAGGAGGACGGGGGACGCGGTGATCG
TGGTCTTTGACGCCAAGGCCCCCTCCTTCCGCCACGAGGCCTACGGGGGGGTACAAGGCGG
CCCGGGCCCCCACGCCGGAGGACTTTCCCCGGCAACTCGCCCTCATCAAGGAGCTGGTGG
ATCTCCTGGGGCTGGCGCGCCTCGAGGTCCCGGGCTACGAGGCGGACGACGTCCTGGCCA
GCCTGGCCAAGAAGGCGGAAAAGGAGGGCTACGAGGTCCGCATCCTCACCGCCGACAAA
GGCCTTTACCAGCTCCTTTCCGACCGCATCCACGTCTCTCCACCCCGAGGGGTACCTCATCA
CCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCAAGTGGGCGGACTACCGGG
CCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGGCATCGGGGAGAAGACG
GCGAGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCAAGAACCTGGACCG
GCTGAAGCCCGCCATCCGGGAGAAGATCCTGGCCCACATGGACGATCTGAAGCTCTCCTG
GGATCTGGCCAAGGTGCGCACCGACCTGCCCCTGGAGGTGGACTTCGCCAAAAGGCGGGA
GCCCCACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCA
CGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGA
AGGGGCCTTCGTGGGCTTTGTCCTTTCCCGCAGGGAGCCCATGTGGGCGGATCTTCTGGCC
CTGGCCGCCGCCAGGGGGGGGCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCAGG
GACCTGAAGGAGGCGCGGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGAA
GGCCTTGGCCTCCCGCCCCGGCGACGACCCCATGCTCCTCGCCTACCTCCTGGACCTTCCA
ACACCACCCCGAGGGGGTGGCCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGG
GAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAG
GAGAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGGCCCTTTCCGCTGTCTGGCCACA
TGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCTGGAGGTGG
CCGAGGAGATCGCCCCGCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACC
TCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCG
GCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCTGGGGGCCCTCCGC
GAGGCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAGC
ACCTACATTGACCCCTTACCGGACCTCATCCACCCAGGACGGGCGCCTCCACACCCGCT
TCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCAGAACA
TCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGGCCTTCATCGCCGAGGAGGGGT
GGCTATTGGTGGTCTCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCG
GCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGGCGGGACATCCACACGGAGACCGCC
AGCTGGATGTTTCGGCGTCCCCCGGGAGGCCGTGGACCCCTGATGCGCCGGGCGGCCAAG
ACCATCAACTTCGGGGTCTCTACGGCATGTGCGGCCACCGCCTCTCCCAGGAGCTAGCCA
TCCCTTACGAGGAGGCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCG
GGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCT
TCGGCCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGGGGCG
GCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCGACCTCATGAAGCTG
GCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGGCCAGGATGCTCCTTCAGGTC
CACGACGAGCTGGTCTCTGAGGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGC
CAAGGAGGTCATGGAGGGGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGAT
AGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

FIG. 1a

SEQ 1

M1 amino acid sequence:

MRGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEDGDAVIVVFDKAPSF
RH EAYGGYKAARAPTEPDFPRQLALIKELVDLLGLARLEVPGYEADDVLA SLAKKAEKEGYEVRLTADKGLYQLLS
DRIHV LHP EGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGECTARKLLEEWGSLEALLKNLDR
LK PAIREKILAHMDDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLERFGLSLLHEFGLLLESPKALEE
APWP PEGAFVGFVLSRREP MWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREG
LGLPPGDDPML LAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERLFANLWGRLEGEERLLWLYRE
VERPLSAVL AHMEATGVR LDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLSRDQLERVLFDELGL
PAIGKTEKTGKRSTSAAVLGALREA HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGR
LHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRRIRRAFIA EEGWLLVVLDYSQIELRVLAHLSG
DENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSA HRLSQELAI
PYEEAQAFIERYFQSFPKVR A WIEKTL EEGRRRGYVETLFGRRRYVPDLEARVKS
VRGAAERMAFN MPVQGTAAADLMK LAMV KLFPRLEEMGARMLLQVHDELVLEAPKERA
EAVARLAK EVM EGVYPLAVPLEVEVGIGE DWLSAKE

FIG. 1b

3 / 29

Figure 2a shows M4 nucleic acid sequence

ATGCTCCCTCTTTATGAGCCCAAGGGCCGCGTCCTCCTGGTGGACGGCCACCACCTGGCCT
ACCGCACCTTCCACGCCCTGAAGGGCCTCACCACCAGCCGGGGGGAGCCGGTGCAGGCGG
TCTACGGCTTCGCCAAGAGCCTCCTCAAGGCCCTCAAGGAGGGCGGGGACGCGGTGATCG
TGGTCTTTGACGCCAAGGCCCCCTCCTTCCCCCATGAGGCCTACGGGGGGTACAAGGCGG
GCCGGGCCCCCACGCCGGAGGACTTTCCCCGACAACCTCGCCCTCATCAAGGAGCTGGTGG
ACCTCCTGGGGCTGACGCGCCTCGAGGTCCCAGGGCTACGAGGCGGACGACGTCCTGGCCA
GCCTGGCCAAGAAGGCGGAAAAGGAGGGCTACGAGGTCCGCATCCTCACCGCCGACAAA
GACCTTTACCAGCTCCTTTCCGACCGCATCCACGTCCTCCACCCCGAGGGGTACCTCATCA
CCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCAGTGGGGCCGACTACCGGG
CCCTGACCGGGGACGAGTCCGACAACCTTCCCAGGGGTCAAGGGCATCGGGGAGAAGACG
GCGAGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAGCCCTCCTCAAGAACCTGGACCG
GCTGAAGCCCGCCATCCGGGAGAAGATCCTGGCCACATGGACGATCTGAAGCTCTCCTG
GGACCGGGCCAAGGTGCGCACCGACCTGCCCCCTGGAGGTGGACTTCGCCAAAAGGCGGG
AGCCCGACCGGGAGAGGCTTAGGGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCC
ACGAGTTCGGCCTTCTGGAAAGCCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGG
AAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCGGATCTTCTAGC
CCTGGCCGCGCCAGGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCGG
GGACCTGAAGGAGGCGCGGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCCTGAGGGA
AGGCCTTGGCCTCCCGCCCCGACGACGACCCCATGCTCCTCGCCTACCTCCTGGACCCTTCC
AACACCACCCCCGAGGGGGTGGCCCCGGCGCTACGGCGGGGAGTGGACGGAGGAGGCAGG
GGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGA
GGAAAGGCTCCTTTGGCTTTACCGGGAGGTGGAGAGGGCCCTTTCCGCTGTCCTGGCCAC
ATGGAGGCCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCTTGGAGGTG
GCCGAGGAGATCGCCCCGCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAAC
CTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATC
GGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCTGGGGGCCCTCCG
CGAGGCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTGAAGAG
CACCTACATTGACCCCTTGCCGGACCTCATCCACCCAGGACGGGGCCGCTCCACACCCGC
TTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCGATCCCAACCTCCAGAGC
ATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGG
TGGCTATTGGTGGCCCTGGACTATAGCCAGATAGAGCTCAGGGTGCTGGCCACCTCTCCG
GCGACGAGAACCTGATCCGGGTCTTCCAGGAGGGGGCGGGACATCCACACGGAGACCGCC
AGCTGGATGTTTCGGCGTCCCCCGGGAGGGCGTGGACCCCTGATGCGCCGGGCGGCCAAG
ACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCACCGCCTCTCCCAGGAGCTAGCCA
TCCCTTACGAGGAGGCCAGGCCTTCATTAAGCGCTACTTTCAGAGCTTCCCCAAGGTGCG
GGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCCTCT
TCGGCCGCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGCGGGAGCCG
GCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGTACCGCCGCCGACCTCATGAAGCTG
GCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTC
CACGACGAGCTGGTCCTCGAGGCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGC
CAAGGAGGTCATGGAGGGGGTGTATCCCCTGGCCGTGCCCTGGAGGTGGAGGTGGGGAT
AGGGGAGGACTGGCTCTCCGCCAAGGAGTGA

FIG. 2a

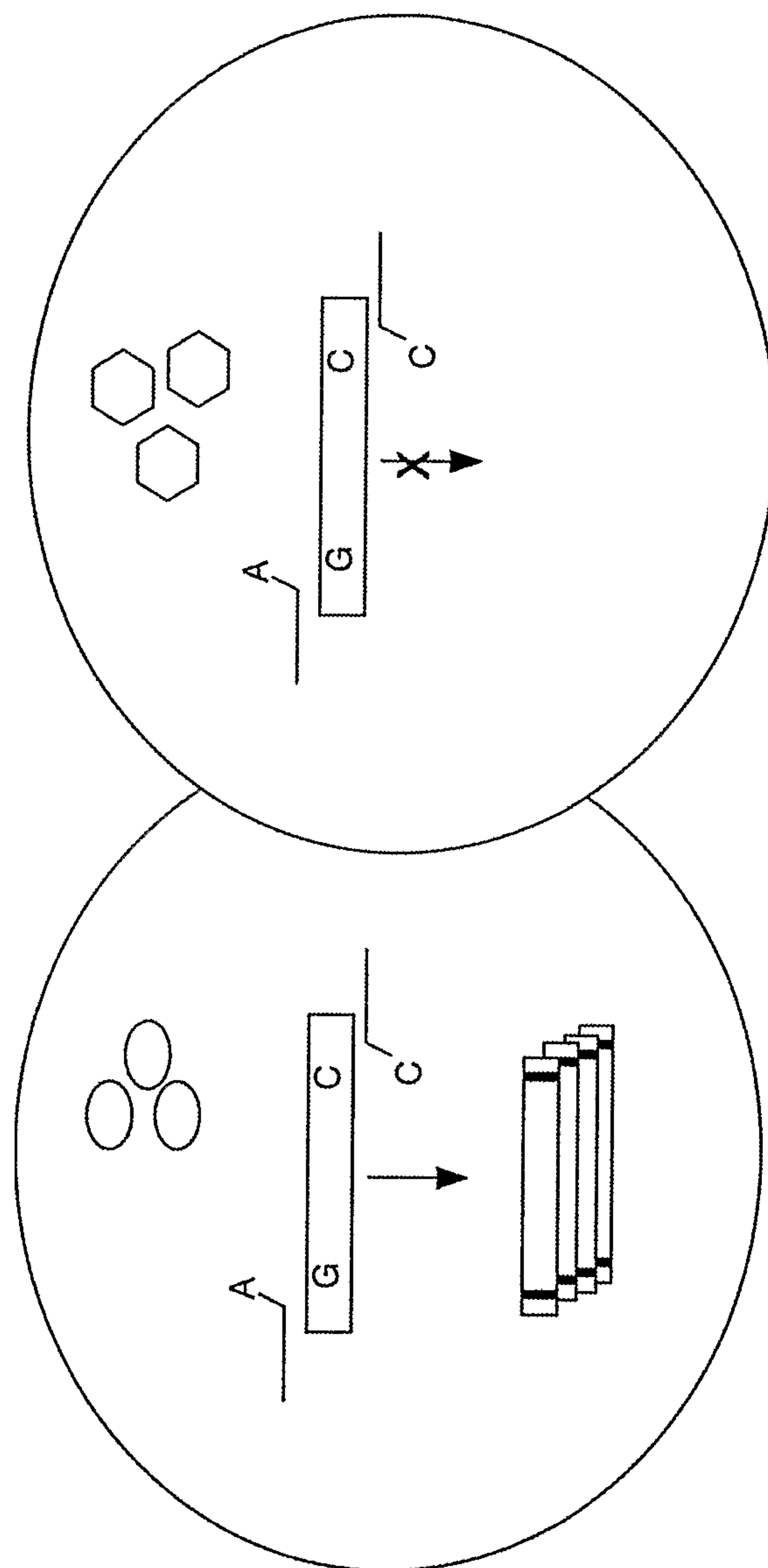
SEQ 2

M4 amino acid sequence:

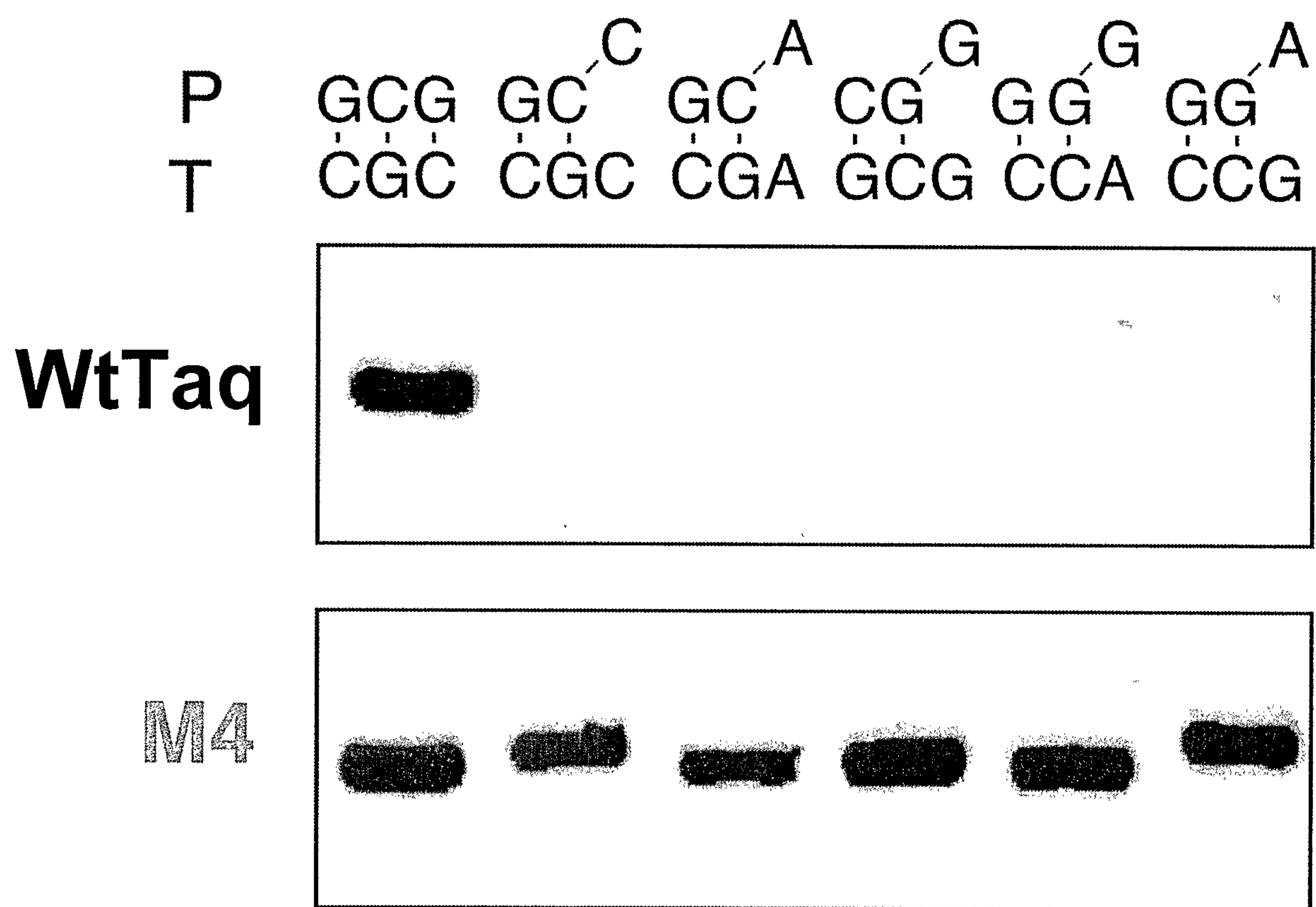
MRGMLPLYEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFAKSLLKALKEGGDAVIVVFDAKAPSFPH
EAYGGYKAGRPTPEDFPRQLALIKELVDLLGLTRLEVPGYEADDVLA SLAKKAKEGYEVRILTADKDL YQLLS
DRIHV LHP EGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIG EKTARKLL EEWGSLEAL LKNLDR LK
PAIREKILAHMDDDLKLSWDR AKVRTDLPLEVDFAKRREPDRERLRAFLEF GSL LHEFGLL ESPKALEEAPWP
PPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALGDLKEARGLLAKDLSVLALREGLGLPPDDDPML
LAYLLDPSNTTPEGVARRYGGGEWTEEAGERAA LSERLFANLWGRLEGEERLLWLYREVERPLSAVL AHMEATGVR
LDVA YLRALSLEVAEEIARLEAEVFRLAGHPFNLSRDQLERVLFD ELGLPAIGKTEKTGKRSTSAAVLGALREA
HPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGR LSSSDPNLQSI PVRTPLGQRIRRAFIA
EEGWLLVALDYSQIELRVL AHLSGDENLIRVFQEGRDIHTETASWMFGVPREA VDP LMRRAAKTINFGVLYGMSA
HRLSQELAI PYEEAQAFIKRYFQSF PKVRAWIEK TLEEGRRRGYVETLFGRRRYVPDLEARVKS VREPAERMAFN
MPVQGTAAADLMK LAMVVKLFPRLEEMGARM LLQVHDEL VLEAPKERAEA VARLAKEVM EGVVYPLA VPLEVEVGIGE
DWLSAKE

FIG. 2b

5 / 29



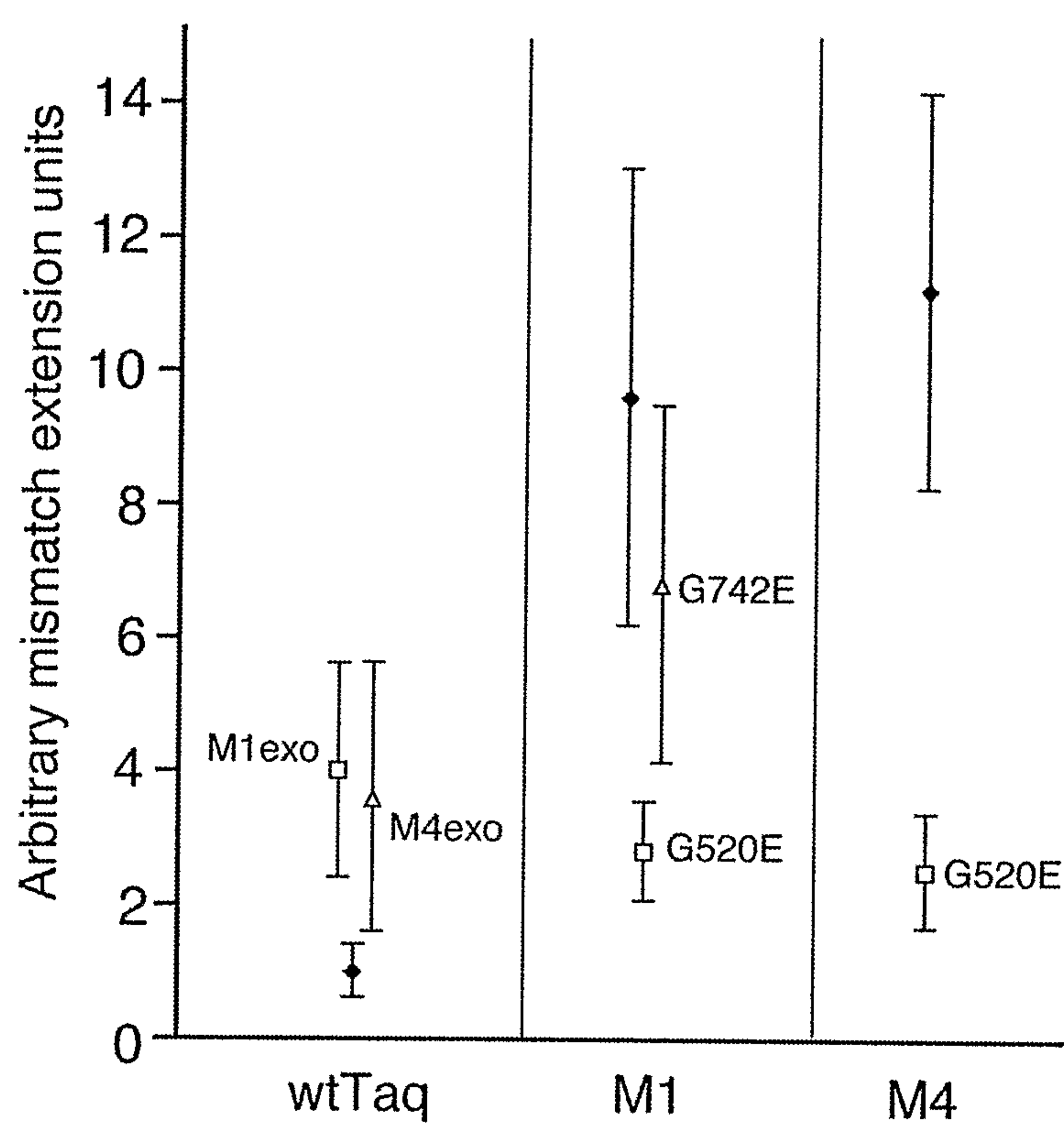
6 / 29



Mismatch extension properties of selected polymerases.

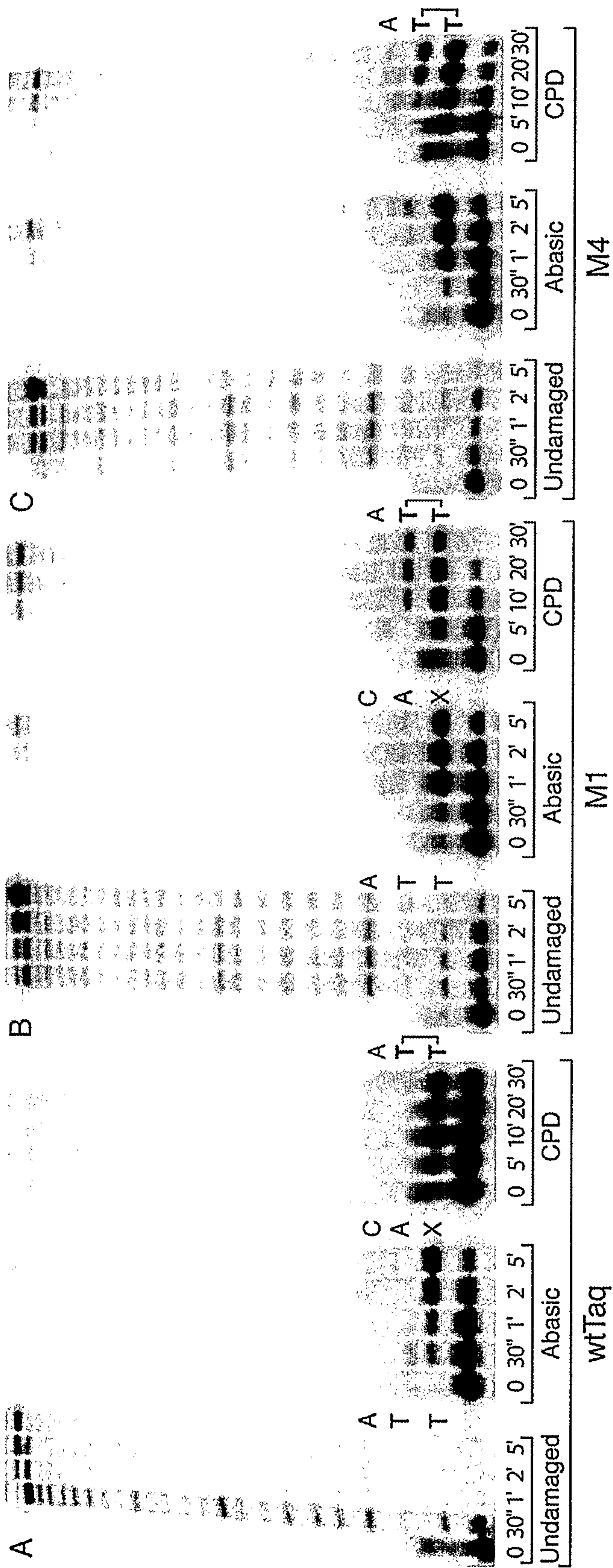
FIG. 4a

7 / 29



Mismatch extension properties of selected polymerases.

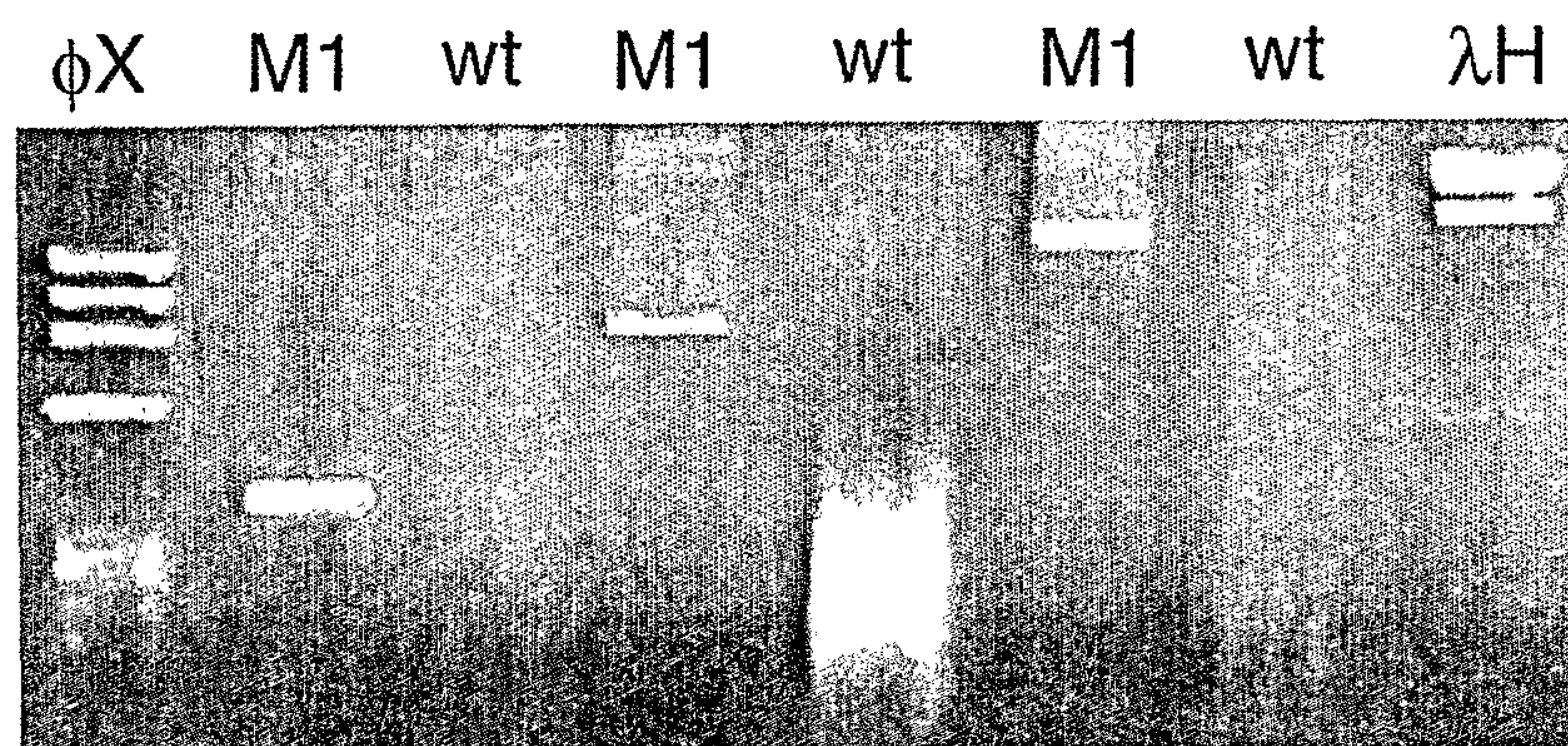
FIG. 4b



Lesion bypass activity

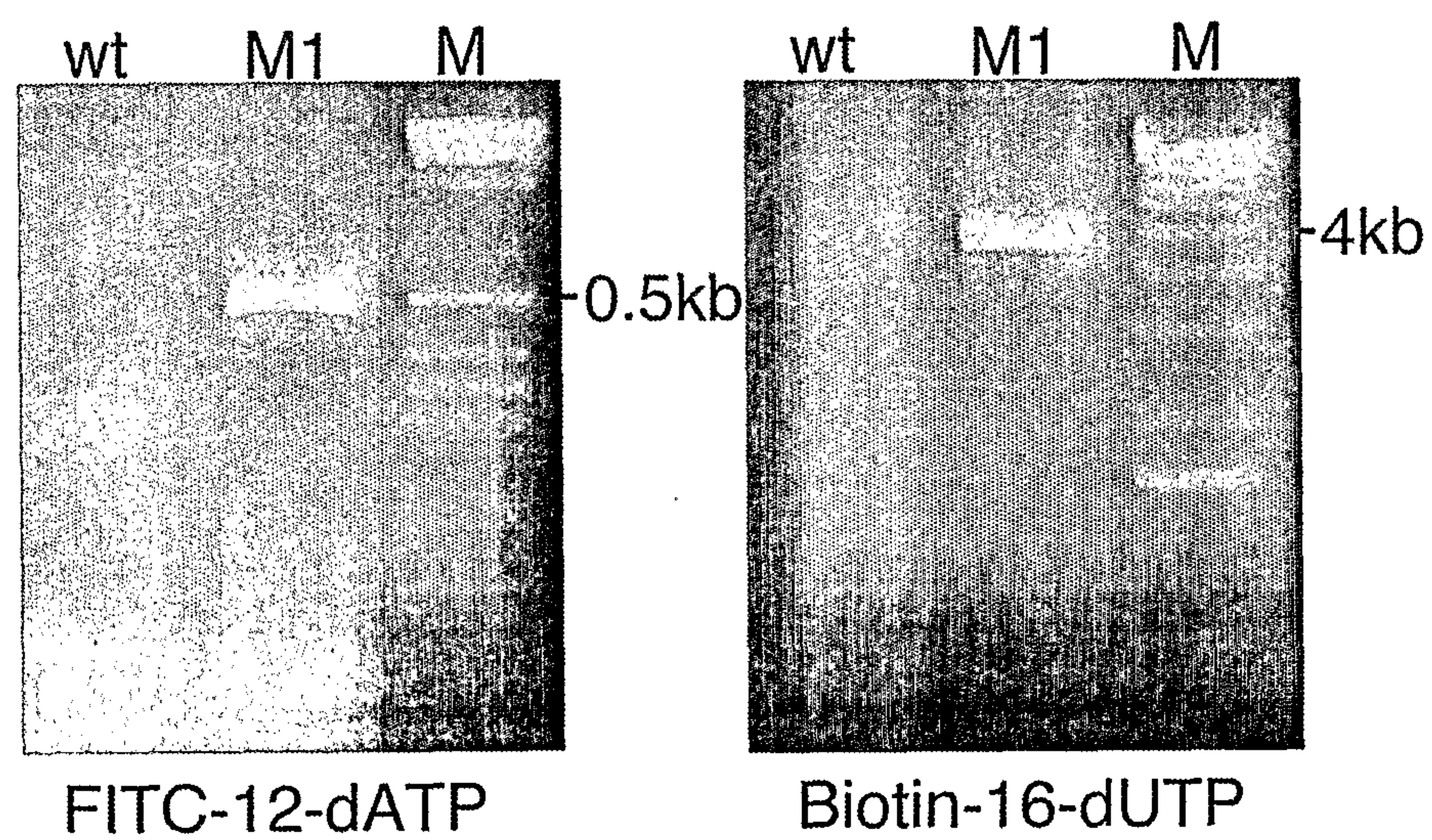
FIG. 5

9 / 29



Polymerase activity on unnatural substrates (α S dNTP).

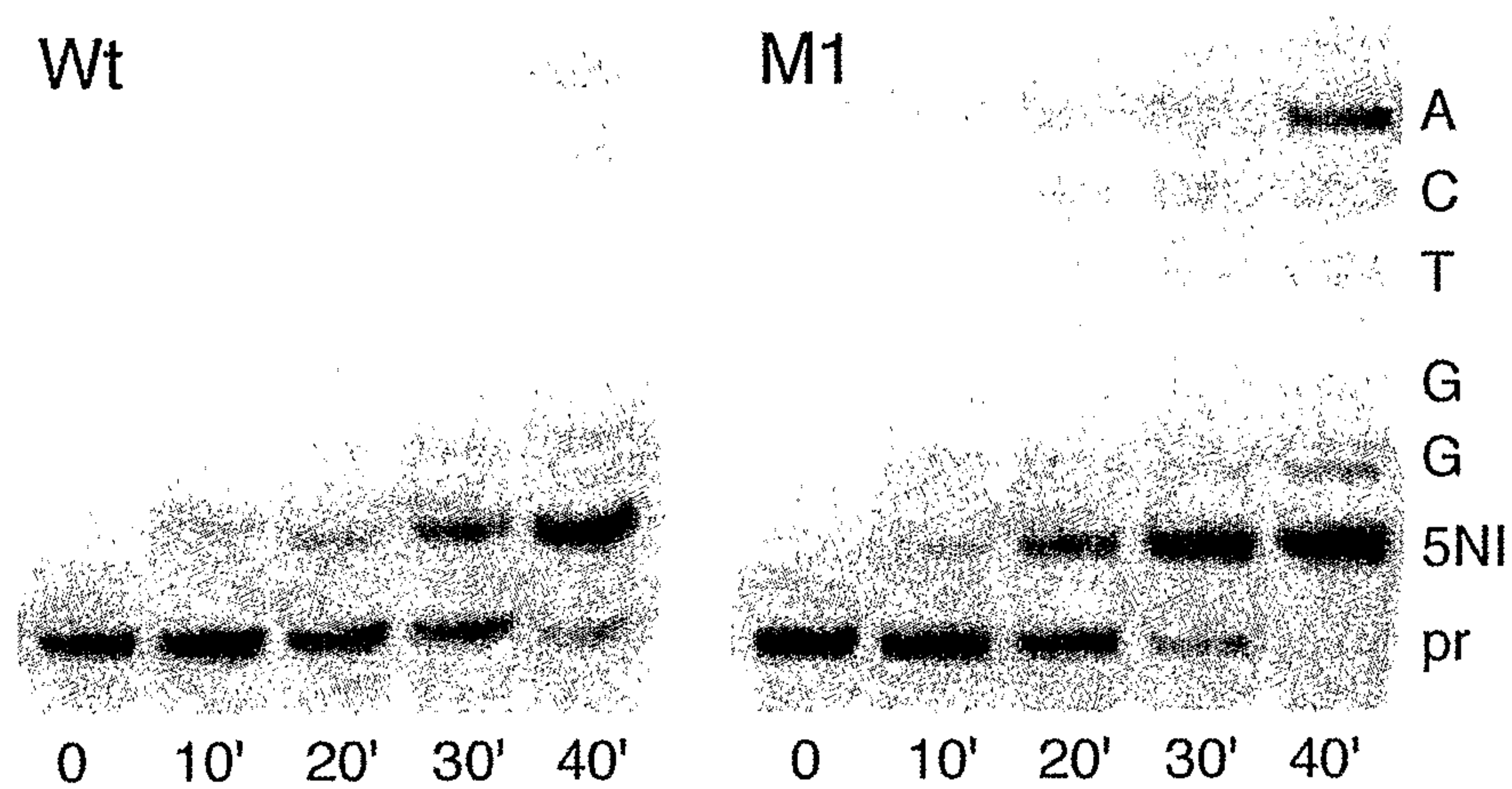
FIG. 6a



Polymerase activity on unnatural substrates.

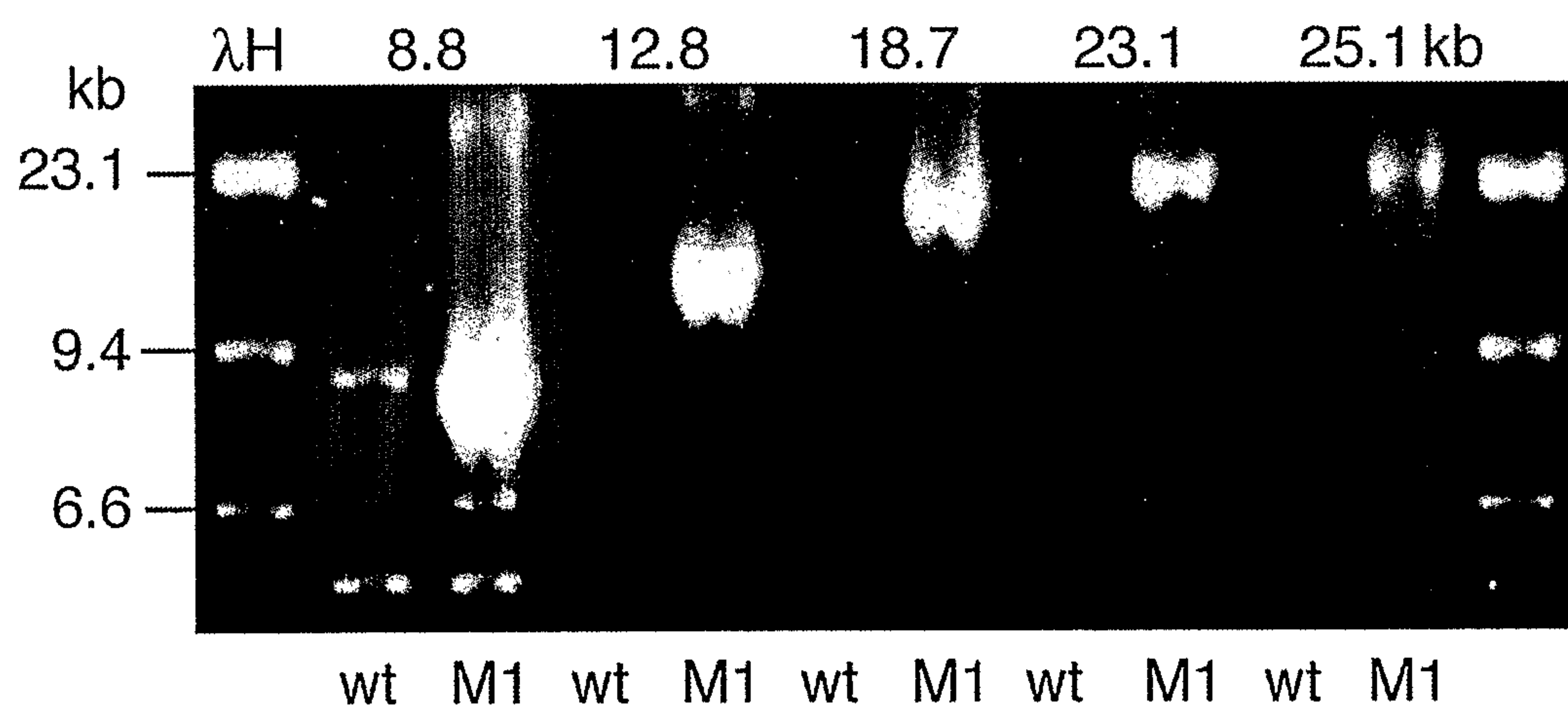
FIG. 6b

10 / 29



Polymerase activity on unnatural substrates (5NI)

FIG. 6c



Long range PCR.

FIG. 7

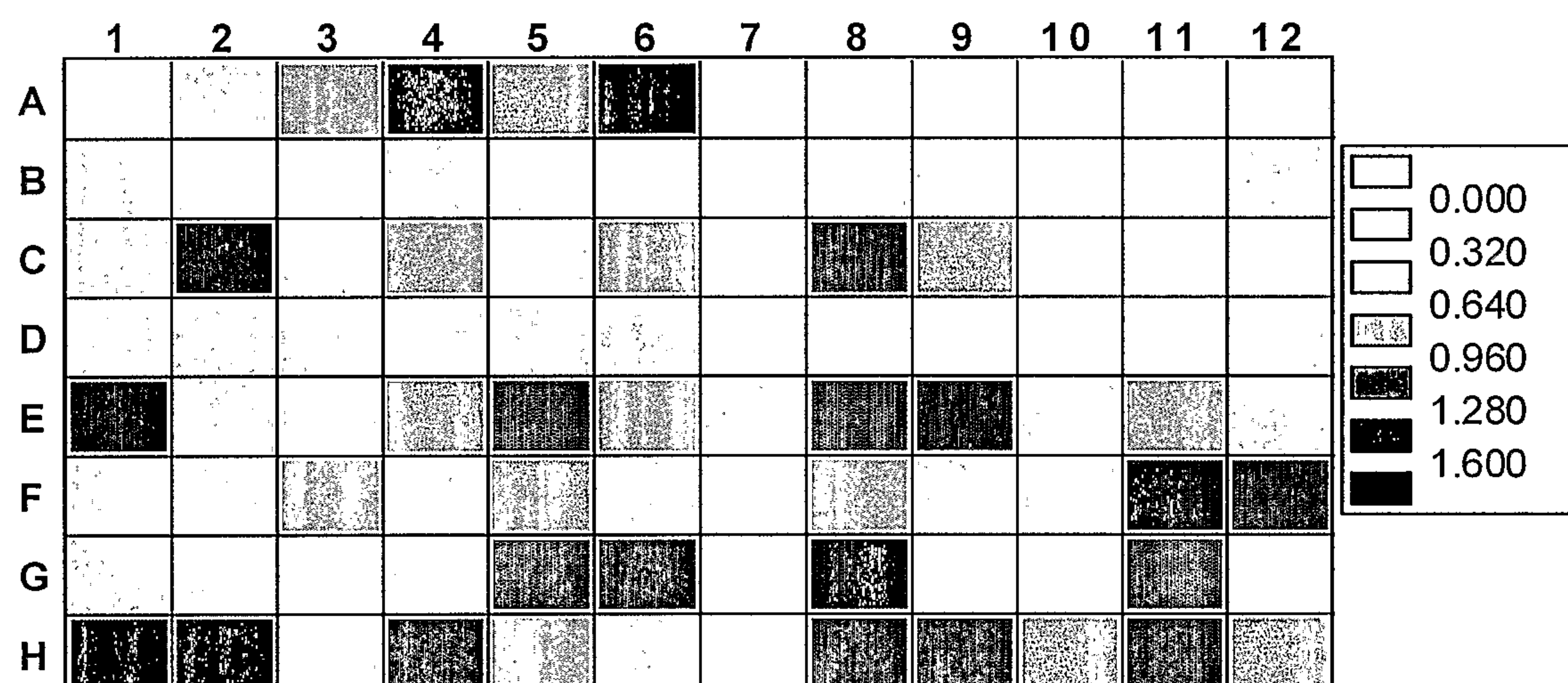
11 / 29

Experiment I.

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

(5'-

TAGCTACCATTTTCGCCGGCTTCCGTCGCGACCACGTP1TTCGTGGTCGCGA
CGGAAGCCG-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

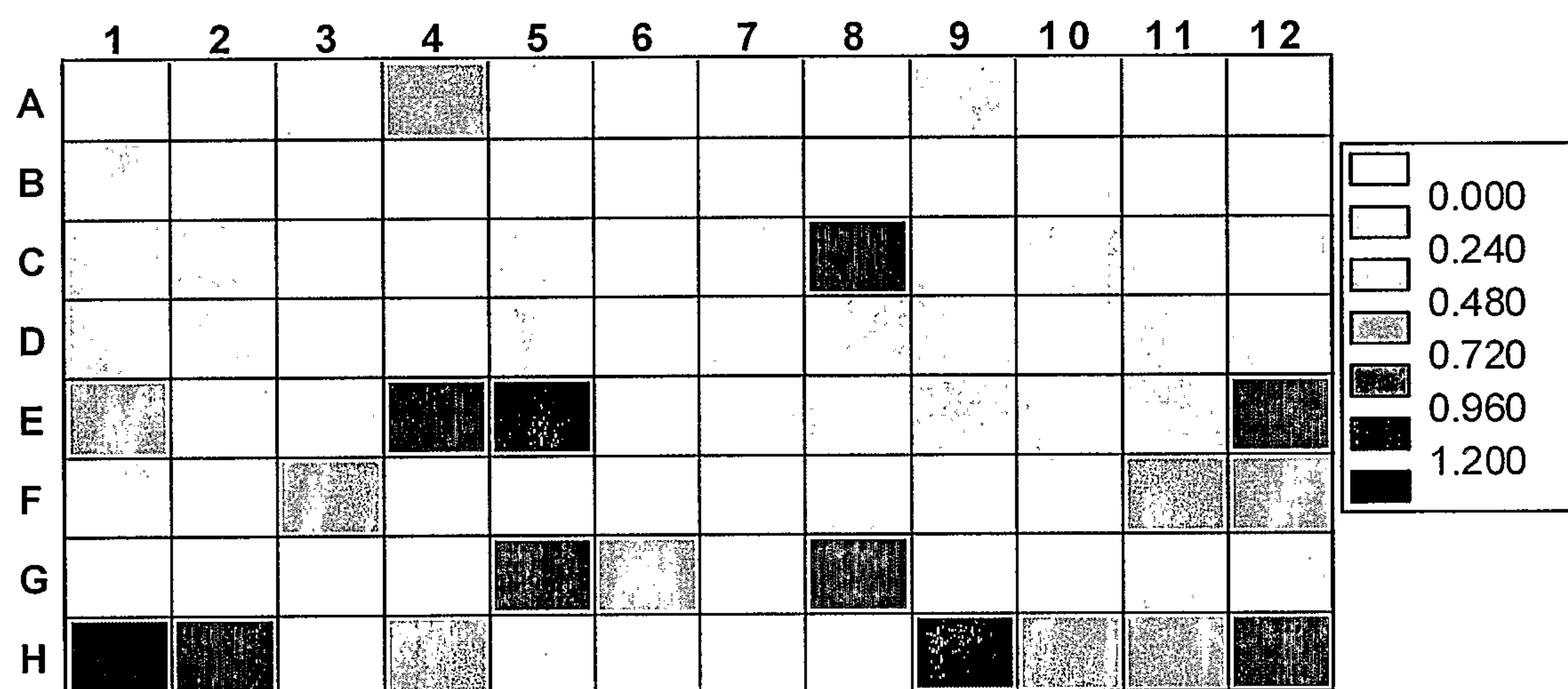
12 / 29

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'-

TAGCTACCATTTTTTTTTTCGCCGGCTTCCGTCGCGACCACGTTTTCGTGGTCGCGACGGAA
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

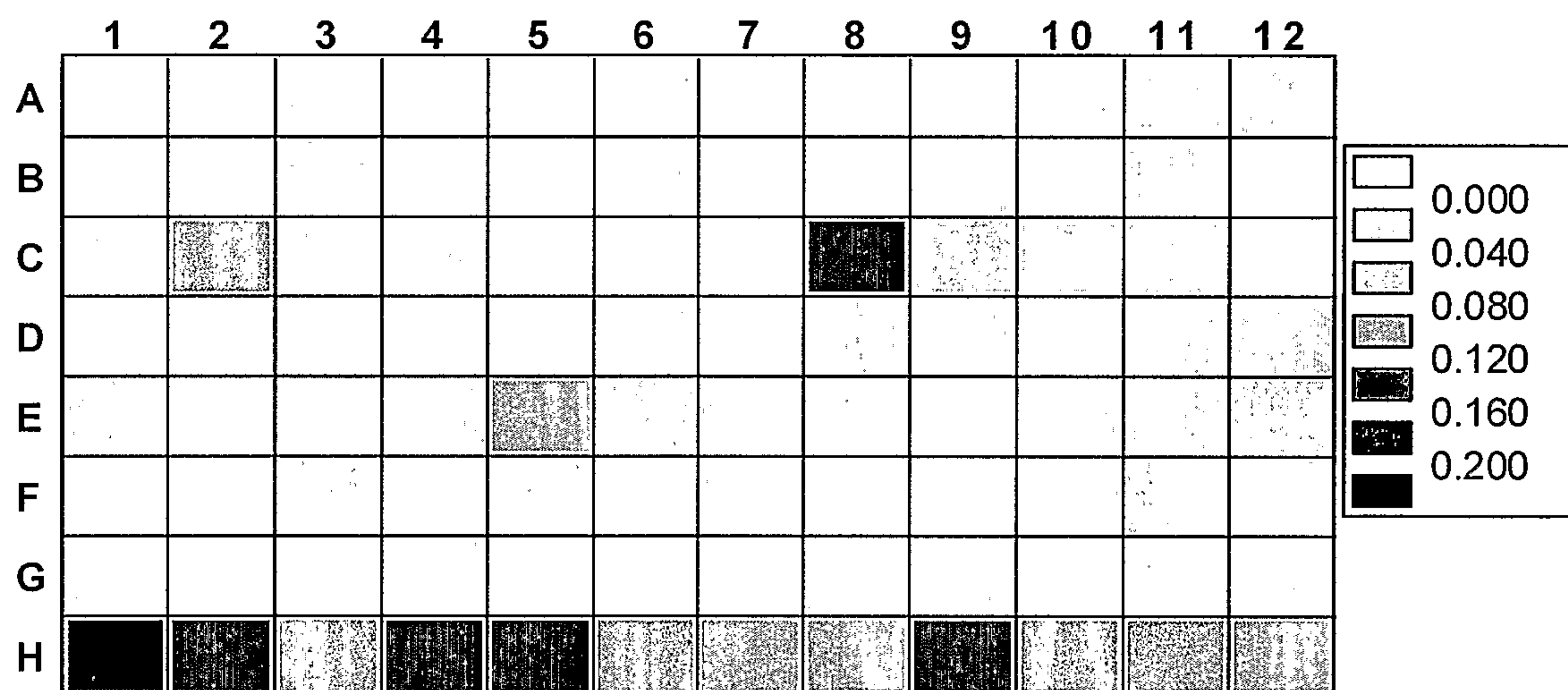
13 / 29

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

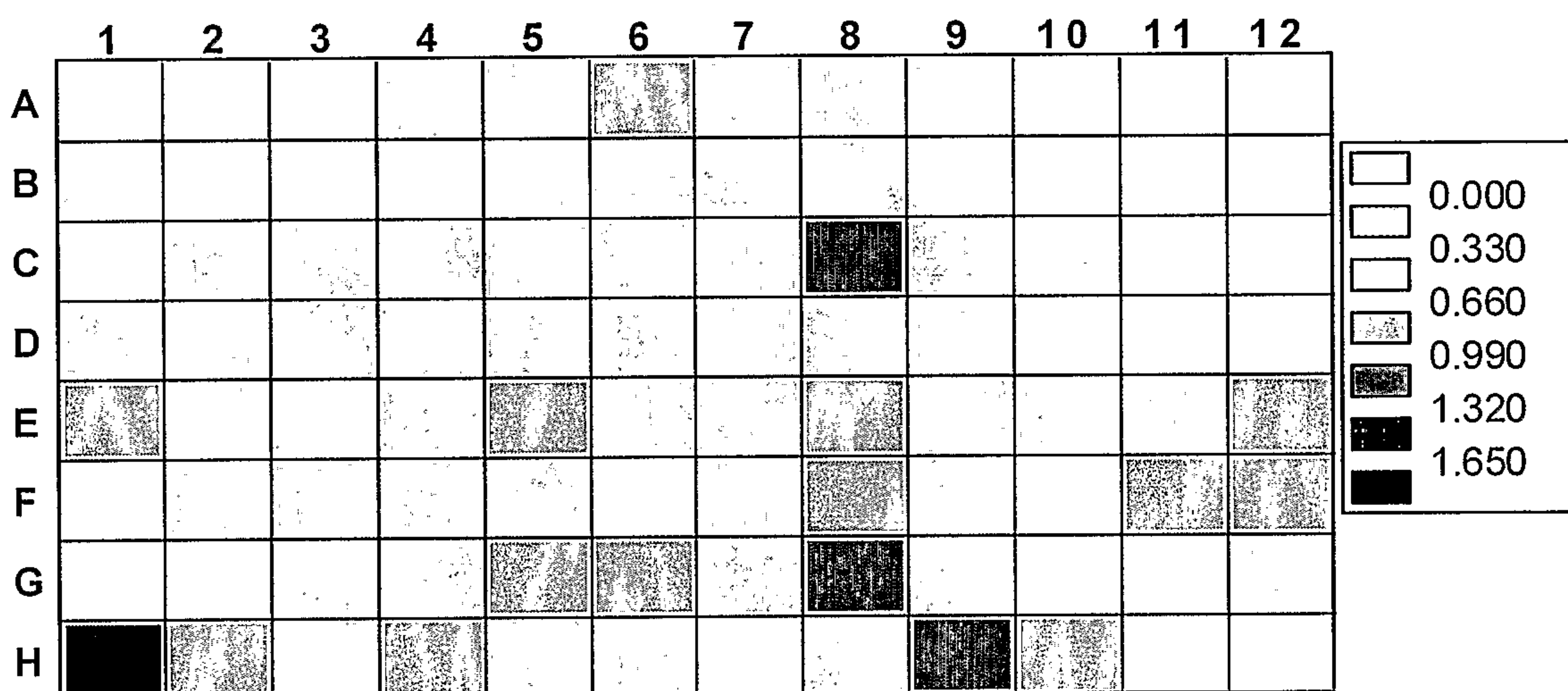
14 / 29

Experiment IV.

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

(5'-

TAGCTCGGTAAACGCCGGCTTCCGTCGCGACCACGTTT5TTCGTGGTCGCGACGGAAGCCG
--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

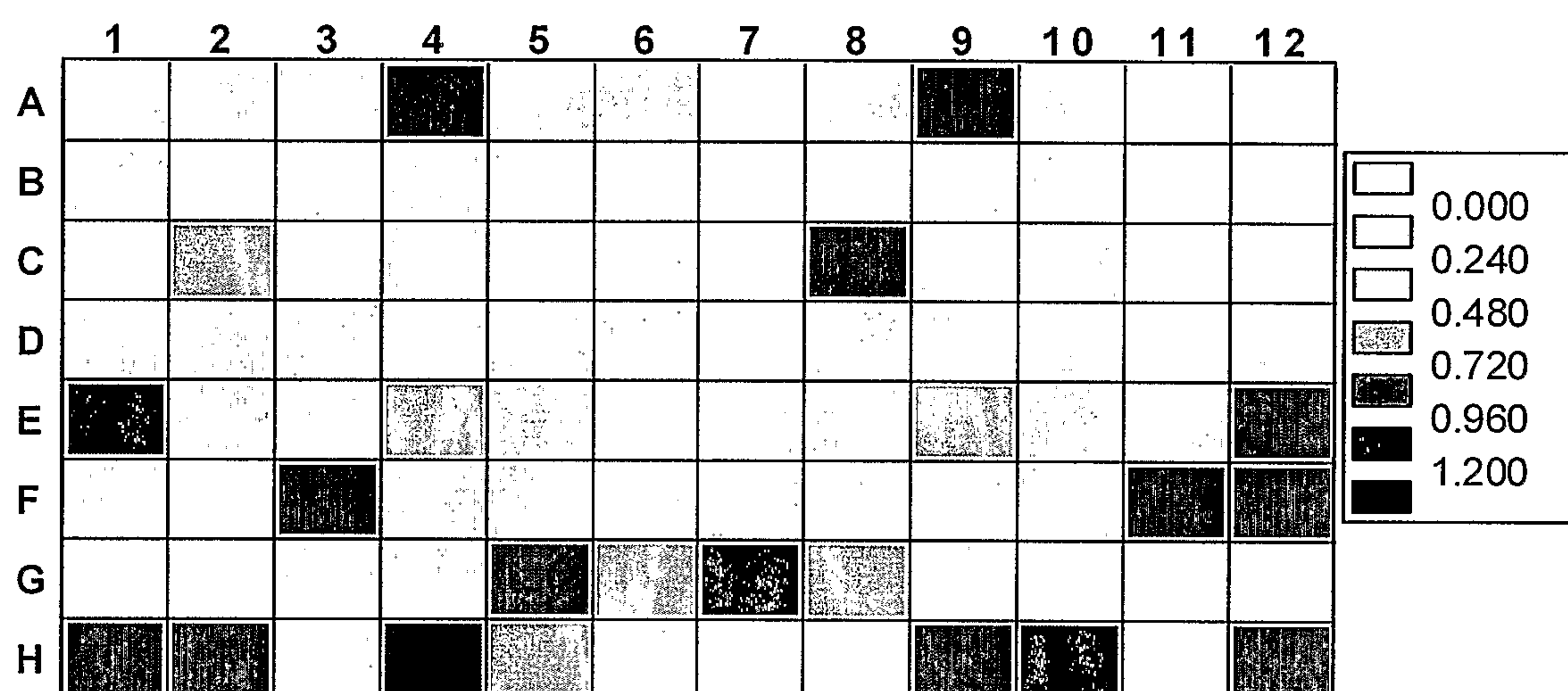
15 / 29

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)



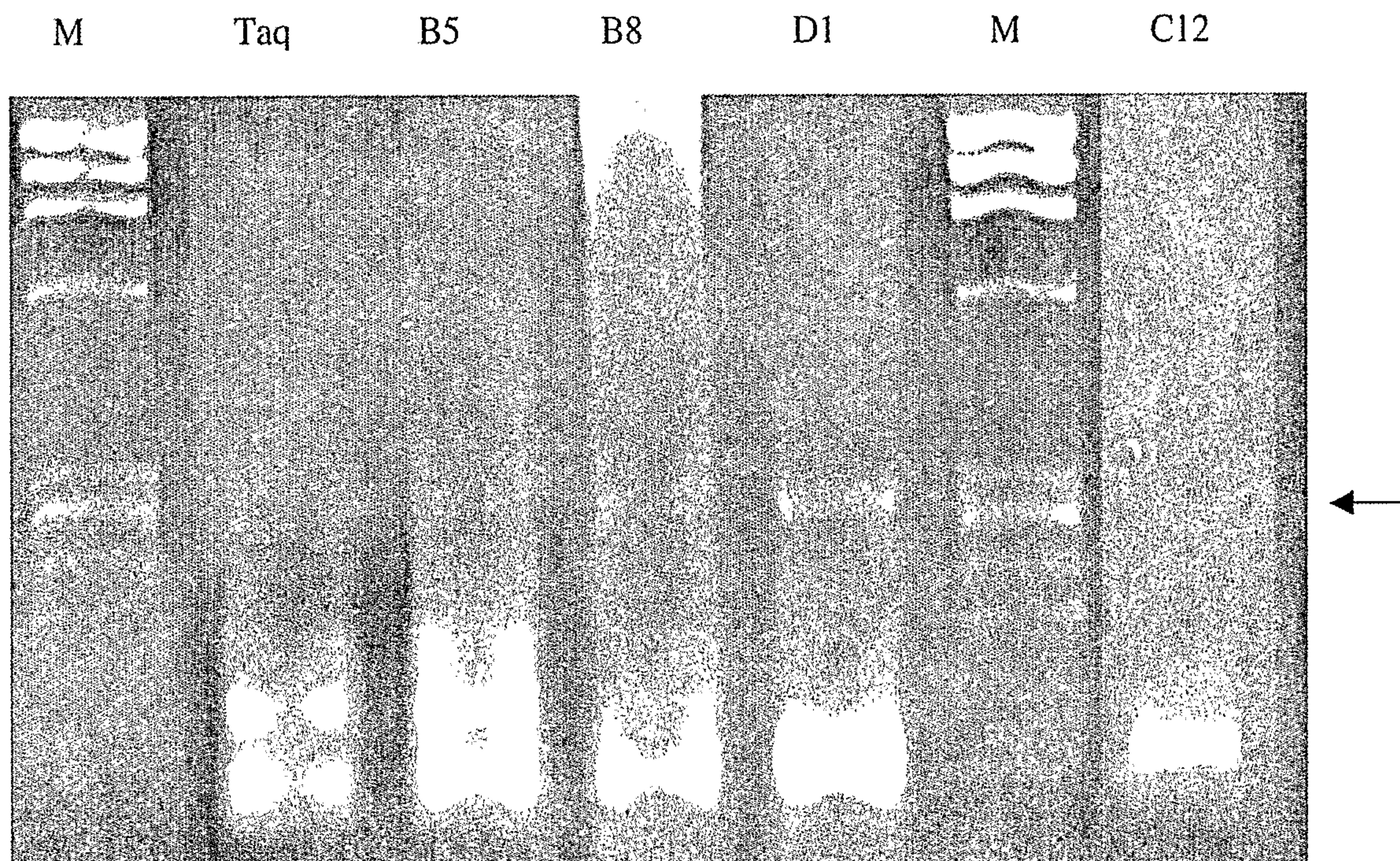
High Limit = 1.200

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

16 / 29



Four Mismatch Extension in PCR.

FIG. 9

17 / 29

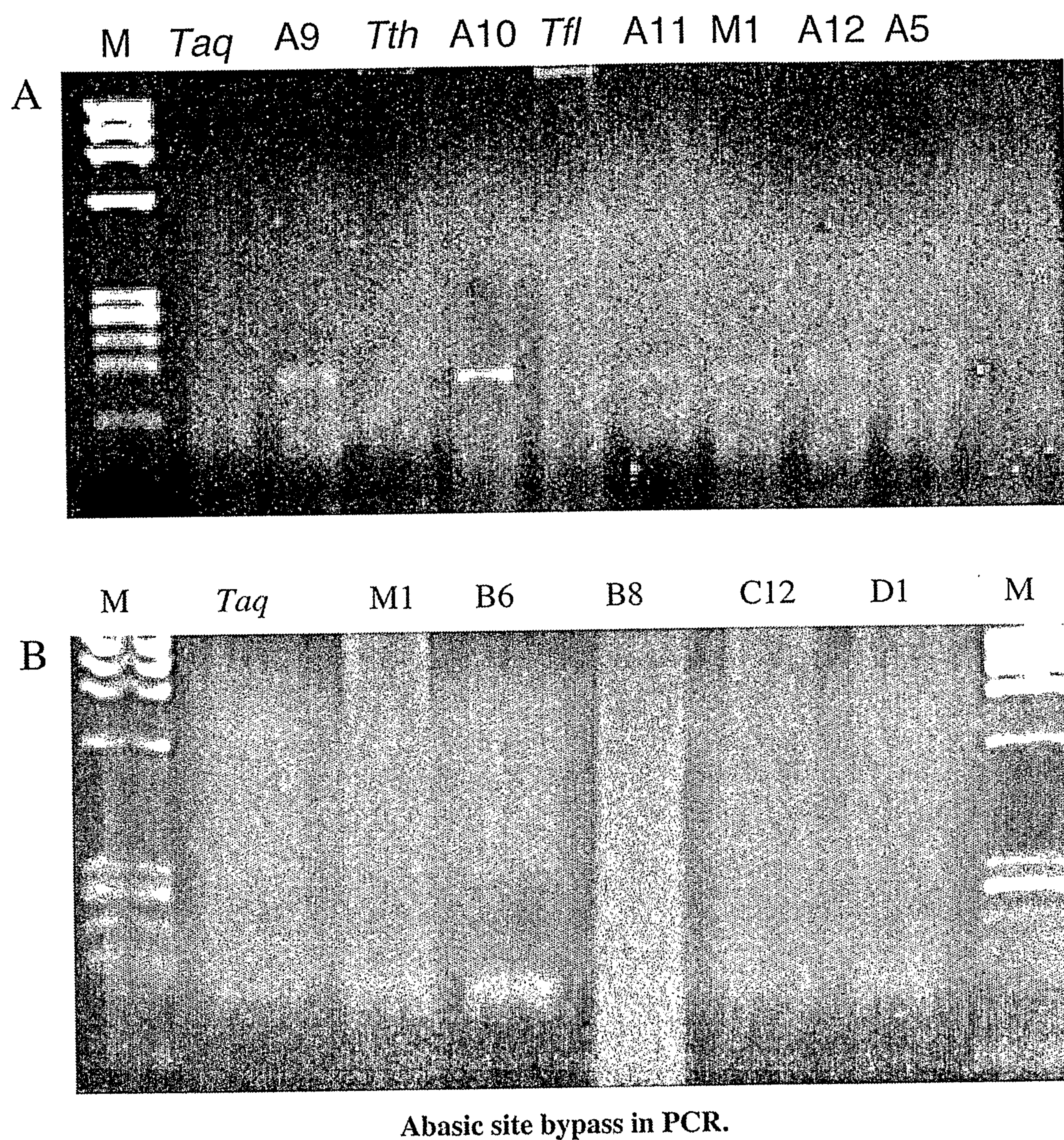


FIG. 10

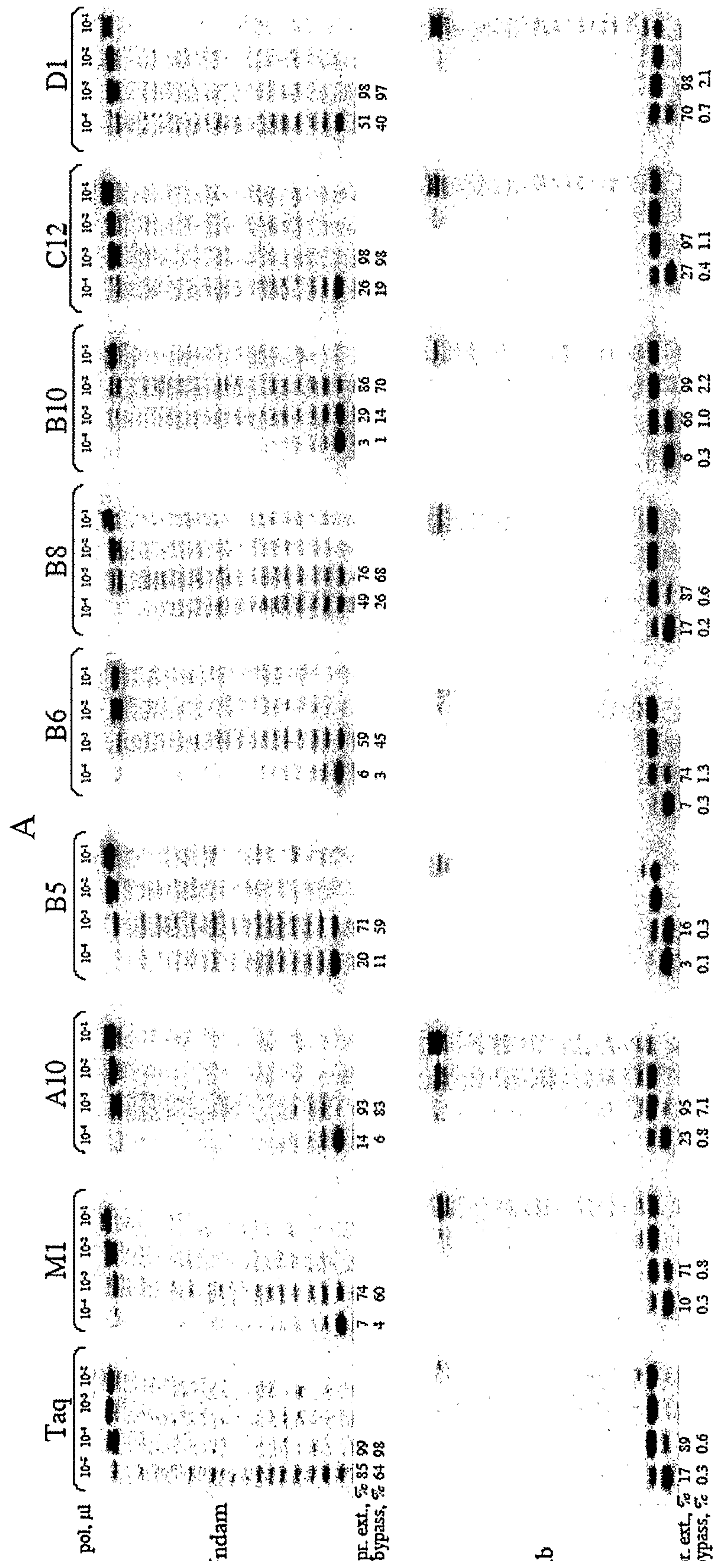
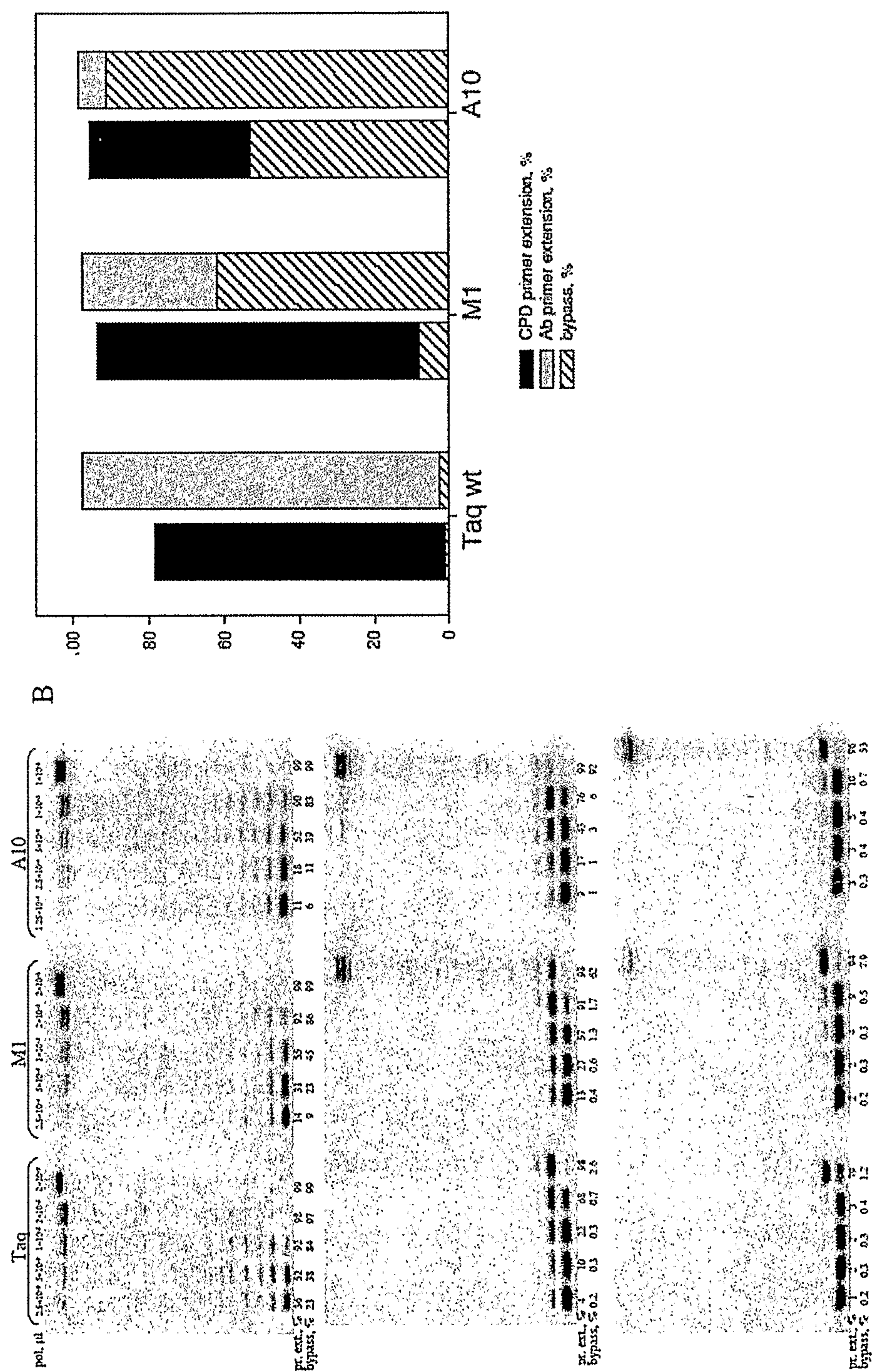
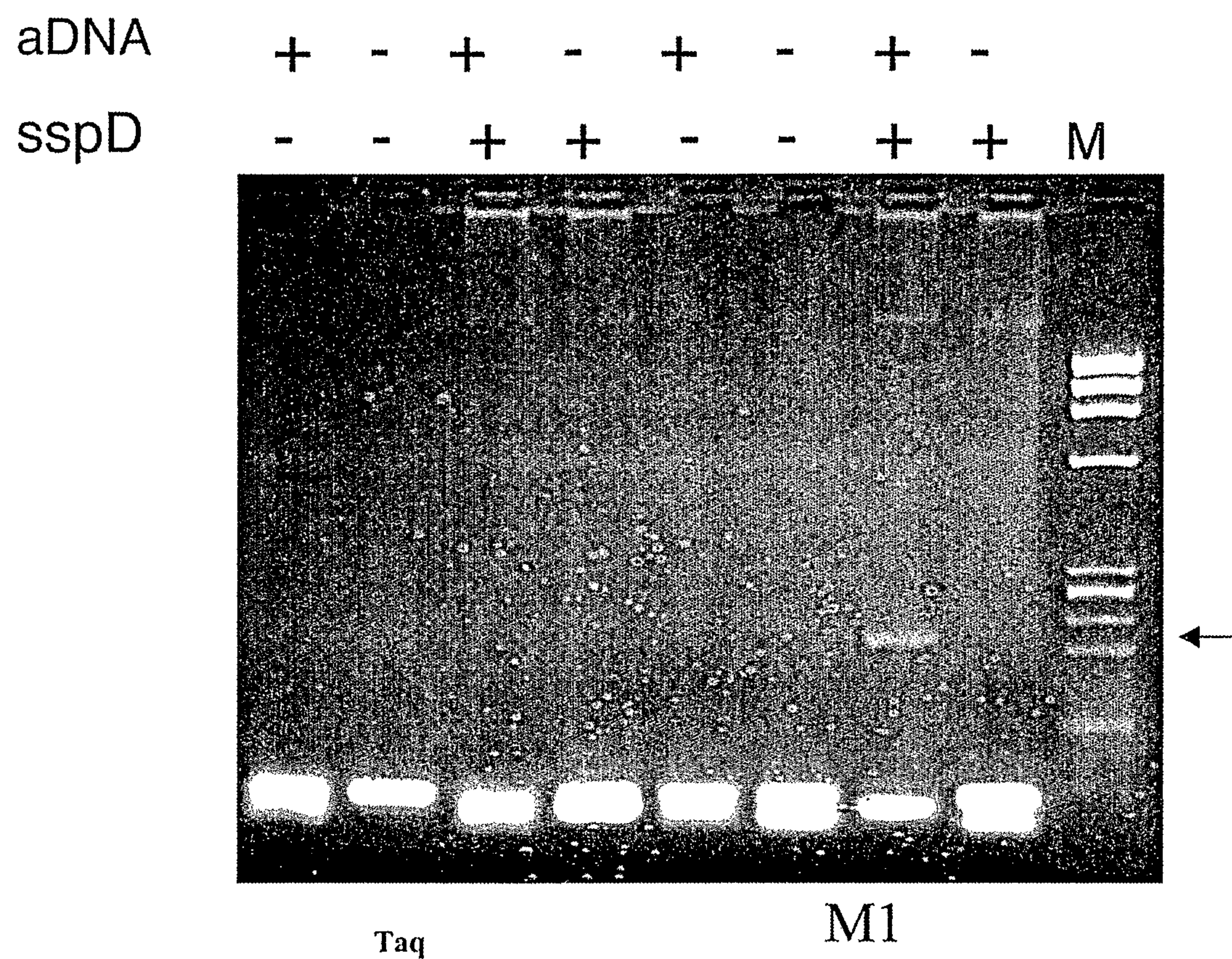


FIG. 11



20 / 29



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

FIG. 12

21 / 29

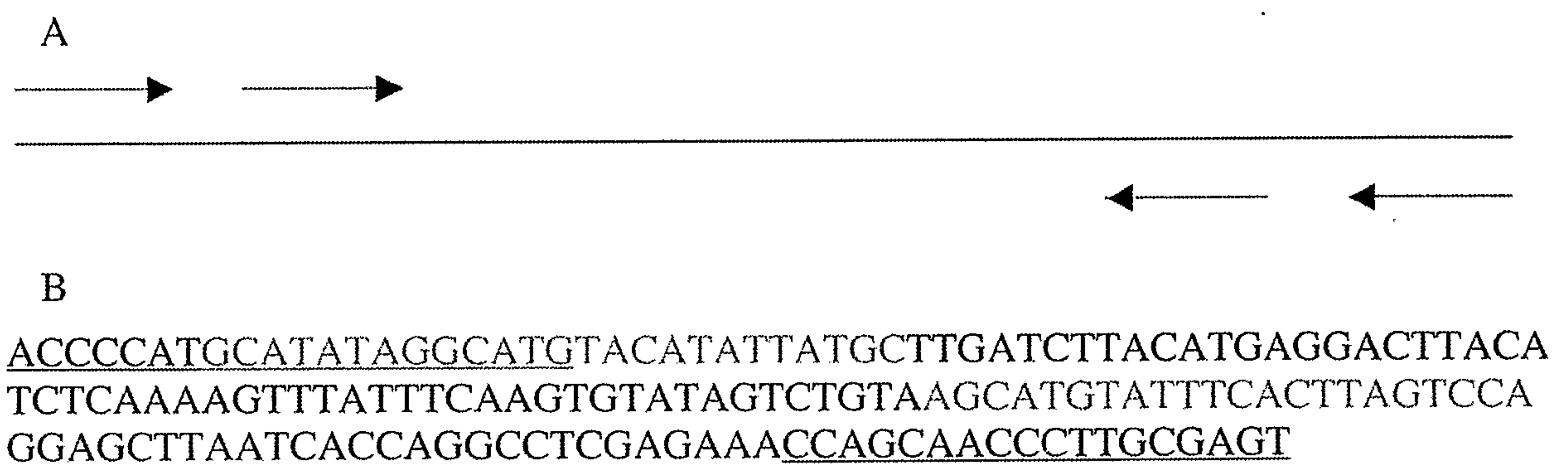


FIG. 13

22 / 29

Experiment	Sample	Dilution	SuperTaq		Mutant blend		Improvement
			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

23 / 29

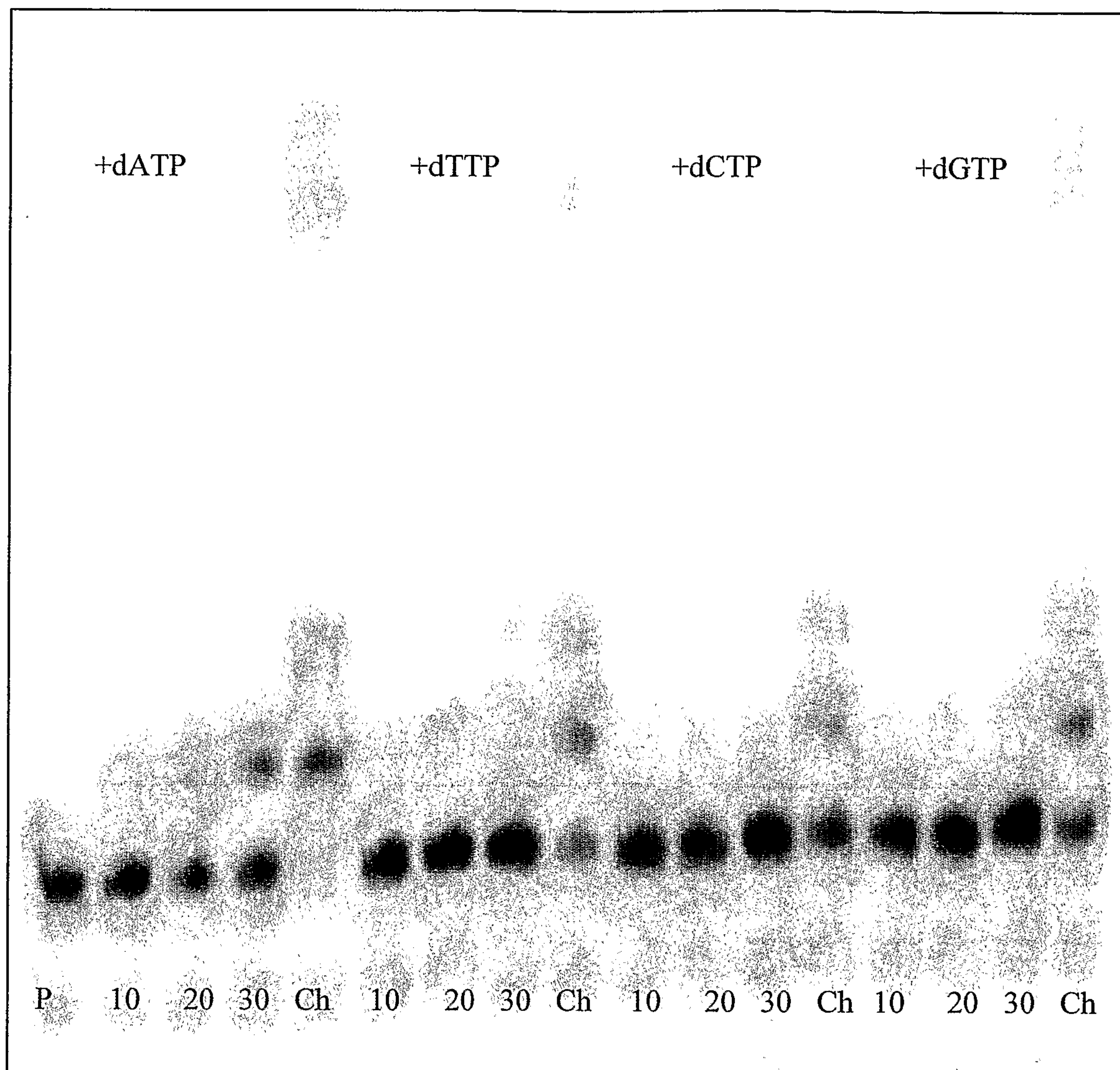
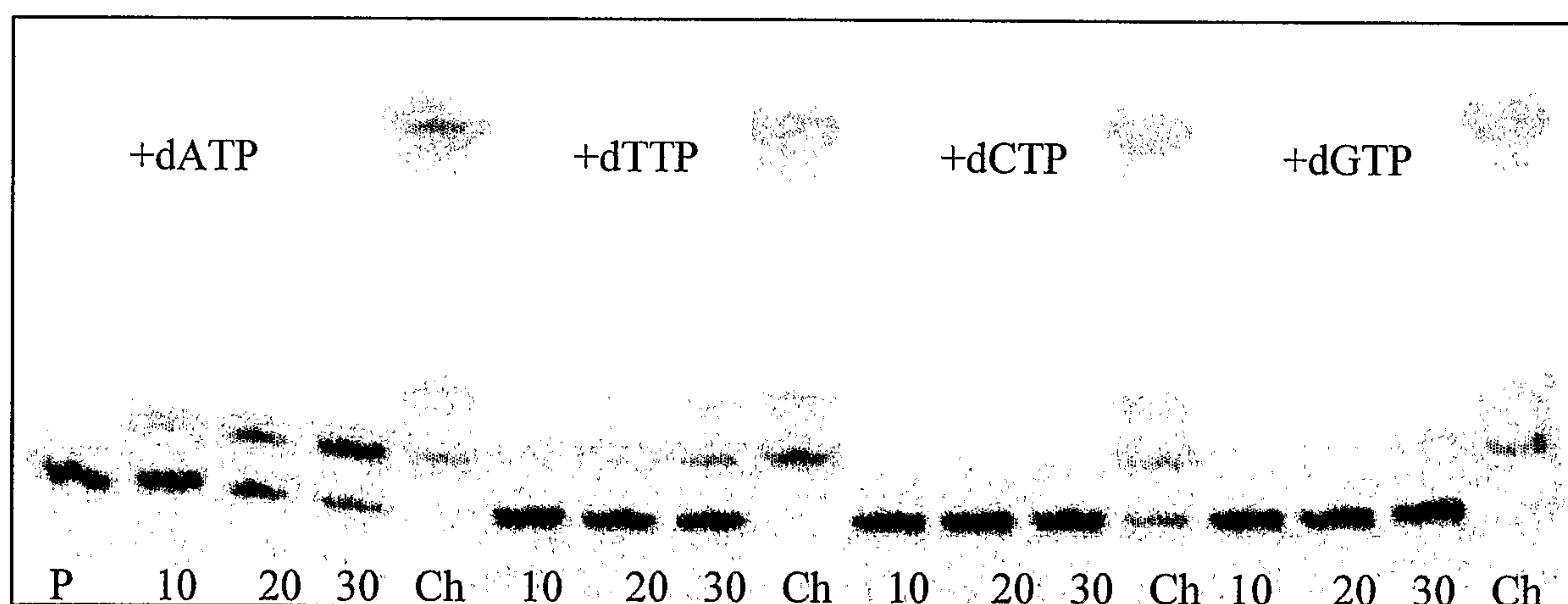


FIG. 15



24 / 29

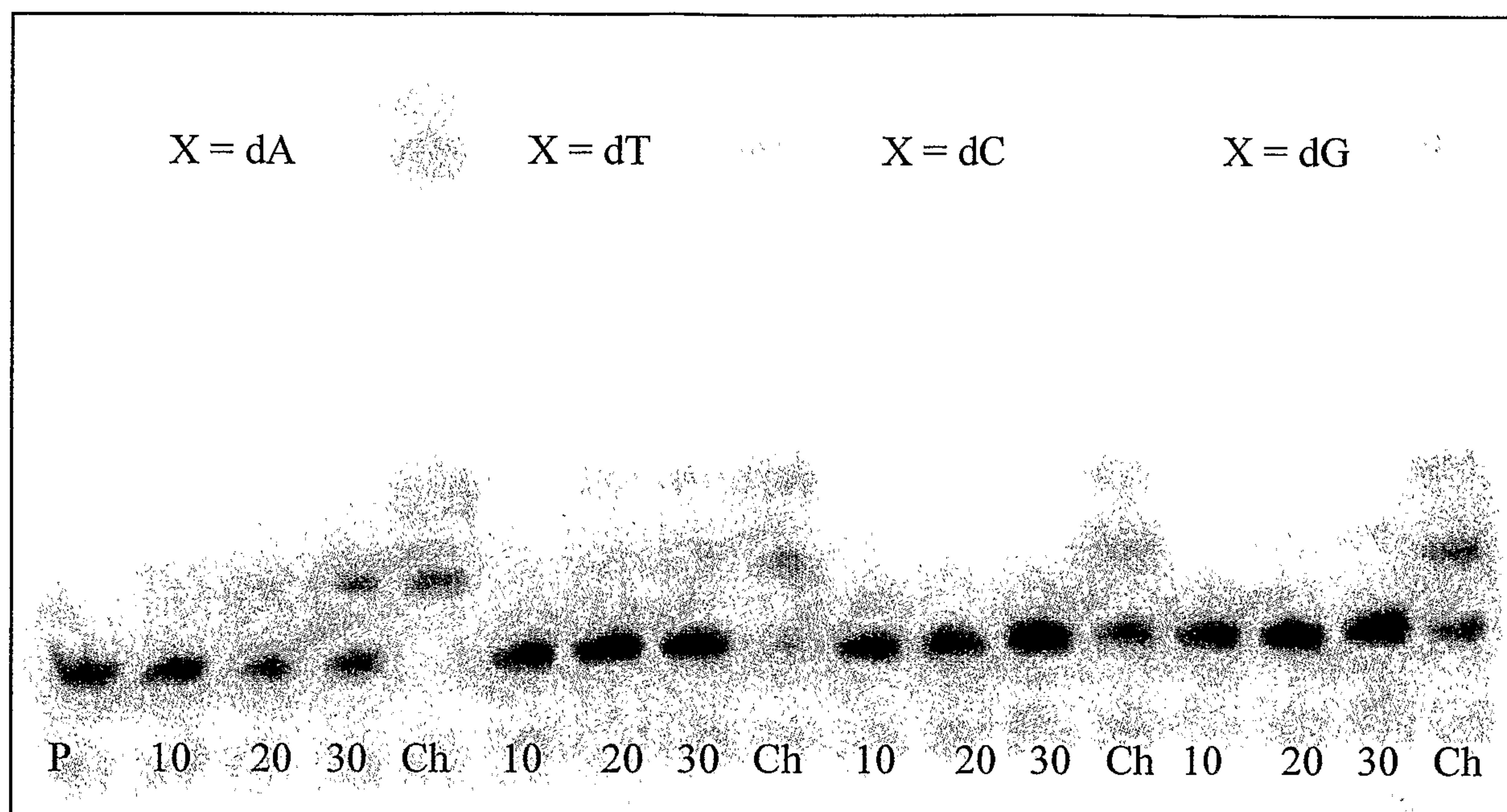


FIG. 17

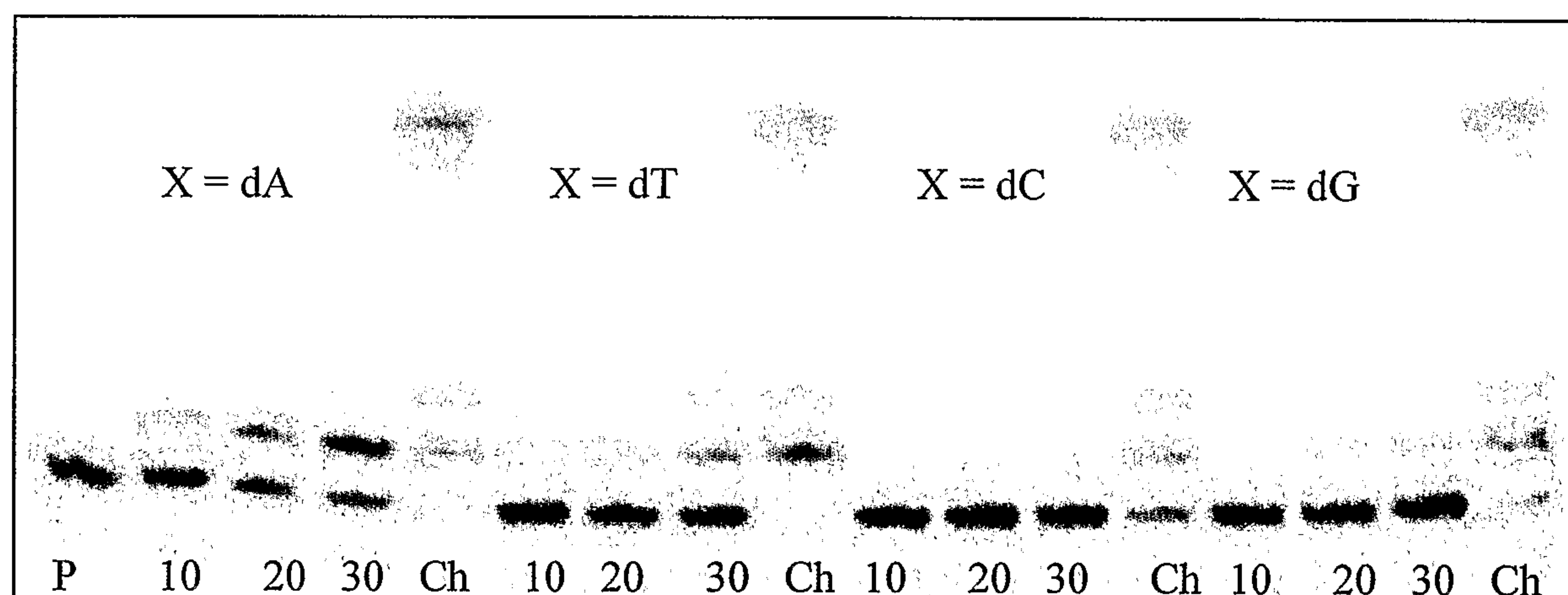


FIG. 18

25 / 29

Microarray hybridisations of FITC-labeled probes

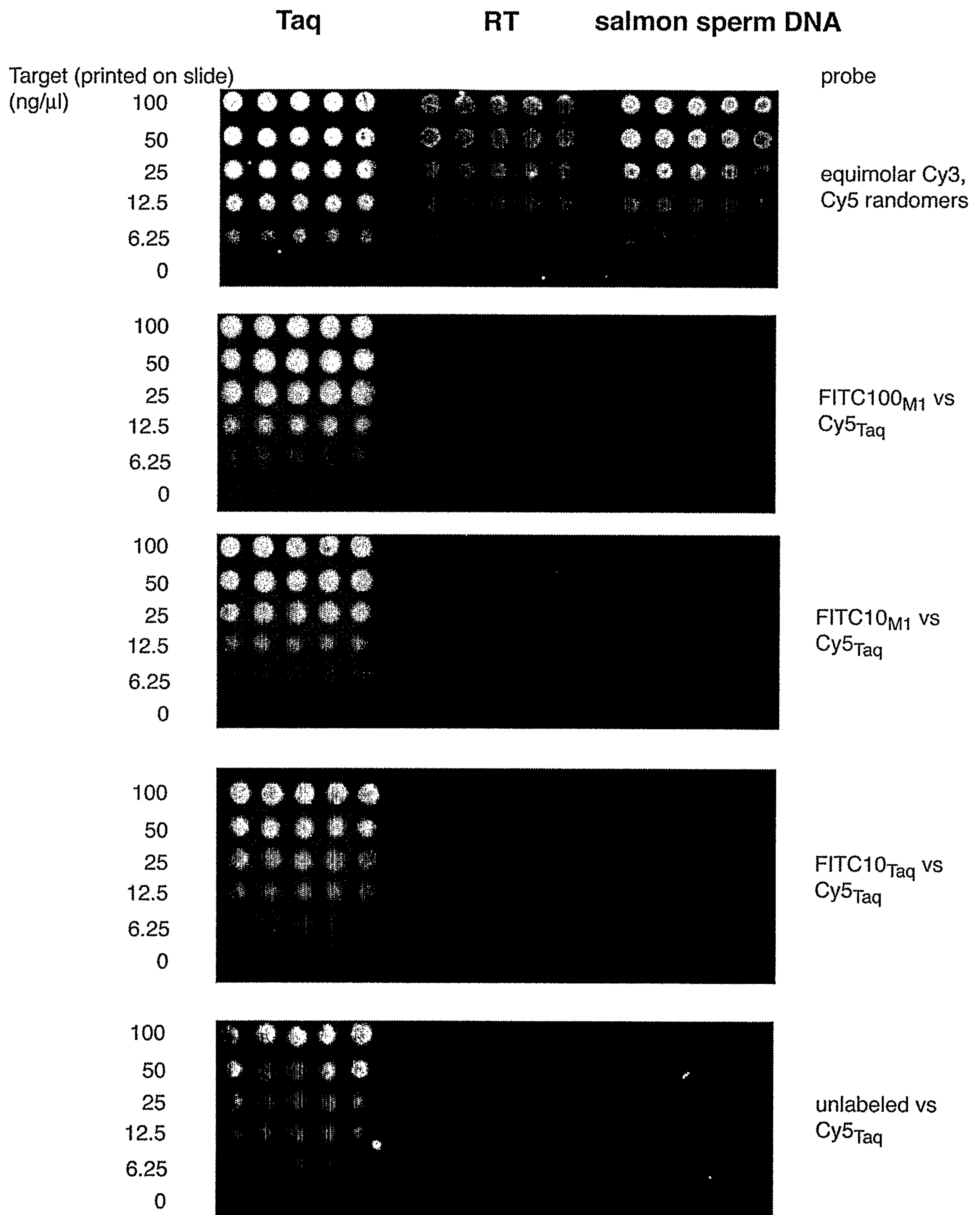
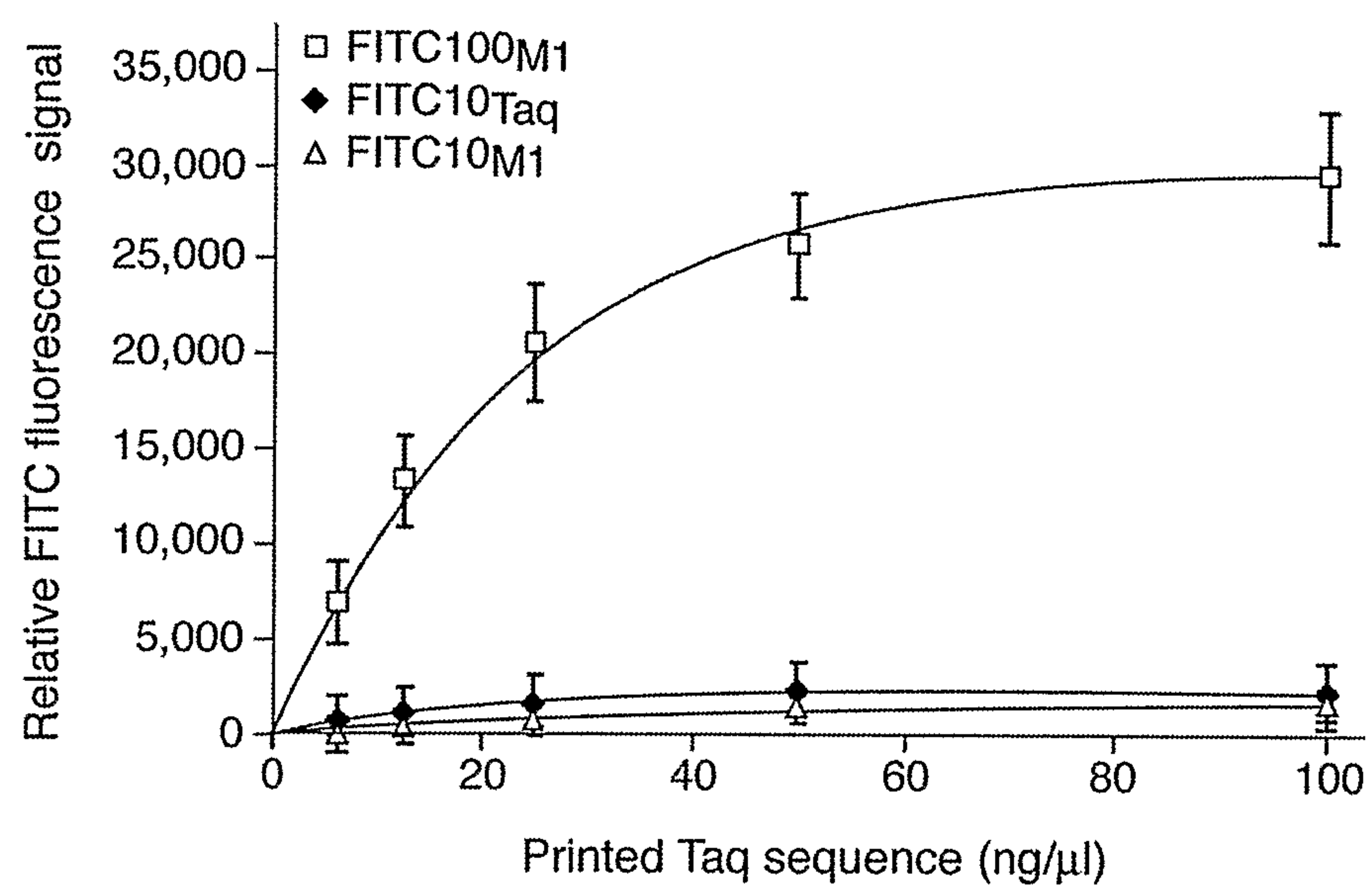


FIG. 19

26 / 29



Microarray analysis of FITC-labeled probes.

FIG. 20

27 / 29

Microarray signals from FITC-labelled probes.

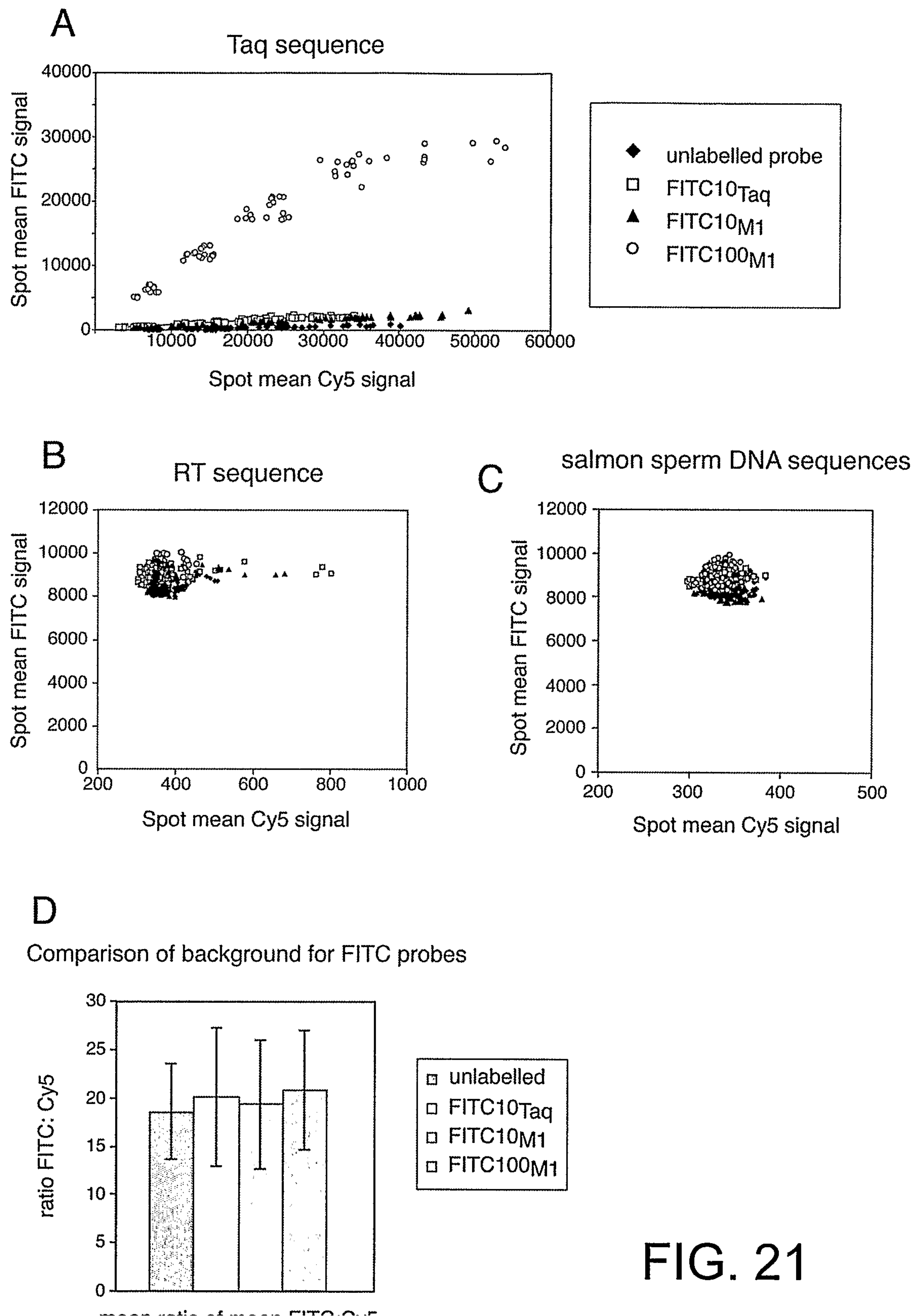
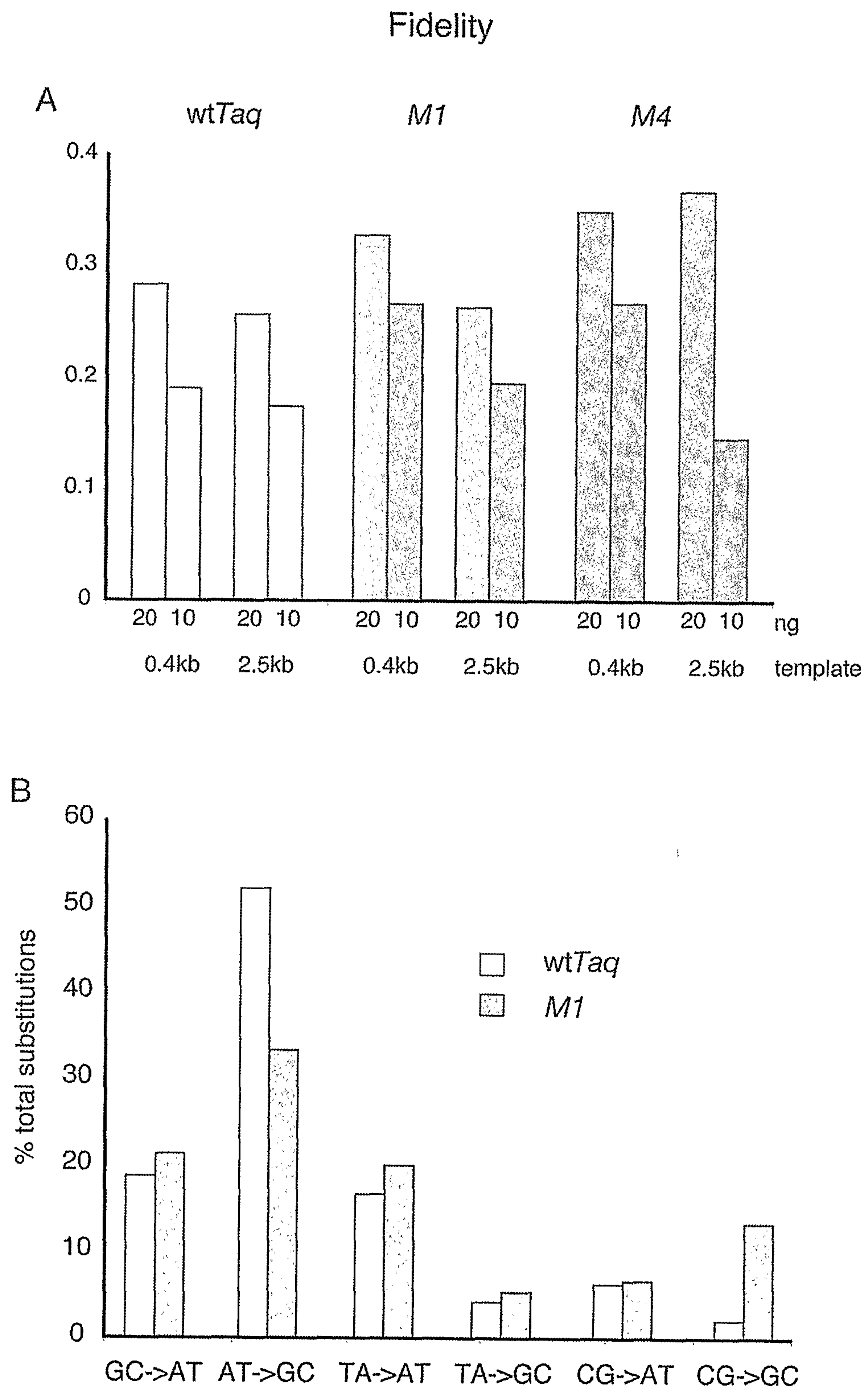


FIG. 21

28 / 29

**FIG. 22**

29 / 29

Processivity

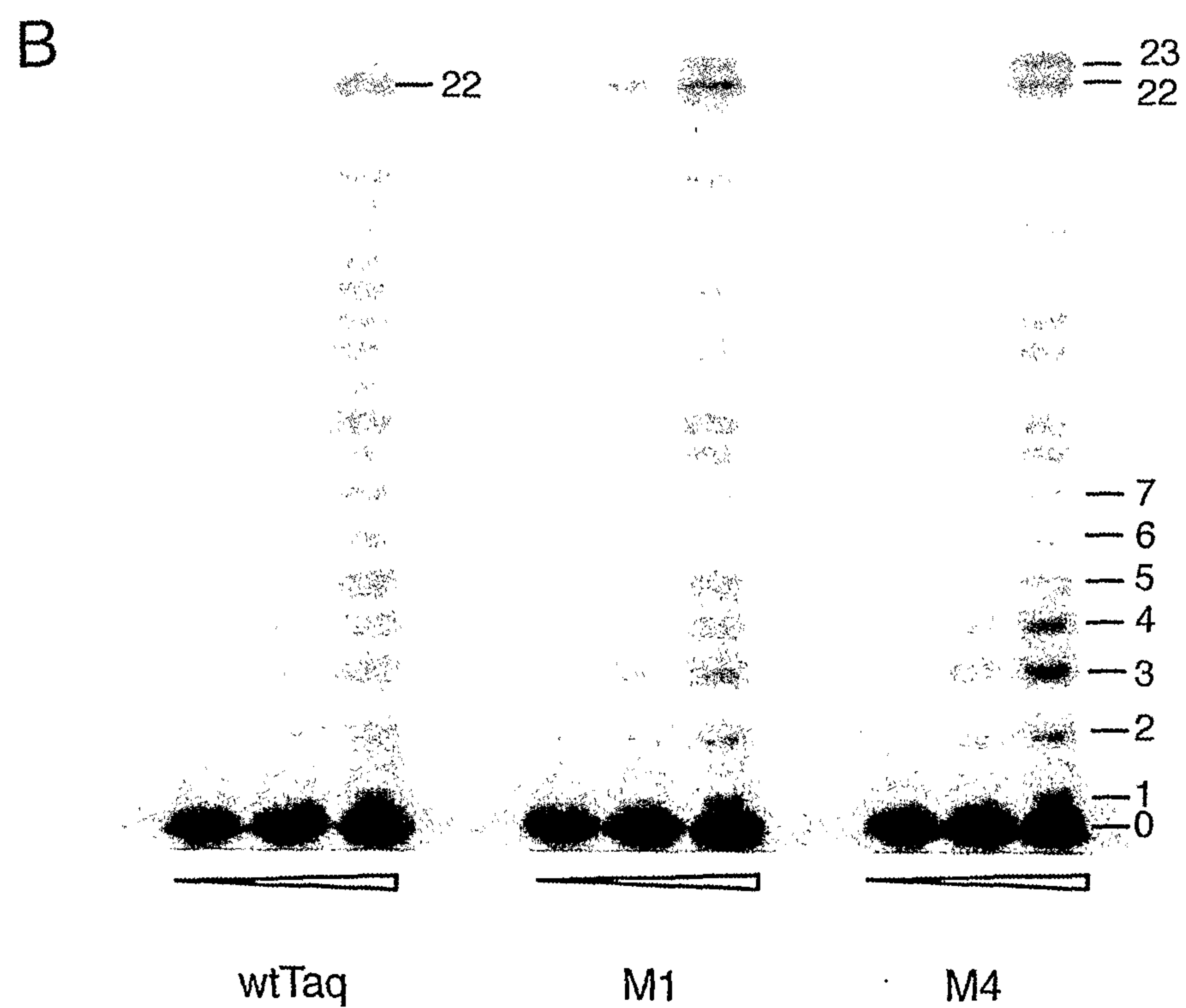
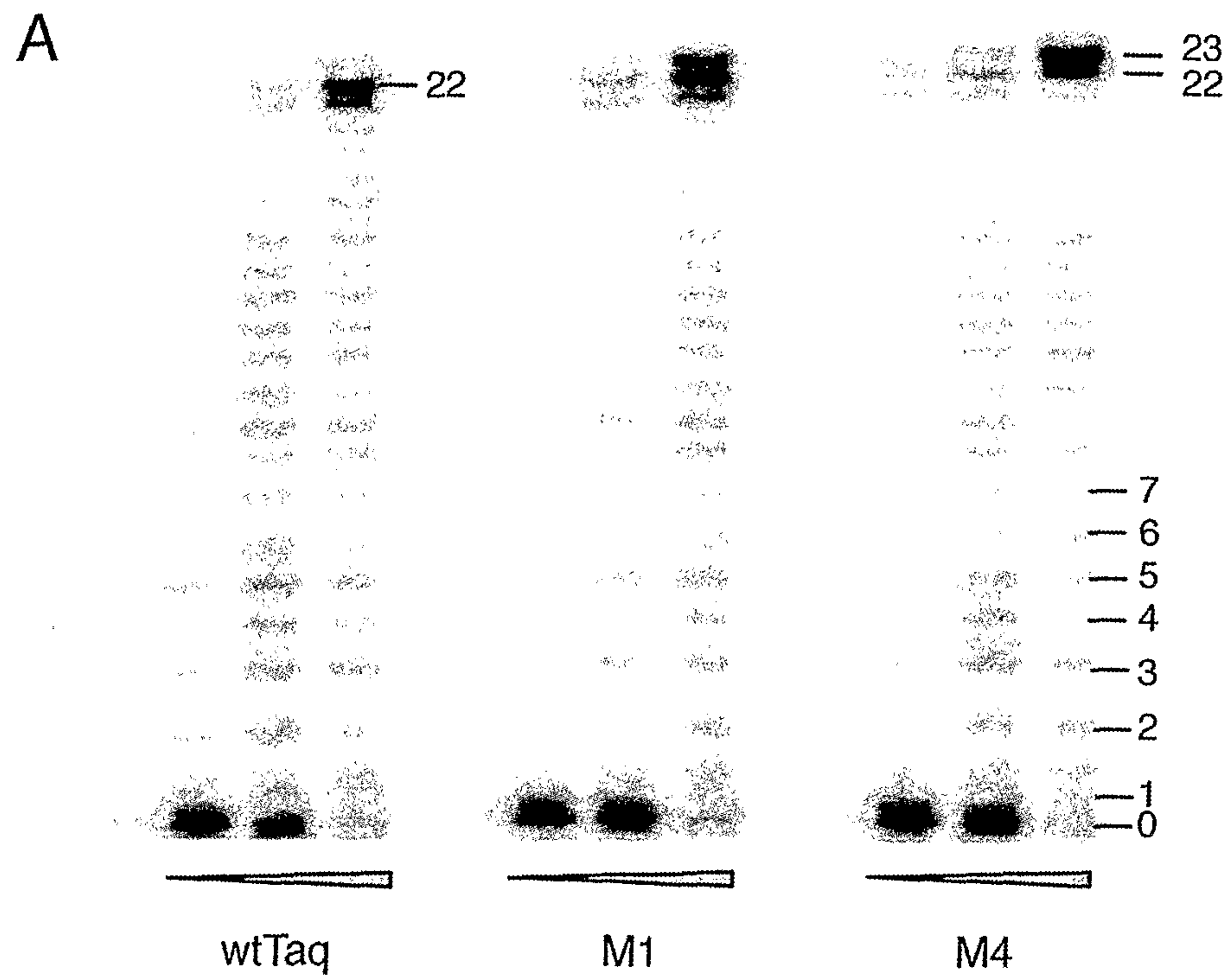


FIG. 29